Computer & Mathematical Sciences
UNIVERSITY OF TORONTO
SCARBOROUGH

STAB22: Final Exam April 19, 2018

Complete this form, then *carefully* tear off this page and place it along with your student card on the surface beside where you are writing.

The form will be verified and collected during the test.

Information collected will be used solely for the purpose of administering this test.

Name (please print): ____

Signature: ___________________________ Seat #: _______________
(if available)

Student Number: ☐☐☐☐☐☐☐☐☐☐

UNIVERSITY OF TORONTO SCARBOROUGH
Department of Computer and Mathematical Sciences
April 19, 2018
APRIL 2018 FINAL EXAMINATIONS

STAB22H3 Statistics I
Duration: 3 hours

Last Name:________________________ First Name: ________________________

Student number: ________________________

Aids allowed:

- Two handwritten letter-sized sheets (both sides) of notes prepared by you
- Non-programmable, non-communicating calculator

The table of normal and binomial distributions are attached at the end.

This test is based on multiple-choice questions. There are 50 questions. All questions carry equal weight. Report all your answers on the Scantron answer sheet before the end of the exam. On the Scantron answer sheet, ensure that you write down and **BUBBLE** your last name, first name (as much of it as fits), and student number (in "Identification").

Mark in each case the best answer out of the alternatives given (**which means the numerically closest answer if the answer is a number and the answer you obtained is not given.**)

Also before you begin, complete the signature sheet, but sign it only when the invigilator collects it. The signature sheet shows that you were present at the exam.

There are 30 pages including this page and statistical tables. Please check to see that you have all the pages.

Good luck!!

ExamVersion: $\boxed{\text{A}}$

1. If there is a very strong linear relationship between two variables then what do you know **for sure** about the correlation coefficient r?

   (a) The correlation coefficient is larger than 1

   (b) The correlation coefficient is close to 0

   (c) The correlation coefficient is close to -1

   (d) The correlation coefficient is close to 1

   [✓ (e) None of the above is known for sure.]

   Solution:
   (e) We know for sure that the correlation coefficient is close to -1 or 1.

2. A box contains 18 balls, some of them are black and the others are white. We do not know how many are black and how many are white. Let n be the number of black balls and so $18 - n$ is the number of white balls. We are going to pick four balls at random from this box, one by one, without replacement (i.e when a ball is picked, we do not put it back in the box). Let A denote the event that the first three balls selected are black and B denote the event that the fourth ball selected is black. If $P(B|A) = 1/5$, what is the value of n?

   [✓ (a) 6]

   (b) 7

   (c) 8

   (d) 9

   (e) 12

$P(B/A) = \frac{P(A \cap B)}{P(A)}$

A = 1st 3 balls black
B = 4th ball is black

$P(B/A) = \frac{n-3}{18-3} = \frac{1}{5} \rightarrow n-3 = 3$
$n = 6$

   Solution:
   a
   After taking out the first 3 balls there are 15 left. $P(B|A) = 1/5$ means the probability of selecting a black ball from these remaining 15 balls is $1/5$. This means there must be 3 $(15 \times \frac{1}{5})$ black balls in these 15 remaining balls. This means at the beginning there were $n = 3 + 3 = 6$ black balls.

   Or second method:
   $P(B|A) = \frac{n-3}{18-3} = 1/5 \implies n = 6$

3. A traditional die is a cube, each face showing an integer in the range from 1 to 6. A **fair** traditional die is rolled 3 times. The first 2 rolls result in the face showing “5” being uppermost. What is the probability the third roll will **not** result in the face showing “5” being uppermost?

   (a) 0

   (b) 0.023

   (c) 0.069

   (d) 0.833

   (e) 1

   Die rolling is independent

   $\underbrace{\frac{1}{6}, \frac{1}{6},}_{5 \text{ happening}}$

   5 not happening $= \frac{5}{6}$

   Solution:
   (d) The rolls are independent. The probability the third roll will not result in the face with “5” being uppermost is $1 - 1/6 = 5/6 = 0.833$.

4. The tail length for a population of female hook-billed kites is Normally distributed with mean 194 mm and standard deviation 11 mm. We select a random sample of 200 kites from this population. Let X be the number of kites in this sample having a tail length between 179 and 209 mm. What is the mean of X?

   (a) 145

   (b) 155

   (c) 165

   (d) 175

   (e) 194

   $n = 200$

   $p = P(179 < L < 209) \rightarrow P(\frac{179-194}{11} < \frac{209-194}{11})$

   $= P(-1.36Z < 1.36)$

   $1 - 2(0.0869) = 1 - 0.1738 = 0.8262$

   $N_X = np = 200 \times 0.8262 = 165.24$

   Solution:
   c
   Letting L denote the tail length of a randomly selected female hook-billed kite, $X \sim Bin(n = 300, p = P(179 < L < 209))$.
   $p = P(179 < L < 209) = P(\frac{179-194}{11} < Z < \frac{209-194}{11}) = P(-1.36Z < 1.36) = 1 - 2 \times 0.0869 = 1 - 0.1738 = 0.8262$
   $\mu_X = np = 200 \times 0.8262 = 165.24$

5. The probability that a patient with a certain disease will be successfully treated with a new medical treatment is 0.80. The success of the treatment is independent from one patient to another. Suppose that the treatment is used on 40 patients with the disease. What is the mean of the number of patients who will **not** be successfully treated?

   [✓ (a) 8]
   (b) 20
   (c) 32
   (d) 40
   (e) Impossible to say given the information provided

   $0.8 \times 40 = 32$

   $40 - 32 = 8$

   Solution:
   (a)
   The mean of the number of patients who will **not** be successfully treated is $n(1-p) = 40 \cdot 0.2 = 8$

6. STA101 and STA102 are two statistics courses both with large classes. We have the following information about the distributions of grades for these two courses:

   - The distribution of grades of STA101 is right skewed. Mean > Median
   - The distribution of grades of STA102 is left skewed. Median > Mean
   - STA101 has lower class average than STA102 (i.e. $\mu_1 < \mu_2$, where μ_1 and μ_2 are the means of the distributions of grades in STA101 and STA102 respectively.)

   Which of the following statements regarding the distributions of grades of these two courses is **NOT** true?

   (a) The median grade of STA101 is less than the median grade of STA102. ✓

   (b) The mean grade of STA102 is less than the median grade of STA101.

   (c) The mean grade of STA101 is less than the median grade of STA102.

   (d) The median grade of STA101 is less than the mean grade of STA102.

   (e) The mean grade of STA102 is less than the median grade of STA102.

   Solution:
   b
   Denoting the medians of STA101 and STA102 by m_1 and m_2 respectively, we have,
   $m_1 < \mu_1$, since the distribution of grades of STA101 is right skewed,
   $\mu_2 < m_2$, since the distribution of grades of STA102 is left skewed,
   and so since we are also given that $\mu_1 < \mu_2$, we have
   $m_1 < \mu_1 < \mu_2 < m_2$ and this implies:

   $m_1 < m_2$, i.e. (a) is true
   $m_1 < \mu_2$, i.e. (b) is false
   $\mu_1 < m_2$, i.e. (c) is true
   $m_1 < \mu_2$, i.e. (d) is true
   $\mu_2 < m_2$, i.e. (e) is true

7. The following table represents the probability distribution of the number of accidents per day in a city.

Number of accidents	0	1	2	3	4 or more
Probability	0.55	0.20	0.10	0.15	0

Mean=0.85

Median= 0

Use this table for this question and the next one.
Which of the following statements about the number of accidents per day in the city is/are true?

(I) The mean number of accidents per day is larger than the median number of accidents per day.

(II) The mean number of accidents per day is smaller than the median number of accidents per day.

(III) The median number of accidents per day is equal to 1.

(a) Only (I) is true

(b) Only (II) is true

(c) Only (III) is true

(d) Only (I) and (II) are true

(e) Only (I) and (III) are true

Solution:
(a)
Right skewed distribution / median is 0 / mean is 0.85

8. Use the probability distribution of the previous question to answer this question.
What is the variance of the number of accidents per day in the city?

(a) 1.23

(b) 1.67

(c) 2.07

(d) 2.50

(e) 3.10

$$(0-0.85)^2(0.55)+(1-0.85)^2(0.20)+(2-0.85)^2(0.10)+(3-0.85)^2(0.15)+(4-0.85)^2(0)$$
$$=1.2275$$

Solution:
(a)
$\mu = 0.85$
$\sigma^2 = \sum(X - 0.85)^2 P(X) = 1.2275$

9. In a large population, 60% of the voters support a particular party. If a random sample of 300 voters were selected from this population, what would be the approximate probability that more than 175 of them would support this party?

   (a) 0.35

   (b) 0.72

   (c) 0.96

   (d) 0.51

   (e) 0.60

$n = 300 \quad p = 0.6$

$N = 300 \times 0.6 = 180$

$\sigma = \sqrt{300 \times 0.6 \times 0.4} = 8.49$

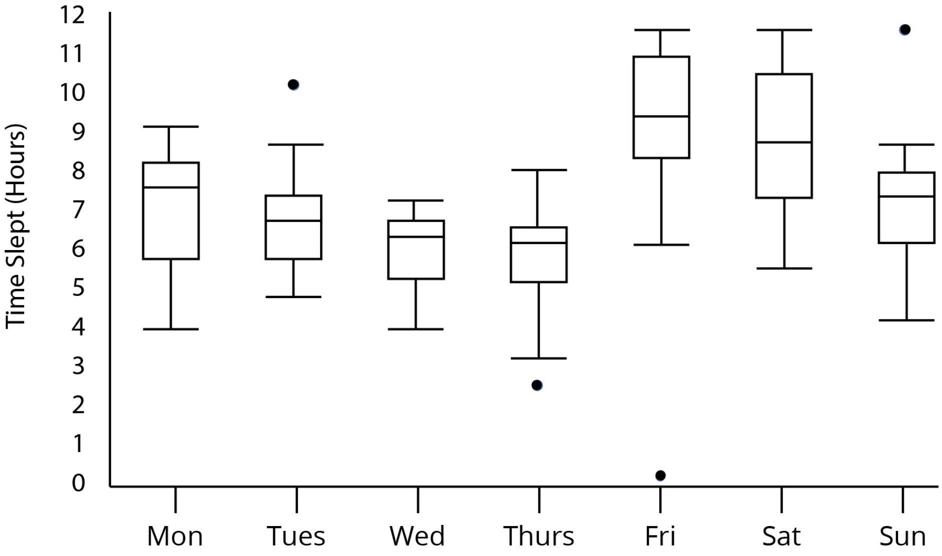
$P(X \leq 175) = P(Z < \frac{175-180}{8.49} = -0.59$

$= 0.2776$

$P(X > 175) = 1 - 0.2776 = 0.7224$

Solution:
b
$X \sim Bin(n = 300, p = 0.6) \overset{approx}{\sim} N(\mu = np = 300 \times 0.6 = 180, \sigma = \sqrt{np(1-p)}) = \sqrt{300 \times 0.6 \times 0.4}) = 8.49.$
$P(X \leq 175) = P(Z < \frac{175-180}{8.49} = -0.59) = 0.2776.$
$P(X > 175) = 1 - P(X \leq 175) = 1 - 0.2776 = 0.7224$

10. The figure below shows time slept (in hours) for each day of the week for a group of 20 students.
Use this figure to answer this question and the following one.



Which of the following statements about time slept on **Monday** is **not correct**?

(a) The distribution of time slept is left skewed

(b) At least 10 students slept more than 7 hours

(c) At least 5 students slept less than 7 hours

(d) At least 15 students slept between 5 and 10 hours

(e) All four statements (a), (b), (c) and (d) are correct

Solution:
(e)

11. Use the figure of the previous question to answer this question.
Which of the following statements is known **for sure**?

(I) The interquartile range (IQR) of time slept on Thursday is smaller than the IQR of time slept on Saturday. ✓

(II) The variance of time slept on Thursday is smaller than the variance of time slept on Saturday.

(III) The mean time slept on Tuesday is smaller than the mean time slept on Monday.

(a) Only (I) is known for sure

(b) Only (II) is known for sure

(c) Only (III) is known for sure

(d) Only (I) and (II) are known for sure

(e) Only (I) and (III) are known for sure

Solution:
(a)
The interquartile range (IQR) of time slept on Thursday is smaller than the IQR of time slept on Saturday because the box is shorter (I correct). We cannot say anything about the mean and variance based on boxplots (II and III not known for sure).

12. A randomized experiment was done by randomly assigning each participant either to walk for half an hour three times a week or to sit quietly reading a book for half an hour three times a week. The participants' blood pressure was measured at the beginning and at the end of the study and the change in participants' blood pressure was computed. The change in participants' blood pressure was compared for the two groups.
This is a randomized experiment rather than an observational study because:

(a) blood pressure was measured at the beginning and end of the study.

(b) the two groups were compared at the end of the study.

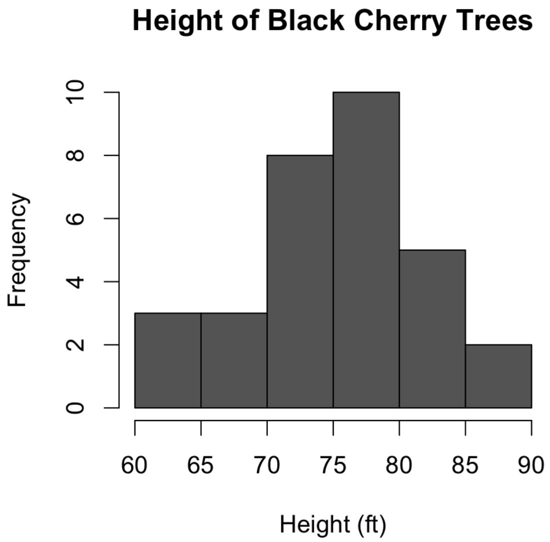
(c) the participants were randomly assigned to either walk or read, rather than choosing their own activity.

(d) a random sample of participants was used.

(e) the study was done by scientists.

Solution:
c

13. The histogram below shows height (in feet) of black cherry trees.



What do we know **for sure** about the median height of these black cherry trees?

(a) The median height is smaller than 75 feet

(b) The median height is 75 feet

(c) The median height is larger than 75 feet but smaller than 80 feet

(d) The median height is 80 feet

(e) The median height is larger than 80 feet

Solution:
(c)
There are 31 observation. The median is the 16th observation, which is between 75 and 80 feet.

14. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\widehat{y} = -0.0127 + 0.0180x.$$

Use this information for this question and the next one.
The above equation implies that:

(a) Each beer consumed increases blood alcohol content by 0.0127
(b) Each beer consumed decreases blood alcohol content by 0.0127
(c) On average it takes 1.8 beers to increase blood alcohol content by 0.01
(d) Each beer consumed increases blood alcohol content by an average of amount of 0.018
(e) Each beer consumed increases blood alcohol content by exactly 0.018

Solution:
(d)

15. Same information as in the previous question.
The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\widehat{y} = -0.0127 + 0.0180x.$$

Suppose that the legal upper limit to drive is a blood alcohol content of 0.08.
Which of the following statements is/are correct?

(I) The model predicts that a driver who drinks 5 beers has a blood content below the legal upper limit to drive.
(II) The model predicts that a driver who drinks 6 beers has a blood content below the legal upper limit to drive.
(III) A driver who drinks 5 beers has **for sure** a blood content that is below the legal upper limit to drive.

(a) Only statement (I) is correct
(b) Only statement (II) is correct
(c) Only statements (I) and (II) are correct
(d) Only statements (I) and (III) are correct
(e) All three statements (I), (II), and (III) are correct

Solution:
(a)
Solving $0.08 = -0.0127 + 0.0180x$ yields $x = 5.15$.

16. In which of the following cases is the normal distribution a reasonable approximation of the sampling distribution of a proportion?

   (I) The probability of an airline flight arriving on time is 90%. The airline operates more than 500 flights per day. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of flights arriving on time in a simple random sample of 30 flights of this airline.

   (II) In a large country 0.2% of the population is affected by a disease. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of individuals affected by the disease in a simple random sample of 2000 individuals in this country.

   (III) A company has 150 workers. The proportion of workers who smoke in this company is 40%. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of workers who smoke in a simple random sample of 30 workers of this company.

   (a) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (I).

   (b) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (II).

   (c) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (III).

   (d) The normal distribution is a reasonable approximation of the sampling distribution of a proportion in at least two of the cases (I), (II), (III).

   (e) The normal distribution is a reasonable approximation of the sampling distribution of a proportion in none of the three cases (I), (II), (III).

   Solution:
   (e)
   Not a reasonable approx in (I) because $n(1-p) = 3 < 10$
   Not a reasonable approx in (II) because $np = 4 < 10$
   Not a reasonable approx in (III) because the sample size (30) is more than 10% of the population (150)

17. Childhood lead poisoning is a public health concern in a country. In this country, 1 child in 10 has a high blood lead level. In a randomly chosen group of 6 children from the population of this country, what is the probability that 3 or more have high blood lead level?

    (a) 0.0146

    (b) 0.5000

    (c) 0.6000

    (d) 0.0159

    (e) 0.9987

    Solution:
    d
    Let X = number of children with high blood level in this sample. Then $X \sim Bin(n = 6, p = 1/10 = 0.1)$. $P(X \geq 3) = 0.0146 + 0.0012 + 0.0001 + 0.0000 = 0.0159$

18. A random experiment consists of five possible outcomes: E_1, E_2, E_3, E_4 and E_5. This means the sample space is $S = \{E_1, E_2, E_3, E_4, E_5\}$. If $P(E_1) = 3 \times P(E_2) = 0.3$ and $P(E_3) = P(E_4) = P(E_5)$, what is $P(E_5)$?

    (a) 0.1

    (b) 0.2

    (c) 0.3

    (d) 0.4

    (e) 0.6

    Solution:
    c
    $P(E_1) = 3P(E_2) = 0.3 \implies P(E_2) = 0.1$ and so $P(E_3) + P(E_4) + P(E_5) = 1 - (0.3 + 0.1) = 0.6$ and since $P(E_3) = P(E_4) = P(E_5)$, $P(E_5) = 0.6/3 = 0.2$

19. In a statistics class, there are 60 men and 40 women. The mean weight of all these 100 students is 160 pounds. If the mean weight of the 60 men is 180 pounds, what is the mean weight of the 40 women?

    (a) 120 pounds
    (b) 125 pounds
    (c) 130 pounds
    (d) 132 pounds
    (e) 135 pounds

    Solution:
    c
    $\frac{100\times160-60\times180}{40} = 130\text{lb}$

20. A business office orders paper supplies from one of three vendors, V1, V2, or V3. Orders are to be placed on three successive days (e.g. next Mon, Tue and Wed), one order per day. On each day, one of the three vendors is selected at random, giving each vendor the same probability (which is 1/3) to be selected independently of the other vendors.
    Use this information for this question and the next one.
    Find the probability that vendor V1 gets at least one order.

    (a) $\frac{1}{3}$
    (b) $\frac{1}{9}$
    (c) $\frac{8}{27}$
    (d) $\frac{19}{27}$
    (e) $\frac{26}{27}$

    Solution:
    d
    $P(\text{ at least one V1}) = 1 - P(\text{ None is V1}) = 1 - P(\text{V1}^c)^3 = 1 - \left(\frac{2}{3}\right)^3 = 1 - \frac{8}{27} = \frac{19}{27}$

21. Same information as in the previous question.

A business office orders paper supplies from one of three vendors, V1, V2, or V3. Orders are to be placed on three successive days (e.g. next Mon, Tue and Wed), one order per day. On each day, one of the three vendors is selected at random, giving each vendor the same probability (which is 1/3) to be selected independently of the other vendors.

Find the probability that vendor V1 gets exactly two of these three orders.

(a) $\frac{1}{3}$

(b) $\frac{1}{9}$

(c) $\frac{2}{9}$

(d) $\frac{2}{27}$

(e) $\frac{3}{27}$

Solution:

c

$P($ exactly two V1$) = P(V1, V1, V1^c) + P(V1, V1^c, V1) + P(V1^c, V1, V1) = 3 \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right) = \frac{2}{9}$

Note: This question does not require the binomial tables or the binomial formula even though this can also be done using the binomial formula.

22. In a clinical study of a new drug, patients are randomly allocated to one of two groups. The patients in the first group are given an injection with the drug in it. The patients in the second group are given an injection, identical to the injection containing the drug, but with no drug in it. Doctors record the symptoms reported by the patients and take measurements, without knowing which patients are in which group. Data are then analysed to determine whether the drug is effective.
Which of the following statements is/are correct?

(I) If the groups end up different in terms of symptoms and measurements, then we know **for sure** that this difference is due to the administered drug, no matter how big or small the difference is.

(II) This is an example of a double-blind experiment.

(III) This is an example of an experiment where the units are assigned to treatments using a randomized block design.

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

(d) Only (I) and (II) are correct

(e) Only (II) and (III) are correct

Solution:
(b) (I) not correct because the difference could be due to chance, specially if it is small. We have to check whether the difference is statistically significant; This is an example of a double-blind experiment where the units are assigned to groups using a completely randomized design, i.e. (II) is correct and (III) is not correct.

23. A randomly selected sample of 1000 college students was asked whether they had ever used the drug Ecstasy. Sixteen percent (16% or 0.16) of the 1000 students surveyed said they had.
Which of the following statements about the number 0.16 is/are correct?

(I) It is the sample proportion of college students who have ever used the drug Ecstasy.

(II) It is the population proportion of college students who have ever used the drug Ecstasy.

(III) It is a value that can be used to estimate the population proportion of college students who have ever used the drug Ecstasy.

(a) Only (I) is correct
(b) Only (II) is correct
(c) Only (III) is correct
(d) Only (I) and (III) are correct
(e) Only (II) and (III) are correct

Solution:
(d) it is a value computed from the sample.

24. Suppose the correlation coefficient between height (measured in feet) and weight (measured in pounds) is 0.40. What is the correlation coefficient between height measured in inches and weight measured in ounces? The conversion is 12 inches = 1 foot and 16 ounces = 1 pound.

(a) 0.002
(b) 0.30
(c) 0.40
(d) 0.533
(e) Impossible to say given the information provided

Solution:
(b)
The correlation coefficient has not units

25. A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

Use this information for this question and the next two questions.
One student in the sample was 73 inches tall with a foot length of 29 cm. What is the residual for this student?

(a) 29 cm
(b) 1.31 cm
(c) 0 cm
(d) 27.69 cm
(e) -1.31 cm

Solution:
b
$29 - (10.9 + 0.23 \times 73) = 1.31$

26. Same information as in the previous question.
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

If we convert the heights of all these 33 students into cm using the conversion factor 1 inch = 2.54 cm and recalculate the regression equation with both the variables measured in cm, what will be the slope of the new regression equation?

(a) 0.09
(b) 0.58
(c) 0.23
(d) 11.48
(e) 10.99

Solution:
a
$b = r\frac{s_y}{s_x}$. r and s_y don't change. s_x gets multiplied by 2.54 and so the new slope is $0.23/2.54 = 0.09055118$

27. Same information as in the previous two questions.
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

If we convert the foot lengths of all these 33 students into inches using the conversion factor 1 cm = 0.3937 inches and recalculate the regression equation with both the variables measured in inches, what will be the y-intercept of the new regression equation?

(a) 11.29

(b) 28.27

(c) 10.9

(d) 27.69

(e) 4.29

Solution:
e
$a = \bar{y} - b\bar{x}, b = r\frac{s_y}{s_x}$. r and $\bar{x}$ don't change. $\bar{y}$,s_y and b get multiplied by 0.3937 and so the new y-intercapt is is $10.9 \times 0.3937 = 4.29133$

28. If the correlation coefficient r between two variables is equal to 1, what do we know **for sure**?

(a) There is a perfect positive linear relationship between the two variables.

(b) There is a very strong, but not perfect, positive linear relationship between the two variables.

(c) There is no relationship between the two variables.

(d) There is a perfect negative linear relationship between the two variables.

(e) It is impossible to know one of the above for sure given the information provided.

Solution:
(a)

29. The probability that a randomly chosen American is a Republican is 0.35. We select a simple random sample of 6 Americans. What is the probability that at least 1 of the selected Americans is a Republican?

   (a) 0.0754
   (b) 0.2437
   (c) 0.3191
   (d) 0.3500
   (e) 0.9246

   Solution:
   (e) 1 - 0.0754 = 0.9246

30. The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.
   Use this information for this question and the next one.
   If I select at random two Canada Post letters, what is the probability that they will both weigh less than 1 oz.?

   (a) 0.0005
   (b) 0.0228
   (c) 0.0456
   (d) 0.4772
   (e) 0.8413

   Solution:
   (a) The probability that one letter weigh less than 1 oz. is $P(X < 1) = P(Z < -2) = 0.0228$. The probability that they will both weigh less than 1 oz is 0.0228^2 (multiplication rule for independent events).

31. Same information as in the previous question.
The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.
If I select at random 1000 Canada Post letters, what is the probability that at least 200 will weigh less than 1 oz.?

   (a) 0.0000
   (b) 0.0046
   (c) 0.1040
   (d) 0.8960
   (e) 1.0000

   Solution:
   (a) The number of letters that weigh less than 1 oz follows a $B(n = 1000, p = 0.0228)$. We can use the normal approximation since $np = 22.8$ and $n(1 - p) = 977.2$. The probability that at least 200 letters will weigh less than 1 oz. is $P(X >= 200) = P(Z >= (200 - 22.8)/\sqrt{22.280}) = 1 - P(Z < 37.54) = 0$

32. A city council of a small city wants to know the proportion of eligible voters that oppose having a garbage incinerator opened just outside the city limits. They randomly select 100 residential numbers from the city's telephone book that contains 3,000 such numbers. Each selected residence is then called and asked to provide the total number of eligible voters in the residence and the number of voters opposed to the incinerator in the residence.
What sampling method is used?

   (a) Simple random sampling
   (b) Stratified sampling
   (c) Cluster sampling
   (d) Multistage sampling
   (e) Systematic sampling

   Solution:
   (c) Clusters = Residences

33. Acme Medicine is conducting an experiment to test a new vaccine, developed to immunize people against the common cold. To test the vaccine, Acme has 1000 participants. Participants are divided into two groups, based on gender. Then, within each group, half of the participants are randomly assigned to treatment groups so that the first half gets the placebo and the second half gets the vaccine.
Which of the following statements is correct?

    (a) This is an example of a prospective observational study.
    (b) This is an example of an experiment where the units are assigned to treatments using a randomized block design.
    (c) This is an example of an experiment where the units are assigned to treatments using a completely randomized design.
    (d) This is an example of a census.
    (e) This is an example of a survey where the participants are selected using stratified sampling.

    Solution:
    (b)

34. A survey of the male students at a junior college reveals that 26% play soccer regularly, 22% are Latino, and half of the Latino students play soccer regularly. If a male student is selected at random, what is the probability that he is neither Latino nor a soccer player?

    (a) 0.89
    (b) 0.52
    (c) 0.63
    (d) 0.26
    (e) 0.41

    Solution:
    c
    P(So) = 0.26, P(L) = 0.22, P(So|L) = 0.5 and so P(So and L) = $0.22 \times 0.5 = 0.11$
    P(L or So) = P(L) + P(So) - P(L and So) = 0.26 +0.22-0.11 = 0.37
    P(neither Latino nor a soccer player) = $1 - 0.37 = 0.63$

35. The table below gives the probability distribution of the number of times (x) a photocopying machine needs repair in a given month.

x	0	1	2	3	4	5
Probability	0.12	0.18	0.23	0.20	0.14	0.13

What is the probability that the machine will have to be repaired at least twice but not more than four times in the given month?

(a) 0.23

(b) 0.24

(c) 0.43

(d) 0.57

(e) 0.87

Solution:
d
$P(2 \leq X \leq 4) = 0.23 + 0.20 + 0.14 = 0.57$

36. Students in an Intro Stats course were asked to describe their political preference as Liberal, Moderate, or Conservative. Here are the results:

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total	85	80	27	192

Use this information for this question and the next one.
We select a student at random. Let L denote the event that the student selected considers himself/herself to be liberal and let F denote the event that the student selected is female. Calculate $P(L|F)$.

(a) 0.18

(b) 0.25

(c) 0.35

(d) 0.41

(e) 0.45

Solution:
e
$P(L|F) = \frac{35}{77} = 0.4545455$

37. Same information as in the previous question.
Students in an Intro Stats course were asked to describe their political preference as Liberal, Moderate, or Conservative. Here are the results:

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total	85	80	27	192

We select a student at random. Let L denote the event that the student selected considers himself/herself to be liberal and let M denote the event that the student selected is male. Calculate $P(L \text{ or } M)$.

(a) 0.18

(b) 0.27

(c) 0.45

(d) 0.78

(e) 1.04

Solution:
c
$P(L \text{ or } M) = P(L) + P(M) - P(L \text{ and } M) = \frac{85}{192} + \frac{115}{192} - \frac{50}{192} = \frac{150}{192} = 0.78125$

38. Using a random sample of 4,000 students, you compute a 95% confidence interval for the proportion of overweight students. You decide to compute another 95% confidence interval using a different sample, this time with only 1,000 students. Suppose that the sample proportion $\widehat{p}$ is the same for both these confidence intervals. What change would you expect from the first confidence interval to the second?

    (a) The margin of error of the second interval will be 4 times as big as the margin of error of the first interval.

    (b) The margin of error of the second interval will be 2 times as big as the margin of error of the first interval.

    (c) The margin of error of the first interval will be 4 times as big as the margin of error of the second interval.

    (d) The margin of error of the first interval will be 2 times as big as the margin of error of the second interval.

    (e) There will be no change in the margin of error since the sample proportion $\widehat{p}$ is the same for both confidence intervals.

    Solution:
    (b)
    The margin of error of the second interval is:

$$1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{1000}} = 2 \cdot 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{4000}}$$

    which is 2 times the margin of error of the first interval.

39. The heights of students in some population have a Normal distribution with mean 172cm. Twenty five percent of the students in this population have heights greater than or equal to 180cm. What is the standard deviation of the heights of the students in this population?

    (a) 7.9cm

    (b) 8.7cm

    (c) 9.9cm

    (d) 10.5cm

    (e) 11.9cm

    Solution:
    e
    $P(X \geq 180) = 0.25 \implies z = 0.67$ and $\frac{180-172}{\sigma} = 0.67 \implies \sigma = \frac{180-172}{0.67} = 11.94$

40. In a game, a wheel with five equal-sized spaces labeled from A to E is spun and a ball randomly lands in one of these spaces. If the ball lands in the space labeled A, the player wins 15 dollars. If the ball lands in any other space, the winner loses 4 dollars. Use this information for this questions and the next one.
A player plays this game 6 times. What is the probability that the ball lands at least one time in the space labeled A? You can assume that games are independent from one another.

(a) 0.2622
(b) 0.3446
(c) 0.3932
(d) 0.6068
(e) 0.7378

Solution:
X = number of times the ball lands in a space labeled A
$X \sim Binom(n = 6, p = 0.2)$
$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.2622 = 0.7378$

41. Same information as in the previous question.
In a game, a wheel with five equal-sized spaces labeled from A to E is spun and a ball randomly lands in one of these spaces. If the ball lands in the space labeled A, the player wins 15 dollars. If the ball lands in any other space, the winner loses 4 dollars. What is the expected gain or loss of a player who plays 40 games? You can assume that games are independent from one another.

(a) Gain of 360 dollars
(b) Gain of 55 dollars
(c) Gain of 8 dollars
(d) Loss of 1 dollar
(e) Loss of 8 dollars

Solution:
Expected loss or gain for one game: $0.2 \cdot 15 + 0.8 \cdot (-4) = -0.2$
Expected loss or gain for 40 games (games are independent): $40 \cdot (-0.2) = -8$

42. Five hundred people used a home test for a HIV. We say that the test is positive if it indicates that the person has HIV, and we say it is negative if it indicates that the person does not have HIV. Out of these 500 people, 40 had HIV and 460 did not have HIV. The table below shows the contingency table of the result of the test (positive or negative) and whether a person has HIV (HIV, no HIV) for these 500 people.

	HIV	No HIV	Total
Positive Test	35	25	60
Negative Test	5	435	440
Total	40	460	500

A false positive is a test result that indicates a person has a disease when the person actually does not have it. What is the probability of a false positive with the home test for a HIV?

(a) 0.054

(b) 0.070

(c) 0.130

(d) 0.417

(e) 0.875

Solution:
(a) 25/460

43. A consumer organization estimates that 34% of the households in a particular community have one television set, 39% have two sets, and 20% have three or more sets. What is the probability that a household chosen at random in this community has no more than one television set?

(a) 0.41

(b) 0.59

(c) 0.07

(d) 0.34

(e) 0.46

Solution:
a
P(no more than 1) = 1- P(more than 1) = 1- P(2)- P(3) = 1-0.39-0.20 = 0.41

44. A randomly selected sample of 1000 college students was asked whether they were in favor of longer hours at the school library. Eighty percent (80%) of the 1000 students surveyed said they were in favor.
Which of the following statements is/are correct?

(I) A 95% confidence interval for the population proportion of college students who were in favor of longer hours at the school library is $[0.78; 0.82]$

(II) The population proportion of college students who were in favor of longer hours at the school library is significantly different from 76% at the level of significance $\alpha = 5\%$.

(III) The population proportion of college students who were in favor of longer hours at the school library is significantly different from 76% at the level of significance $\alpha = 1\%$.

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

(d) Only (I) and (II) are correct

(e) The three statements (I), (II), and (III) are correct

Solution:
(e)
A 95% confidence interval is $0.8 \pm 1.96\sqrt{0.2 \cdot 0.8/1000} = [0.78; 0.82]$. The value of the test statistic is z-obs $= 2.96$ and the p-value is $2 \cdot 0.0015 = 0.003$.

45. After menopause, some women take supplemental oestrogen. There is some concern that if these women also drink alcohol, their oestrogen levels will rise too high. Twelve volunteers who were receiving supplemental oestrogen were randomly divided into two groups, as were 12 other volunteers who were not receiving supplemental oestrogen. In each case, one group drank an alcoholic beverage, the other a non-alcoholic beverage. An hour later, everyone's oestrogen level was checked. Only those on supplemental oestrogen who drank alcohol showed a marked increase in oestrogen levels.
Use this information for this questions and the next one.
What is the response variable in this study?

(a) "Menopause" is the response variable

(b) "Amount of supplemental oestrogen received" is the response variable

(c) "Type of beverage" (alcoholic beverage, non-alcoholic beverage) is the response variable

(d) "Oestrogen level" is the response variable

(e) "Gender" is the response variable

Solution:
(d)

46. Same information as in the previous question.
After menopause, some women take supplemental oestrogen. There is some concern that if these women also drink alcohol, their oestrogen levels will rise too high. Twelve volunteers who were receiving supplemental oestrogen were randomly divided into two groups, as were 12 other volunteers who were not receiving supplemental oestrogen. In each case, one group drank an alcoholic beverage, the other a non-alcoholic beverage. An hour later, everyone's oestrogen level was checked. Only those on supplemental oestrogen who drank alcohol showed a marked increase in oestrogen levels.
Which of the following statements is/are correct?

(I) This is an example of an experiment where the units are assigned to treatments using a randomized block design

(II) "Amount of supplemental oestrogen received" is a blocking variable

(III) This is an example of an observational study since patients were not randomly assigned to treatments

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

(d) Only (I) and (II) are correct

(e) None of the three statements is correct

Solution:
(d)

47. A box contains 3 yellow, 2 red, 4 green and 3 black marbles. Two marbles are taken one after the other at random from the box without replacement. What is the probability that both marbles are red?

(a) $\frac{1}{50}$

(b) $\frac{1}{60}$

(c) $\frac{1}{66}$

(d) $\frac{1}{24}$

(e) $\frac{1}{18}$

Solution:
c
Let $R1$ be the event that the first marble selected is red and $R2$ be the event that the second is red.
We want $P(R1 \text{ and } R2)$
Using general multiplication rule we have,
$P(R1 \text{ and } R2) = P(R1) \times P(R2|R1) = \frac{2}{3+2+4+3} \times \frac{2-1}{3+2+4+3-1} = \frac{2}{12} \times \frac{1}{11} = \frac{1}{66}$

48. A local bakery has determined a probability distribution for the number of cheesecakes it sells in a given day. The distribution is as follows:

Number sold in a day	10	20	30	35	x
Probability	0.1	0.5	0.1	0.2	0.1

If the mean number of cheesecakes that this local bakery sells in a day is 25, what is the value of x?

(a) 30

(b) 35

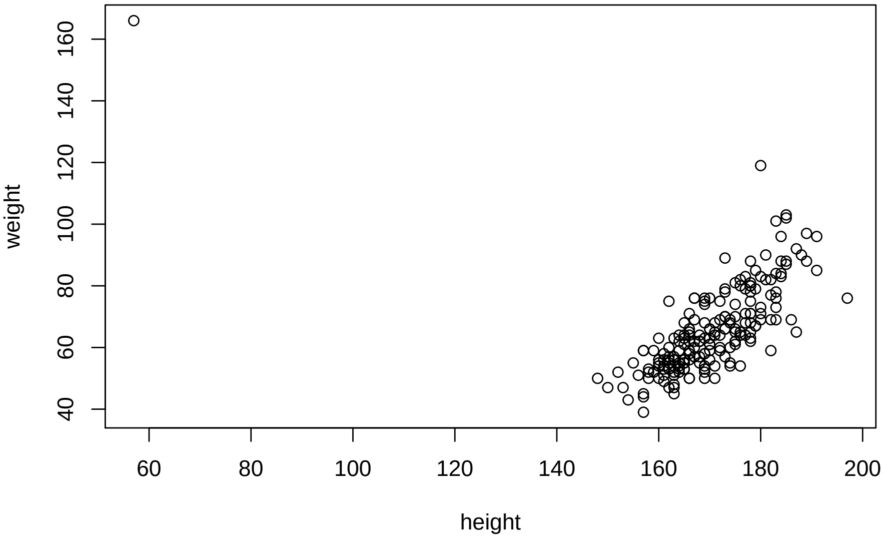
(c) 38

(d) 40

(e) 45

Solution:
d
$(10 \times 0.1) + (20 \times 0.5) + (30 \times 0.1) + (35 \times 0.2) + (x \times 0.1) = 25 \implies x \times 0.1 = 25 - 21 \implies x = \frac{4}{0.1} = 40$

49. The scatterplot below shows the self-reported height and weight of 200 subjects. The subjects were men and women engaged in regular exercise. Suppose we fit a linear regression of weight on height.
Use this scatterplot for this question and the next one.



Which of the following statements about the data point with height = 57 is/are true?

(I) The data point with height = 57 is an outlier.

(II) The data point with height = 57 is a high leverage point.

(III) The data point with height = 57 is an influential observation.

(a) Only (I) is true

(b) Only (II) is true

(c) Only (III) is true

(d) All three statements (I), (II) and (III) are true

(e) None of the three statements is true

Solution:
(d)

50. Use the scatterplot of the previous question to answer this question.
Suppose we fit a linear regression of weight on height.
Which of the following statements is known **for sure**?

- (I) If we remove the data point with height = 57, the slope of the regression line increases.
- (II) If we remove the data point with height = 57, the mean height increases.
- (III) If we remove the data point with height = 57, the variance of height increases.

- (a) Only (II) is known for sure
- (b) Only (I) and (II) are known for sure
- (c) Only (I) and (III) are known for sure
- (d) Only (II) and (III) are known for sure
- (e) None of the three statements is known for sure

Solution:
(b)
If we remove the data point with height = 57, the slope of the regression line increases (I true), the mean height increases (II true), the variance of height decreases (III not true).

END OF EXAM

Tables of normal, binomial and t-distributions

prepared by Ken Butler

2015-06-03

Table Z (normal distribution)

Values of z greater than 0 are on the next page. Second decimal place is at the top of the column.

	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
-3.3	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005
-3.2	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007
-3.1	0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010
-3.0	0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013
-2.9	0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019
-2.8	0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026
-2.7	0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035
-2.6	0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047
-2.5	0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062
-2.4	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
-2.3	0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
-2.2	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139
-2.1	0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179
-2.0	0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228
-1.9	0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287
-1.8	0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359
-1.7	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
-1.6	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
-1.5	0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668
-1.4	0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808
-1.3	0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968
-1.2	0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151
-1.1	0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357
-1.0	0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587
-0.9	0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841
-0.8	0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119
-0.7	0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420
-0.6	0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743
-0.5	0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085
-0.4	0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446
-0.3	0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821
-0.2	0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207
-0.1	0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602
0.0	0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000

Table Z (continued)

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table of the Binomial Distribution

(Table shows probability of exactly k successes in n trials with success probability p. Values of p are shown along the top of the table.)

		p									
n	k	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
2	1	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
2	2	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
3	1	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
3	2	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
3	3	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
4	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
4	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
4	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
4	4	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
5	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1563
5	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
5	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1812	0.2304	0.2757	0.3125
5	4	0.0000	0.0004	0.0022	0.0064	0.0147	0.0284	0.0488	0.0768	0.1128	0.1563
5	5	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0313
6	0	0.7351	0.5315	0.3772	0.2622	0.1780	0.1177	0.0754	0.0467	0.0277	0.0156
6	1	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0938
6	2	0.0306	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344
6	3	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125
6	4	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344
6	5	0.0000	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938
6	6	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156

Computer & Mathematical Sciences
UNIVERSITY OF TORONTO
SCARBOROUGH

STAB22: Final Exam April 19, 2018

Complete this form, then *carefully* tear off this page and place it along with your student card on the surface beside where you are writing.

The form will be verified and collected during the test.

Information collected will be used solely for the purpose of administering this test.

Name (please print): ____

Signature: ___________________________ Seat #: ________________
(if available)

Student Number: ☐☐☐☐☐☐☐☐☐☐

UNIVERSITY OF TORONTO SCARBOROUGH
Department of Computer and Mathematical Sciences
April 19, 2018
APRIL 2018 FINAL EXAMINATIONS

STAB22H3 Statistics I
Duration: 3 hours

Last Name:________________________ First Name: ________________________

Student number: ________________________

Aids allowed:

- Two handwritten letter-sized sheets (both sides) of notes prepared by you
- Non-programmable, non-communicating calculator

The table of normal and binomial distributions are attached at the end.

This test is based on multiple-choice questions. There are 50 questions. All questions carry equal weight. Report all your answers on the Scantron answer sheet before the end of the exam. On the Scantron answer sheet, ensure that you write down and **BUBBLE** your last name, first name (as much of it as fits), and student number (in "Identification").

Mark in each case the best answer out of the alternatives given (**which means the numerically closest answer if the answer is a number and the answer you obtained is not given.**)

Also before you begin, complete the signature sheet, but sign it only when the invigilator collects it. The signature sheet shows that you were present at the exam.

There are 30 pages including this page and statistical tables. Please check to see that you have all the pages.

Good luck!!

Answer Key for Exam $\boxed{\text{A}}$

1. If there is a very strong linear relationship between two variables then what do you know **for sure** about the correlation coefficient r?

   (a) The correlation coefficient is larger than 1
   (b) The correlation coefficient is close to 0
   (c) The correlation coefficient is close to -1
   (d) The correlation coefficient is close to 1
   [✓ (e)] None of the above is known for sure.

   Solution:
   (e) We know for sure that the correlation coefficient is close to -1 or 1.

2. A box contains 18 balls, some of them are black and the others are white. We do not know how many are black and how many are white. Let n be the number of black balls and so $18 - n$ is the number of white balls. We are going to pick four balls at random from this box, one by one, without replacement (i.e when a ball is picked, we do not put it back in the box). Let A denote the event that the first three balls selected are black and B denote the event that the fourth ball selected is black. If $P(B|A) = 1/5$, what is the value of n?

   [✓ (a)] 6
   (b) 7
   (c) 8
   (d) 9
   (e) 12

   Solution:
   a
   After taking out the first 3 balls there are 15 left. $P(B|A) = 1/5$ means the probability of selecting a black ball from these remaining 15 balls is $1/5$. This means there must be 3 $(15 \times \frac{1}{5})$ black balls in these 15 remaining balls. This means at the beginning there were $n = 3 + 3 = 6$ black balls.

   Or second method:
   $P(B|A) = \frac{n-3}{18-3} = 1/5 \implies n = 6$

3. A traditional die is a cube, each face showing an integer in the range from 1 to 6. A **fair** traditional die is rolled 3 times. The first 2 rolls result in the face showing "5" being uppermost. What is the probability the third roll will **not** result in the face showing "5" being uppermost?

   (a) 0

   (b) 0.023

   (c) 0.069

   $\boxed{\text{(d)}}$ 0.833

   (e) 1

   Solution:
   (d) The rolls are independent. The probability the third roll will not result in the face with "5" being uppermost is $1 - 1/6 = 5/6 = 0.833$.

4. The tail length for a population of female hook-billed kites is Normally distributed with mean 194 mm and standard deviation 11 mm. We select a random sample of 200 kites from this population. Let X be the number of kites in this sample having a tail length between 179 and 209 mm. What is the mean of X?

   (a) 145

   (b) 155

   $\boxed{\text{(c)}}$ 165

   (d) 175

   (e) 194

   Solution:
   c
   Letting L denote the tail length of a randomly selected female hook-billed kite, $X \sim Bin(n = 300, p = P(179 < L < 209))$.
   $p = P(179 < L < 209) = P(\frac{179-194}{11} < Z < \frac{209-194}{11}) = P(-1.36Z < 1.36) = 1 - 2 \times 0.0869 = 1 - 0.1738 = 0.8262$
   $\mu_X = np = 200 \times 0.8262 = 165.24$

5. The probability that a patient with a certain disease will be successfully treated with a new medical treatment is 0.80. The success of the treatment is independent from one patient to another. Suppose that the treatment is used on 40 patients with the disease. What is the mean of the number of patients who will **not** be successfully treated?

   [✓ (a)] 8
   (b) 20
   (c) 32
   (d) 40
   (e) Impossible to say given the information provided

   Solution:
   (a)
   The mean of the number of patients who will **not** be successfully treated is $n(1-p) = 40 \cdot 0.2 = 8$

6. STA101 and STA102 are two statistics courses both with large classes. We have the following information about the distributions of grades for these two courses:

   - The distribution of grades of STA101 is right skewed.
   - The distribution of grades of STA102 is left skewed.
   - STA101 has lower class average than STA102 (i.e. $\mu_1 < \mu_2$, where μ_1 and μ_2 are the means of the distributions of grades in STA101 and STA102 respectively.)

   Which of the following statements regarding the distributions of grades of these two courses is **NOT** true?

   (a) The median grade of STA101 is less than the median grade of STA102.
   [✓ (b)] The mean grade of STA102 is less than the median grade of STA101.
   (c) The mean grade of STA101 is less than the median grade of STA102.
   (d) The median grade of STA101 is less than the mean grade of STA102.
   (e) The mean grade of STA102 is less than the median grade of STA102.

   Solution:
   b
   Denoting the medians of STA101 and STA102 by m_1 and m_2 respectively, we have,
   $m_1 < \mu_1$, since the distribution of grades of STA101 is right skewed,
   $\mu_2 < m_2$, since the distribution of grades of STA102 is left skewed,
   and so since we are also given that $\mu_1 < \mu_2$, we have
   $m_1 < \mu_1 < \mu_2 < m_2$ and this implies:

   $m_1 < m_2$, i.e. (a) is true
   $m_1 < \mu_2$, i.e. (b) is false
   $\mu_1 < m_2$, i.e. (c) is true
   $m_1 < \mu_2$, i.e. (d) is true
   $\mu_2 < m_2$, i.e. (e) is true

7. The following table represents the probability distribution of the number of accidents per day in a city.

Number of accidents	0	1	2	3	4 or more
Probability	0.55	0.20	0.10	0.15	0

Use this table for this question and the next one.
Which of the following statements about the number of accidents per day in the city is/are true?

(I) The mean number of accidents per day is larger than the median number of accidents per day.

(II) The mean number of accidents per day is smaller than the median number of accidents per day.

(III) The median number of accidents per day is equal to 1.

[✓ (a)] Only (I) is true
(b) Only (II) is true
(c) Only (III) is true
(d) Only (I) and (II) are true
(e) Only (I) and (III) are true

Solution:
(a)
Right skewed distribution / median is 0 / mean is 0.85

8. Use the probability distribution of the previous question to answer this question.
What is the variance of the number of accidents per day in the city?

[✓ (a)] 1.23
(b) 1.67
(c) 2.07
(d) 2.50
(e) 3.10

Solution:
(a)
$\mu = 0.85$
$\sigma^2 = \sum(X - 0.85)^2 P(X) = 1.2275$

9. In a large population, 60% of the voters support a particular party. If a random sample of 300 voters were selected from this population, what would be the approximate probability that more than 175 of them would support this party?

   (a) 0.35

   $\boxed{\text{(b)}}$ 0.72

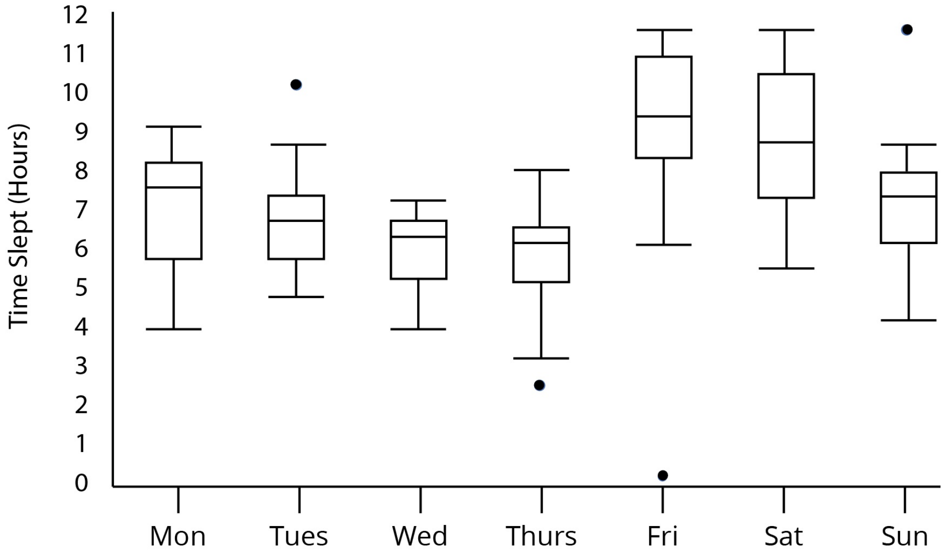
   (c) 0.96

   (d) 0.51

   (e) 0.60

   Solution:
   b
   $X \sim Bin(n = 300, p = 0.6) \overset{approx}{\sim} N(\mu = np = 300 \times 0.6 = 180, \sigma = \sqrt{np(1-p)}) = \sqrt{300 \times 0.6 \times 0.4}) = 8.49.$
   $P(X \leq 175) = P(Z < \frac{175-180}{8.49} = -0.59) = 0.2776.$
   $P(X > 175) = 1 - P(X \leq 175) = 1 - 0.2776 = 0.7224$

10. The figure below shows time slept (in hours) for each day of the week for a group of 20 students.
Use this figure to answer this question and the following one.



Which of the following statements about time slept on **Monday** is **not correct**?

(a) The distribution of time slept is left skewed

(b) At least 10 students slept more than 7 hours

(c) At least 5 students slept less than 7 hours

(d) At least 15 students slept between 5 and 10 hours

[✓ (e)] All four statements (a), (b), (c) and (d) are correct

Solution:
(e)

11. Use the figure of the previous question to answer this question.
Which of the following statements is known **for sure**?

(I) The interquartile range (IQR) of time slept on Thursday is smaller than the IQR of time slept on Saturday.

(II) The variance of time slept on Thursday is smaller than the variance of time slept on Saturday.

(III) The mean time slept on Tuesday is smaller than the mean time slept on Monday.

[✓ (a)] Only (I) is known for sure
(b) Only (II) is known for sure
(c) Only (III) is known for sure
(d) Only (I) and (II) are known for sure
(e) Only (I) and (III) are known for sure

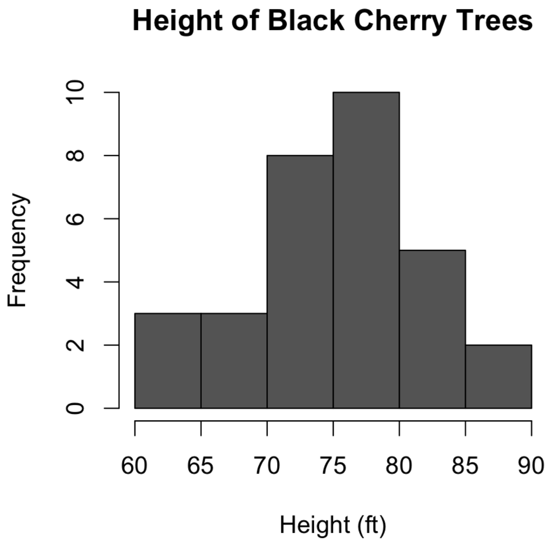
Solution:
(a)
The interquartile range (IQR) of time slept on Thursday is smaller than the IQR of time slept on Saturday because the box is shorter (I correct). We cannot say anything about the mean and variance based on boxplots (II and III not known for sure).

12. A randomized experiment was done by randomly assigning each participant either to walk for half an hour three times a week or to sit quietly reading a book for half an hour three times a week. The participants' blood pressure was measured at the beginning and at the end of the study and the change in participants' blood pressure was computed. The change in participants' blood pressure was compared for the two groups.
This is a randomized experiment rather than an observational study because:

(a) blood pressure was measured at the beginning and end of the study.
(b) the two groups were compared at the end of the study.
[✓ (c)] the participants were randomly assigned to either walk or read, rather than choosing their own activity.
(d) a random sample of participants was used.
(e) the study was done by scientists.

Solution:
c

13. The histogram below shows height (in feet) of black cherry trees.



What do we know **for sure** about the median height of these black cherry trees?

(a) The median height is smaller than 75 feet

(b) The median height is 75 feet

[✓ (c)] The median height is larger than 75 feet but smaller than 80 feet

(d) The median height is 80 feet

(e) The median height is larger than 80 feet

Solution:
(c)
There are 31 observation. The median is the 16th observation, which is between 75 and 80 feet.

14. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\widehat{y} = -0.0127 + 0.0180x.$$

Use this information for this question and the next one.
The above equation implies that:

(a) Each beer consumed increases blood alcohol content by 0.0127

(b) Each beer consumed decreases blood alcohol content by 0.0127

(c) On average it takes 1.8 beers to increase blood alcohol content by 0.01

[✓ (d)] Each beer consumed increases blood alcohol content by an average of amount of 0.018

(e) Each beer consumed increases blood alcohol content by exactly 0.018

Solution:
(d)

15. Same information as in the previous question.
The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\widehat{y} = -0.0127 + 0.0180x.$$

Suppose that the legal upper limit to drive is a blood alcohol content of 0.08.
Which of the following statements is/are correct?

(I) The model predicts that a driver who drinks 5 beers has a blood content below the legal upper limit to drive.

(II) The model predicts that a driver who drinks 6 beers has a blood content below the legal upper limit to drive.

(III) A driver who drinks 5 beers has **for sure** a blood content that is below the legal upper limit to drive.

$\boxed{\text{(a)}}$ Only statement (I) is correct
(b) Only statement (II) is correct
(c) Only statements (I) and (II) are correct
(d) Only statements (I) and (III) are correct
(e) All three statements (I), (II), and (III) are correct

Solution:
(a)
Solving $0.08 = -0.0127 + 0.0180x$ yields $x = 5.15$.

16. In which of the following cases is the normal distribution a reasonable approximation of the sampling distribution of a proportion?

(I) The probability of an airline flight arriving on time is 90%. The airline operates more than 500 flights per day. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of flights arriving on time in a simple random sample of 30 flights of this airline.

(II) In a large country 0.2% of the population is affected by a disease. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of individuals affected by the disease in a simple random sample of 2000 individuals in this country.

(III) A company has 150 workers. The proportion of workers who smoke in this company is 40%. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of workers who smoke in a simple random sample of 30 workers of this company.

(a) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (I).

(b) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (II).

(c) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (III).

(d) The normal distribution is a reasonable approximation of the sampling distribution of a proportion in at least two of the cases (I), (II), (III).

[✓ (e)] The normal distribution is a reasonable approximation of the sampling distribution of a proportion in none of the three cases (I), (II), (III).

Solution:
(e)
Not a reasonable approx in (I) because $n(1-p) = 3 < 10$
Not a reasonable approx in (II) because $np = 4 < 10$
Not a reasonable approx in (III) because the sample size (30) is more than 10% of the population (150)

17. Childhood lead poisoning is a public health concern in a country. In this country, 1 child in 10 has a high blood lead level. In a randomly chosen group of 6 children from the population of this country, what is the probability that 3 or more have high blood lead level?

    (a) 0.0146
    (b) 0.5000
    (c) 0.6000
    $\boxed{\text{(d)}}$ 0.0159
    (e) 0.9987

    Solution:
    d
    Let X = number of children with high blood level in this sample. Then $X \sim Bin(n = 6, p = 1/10 = 0.1)$. $P(X \geq 3) = 0.0146 + 0.0012 + 0.0001 + 0.0000 = 0.0159$

18. A random experiment consists of five possible outcomes: E_1, E_2, E_3, E_4 and E_5. This means the sample space is $S = \{E_1, E_2, E_3, E_4, E_5\}$. If $P(E_1) = 3 \times P(E_2) = 0.3$ and $P(E_3) = P(E_4) = P(E_5)$, what is $P(E_5)$?

    (a) 0.1
    $\boxed{\text{(b)}}$ 0.2
    (c) 0.3
    (d) 0.4
    (e) 0.6

    Solution:
    c
    $P(E_1) = 3P(E_2) = 0.3 \implies P(E_2) = 0.1$ and so $P(E_3) + P(E_4) + P(E_5) = 1 - (0.3 + 0.1) = 0.6$ and since $P(E_3) = P(E_4) = P(E_5)$, $P(E_5) = 0.6/3 = 0.2$

19. In a statistics class, there are 60 men and 40 women. The mean weight of all these 100 students is 160 pounds. If the mean weight of the 60 men is 180 pounds, what is the mean weight of the 40 women?

    (a) 120 pounds
    (b) 125 pounds
    [✓ (c)] 130 pounds
    (d) 132 pounds
    (e) 135 pounds

    Solution:
    c
    $\frac{100\times160-60\times180}{40} = 130\text{lb}$

20. A business office orders paper supplies from one of three vendors, V1, V2, or V3. Orders are to be placed on three successive days (e.g. next Mon, Tue and Wed), one order per day. On each day, one of the three vendors is selected at random, giving each vendor the same probability (which is 1/3) to be selected independently of the other vendors.
    Use this information for this question and the next one.
    Find the probability that vendor V1 gets at least one order.

    (a) $\frac{1}{3}$
    (b) $\frac{1}{9}$
    (c) $\frac{8}{27}$
    [✓ (d)] $\frac{19}{27}$
    (e) $\frac{26}{27}$

    Solution:
    d
    $P(\text{ at least one V1}) = 1 - P(\text{ None is V1}) = 1 - P(\text{V1}^c)^3 = 1 - \left(\frac{2}{3}\right)^3 = 1 - \frac{8}{27} = \frac{19}{27}$

21. Same information as in the previous question.

A business office orders paper supplies from one of three vendors, V1, V2, or V3. Orders are to be placed on three successive days (e.g. next Mon, Tue and Wed), one order per day. On each day, one of the three vendors is selected at random, giving each vendor the same probability (which is 1/3) to be selected independently of the other vendors.

Find the probability that vendor V1 gets exactly two of these three orders.

(a) $\frac{1}{3}$

(b) $\frac{1}{9}$

[✓ (c)] $\frac{2}{9}$

(d) $\frac{2}{27}$

(e) $\frac{3}{27}$

Solution:

c

$P(\text{ exactly two V1}) = P(V1, V1, V1^c) + P(V1, V1^c, V1) + P(V1^c, V1, V1) = 3 \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right) = \frac{2}{9}$

Note: This question does not require the binomial tables or the binomial formula even though this can also be done using the binomial formula.

22. In a clinical study of a new drug, patients are randomly allocated to one of two groups. The patients in the first group are given an injection with the drug in it. The patients in the second group are given an injection, identical to the injection containing the drug, but with no drug in it. Doctors record the symptoms reported by the patients and take measurements, without knowing which patients are in which group. Data are then analysed to determine whether the drug is effective.
Which of the following statements is/are correct?

(I) If the groups end up different in terms of symptoms and measurements, then we know **for sure** that this difference is due to the administered drug, no matter how big or small the difference is.

(II) This is an example of a double-blind experiment.

(III) This is an example of an experiment where the units are assigned to treatments using a randomized block design.

(a) Only (I) is correct
[✓ (b)] Only (II) is correct
(c) Only (III) is correct
(d) Only (I) and (II) are correct
(e) Only (II) and (III) are correct

Solution:
(b) (I) not correct because the difference could be due to chance, specially if it is small. We have to check whether the difference is statistically significant; This is an example of a double-blind experiment where the units are assigned to groups using a completely randomized design, i.e. (II) is correct and (III) is not correct.

23. A randomly selected sample of 1000 college students was asked whether they had ever used the drug Ecstasy. Sixteen percent (16% or 0.16) of the 1000 students surveyed said they had.
Which of the following statements about the number 0.16 is/are correct?

(I) It is the sample proportion of college students who have ever used the drug Ecstasy.

(II) It is the population proportion of college students who have ever used the drug Ecstasy.

(III) It is a value that can be used to estimate the population proportion of college students who have ever used the drug Ecstasy.

(a) Only (I) is correct
(b) Only (II) is correct
(c) Only (III) is correct
$\boxed{\text{(d)}}$ Only (I) and (III) are correct
(e) Only (II) and (III) are correct

Solution:
(d) it is a value computed from the sample.

24. Suppose the correlation coefficient between height (measured in feet) and weight (measured in pounds) is 0.40. What is the correlation coefficient between height measured in inches and weight measured in ounces? The conversion is 12 inches = 1 foot and 16 ounces = 1 pound.

(a) 0.002
(b) 0.30
$\boxed{\text{(c)}}$ 0.40
(d) 0.533
(e) Impossible to say given the information provided

Solution:
(b)
The correlation coefficient has not units

25. A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

Use this information for this question and the next two questions.
One student in the sample was 73 inches tall with a foot length of 29 cm. What is the residual for this student?

(a) 29 cm

[✓ (b)] 1.31 cm

(c) 0 cm

(d) 27.69 cm

(e) -1.31 cm

Solution:
b
$29 - (10.9 + 0.23 \times 73) = 1.31$

26. Same information as in the previous question.
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

If we convert the heights of all these 33 students into cm using the conversion factor 1 inch = 2.54 cm and recalculate the regression equation with both the variables measured in cm, what will be the slope of the new regression equation?

[✓ (a)] 0.09

(b) 0.58

(c) 0.23

(d) 11.48

(e) 10.99

Solution:
a
$b = r\frac{s_y}{s_x}$. r and s_y don't change. s_x gets multiplied by 2.54 and so the new slope is $0.23/2.54 = 0.09055118$

27. Same information as in the previous two questions.
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

If we convert the foot lengths of all these 33 students into inches using the conversion factor 1 cm = 0.3937 inches and recalculate the regression equation with both the variables measured in inches, what will be the y-intercept of the new regression equation?

(a) 11.29

(b) 28.27

(c) 10.9

(d) 27.69

[✓ (e)] 4.29

Solution:
e
$a = \bar{y} - b\bar{x}, b = r\frac{s_y}{s_x}$. r and $\bar{x}$ don't change. $\bar{y}$,s_y and b get multiplied by 0.3937 and so the new y-intercapt is is $10.9 \times 0.3937 = 4.29133$

28. If the correlation coefficient r between two variables is equal to 1, what do we know **for sure**?

[✓ (a)] There is a perfect positive linear relationship between the two variables.

(b) There is a very strong, but not perfect, positive linear relationship between the two variables.

(c) There is no relationship between the two variables.

(d) There is a perfect negative linear relationship between the two variables.

(e) It is impossible to know one of the above for sure given the information provided.

Solution:
(a)

29. The probability that a randomly chosen American is a Republican is 0.35. We select a simple random sample of 6 Americans. What is the probability that at least 1 of the selected Americans is a Republican?

    (a) 0.0754

    (b) 0.2437

    (c) 0.3191

    (d) 0.3500

    $\boxed{\text{(e)}}$ 0.9246

    Solution:
    (e) 1 - 0.0754 = 0.9246

30. The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.
    Use this information for this question and the next one.
    If I select at random two Canada Post letters, what is the probability that they will both weigh less than 1 oz.?

    $\boxed{\text{(a)}}$ 0.0005

    (b) 0.0228

    (c) 0.0456

    (d) 0.4772

    (e) 0.8413

    Solution:
    (a) The probability that one letter weigh less than 1 oz. is $P(X < 1) = P(Z < -2) = 0.0228$. The probability that they will both weigh less than 1 oz is 0.0228^2 (multiplication rule for independent events).

31. Same information as in the previous question.
The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.
If I select at random 1000 Canada Post letters, what is the probability that at least 200 will weigh less than 1 oz.?

[✓ (a)] 0.0000
(b) 0.0046
(c) 0.1040
(d) 0.8960
(e) 1.0000

Solution:
(a) The number of letters that weigh less than 1 oz follows a $B(n = 1000, p = 0.0228)$. We can use the normal approximation since $np = 22.8$ and $n(1 - p) = 977.2$. The probability that at least 200 letters will weigh less than 1 oz. is $P(X >= 200) = P(Z >= (200 - 22.8)/\sqrt{22.280}) = 1 - P(Z < 37.54) = 0$

32. A city council of a small city wants to know the proportion of eligible voters that oppose having a garbage incinerator opened just outside the city limits. They randomly select 100 residential numbers from the city's telephone book that contains 3,000 such numbers. Each selected residence is then called and asked to provide the total number of eligible voters in the residence and the number of voters opposed to the incinerator in the residence.
What sampling method is used?

(a) Simple random sampling
(b) Stratified sampling
[✓ (c)] Cluster sampling
(d) Multistage sampling
(e) Systematic sampling

Solution:
(c) Clusters = Residences

33. Acme Medicine is conducting an experiment to test a new vaccine, developed to immunize people against the common cold. To test the vaccine, Acme has 1000 participants. Participants are divided into two groups, based on gender. Then, within each group, half of the participants are randomly assigned to treatment groups so that the first half gets the placebo and the second half gets the vaccine.
Which of the following statements is correct?

(a) This is an example of a prospective observational study.

$\boxed{\text{(b)}}$ This is an example of an experiment where the units are assigned to treatments using a randomized block design.

(c) This is an example of an experiment where the units are assigned to treatments using a completely randomized design.

(d) This is an example of a census.

(e) This is an example of a survey where the participants are selected using stratified sampling.

Solution:
(b)

34. A survey of the male students at a junior college reveals that 26% play soccer regularly, 22% are Latino, and half of the Latino students play soccer regularly. If a male student is selected at random, what is the probability that he is neither Latino nor a soccer player?

(a) 0.89

(b) 0.52

$\boxed{\text{(c)}}$ 0.63

(d) 0.26

(e) 0.41

Solution:
c
P(So) = 0.26, P(L) = 0.22, P(So|L) = 0.5 and so P(So and L) = 0.22×0.5 = 0.11
P(L or So) = P(L) + P(So) - P(L and So) = 0.26 +0.22-0.11 = 0.37
P(neither Latino nor a soccer player) = $1 - 0.37 = 0.63$

35. The table below gives the probability distribution of the number of times (x) a photocopying machine needs repair in a given month.

x	0	1	2	3	4	5
Probability	0.12	0.18	0.23	0.20	0.14	0.13

What is the probability that the machine will have to be repaired at least twice but not more than four times in the given month?

(a) 0.23
(b) 0.24
(c) 0.43
$\boxed{\text{(d)}}$ 0.57
(e) 0.87

Solution:
d
$P(2 \leq X \leq 4) = 0.23 + 0.20 + 0.14 = 0.57$

36. Students in an Intro Stats course were asked to describe their political preference as Liberal, Moderate, or Conservative. Here are the results:

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total	85	80	27	192

Use this information for this question and the next one.
We select a student at random. Let L denote the event that the student selected considers himself/herself to be liberal and let F denote the event that the student selected is female. Calculate $P(L|F)$.

(a) 0.18

(b) 0.25

(c) 0.35

(d) 0.41

[✓ (e)] 0.45

Solution:
e
$P(L|F) = \frac{35}{77} = 0.4545455$

37. Same information as in the previous question.
Students in an Intro Stats course were asked to describe their political preference as Liberal, Moderate, or Conservative. Here are the results:

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total	85	80	27	192

We select a student at random. Let L denote the event that the student selected considers himself/herself to be liberal and let M denote the event that the student selected is male. Calculate $P(L \text{ or } M)$.

(a) 0.18

(b) 0.27

(c) 0.45

[✓ (d)] 0.78

(e) 1.04

Solution:
c
$P(L \text{ or } M) = P(L) + P(M) - P(L \text{ and } M) = \frac{85}{192} + \frac{115}{192} - \frac{50}{192} = \frac{150}{192} = 0.78125$

38. Using a random sample of 4,000 students, you compute a 95% confidence interval for the proportion of overweight students. You decide to compute another 95% confidence interval using a different sample, this time with only 1,000 students. Suppose that the sample proportion $\widehat{p}$ is the same for both these confidence intervals. What change would you expect from the first confidence interval to the second?

    (a) The margin of error of the second interval will be 4 times as big as the margin of error of the first interval.

    $\boxed{\text{(b)}}$ The margin of error of the second interval will be 2 times as big as the margin of error of the first interval.

    (c) The margin of error of the first interval will be 4 times as big as the margin of error of the second interval.

    (d) The margin of error of the first interval will be 2 times as big as the margin of error of the second interval.

    (e) There will be no change in the margin of error since the sample proportion $\widehat{p}$ is the same for both confidence intervals.

    Solution:
    (b)
    The margin of error of the second interval is:

$$1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{1000}} = 2 \cdot 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{4000}}$$

    which is 2 times the margin of error of the first interval.

39. The heights of students in some population have a Normal distribution with mean 172cm. Twenty five percent of the students in this population have heights greater than or equal to 180cm. What is the standard deviation of the heights of the students in this population?

    (a) 7.9cm

    (b) 8.7cm

    (c) 9.9cm

    (d) 10.5cm

    $\boxed{\text{(e)}}$ 11.9cm

    Solution:
    e
    $P(X \geq 180) = 0.25 \implies z = 0.67$ and $\frac{180-172}{\sigma} = 0.67 \implies \sigma = \frac{180-172}{0.67} = 11.94$

40. In a game, a wheel with five equal-sized spaces labeled from A to E is spun and a ball randomly lands in one of these spaces. If the ball lands in the space labeled A, the player wins 15 dollars. If the ball lands in any other space, the winner loses 4 dollars. Use this information for this questions and the next one.
A player plays this game 6 times. What is the probability that the ball lands at least one time in the space labeled A? You can assume that games are independent from one another.

    (a) 0.2622
    (b) 0.3446
    (c) 0.3932
    (d) 0.6068
    $\boxed{\text{(e)}}$ 0.7378

    Solution:
    X = number of times the ball lands in a space labeled A
    $X \sim Binom(n = 6, p = 0.2)$
    $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.2622 = 0.7378$

41. Same information as in the previous question.
In a game, a wheel with five equal-sized spaces labeled from A to E is spun and a ball randomly lands in one of these spaces. If the ball lands in the space labeled A, the player wins 15 dollars. If the ball lands in any other space, the winner loses 4 dollars. What is the expected gain or loss of a player who plays 40 games? You can assume that games are independent from one another.

    (a) Gain of 360 dollars
    (b) Gain of 55 dollars
    (c) Gain of 8 dollars
    (d) Loss of 1 dollar
    $\boxed{\text{(e)}}$ Loss of 8 dollars

    Solution:
    Expected loss or gain for one game: $0.2 \cdot 15 + 0.8 \cdot (-4) = -0.2$
    Expected loss or gain for 40 games (games are independent): $40 \cdot (-0.2) = -8$

42. Five hundred people used a home test for a HIV. We say that the test is positive if it indicates that the person has HIV, and we say it is negative if it indicates that the person does not have HIV. Out of these 500 people, 40 had HIV and 460 did not have HIV. The table below shows the contingency table of the result of the test (positive or negative) and whether a person has HIV (HIV, no HIV) for these 500 people.

	HIV	No HIV	Total
Positive Test	35	25	60
Negative Test	5	435	440
Total	40	460	500

A false positive is a test result that indicates a person has a disease when the person actually does not have it. What is the probability of a false positive with the home test for a HIV?

$\boxed{\text{(a)}}$ 0.054

(b) 0.070

(c) 0.130

(d) 0.417

(e) 0.875

Solution:
(a) 25/460

43. A consumer organization estimates that 34% of the households in a particular community have one television set, 39% have two sets, and 20% have three or more sets. What is the probability that a household chosen at random in this community has no more than one television set?

$\boxed{\text{(a)}}$ 0.41

(b) 0.59

(c) 0.07

(d) 0.34

(e) 0.46

Solution:
a
P(no more than 1) = 1- P(more than 1) = 1- P(2)- P(3) = 1-0.39-0.20 = 0.41

44. A randomly selected sample of 1000 college students was asked whether they were in favor of longer hours at the school library. Eighty percent (80%) of the 1000 students surveyed said they were in favor.
Which of the following statements is/are correct?

(I) A 95% confidence interval for the population proportion of college students who were in favor of longer hours at the school library is $[0.78; 0.82]$

(II) The population proportion of college students who were in favor of longer hours at the school library is significantly different from 76% at the level of significance $\alpha = 5\%$.

(III) The population proportion of college students who were in favor of longer hours at the school library is significantly different from 76% at the level of significance $\alpha = 1\%$.

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

(d) Only (I) and (II) are correct

$\boxed{\text{(e)}}$ The three statements (I), (II), and (III) are correct

Solution:
(e)
A 95% confidence interval is $0.8 \pm 1.96\sqrt{0.2 \cdot 0.8/1000} = [0.78; 0.82]$. The value of the test statistic is z-obs $= 2.96$ and the p-value is $2 \cdot 0.0015 = 0.003$.

45. After menopause, some women take supplemental oestrogen. There is some concern that if these women also drink alcohol, their oestrogen levels will rise too high. Twelve volunteers who were receiving supplemental oestrogen were randomly divided into two groups, as were 12 other volunteers who were not receiving supplemental oestrogen. In each case, one group drank an alcoholic beverage, the other a non-alcoholic beverage. An hour later, everyone's oestrogen level was checked. Only those on supplemental oestrogen who drank alcohol showed a marked increase in oestrogen levels.
Use this information for this questions and the next one.
What is the response variable in this study?

(a) "Menopause" is the response variable

(b) "Amount of supplemental oestrogen received" is the response variable

(c) "Type of beverage" (alcoholic beverage, non-alcoholic beverage) is the response variable

$\boxed{\text{(d)}}$ "Oestrogen level" is the response variable

(e) "Gender" is the response variable

Solution:
(d)

46. Same information as in the previous question.
After menopause, some women take supplemental oestrogen. There is some concern that if these women also drink alcohol, their oestrogen levels will rise too high. Twelve volunteers who were receiving supplemental oestrogen were randomly divided into two groups, as were 12 other volunteers who were not receiving supplemental oestrogen. In each case, one group drank an alcoholic beverage, the other a non-alcoholic beverage. An hour later, everyone's oestrogen level was checked. Only those on supplemental oestrogen who drank alcohol showed a marked increase in oestrogen levels.
Which of the following statements is/are correct?

(I) This is an example of an experiment where the units are assigned to treatments using a randomized block design

(II) "Amount of supplemental oestrogen received" is a blocking variable

(III) This is an example of an observational study since patients were not randomly assigned to treatments

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

$\boxed{\text{(d)}}$ Only (I) and (II) are correct

(e) None of the three statements is correct

Solution:
(d)

47. A box contains 3 yellow, 2 red, 4 green and 3 black marbles. Two marbles are taken one after the other at random from the box without replacement. What is the probability that both marbles are red?

(a) $\frac{1}{50}$

(b) $\frac{1}{60}$

$\boxed{\text{(c)}}$ $\frac{1}{66}$

(d) $\frac{1}{24}$

(e) $\frac{1}{18}$

Solution:
c
Let $R1$ be the event that the first marble selected is red and $R2$ be the event that the second is red.
We want $P(R1 \text{ and } R2)$
Using general multiplication rule we have,
$P(R1 \text{ and } R2) = P(R1) \times P(R2|R1) = \frac{2}{3+2+4+3} \times \frac{2-1}{3+2+4+3-1} = \frac{2}{12} \times \frac{1}{11} = \frac{1}{66}$

48. A local bakery has determined a probability distribution for the number of cheesecakes it sells in a given day. The distribution is as follows:

Number sold in a day	10	20	30	35	x
Probability	0.1	0.5	0.1	0.2	0.1

If the mean number of cheesecakes that this local bakery sells in a day is 25, what is the value of x?

(a) 30

(b) 35

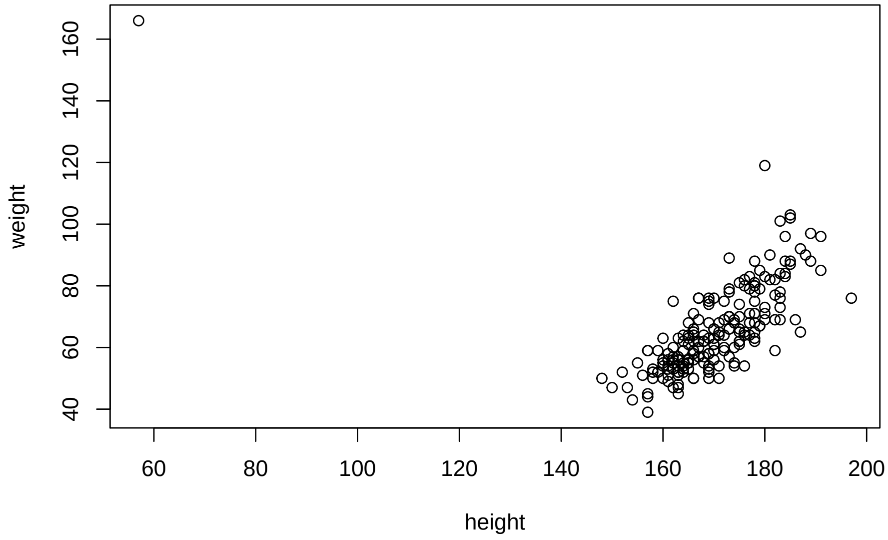
(c) 38

$\boxed{\text{(d)}}$ 40

(e) 45

Solution:
d
$(10 \times 0.1) + (20 \times 0.5) + (30 \times 0.1) + (35 \times 0.2) + (x \times 0.1) = 25 \implies x \times 0.1 = 25 - 21 \implies x = \frac{4}{0.1} = 40$

49. The scatterplot below shows the self-reported height and weight of 200 subjects. The subjects were men and women engaged in regular exercise. Suppose we fit a linear regression of weight on height.
Use this scatterplot for this question and the next one.



Which of the following statements about the data point with height = 57 is/are true?

(I) The data point with height = 57 is an outlier.

(II) The data point with height = 57 is a high leverage point.

(III) The data point with height = 57 is an influential observation.

(a) Only (I) is true

(b) Only (II) is true

(c) Only (III) is true

$\boxed{\text{(d)}}$ All three statements (I), (II) and (III) are true

(e) None of the three statements is true

Solution:
(d)

50. Use the scatterplot of the previous question to answer this question.
Suppose we fit a linear regression of weight on height.
Which of the following statements is known **for sure**?

(I) If we remove the data point with height = 57, the slope of the regression line increases.

(II) If we remove the data point with height = 57, the mean height increases.

(III) If we remove the data point with height = 57, the variance of height increases.

(a) Only (II) is known for sure
[✓ (b)] Only (I) and (II) are known for sure
(c) Only (I) and (III) are known for sure
(d) Only (II) and (III) are known for sure
(e) None of the three statements is known for sure

Solution:
(b)
If we remove the data point with height = 57, the slope of the regression line increases (I true), the mean height increases (II true), the variance of height decreases (III not true).

Tables of normal, binomial and t-distributions

prepared by Ken Butler

2015-06-03

Table Z (normal distribution)

Values of z greater than 0 are on the next page. Second decimal place is at the top of the column.

	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
-3.3	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005
-3.2	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007
-3.1	0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010
-3.0	0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013
-2.9	0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019
-2.8	0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026
-2.7	0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035
-2.6	0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047
-2.5	0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062
-2.4	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
-2.3	0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
-2.2	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139
-2.1	0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179
-2.0	0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228
-1.9	0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287
-1.8	0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359
-1.7	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
-1.6	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
-1.5	0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668
-1.4	0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808
-1.3	0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968
-1.2	0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151
-1.1	0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357
-1.0	0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587
-0.9	0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841
-0.8	0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119
-0.7	0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420
-0.6	0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743
-0.5	0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085
-0.4	0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446
-0.3	0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821
-0.2	0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207
-0.1	0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602
0.0	0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000

Table Z (continued)

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table of the Binomial Distribution

(Table shows probability of exactly k successes in n trials with success probability p. Values of p are shown along the top of the table.)

		p									
n	k	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
2	1	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
2	2	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
3	1	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
3	2	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
3	3	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
4	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
4	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
4	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
4	4	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
5	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1563
5	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
5	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1812	0.2304	0.2757	0.3125
5	4	0.0000	0.0004	0.0022	0.0064	0.0147	0.0284	0.0488	0.0768	0.1128	0.1563
5	5	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0313
6	0	0.7351	0.5315	0.3772	0.2622	0.1780	0.1177	0.0754	0.0467	0.0277	0.0156
6	1	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0938
6	2	0.0306	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344
6	3	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125
6	4	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344
6	5	0.0000	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938
6	6	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156

Computer & Mathematical Sciences
UNIVERSITY OF TORONTO
SCARBOROUGH

STAB22: Final Exam April 19, 2018

Complete this form, then *carefully* tear off this page and place it along with your student card on the surface beside where you are writing.

The form will be verified and collected during the test.

Information collected will be used solely for the purpose of administering this test.

Name (please print): ____

Signature: ___________________________ Seat #: _______________
(if available)

Student Number: ☐☐☐☐☐☐☐☐☐☐

UNIVERSITY OF TORONTO SCARBOROUGH
Department of Computer and Mathematical Sciences
April 19, 2018
APRIL 2018 FINAL EXAMINATIONS

STAB22H3 Statistics I
Duration: 3 hours

Last Name:________________________ First Name: ________________________

Student number: ________________________

Aids allowed:

- Two handwritten letter-sized sheets (both sides) of notes prepared by you
- Non-programmable, non-communicating calculator

The table of normal and binomial distributions are attached at the end.

This test is based on multiple-choice questions. There are 50 questions. All questions carry equal weight. Report all your answers on the Scantron answer sheet before the end of the exam. On the Scantron answer sheet, ensure that you write down and **BUBBLE** your last name, first name (as much of it as fits), and student number (in "Identification").

Mark in each case the best answer out of the alternatives given (**which means the numerically closest answer if the answer is a number and the answer you obtained is not given.**)

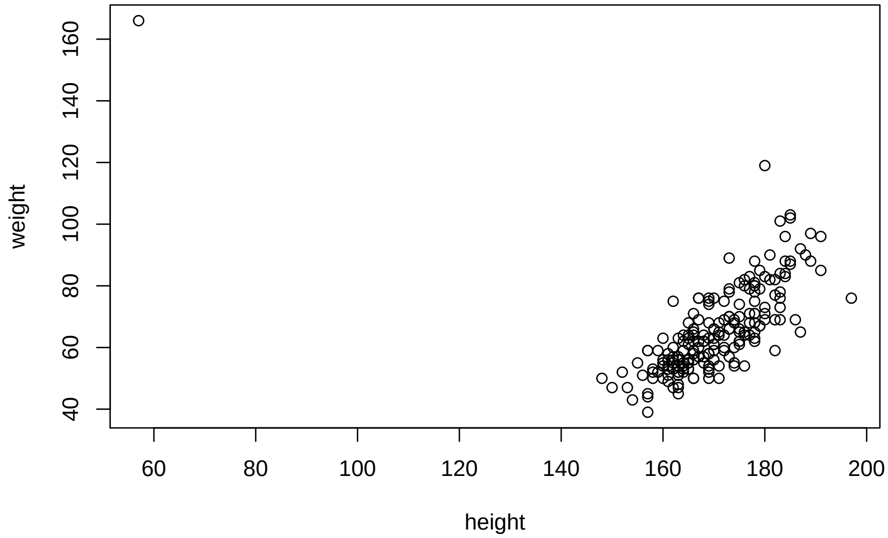
Also before you begin, complete the signature sheet, but sign it only when the invigilator collects it. The signature sheet shows that you were present at the exam.

There are 30 pages including this page and statistical tables. Please check to see that you have all the pages.

Good luck!!

ExamVersion: B

1. The scatterplot below shows the self-reported height and weight of 200 subjects. The subjects were men and women engaged in regular exercise. Suppose we fit a linear regression of weight on height.
Use this scatterplot for this question and the next one.



Which of the following statements about the data point with height = 57 is/are true?

(I) The data point with height = 57 is an outlier.

(II) The data point with height = 57 is a high leverage point.

(III) The data point with height = 57 is an influential observation.

(a) Only (I) is true

(b) Only (II) is true

(c) Only (III) is true

(d) All three statements (I), (II) and (III) are true

(e) None of the three statements is true

Solution:
(d)

2. Use the scatterplot of the previous question to answer this question.
   Suppose we fit a linear regression of weight on height.
   Which of the following statements is known **for sure**?

   (I) If we remove the data point with height = 57, the slope of the regression line increases.

   (II) If we remove the data point with height = 57, the mean height increases.

   (III) If we remove the data point with height = 57, the variance of height increases.

   (a) Only (II) is known for sure
   (b) Only (I) and (II) are known for sure
   (c) Only (I) and (III) are known for sure
   (d) Only (II) and (III) are known for sure
   (e) None of the three statements is known for sure

   Solution:
   (b)
   If we remove the data point with height = 57, the slope of the regression line increases (I true), the mean height increases (II true), the variance of height decreases (III not true).

3. The tail length for a population of female hook-billed kites is Normally distributed with mean 194 mm and standard deviation 11 mm. We select a random sample of 200 kites from this population. Let X be the number of kites in this sample having a tail length between 179 and 209 mm. What is the mean of X?

   (a) 145
   (b) 155
   (c) 165
   (d) 175
   (e) 194

   Solution:
   c
   Letting L denote the tail length of a randomly selected female hook-billed kite, $X \sim Bin(n = 300, p = P(179 < L < 209))$.
   $p = P(179 < L < 209) = P(\frac{179-194}{11} < Z < \frac{209-194}{11}) = P(-1.36Z < 1.36) = 1 - 2 \times 0.0869 = 1 - 0.1738 = 0.8262$
   $\mu_X = np = 200 \times 0.8262 = 165.24$

4. The heights of students in some population have a Normal distribution with mean 172cm. Twenty five percent of the students in this population have heights greater than or equal to 180cm. What is the standard deviation of the heights of the students in this population?

   (a) 7.9cm
   (b) 8.7cm
   (c) 9.9cm
   (d) 10.5cm
   (e) 11.9cm

   Solution:
   e
   $P(X \geq 180) = 0.25 \implies z = 0.67$ and $\frac{180-172}{\sigma} = 0.67 \implies \sigma = \frac{180-172}{0.67} = 11.94$

5. Childhood lead poisoning is a public health concern in a country. In this country, 1 child in 10 has a high blood lead level. In a randomly chosen group of 6 children from the population of this country, what is the probability that 3 or more have high blood lead level?

   (a) 0.0146
   (b) 0.5000
   (c) 0.6000
   (d) 0.0159
   (e) 0.9987

   Solution:
   d
   Let X = number of children with high blood level in this sample. Then $X \sim Bin(n = 6, p = 1/10 = 0.1)$. $P(X \geq 3) = 0.0146 + 0.0012 + 0.0001 + 0.0000 = 0.0159$

6. Using a random sample of 4,000 students, you compute a 95% confidence interval for the proportion of overweight students. You decide to compute another 95% confidence interval using a different sample, this time with only 1,000 students. Suppose that the sample proportion $\widehat{p}$ is the same for both these confidence intervals. What change would you expect from the first confidence interval to the second?

   (a) The margin of error of the second interval will be 4 times as big as the margin of error of the first interval.

   (b) The margin of error of the second interval will be 2 times as big as the margin of error of the first interval.

   (c) The margin of error of the first interval will be 4 times as big as the margin of error of the second interval.

   (d) The margin of error of the first interval will be 2 times as big as the margin of error of the second interval.

   (e) There will be no change in the margin of error since the sample proportion $\widehat{p}$ is the same for both confidence intervals.

   Solution:
   (b)
   The margin of error of the second interval is:

$$1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{1000}} = 2 \cdot 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{4000}}$$

   which is 2 times the margin of error of the first interval.

7. A random experiment consists of five possible outcomes: E_1, E_2, E_3, E_4 and E_5. This means the sample space is $S = \{E_1, E_2, E_3, E_4, E_5\}$. If $P(E_1) = 3 \times P(E_2) = 0.3$ and $P(E_3) = P(E_4) = P(E_5)$, what is $P(E_5)$?

   (a) 0.1

   (b) 0.2

   (c) 0.3

   (d) 0.4

   (e) 0.6

   Solution:
   c
   $P(E_1) = 3P(E_2) = 0.3 \implies P(E_2) = 0.1$ and so $P(E_3) + P(E_4) + P(E_5) = 1-(0.3+0.1) = 0.6$ and since $P(E_3) = P(E_4) = P(E_5)$, $P(E_5) = 0.6/3 = 0.2$

8. A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

Use this information for this question and the next two questions.
One student in the sample was 73 inches tall with a foot length of 29 cm. What is the residual for this student?

(a) 29 cm

(b) 1.31 cm

(c) 0 cm

(d) 27.69 cm

(e) -1.31 cm

Solution:
b
$29 - (10.9 + 0.23 \times 73) = 1.31$

9. Same information as in the previous question.
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

If we convert the heights of all these 33 students into cm using the conversion factor 1 inch = 2.54 cm and recalculate the regression equation with both the variables measured in cm, what will be the slope of the new regression equation?

(a) 0.09

(b) 0.58

(c) 0.23

(d) 11.48

(e) 10.99

Solution:
a
$b = r\frac{s_y}{s_x}$. r and s_y don't change. s_x gets multiplied by 2.54 and so the new slope is $0.23/2.54 = 0.09055118$

10. Same information as in the previous two questions.
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

If we convert the foot lengths of all these 33 students into inches using the conversion factor 1 cm = 0.3937 inches and recalculate the regression equation with both the variables measured in inches, what will be the y-intercept of the new regression equation?

(a) 11.29

(b) 28.27

(c) 10.9

(d) 27.69

(e) 4.29

Solution:
e
$a = \bar{y} - b\bar{x}, b = r\frac{s_y}{s_x}$. r and $\bar{x}$ don't change. $\bar{y}$,s_y and b get multiplied by 0.3937 and so the new y-intercapt is is $10.9 \times 0.3937 = 4.29133$

11. A traditional die is a cube, each face showing an integer in the range from 1 to 6. A **fair** traditional die is rolled 3 times. The first 2 rolls result in the face showing "5" being uppermost. What is the probability the third roll will **not** result in the face showing "5" being uppermost?

(a) 0

(b) 0.023

(c) 0.069

(d) 0.833

(e) 1

Solution:
(d) The rolls are independent. The probability the third roll will not result in the face with "5" being uppermost is $1 - 1/6 = 5/6 = 0.833$.

12. A business office orders paper supplies from one of three vendors, V1, V2, or V3. Orders are to be placed on three successive days (e.g. next Mon, Tue and Wed), one order per day. On each day, one of the three vendors is selected at random, giving each vendor the same probability (which is 1/3) to be selected independently of the other vendors.

    Use this information for this question and the next one.

    Find the probability that vendor V1 gets at least one order.

    (a) $\frac{1}{3}$

    (b) $\frac{1}{9}$

    (c) $\frac{8}{27}$

    (d) $\frac{19}{27}$

    (e) $\frac{26}{27}$

    Solution:
    d
    $P(\text{ at least one V1}) = 1 - P(\text{ None is V1}) = 1 - P(\text{V1}^c)^3 = 1 - \left(\frac{2}{3}\right)^3 = 1 - \frac{8}{27} = \frac{19}{27}$

13. Same information as in the previous question.

    A business office orders paper supplies from one of three vendors, V1, V2, or V3. Orders are to be placed on three successive days (e.g. next Mon, Tue and Wed), one order per day. On each day, one of the three vendors is selected at random, giving each vendor the same probability (which is 1/3) to be selected independently of the other vendors.

    Find the probability that vendor V1 gets exactly two of these three orders.

    (a) $\frac{1}{3}$

    (b) $\frac{1}{9}$

    (c) $\frac{2}{9}$

    (d) $\frac{2}{27}$

    (e) $\frac{3}{27}$

    Solution:
    c
    $P(\text{ exactly two V1}) = P(V1, V1, V1^c) + P(V1, V1^c, V1) + P(V1^c, V1, V1) = 3 \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right) = \frac{2}{9}$
    Note: This question does not require the binomial tables or the binomial formula even though this can also be done using the binomial formula.

14. A randomized experiment was done by randomly assigning each participant either to walk for half an hour three times a week or to sit quietly reading a book for half an hour three times a week. The participants' blood pressure was measured at the beginning and at the end of the study and the change in participants' blood pressure was computed. The change in participants' blood pressure was compared for the two groups.
This is a randomized experiment rather than an observational study because:

    (a) blood pressure was measured at the beginning and end of the study.
    (b) the two groups were compared at the end of the study.
    (c) the participants were randomly assigned to either walk or read, rather than choosing their own activity.
    (d) a random sample of participants was used.
    (e) the study was done by scientists.

    Solution:
    c

15. If there is a very strong linear relationship between two variables then what do you know **for sure** about the correlation coefficient r?

    (a) The correlation coefficient is larger than 1
    (b) The correlation coefficient is close to 0
    (c) The correlation coefficient is close to -1
    (d) The correlation coefficient is close to 1
    (e) None of the above is known for sure.

    Solution:
    (e) We know for sure that the correlation coefficient is close to -1 or 1.

16. In a game, a wheel with five equal-sized spaces labeled from A to E is spun and a ball randomly lands in one of these spaces. If the ball lands in the space labeled A, the player wins 15 dollars. If the ball lands in any other space, the winner loses 4 dollars. Use this information for this questions and the next one.
A player plays this game 6 times. What is the probability that the ball lands at least one time in the space labeled A? You can assume that games are independent from one another.

    (a) 0.2622
    (b) 0.3446
    (c) 0.3932
    (d) 0.6068
    (e) 0.7378

    Solution:
    X = number of times the ball lands in a space labeled A
    $X \sim Binom(n = 6, p = 0.2)$
    $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.2622 = 0.7378$

17. Same information as in the previous question.
In a game, a wheel with five equal-sized spaces labeled from A to E is spun and a ball randomly lands in one of these spaces. If the ball lands in the space labeled A, the player wins 15 dollars. If the ball lands in any other space, the winner loses 4 dollars. What is the expected gain or loss of a player who plays 40 games? You can assume that games are independent from one another.

    (a) Gain of 360 dollars
    (b) Gain of 55 dollars
    (c) Gain of 8 dollars
    (d) Loss of 1 dollar
    (e) Loss of 8 dollars

    Solution:
    Expected loss or gain for one game: $0.2 \cdot 15 + 0.8 \cdot (-4) = -0.2$
    Expected loss or gain for 40 games (games are independent): $40 \cdot (-0.2) = -8$

18. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\widehat{y} = -0.0127 + 0.0180x.$$

Use this information for this question and the next one.
The above equation implies that:

(a) Each beer consumed increases blood alcohol content by 0.0127

(b) Each beer consumed decreases blood alcohol content by 0.0127

(c) On average it takes 1.8 beers to increase blood alcohol content by 0.01

(d) Each beer consumed increases blood alcohol content by an average of amount of 0.018

(e) Each beer consumed increases blood alcohol content by exactly 0.018

Solution:
(d)

19. Same information as in the previous question.
The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\widehat{y} = -0.0127 + 0.0180x.$$

Suppose that the legal upper limit to drive is a blood alcohol content of 0.08.
Which of the following statements is/are correct?

(I) The model predicts that a driver who drinks 5 beers has a blood content below the legal upper limit to drive.

(II) The model predicts that a driver who drinks 6 beers has a blood content below the legal upper limit to drive.

(III) A driver who drinks 5 beers has **for sure** a blood content that is below the legal upper limit to drive.

(a) Only statement (I) is correct

(b) Only statement (II) is correct

(c) Only statements (I) and (II) are correct

(d) Only statements (I) and (III) are correct

(e) All three statements (I), (II), and (III) are correct

Solution:
(a)
Solving $0.08 = -0.0127 + 0.0180x$ yields $x = 5.15$.

20. A consumer organization estimates that 34% of the households in a particular community have one television set, 39% have two sets, and 20% have three or more sets. What is the probability that a household chosen at random in this community has no more than one television set?

    (a) 0.41
    (b) 0.59
    (c) 0.07
    (d) 0.34
    (e) 0.46

    Solution:
    a
    P(no more than 1) = 1- P(more than 1) = 1- P(2)- P(3) = 1-0.39-0.20 = 0.41

21. The following table represents the probability distribution of the number of accidents per day in a city.

Number of accidents	0	1	2	3	4 or more
Probability	0.55	0.20	0.10	0.15	0

Use this table for this question and the next one.
Which of the following statements about the number of accidents per day in the city is/are true?

(I) The mean number of accidents per day is larger than the median number of accidents per day.

(II) The mean number of accidents per day is smaller than the median number of accidents per day.

(III) The median number of accidents per day is equal to 1.

(a) Only (I) is true

(b) Only (II) is true

(c) Only (III) is true

(d) Only (I) and (II) are true

(e) Only (I) and (III) are true

Solution:
(a)
Right skewed distribution / median is 0 / mean is 0.85

22. Use the probability distribution of the previous question to answer this question.
What is the variance of the number of accidents per day in the city?

(a) 1.23

(b) 1.67

(c) 2.07

(d) 2.50

(e) 3.10

Solution:
(a)
$\mu = 0.85$
$\sigma^2 = \sum(X - 0.85)^2 P(X) = 1.2275$

23. Acme Medicine is conducting an experiment to test a new vaccine, developed to immunize people against the common cold. To test the vaccine, Acme has 1000 participants. Participants are divided into two groups, based on gender. Then, within each group, half of the participants are randomly assigned to treatment groups so that the first half gets the placebo and the second half gets the vaccine.
Which of the following statements is correct?

   (a) This is an example of a prospective observational study.
   (b) This is an example of an experiment where the units are assigned to treatments using a randomized block design.
   (c) This is an example of an experiment where the units are assigned to treatments using a completely randomized design.
   (d) This is an example of a census.
   (e) This is an example of a survey where the participants are selected using stratified sampling.

   Solution:
   (b)

24. A randomly selected sample of 1000 college students was asked whether they were in favor of longer hours at the school library. Eighty percent (80%) of the 1000 students surveyed said they were in favor.
Which of the following statements is/are correct?

   (I) A 95% confidence interval for the population proportion of college students who were in favor of longer hours at the school library is $[0.78; 0.82]$
   (II) The population proportion of college students who were in favor of longer hours at the school library is significantly different from 76% at the level of significance $\alpha = 5\%$.
   (III) The population proportion of college students who were in favor of longer hours at the school library is significantly different from 76% at the level of significance $\alpha = 1\%$.

   (a) Only (I) is correct
   (b) Only (II) is correct
   (c) Only (III) is correct
   (d) Only (I) and (II) are correct
   (e) The three statements (I), (II), and (III) are correct

   Solution:
   (e)
   A 95% confidence interval is $0.8 \pm 1.96\sqrt{0.2 \cdot 0.8/1000} = [0.78; 0.82]$. The value of the test statistic is z-obs $= 2.96$ and the p-value is $2 \cdot 0.0015 = 0.003$.

25. STA101 and STA102 are two statistics courses both with large classes. We have the following information about the distributions of grades for these two courses:

    - The distribution of grades of STA101 is right skewed.
    - The distribution of grades of STA102 is left skewed.
    - STA101 has lower class average than STA102 (i.e. $\mu_1 < \mu_2$, where μ_1 and μ_2 are the means of the distributions of grades in STA101 and STA102 respectively.)

    Which of the following statements regarding the distributions of grades of these two courses is **NOT** true?

    (a) The median grade of STA101 is less than the median grade of STA102.
    (b) The mean grade of STA102 is less than the median grade of STA101.
    (c) The mean grade of STA101 is less than the median grade of STA102.
    (d) The median grade of STA101 is less than the mean grade of STA102.
    (e) The mean grade of STA102 is less than the median grade of STA102.

Solution:
b
Denoting the medians of STA101 and STA102 by m_1 and m_2 respectively, we have,
$m_1 < \mu_1$, since the distribution of grades of STA101 is right skewed,
$\mu_2 < m_2$, since the distribution of grades of STA102 is left skewed,
and so since we are also given that $\mu_1 < \mu_2$, we have
$m_1 < \mu_1 < \mu_2 < m_2$ and this implies:

$m_1 < m_2$, i.e. (a) is true
$m_1 < \mu_2$, i.e. (b) is false
$\mu_1 < m_2$, i.e. (c) is true
$m_1 < \mu_2$, i.e. (d) is true
$\mu_2 < m_2$, i.e. (e) is true

26. In a clinical study of a new drug, patients are randomly allocated to one of two groups. The patients in the first group are given an injection with the drug in it. The patients in the second group are given an injection, identical to the injection containing the drug, but with no drug in it. Doctors record the symptoms reported by the patients and take measurements, without knowing which patients are in which group. Data are then analysed to determine whether the drug is effective.
Which of the following statements is/are correct?

(I) If the groups end up different in terms of symptoms and measurements, then we know **for sure** that this difference is due to the administered drug, no matter how big or small the difference is.

(II) This is an example of a double-blind experiment.

(III) This is an example of an experiment where the units are assigned to treatments using a randomized block design.

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

(d) Only (I) and (II) are correct

(e) Only (II) and (III) are correct

Solution:
(b) (I) not correct because the difference could be due to chance, specially if it is small. We have to check whether the difference is statistically significant; This is an example of a double-blind experiment where the units are assigned to groups using a completely randomized design, i.e. (II) is correct and (III) is not correct.

27. A local bakery has determined a probability distribution for the number of cheesecakes it sells in a given day. The distribution is as follows:

Number sold in a day	10	20	30	35	x
Probability	0.1	0.5	0.1	0.2	0.1

If the mean number of cheesecakes that this local bakery sells in a day is 25, what is the value of x?

(a) 30

(b) 35

(c) 38

(d) 40

(e) 45

Solution:
d
$(10 \times 0.1) + (20 \times 0.5) + (30 \times 0.1) + (35 \times 0.2) + (x \times 0.1) = 25 \implies x \times 0.1 = 25 - 21 \implies x = \frac{4}{0.1} = 40$

28. In a large population, 60% of the voters support a particular party. If a random sample of 300 voters were selected from this population, what would be the approximate probability that more than 175 of them would support this party?

(a) 0.35

(b) 0.72

(c) 0.96

(d) 0.51

(e) 0.60

Solution:
b
$X \sim Bin(n = 300, p = 0.6) \overset{approx}{\sim} N(\mu = np = 300 \times 0.6 = 180, \sigma = \sqrt{np(1-p)}) = \sqrt{300 \times 0.6 \times 0.4}) = 8.49.$
$P(X \leq 175) = P(Z < \frac{175-180}{8.49} = -0.59) = 0.2776.$
$P(X > 175) = 1 - P(X \leq 175) = 1 - 0.2776 = 0.7224$

29. In which of the following cases is the normal distribution a reasonable approximation of the sampling distribution of a proportion?

    (I) The probability of an airline flight arriving on time is 90%. The airline operates more than 500 flights per day. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of flights arriving on time in a simple random sample of 30 flights of this airline.

    (II) In a large country 0.2% of the population is affected by a disease. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of individuals affected by the disease in a simple random sample of 2000 individuals in this country.

    (III) A company has 150 workers. The proportion of workers who smoke in this company is 40%. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of workers who smoke in a simple random sample of 30 workers of this company.

    (a) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (I).

    (b) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (II).

    (c) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (III).

    (d) The normal distribution is a reasonable approximation of the sampling distribution of a proportion in at least two of the cases (I), (II), (III).

    (e) The normal distribution is a reasonable approximation of the sampling distribution of a proportion in none of the three cases (I), (II), (III).

    Solution:
    (e)
    Not a reasonable approx in (I) because $n(1-p) = 3 < 10$
    Not a reasonable approx in (II) because $np = 4 < 10$
    Not a reasonable approx in (III) because the sample size (30) is more than 10% of the population (150)

30. A survey of the male students at a junior college reveals that 26% play soccer regularly, 22% are Latino, and half of the Latino students play soccer regularly. If a male student is selected at random, what is the probability that he is neither Latino nor a soccer player?

    (a) 0.89

    (b) 0.52

    (c) 0.63

    (d) 0.26

    (e) 0.41

    Solution:
    c
    P(So) $= 0.26$, P(L) $= 0.22$, P(So$|$L) $= 0.5$ and so P(So and L) $= 0.22 \times 0.5 = 0.11$
    P(L or So) = P(L) + P(So) - P(L and So) = 0.26 +0.22-0.11 = 0.37
    P(neither Latino nor a soccer player) $= 1 - 0.37 = 0.63$

31. After menopause, some women take supplemental oestrogen. There is some concern that if these women also drink alcohol, their oestrogen levels will rise too high. Twelve volunteers who were receiving supplemental oestrogen were randomly divided into two groups, as were 12 other volunteers who were not receiving supplemental oestrogen. In each case, one group drank an alcoholic beverage, the other a non-alcoholic beverage. An hour later, everyone's oestrogen level was checked. Only those on supplemental oestrogen who drank alcohol showed a marked increase in oestrogen levels.
Use this information for this questions and the next one.
What is the response variable in this study?

(a) "Menopause" is the response variable

(b) "Amount of supplemental oestrogen received" is the response variable

(c) "Type of beverage" (alcoholic beverage, non-alcoholic beverage) is the response variable

(d) "Oestrogen level" is the response variable

(e) "Gender" is the response variable

Solution:
(d)

32. Same information as in the previous question.
After menopause, some women take supplemental oestrogen. There is some concern that if these women also drink alcohol, their oestrogen levels will rise too high. Twelve volunteers who were receiving supplemental oestrogen were randomly divided into two groups, as were 12 other volunteers who were not receiving supplemental oestrogen. In each case, one group drank an alcoholic beverage, the other a non-alcoholic beverage. An hour later, everyone's oestrogen level was checked. Only those on supplemental oestrogen who drank alcohol showed a marked increase in oestrogen levels.
Which of the following statements is/are correct?

(I) This is an example of an experiment where the units are assigned to treatments using a randomized block design

(II) "Amount of supplemental oestrogen received" is a blocking variable

(III) This is an example of an observational study since patients were not randomly assigned to treatments

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

(d) Only (I) and (II) are correct

(e) None of the three statements is correct

Solution:
(d)

33. Five hundred people used a home test for a HIV. We say that the test is positive if it indicates that the person has HIV, and we say it is negative if it indicates that the person does not have HIV. Out of these 500 people, 40 had HIV and 460 did not have HIV. The table below shows the contingency table of the result of the test (positive or negative) and whether a person has HIV (HIV, no HIV) for these 500 people.

	HIV	No HIV	Total
Positive Test	35	25	60
Negative Test	5	435	440
Total	40	460	500

A false positive is a test result that indicates a person has a disease when the person actually does not have it. What is the probability of a false positive with the home test for a HIV?

(a) 0.054

(b) 0.070

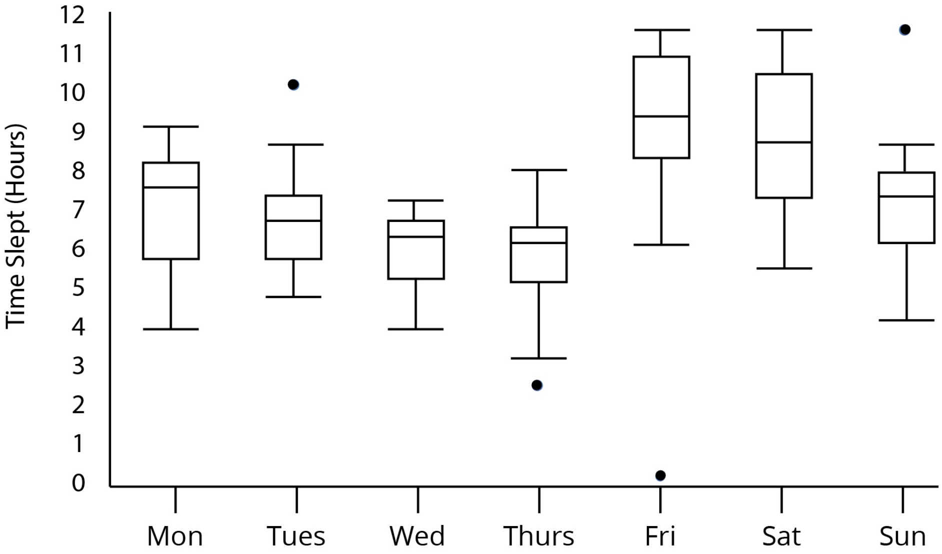
(c) 0.130

(d) 0.417

(e) 0.875

Solution:
(a) 25/460

34. The figure below shows time slept (in hours) for each day of the week for a group of 20 students.
Use this figure to answer this question and the following one.



Which of the following statements about time slept on **Monday** is **not correct**?

(a) The distribution of time slept is left skewed

(b) At least 10 students slept more than 7 hours

(c) At least 5 students slept less than 7 hours

(d) At least 15 students slept between 5 and 10 hours

(e) All four statements (a), (b), (c) and (d) are correct

Solution:
(e)

35. Use the figure of the previous question to answer this question.
Which of the following statements is known **for sure**?

(I) The interquartile range (IQR) of time slept on Thursday is smaller than the IQR of time slept on Saturday.

(II) The variance of time slept on Thursday is smaller than the variance of time slept on Saturday.

(III) The mean time slept on Tuesday is smaller than the mean time slept on Monday.

(a) Only (I) is known for sure
(b) Only (II) is known for sure
(c) Only (III) is known for sure
(d) Only (I) and (II) are known for sure
(e) Only (I) and (III) are known for sure

Solution:
(a)
The interquartile range (IQR) of time slept on Thursday is smaller than the IQR of time slept on Saturday because the box is shorter (I correct). We cannot say anything about the mean and variance based on boxplots (II and III not known for sure).

36. A city council of a small city wants to know the proportion of eligible voters that oppose having a garbage incinerator opened just outside the city limits. They randomly select 100 residential numbers from the city's telephone book that contains 3,000 such numbers. Each selected residence is then called and asked to provide the total number of eligible voters in the residence and the number of voters opposed to the incinerator in the residence.
What sampling method is used?

(a) Simple random sampling
(b) Stratified sampling
(c) Cluster sampling
(d) Multistage sampling
(e) Systematic sampling

Solution:
(c) Clusters = Residences

37. Suppose the correlation coefficient between height (measured in feet) and weight (measured in pounds) is 0.40. What is the correlation coefficient between height measured in inches and weight measured in ounces? The conversion is 12 inches = 1 foot and 16 ounces = 1 pound.

    (a) 0.002
    (b) 0.30
    (c) 0.40
    (d) 0.533
    (e) Impossible to say given the information provided

    Solution:
    (b)
    The correlation coefficient has not units

38. The probability that a randomly chosen American is a Republican is 0.35. We select a simple random sample of 6 Americans. What is the probability that at least 1 of the selected Americans is a Republican?

    (a) 0.0754
    (b) 0.2437
    (c) 0.3191
    (d) 0.3500
    (e) 0.9246

    Solution:
    (e) 1 - 0.0754 = 0.9246

39. The table below gives the probability distribution of the number of times (x) a photocopying machine needs repair in a given month.

x	0	1	2	3	4	5
Probability	0.12	0.18	0.23	0.20	0.14	0.13

What is the probability that the machine will have to be repaired at least twice but not more than four times in the given month?

(a) 0.23

(b) 0.24

(c) 0.43

(d) 0.57

(e) 0.87

Solution:
d
$P(2 \leq X \leq 4) = 0.23 + 0.20 + 0.14 = 0.57$

40. A box contains 3 yellow, 2 red, 4 green and 3 black marbles. Two marbles are taken one after the other at random from the box without replacement. What is the probability that both marbles are red?

(a) $\frac{1}{50}$

(b) $\frac{1}{60}$

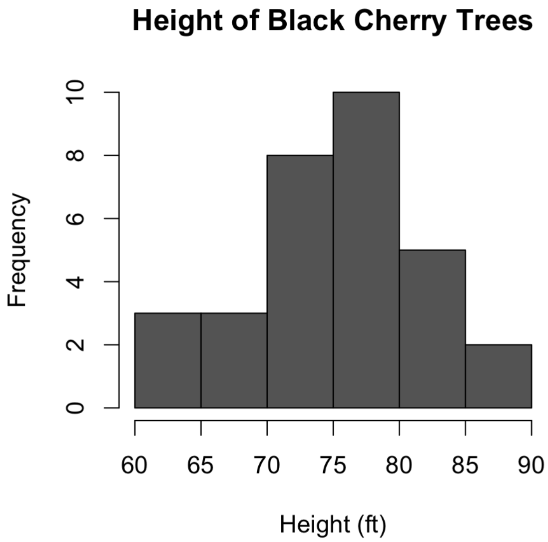
(c) $\frac{1}{66}$

(d) $\frac{1}{24}$

(e) $\frac{1}{18}$

Solution:
c
Let $R1$ be the event that the first marble selected is red and $R2$ be the event that the second is red.
We want $P(R1 \text{ and } R2)$
Using general multiplication rule we have,
$P(R1 \text{ and } R2) = P(R1) \times P(R2|R1) = \frac{2}{3+2+4+3} \times \frac{2-1}{3+2+4+3-1} = \frac{2}{12} \times \frac{1}{11} = \frac{1}{66}$

41. The histogram below shows height (in feet) of black cherry trees.



What do we know **for sure** about the median height of these black cherry trees?

(a) The median height is smaller than 75 feet

(b) The median height is 75 feet

(c) The median height is larger than 75 feet but smaller than 80 feet

(d) The median height is 80 feet

(e) The median height is larger than 80 feet

Solution:
(c)
There are 31 observation. The median is the 16th observation, which is between 75 and 80 feet.

42. A randomly selected sample of 1000 college students was asked whether they had ever used the drug Ecstasy. Sixteen percent (16% or 0.16) of the 1000 students surveyed said they had.
Which of the following statements about the number 0.16 is/are correct?

(I) It is the sample proportion of college students who have ever used the drug Ecstasy.

(II) It is the population proportion of college students who have ever used the drug Ecstasy.

(III) It is a value that can be used to estimate the population proportion of college students who have ever used the drug Ecstasy.

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

(d) Only (I) and (III) are correct

(e) Only (II) and (III) are correct

Solution:
(d) it is a value computed from the sample.

43. Students in an Intro Stats course were asked to describe their political preference as Liberal, Moderate, or Conservative. Here are the results:

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total	85	80	27	192

Use this information for this question and the next one.
We select a student at random. Let L denote the event that the student selected considers himself/herself to be liberal and let F denote the event that the student selected is female. Calculate $P(L|F)$.

(a) 0.18

(b) 0.25

(c) 0.35

(d) 0.41

(e) 0.45

Solution:
e
$P(L|F) = \frac{35}{77} = 0.4545455$

44. Same information as in the previous question.
Students in an Intro Stats course were asked to describe their political preference as Liberal, Moderate, or Conservative. Here are the results:

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total	85	80	27	192

We select a student at random. Let L denote the event that the student selected considers himself/herself to be liberal and let M denote the event that the student selected is male. Calculate $P(L \text{ or } M)$.

(a) 0.18

(b) 0.27

(c) 0.45

(d) 0.78

(e) 1.04

Solution:
c
$P(L \text{ or } M) = P(L) + P(M) - P(L \text{ and } M) = \frac{85}{192} + \frac{115}{192} - \frac{50}{192} = \frac{150}{192} = 0.78125$

45. A box contains 18 balls, some of them are black and the others are white. We do not know how many are black and how many are white. Let n be the number of black balls and so $18 - n$ is the number of white balls. We are going to pick four balls at random from this box, one by one, without replacement (i.e when a ball is picked, we do not put it back in the box). Let A denote the event that the first three balls selected are black and B denote the event that the fourth ball selected is black. If $P(B|A) = 1/5$, what is the value of n?

   (a) 6
   (b) 7
   (c) 8
   (d) 9
   (e) 12

   Solution:
   a
   After taking out the first 3 balls there are 15 left. $P(B|A) = 1/5$ means the probability of selecting a black ball from these remaining 15 balls is 1/5. This means there must be 3 $(15 \times \frac{1}{5})$ black balls in these 15 remaining balls. This means at the beginning there were $n = 3 + 3 = 6$ black balls.

   Or second method:
   $P(B|A) = \frac{n-3}{18-3} = 1/5 \implies n = 6$

46. The probability that a patient with a certain disease will be successfully treated with a new medical treatment is 0.80. The success of the treatment is independent from one patient to another. Suppose that the treatment is used on 40 patients with the disease. What is the mean of the number of patients who will **not** be successfully treated?

   (a) 8
   (b) 20
   (c) 32
   (d) 40
   (e) Impossible to say given the information provided

   Solution:
   (a)
   The mean of the number of patients who will **not** be successfully treated is $n(1 - p) = 40 \cdot 0.2 = 8$

47. The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.
Use this information for this question and the next one.
If I select at random two Canada Post letters, what is the probability that they will both weigh less than 1 oz.?

   (a) 0.0005
   (b) 0.0228
   (c) 0.0456
   (d) 0.4772
   (e) 0.8413

   Solution:
   (a) The probability that one letter weigh less than 1 oz. is $P(X < 1) = P(Z < -2) = 0.0228$. The probability that they will both weigh less than 1 oz is 0.0228^2 (multiplication rule for independent events).

48. Same information as in the previous question.
The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.
If I select at random 1000 Canada Post letters, what is the probability that at least 200 will weigh less than 1 oz.?

   (a) 0.0000
   (b) 0.0046
   (c) 0.1040
   (d) 0.8960
   (e) 1.0000

   Solution:
   (a) The number of letters that weigh less than 1 oz follows a $B(n = 1000, p = 0.0228)$. We can use the normal approximation since $np = 22.8$ and $n(1 - p) = 977.2$. The probability that at least 200 letters will weigh less than 1 oz. is $P(X >= 200) = P(Z >= (200 - 22.8)/\sqrt{22.280}) = 1 - P(Z < 37.54) = 0$

49. If the correlation coefficient r between two variables is equal to 1, what do we know **for sure**?

   (a) There is a perfect positive linear relationship between the two variables.

   (b) There is a very strong, but not perfect, positive linear relationship between the two variables.

   (c) There is no relationship between the two variables.

   (d) There is a perfect negative linear relationship between the two variables.

   (e) It is impossible to know one of the above for sure given the information provided.

   Solution:
   (a)

50. In a statistics class, there are 60 men and 40 women. The mean weight of all these 100 students is 160 pounds. If the mean weight of the 60 men is 180 pounds, what is the mean weight of the 40 women?

   (a) 120 pounds

   (b) 125 pounds

   (c) 130 pounds

   (d) 132 pounds

   (e) 135 pounds

   Solution:
   c
   $\frac{100\times160-60\times180}{40} = 130\text{lb}$

END OF EXAM

Tables of normal, binomial and t-distributions

prepared by Ken Butler

2015-06-03

Table Z (normal distribution)

Values of z greater than 0 are on the next page. Second decimal place is at the top of the column.

	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
-3.3	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005
-3.2	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007
-3.1	0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010
-3.0	0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013
-2.9	0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019
-2.8	0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026
-2.7	0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035
-2.6	0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047
-2.5	0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062
-2.4	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
-2.3	0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
-2.2	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139
-2.1	0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179
-2.0	0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228
-1.9	0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287
-1.8	0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359
-1.7	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
-1.6	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
-1.5	0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668
-1.4	0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808
-1.3	0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968
-1.2	0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151
-1.1	0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357
-1.0	0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587
-0.9	0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841
-0.8	0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119
-0.7	0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420
-0.6	0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743
-0.5	0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085
-0.4	0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446
-0.3	0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821
-0.2	0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207
-0.1	0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602
0.0	0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000

Table Z (continued)

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table of the Binomial Distribution

(Table shows probability of exactly k successes in n trials with success probability p. Values of p are shown along the top of the table.)

		p									
n	k	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
2	1	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
2	2	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
3	1	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
3	2	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
3	3	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
4	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
4	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
4	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
4	4	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
5	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1563
5	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
5	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1812	0.2304	0.2757	0.3125
5	4	0.0000	0.0004	0.0022	0.0064	0.0147	0.0284	0.0488	0.0768	0.1128	0.1563
5	5	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0313
6	0	0.7351	0.5315	0.3772	0.2622	0.1780	0.1177	0.0754	0.0467	0.0277	0.0156
6	1	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0938
6	2	0.0306	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344
6	3	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125
6	4	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344
6	5	0.0000	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938
6	6	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156

Computer & Mathematical Sciences
UNIVERSITY OF TORONTO
SCARBOROUGH

STAB22: Final Exam April 19, 2018

Complete this form, then *carefully* tear off this page and place it along with your student card on the surface beside where you are writing.

The form will be verified and collected during the test.

Information collected will be used solely for the purpose of administering this test.

Name (please print): ____

Signature: ____________________________ Seat #: ________________
(if available)

Student Number: ☐☐☐☐☐☐☐☐☐☐

UNIVERSITY OF TORONTO SCARBOROUGH
Department of Computer and Mathematical Sciences
April 19, 2018
APRIL 2018 FINAL EXAMINATIONS

STAB22H3 Statistics I
Duration: 3 hours

Last Name:________________________ First Name: ________________________

Student number: ________________________

Aids allowed:

- Two handwritten letter-sized sheets (both sides) of notes prepared by you
- Non-programmable, non-communicating calculator

The table of normal and binomial distributions are attached at the end.

This test is based on multiple-choice questions. There are 50 questions. All questions carry equal weight. Report all your answers on the Scantron answer sheet before the end of the exam. On the Scantron answer sheet, ensure that you write down and **BUBBLE** your last name, first name (as much of it as fits), and student number (in "Identification").

Mark in each case the best answer out of the alternatives given (**which means the numerically closest answer if the answer is a number and the answer you obtained is not given.**)

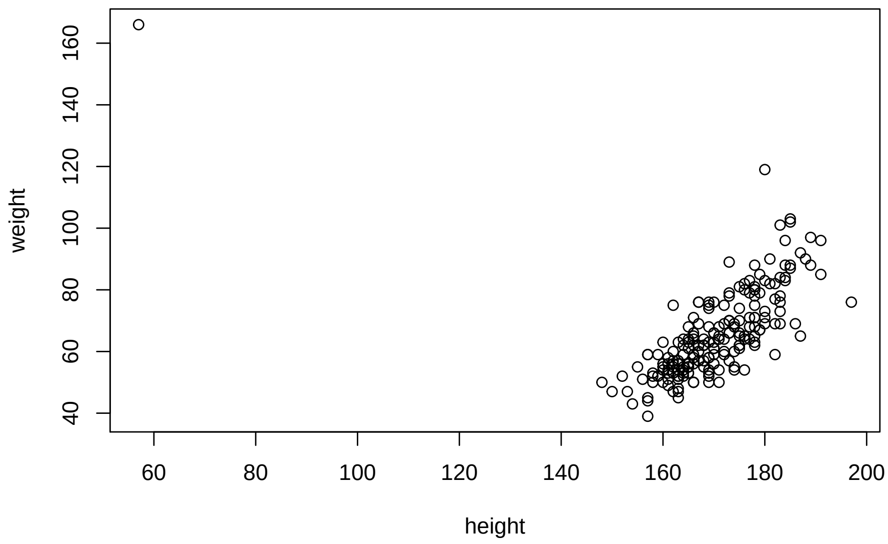
Also before you begin, complete the signature sheet, but sign it only when the invigilator collects it. The signature sheet shows that you were present at the exam.

There are 30 pages including this page and statistical tables. Please check to see that you have all the pages.

Good luck!!

Answer Key for Exam $\boxed{\text{B}}$

1. The scatterplot below shows the self-reported height and weight of 200 subjects. The subjects were men and women engaged in regular exercise. Suppose we fit a linear regression of weight on height.
   Use this scatterplot for this question and the next one.



Which of the following statements about the data point with height = 57 is/are true?

(I) The data point with height = 57 is an outlier.
(II) The data point with height = 57 is a high leverage point.
(III) The data point with height = 57 is an influential observation.

(a) Only (I) is true
(b) Only (II) is true
(c) Only (III) is true
[✓ (d)] All three statements (I), (II) and (III) are true
(e) None of the three statements is true

Solution:
(d)

2. Use the scatterplot of the previous question to answer this question.
Suppose we fit a linear regression of weight on height.
Which of the following statements is known **for sure**?

   (I) If we remove the data point with height = 57, the slope of the regression line increases.

   (II) If we remove the data point with height = 57, the mean height increases.

   (III) If we remove the data point with height = 57, the variance of height increases.

   (a) Only (II) is known for sure
   [✓ (b)] Only (I) and (II) are known for sure
   (c) Only (I) and (III) are known for sure
   (d) Only (II) and (III) are known for sure
   (e) None of the three statements is known for sure

   Solution:
   (b)
   If we remove the data point with height = 57, the slope of the regression line increases (I true), the mean height increases (II true), the variance of height decreases (III not true).

3. The tail length for a population of female hook-billed kites is Normally distributed with mean 194 mm and standard deviation 11 mm. We select a random sample of 200 kites from this population. Let X be the number of kites in this sample having a tail length between 179 and 209 mm. What is the mean of X?

   (a) 145
   (b) 155
   [✓ (c)] 165
   (d) 175
   (e) 194

   Solution:
   c
   Letting L denote the tail length of a randomly selected female hook-billed kite, $X \sim Bin(n = 300, p = P(179 < L < 209))$.
   $p = P(179 < L < 209) = P(\frac{179-194}{11} < Z < \frac{209-194}{11}) = P(-1.36Z < 1.36) = 1 - 2 \times 0.0869 = 1 - 0.1738 = 0.8262$
   $\mu_X = np = 200 \times 0.8262 = 165.24$

4. The heights of students in some population have a Normal distribution with mean 172cm. Twenty five percent of the students in this population have heights greater than or equal to 180cm. What is the standard deviation of the heights of the students in this population?

   (a) 7.9cm
   (b) 8.7cm
   (c) 9.9cm
   (d) 10.5cm
   [✓ (e)] 11.9cm

   Solution:
   e
   $P(X \geq 180) = 0.25 \implies z = 0.67$ and $\frac{180-172}{\sigma} = 0.67 \implies \sigma = \frac{180-172}{0.67} = 11.94$

5. Childhood lead poisoning is a public health concern in a country. In this country, 1 child in 10 has a high blood lead level. In a randomly chosen group of 6 children from the population of this country, what is the probability that 3 or more have high blood lead level?

   (a) 0.0146
   (b) 0.5000
   (c) 0.6000
   [✓ (d)] 0.0159
   (e) 0.9987

   Solution:
   d
   Let X = number of children with high blood level in this sample. Then $X \sim Bin(n = 6, p = 1/10 = 0.1)$. $P(X \geq 3) = 0.0146 + 0.0012 + 0.0001 + 0.0000 = 0.0159$

6. Using a random sample of 4,000 students, you compute a 95% confidence interval for the proportion of overweight students. You decide to compute another 95% confidence interval using a different sample, this time with only 1,000 students. Suppose that the sample proportion $\widehat{p}$ is the same for both these confidence intervals. What change would you expect from the first confidence interval to the second?

   (a) The margin of error of the second interval will be 4 times as big as the margin of error of the first interval.

   [✓ (b)] The margin of error of the second interval will be 2 times as big as the margin of error of the first interval.

   (c) The margin of error of the first interval will be 4 times as big as the margin of error of the second interval.

   (d) The margin of error of the first interval will be 2 times as big as the margin of error of the second interval.

   (e) There will be no change in the margin of error since the sample proportion $\widehat{p}$ is the same for both confidence intervals.

   Solution:
   (b)
   The margin of error of the second interval is:

$$1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{1000}} = 2 \cdot 1.96\sqrt{\frac{\widehat{p}(1-\widehat{p})}{4000}}$$

   which is 2 times the margin of error of the first interval.

7. A random experiment consists of five possible outcomes: E_1, E_2, E_3, E_4 and E_5. This means the sample space is $S = \{E_1, E_2, E_3, E_4, E_5\}$. If $P(E_1) = 3 \times P(E_2) = 0.3$ and $P(E_3) = P(E_4) = P(E_5)$, what is $P(E_5)$?

   (a) 0.1

   [✓ (b)] 0.2

   (c) 0.3

   (d) 0.4

   (e) 0.6

   Solution:
   c
   $P(E_1) = 3P(E_2) = 0.3 \implies P(E_2) = 0.1$ and so $P(E_3) + P(E_4) + P(E_5) = 1 - (0.3 + 0.1) = 0.6$ and since $P(E_3) = P(E_4) = P(E_5)$, $P(E_5) = 0.6/3 = 0.2$

8. A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

Use this information for this question and the next two questions.
One student in the sample was 73 inches tall with a foot length of 29 cm. What is the residual for this student?

(a) 29 cm

[✓ (b)] 1.31 cm

(c) 0 cm

(d) 27.69 cm

(e) -1.31 cm

Solution:
b
$29 - (10.9 + 0.23 \times 73) = 1.31$

9. Same information as in the previous question.
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

If we convert the heights of all these 33 students into cm using the conversion factor 1 inch = 2.54 cm and recalculate the regression equation with both the variables measured in cm, what will be the slope of the new regression equation?

[✓ (a)] 0.09

(b) 0.58

(c) 0.23

(d) 11.48

(e) 10.99

Solution:
a
$b = r\frac{s_y}{s_x}$. r and s_y don't change. s_x gets multiplied by 2.54 and so the new slope is $0.23/2.54 = 0.09055118$

10. Same information as in the previous two questions.
A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

If we convert the foot lengths of all these 33 students into inches using the conversion factor 1 cm = 0.3937 inches and recalculate the regression equation with both the variables measured in inches, what will be the y-intercept of the new regression equation?

(a) 11.29

(b) 28.27

(c) 10.9

(d) 27.69

$\boxed{\text{(e)}}$ 4.29

Solution:
e
$a = \bar{y} - b\bar{x}, b = r\frac{s_y}{s_x}$. r and $\bar{x}$ don't change. $\bar{y}$,s_y and b get multiplied by 0.3937 and so the new y-intercapt is is $10.9 \times 0.3937 = 4.29133$

11. A traditional die is a cube, each face showing an integer in the range from 1 to 6. A **fair** traditional die is rolled 3 times. The first 2 rolls result in the face showing "5" being uppermost. What is the probability the third roll will **not** result in the face showing "5" being uppermost?

(a) 0

(b) 0.023

(c) 0.069

$\boxed{\text{(d)}}$ 0.833

(e) 1

Solution:
(d) The rolls are independent. The probability the third roll will not result in the face with "5" being uppermost is $1 - 1/6 = 5/6 = 0.833$.

12. A business office orders paper supplies from one of three vendors, V1, V2, or V3. Orders are to be placed on three successive days (e.g. next Mon, Tue and Wed), one order per day. On each day, one of the three vendors is selected at random, giving each vendor the same probability (which is 1/3) to be selected independently of the other vendors.

    Use this information for this question and the next one.

    Find the probability that vendor V1 gets at least one order.

    (a) $\frac{1}{3}$

    (b) $\frac{1}{9}$

    (c) $\frac{8}{27}$

    $\boxed{\text{(d)}}$ $\frac{19}{27}$

    (e) $\frac{26}{27}$

    Solution:

    d

    $P(\text{ at least one V1}) = 1 - P(\text{ None is V1}) = 1 - P(\text{V1}^c)^3 = 1 - \left(\frac{2}{3}\right)^3 = 1 - \frac{8}{27} = \frac{19}{27}$

13. Same information as in the previous question.

    A business office orders paper supplies from one of three vendors, V1, V2, or V3. Orders are to be placed on three successive days (e.g. next Mon, Tue and Wed), one order per day. On each day, one of the three vendors is selected at random, giving each vendor the same probability (which is 1/3) to be selected independently of the other vendors.

    Find the probability that vendor V1 gets exactly two of these three orders.

    (a) $\frac{1}{3}$

    (b) $\frac{1}{9}$

    $\boxed{\text{(c)}}$ $\frac{2}{9}$

    (d) $\frac{2}{27}$

    (e) $\frac{3}{27}$

    Solution:

    c

    $P(\text{ exactly two V1}) = P(V1, V1, V1^c) + P(V1, V1^c, V1) + P(V1^c, V1, V1) = 3 \times \left(\frac{1}{3}\right)^2 \times \left(\frac{2}{3}\right) = \frac{2}{9}$

    Note: This question does not require the binomial tables or the binomial formula even though this can also be done using the binomial formula.

14. A randomized experiment was done by randomly assigning each participant either to walk for half an hour three times a week or to sit quietly reading a book for half an hour three times a week. The participants' blood pressure was measured at the beginning and at the end of the study and the change in participants' blood pressure was computed. The change in participants' blood pressure was compared for the two groups.
    This is a randomized experiment rather than an observational study because:

    (a) blood pressure was measured at the beginning and end of the study.

    (b) the two groups were compared at the end of the study.

    $\boxed{\text{(c)}}$ the participants were randomly assigned to either walk or read, rather than choosing their own activity.

    (d) a random sample of participants was used.

    (e) the study was done by scientists.

    Solution:
    c

15. If there is a very strong linear relationship between two variables then what do you know **for sure** about the correlation coefficient r?

    (a) The correlation coefficient is larger than 1

    (b) The correlation coefficient is close to 0

    (c) The correlation coefficient is close to -1

    (d) The correlation coefficient is close to 1

    $\boxed{\text{(e)}}$ None of the above is known for sure.

    Solution:
    (e) We know for sure that the correlation coefficient is close to -1 or 1.

16. In a game, a wheel with five equal-sized spaces labeled from A to E is spun and a ball randomly lands in one of these spaces. If the ball lands in the space labeled A, the player wins 15 dollars. If the ball lands in any other space, the winner loses 4 dollars. Use this information for this questions and the next one.
A player plays this game 6 times. What is the probability that the ball lands at least one time in the space labeled A? You can assume that games are independent from one another.

(a) 0.2622
(b) 0.3446
(c) 0.3932
(d) 0.6068
[✓ (e)] 0.7378

Solution:
X = number of times the ball lands in a space labeled A
$X \sim Binom(n = 6, p = 0.2)$
$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.2622 = 0.7378$

17. Same information as in the previous question.
In a game, a wheel with five equal-sized spaces labeled from A to E is spun and a ball randomly lands in one of these spaces. If the ball lands in the space labeled A, the player wins 15 dollars. If the ball lands in any other space, the winner loses 4 dollars. What is the expected gain or loss of a player who plays 40 games? You can assume that games are independent from one another.

(a) Gain of 360 dollars
(b) Gain of 55 dollars
(c) Gain of 8 dollars
(d) Loss of 1 dollar
[✓ (e)] Loss of 8 dollars

Solution:
Expected loss or gain for one game: $0.2 \cdot 15 + 0.8 \cdot (-4) = -0.2$
Expected loss or gain for 40 games (games are independent): $40 \cdot (-0.2) = -8$

18. The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\widehat{y} = -0.0127 + 0.0180x.$$

Use this information for this question and the next one.
The above equation implies that:

(a) Each beer consumed increases blood alcohol content by 0.0127

(b) Each beer consumed decreases blood alcohol content by 0.0127

(c) On average it takes 1.8 beers to increase blood alcohol content by 0.01

[✓ (d)] Each beer consumed increases blood alcohol content by an average of amount of 0.018

(e) Each beer consumed increases blood alcohol content by exactly 0.018

Solution:
(d)

19. Same information as in the previous question.
The relationship between number of beers consumed (x) and blood alcohol content (y) was studied in 16 male college students by using least squares regression. The following regression equation was obtained from this study:

$$\widehat{y} = -0.0127 + 0.0180x.$$

Suppose that the legal upper limit to drive is a blood alcohol content of 0.08.
Which of the following statements is/are correct?

(I) The model predicts that a driver who drinks 5 beers has a blood content below the legal upper limit to drive.

(II) The model predicts that a driver who drinks 6 beers has a blood content below the legal upper limit to drive.

(III) A driver who drinks 5 beers has **for sure** a blood content that is below the legal upper limit to drive.

$\boxed{\text{(a)}}$ Only statement (I) is correct
(b) Only statement (II) is correct
(c) Only statements (I) and (II) are correct
(d) Only statements (I) and (III) are correct
(e) All three statements (I), (II), and (III) are correct

Solution:
(a)
Solving $0.08 = -0.0127 + 0.0180x$ yields $x = 5.15$.

20. A consumer organization estimates that 34% of the households in a particular community have one television set, 39% have two sets, and 20% have three or more sets. What is the probability that a household chosen at random in this community has no more than one television set?

$\boxed{\text{(a)}}$ 0.41
(b) 0.59
(c) 0.07
(d) 0.34
(e) 0.46

Solution:
a
P(no more than 1) = 1- P(more than 1) = 1- P(2)- P(3) = 1-0.39-0.20 = 0.41

21. The following table represents the probability distribution of the number of accidents per day in a city.

Number of accidents	0	1	2	3	4 or more
Probability	0.55	0.20	0.10	0.15	0

Use this table for this question and the next one.
Which of the following statements about the number of accidents per day in the city is/are true?

(I) The mean number of accidents per day is larger than the median number of accidents per day.

(II) The mean number of accidents per day is smaller than the median number of accidents per day.

(III) The median number of accidents per day is equal to 1.

$\boxed{\text{(a)}}$ Only (I) is true
(b) Only (II) is true
(c) Only (III) is true
(d) Only (I) and (II) are true
(e) Only (I) and (III) are true

Solution:
(a)
Right skewed distribution / median is 0 / mean is 0.85

22. Use the probability distribution of the previous question to answer this question.
What is the variance of the number of accidents per day in the city?

$\boxed{\text{(a)}}$ 1.23
(b) 1.67
(c) 2.07
(d) 2.50
(e) 3.10

Solution:
(a)
$\mu = 0.85$
$\sigma^2 = \sum(X - 0.85)^2 P(X) = 1.2275$

23. Acme Medicine is conducting an experiment to test a new vaccine, developed to immunize people against the common cold. To test the vaccine, Acme has 1000 participants. Participants are divided into two groups, based on gender. Then, within each group, half of the participants are randomly assigned to treatment groups so that the first half gets the placebo and the second half gets the vaccine.
Which of the following statements is correct?

(a) This is an example of a prospective observational study.

$\boxed{\text{(b)}}$ This is an example of an experiment where the units are assigned to treatments using a randomized block design.

(c) This is an example of an experiment where the units are assigned to treatments using a completely randomized design.

(d) This is an example of a census.

(e) This is an example of a survey where the participants are selected using stratified sampling.

Solution:
(b)

24. A randomly selected sample of 1000 college students was asked whether they were in favor of longer hours at the school library. Eighty percent (80%) of the 1000 students surveyed said they were in favor.
Which of the following statements is/are correct?

(I) A 95% confidence interval for the population proportion of college students who were in favor of longer hours at the school library is $[0.78; 0.82]$

(II) The population proportion of college students who were in favor of longer hours at the school library is significantly different from 76% at the level of significance $\alpha = 5\%$.

(III) The population proportion of college students who were in favor of longer hours at the school library is significantly different from 76% at the level of significance $\alpha = 1\%$.

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

(d) Only (I) and (II) are correct

$\boxed{\text{(e)}}$ The three statements (I), (II), and (III) are correct

Solution:
(e)
A 95% confidence interval is $0.8 \pm 1.96\sqrt{0.2 \cdot 0.8/1000} = [0.78; 0.82]$. The value of the test statistic is z-obs $= 2.96$ and the p-value is $2 \cdot 0.0015 = 0.003$.

25. STA101 and STA102 are two statistics courses both with large classes. We have the following information about the distributions of grades for these two courses:

    - The distribution of grades of STA101 is right skewed.
    - The distribution of grades of STA102 is left skewed.
    - STA101 has lower class average than STA102 (i.e. $\mu_1 < \mu_2$, where μ_1 and μ_2 are the means of the distributions of grades in STA101 and STA102 respectively.)

    Which of the following statements regarding the distributions of grades of these two courses is **NOT** true?

    (a) The median grade of STA101 is less than the median grade of STA102.
    $\boxed{\text{(b)}}$ The mean grade of STA102 is less than the median grade of STA101.
    (c) The mean grade of STA101 is less than the median grade of STA102.
    (d) The median grade of STA101 is less than the mean grade of STA102.
    (e) The mean grade of STA102 is less than the median grade of STA102.

Solution:
b
Denoting the medians of STA101 and STA102 by m_1 and m_2 respectively, we have,
$m_1 < \mu_1$, since the distribution of grades of STA101 is right skewed,
$\mu_2 < m_2$, since the distribution of grades of STA102 is left skewed,
and so since we are also given that $\mu_1 < \mu_2$, we have
$m_1 < \mu_1 < \mu_2 < m_2$ and this implies:

$m_1 < m_2$, i.e. (a) is true
$m_1 < \mu_2$, i.e. (b) is false
$\mu_1 < m_2$, i.e. (c) is true
$m_1 < \mu_2$, i.e. (d) is true
$\mu_2 < m_2$, i.e. (e) is true

26. In a clinical study of a new drug, patients are randomly allocated to one of two groups. The patients in the first group are given an injection with the drug in it. The patients in the second group are given an injection, identical to the injection containing the drug, but with no drug in it. Doctors record the symptoms reported by the patients and take measurements, without knowing which patients are in which group. Data are then analysed to determine whether the drug is effective.
Which of the following statements is/are correct?

(I) If the groups end up different in terms of symptoms and measurements, then we know **for sure** that this difference is due to the administered drug, no matter how big or small the difference is.

(II) This is an example of a double-blind experiment.

(III) This is an example of an experiment where the units are assigned to treatments using a randomized block design.

(a) Only (I) is correct
[✓ (b)] Only (II) is correct
(c) Only (III) is correct
(d) Only (I) and (II) are correct
(e) Only (II) and (III) are correct

Solution:
(b) (I) not correct because the difference could be due to chance, specially if it is small. We have to check whether the difference is statistically significant; This is an example of a double-blind experiment where the units are assigned to groups using a completely randomized design, i.e. (II) is correct and (III) is not correct.

27. A local bakery has determined a probability distribution for the number of cheesecakes it sells in a given day. The distribution is as follows:

Number sold in a day	10	20	30	35	x
Probability	0.1	0.5	0.1	0.2	0.1

If the mean number of cheesecakes that this local bakery sells in a day is 25, what is the value of x?

(a) 30

(b) 35

(c) 38

[✓ (d)] 40

(e) 45

Solution:

d

$(10 \times 0.1) + (20 \times 0.5) + (30 \times 0.1) + (35 \times 0.2) + (x \times 0.1) = 25 \implies x \times 0.1 = 25 - 21 \implies x = \frac{4}{0.1} = 40$

28. In a large population, 60% of the voters support a particular party. If a random sample of 300 voters were selected from this population, what would be the approximate probability that more than 175 of them would support this party?

(a) 0.35

[✓ (b)] 0.72

(c) 0.96

(d) 0.51

(e) 0.60

Solution:

b

$X \sim Bin(n = 300, p = 0.6) \overset{approx}{\sim} N(\mu = np = 300 \times 0.6 = 180, \sigma = \sqrt{np(1-p)}) = \sqrt{300 \times 0.6 \times 0.4}) = 8.49.$

$P(X \leq 175) = P(Z < \frac{175-180}{8.49} = -0.59) = 0.2776.$

$P(X > 175) = 1 - P(X \leq 175) = 1 - 0.2776 = 0.7224$

29. In which of the following cases is the normal distribution a reasonable approximation of the sampling distribution of a proportion?

(I) The probability of an airline flight arriving on time is 90%. The airline operates more than 500 flights per day. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of flights arriving on time in a simple random sample of 30 flights of this airline.

(II) In a large country 0.2% of the population is affected by a disease. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of individuals affected by the disease in a simple random sample of 2000 individuals in this country.

(III) A company has 150 workers. The proportion of workers who smoke in this company is 40%. The normal distribution is a reasonable approximation of the sampling distribution of the proportion of workers who smoke in a simple random sample of 30 workers of this company.

(a) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (I).

(b) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (II).

(c) The normal distribution is a reasonable approximation of the sampling distribution of a proportion only in case (III).

(d) The normal distribution is a reasonable approximation of the sampling distribution of a proportion in at least two of the cases (I), (II), (III).

[✓ (e)] The normal distribution is a reasonable approximation of the sampling distribution of a proportion in none of the three cases (I), (II), (III).

Solution:
(e)
Not a reasonable approx in (I) because $n(1-p) = 3 < 10$
Not a reasonable approx in (II) because $np = 4 < 10$
Not a reasonable approx in (III) because the sample size (30) is more than 10% of the population (150)

30. A survey of the male students at a junior college reveals that 26% play soccer regularly, 22% are Latino, and half of the Latino students play soccer regularly. If a male student is selected at random, what is the probability that he is neither Latino nor a soccer player?

    (a) 0.89

    (b) 0.52

    [✓ (c)] 0.63

    (d) 0.26

    (e) 0.41

    Solution:
    c
    P(So) $= 0.26$, P(L) $= 0.22$, P(So$|$L) $= 0.5$ and so P(So and L) $= 0.22 \times 0.5 = 0.11$
    P(L or So) = P(L) + P(So) - P(L and So) = 0.26 +0.22-0.11 = 0.37
    P(neither Latino nor a soccer player) $= 1 - 0.37 = 0.63$

31. After menopause, some women take supplemental oestrogen. There is some concern that if these women also drink alcohol, their oestrogen levels will rise too high. Twelve volunteers who were receiving supplemental oestrogen were randomly divided into two groups, as were 12 other volunteers who were not receiving supplemental oestrogen. In each case, one group drank an alcoholic beverage, the other a non-alcoholic beverage. An hour later, everyone's oestrogen level was checked. Only those on supplemental oestrogen who drank alcohol showed a marked increase in oestrogen levels.
Use this information for this questions and the next one.
What is the response variable in this study?

(a) "Menopause" is the response variable

(b) "Amount of supplemental oestrogen received" is the response variable

(c) "Type of beverage" (alcoholic beverage, non-alcoholic beverage) is the response variable

$\boxed{\text{(d)}}$ "Oestrogen level" is the response variable

(e) "Gender" is the response variable

Solution:
(d)

32. Same information as in the previous question.
After menopause, some women take supplemental oestrogen. There is some concern that if these women also drink alcohol, their oestrogen levels will rise too high. Twelve volunteers who were receiving supplemental oestrogen were randomly divided into two groups, as were 12 other volunteers who were not receiving supplemental oestrogen. In each case, one group drank an alcoholic beverage, the other a non-alcoholic beverage. An hour later, everyone's oestrogen level was checked. Only those on supplemental oestrogen who drank alcohol showed a marked increase in oestrogen levels.
Which of the following statements is/are correct?

(I) This is an example of an experiment where the units are assigned to treatments using a randomized block design

(II) "Amount of supplemental oestrogen received" is a blocking variable

(III) This is an example of an observational study since patients were not randomly assigned to treatments

(a) Only (I) is correct

(b) Only (II) is correct

(c) Only (III) is correct

$\boxed{\text{(d)}}$ Only (I) and (II) are correct

(e) None of the three statements is correct

Solution:
(d)

33. Five hundred people used a home test for a HIV. We say that the test is positive if it indicates that the person has HIV, and we say it is negative if it indicates that the person does not have HIV. Out of these 500 people, 40 had HIV and 460 did not have HIV. The table below shows the contingency table of the result of the test (positive or negative) and whether a person has HIV (HIV, no HIV) for these 500 people.

	HIV	No HIV	Total
Positive Test	35	25	60
Negative Test	5	435	440
Total	40	460	500

A false positive is a test result that indicates a person has a disease when the person actually does not have it. What is the probability of a false positive with the home test for a HIV?

[✓ (a)] 0.054

(b) 0.070

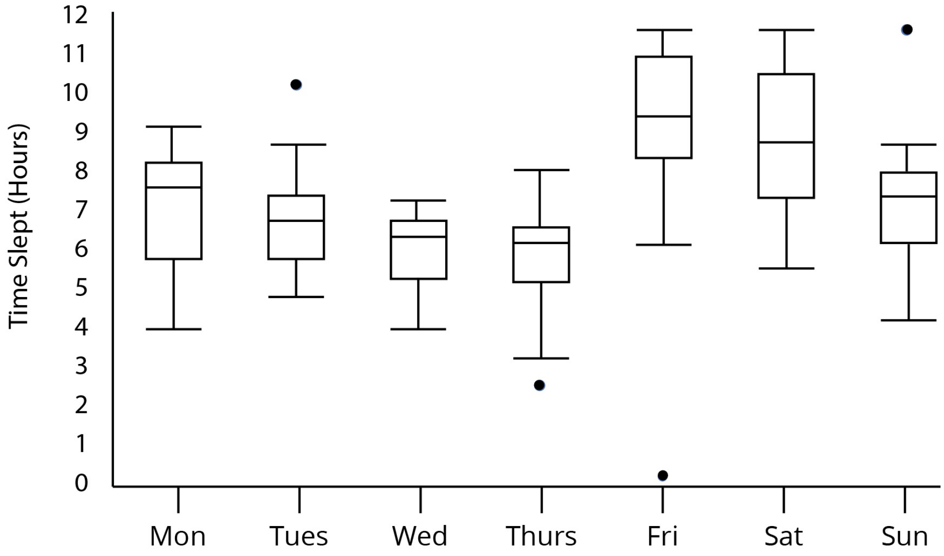
(c) 0.130

(d) 0.417

(e) 0.875

Solution:
(a) 25/460

34. The figure below shows time slept (in hours) for each day of the week for a group of 20 students.
Use this figure to answer this question and the following one.



Which of the following statements about time slept on **Monday** is **not correct**?

- (a) The distribution of time slept is left skewed
- (b) At least 10 students slept more than 7 hours
- (c) At least 5 students slept less than 7 hours
- (d) At least 15 students slept between 5 and 10 hours
- [✓ (e)] All four statements (a), (b), (c) and (d) are correct

Solution:
(e)

35. Use the figure of the previous question to answer this question.
Which of the following statements is known **for sure**?

(I) The interquartile range (IQR) of time slept on Thursday is smaller than the IQR of time slept on Saturday.

(II) The variance of time slept on Thursday is smaller than the variance of time slept on Saturday.

(III) The mean time slept on Tuesday is smaller than the mean time slept on Monday.

[✓ (a)] Only (I) is known for sure
(b) Only (II) is known for sure
(c) Only (III) is known for sure
(d) Only (I) and (II) are known for sure
(e) Only (I) and (III) are known for sure

Solution:
(a)
The interquartile range (IQR) of time slept on Thursday is smaller than the IQR of time slept on Saturday because the box is shorter (I correct). We cannot say anything about the mean and variance based on boxplots (II and III not known for sure).

36. A city council of a small city wants to know the proportion of eligible voters that oppose having a garbage incinerator opened just outside the city limits. They randomly select 100 residential numbers from the city's telephone book that contains 3,000 such numbers. Each selected residence is then called and asked to provide the total number of eligible voters in the residence and the number of voters opposed to the incinerator in the residence.
What sampling method is used?

(a) Simple random sampling
(b) Stratified sampling
[✓ (c)] Cluster sampling
(d) Multistage sampling
(e) Systematic sampling

Solution:
(c) Clusters = Residences

37. Suppose the correlation coefficient between height (measured in feet) and weight (measured in pounds) is 0.40. What is the correlation coefficient between height measured in inches and weight measured in ounces? The conversion is 12 inches = 1 foot and 16 ounces = 1 pound.

    (a) 0.002

    (b) 0.30

    $\boxed{\text{(c)}}$ 0.40

    (d) 0.533

    (e) Impossible to say given the information provided

    Solution:
    (b)
    The correlation coefficient has not units

38. The probability that a randomly chosen American is a Republican is 0.35. We select a simple random sample of 6 Americans. What is the probability that at least 1 of the selected Americans is a Republican?

    (a) 0.0754

    (b) 0.2437

    (c) 0.3191

    (d) 0.3500

    $\boxed{\text{(e)}}$ 0.9246

    Solution:
    (e) 1 - 0.0754 = 0.9246

39. The table below gives the probability distribution of the number of times (x) a photocopying machine needs repair in a given month.

x	0	1	2	3	4	5
Probability	0.12	0.18	0.23	0.20	0.14	0.13

    What is the probability that the machine will have to be repaired at least twice but not more than four times in the given month?

    (a) 0.23

    (b) 0.24

    (c) 0.43

    $\boxed{\text{(d)}}$ 0.57

    (e) 0.87

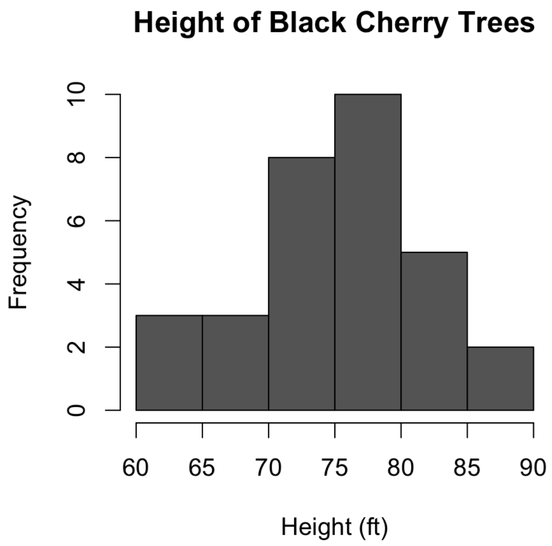
    Solution:
    d
    $P(2 \leq X \leq 4) = 0.23 + 0.20 + 0.14 = 0.57$

40. A box contains 3 yellow, 2 red, 4 green and 3 black marbles. Two marbles are taken one after the other at random from the box without replacement. What is the probability that both marbles are red?

   (a) $\frac{1}{50}$
   (b) $\frac{1}{60}$
   [✓ (c)] $\frac{1}{66}$
   (d) $\frac{1}{24}$
   (e) $\frac{1}{18}$

Solution:
c
Let $R1$ be the event that the first marble selected is red and $R2$ be the event that the second is red.
We want $P(R1 \text{ and } R2)$
Using general multiplication rule we have,
$P(R1 \text{ and } R2) = P(R1) \times P(R2|R1) = \frac{2}{3+2+4+3} \times \frac{2-1}{3+2+4+3-1} = \frac{2}{12} \times \frac{1}{11} = \frac{1}{66}$

41. The histogram below shows height (in feet) of black cherry trees.



What do we know **for sure** about the median height of these black cherry trees?

(a) The median height is smaller than 75 feet

(b) The median height is 75 feet

[✓ (c)] The median height is larger than 75 feet but smaller than 80 feet

(d) The median height is 80 feet

(e) The median height is larger than 80 feet

Solution:
(c)
There are 31 observation. The median is the 16th observation, which is between 75 and 80 feet.

42. A randomly selected sample of 1000 college students was asked whether they had ever used the drug Ecstasy. Sixteen percent (16% or 0.16) of the 1000 students surveyed said they had.
Which of the following statements about the number 0.16 is/are correct?

(I) It is the sample proportion of college students who have ever used the drug Ecstasy.

(II) It is the population proportion of college students who have ever used the drug Ecstasy.

(III) It is a value that can be used to estimate the population proportion of college students who have ever used the drug Ecstasy.

(a) Only (I) is correct
(b) Only (II) is correct
(c) Only (III) is correct
[✓ (d)] Only (I) and (III) are correct
(e) Only (II) and (III) are correct

Solution:
(d) it is a value computed from the sample.

43. Students in an Intro Stats course were asked to describe their political preference as Liberal, Moderate, or Conservative. Here are the results:

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total	85	80	27	192

Use this information for this question and the next one.
We select a student at random. Let L denote the event that the student selected considers himself/herself to be liberal and let F denote the event that the student selected is female. Calculate $P(L|F)$.

(a) 0.18
(b) 0.25
(c) 0.35
(d) 0.41
[✓ (e)] 0.45

Solution:
e
$P(L|F) = \frac{35}{77} = 0.4545455$

44. Same information as in the previous question.
Students in an Intro Stats course were asked to describe their political preference as Liberal, Moderate, or Conservative. Here are the results:

	Liberal	Moderate	Conservative	Total
Female	35	36	6	77
Male	50	44	21	115
Total	85	80	27	192

We select a student at random. Let L denote the event that the student selected considers himself/herself to be liberal and let M denote the event that the student selected is male. Calculate $P(L \text{ or } M)$.

(a) 0.18
(b) 0.27
(c) 0.45
[✓ (d)] 0.78
(e) 1.04

Solution:
c
$P(L \text{ or } M) = P(L) + P(M) - P(L \text{ and } M) = \frac{85}{192} + \frac{115}{192} - \frac{50}{192} = \frac{150}{192} = 0.78125$

45. A box contains 18 balls, some of them are black and the others are white. We do not know how many are black and how many are white. Let n be the number of black balls and so $18 - n$ is the number of white balls. We are going to pick four balls at random from this box, one by one, without replacement (i.e when a ball is picked, we do not put it back in the box). Let A denote the event that the first three balls selected are black and B denote the event that the fourth ball selected is black. If $P(B|A) = 1/5$, what is the value of n?

    $\boxed{\text{(a)}}$ 6
    (b) 7
    (c) 8
    (d) 9
    (e) 12

    Solution:
    a
    After taking out the first 3 balls there are 15 left. $P(B|A) = 1/5$ means the probability of selecting a black ball from these remaining 15 balls is $1/5$. This means there must be 3 $(15 \times \frac{1}{5})$ black balls in these 15 remaining balls. This means at the beginning there were $n = 3 + 3 = 6$ black balls.

    Or second method:
    $P(B|A) = \frac{n-3}{18-3} = 1/5 \implies n = 6$

46. The probability that a patient with a certain disease will be successfully treated with a new medical treatment is 0.80. The success of the treatment is independent from one patient to another. Suppose that the treatment is used on 40 patients with the disease. What is the mean of the number of patients who will **not** be successfully treated?

    $\boxed{\text{(a)}}$ 8
    (b) 20
    (c) 32
    (d) 40
    (e) Impossible to say given the information provided

    Solution:
    (a)
    The mean of the number of patients who will **not** be successfully treated is $n(1-p) = 40 \cdot 0.2 = 8$

47. The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.
Use this information for this question and the next one.
If I select at random two Canada Post letters, what is the probability that they will both weigh less than 1 oz.?

[✓ (a)] 0.0005
(b) 0.0228
(c) 0.0456
(d) 0.4772
(e) 0.8413

Solution:
(a) The probability that one letter weigh less than 1 oz. is $P(X < 1) = P(Z < -2) = 0.0228$. The probability that they will both weigh less than 1 oz is 0.0228^2 (multiplication rule for independent events).

48. Same information as in the previous question.
The weight of Canada Post letters is normally distributed with a mean of 2 oz. and a standard deviation of 0.5 oz.
If I select at random 1000 Canada Post letters, what is the probability that at least 200 will weigh less than 1 oz.?

[✓ (a)] 0.0000
(b) 0.0046
(c) 0.1040
(d) 0.8960
(e) 1.0000

Solution:
(a) The number of letters that weigh less than 1 oz follows a $B(n = 1000, p = 0.0228)$. We can use the normal approximation since $np = 22.8$ and $n(1-p) = 977.2$. The probability that at least 200 letters will weigh less than 1 oz. is $P(X >= 200) = P(Z >= (200 - 22.8)/\sqrt{22.280}) = 1 - P(Z < 37.54) = 0$

49. If the correlation coefficient r between two variables is equal to 1, what do we know **for sure**?

    $\boxed{\text{(a)}}$ There is a perfect positive linear relationship between the two variables.

    (b) There is a very strong, but not perfect, positive linear relationship between the two variables.

    (c) There is no relationship between the two variables.

    (d) There is a perfect negative linear relationship between the two variables.

    (e) It is impossible to know one of the above for sure given the information provided.

    Solution:
    (a)

50. In a statistics class, there are 60 men and 40 women. The mean weight of all these 100 students is 160 pounds. If the mean weight of the 60 men is 180 pounds, what is the mean weight of the 40 women?

    (a) 120 pounds

    (b) 125 pounds

    $\boxed{\text{(c)}}$ 130 pounds

    (d) 132 pounds

    (e) 135 pounds

    Solution:
    c
    $\frac{100\times160-60\times180}{40} = 130\text{lb}$

Tables of normal, binomial and t-distributions

prepared by Ken Butler

2015-06-03

Table Z (normal distribution)

Values of z greater than 0 are on the next page. Second decimal place is at the top of the column.

	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00
-3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.6	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0002	0.0002
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
-3.3	0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005
-3.2	0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007
-3.1	0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010
-3.0	0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013
-2.9	0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019
-2.8	0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026
-2.7	0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035
-2.6	0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047
-2.5	0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062
-2.4	0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082
-2.3	0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107
-2.2	0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139
-2.1	0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179
-2.0	0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228
-1.9	0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287
-1.8	0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359
-1.7	0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446
-1.6	0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548
-1.5	0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668
-1.4	0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808
-1.3	0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968
-1.2	0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151
-1.1	0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357
-1.0	0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587
-0.9	0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841
-0.8	0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119
-0.7	0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420
-0.6	0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743
-0.5	0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085
-0.4	0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446
-0.3	0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821
-0.2	0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207
-0.1	0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602
0.0	0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000

Table Z (continued)

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

Table of the Binomial Distribution

(Table shows probability of exactly k successes in n trials with success probability p. Values of p are shown along the top of the table.)

		p									
n	k	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
2	1	0.0950	0.1800	0.2550	0.3200	0.3750	0.4200	0.4550	0.4800	0.4950	0.5000
2	2	0.0025	0.0100	0.0225	0.0400	0.0625	0.0900	0.1225	0.1600	0.2025	0.2500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
3	1	0.1354	0.2430	0.3251	0.3840	0.4219	0.4410	0.4436	0.4320	0.4084	0.3750
3	2	0.0071	0.0270	0.0574	0.0960	0.1406	0.1890	0.2389	0.2880	0.3341	0.3750
3	3	0.0001	0.0010	0.0034	0.0080	0.0156	0.0270	0.0429	0.0640	0.0911	0.1250
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
4	1	0.1715	0.2916	0.3685	0.4096	0.4219	0.4116	0.3845	0.3456	0.2995	0.2500
4	2	0.0135	0.0486	0.0975	0.1536	0.2109	0.2646	0.3105	0.3456	0.3675	0.3750
4	3	0.0005	0.0036	0.0115	0.0256	0.0469	0.0756	0.1115	0.1536	0.2005	0.2500
4	4	0.0000	0.0001	0.0005	0.0016	0.0039	0.0081	0.0150	0.0256	0.0410	0.0625
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
5	1	0.2036	0.3281	0.3915	0.4096	0.3955	0.3602	0.3124	0.2592	0.2059	0.1563
5	2	0.0214	0.0729	0.1382	0.2048	0.2637	0.3087	0.3364	0.3456	0.3369	0.3125
5	3	0.0011	0.0081	0.0244	0.0512	0.0879	0.1323	0.1812	0.2304	0.2757	0.3125
5	4	0.0000	0.0004	0.0022	0.0064	0.0147	0.0284	0.0488	0.0768	0.1128	0.1563
5	5	0.0000	0.0000	0.0001	0.0003	0.0010	0.0024	0.0053	0.0102	0.0185	0.0313
6	0	0.7351	0.5315	0.3772	0.2622	0.1780	0.1177	0.0754	0.0467	0.0277	0.0156
6	1	0.2321	0.3543	0.3993	0.3932	0.3560	0.3025	0.2437	0.1866	0.1359	0.0938
6	2	0.0306	0.0984	0.1762	0.2458	0.2966	0.3241	0.3280	0.3110	0.2780	0.2344
6	3	0.0021	0.0146	0.0415	0.0819	0.1318	0.1852	0.2355	0.2765	0.3032	0.3125
6	4	0.0001	0.0012	0.0055	0.0154	0.0330	0.0595	0.0951	0.1382	0.1861	0.2344
6	5	0.0000	0.0001	0.0004	0.0015	0.0044	0.0102	0.0205	0.0369	0.0609	0.0938
6	6	0.0000	0.0000	0.0000	0.0001	0.0002	0.0007	0.0018	0.0041	0.0083	0.0156