MAT 222 (Spring 2017) HW Problems (for chapters 9, 10, and 11)

Chapter 9

From the Textbook:

9.25, 9.26, 9.27, 9.28, 9.31, 9.32, 9.33, 9.34, 9.40, 9.42, 9.55

Additional Problems:

1. (From MAT222 Spring 2006 Final Exam) A company's workforce was cross-classified according to job type and gender. That classification appears in the following contingency (two-way) table.

	Men	Women	Total
Management/Sales	39	47	**86**
Research/Development	28	22	**50**
Labor	188	116	**304**
Total	**255**	**185**	**440**

Using a Chi-Square test and the 0.05 level of significance, determine if there is a significant association between gender and the job type. Your test should include appropriate hypotheses, the computation of a test statistic, mention of the number of degrees of freedom involved, the p-value, a decision (reject H_0 or not) with justification, and a conclusion.

2. A research study about student retention is interested in assessing whether there is an association between the students' initial major and the program they transferred to. The data collected by the study is as follows:

	Program transferred to				
Initial Major	**Engineering**	**Management**	**Liberal Arts**	**Other**	**Total**
Biology	13	25	158	202	**398**
Chemistry	16	15	19	64	**114**
Mathematics	3	11	20	38	**72**
Physics	9	5	14	33	**61**
Total	**41**	**56**	**211**	**337**	**645**

a. Complete the following table of expected counts:

	Program transferred to			
Initial Major	**Engineering**	**Management**	**Liberal Arts**	**Other**
Biology	25.30		130.20	207.95
Chemistry		9.90	37.29	59.56
Mathematics	4.58	6.25		37.62
Physics	3.88	5.30	19.96	

b. Use Chi-Square test with the significance level 1% to check whether there is a significant relationship between the students' initial major and the program they transferred to. Provide the hypotheses, the test statistic, its degrees of freedom, the p-value, and the conclusion in context.

3. If both parents have hybrid genotypes Aa, a simplified genetic theory predicts that the child will have exactly one of the genotypes AA, Aa, and aa with respective probabilities 0.25, 0.50, and 0.25. In a random sample of 24 offsprings each having both parents with hybrid genotype Aa, it was found that 10 had genotype AA, 10 had genotype Aa, and 4 had genotype aa. By carrying out a Chi-Square goodness of fit test on this sample data, at level of significance $\alpha = 0.10$, investigate whether the simplified genetic theory is incorrect. Clearly indicate the hypotheses, the test statistic, its degrees of freedom, the p-value, and the conclusion in context.

4. A forestry researcher is interested in assessing whether the longleaf pine trees are distributed randomly across a certain tract. The tract was divided into four equal quadrants. A random sample of 100 longleaf pine trees was taken and the number of trees in each quadrant was counted. Below are the data:

Quadrant	Number of Trees
Q1	18
Q2	22
Q3	39
Q4	21

   a. If the longleaf pine trees are randomly distributed, we expect to find equal number of trees in each quadrant. Find the expected count for each quadrant.
   b. Perform a goodness of fit test to determine if the longleaf pine trees are randomly distributed. Clearly indicate the null hypothesis, the test statistic, its degrees of freedom, and the p-value.
   c. Based on the answer to the previous part, is the evidence significant at level 10%? At level 5%? At level 1%?

Chapter 10

From the Textbook (problems marked with * are recommended to be done with MINITAB):

10.8, 10.9, 10.16*, 10.17*, 10.18*, 10.19*, 10.21*, 10.22*, 10.23*, 10.30*, 10.39*, 10.40*, 10.44, 10.45, 10.46, 10.47, 10.51*, 10.52*

Additional Problems:

1. A regression model was run to predict the SAT score based on the ACT score. Partial minitab output is below:

```
The regression equation is

sat = ____________________________________________________

Predictor    Coef  SE Coef      T      P
Constant   253.19    62.67   4.04  0.000
act        31.206    2.895  _____  0.000

S = _______________   R-Sq = ___________%   R-Sq(adj) = 66.1%

Analysis of Variance

Source          DF       SS       MS       F      P
Regression      __  1276586  _______  _______  0.000
Residual Error  __  _______    10989
Total           59  1913973
```

a. Fill in the missing parts in the output above.
b. How many observations were included in this analysis?
c. Provide the details of a t-test to assess whether ACT score is a significant predictor of SAT score. State the hypotheses. Provide the t-statistic, its degrees of freedom, and the p-value. State the conclusion in context at α=5%.
d. Find the predicted SAT score for an ACT score of 20.
e. Find the 95% confidence interval for the mean response of ACT score of 20 (i.e., for the average SAT score among all students having ACT score $x^* = 20$).
   Note 1: For degrees of freedom = 58 and 95% confidence level, t^* is 2.002.
   Note 2: The following descriptive statistics provide the sample mean and the sample standard deviation of the predictor variable:

Descriptive Statistics: act

```
Variable     Mean   StDev
act        21.133   4.714
```

2. A regression model was run to predict the "Assessed Value" of a house based on its "Sales Price". Partial minitab output is below:

Regression Analysis: Assessed Value versus Sales Price

The regression equation is
Assessed Value = _____

Predictor	Coef	SE Coef	T	P
Constant	49.87	28.53	1.75	0.091
Sales Price	0.7312	0.1386	_____	0.000

S = _____ R-Sq = _____%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	__	29462	29462	27.85	0.000
Residual Error	__	_____	1058		
Total	29	59084			

a) Fill in the missing parts in the output above.

b) How many observations were used in this analysis?

c) What is the correlation coefficient "r" between the variables "Assessed Value" and "Sales Price"

d) Is the variable "Sales Price"a significant predictor of "Assessed Value"? Answer this question by providing the details of a t-test. State the hypotheses to test, provide the t-statistic, and provide the p-value. State the conclusion of whether or not "Sales Price" is a significant predictor of "Assessed Value" at α=5%.

e) Provide a 95% confidence interval for β_1. Clearly indicate the degrees of freedom and t*.

Chapter 11

From the Textbook (problems marked with * are recommended to be done with MINITAB):

11.7, 11.8, 11.9, 11.10, 11.11, 11.12, 11.17, 11.23*, 11.24*,
11.25*, 11.27*, 11.28*, 11.33*, 11.34*, 11.35*

Additional Problems:

1. A regression model was run to predict GPA based on HSM, HSE, and SATM. Partial minitab output is below:

Regression Analysis: GPA versus HSM, HSE, SATM

Predictor	Coef	SE Coef	T	P
Constant	-1.0488	0.6024	-1.74	0.084
HSM	0.13765	0.04801	____	0.005
HSE	0.13555	0.05290	____	0.011
SATM	0.0024241	0.0008276	____	0.004

S = _____ R-Sq = 25.2% R-Sq(adj) = _____%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	___	25.0856	8.3619	_____	0.000
Residual Error	___	74.5893	0.5109		
Total	149	99.6749			

a) Fill in the missing parts in the output above.

b) Using the model given above, find the predicted GPA for a student with HSM=10, HSE=10, and SATM=640.

c) State the null and the alternate hypotheses for the F-test in the Analysis of Variance table.

d) Is the variable "HSE"a significant predictor of "GPA"? Answer this question by providing the details of a t-test. State the hypotheses to test (in terms of one of β_1, β_2, or β_3), provide the t-statistic, and provide the p-value. State the conclusion of whether or not "HSE" is a significant predictor of "GPA" at α=5% and at α=1%.

e) Provide a 95% confidence interval for β_3, the slope of the predictor "SATM". Clearly indicate the degrees of freedom. Then circle the appropriate cell in the table below that can be used as t*. Then compute the confidence interval.

2. A study[1] about the birth weight analyzed related variables. Study was based on a sample of all births occurring in Philadelphia in 1990. The following regression model was run on two predictor variables ("YrsEduc" - Mother's years of education and "GestWks" -Gestational age in weeks).

Regression Analysis: BirthWt versus YrsEduc, GestWks

The regression equation is

BirthWt = _____

Predictor	Coef	SE Coef	T	P
Constant	-3.4806	0.2070	-16.81	0.000
YrsEduc	0.022696	0.006429	3.53	0.000
GestWks	0.165333	0.005044	32.78	0.000

S = _____ R-Sq = 49.9% R-Sq(adj) = 49.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	__	222.90	_____	_____	0.000
Residual Error	1112	_____	_____		
Total	____	447.10			

a) Find the missing parts in the output above.
b) How many observations were used in this study?
c) One of the data points had YrsEduc = 8, GestWks = 40, and BirthWt = 3.51. Find the predicted value and the residual corresponding to this observation.

[1] I.T. Elo, G. Rodríguez and H. Lee (2001). *Racial and Neighborhood Disparities in Birthweight in Philadelphia*. Paper presented at the Annual Meeting of the Population Association of America, Washington, DC 2001. Downloaded from: http://data.princeton.edu/wws509/datasets/