YORK UNIVERSITY
ATKINSON COLLEGE
Department of Administrative Studies
AK/ADMS 2320.3.0 - Quantitative Methods 1

FALL 2005 - FINAL EXAMINATION

December 10, 2005 **Time Allowed: 3 hours**

H. Bartel	**Section B**	☐	**Section C**	☐						
A. Marshall	**Section A**	☐	**Section K**	☐						
P. Ng	**Section I**	☐								
R. Huang	**Section H**	☐								
S. Abdullah	**Section L**	☐								
G. O'Keefe	**Section D**	☐	**Section E**	☐	**Section F**	☐	**Section G**	☐	**Section J**	☐

INSTRUCTIONS:

1. **You must write this exam in the section in which you are enrolled.**
2. **This is a closed book examination.**
3. Answer all questions on examination paper where space is provided. **YOU MUST SHOW ALL YOUR WORK.**
4. **Record Multiple Choice answers in pencil on the Scantron sheet provided.**
5. Aids allowed: writing utensils, eraser, silent, **non-programmable** calculator, two sides of 8 ½" x 11" formula sheet, and English paper-based dictionary.
6. Calculations are to be to the **4th decimal place.**
7. **YOU MUST HAND IN THIS EXAM BEFORE LEAVING THE EXAMINATION ROOM. FAILURE TO DO SO WILL RESULT IN AN AUTOMATIC F GRADE.**

____
Student's Name (Print)

____
Student Number

____
Signature

MARKS:	
Question 1: (7)	
Question 2: (25)	
Question 3: (15)	
Question 4: (13)	
Question 5: (20)	
Question 6: (20)	
TOTAL: (100)	

THIS EXAM CONTAINS 28 PAGES (INCLUDING COVER PAGE AND TABLES).

DO NOT REMOVE ANY PAGES

Question 1 (7 marks)

As a manufacturer of golf clubs, a major corporation wants to estimate the proportion of golfers who are right-handed. The manufacturer has information from a previous study which suggested that 75% of golfers are right-handed.

How many golfers must be surveyed if they want to be within 0.02, with a 95% confidence?

Question 2 (25 marks)

Random samples of two brands of whole milk are checked for their fat content as follows: 33 half-gallon containers of each brand are selected, and the fat content is measured (in grams).

The resulting data are shown in the accompanying table.

BRAND A				BRAND B		
30	26	36		24	33	17
26	33	35		27	20	21
31	20	32		22	18	18
27	28	27		31	26	25
37	27	29		30	25	27
28	31	33		25	20	24
31	35	27		22	22	24
29	30	30		26	24	20
25	26	36		29	33	18
27	29	30		22	25	20
25	31	27		30	26	27
$\sum x_A = 974$ $\sum x_A^2 = 29{,}200$				$\sum x_B = 801$ $\sum x_B^2 = 20{,}037$		

a. Can we conclude at the 5% significance level that the variability in fat content of Brand A is different from that of Brand B?

b. Estimate with 95% confidence the difference in mean fat content between Brand A and Brand B.

c. Any container that contains 30 g or more of fat is considered unacceptable.

Is there enough evidence to conclude that Brand A has a higher fraction of unacceptable containers than Brand B? (Use $\alpha = .01$)

d. What is the p-value of the test in part (c)?

e. The owner of Brand B claims that no more than 10% of his containers are unacceptable.

Is there sufficient evidence at the 5% significance level to refute this claim?

f. Estimate with 90% confidence the fraction of unacceptable Brand A containers.

Question 3 (15 marks)

A major insurance firm interviewed a random sample of 1200 university students to find out the type of life insurance preferred, if any.

The results follow:

	Insurance Preference		
Gender	**Term**	**Whole Life**	**No Insurance**
Female	100	80	325
Male	160	60	475

Is there evidence that life insurance preference of male students is different than that of female students? Test using the 5% level of significance.

Question 4 (13 marks)

Calculate the probability of a Type II error for the following test of hypothesis $H_0 : \mu = 50$ vs. $H_1 : \mu > 50$, given that $\mu = 55$, $\alpha = 0.05, \sigma = 10,$ and n = 16.

Question 5 (20 marks)

A financier whose specialty is investing in movie productions has observed that, in general, movies with “big-name” stars seem to generate more revenue than those movies whose stars are less well known. To examine his belief he records the gross revenue and the payment (in $ millions) given to the two highest-paid performers in the movie for ten recently released movies.

Movie	Cost of 2 Highest Paid Performers ($millions)	Gross Revenue ($millions)
1	5.3	48
2	7.2	65
3	1.3	18
4	1.8	20
5	3.5	31
6	2.6	26
7	8.0	73
8	2.4	23
9	4.5	39
10	6.7	58

a. Find the sample regression line

b. Interpret the coefficients.

c. Determine the standard error of estimate and describe what this statistic tells you about the regression line.

d. Determine the coefficient of determination and discuss what its value tells you about the two variables.

e. Is there a linear relationship between payment to the two highest-paid performers and gross revenue?

b. Predict with 95% confidence the Gross Revenue if the cost of two highest paid performers is $5 million.

Question 6: MULTIPLE CHOICE – *Answer on Scantron Sheet*

(ONLY use PENCIL and PRESS HARD, fill in LAST NAME, FIRST NAME, STUDENT NUMBER, Answers)

1. From a population that is not normally distributed and whose standard deviation is not known, a sample of 6 items is selected to develop an interval estimate for the mean of the population (μ).
   a. The normal distribution can be used.
   b. The t distribution with 5 degrees of freedom must be used.
   c. The t distribution with 6 degrees of freedom must be used.
   d. The sample size must be increased.

2. In which of the following are the two events A and B, always independent?
   a. A and B are mutually exclusive.
   b. The probability of event A is not influenced by the probability of event B.
   c. $P(A|B) = P(A)$.
   d. b and c.

3. A fair die is rolled 10 times. What is the probability that an odd number (1, 3, or 5) will occur less than 3 times?
   a. .055
   b. .117
   c. .775
   d. .155

4. All of the following are assumptions of the error terms in the simple linear regression model except
   a. Errors are normally distributed.
   b. Error terms have a mean of zero.
   c. Error terms have a constant variance.
   d. Error terms indicate a positive autocorrelation.

5. The simple linear regression (least squares method) minimizes:
   a. The explained variation
   b. SSR
   c. SSE
   d. Total variation

6. After computing a confidence interval, the user believes the results are meaningless because the width of the interval is too large. Which one of the following is the best recommendation?
   a. Increase the level of confidence for the interval.
   b. Decrease the sample size.
   c. Increase the sample size.
   d. Reduce the population variance.

7. The level of significance is the
   a. maximum allowable probability of Type II error
   b. maximum allowable probability of Type I error
   c. same as the confidence coefficient
   d. same as the *p*-value

8. The regression line $\hat{y} = 3 + 2x$ has been fitted to the data points (4, 8), (2, 5), and (1, 2). The sum of the squared residuals will be:
   a. 7
   b. 15
   c. 8
   d. 22

9. In a simple linear regression problem, the following sum of squares are produced: $\sum(y_i - \bar{y})^2 = 200$, $\sum(y_i - \hat{y}_i)^2 = 50$, and $\sum(\hat{y}_i - \bar{y})^2 = 150$. The percentage of the variation in *y* that is explained by the variation in *x* is:
   a. 25%
   b. 75%
   c. 33%
   d. 50%

10. If the coefficient of determination is 0.975, then the slope of the regression line:
    a. must be positive
    b. must be negative
    c. could be either positive or negative
    d. None of the above.

11. Which of the following statistics and procedures can be used to determine whether a linear model should be employed?
    a. The standard error of estimate
    b. The coefficient of determination
    c. The *t*-test of the slope
    d. All of the above

12. In regression analysis, the coefficient of determination R^2 measures the amount of variation in *y* that is:
    a. caused by the variation in *x*
    b. explained by the variation in *x*
    c. unexplained by the variation in *x*
    d. None of the above

13. Given a specific value of x and confidence level, which of the following statements is correct?
    a. The confidence interval estimate of the expected value of y can be calculated but the prediction interval of y for the given value of x cannot be calculated.
    b. The confidence interval estimate of the expected value of y will be wider than the prediction interval.
    c. The prediction interval of y for the given value of x can be calculated but the confidence interval estimate of the expected value of y cannot be calculated.
    d. The confidence interval estimate of the expected value of y will be narrower than the prediction interval.

14. In an ANOVA test, the test statistic is $F = 6.75$. The rejection region is $F > 3.97$ for the 5% level of significance, $F > 5.29$ for the 2.5% level, and $F > 7.46$ for the 1% level. For this test, the p-value is
    a. greater than 0.05
    b. between 0.025 and 0.05
    c. between 0.01 and 0.025
    d. approximately 0.05

15. If we want to conduct a test to determine whether a population mean is greater than another population mean, we
    a. can use the analysis of variance
    b. can use the independent samples t-test for difference between two means
    c. can use the chi-squared test
    d. All of the above.

16. One-way ANOVA is applied to three independent samples having means 10, 13, and 18, respectively. If each observation in the third sample were increased by 30, the value of the F-statistics would:
    a. increase
    b. decrease
    c. remain unchanged
    d. increase by 30

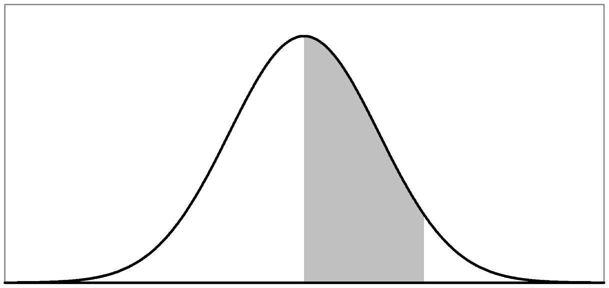
17. In chi-squared tests, the conventional and conservative rule – known as the *rule of five* – is to require that the:
    a. observed frequency for each cell be at least five
    b. degrees of freedom for the test be at least five
    c. expected frequency for each cell be at least five
    d. difference between the observed and expected frequency for each cell be at least five

18. One-way ANOVA is performed on independent samples taken from three normally distributed populations with equal variances. The following summary statistics were calculated: $n_1 = 6$, $\bar{x}_1 = 50$, $s_1 = 5.2$, $n_2 = 8$, $\bar{x}_2 = 55$, $s_2 = 4.9$, $n_3 = 6$, $\bar{x}_3 = 51$, and $s_3 = 5.4$.
The grand mean equals
a. 50.0
b. 52.0
c. 52.3
d. 53.0

19. To determine whether a single coin is fair, the coin was tossed 100 times, and head was observed 60 times. The value of the test statistic could be
a. 2
b. 4
c. 6 or 4
d. 4 or 2

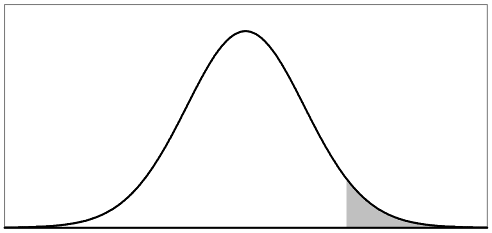
20. A chi-squared goodness-of-fit test is always conducted as:
a. an upper-tail test
b. a lower-tail test
c. a two-tail test
d. All of the above

Normal Probabilities



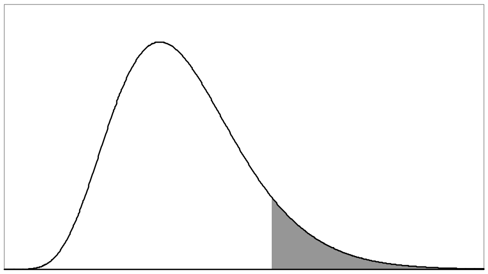
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Critical Values of *t*



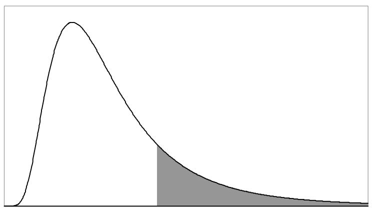
DEGREES OF FREEDOM	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	DEGREES OF FREEDOM	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.656	24	1.318	1.711	2.064	2.492	2.797
2	1.886	2.920	4.303	6.965	9.925	25	1.316	1.708	2.060	2.485	2.787
3	1.638	2.353	3.182	4.541	5.841	26	1.315	1.706	2.056	2.479	2.779
4	1.533	2.132	2.776	3.747	4.604	27	1.314	1.703	2.052	2.473	2.771
5	1.476	2.015	2.571	3.365	4.032	28	1.313	1.701	2.048	2.467	2.763
6	1.440	1.943	2.447	3.143	3.707	29	1.311	1.699	2.045	2.462	2.756
7	1.415	1.895	2.365	2.998	3.499	30	1.310	1.697	2.042	2.457	2.750
8	1.397	1.860	2.306	2.896	3.355	35	1.306	1.690	2.030	2.438	2.724
9	1.383	1.833	2.262	2.821	3.250	40	1.303	1.684	2.021	2.423	2.704
10	1.372	1.812	2.228	2.764	3.169	45	1.301	1.679	2.014	2.412	2.690
11	1.363	1.796	2.201	2.718	3.106	50	1.299	1.676	2.009	2.403	2.678
12	1.356	1.782	2.179	2.681	3.055	60	1.296	1.671	2.000	2.390	2.660
13	1.350	1.771	2.160	2.650	3.012	70	1.294	1.667	1.994	2.381	2.648
14	1.345	1.761	2.145	2.624	2.977	80	1.292	1.664	1.990	2.374	2.639
15	1.341	1.753	2.131	2.602	2.947	90	1.291	1.662	1.987	2.368	2.632
16	1.337	1.746	2.120	2.583	2.921	100	1.290	1.660	1.984	2.364	2.626
17	1.333	1.740	2.110	2.567	2.898	120	1.289	1.658	1.980	2.358	2.617
18	1.330	1.734	2.101	2.552	2.878	140	1.288	1.656	1.977	2.353	2.611
19	1.328	1.729	2.093	2.539	2.861	160	1.287	1.654	1.975	2.350	2.607
20	1.325	1.725	2.086	2.528	2.845	180	1.286	1.653	1.973	2.347	2.603
21	1.323	1.721	2.080	2.518	2.831	200	1.286	1.653	1.972	2.345	2.601
22	1.321	1.717	2.074	2.508	2.819	Inf.	1.282	1.645	1.960	2.326	2.576
23	1.319	1.714	2.069	2.500	2.807						

Critical Values of χ^2



DEGREES OF FREEDOM	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.0000	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2104	10.5965
3	0.0717	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449	12.8381
4	0.2070	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767	14.8602
5	0.4118	0.5543	0.8312	1.1455	1.6103	9.2363	11.0705	12.8325	15.0863	16.7496
6	0.6757	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916	14.4494	16.8119	18.5475
7	0.9893	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671	16.0128	18.4753	20.2777
8	1.3444	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073	17.5345	20.0902	21.9549
9	1.7349	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190	19.0228	21.6660	23.5893
10	2.1558	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070	20.4832	23.2093	25.1881
11	2.6032	3.0535	3.8157	4.5748	5.5778	17.2750	19.6752	21.9200	24.7250	26.7569
12	3.0738	3.5706	4.4038	5.2260	6.3038	18.5493	21.0261	23.3367	26.2170	28.2997
13	3.5650	4.1069	5.0087	5.8919	7.0415	19.8119	22.3620	24.7356	27.6882	29.8193
14	4.0747	4.6604	5.6287	6.5706	7.7895	21.0641	23.6848	26.1189	29.1412	31.3194
15	4.6009	5.2294	6.2621	7.2609	8.5468	22.3071	24.9958	27.4884	30.5780	32.8015
16	5.1422	5.8122	6.9077	7.9616	9.3122	23.5418	26.2962	28.8453	31.9999	34.2671
17	5.6973	6.4077	7.5642	8.6718	10.0852	24.7690	27.5871	30.1910	33.4087	35.7184
18	6.2648	7.0149	8.2307	9.3904	10.8649	25.9894	28.8693	31.5264	34.8052	37.1564
19	6.8439	7.6327	8.9065	10.1170	11.6509	27.2036	30.1435	32.8523	36.1908	38.5821
20	7.4338	8.2604	9.5908	10.8508	12.4426	28.4120	31.4104	34.1696	37.5663	39.9969
21	8.0336	8.8972	10.2829	11.5913	13.2396	29.6151	32.6706	35.4789	38.9322	41.4009
22	8.6427	9.5425	10.9823	12.3380	14.0415	30.8133	33.9245	36.7807	40.2894	42.7957
23	9.2604	10.1957	11.6885	13.0905	14.8480	32.0069	35.1725	38.0756	41.6383	44.1814
24	9.8862	10.8563	12.4011	13.8484	15.6587	33.1962	36.4150	39.3641	42.9798	45.5584
25	10.5196	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3140	46.9280
26	11.1602	12.1982	13.8439	15.3792	17.2919	35.5632	38.8851	41.9231	45.6416	48.2898
27	11.8077	12.8785	14.5734	16.1514	18.1139	36.7412	40.1133	43.1945	46.9628	49.6450
28	12.4613	13.5647	15.3079	16.9279	18.9392	37.9159	41.3372	44.4608	48.2782	50.9936
29	13.1211	14.2564	16.0471	17.7084	19.7677	39.0875	42.5569	45.7223	49.5878	52.3355
30	13.7867	14.9535	16.7908	18.4927	20.5992	40.2560	43.7730	46.9792	50.8922	53.6719
35	17.1917	18.5089	20.5694	22.4650	24.7966	46.0588	49.8018	53.2033	57.3420	60.2746
40	20.7066	22.1642	24.4331	26.5093	29.0505	51.8050	55.7585	59.3417	63.6908	66.7660
45	24.3110	25.9012	28.3662	30.6123	33.3504	57.5053	61.6562	65.4101	69.9569	73.1660
50	27.9908	29.7067	32.3574	34.7642	37.6886	63.1671	67.5048	71.4202	76.1538	79.4898
60	35.5344	37.4848	40.4817	43.1880	46.4589	74.3970	79.0820	83.2977	88.3794	91.9518
70	43.2753	45.4417	48.7575	51.7393	55.3289	85.5270	90.5313	95.0231	100.4251	104.2148
80	51.1719	53.5400	57.1532	60.3915	64.2778	96.5782	101.8795	106.6285	112.3288	116.3209
90	59.1963	61.7540	65.6466	69.1260	73.2911	107.5650	113.1452	118.1359	124.1162	128.2987
100	67.3275	70.0650	74.2219	77.9294	82.3581	118.4980	124.3421	129.5613	135.8069	140.1697

Critical Values of F, A = .05



ν_2 \ ν_1 (DENOMINATOR DEGREES OF FREEDOM \ NUMERATOR DEGREES OF FREEDOM)	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
Inf	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Critical Values of F, A = .05 (Continued)

ν_2 \ ν_1 (NUMERATOR DEGREES OF FREEDOM)	10	12	15	20	24	30	40	60	120	Inf
1	241.88	243.90	245.95	248.02	249.05	250.10	251.14	252.20	253.25	254.32
2	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
Inf	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

DENOMINATOR DEGREES OF FREEDOM

Critical Values of F, A = .025

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10	12
	NUMERATOR DEGREES OF FREEDOM										
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	963.3	968.6	976.7
2	38.51	39	39.17	39.25	39.3	39.33	39.36	39.37	39.39	39.4	39.41
3	17.44	16.04	15.44	15.1	14.88	14.73	14.62	14.54	14.47	14.42	14.34
4	12.22	10.65	9.98	9.6	9.36	9.2	9.07	8.98	8.9	8.84	8.75
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52
6	8.81	7.26	6.6	6.23	5.99	5.82	5.7	5.6	5.52	5.46	5.37
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.9	4.82	4.76	4.67
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.3	4.2
9	7.21	5.71	5.08	4.72	4.48	4.32	4.2	4.1	4.03	3.96	3.87
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43
12	6.55	5.1	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28
13	6.41	4.97	4.35	4	3.77	3.6	3.48	3.39	3.31	3.25	3.15
14	6.3	4.86	4.24	3.89	3.66	3.5	3.38	3.29	3.21	3.15	3.05
15	6.2	4.77	4.15	3.8	3.58	3.41	3.29	3.2	3.12	3.06	2.96
16	6.12	4.69	4.08	3.73	3.5	3.34	3.22	3.12	3.05	2.99	2.89
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82
18	5.98	4.56	3.95	3.61	3.38	3.22	3.1	3.01	2.93	2.87	2.77
19	5.92	4.51	3.9	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.8	2.73	2.64
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.7	2.6
23	5.75	4.35	3.75	3.41	3.18	3.02	2.9	2.81	2.73	2.67	2.57
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.7	2.64	2.54
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51
26	5.66	4.27	3.67	3.33	3.1	2.94	2.82	2.73	2.65	2.59	2.49
27	5.63	4.24	3.65	3.31	3.08	2.92	2.8	2.71	2.63	2.57	2.47
28	5.61	4.22	3.63	3.29	3.06	2.9	2.78	2.69	2.61	2.55	2.45
29	5.59	4.2	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41
40	5.42	4.05	3.46	3.13	2.9	2.74	2.62	2.53	2.45	2.39	2.29
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.22
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17
120	5.15	3.8	3.23	2.89	2.67	2.52	2.39	2.3	2.22	2.16	2.05
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94

Critical Values of F, A = .025

v_2 \ v_1	NUMERATOR DEGREES OF FREEDOM										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	963.3	968.6	976.7
2	38.51	39	39.17	39.25	39.3	39.33	39.36	39.37	39.39	39.4	39.41
3	17.44	16.04	15.44	15.1	14.88	14.73	14.62	14.54	14.47	14.42	14.34
4	12.22	10.65	9.98	9.6	9.36	9.2	9.07	8.98	8.9	8.84	8.75
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52
6	8.81	7.26	6.6	6.23	5.99	5.82	5.7	5.6	5.52	5.46	5.37
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.9	4.82	4.76	4.67
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.3	4.2
9	7.21	5.71	5.08	4.72	4.48	4.32	4.2	4.1	4.03	3.96	3.87
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43
12	6.55	5.1	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28
13	6.41	4.97	4.35	4	3.77	3.6	3.48	3.39	3.31	3.25	3.15
14	6.3	4.86	4.24	3.89	3.66	3.5	3.38	3.29	3.21	3.15	3.05
15	6.2	4.77	4.15	3.8	3.58	3.41	3.29	3.2	3.12	3.06	2.96
16	6.12	4.69	4.08	3.73	3.5	3.34	3.22	3.12	3.05	2.99	2.89
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82
18	5.98	4.56	3.95	3.61	3.38	3.22	3.1	3.01	2.93	2.87	2.77
19	5.92	4.51	3.9	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.8	2.73	2.64
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.7	2.6
23	5.75	4.35	3.75	3.41	3.18	3.02	2.9	2.81	2.73	2.67	2.57
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.7	2.64	2.54
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.51
26	5.66	4.27	3.67	3.33	3.1	2.94	2.82	2.73	2.65	2.59	2.49
27	5.63	4.24	3.65	3.31	3.08	2.92	2.8	2.71	2.63	2.57	2.47
28	5.61	4.22	3.63	3.29	3.06	2.9	2.78	2.69	2.61	2.55	2.45
29	5.59	4.2	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53	2.43
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41
40	5.42	4.05	3.46	3.13	2.9	2.74	2.62	2.53	2.45	2.39	2.29
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.22
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17
120	5.15	3.8	3.23	2.89	2.67	2.52	2.39	2.3	2.22	2.16	2.05
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94

MCQs

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
D	D	A	D	C	C	B	D	B	C	D	B	D	C	B	A	C	C	D	A

Question 1 (7 marks)

As a manufacturer of golf clubs, a major corporation wants to estimate the proportion of golfers who are right-handed. The manufacturer has information from a previous study which suggested that 75% of golfers are right-handed.
How many golfers must be surveyed if they want to be within 0.02, with a 95% confidence?

Solution:

We apply the *sample size to estimate proportion* formula: $\hat{p}$ = 0.75, W = 0.02.

For 95% confidence, $z_{\alpha/2}$ = 1.96

$$n = \left(\frac{z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})}}{W}\right)^2 = \left(\frac{1.96\sqrt{0.75(0.25)}}{0.02}\right)^2 = 1800.75 \text{ rounded up to 1801.}$$

1801 golfers must be surveyed if they want to be within 0.02, with a 95% confidence.

Question 2 (25 marks)

Random samples of two brands of whole milk are checked for their fat content as follows: 33 half-gallon containers of each brand are selected, and the fat content is measured (in grams).
The resulting data are shown in the accompanying table.

	BRAND A				**BRAND B**	
30	26	36		24	33	17
26	33	35		27	20	21
31	20	32		22	18	18
27	28	27		31	26	25
37	27	29		30	25	27
28	31	33		25	20	24
31	35	27		22	22	24
29	30	30		26	24	20
25	26	36		29	33	18
27	29	30		22	25	20
25	31	27		30	26	27
$\sum x_A = 974$ $\sum x_A^2 = 29,200$				$\sum x_B = 801$ $\sum x_B^2 = 20.037$		

a. *Can we conclude at the 5% significance level that the variability in fat content of Brand A is different from that of Brand B?*

H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$ $\qquad$ H_1: $\frac{\sigma_1^2}{\sigma_2^2} \neq 1$

$$s_1^2 = \frac{\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n}}{n-1} = \frac{29200 - \frac{974^2}{33}}{32} = 14.1326 \qquad s_2^2 = \frac{\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n}}{n-1} = \frac{20037 - \frac{801^2}{33}}{32} = 18.5795$$

$$F = \frac{s_1^2}{s_2^2} = \frac{14.1326}{18.5795} = 0.761$$

At α = 0.05 the rejection region is

$$F < \frac{1}{F_{\alpha/2,\nu 2,\nu 1}} = \frac{1}{F_{0.025,32,32}} \approx \frac{1}{2.07} = 0.483 \qquad \text{or} \qquad F > F_{\alpha/2,\nu 1,\nu 2} = F_{0.025,32,32} = 2.07$$

Since F stat = 0.761 is not in the rejection region, we fail to reject the null hypothesis.

There is not enough evidence to infer that the variability in fat content of Brand A is different from that of Brand B at the 5% significance level.

b. Estimate with 95% confidence the difference in mean fat content between Brand A and Brand B.

Since we do not reject H_0 in part a), we assume equal variances.

$$\bar{x}_1 = \frac{\sum x_A}{n_A} = \frac{974}{33} = 29.5152 \qquad \bar{x}_2 = \frac{\sum x_B}{n_B} = \frac{801}{33} = 24.2727$$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{32(14.1326)+32(18.5795)}{33+33-2} = 16.3561$$

Thus the interval estimate is

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = (29.5152 - 24.2727) \pm 1.997\sqrt{16.3561\left(\frac{1}{33}+\frac{1}{33}\right)}$$

5.2425 $\pm$ 1.9890

(3.2535, 7.2315)

LCL = 3.2535 $\qquad$ **UCL = 7.2315**

c. Any container that contains 30 g or more of fat is considered unacceptable. Is there enough evidence to conclude that Brand A has a higher fraction of unacceptable containers than Brand B? (Use α = .01)

The word "fraction" implies that we are comparing proportions. We apply z-test for 2 population proportions. Since the amount of difference is not specified, we apply Case 1;

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

p_1 = population proportion of unacceptable Brand A containers

p_2 = population proportion of unacceptable Brand B containers

H_0: $(p_1 - p_2) = 0$

H_1: $(p_1 - p_2) > 0$ (*higher*)

	Brand A	*Brand B*
Unacceptable	*16*	*5*
Sample Size	*33*	*33*

At $\alpha = 0.01$, the rejection region is $z > z_\alpha = z_{0.01} = 2.33$

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{16}{33} = 0.4848 \qquad \hat{p}_2 = \frac{x_2}{n_2} = \frac{5}{33} = 0.1515$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{16+5}{33+33} = \frac{21}{66} = 0.3182$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.4848 - 0.1515}{\sqrt{0.3182 * 0.6818\left(\frac{1}{33} + \frac{1}{33}\right)}} = 2.91$$

Since computed z = 2.91 > 2.33 (critical value), we reject the null hypothesis. There is enough evidence to infer that Brand A has a higher fraction of unacceptable containers than Brand B at α = .01.

d. What is the p-value of the test in part (c)?

p-value = P(z > 2.91) = 1 – 0.9982 = 0.00081 *e. The owner of Brand B claims that no more than 10% of his containers are unacceptable. Is there sufficient evidence at the 5% significance level to refute this claim?*

p_B = population proportion of unacceptable Brand B containers

H_0: $p_B \le 0.1$

H_1: $p_B > 0.1$

	Brand B
Unacceptable	*5*
Sample Size	*33*

At $\alpha = 0.05$, the rejection region is $z > z_{\alpha} = z_{0.05} = 1.645$

$$\hat{p}_B = \frac{x_B}{n_B} = \frac{5}{33} = 0.1515$$

$$z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \frac{0.1515 - 0.1}{\sqrt{0.1(0.9)/33}} = 0.99$$

Since z = 0.99 is less than 1.645, we do not reject H_0. There is not enough evidence at the 5% significance level to refute the owner's claim.

f. Estimate with 90% confidence the fraction of unacceptable Brand A containers.

The sample proportion $\hat{p}_A = \frac{x_A}{n_A} = \frac{16}{33} = 0.4848$

For 90% confidence, $\alpha = 0.1$, $\alpha/2 = 0.05$, and $z_{\alpha/2} = z_{0.05} = 1.645$.

The 90% confidence interval estimate is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad = \quad 0.4848 \pm 1.645\sqrt{\frac{0.4848(0.5152)}{33}} \quad = \quad 0.4848 \pm 0.1431$$

LCL = 0.3417 UCL = 0.6279

The 90% confidence interval estimate of the proportion of unacceptable Brand A containers is between 34.17% and 62.79%.

Question 3 (15 marks)

A major insurance firm interviewed a random sample of 1200 university students to find out the type of life insurance preferred, if any.

The results follow:

Insurance Preference

Gender	***Term***	***Whole Life***	***No Insurance***
Female	*100*	*80*	*325*
Male	*160*	*60*	*475*

Is there evidence that life insurance preference of male students is different than that of female students? Test using the 5% level of significance.

Solution:

Row and Column Totals

Insurance Preference

Gender	**Term**	**Whole Life**	**No Insurance**	***Total***
Female	100	80	325	*505*
Male	160	60	475	*695*
Total	*260*	*140*	*800*	*1200*

Expected Values (row total x column total/sample size)

Gender	**Term**	**Whole Life**	**No Insurance**	***Total***
Female	$\frac{505\times260}{1200} = 109.4167$	$\frac{505\times140}{1200} = 58.9167$	$\frac{505\times800}{1200} = 336.6667$	505
Male	$\frac{695\times260}{1200} = 150.5833$	$\frac{695\times140}{1200} = 81.0833$	$\frac{695\times800}{1200} = 463.3333$	695
Total	260	140	800	1200

H_0: insurance preference of male students is NOT different than that of female students

H_1: insurance preference of male students is different than that of female students

OR

H_0: insurance preference is independent of gender

H_1: insurance preference is dependent on gender

Df = (r – 1)(c – 1) = (2 – 1)(3 – 1) = 1(2) = 2

$\alpha = 0.05$

Rejection region: $\chi^2 > \chi^2_{\alpha,(r-1)(c-1)} = \chi^2_{.05,2} = 5.99$

Cell	f_i	e_i	$(f_i - e_i)$	$(fi - ei)^2/ei$
1	100	109.4167	-9.41667	0.8104
2	80	58.9167	21.08333	7.5447
3	325	336.6667	-11.6667	0.4043
4	160	150.5833	9.416667	0.5889
5	60	81.0833	-21.0833	5.4821
6	475	463.3333	11.66667	0.2938
Total	1200	1200		χ^2 = 15.1241

Since χ^2 = 15.1241 > 5.99, we reject the null hypothesis.

There is enough evidence to infer that insurance preference of male students is different than that of female students.

Question 4 (13 marks)

Calculate the probability of a Type II error for the following test of hypothesis
$H_o : \mu = 50$ vs. $H_1 : \mu = 50$, given that $\mu = 55$, $\alpha = 0.05$, $\sigma = 10$, and $n = 16$.

Solution:

H_0: μ = 50
H_1: μ > 50

n = 16 and σ = 10

Our objective is to find the probability of a Type II error, β, given μ is in fact 55. That is, the probability of failing to reject H_0: μ = 50 given it is false.

The appropriate test statistic is $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$ since σ is known

For α = 0.05, the rejection region is $z > z_\alpha = z_{0.05} = 1.645$ (from z-tables)

We now find the critical value for the sample mean, $\bar{x}$, that corresponds to $z_{0.05}$.

That is we solve for $\bar{x}$ if $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} > 1.645$

$$\frac{\bar{x} - 50}{\frac{10}{\sqrt{6}}} > 1.645$$

$$\bar{x} - 50 > 1.645\left(\frac{10}{\sqrt{16}}\right) = 4.1125$$

$$\bar{x} > 50 + 4.1125$$

$$\bar{x} > 54.1125$$

Suppose μ = 55 and we fail to reject H_0: μ = 2.25 ($\bar{x} < 54.1125$) then the probability of a Type II error is

$$\beta = P(\bar{x} < 54.1125, \text{ given that } \mu = 55)$$

$$\beta = P\left[z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{54.1125 - 55}{\frac{10}{\sqrt{16}}}\right]$$

$$= P(z < -0.36) = 0.3594$$

β = 0.3594

Question 5 (20 marks)

A financier whose specialty is investing in movie productions has observed that, in general, movies with "big-name" stars seem to generate more revenue than those movies whose stars are less well known. To examine his belief he records the gross revenue and the payment (in $ millions) given to the two highest-paid performers in the movie for ten recently released movies.

Movie	Cost of 2 Highest Paid Performer ($millions)	Gross Revenue ($millions)
1	5.3	48
2	7.2	65
3	1.3	18
4	1.8	20
5	3.5	31
6	2.6	26
7	8	73
8	2.4	23
9	4.5	39
10	6.7	58

a. *Find the sample regression line*

The independent variable is Cost because Revenue depends on it (cost).

Movie	x	y	x^2	y^2	xy
1	5.3	48	28.09	2304	254.4
2	7.2	65	51.84	4225	468
3	1.3	18	1.69	324	23.4
4	1.8	20	3.24	400	36
5	3.5	31	12.25	961	108.5
6	2.6	26	6.76	676	67.6
7	8	73	64	5329	584
8	2.4	23	5.76	529	55.2
9	4.5	39	20.25	1521	175.5
10	6.7	58	44.89	3364	388.6
Total	**43.3**	**401**	**238.77**	**19633**	**2161.2**

$\sum x = 43.3 \quad \sum x^2 = 238.77 \quad \sum y = 401 \quad \sum y^2 = 19633 \quad \sum xy = 2161.2$

$$\bar{x} = \frac{\sum x}{n} = \frac{43.3}{10} = 4.33 \qquad \bar{y} = \frac{\sum y}{n} = \frac{401}{10} = 40.1$$

$$s_x^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{\left(\sum x\right)^2}{n}\right] = \frac{1}{9}\left[238.77 - \frac{(43.3)^2}{10}\right] = 5.6979$$

$$s_y^2 = \frac{1}{n-1}\left[\sum y^2 - \frac{\left(\sum y\right)^2}{n}\right] = \frac{1}{9}\left[19633\text{-}\frac{(401)^2}{10}\right] = 394.7667$$

$$s_{xy} = \frac{1}{n-1}\left[\sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}\right] = \frac{1}{9}\left[2161.2 - \frac{(43.3)(401)}{10}\right] = 47.2078$$

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{47.2078}{5.6979} = 8.2851$$

The sample regression line is defined by $\hat{y} = 4.2255 + 8.2851x$

b. *Interpret the coefficients*

Intercept b_0 = 4.2255: When the Cost of 2 highest paid performers is $0, Gross Revenue is expected to be $4.2255 million, on average.

Slope b_1 = 8.2851; for every additional million dollar increase in the cost of 2 highest paid performers, Gross Revenue increases by $8.2851 million, on average.

c. *Determine the standard error of estimate and describe what this statistic tells you about the regression line.*

$$SSE = (n-1)\left[s_y^2 - \frac{s_{xy}^{\ 2}}{s_x^2}\right] = 9\left(394.7667 - \frac{47.2078^2}{5.6979}\right) = 32.7986$$

$$s_\varepsilon = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{32.7986}{8}} = 2.0248$$

Since Sε (2.0248) is small compared to $\bar{y} = 40.1$ the model fits the data well.

d. *Determine the coefficient of determination and discuss what its value tells you about the two variables.*

Coefficient of Determination $R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = \frac{47.2078^2}{5.6979*394.7667} = 0.9908$

99.08% of the variation in gross revenue is explained by the variation in the cost of 2 highest paid performers while 0.92% remains unexplained.

e. *Is there a linear relationship between payment to the two highest-paid performers and gross revenue?*

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}} = \frac{2.0248}{\sqrt{9(5.6979)}} = 0.2828$$

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

$\alpha = 0.05$

Rejection Region: $|t| > t_{\alpha,n-2} = t_{0.025,8} = 2.306$

$$t = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{8.2851 - 0}{0.2828} = 29.2967$$

Since t =29.2967 > 2.306, we reject H_0. There is enough evidence to infer that a linear relationship between payment to the two highest-paid performers and gross revenue.

f. *Predict with 95% confidence the Gross Revenue if the cost of two highest paid performers is $5 million.*

$\hat{y} = 4.2255 + 8.2851x$

Point estimate: ŷ = 4.2255 + 8.8251*5 = 48.351

Df = n – 2 =10 - 2 =8; $t_{\alpha/2} = t_{0.025} = 2.306$

Prediction Interval: $\hat{y} \pm t_{\alpha/2} s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}} = 48.351 \pm 2.306(2.0248)\sqrt{1 + \frac{1}{10} + \frac{(5-4.33)^2}{9(5.6979)}}$

$$48.351 \pm 2.306(2.0248)\sqrt{1 + \frac{1}{10} + \frac{(5-4.33)^2}{9(5.6979)}}$$

48.351 ± 2.306(2.0248)(1.053)

48.351 ± 4.917

(43.434, 53.268)