Ma 320 Test 3 with Answers – Section 2

S. Sawyer — Washington University — April 27, 2001

Calculators and a single $8^1\!/_2$ by 11 cribsheet can be used. All problems are 15 points each unless indicated otherwise.

1. A tire store wants to compare the performances of two types of automobile tires. They randomly select 21 tires of Brand A and 21 tires of Brand B from their records of tires with recorded lifetimes. The brand A tires lasted an average of 41000 miles with standard deviation of 3000 miles. The brand B tires lasted an average of 45000 miles with a standard deviation of 5000 miles.

Assuming that the tire lifetimes are normally distributed, first test to see if the variances of the two groups are equal ($\alpha = 0.10$). Based on this result, apply either a classical two-sample t-test or else the Satterthwaite t-test to determine if the means of the two techniques are equal, and find the P-value.

Answer: First, to test equality of variances: $F = (5000)^2/(3000)^2 = 2.778$. The upper critical value for a two-sided test with $\alpha = 0.10$ is $F(0.95, 20, 20) = 2.124 < 2.778$, so that we reject the hypothesis of equality of variances and assume that the variances are difference. (In fact, $P = 0.027$, two-sided.)

Applying the Satterthwaite t-test, the formula for the degrees of freedom gives df=32.748. The unpooled $T = (45000-41000)/\sqrt{5000^2/21 + 3000^2/21}$ $= 3.1436$. Since $t(0.975, 32) = 2.0369 < 3.1436$, we conclude $P < 0.05$ (two-sided) and we reject the hypothesis that the means are equal. In fact, $P = 0.00353$ (two-sided).

2. A poll was conducted to determine if perceptions of the hazards of smoking were dependent on whether or not the person smoked. One hundred (100) smokers and 100 nonsmokers were randomly selected and surveyed. The results are given below. Test the hypothesis that smoking status does not affect perception of the dangers of smoking at $\alpha = 0.05$. Find the P-value.

Perceptions of Dangers of Smoking

	Dangerous	Not Dangerous
Smokers	80	20
Nonsmokers	91	9

Answer: The row sums are 100 and 100, the column sums are 171 and 29, and the total sum is 200. Thus the Pearson $T = 200(80 \times 9 - 91 \times 20)^2/(171 \times 29 \times 100 \times 100) = 4.8800$. Since $\chi^2(0.95, 1) = 3.841 < 4.8800$, we reject the hypothesis that the row proportions are equal and conclude that smoking status does affect perception of the dangers of smoking. Here $P = \Pr(\chi_1^2 \geq 4.8800) = 0.0272$.

3. A political consultant wants to examine the relationship between voters' perceptions of a particular political candidate and the distance in miles between the residences of the voter and the candidate. Twelve (12) randomly chosen voters were asked to rate the candidate on a scale of 1 (bad) to 20 (excellent). The results are presented below. Assume that both the ratings and the distances are normally distributed. Find the (Pearson) correlation coefficient r for these data and determine whether or not the voter ratings and the distances are correlated. (Use a two-sided test with $\alpha = 0.05$.)

Voter	Rating	Distance
1	15	273
2	9	184
3	11	188
4	14	168
5	18	351
6	17	259
7	12	116
8	4	47
9	19	367
10	12	94
11	7	160
12	18	246

Answer: The Pearson $r = 0.82834$, $T = r\sqrt{(n-2)/(1-r^2)} = 4.6756$ for $n = 12$ and $n - 2 = 10$, and $t(0.975, 10) = 2.2281 < 4.6756$. Thus we reject the hypothesis that the voter ratings and distances are uncorrelated and conclude that they are correlated. (In fact, $P = 2\Pr(t(10) \geq 4.6756) = 0.000873$.)

4. A scientist studies a particular species of mollusk. The heights (X) and diameters (Y) of ten mollusks were measured and are recorded below. Assume that the data are normally distributed. Find the least-squares regression of Y on X and use it to find a 95% confidence interval for the mean diameter of shells whose height is 70 mm. (***Warning:*** Be careful that you and your calculator agree on which numbers are Xs and which numbers are Ys!)

Mollusk Shell Measurements (millimeters)

Observ.	Diameter (Y)	Height (X)
1	185	78
2	194	65
3	173	77
4	200	76
5	179	72
6	213	76
7	134	75
8	191	77
9	177	69
10	199	65

Answer: The least-squares regression of Y on X is $Y = 231.1027 - 0.6384X$. The predicted value at $X = 70$ is $Y(70) = 231.1027 - 0.6384 \times 70 = 186.4147$. The standard deviation of the regression errors is $s = \widehat{\sigma} = 22.56$. Here $n = 10$ and $t(0.975, 8) = 2.306$, $\overline{X} = 73$, and $s_X = 4.99$. The formula for the 95% confidence interval for $E(Y \mid X = 70)$ yields (166.94, 205.89).

5. A physician wants to study the time to recovery in days for students from a newly discovered long-lasting disease. Observations for 15 male and 15 female students were collected along with the Vitamin D and Vitamin E concentrations in the blood when the patient was first examined. It is thought that these vitamins might impede the recovery process. Two regression models were studied:

(A) Recovery time = β_0 + β_1*Sex + β_2*VitD + β_3*VitE and

(B) Recovery time = β_0 + β_1*Sex + β_2*VitD + β_3*VitE + β_4*VitD*VitE

where VitD and VitE are vitamin concentrations in micrograms per liter.

Attached to the test is a printout of an Excel spreadsheet with the data and with output from Excel's Regression procedure for the two models. You can use this output to answer the following questions:

(i) For the output from model (A), does the model fit the observed recovery time data better than with no variables at all? What is the P-value of the resulting F-test? What are the two degrees of freedom for the F-distribution involved in the F-test?

(ii) Which of the variables in model (A) have coefficients that are significantly different from zero? What are their P-values? How many degrees of freedom are involved in the T-tests that these P-values are based on?

(iii) As predicted by the estimated regression function for model (A), which sex recovers sooner on the average from the disease, as determined by the regression coefficient for Sex? Is the difference significant?

(iv) Suppose that a female student with the disease is examined with a Vitamin D level of 17 and a Vitamin E level of 10. What recovery time would the regression model (A) predict? What recovery time would regression model (B) predict?

Answer: (i) From the regression output for model (A), the three covariates (Sex, VitD, and VitE) fit better than no fit at all: From the ANOVA table of the regression, $P = 8.21 \times 10^{-10} < 0.05$. Again from the ANOVA table, the degrees of freedom is 3 in the numerator and 26 in the denominator.

(ii) From the coefficient table in the regression output, Sex, VitD, and VitE all have coefficients that are significantly different from zero. The P-values are $P = 0.0116$ for Sex, $P = 1.57 \times 10^{-10}$ for VitD, and $P = 0.0102$ for VitE. The degrees of freedom of these t-tests is the same as that of the "residual" line in the ANOVA table, namely 26.

(iii) The regression coefficient for Sex is 55.477. Since Sex is coded as Female=0 and Male=1, this means that the fitted lines for males have an additional additive term of 55.477, which says that they recover (as predicted by the regression equation) 55.477 days later, so that women recover sooner (as measured by this term in the regression equation). The difference is statistically significant, since $P = 0.0116$ for the coefficient for Sex in the coefficient table.

(iv) For Sex=Female=0, VitD=17, and VitE=10, model (A) predicts a recovery time of 246.1045 days and model (B) predicts a recovery time of 286.0171 days.

6. (Essay – 5 points) Which of models (A) and (B) in Question 5 is better? Why?

7. (Extra Credit – 10 points) An explorer finds a six-sided die on a newly discovered planet and wants to test whether or not the die is fair. (That is, whether or not each side of the die is equally likely to be on top when the die is rolled.) The explorer tosses the die 60 times. The results are recorded below. Test the hypothesis that the die is fair ($\alpha = 0.05$) and find the P-value.

Face	1	2	3	4	5	6	
Count	5	11	14	15	9	6	Total: 60

Answer: If n_i are the counts in the table, the goodness-of-fit statistic is $T = \left(\sum_{i=1}^{6} n_i^2/10\right) - 60 = 8.40$. Since $\chi^2(0.95, 5) = 11.07$, this is not significant at $\alpha = 0.05$ and we accept the hypothesis that the face probabilities are equal. Here $P = \Pr(\chi^2(5) \geq 8.04) = 0.1355$.