STA 6934 – Fall 2002 – Quiz 4

Print Name: **SSN:**

1) A government health worker was interested in the relationship between the weekly number of death registered as due to bronchitis and pneumonia (Y) and mean weekly temperature (X, in degrees Celsius) in a large city. She samples $n = 26$ weeks from recent winters, observing Y and X for each week (values of X have ranged between 2 and 8 over this period). She believes that the mean number of deaths is linearly related to temperature, and fits the model (and obtains the following estimates):

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad \hat{\beta}_0 = 380.0 \quad \hat{\beta}_1 = -42.0 \quad \hat{\sigma}_{\hat{\beta}} = 3.0$$

a) Give the predicted number of deaths when the mean weekly temperature is 3 degrees celsius. When it is 7 degrees celsius.

b) Give a 95% confidence interval for β_1.

c) Based on your confidence interval, which statement best describes your conclusion (based on $\alpha = 0.05$ significance level).

i) Cannot conclude that the mean number of deaths is related to mean weekly temperature in the range of 2–8 degrees

ii) Conclude that the mean number of deaths increases as mean weekly temperature (in the range of 2–8 degrees) increases

iii) Conclude that the mean number of deaths decreases as mean weekly temperature (in the range of 2–8 degrees) increases

d) If the regression equation "explains" 81% of the variation in the weekly number of deaths due to bronchitis and pneumonia, give the correlation coefficient.

2) A study is conducted to compare the pharmacokinetics of three formulations of an OTC medication. One response of interest was the maximum concentration (C_{max}). A sample of 5 subjects was taken and each subject received each formulation (in random order). The following table gives their results.

Subject	Form 1	Form 2	Form 3
Amber	12.0	14.0	16.0
Ben	8.0	11.0	14.0
Chris	15.0	17.0	16.0
Darryl	10.0	12.0	14.0
Ellen	10.0	11.0	15.0

a) Test whether true (populatuion) mean C_{max} measurements differ among these three formulations based on Friedman's test at $\alpha = 0.05$ significance level. Give the test statistic, rejection region, and conclusion for the test:

H_0 : No difference in formulation mean C_{max} $\qquad$ H_A : Differences in formulation mean C_{max}

b) The (partial) Analysis of Variance is given below. Use the F-test to compare the formulations at $\alpha = 0.05$ significance level. Give the test statistic, rejection region, and conclusion for the test.

Source	Sum of Squares
Treatments (Forms)	40
Blocks (Subjects)	48
Error	10
Total	98

3) A regression model is fit, relating Geritol sales (Y, in \$1000s in the previous month) to the following independent variables for a large retail pharmacy chain (where individual store locations represent units): number of people over 65 living in 3-mile radius (X_1, in 1000s), average household income for households in 3-mile radius (X_2, in \$1000s), and competition ($X_3 = 1$ if there's another pharmacy within 3 miles, $X_3 = 0$ if not). The following estimates and standard errors for the regression coefficients were obtained, based on a sample of $n = 20$ store locations:

Variable	$\hat{\beta}_i$	$\hat{\sigma}_{\hat{\beta}_i}$
X_1	1.50	0.20
X_2	0.10	0.30
X_3	-10.0	2.0

a) Which, if any, of the independent variables are significantly related to Geritol sales, controlling for the remaining independent variables (base each appropriate test at the $\alpha = 0.05$ significance level. That is, test $H_0 : \beta_i = 0$ vs $H_A : \beta_i \neq 0$ for each independent variable.

b) If $\sum(\hat{y} - \overline{y})^2 = 20000$ and $\sum(y - \hat{y})^2 = 6000$,

i) Test $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ vs H_A : Not all $\beta_i = 0$ at $\alpha = 0.05$ significance level.

ii) What proportion of the variation in Geritol sales is "explained" by these three independent variables?