

STA 6934 – Fall 2001 – Quiz 5

Print Name:

SSN:

1) The National Institutes of Health is interested in the relationship between the number of general practitioners (Y) and the number of residents (X , in 1000s) among counties in the United States. A random sample of $n = 20$ counties are selected, and Y and X are obtained. The estimated linear regression equation is:

$$\hat{Y} = 8.0 + 10.0X \quad \hat{\sigma}_{\hat{\beta}_1} = 1.2$$

a) Which is the best interpretation of β_1 ? **(3 Points)**

- i) The mean number of physicians among U.S. Counties.
- ii) The increase in the mean number of physicians per unit increase in residents in a county.
- iii) The increase in the mean number of physicians per 1000 increase in residents in a county.
- iv) The mean number of physicians among counties with 1000 residents.

b) Give a 95% confidence interval for β_1 . **(6 Points)**

c) Based on your interval, what is your conclusion (at the 95% level of confidence)? **(3 Points)**

- i) There is a negative association between the number of general practitioners and residents among US counties.
- ii) There is a positive association between the number of general practitioners and residents among US counties.
- iii) There is no association between the number of general practitioners and residents among US counties.

2) A mail order pharmacy is interested in the relationship between annual sales to its customers (Y , in dollars) and the following two predictor variables: number of catalogs sent to the household during the year (X_1 , which they vary from 5 to 20 among households) and household income (X_2 , in \$1000s). They obtain the fitted equation and partial Analysis of Variance based on $n = 63$ households:

$$\hat{Y} = 15 + 25X_1 + 3X_2$$

Source	Sum of Squares
Model	
Error	28000
Total	100000

a) Give the predicted sales to a household that received 15 catalogs and has an income of \$30000 (note, this means $X_2 = 30$). **(5 Points)**

b) What proportion of the variation in annual sales is “explained” by using number of catalogs and household income to predict sales? **(3 Points)**

3) A sample of 54 elderly men are given a psychological examination to determine whether symptoms of senility are present. The Wechsler Adult Intelligence scale (WAIS) is given to each subject. Independently of this scale, it is determined whether or not the subject suffers from senility ($Y = 1$ if senile, $Y = 0$ if not). A logistic regression model is fit, relating presence/absence of senility to the WAIS score (X). The model parameter estimates and standard errors are given below.

$$\hat{\alpha} = 2.0 \quad \hat{\sigma}_{\hat{\alpha}} = 1.1 \quad \hat{\beta} = -0.36 \quad \hat{\sigma}_{\hat{\beta}} = 0.12$$

a) Give the estimated probability a subject is senile if he scores 4.0 on the WAIS test. What if he scores 16.0? **(8 Points)**

b) Test whether there is an association between WAIS score and presence/absence of senility at the $\alpha = 0.05$ significance level: **(12 Points)**

i) Null Hypothesis:

ii) Alternative Hypothesis:

iii) Test Statistic:

iv) Rejection Region:

v) Conclude:

vi) Is the P -value for this test **larger** or **smaller** than 0.05? (Circle one)

c) Give the ratio of the odds for being senile among people scoring 11 to the odds of being senile among people scoring 10. **(3 Points)**

4) On *Temptation Island*, a CDC epidemiologist is studying the time it takes until inhabitants contract a sexually transmitted disease. Every day, she tests each inhabitant (except for the known carriers (one male and one female), who have been placed on the island in her experiment). Besides the carriers, there are 10 men and 10 women. The following table gives the dates of occurrences of diagnosed STD's for males and females.

Males						Females					
i	$t_{(i)}$	n_i	d_i	$\hat{\lambda}_i$	$\hat{S}(t_{(i)})$	i	$t_{(i)}$	n_i	d_i	$\hat{\lambda}_i$	$\hat{S}(t_{(i)})$
1	2	10	1	.100	.900	1	8	10	1	.100	.900
2	6	9	1	.111	.800	2	12	9	1	.111	.800
3	10	8	3			3	14	8	1		
4	12	5	1			4	18	7	3		

a) Complete the table, obtaining the Kaplan–Meier estimates of the survival functions. **(8 Points)**

b) On the plot, identify the survival functions for males and females. **(4 Points)**

5) A drugmaker wishes to determine whether its new compound reduces risk of stroke in the elderly. They run a controlled trial, randomly assigning 1000 subjects to treatment (their drug or placebo). However, they acknowledge that gender, age, and cholesterol level are also related to risk of stroke. They fit a proportional hazards regression model relating the following independent (predictor) variables to time to fracture:

$$RR(t; x_1, x_2, x_3, x_4) = e^{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4}$$

x_1 — Treatment (1=Active Drug, 0=Placebo)

x_2 — Gender (1=Male, 0=Female)

x_3 — Age (1=65+, 0=64-)

x_4 — Cholesterol (1=250+, 0=250-)

a) Describe individuals in the baseline group? **(4 Points)**

- Treatment:
- Gender:
- Age:
- Cholesterol Level:

b) Based on the following table, give a 95% confidence interval for the relative risk of stroke for subjects receiving active drug relative to those receiving the placebo, controlling for gender, age, and cholesterol level. **(3 Points)**

Variable(x_i)	Estimate ($\hat{\beta}_i$)	Std Error $\hat{\sigma}_{\beta}$	Rel Risk ($e^{\hat{\beta}_i}$)	95% CI
Trt (x_1)	-0.80	0.20	0.45	(0.30,0.67)
Gender (x_2)	1.00	0.30	2.72	(1.49,4.95)
Age (x_3)	0.05	0.02	1.05	(1.01,1.09)
Cholesterol (x_4)	-0.20	0.08	0.82	(0.70,0.96)

c) Based on your previous answer, which statement is appropriate at the $\alpha = 0.05$ significance level? **(5 Points)**

- i) Conclude that subjects have a higher risk of stroke on placebo.
- ii) Conclude that subjects have a higher risk of stroke on the active drug.
- iii) Cannot conclude that risks of stroke differ between the active drug and placebo groups.

d) Give the highest and lowest risk groups, based on the estimated relative risks. **(4 Points)**

- Highest Risk:
- Lowest Risk:

HAVE A SAFE AND HAPPY VACATION/NEW YEAR!!!!!!!!!!!!!!!!!!!!