

STATS 2035 FINAL EXAM PRACTICE QUESTIONS
(Chapters 10, 11, 12, and 14)

1. A dummy variable
 - a. varies continuously over the range 0 to 1
 - b. is not a true variable, because it has a zero variance
 - c. can be used to measure a numerical value such as salary
 - d. can be used to measure, for example, whether or not a person owns a driver's license.

2. If Y rises linearly with X over a wide range, we can reduce the standard error of the slope by choosing the X values
 - a. Wide apart, so that $\sum(X - \bar{X})^2$ is large
 - b. Close together, so that $\sum(X - \bar{X})^2$ is large
 - c. Close together, so that $\sum(X - \bar{X})^2$ is small, and consequently the SE is small
 - d. So that $\bar{X} = 0$, and consequently $\sum X = 0$.

3. Income Y was regressed against experience E and sex M (0 for women, 1 for men):
$$Y = 21 + 0.3E - 0.7M + 0.1ME$$
The graph of Y against E would show a line for men and a line for women, where
 - a. both slope down.
 - b. both slope up.
 - c. the men's slopes up, the women's slopes down.
 - d. the men's slopes down, the women's slopes up.

4. Each year for 30 years the average values of hay yield H, rainfall R, and temperature T were recorded. Then a multiple regression was computed:
$$H = 10 + 0.07T - 7.3R$$
To clarify the relation of H to T, we could draw several "level curves" as follows: H is plotted against T for R = 10; then H is plotted against T for R = 12; similarly for several more values, say R = 14, 16, and 18 -- five altogether.

The five curves we thus obtain would be parallel straight lines,

 - a. positively sloped, and equally spaced.
 - b. positively sloped, and more widely spaced as we move up.
 - c. negatively sloped, and equally spaced.
 - d. negatively sloped, and more widely spaced as we move up.

5. To determine how sales were related to four important characteristics of the sales staff, a multiple regression was run on the data from 15 sales agents randomly drawn from a large farm machinery corporation:

$$Y = 11.2 + 2.1X_1 + 0.8X_2 - 7.6X_3 + 1.2X_4$$

(SE) (0.3) (1.2) (1.3) (0.7)

residual s = 4.52

The 95% confidence interval for β_2 is approximately

- a. $0.8 \pm 1.96 \cdot 1.2$
 - b. $0.8 \pm 2.16 \cdot 1.2$
 - c. $0.8 \pm 2.23 \cdot 1.2$
 - d. not a meaningful concept, because the population is not infinite.
6. A regression, with summary statistics, was computed from a random sample of $n = 500$ observations of (X, Y) :

$$Y = -1.9 + 0.335X$$

$\bar{X} = 46$; sum of $(X - \bar{X})^2 = 29,000$, hence $s_x^2 = 58$
 regression residual $s_e^2 = 7.1$

The 95% confidence interval for the population slope is therefore

- a. 0.335 ± 0.234 b. 0.335 ± 0.668
 - c. 0.335 ± 0.031 d. 0.335 ± 0.088
7. For 30 randomly selected areas, values of hay yield H , rainfall R , and temperature T were recorded. Then the following regression was computed, along with the standard errors of the coefficients:

$$H = 15.2 + 0.261T + 3.65R$$

(SE) (0.086) (2.11)

What is t and the p -value (one sided) for the null hypothesis that the population coefficient of T is zero?

- a. $t = 3.03$, $p < 0.005$
 - b. $t = 3.03$, $p < 0.0025$
 - c. $t = 0.33$, $p > 0.25$
 - d. $t = 16.6$, $p < 0.0005$
8. Suppose U and W increase 10 units, while V remains unchanged. If T is related to all these by the multiple regression $T = 15 + 2U + 6V - 12W$, we estimate the accompanying change in T would be
- a. down 85
 - b. down 100
 - c. no change
 - d. down 85, or stay unchanged (with chances 67% and 33%)

9. For a random sample of 30 baseball players from the National League, batting averages in 2001 (X) and 2002 (Y) were run through a computer program to give the following regression equation and summary statistics: $Y = 0.109 + 0.620X$

$$\text{residual } s_e = 0.013 \qquad \bar{X} = 0.262 \qquad \bar{Y} = 0.271$$

$$\text{sum of } (X - \bar{X})^2 = 0.0094, \text{ hence } s_x = 0.018$$

$$\text{sum of } (Y - \bar{Y})^2 = 0.0083, \text{ hence } s_y = 0.017$$

To predict the batting average Y for a player (drawn at random) who batted $X = 0.310$ in 2001, we have 95% confidence in the interval:

- a. 0.301 ± 0.014 b. 0.310 ± 0.014 c. 0.301 ± 0.030 d. 0.310 ± 0.030

10. A random sample of $n = 25$ daily production figures at a hosiery mill in North Carolina in 2001 gave some interesting statistics on output X and cost Y:—

$$\bar{X} = 53; \text{ sum of } (X - \bar{X})^2 = 8,660, \text{ hence } s_x = 19.0$$

$$\bar{Y} = 0.87; \text{ sum of } (Y - \bar{Y})^2 = 0.135, \text{ hence } s_y = 0.075$$

As well, the computer gave the regression equation:

$$Y = 1.05 - 0.0034X \qquad \text{residual } s_e = 0.039$$

Suppose we wanted to predict the production cost for an average output of 70 units. What would the 95% confidence interval be for the cost associated with this long-run average output?

- a. 0.812 ± 0.0836 b. 0.812 ± 0.0219
 c. 0.812 ± 0.0207 d. 0.812 ± 0.0792

11. What is the best interpretation of a multiple regression coefficient? It is the change in Y that occurs when

- a. that regressor changes one unit, and the other regressors change correspondingly.
 b. that regressor changes one unit, and all the other regressors are constant.
 c. that regressor remains constant, and all the other regressors change one unit.
 d. that regressor changes by the amount s_y/s_x -- where s_y and s_x are the standard deviations of Y and the given regressor X, respectively.

12. In simple regression and correlation, as X increases one unit, by how much does Y increase on average?

- a. r b. $r*s_x$ c. $r*s_y/s_x$ d. $r*s_x/s_y$

13. Y will be most accurately predicted from another variable X when the correlation of X with Y is

- a. 0.60 b. -0.20 c. -0.70 d. 0.50

14. If Y tends to increase as X increases, what can we conclude about the correlation r?

- a. $r > 1$ b. $r < -1$ c. $r > 0$ d. $r < 0$

15. The correlation between age of an auto and money spent for repairs is 0.64. Of the variation in money spent for repairs, the proportion accounted for by the age of the auto is therefore:
- a. 64% b. 41% c. 80% d. 36%
16. When is chi-square suitable, instead of some other technique such as t-tests or regression?
- a. Chi-square is an extension of regression, just as the multinomial is an extension of the binomial.
 b. Chi-square is an extension of the two-sample t-test, just as the multinomial is an extension of the binomial.
 c. Chi-square is suitable for data counted off into levels of a categorical variable such as nationality or religion.
 d. Chi-square is suitable for data counted off into levels of a numerical variable such as volume of sales.
17. Suppose men always married women who were exactly 4 inches shorter than themselves. What would be the correlation between husbands' and wives' heights?
- a. undefined b. 0.25 c. -0.25 d. 1.00
18. To determine whether the percentage of cars with defective trim varied from day to day on an assembly line, a random sample of 200 cars was taken on each of the 5 days of the week, producing the following statistics:

Day	Number With Major Defects	Number With Minor Defects	Number O.K.	Totals
Mon.	8	22	170	200
Tues.	2	10	188	200
Wed.	6	16	178	200
Thurs.	2	8	190	200
Fri.	12	34	154	200
Totals	30	90	880	1,000

What is the appropriate chi-square test of the null hypothesis?

- a. The test of independence--a test of whether the sample proportions can be expressed as $P_{ij} = P_i * P_j$
 b. The test of homogeneity (independence) -- a test of whether all five days are equal in their proportions of defects.
 c. The goodness-of-fit test of the multinomial distribution (probabilities of 1/5 in each cell).
 d. The goodness-of-fit test of the multinomial distribution (probabilities of 1/5 for each day).

19.

		average		st. dev
X	23	16	20	21
Y	8	13	9	10
		20		2.94
		10		2.16

The correlation r is

- a. -2.83 b. -0.69 c. -0.05 d. -0.94

20. For her doctoral dissertation, a business student gathered data on 130 corporations. She calculated the following statistics for annual earnings E and annual charitable donations D :

$$\bar{E} = 12 \quad \text{sum of } (E - \bar{E})^2 = 12,000$$

$$\bar{D} = 1.6 \quad \text{sum of } (D - \bar{D})^2 = 340$$

$$\text{sum of } (E - \bar{E})(D - \bar{D}) = 760$$

Using linear regression, estimate the annual donation for a corporation whose annual earning is 20.

- a. 1.1 b. 19.5 c. 2.1 d. 2.9

21. If X and Y have a correlation of 0.81, the variation of Y that is accounted for by the variation in X is:

- a. 90% b. 66% c. 81% d. 19%

22. Use the following information to answer the next TWO questions:
In a randomized block design, data was collected on five treatments and four times

source	SS	df
between treatments	80	4
between times	300	3
residual		
total	480	

Assume a simple additive model as usual (no interaction). By completing the ANOVA table above, find the F ratio to test the null hypothesis of no difference between treatments.

- a. $(80/4) / (100/19)$ b. $(80/4) / (100/12)$
c. $(80/4) / (480/19)$ d. $(80/4) / (300/3)$

23. After conducting the test in the last question, you would decide to:

- a. Reject H_0 and conclude that the result is statistically significant.
b. Do not reject H_0 and conclude that the result is not statistically significant.
c. Reject H_0 and conclude that the result is not statistically significant.
d. Do not reject H_0 and conclude that the result is statistically significant.

24. The correlation of X and Y is equal to zero when

- a. the residual variance s_e^2 is zero.
- b. the coefficient of determination is 1
- c. $\bar{X} = 0$ and $\bar{Y} = 0$
- d. the regression slope b is zero.

25. For 30 randomly selected areas, values of hay yield H, rainfall R, and temperature T were recorded. Then the following regression was computed, along with the standard errors of the coefficients:

$$\begin{array}{rcccc}
 H & = & 15.2 & + & 0.261T & + & 3.65R \\
 (SE) & & & & (0.086) & & (2.11)
 \end{array}$$

What is t and the p-value (one sided) for the null hypothesis that the population coefficient of T is zero?

- a. $t = 3.03, p < 0.005$
- b. $t = 3.03, p < 0.0025$
- c. $t = 0.33, p > 0.25$
- d. $t = 16.6, p < 0.0005$

Use the following information for the next TWO questions:

To determine how effective 3 different treatments were against 5 different poisons, each of 15 rats were randomly assigned to one of the 15 combinations. The survival times (minutes) are given below:

Treatment	Poison					Means
	1	2	3	4	5	
A						42
B						30
C						48
Means	46	49	43	30	32	40

26. Treating the poisons as the blocking factor, what is the sum of squares for poisons (SSBL)?

- a. 290
- b. 840
- c. 870
- d. 1450

27. For comparing treatments, the observed test statistic is $F_{tr} = 4.91$. For comparing poisons, the observed test statistic is $F_{bl} = 2.55$. If you are testing the null hypotheses of equality of the treatment means and the equality of the poison means at the 5% level of significance, which of the following are the correct decisions?

- | | <u>Equality of Treatments</u> | <u>Equality of Poisons</u> |
|----|-------------------------------|----------------------------|
| a. | Reject H_0 | Reject H_0 |
| b. | Do not reject H_0 | Do not reject H_0 |
| c. | Do not reject H_0 | Reject H_0 |
| d. | Reject H_0 | Do not reject H_0 |

28. Three new incentive plans for sales staff were tried out, each on 25 randomly selected new employees. The first year's sales figures of these 75 people were analyzed in the following ANOVA table:

Source	Sum of Squares
Differences between plans (treatments)	107
Error	
Total	1087

What is the value of the test statistic to test the hypothesis of no difference between the 3 plans?

- a. 2.62 b. 3.64 c. 3.93 d. 9.16
- 29 It is thought that stress may increase susceptibility to illness through suppression of the immune system. To investigate this theory, an experiment was performed on 38 rats randomly allocated to four treatment groups: no stress, mild stress, moderate stress and high stress. Susceptibility to illness was measured by measuring the concentration of lymphocytes in each rat's blood (low concentration means more susceptible to illness). The summary statistics are given below:

	No Stress	Mild Stress	Moderate Stress	High Stress
Mean	6.64	5.42	3.84	2.50
Std. Dev.	2.77	2.42	3.07	1.56
sample size	9	11	10	8

In an ANOVA table, what is the value of the error sum of squares?

- a. 84.91 b. 94.73 c. 221.81 d. 247.19
30. A university had four candidates running for the position of student council president. A random sample of 100 students gave the following distribution of votes for the candidates:

John	Jim	Karen	Sally
15	20	35	30

When the campaign started, the student newspaper predicted that the proportion of votes for the candidates would be 10% each for John and Jim and 40% each for Karen and Sally. In order to test whether the newspaper's hypothesis is correct, what would be the value of the appropriate test statistic and what would be the decision if testing at the 1% level of significance?

- a. 10.714, do not reject the paper's hypothesis
 b. 10.714, reject the paper's hypothesis
 c. 15.625, do not reject the paper's hypothesis
 d. 15.625, reject the paper's hypothesis

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
d	a	b	a	c	c	a	b	c	b	b	c	c	c	b
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
c	d	b	d	c	b	b	b	d	a	c	d	c	c	d