Ma 320 Test 3 with Answers – Section 1

J. Herdlick — 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. Forty two (42) randomly selected students were random divided into two groups of 21 students each. The first group was taught calculus by a live instructor and the second group was shown movies covering the same material. On identical exams given to the two groups after the two courses, the first group had a mean of 80 with a standard deviation of 10. The second group had a mean of 84 with a standard deviation of 6.

Assuming that the grades 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 the 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: $F = (10)^2/6^2 = 2.778$. 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 $T = (84 - 80)/\sqrt{100/21 + 36/21} = 1.5718$. Since $t(0.975, 32) = 2.0369$ and $t(0.95, 32) = 1.6939$, we conclude that $P \geq 0.10$, two-sided, and that there is insufficient evidence to conclude that the means are difference, so we accept the hypothesis that the means are equal. In fact, $P = 0.1256$, two-sided.

2. The EPA wants to test the effect of a particular industrial chemical on a certain species of plant. One group of plants was sprayed with a solution containing 2% of the chemical. A second group was sprayed with a solution containing 15% of this chemical. The results after three days are given in the following table. Test the hypothesis that the two concentrations of the chemical have the same effect on the plant at $\alpha = 0.05$. Find the P-value.

Survival of Test Plants

	No. Alive	No. Dead
2% solution	19	6
15% solution	12	13

Answer: The row sums are 25 and 25, the column sums are 31 and 19, and the total sum is 50. Thus the Pearson $T = 50(19 \times 13 - 12 \times 6)^2/(31 \times 19 \times 25 \times 25) = 4.1596$. Since $\chi^2(0.95, 1) = 3.841 < 4.1596$, we reject the hypothesis that the row proportions are equal and conclude that the two chemical concentrations do not have the same effect. Here $P = \Pr(\chi_1^2 \geq 4.1596) = 0.0414$.

3. A cookie company conducts an experiment to determine if there is a relationship between the amount of a food preservative and the flavor of the cookies. Twelve batches of cookies are made with variable amounts of preservative. A panel of professional cookie tasters rate the cookies from 1 (bad) to 5 (good). The results are presented below. Assume that both the amounts of preservative and the panel ratings are normally distributed. Find the (Pearson) correlation coefficient r for these data and determine whether or not the food preservative levels and taste ratings are correlated. (Use a two-sided test with $\alpha = 0.05$.)

Cookie Batch	Food Preservative	Average Panel Rating
1	30	4.3
2	47	3.6
3	26	4.5
4	94	2.8
5	67	3.3
6	83	2.7
7	36	4.2
8	77	3.9
9	43	3.6
10	109	2.2
11	56	3.1
12	70	2.9

Answer: The Pearson $r = -0.87712$, $T = r\sqrt{(n-2)/(1-r^2)} = -5.775$ for $n = 12$ and $n - 2 = 10$, and $t(0.975, 10) = 2.2281 < 5.775$. Thus we reject the hypothesis that the amount of food preservative and flavor are uncorrelated and conclude that they are correlated. (In fact, $P = 2\Pr(t(10) \geq 5.775) = 0.000179$.)

4. A company wants to study the properties of a new filter for water purification systems. An experiment is conducted by varying levels of iron in water, passing it through the filter, and recording the level of iron in the water after the filter. The data are recorded below. Assume that the data is normally distributed. Find the least-squares regression line of the Y column on the X column. Use this to find a 95% confidence interval for the mean concentration of iron after filtration of water containing 22 μg/L of iron. (***Warning:*** Be careful that you and your calculator agree on which numbers are Xs and which numbers are Ys!)

Iron Concentrations in Water ($\mu g/L$)

Observ.	After (Y)	Before (X)
1	0.28	0.45
2	0.15	0.64
3	0.39	0.77
4	0.51	2.36
5	0.75	2.62
6	9.70	17.50
7	9.40	22.20
8	28.00	32.00
9	22.30	37.50
10	23.00	39.00

Answer: The least-squares regression of Y on X is $Y = -0.7110{+}0.6553X$. The predicted value at $X = 22$ is $Y(22) = -0.7110 + 0.6553 \times 22 = 13.7056$. The standard deviation of the regression errors is $s = \widehat{\sigma} = 3.3099$. Here $n = 10$ and $t(0.975, 8) = 2.306$, $\overline{X} = 15.504$, and $s_X = 16.1941$. The formula for the 95% confidence interval for $E(Y \mid X = 22)$ yields (11.0845, 16.3255).

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, $P = 1.34 \times 10^{-9} < 0.05$. 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, VitD and VitE (but not Sex) have coefficients that are significantly different from zero. The P-values are $P = 0.00115$ for VitD and $P = 2.53 \times 10^{-10}$ 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 32.770. Since Sex is coded as Female=0 and Male=1, this means that the fitted lines for males have an additional additive term of 32.770, which says that they recover (as predicted by the regression equation) 32.770 days later, so that women recover sooner (as measured by this term in the regression equation). The difference is not statistically significant, since $P = 0.1073$ 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 28.9053 days and model (B) predicts a recovery time of 165.412 days.

6. (5 points) Which model 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	4	11	15	16	9	5	Total: 60

Answer: If n_i are the counts in the table, the goodness-of-fit statistic is $T = \left(\sum_{i=1}^{6} n_i^2/10\right) - 60 = 12.40$. Since $\chi^2(0.95, 5) = 11.07$, this is significant at $\alpha = 0.05$ and we reject the hypothesis that the face probabilities are equal. Here $P = \Pr(\chi^2(5) \geq 12.04) = 0.0297$.