MAT 222 (Spring 2017) HW Problems (for chapters 6, 7, and 8)

Chapter 6

From the Textbook:

Section 6.1	6.12, 6.27, 6.28, 6.30, 6.36
Section 6.2	6.52, 6.53, 6.54, 6.58, 6.59, 6.70, 6.71, 6.73, 6.75
Section 6.4	6.118, 6.119

Additional Problems:

1. A machine is used to fill soda bottles. The amount of soda dispensed into each bottle varies slightly and is known to have a normal distribution with population standard deviation σ = 2.62 ml. A random sample of 25 bottles filled by the machine is taken and the amount of soda filled in each bottle was measured. From this sample data, the sample mean was calculated to be 591.2 ml.
    a. Find the 90% confidence interval for the population mean amount of soda filled by the machine.
    b. Find the 98% confidence interval for the population mean amount of soda filled by the machine.
2. A firm that administers flexible spending plans would like to estimate the population mean of the unused yearly dollars within $0.30 with 99% confidence. The standard deviation of the unused yearly dollars is known to be $2.4. What should be the number of flexible spending plan members that the firm should sample?
3. The level of calcium in the blood of healthy young adults follows a normal distribution with mean μ = 10 milligrams per deciliter and standard deviation σ = 0.5. A clinic measures the blood calcium of 27 healthy pregnant young women at their first visit for prenatal care. The mean of these 27 measurements is $\bar{x}$ = 9.8. Is this evidence that the mean calcium level in the population of healthy pregnant young women is less than 10?
    a. Set up the null and alternative hypotheses that you will be testing regarding the above question.
    b. Assuming that the distribution of calcium level measurements for pregnant young women is a normal distribution with σ = 0.5, find the value of the test statistic and compute the P-value.
    c. Is this statistically significant (circle your answers)
        at the 5% level? (Yes, No); at the 1% level? (Yes, No)
    d. At the 5% level of significance, state your conclusion in words.
4. A car brand XYZ states on their stickers and their advertisements that the highway MPG is 27. A consumer watchdog group believes that XYZ is overstating the highway MPG.
    a. If μ denotes the average highway MPG for all XYZ cars, state the null and alternate hypotheses for this scenario in terms of μ.
    b. From a simple random sample of 140 cars of the brand XYZ, sample average highway MPG was computed to be 25.9. Assuming the population standard deviation is σ = 7, compute the test statistic and the p-value.
    c. State the decisions at significance levels α=10%, α=5%, and α=1%.
5. A machine is used to fill soda bottles. The amount of soda dispensed into each bottle varies slightly and is known to have a normal distribution with population standard deviation σ = 2.62 ml. A random sample of 25 bottles filled by the machine is taken and the amount of soda filled in each bottle was measured. From this sample data, the sample mean was calculated to be 591.2 ml.
    a. Consider the hypotheses H_0 : μ = 592.4 and H_a : μ < 592.4. Compute the test statistic z and the p-value.
    b. State the decisions at significance levels α=10%, α=5%, and α=1%.
6. Consider the test H_0: μ = 450 against the alternative hypothesis H_a: μ > 450 at the α = 1% level of significance. Assume that the population standard deviation is σ = 100 and a simple random sample of size 500 is to be taken.
    a. Calculate the critical value. For what values of the sample mean will H_0 be rejected?
    b. Calculate P(Type II Error) and the power at the particular alternative value μ = 460.

Chapter 7

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

Section 7.1 7.17, 7.20, 7.21, 7.22, 7.23, 7.24, 7.30*, 7.31*, 7.32, 7.35
Section 7.2 7.63, 7.64, 7.67, 7.74, 7.80, 7.84, 7.85*
Section 7.3 7.102, 7.103, 7.107*, 7.109

Additional Problems:

1. A machine is used to make a particular type of HDMI cord. The length of each cord made varies slightly and is known to have a normal distribution. A random sample of 25 cords made by the machine is taken and the exact length of each cord in the sample was measured. From this sample data, the sample mean was calculated to be 591.2 mm with a sample standard deviation 2.62 mm.
    a. Find a 95% confidence interval for the population mean cord length made by the machine.
    b. Consider the hypotheses $H_0 : \mu = 592.4$ and $H_a : \mu < 592.4$. Compute the t-statistic and the p-value.
    c. State the decisions at significance levels α=10%, α=5%, and α=1%.
2. A faster loan processing time produces higher productivity and greater customer satisfaction. A financial services institution wants to determine if their mean loan processing time is less than a competitor's claim of 6 hours. A financial analyst randomly selects 7 loan applications and manually calculates the time between loan initiation and when the customer receives the institution's decision. From the sample data, the sample mean of the loan processing time was 5.079 hours with a sample standard deviation of 1.319 hours. Assume that the loan processing times follow a normal distribution.
    a. Compute the test statistic and the p-value that can be used to determine if their mean loan processing time is less than a competitor's claim of 6 hours.
    b. State the decisions at significance levels α=10%, α=5%, and α=1%.
    c. Find a 95% confidence interval for the population mean loan processing time.
3. An inspector wants to compare the bolts produced by company A and company B, based on the length of the bolts. A random sample of 16 bolts by company A yields a sample mean of 4.488 and a sample standard deviation of 0.035 inches. Random sample of size 10 from company B yields a sample mean of 4.512 inches and a sample standard deviation of 0.029 inches. Assume the population standard deviations are equal.
    a. Consider testing the research hypothesis that the population mean length of the bolts from company A is *smaller* than the population mean length of the bolts from company B. Calculate the appropriate test statistic, specify the degrees of freedom of the test statistic, and find the p-value.
    b. State the conclusions at significance levels 1%, 5%, and 10%.
    c. Consider instead testing the research hypothesis that the mean lengths of the bolts from company A and company B are *different*. Find the p-value. State the conclusions at significance levels 1%, 5%, and 10%.
    d. Provide a 95% confidence interval for the difference in the population mean lengths.
4. Repeat the previous problem assuming that the population standard deviations are not equal.
5. Using the data from problem 3, test whether or not the population standard deviations are equal.
6. Sample standard deviation from a random sample of size 7 was computed to be 4.876. From another independent random sample of size 13, sample standard deviation was computed to be 2.371.
    a. Consider the hypotheses $H_0: \sigma_1 = \sigma_2$ against $H_a: \sigma_1 \neq \sigma_2$. Find the appropriate test statistic for this hypothesis testing scenario, specify the numerator and denominator degrees of freedom of the test statistic, and find the p-value. Work must be shown by providing the bracketing critical values from the table and the corresponding upper tail probabilities.
    b. State the conclusions at significance levels 1%, 5%, and 10%.

Chapter 8

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

Section 8.1 8.15, 8.18, 8.19, 8.20, 8.37, 8.39
Section 8.2 8.60, 8.61, 8.62, 8.63, 8.64, 8.69, 8.71

Additional Problems:

1. A software developer is interested in analyzing the population proportion of doctors that use a certain medical software.
    a. How many observations (what sample size) should be taken to estimate, at 99% confidence level, the population proportion within ±5% if no prior estimate is available?
    Note: Here desired margin of error is at most ± 0.05 (i.e., ± 5% expressed in decimal form).
    b. A random sample of 190 doctors was collected. 85 doctors out of the sample reported using the specific software.
    c. Find a 90% confidence interval for the proportion of doctors that use that particular software.
2. Delivery department of a manufacturing company A&B uses a company CDE for its shipping needs. CDE advertises that they deliver shipments on schedule (on or before the scheduled delivery date) 85% of the time. If the on time delivery rate of CDE is less than 85%, A&B will have to consider other shippers or alternative options. A&B collects a random sample of size 625 from their shipments across a year and finds that 515 shipments were delivered on time. Let p denote the population proportion on time delivery rate of CDE.
    a. State the appropriate null and alternate hypotheses to test (in terms of p)
    b. Find the appropriate test statistic and the p-value.
    c. Is there significant evidence at level 1% to believe that the on time delivery of CDE is less than 85%?

        YES NO (circle the correct option)
    d. Is there significant evidence at level 5% to believe that the on time delivery of CDE is less than 85%?

        YES NO (circle the correct option)
    e. In order to explore their options, another random sample of 625 shipments across a year were sent via a different shipping carrier FGH. 530 of those shipments were delivered on time. Is the performance of FGH significantly better than that of CDE? Calculate the appropriate test statistic to answer this question and find the p-value. State whether the performance of FGH significantly better than that of CDE at significance level 5%.
3. Using the data from the previous question, answer the following:
    a. Find the 95% confidence interval for the CDE's on time delivery rate (i.e., the population proportion of time packages delivered by CDE before the scheduled delivery date).
    b. Find the 95% confidence interval for the FGH's on time delivery rate (i.e., the population proportion of time packages delivered by FGH before the scheduled delivery date).
    c. Find the 95% confidence interval for the difference between CDE's on time delivery rate and FGH's on time delivery rate.

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/

MAT 222 (Spring 2017) HW Problems (for chapters 12 and 13)

Chapter 12

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

12.9, 12.10, 12.13, 12.15, 12.25, 12.40, 12.41, 12.42, 12.46*, 12.47*, 12.54*, 12.55*, 12.56*, 12.57*

Additional Problems:

1. A One-Way ANOVA was done to analyze the quantitative variable "Score" against a categorical variable "Food" (representing three food groups: "Comfort", "Control", "Organic").

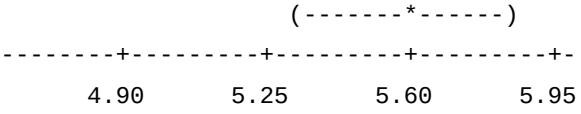
One-way ANOVA: Score versus Food

Source	DF	SS	MS	F	P
Food	__	5.330	_____	_____	0.001
Error	__	20.930	_____		
Total	61	26.260			

S = _____ R-Sq = 20.30% R-Sq(adj) = 17.59%

Level	N	Mean	StDev
Comfort	22	4.8873	0.5729
Control	20	5.0825	0.6217
Organic	20	5.5835	0.5936

Individual 95% CIs For Mean Based on Pooled StDev



Pooled StDev = _____

a. Fill in the missing parts in the output above.
b. State the null and the alternate hypotheses for the F-test in the Analysis of Variance table.
c. Using an appropriate contrast, we would like to compare the mean score of the control group with the average of the other groups. Compute the sample contrast, the standard error of the sample contrast, the t-statistic, and its degrees of freedom. You don't need to compute the p-value or provide the conclusion.

2. (From MAT222 Spring 2010 Final Exam[1] – see question 5. Ignore the points reference)

(20 pts) Battery life of MP3 players is of great concern to customers. A consumer group has tested three brands of such players to determine the battery life. Samples of players of each brand were fully charged and left to run at medium volume until the battery died. The number of hours that each player ran was recorded.

Consider the following summary statistics obtained:

Brand	n_i	$\bar{x}_i$	s_i
A	14	24.63	3.06
B	15	27.71	2.74
C	14	24.87	2.95

(a) If one wanted to calculate the SSG, the between-group sum of squares, the value of the overall mean, $\bar{x}$, needs to be determined. Using the above information, show that $\bar{x} = 25.78$ and calculate SSG.

(b) Choose the correct F value and its associate P-value to test $H_0 : \mu_A = \mu_B = \mu_C$, where $F(d_1, d_2)$ denotes the value of the test statistic with the numerator and denominator degrees of freedom, d_1 and d_2, respectively. You do not need to calculate the F value.

(i) $F(2, 43) = 5.06;\ P < 0.025$ (ii) $F(2, 40) = 5.06;\ P < 0.025$

(iii) $F(3, 43) = 5.06;\ P < 0.01$ (iv) $F(3, 40) = 5.06;\ P < 0.01$

(c) What conclusion might you reach based on your answer in (b)?

(i) There is no evidence to suggest that the null hypothesis should be rejected.

(ii) There is sufficient evidence to conclude that the sample means are not equal.

(iii) There is sufficient evidence to conclude that not all the population means are equal.

(iv) There is evidence to support the conclusion that all the population means are different from each other.

(d) Because one of the brands of MP3 players (Brand B) is known to be the most popular among customers, it was decided before the data were gathered to compare this brand against the other two to see if it had a longer battery life. What would be appropriate null and alternative hypotheses to establish to do such a test?

(i) $H_0 : \mu_A = \mu_B = \mu_C$ versus $H_a : \mu_B > \mu_A = \mu_C$

(ii) $H_0 : \frac{\mu_A + \mu_C}{2} = \mu_B$ versus $H_a : \frac{\mu_A + \mu_C}{2} < \mu_B$

(iii) $H_0 : \mu_A = \mu_B = \mu_C$ versus $H_a : \mu_B > \mu_A$ and $\mu_B > \mu_C$

(iv) H_0 : all of the means are equal,
H_a : all the other means are different from μ_B.

(e) Calculate the test statistic to test the contrast in (d) and specify its degrees of freedom. Draw your conclusion at 0.05 level of significance. Note that $s_p^2 = 8.50$.

[1] See http://researchguides.library.syr.edu/mathexams for old final exams.

3. (From MAT222 Spring 2011 Final Exam[2] – see question 4. Ignore the question number & points reference)

4. [20 points] In a study of effective weight loss programs, 24 subjects who were at least 20% overweight took part in a three month group support program. The subjects were divided evenly into one of three different programs. Private weightings determined each subjects weight at the beginning of the program and four months after the program's end. The table below summarized the mean weight loss of each group, and the group's standard deviation of weight loss.

	$\bar{x}$	s
Program 1	13.2	4.23
Program 2	7.9	8.02
Program 3	17.3	4.55

a) [5 points] Is the assumption of equal population standard deviations reasonable? Explain.

b) [10 points] Calculate the pooled sample standard deviation, s_p.

c) [5 points] Using an ANOVA analysis, the F statistic is 4.11. Give the degrees of freedom and an approximate value of the p-value. What are your conclusions at the 5% significance level?

4. Iron-deficiency is a cause of concern in many developing countries. Some research has suggested that food cooked in iron pots contain more iron than food cooked in other types of pots. A study was conducted to analyze the effect of the type of pot used in cooking on the iron content in the food prepared. Result of the One-way ANOVA on the dataset is given below:

```
Source  DF      SS      MS      F      P
pot     __  ______  ______  _____  0.000
Error   __  15.580  ______
Total   __  40.474

S = ______   R-Sq = 61.51%   R-Sq(adj) = 59.17%

                                 Individual 95% CIs For Mean Based on
                                 Pooled StDev
Level      N    Mean   StDev  ---------+---------+---------+---------+
Aluminum  12  1.8733  0.5223  (-----*-----)
Clay      12  2.0367  0.6013    (-----*-----)
Iron      12  3.7133  0.8843                          (-----*-----)
                                 ---------+---------+---------+---------+
                                        2.10      2.80      3.50      4.20
Pooled StDev = ______
```

a. Fill in the missing parts in the output above.
b. State the null and the alternate hypotheses for the F-test in the Analysis of Variance table.
c. Using the output above and an appropriate contrast, we would like to test whether the mean iron content of the iron pot group is higher than the average of the other groups. Compute the sample contrast, the standard error of the sample contrast, the t-statistic, its degrees of freedom, and the p-value. Provide the conclusion in context using significance level α=5%.

[2] See http://researchguides.library.syr.edu/mathexams for old final exams.

Chapter 13

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

13.4, 13.4, 13.5, 13.6, 13.7, 13.14, 13.15, 13.18, 13.22, 13.23, 13.24, 13.39*, 13.40*

Additional Problems:

1. The research study in the additional problem #4 from chapter 12 (see previous page) also considered the effect of the type of dish prepared, in conjunction with the potential effect of the type of pot used, via a Two-way ANOVA. Three pot types (aluminum, clay, iron) and three dish types (meat, legume, vegetables) were used as the factors with the iron content as the response.

Source	DF	SS	MS	F	P
pot	__	24.8940	12.4470	92.26	0.000
dish	__	_____	_____	_____	0.000
Interaction	__	2.6404	_____	_____	0.004
Error	__	3.6425	0.1349		
Total	35	40.4738			

S = _____ R-Sq = 91.00% R-Sq(adj) = 88.33%

   a. Fill in the missing parts in the output above.
   b. Consider assessing the effect of the dish type. Specify the F-statistic for this test. In addition specify the numerator and denominator degrees of freedom of that F-statistic. State the conclusion in context.
   c. Consider assessing the interaction effect of pot and dish types. Specify the F-statistic for this test. In addition specify the numerator and denominator degrees of freedom of that F-statistic. State the conclusion in context.

2. (From MAT222 Spring 2011 Final Exam[3] – see question 5. Ignore the question number & points reference)

5. [20 points] An agricultural researcher conducted a Two-Way ANOVA to assess the effect of seeds and fertilizers on growing corn. Equal number of samples from each treatment group was taken. Below is the ANOVA output.

Source	*df*	*SS*	*MS*	*F*	*P-value*
Seed	2	512.8667	256.4333	28.283	0.000008
Fertilizer	4	449.4667	112.3667	12.393	0.000119
Interaction	__	143.1333	17.8917	1.973	0.122090
Error	__	136.0000	_____		
Total	**29**	**1241.4667**			

Based on this output, answer the following questions.

a) [4 points] Specify how many types of seeds, how many types of fertilizers, how many treatment groups, and how many samples from each treatment group were used in this study.
b) [8 points] Consider the F-statistic = 12.393. State the null and alternate hypotheses tested by this statistic. Also, specify the numerator and denominator degrees of freedom for this statistic.
c) [8 points] What is the pooled sample standard deviation?

[3] See http://researchguides.library.syr.edu/mathexams for old final exams.

3. (From MAT222 Spring 2010 Final Exam[4] – see question 6. Ignore the points reference)

(15 pts) Twenty high school-aged students are randomly selected from three different school districts: a district in the city, a district in the suburbs, and a district in a rural area. Each group of twenty students consisted of 10 boys and 10 girls. Each of the students was asked what price they paid for their last haircut. The data were entered into statistical software and a partial ANOVA table is obtained below.

(a) Compete the ANOVA table.

Source	DF	Sum of Squares	Mean Square	F
Sex		2674.0	2674.0	67.37
Region			351.8	
Sex* Region		231.8		
Error	54	2143.0	39.69	***
Total	59	5752.4	***	***

(b) What is the value of the pooled standard deviation, s_p?

(c) Is the interaction effect statistically significant at the 5% significance level? Include the P-value (or the range of the P-value) and the degrees of freedom.

(d) Is the main effect for region statistically significant at the 5% significance level? Include the P-value (or the range of the P-value) and the degrees of freedom.

[4] See http://researchguides.library.syr.edu/mathexams for old final exams.