POS 3713
Spring 2001
Key for Review Problems for Final Exam

1.

a. Write down the appropriate null and alternative hypotheses for Dr. Bailey's model. Do the results support his theoretical hypotheses?

Solution: Numplants = a + b_1*Tax Rate + b_2*Infrastructure

$H_o: b_1 = 0$ $H_o: b_2 = 0$

$H_1: b_1 < 0$ $H_1: b_2 > 0$

df = N - K - 1 = 31 - 2 - 1 = 28

Tax Rate: t = b/s.e. = -1.494506/0.625466 = -2.389

-2.389 < -1.70 (critical t) so reject H_o

Infrastructure: t = b/s.e. = 0.001716/.000666 = 2.577

2.577 > 1.70 (critical t) so reject H_o

Conclusion: Both the tax rate and the amount of money spent by the city on infrastructure have a statistically significant effect on the number of businesses locating in a city. The results support Dr. Bailey's theoretical hypotheses (the signs are in the predicted direction and the results are significant).

b. What is the substantive impact of the tax rate and money spent on infrastructure on the number of businesses that locate within a city? In Dr. Bailey's model, what impact does a one percent increase in the tax rate have on the number of businesses that locate in a city? What impact does a one unit increase in money spent on infrastructure (1 unit = $1000) have on the number of businesses that locate in a city?

Solution: Tax Rate: For every 1% increase in the tax rate, we expect 1.495 fewer businesses to locate in a city. Infrastructure: For every $1000 spent on infrastructure, we expect .0017 more businesses to locate in a city. An increase of $1 million for infrastructure would attract 1.716 more businesses.

c. What does the reported R^2 tell you about the overall fit of Dr. Bailey's model? What does the reported F-statistic tell you about the overall fit of the model? Does his model do a good job in explaining the number of businesses that locate in a city?

Solution: R^2 = 0.648 which means that the tax rate and amount of money spent on infrastructure can account for 64.8% of the variance in the number of businesses locating in a city. This R^2 value is quite high (the maximum is one).

The reported F statistic is 5.45, which has a reported p-value of .01. Since the p-value is less than .05 we can reject the null hypothesis ($H_o: b_1 = b_2 = 0$) and conclude that

the model as a whole is a good one. In other words, at least one of the slope coefficients is significantly different from zero.

d. The city of Tallahassee wants to use these results to make a prediction of the number of plants that will locate in our city in the next five years based on this regression model. The tax rate in Tallahassee is 5% (tax rate = 5), while the money spent by the city on infrastructure in the past 5 years is $5 million (infrastructure = 5000). How many businesses do you expect to locate in Tallahassee based on this information?

Solution: Numplants = 5.621624 + -1.494506*Tax Rate + 0.001716*Infrastructure
Numplants = 5.621624 + (-1.494506*5) + (0.001716*5000)
Numplants = 5.621624 + (-7.47253) + 8.58 = 6.729
Thus we expect 6.729 businesses to locate in Tallahassee based on the current tax rate and money allocated to infrastructure.

2.

a. Conduct a χ^2 test for independence. State the null and alternative hypotheses. Assume a 95% level of confidence, i.e., $\alpha = .05$. Do your results support Mitchell and Prins theoretical hypothesis that newer democracies are more likely to fight over territorial issues?

Solution: **H_o: Democracy level and the issues at stake in militarized disputes are independent.**
H_1: More established democracies are less likely to dispute over territory, i.e., the issues at stake in militarized disputes depend on democracy level.

Df = (r - 1)(c - 1) = (2 - 1)(2 - 1) = 1, Critical c^2 = 3.841. Since 8.61 > 3.841 we can reject the null hypothesis of independence. We do find support for Mitchell and Prins' hypothesis.

b. Calculate ϕ (phi) and interpret your results. Is this consistent with what you found in (a)?

Solution: f = Ö(c^2/N) = Ö(8.61/97) = Ö.08876 = 0.2979

Phi shows a moderately weak relationship (maximum is one). The relationship between these two variables is statistically significant but not particularly strong. This is consistent with what we found in part a.

3.

a. Calculate the value of gamma for this table. How do you interpret this number? Does it support the contention that higher levels of education in general promote greater civic education? In other words, are people with more education more likely to be interested in politics?

Solution: Gamma = (N_s - N_d)/(N_s + N_d)

N_s = 78(43 + 34 + 40 + 76) + 69(34 + 76) + 25(40 + 76) + 43(76)
= 78(193) + 69(110) + 25(116) + 3268
= 15054 + 7590 + 2900 + 3268
= 28812

N_d = 37(25 + 43 + 15 + 40) + 69(25 + 15) + 34(15 + 40) + 43(15)
= 37(123) + 69(40) + 34(55) + 645
= 4551 + 2760 + 1870 + 645
= 9826

Gamma = (28812 - 9826)/(28812 + 9826) = 18986/38638 = 0.4914

A gamma of 0.4914 shows a moderately strong, positive relationship between education levels and interest in politics. It demonstrates that people with more education are more likely to be interested in politics.

b. Conduct a statistical hypothesis test for the value of Gamma reported in (a). Suppose that the calculated t statistic equals 3.772 with a p-value less than .0001. State the null and alternative hypotheses. What can you conclude about the relationship between education and interest in politics based on this result?

Solution: **H_o: Gamma = 0**
H_1: Gamma ¹ 0

The reported p-value of .0001 is less than alpha (.05), thus we can reject the null hypothesis and conclude that gamma is significantly different from zero. Education has a significant impact on a person's interest in politics.

4.

a) Calculate the regression line.

Solution:

$$b = \frac{NSXY - (SX)(SY)}{NSX^2 - (SX)^2} = \frac{7(2157.56) - (75)(213.66)}{7(1041) - (75)^2} = \frac{-921.58}{1662} = -.5545$$

a = Ybar - b(Xbar) = 30.5 - (-.5545)(10.7) = 36.43

Y = 36.43 - .5545X

b) How do you interpret the coefficient for X (i.e., b), the number of chocolate donuts Bob eats?

Solution: For every additional donut Bob eats, his running time decreases by 0.5545 minutes.

c) If Bob eats 9 donuts, how fast can he expect to run?

Solution: Y = 36.43 - .5545X = 36.43 - .5545(9) = 31.44

d) Calculate the correlation, or Pearson's *r*, between the # of donuts consumed and the recorded running time. Interpret your results. In other words, is Bob's conclusion that the more donuts he eats, the faster he runs supported?

Solution: $$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} = \frac{7(2157.56) - (75)(213.66)}{\sqrt{[7(1041) - (75)^2][7(6600.105) - (213.66)^2]}}$$

$$= \frac{15102.92 - 16024.5}{\sqrt{[1662][550.1394]}} = \frac{-921.58}{\sqrt{914331.6828}} = \frac{-921.58}{956.207} = -0.9638$$

A correlation of -0.9638 is an extremely strong negative correlation which means that the more donuts Bob eats, the faster he runs. This does support his contention.

5.

a. State the null and alternative hypotheses.

Solution: $H_o: \mu_1 = \mu_2 = \mu_3$

$H_1: \mu_1 \neq \mu_2 \neq \mu_3$

b. Compare the means. Do they seem different just by looking at the mean infant mortality rate (you do not need to do a test here, just tell me if the mean levels look different)?

Solution: Yes, the means do seem different. The mean infant mortality rate in rural areas is 11.77, which is over 2 points higher than the mean rate in suburban areas (9.12). Rural areas experience higher rates of infant mortality on average.

c. Test your hypothesis for a difference in means with analysis of variance (ANOVA), i.e., conduct an F test assuming $\alpha = .05$. Does the infant mortality rate (number of infant deaths per 1000 live births) in these countries vary significantly by the level of urbanization?

Solution: $$F = \frac{SSB/(k-1)}{SSW/(N-k)} = \frac{41.45/2}{164.098/27} = 20.725/6.0777 = 3.41$$

$N_1 = 2$, $N_2 = 27$, Critical F = 3.35

Since 3.41 > 3.35, we can reject the null hypothesis that the means are equal across these groups and conclude that infant mortality does vary based on the level of urbanization.

6. Calculate λ (Lambda) for the data in the table above (assume party identification is the dependent variable). What does Lambda tell us about the relationship between religion and party identification, in other words, how do you interpret your results? If you know that the calculated t value for this statistic is 3.231, with a p-value of .001, can you

conclude that religion has a significant impact on party identification (i.e., is it significantly different from zero)?

Solution: $\lambda = \dfrac{L - M}{L} = \dfrac{84 - 56}{84} = 28/84 = 0.333$

L = 64 + 20 = 84
M = (11 + 0) + (15 + 10) + (10 + 10) = 56

We would make 33.3% fewer errors predicting a person' party identification on the basis of their religion. The p-value for the calculated t statistic is .001, which is less than .05. This means we can reject the null hypothesis that lambda equals zero and conclude that religion is a significant predictor of party identification.