Review for Test 4 STA-3123

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

SOURCE	DF	SS	MS	F	P
REGRESSION	1	1151.55	1151.55	99.12	0.0000
RESIDUAL	38	441.464	11.6175		
TOTAL	39	1593.01			

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

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

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

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

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

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

7. Interpret the 95% Prediction interval (17.753, 31.755) shown on the printout.

8. Interpret the 95% Confidence interval (23.568, 25.940) shown on the printout.

PREDICTED/FITTED VALUES OF TIME

LOWER PREDICTED BOUND	17.753	LOWER FITTED BOUND	23.568
PREDICTED VALUE	24.754	FITTED VALUE	24.754
UPPER PREDICTED BOUND	31.755	UPPER FITTED BOUND	25.940
SE (PREDICTED VALUE)	3.4584	SE (FITTED VALUE)	0.5857

Predictor values: orders = 250

Answers:

The mean total order time for all days with 250 telephone orders falls between 23.568 and 25.94 hours.

The total time for a day with 250 telephone orders falls between 17.7 and 31.7 hours

9. Give a practical interpretation of the estimate of the ***y*-intercept** of the least squares line.

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

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	
MAX	- 0.12	0.00491	-17.05	0.000	R^2 = 0.75

SOURCE	DF	SS	MS	F	P
----------	---	----------	----------	-----	------
REGRESSION	1	374.285	374.285	290.83	0.0000
RESIDUAL	123	183.443	1.425		
TOTAL	124	567.728			

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.

7. Choosing the correct answer to the next three questions use the output below.

PREDICTED/FITTED VALUES OF TIME 60 (95% level)

LOWER PREDICTED BOUND	**4.7493**	**LOWER FITTED BOUND**	**6.7776**
PREDICTED VALUE	**7.0057**	**FITTED VALUE**	**7.0057**
UPPER PREDICTED BOUND	**9.2621**	**UPPER FITTED BOUND**	**7.2338**
SE (PREDICTED VALUE)	**1.1403**	**SE (FITTED VALUE)**	**0.1153**

Predictor values: max speed = 140

A) Interpret the 95% confidence interval (6.78, 7.23) shown on the printout above.
Each answer begins with "We are 95% confident that ..."

a. the mean acceleration time for all new cars falls between 6.78 and 7.23 seconds.

b. the increase in acceleration time for every 1 mile per hr increase in maximum speed falls between 6.78 and 7.23 sec.

c. the mean acceleration time for all new cars with a max. speed of 140 miles per hour falls between 6.78 and 7.23 sec.

d. the acceleration time for a new car with a max. speed of 140 miles per hour falls between 6.78 and 7.23 sec.

B) Suppose we conduct a test for a car with a maximum speed of 140 miles per hour. Predict the acceleration time for this particular car with 95% confidence.

C) Give a theoretical interpretation of the phrase "95% confident" in the question above.

a. In repeated sampling, 95% of all similarly constructed intervals will equal (6.78, 7.23).

b. If we repeatedly sample from the population of new cars and compute a similar interval for each sample, 95% of all intervals constructed would capture the true mean acceleration time.

c. 95% of the acceleration times in the sample will fall within the interval (6.78, 7.23); 5% will fall outside the interval.

d. If we repeatedly sample from the population of new cars and compute a sample mean acceleration time for each, 95% of the sample means will fall within the interval constructed.

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%   R-Sq(adj) = 91.4%

Analysis of Variance

Source          DF      SS      MS       F     P
Regression       1  6.6813  6.6813  202.23 0.000
Residual Error  18  0.5947  0.0330
Total           19  7.2760
```

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.

VII. Find the correct interpretation of the 95% CI (3.1, 3.4) shown in the printout below: **"We are 95% confident that"**

a) the mean GPA score for all students fall between 3.1 and 3.4.
b) the increase in GPA for every 1 point increase in the entrance test score falls between 3.1 and 3.4points.
c) the average GPA score for all students with a score on the entrance test of 90 points falls between 3.1 and 3.4.
d) the GPA score for a new student with entrance test 90 points falls between 3.1 and 3.4.

IX. Locate from the printout the 95% prediction interval for the GPA score when the entrance score is 90, and interpret it.

```
Predicted Values for New Observations

Obs     Fit  SE Fit       95% CI     95% PI

  1  3.2345  0.0653  (3.1, 3.4)  (2.9, 3.6)

Values of Predictors for New Observations

Obs  Score
  1   90.0
```

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

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1921.143	1	1921.143	88.664	.000[a]
	Residual	476.690	22	21.668		
	Total	2397.833	23			

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. Dependent Variable: YIELD

FERTILIZER	YIELD	L_CI	U_CI	L_PI	U_PI
⋮	⋮	⋮	⋮	⋮	⋮
50.00	.	14.4	20.1	5.21	29.21

A. Which of the following is the **least squares line** relating the amount of nitrogen 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 practical **interpretation** of this interval. 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.

K. 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.

M. Find and interpret a 95% **PI** to predict yield of wheat when 50 pounds of fertilizer were used per plot.

a) When the amount of fertilizer is 50 pounds, we can be 95% confident that the predicted yield of wheat for all possible use of fertilizer will range from 5.2 pounds and 29.2 pounds.

b) When the yield of wheat is 50 pounds, we can be 95% confident that the predicted amount of fertilizer for all possible use of fertilizer will range from 5.2 pounds and 29.2 pounds.

c) We can be 95% confident that the true mean amount of fertilizer is between 14.4 pounds and 20.1 pounds.

d) We can be 95% confident that the true mean yield of wheat is between 5.2 pounds and 29.2 pounds.

e) When the amount of fertilizer is 50 pounds, we can be 95% confident that the predicted yield of wheat for all possible use of fertilizers will range from 14.4 pounds and 20.1 pounds.

VI. **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					

VII. A random sample of students of a certain university were classified according to the college in which they were enrolled and also according to whether they graduated from a high school in the state or out the state. The results are shown in the contingency table:

	Engineering	Arts and Sciences	Home Economics	Other	TOTALS
In State	16	14	13	13	56
Out of State	14	6	10	8	38
TOTALS	30	20	23	21	94

VIII. 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

IX. 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 the 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. = _____ **$\alpha = .05$** **RR:** _________________ **Test Statistic:** _____________,

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

X. 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: ____

Expected counts are printed below observed counts

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

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