Review for Test 3 STA-3145

I. Operations managers often use work sampling to estimate how much time workers spend on each operation. Work sampling, which involves observing workers at random points in time, was applied to the staff of the catalog sales department of a clothing manufacturer. The department applied regression to the following data collected for 40 consecutive working days:

TIME: y = Time spent (in hours) taking telephone orders during the day

ORDERS: x = Number of telephone orders received during the day

Initially, the simple linear model $E(y) = \beta_0 + \beta_1 x$ was fit to the data.

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	STUDENT'S T	P
CONSTANT	10.1639	1.77844	5.72	0.0000
ORDERS	0.05836	0.00586	9.96	0.0000

```
R-SQUARED             0.7229   RESID. MEAN SQUARE (MSE)   11.6175
ADJUSTED R-SQUARED    0.7156   STANDARD DEVIATION         3.40844
```

1. Conduct a test of hypothesis to determine if time spent (in hours) taking telephone orders during the day (Y) and the number of telephone orders received during the day (X) are positively linearly related.

Ho: $\beta_1 = 0$ **T.S.: t = 9.96, P-val. = 0.000. We have sufficient evidence that X and Y are positively linearly related.**
Ha: $\beta_1 > 0$

2. Give a practical interpretation of the correlation coefficient for the above output.

R = 0.88 Strong positive linear relationship between X and Y.

3. Give a practical interpretation of the coefficient of determination, R^2.

About 72% in sample variability in Y could be explained by linear relationship between X and Y.

4. Give a practical interpretation of the estimated slope of the least squares line.

For every additional telephone order the estimated time will increase by β_1 (0.06) hours.

5. Find a **90% confidence interval** for **β_1**. Give a practical interpretation.

$\hat{\beta}_1 \pm t_{\alpha/2} \bullet S_{\beta 1}$ **$0.05836 \pm 1.69 \bullet 0.00586$** **(0.048, 0.068)**

We are 90% confident that average increase in time for every additional telephone order will fall within (0.048, 0.068) hrs.

6. Give a practical interpretation of the model standard deviation, ***s***.

About 95% of all observed time values will fall within 2 S (6.8 hrs.) from corresponding predicted values.

7. Give a practical interpretation of the estimate of the ***y*-intercept** of the least squares line. **No practical interpretation.**

8. Based on the value of the test statistic given in the problem, make the proper conclusion.

T.S.: t = 9.96, P-val. = 0.000. Model is useful for estimations and predictions.

II. Car & Driver conducts road tests on all new car models. One variable measured is the time it takes a car to accelerate from 0 to 70 miles per hour. To model acceleration time, a regression analysis is conducted on the data collected for a random sample of 125 new cars:

TIME 60: y = Elapsed time (in seconds) from 0 mph to 60 mph
MAX: x_1 = Maximum speed attained (miles per hour)

The simple linear model $E(y) = \beta_0 + \beta_1 x_1$ was fit to the data.

PREDICTOR VARIABLES	COEFFICIENT	STD ERROR	T	P	
CONSTANT	16.71	0.63708	29.38	0.0000	S = 1.20
MAX	- 0.12	0.00491	-17.05	0.000	**R^2 = 0.75**

1. Give the correlation coefficient for the above output. **r = __________.**

2. Describe the nature of the relationship (if any) that exist between maximum speed and acceleration time.

3. Approximately what percentage of the sample variation in time can be explained by the linear model?

4. Complete the sentence: "About 95% of the sampled cars have acceleration times that fall within _______ seconds of their ____________________ values."

5. Find a 95% confidence interval for the true slope of the regression line.

6. Choose correct practical interpretation of this interval and fill in the blanks.

Each answer begins with "We are 95% confident that ..."

a. acceleration time will fall between ____ and ____ second.

b. acceleration time will decrease between ____ and ____ second.

c. for every 1 sec. increase in acceleration time, max. speed will decrease between ____ and ____ mile per hour.

d. for a new car with a max. speed of 1 mile per hour, acceleration time will fall between ____ and ____ second.

e. for every 1 mile per hour increase in max. speed, acceleration time will decrease between ____ and ___ second.

III. Cocoon Problem

Researchers investigated the relationship between the mean daily air temperature and the cocoon temperature of wooly-bear caterpillars of the High arctic.

The regression equation is

Cocoon = 3.37 + 1.20 Air

Predictor	Coef	Stdev	t	p
Constant	3.3747	0.4708	7.17	0.000
Air	1.20086	0.09375	12.81	0.000

s = 0.8558 R-sq = 94.3%

Obs.	Air	Cocoon	Fit	95% C.I.	95% P.I.
1	1.7	3.600	5.416	(4.646, 6.186)	(3.359, 7.473)
2	2.0	5.300	5.776	(5.049, 6.504)	(3.735, 7.818)
3	2.2	6.800	6.017	(5.316, 6.717)	(3.984, 8.049)
4	2.6	6.800	6.497	(5.844, 7.149)	(4.481, 8.513)
5	3.0	7.000	6.977	(6.366, 7.589)	(4.974, 8.980)
6	3.5	7.100	7.578	(7.004, 8.152)	(5.586, 9.570)
7	3.7	8.700	7.818	(7.254, 8.381)	(5.829, 9.807)
8	4.1	8.000	8.298	(7.746, 8.850)	(6.313, 10.284)
9	4.4	9.500	8.658	(8.107, 9.210)	(6.673, 10.644)
10	4.5	9.600	8.779	(8.226, 9.331)	(6.793, 10.764)
11	9.2	14.600	14.423	(13.256, 15.590)	(12.186, 16.659)
12	10.4	15.100	15.864	(14.470, 17.257)	(13.502, 18.226)

1) According to MINITAB, the least squares equation is ____________ .

2) When the air temperature was 4.4°C the cocoon temperature was______, and the estimated cocoon temperature is ___________.

3) Since t = ____________ with p-value ____________, there __enough evidence at the 5% level to indicate that the t° of the cocoon is linearly related to the air temperature, for air t° ______

4) The estimated slope of the regression line is ___________.

5) The correlation coefficient for this data is r = ______.; r^2 = ____________.

6) Hence, we can conclude that ___________% of the variability in the cocoon temperatures is explained by the estimated least squares line relating cocoon temperature to air temperature.

7) Suppose we put a single woolly-bear caterpillar cocoon in a controlled environment with the air temperature set at 7°C. Predict the cocoon temperature.

IV. The Director of a small college conducted an entrance test to 20 randomly selected students from the new freshman class in a study to determine whether a student's GPA (y) at the end of the freshmen year can be predicted from the entrance test score (x).

```
Regression Analysis: GPA versus Score
The regression equation is
GPA = - 1.77 + 0.055635 Score

Predictor       Coef   SE Coef       T      P
Constant     -1.7726    0.3037   -5.84  0.000
Score       0.055635  0.003912   14.22  0.000

S = 0.181766   R-Sq = 91.8%
```

I. Find the estimates of β_1 and give a practical interpretation in context of the problem.

II. Give the correlation coefficient for the above output. Describe the nature of the relationship (if any) that exist between GPA and entrance test score.

III. Approximately what percentage of the sample variation in the GPA can be explained by the linear model?

IV. We expect approximately 95% of the observed GPAs to lie within _______points of their ________ values.

V. Find and interpret the 95% confidence interval for β_1.

VI. Predict the estimated average GPA score for all freshmen that have an entrance test score of 85.

V. John Breathe suspects that the amount of nitrogen fertilizer used per acre has a direct effect on the amount of wheat produced. The amount (in pounds) of nitrogen fertilizer (X) ranging from 30 to 100 pounds used per test plot and the amount (in pounds) of wheat (Y) harvested per test plot have been collected and used to fit the model. The SPSS outputs of the simple linear regression are provided below.

Model	R Square	Adjusted R Square	Std. Error of the Estimate
1	.801	.792	4.65486

a. Predictors: (Constant), FERTILIZER

b. Dependent Variable: YIELD

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	98.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-2.298	2.858		-.804	.430	-9.467	4.871
	FERTILIZER	.390	.041	.895	9.416	.000	.286	.494

A. Which of the following is the least squares line relating the amount of fertilizer and yield of wheat?

a) $\hat{y} = -2.298 + 0.390x + 0.895$ b) $\hat{y} = 0.390 - 2.298x + 0.895$

c) $\hat{y} = 0.895 - 2.298x$ d) $\hat{y} = -2.298 + 0.390x$ e) $\hat{y} = 0.390 - 2.298x$

B. Give the **correlation coefficient** from the SPSS output, **r** = _________.

C. Interpret the estimated slope of the regression line.

a) For each additional pound in nitrogen fertilizer, the mean yield of wheat is estimated to decrease by 2.298 pounds.

b) For each additional pound in nitrogen fertilizer, the mean yield of wheat is estimated to increase by 0.390 pounds.

c) The mean yield of wheat for a 1 pound of nitrogen fertilizer, is estimated to be 0.895 pounds.

d) For each 1 pound in yield of wheat, the mean amount of nitrogen fertilizer is estimated to increase by 0.390 pounds.

e) The mean yield of wheat for a 1 pound of nitrogen fertilizer, is estimated to be 2.298 pounds.

D. Describe the **nature of the relationship** that exists between the amount of nitrogen fertilizer (x) and yield of wheat (y).

E. Approximately what percentage of the **sample variation in the yield** of wheat can be explained by the linear model?

F. Complete the following sentence: "About 95% of the sampled amount of nitrogen fertilizer have yield of wheat that fall within ________ percentage points of their ___________ values.

G. Suppose the 50 pounds of fertilizer were used per plot. Predict the yield of wheat.

H. From the output, find a 98% confidence interval for β_1 and choose correct interpretation of this CI. Each answer begins with "We are 95% confident that ...". For the correct choice fill in the blanks.

a) the yield of wheat will fall between ____ and ____ pounds.

b) the yield of wheat will increase between ____ and ____ pounds.

c) for every 1 pound increase in nitrogen fertilizer, yield of wheat will increase between ____ and ____ pounds.

d) for a 1 pound of nitrogen fertilizer, yield of wheat will fall between ____ and ____ pounds.

e) for every 1 pound increase in the yield of wheat, nitrogen fertilizer will increase between ___and ___ pounds.

I. Based on the value of the test statistic given in the problem, make the proper conclusion.

1. We are 95% confident that there is no relationship between the amount of fertilizer used per plot and the yield of wheat.

2. There is sufficient evidence (at $\alpha = .02$) that use of nitrogen fertilizer is a useful predictor of the yield of wheat.
3. There is enough evidence (at $\alpha = .02$) that the use of nitrogen fertilizer increases linearly as yield of wheat increases.

4. There is insufficient evidence (at $\alpha = .05$) to conclude that the use of nitrogen fertilizer increases linearly as yield of wheat increases.

5. There is enough evidence (at $\alpha = .02$) that the use of nitrogen fertilizer increases linearly as yield of wheat decreases.

Chapter 13 Review

I. Last year, Dr. Johnson, an entomologist graduated from FIU, was given a brand new indoor garden for his research. He decided to include in the indoor garden the following percentage of insects: 30% were butterflies, 10% were ladybugs, 20% were fireflies, 25% were moths, and 15% of other types. A year has passed and he wants to conduct a test at 1% level of significance to see whether the proportions of insects differ significantly from the proportions that the entomologist started last year.

He observed 200 insects classified in the following table:

INSECTS IN THE INDOOR GARDEN

	Butterflies	Ladybugs	Fireflies	Moths	Other
Observed Counts	57	22	36	53	32
Expected Counts					

II. The study designed to test effectiveness of two types of frontier medicine – music, imaginary, and touch &MIT) therapy and therapeutic prayer – in healing cardiac care patients.

Tabulated statistics: THERAPY, EVENT (e.g., a heart attack)

```
Rows: THERAPY   Columns: EVENT

                  No       Yes     All

MIT               138       47     185
                140.7     44.3

Prayer            139       43     182
                138.4     43.6

Prayer & MIT      150       39      189
                143.8     45.2

Standard          142       50     192
                146.1     45.9


All               569      179     748
```

χ^2 = **0.05291 + 0.16817 + 0.00221 + 0.00703 + 0.26984 + 0.85777 + 0.11250 + 0.35760 + 0.11250 + 0.35760 = 1.828**

DF = 3, **P-Value = 0.609**

III. Suppose an educational TV station has broadcast a series of programs on the physiological and psychological effects of smoking marijuana. Before the series was shown, it was determined that 7% of the citizens favored legalization, 18% favored decriminalization, 65% favored the existing law, and 10% had no opinion. Test at $\alpha = .05$ level to see whether these data indicate that the distribution of opinions differs significantly from the proportions that existed before the educational series was aired.

H_o: ______________ **H_a**: __________________

D.F. = _____ **RR:** ________________

Decision: ____

$E(n_1)$ = ________, $E(n_2)$ = ________, $E(n_3)$ = ________, $E(n_4)$ = ________.

prob	observed	expected	O - E	O-E sq	terms
0.07	39		4	16	0.4571
0.18	99		9	81	0.9000
0.65	336		11	121	0.3723
0.10	26		-24	576	11.5200

chi-sq = __________ **p-value** = 0.00412732

IV. The researchers investigated the relationship between the gender of a viewer and the viewer's brand awareness. 300 TV viewers were asked to identify products advertised by male celebrity spokespersons.

H_o: ________________________ **H_a:** ___________________________

D.F. = _______ $\alpha = .01$ **RR:** ______________

χ^2 = ___________________________, **$E(n_{21})$ =** ________, **$E(n_{12})$ =** ________.

Decision: ____

	Male	**female**	**Total**
1	95	41	136
2	55	109	164
Total	150	150	300

$\chi^2 = 10.721 + 10.721 + 8.890 + 8.890 = 39.222$

CH. 9 Review

I. Magnets are often used by people to treat a variety of disorders. Researchers recently treated a group of patients with magnets and another group of patients with a fake magnet treatment. The results are given below. Test the claim that the magnet treatment is more effective at lowering pain in arthritis patients. Use a 5% level of significance.

	Placebo Group	Treatment Group
Sample size	25	25
Mean Pain Reduction	0.20	0.50
Standard Deviation	0.4	0.6

1. State the null and alternate hypotheses. **$H_0: \mu_1 - \mu_2 = 0$** **$H_a: \mu_1 - \mu_2 < 0$**

2. What is the **name of the distribution** to use to calculate the Test statistic? **Student (t) distribution**

3. What is the **critical value** for the test? **d.f. = 25 +25 -2 = 48 $t_{.05,\,48} = 1.677$**

4. Under what **assumptions**, the distribution from part 2 is used?

- two population variances are equal
- samples are randomly and independently selected
- populations are normally distributed

5. Compute the value of the **pooled variance estimator**.

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(25-1)0.4^2 + (25-1)0.6^2}{25+25-2} = 0.26$$

6. The **Test Statistic** for this test was reported as equal to **-2.08**, calculate the interval that contains P – value, give a conclusion and explain it in **simple non-technical terms in contest of the problem**.

1% $\leq$ P-value $\leq$ 2.5% **Decision: Reject H_0**

Conclusion: At 5% level of significance we have sufficient evidence to support claim that the magnet treatment is more effective at lowering pain in arthritis patients than placebo.

7. Suppose, a **95% CI** for the difference in Mean Pain Reduction between Treatment Group and Placebo Group patients turns out to be (-0.30 ± 0.29). Give a **practical interpretation** of this CI in terms of the problem.

CI : (-0.59, -0.01).

We are 95% confident that the magnet treatment is more effective at lowering pain in arthritis patients than placebo.

II. A new insect spray, type A, is to be compared with a spray, type B, that is currently in use. Two rooms of equal size are sprayed with sprays A and B. Two hundred insects are released into each room, and after 1 hour the numbers of dead insects are counted." There are 120 dead insects in the room sprayed with A and 90 in the room sprayed with B. Do the data provide enough evidence to indicate that spray A is more effective than spray B? Use α = .05.

1. State the null and alternate hypotheses. **H_o: $p_A - p_B \le 0$** **H_a: $p_A - p_B > 0$**

2. Compute the value of the **best estimate of true proportion** of dead insects for both sprays.

$$\hat{p} = \frac{120+90}{200+200} = 0.525$$

3. **Test statistic**: (Provide formula and substitution only, do not calculate!)

$$Z = \frac{0.60-0.45}{\sqrt{0.525\times0.475\left(\frac{1}{200}+\frac{1}{200}\right)}} = 3.038$$

4. The **P-value** for this test was **.0012**, give a conclusion and explain it in **context of the problem**.

The p-value = P(z > 3.04) = .5 - .4988 = .0012. **P-value < 0.05**

Decision: Reject the null hypothesis.

Conclusion: At level of significance α = .05 there is enough evidence to indicate that Spray A is more effective than Spray B.

5. Suppose, a **90%** confidence interval for the difference between population proportions of dead insects for spray A and B turns out to be (**0.15 ± 0.08**). **Interpret this CI in the context of this problem**.

7% < P_A - P_B < 23%

We are 90% confident that Spray A kills between 7% and 23% more insects than Spray B.

III. A manufacturer of shock absorbers claims that their shock absorbers last longer than those produced by its biggest competitor. To see if there is support for such a claim, six of the manufacturer's shocks and six of the competitor's shocks were randomly selected, and one of each brand was installed on the rear wheels of each of six cars. After the cars had been driven 20,000 miles, the strength of each shock absorber was measured.

Paired T-Test and CI: Manufacturer, Competitor

Paired T for Manufacturer - Competitor

	N	Mean	StDev	SE Mean
Manufacturer	6	10.717	1.752	0.715
Competitor	6	10.300	1.818	0.742
Difference	6	0.4167	0.1329	0.0543

95% lower bound for mean difference: 0.3073
T-Test of mean difference = 0 (vs > 0): **T-Value = 7.68 P-Value = 0.000**

1. State the hypotheses using the **appropriate symbols**. **H_0: $\mu_d = 0$** **H_a: $\mu_d > 0$**

2. What is the **critical value** for the test? Use **$\alpha = 0.01$**.

$t_{.01,5} = 3.365$

3. What is the **value** of the **point estimator** for the mean difference (Man - Com)? Use **appropriate symbols**

$\bar{x}_d = 0.4167$

4. Using output, **estimate the mean difference** of mean shock absorbers strength for two brands using 95% level of confidence. **Interpret your result in the context of this problem.**

CI: (0.3073, 0.5261)

We are 95% confident that the manufacturer's shock absorbers have a greater mean strength after being on cars for 20,000 miles than the competitor's shock absorbers.

5. What conditions are required for valid inferences from this experiment?

a. We must assume that the sample of differences was chosen randomly from the target population.
b. We must assume that the population of differences has a normal distribution.

IV. It is desired to compare the average test scores at the two schools. Suppose that simple random samples of college freshman are selected from two universities - 15 students from school A and 20 students from school B. On a standardized test, the sample from school A has an average score of 1000 with a standard deviation of 100. The sample from school B has an average score of 950 with a standard deviation of 90. Is there the difference in test scores at the two schools, assuming that test scores came from normal distributions?

Two-Sample T-Test and CI

Difference = $\mu_1 - \mu_2$
Estimate for difference: 50.0
95% CI for difference: (-15.6, 115.6)
T-Test of difference = 0 (vs not =): T-Value = 1.55 P-Value = 0.130 DF = 33
Both use Pooled StDev = 94.3719

a) State the hypotheses using the **appropriate symbols**. **H_0 : $\mu_1 - \mu_2 = 0$** **H_a :** $\mu_1 - \mu_2 \neq 0$

b) Using output, **estimate the difference** in the test scores at the two schools using 95% confidence. **Interpret your result in the context of this problem.**

CI: (-15.6, 115.6) We are 95% confident that the true difference in average test scores at the two schools is within: -15.6 to 115.6. Although interval suggest that school A has a higher test score, than school B, since interval contains 0 the result is statistically insignificant.

c) Using 5% level of significance what would be your inference from the above output?

Based on P-value = 0.13 we have insufficient evidence to support claim about the difference in average scores for two schools.

V. The newspaper recently ran an article indicating differences in perception of sexual harassment on the job between men and women. The article claimed that women perceived the problem to be much more prevalent than did the men. One question asked of both men and women was "Do you think sexual harassment is a major problem in the American workplace?" 24% of the men and 62% of the women responded "Yes." The newspaper created a 99% confidence interval for the true difference in proportions and reported it to be -.28 to -.48.

a) What is the **value** of the **point estimator** for the parameter of interest? Use **appropriate symbols.**

$\hat{P}_M - \hat{P}_W = .24 - .62 = -.38$

b) What does this CI suggest about perception of sexual harassment on the job between men and women?

We are 99% confident that percentage of women who think sexual harassment is a major problem in the American workplace between 28 and 48 percent higher than percentage of men.