

Statistics
Ph.D. Preliminary Examination
2002

(Open book, open notes, but No old Exams and solutions)

Name _____

1. (20 points) Assume that Y_1, Y_2, \dots, Y_n are a random sample from a normal population $Y \sim N(0, \sigma^2)$.

a) If $SS_{YY} = Y_1^2 + Y_2^2 + \dots + Y_n^2$, then what is the probability distribution of $\frac{SS_{YY}}{\sigma^2}$? And why?

b) If $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$, then what is the probability distribution of $\frac{\bar{Y}}{\sqrt{SS_{YY}}}$? And why?

2. (20 points) A public opinion survey company wants to conduct a poll on the following issue:

'Do you support an increase in federal cigarette tax?'

- a) *Yes*
- b) *No*

The company believes that the percentage of people who support the tax increase is very close to the percentage of people who do not support the tax increase. The company wants to survey enough people such that the margin of error is less than +/- 2% with 95% confidence level.

At least how many people the company needs to survey with?

3. (20 points) Painkiller is the medicine that suppresses pains, such as headache. Two kinds of new painkillers, say A and B are going to be in the market. Both painkillers have very little side effects, if taken according to the recommended daily dosage. The effectiveness of a painkiller can be measured by the effective duration within which the pain is fully suppressed.

Two groups of volunteers are used to test the effectiveness of the new drug, one group is for A and another group is for B. Each group member took a prescribed dosage and the effective durations (in hours) are recorded. The result is as follows:

Group A (for painkiller A).

Group A member	1	2	3	4	5	6	7	8	9	10	11	12
Hours	7.5	6.8	5.9	8.4	8.0	7.2	7.7	8.6	7.1	6.4	8.2	7.7

Group B (For painkiller B)

Group B member	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Hours	3.8	10.4	11.2	5.1	4.1	12.8	13.5	8.3	6.2	2.8	15.9	10.8	9.4	20.8	3.9

Both painkillers will soon be available to general public, given above testing data, and using appropriate statistical analysis, assuming that effective duration is normally distributed, can you conclude the following to the general public:

- a) Painkiller B has longer average effective duration than that of painkiller A?
- b) Painkiller B has larger variation in effective duration than that of painkiller A?

4. (20 points) (Just provide detailed step by step approach to solve the following problems, do not do data analysis)

An independent automobile evaluation company wants to compare the qualities of 4 different cars, Ford Taurus, Chevy Lumina, Honda Accord and Toyota Camry. Evaluation cars are driven and evaluated by 7 evaluators, each evaluator is asked to give a composite score ranging from 1-10, 10 is the best and 1 is the worst. The result is as follows:

Car Models	Evaluation scores by evaluators						
	Evaluator1	Evaluator2	Evaluator3	Evaluator4	Evaluator5	Evaluator6	Evaluator7
Ford Taurus	6.5	8.2	3.9	9.4	7.5	5.8	7.9
Chevy Lumina	6.3	8.3	3.5	9.2	7.6	5.4	7.4
Honda Accord	7.2	8.6	3.9	9.5	7.3	6.1	8.2
Toyota Camry	7.3	8.5	4.1	9.4	7.2	6.3	8.1

- a) Design a statistical analysis method to conclude that if those 4 cars have comparable quality or not, please give details. (Design method only, no data analysis)
- b) Design a statistical tests to compare
 - Us cars vs Japanese cars (to see if significant quality gap exists)
 - Ford car vs competitors (to see if significant quality gap exists)

Design method only, no data analysis, but give details.

5. (20 points) Suppose that in a problem of simple linear regression, the ten pairs of observed values of x_i and y_i for $i=1, \dots, 10$ are obtained.

i	1	2	3	4	5	6	7	8	9	10
x_i	0.3	1.4	1.0	-0.3	-0.2	1.0	2.0	-1.0	-0.7	0.7
y_i	0.4	0.9	0.4	-0.3	0.3	0.8	0.7	-0.4	-0.2	0.7

Linear regression model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

is assumed.

Also, the following computer printout is obtained:

ANOVA table

Source	DF	Sum of Squares	Mean squares	F-test
Regression	1	1.59	1.59	28.224
Residual	8	0.451	0.056	
Total	9	2.041		

R-Squared: 0.779

Parameter	Beta Coefficient table		
	Value	StDev	t-Value
Intercept	0.147	0.083	1.78
Slope	0.435	0.082	5.313

a) Test the hypothesis $H_0: \beta_1 = 0$,

$$H_1: \beta_1 \neq 0$$

- a) Construct a 90% confidence interval for β_1
- b) Construct 95% prediction interval for new observation when $x=1.5$.