

Statistics 224 Practice exam 4 FINAL Your Name

Friday 12/21/07

Professor Michael Iltis (Lecture 2)

Discussion section (circle yours) :

section: 321 (3:30 pm M)
322 (2:25 pm M)
323 (4:35 pm M)

Problem	max points	points scored
1		
2		
3		
4		
5		
6		
Total	120	

Do all 6 problems.

Rules :

1. No notes allowed
2. Standard hand calculator allowed
3. Numerical answers without supporting work (or rationale) may receive no points
4. Failure to follow rules may result in lost points

1. In an effort to determine the most effective way to teach safety principles to a group of employees at Weedco, four different methods were used. A sample of 20 employees were randomly assigned to one of the four groups. The first group was given programmed instruction booklets and worked through the course at their own pace. The second group attended lectures. The third group watched television presentations, and a fourth was divided into small discussion groups. At the end of the session, a test was given to the four groups. A high score of 10 was possible. The results were :

TEST GRADES

Programmed instruction	Lecture	Group TV	discussion
6	8	7	5
5	7	9	5
6	8	6	6
5	8	8	6
6	8	9	5

The following is an Analysis of Variance Mini-tab software output with missing information:
ANALYSIS OF VARIANCE ON GRADES

SOURCE	DF	SS	MS	F
TREAT	_____	26.550	8.850	_____
ERROR	_____	_____	_____	
TOTAL	_____	36.550		

a) Complete the missing values :

b) Test at the .05 level that there is no difference among the four means.

H_0 :

H_a :

Reject H_0 if

Decision :

2. W&A Beverages has observed the following Overhead Costs associated with Gallons of Output over the past twelve months. Additional columns, their sums and relevant formulas are included to aid in answering questions

Month	x = Thousands of Gallons of Output	y = Overhead cost (in \$1000)	x^2	y^2	xy
January	18	45.6	324	2079.36	820.8
February	44	62.4	1936	3893.76	2745.6
⋮	⋮	⋮	⋮	⋮	⋮
December	40	60.0	1600	3600.00	2400
TOTAL	408	684	14900	39376.00	23600

Additional information

$$S_{xx} = \sum (x_i - \bar{x})^2 = 14900 - \frac{408^2}{12} = 1028, \quad S_{yy} = SST = 39376 - \frac{684^2}{12} = 388$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = 23600 - \frac{(408)(684)}{12} = 344, \quad b = \hat{\beta} = \frac{S_{xy}}{S_{xx}}, \quad a = \hat{\alpha} = \bar{Y} - b \bar{X}$$

$$SSR = b^2 S_{xx}, \quad SST = SSR + SSE, \quad MSE = \frac{SSE}{n-2} = s_e^2$$

100(1- α)% confidence interval for β : $b \pm t_{\alpha/2, n-2} \frac{s_e}{\sqrt{S_{xx}}}$. 100(1- α)% prediction

interval for an individual Y when $x = x_0$: $\hat{y}(x_0) \pm t_{\alpha/2, n-2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$

a) An estimate of the variable overhead cost (in \$1000) per thousand gallons of output is _____

b) An estimate of the monthly fixed cost (in \$1000) is _____

c) The fraction of the total sum of squares error from the mean monthly overhead cost explained by the gallons produced variable and the least square line is _____

d) A 95% confidence interval for the true variable overhead cost estimated in (a) is _____

e) If 30 thousand gallons of output are planned for next month, what would you predict for the overhead cost (to the nearest \$10) Answer : _____

f) A 95% prediction interval (in \$) for the prediction of overhead cost in the previous problem is (to the nearest \$100) _____

3. Let

X = amount of raw material added to a chemical process (in gms)

Y = amount of usable final product (in gms)

Suppose the true regression equation

$$E[Y|x] = -1 + 2.5x$$

is known to hold for $6 \leq x \leq 14$ with error standard deviation $\sigma = .1$ gm.

a) What is the expected usable final product when 10 gms of raw material is added to the chemical process ?

b) By how many gms do we expect usable final product to increase for each additional gm of raw material added ?

c) What is the probability that the amount of usable final product would exceed 24.2 gms if 10 gms of raw material is added ?

4. A physical anthropology study of the ability of individuals to walk in a straight line reported that in a sample of $n = 20$ randomly selected healthy men their cadence (which is the number of strides per second) data had a sample mean of

$$\bar{x} = .9255 \quad \text{and standard deviation} \quad s = .0809 .$$

A normal probability plot yielded substantial support to the assumption that the population distribution of cadence is (approximately) normal.

a) (8 points) Calculate and **interpret** a 95% confidence interval for population mean cadence. Your interpretation should answer the question of whether you know or in what sense the actual population mean cadence lies in the interval you found.

4. b) (6 points) Calculate a 95% prediction interval for a future single value X_{21} :

c) (6 points) Calculate a 95% confidence interval for the true population standard deviation σ of men's cadence.

5. A family that owns two automobiles is selected at random. Let

A = older auto is American

B = the newer auto is American

Suppose these probabilities are known : $P(A)=.8, P(B)=.5$ and $P(A \cap B)=0.4$.

Determine :

a) Draw a Venn diagram for the problem, then find the probability that at least one auto is American.

b) The probability that neither auto is American

c) The probability that the newer auto is American given that the older auto is American.

6. Customers at a gas station select either regular (A) or premium (B) or diesel fuel (C) assume that successive customers make independent choices with $P(A) = .5$, $P(B) = .4$, $P(C) = .1$. Let X_1, X_2 , and X_3 be the number of customers who select regular, premium or diesel fuel respectively (these will depend on the total number of customers under consideration).

a) Among the next 10 customers, how many ways are there for exactly 3 to purchase regular fuel, 5 to purchase premium and 2 to purchase diesel if order does not matter ? Hint : note that there are 10 choose 3 ways for 3 customers to purchase regular and of the remaining 7 customers there are 7 choose 5 ways for 5 to purchase premium (so the remaining 2 purchase diesel).

b) Among the next 10 customers, what is the joint probability $P(X_1=3, X_2=5, X_3=2)$ that exactly 3 will purchase regular fuel, 5 will purchase premium and 2 diesel ? Hint: All selections are made independently. What kind of joint distribution is this ?

b) Among the next 25 customers, what are the mean and variance of the number who select premium fuel ? What kind of random variable is this number and with what parameters ? Hint : either a customer does or doesn't select premium. The probability that premium is not selected is $1 - P(B) = .6$. Explain your reasoning.