Stat 5021, Final exam, Spring '04 Name ______________________

Thirty seven people took this exam. Their mean score was 73.9 with a standard deviation of 17.6.

1. Let X be a discrete random variable whose probability distribution is given in the table just below.

x	0	1	2	4
$p(x)$	0.6	0.2	0.1	0.1

Find the mean or expected value of the random variable $Y = 3X^2 + 12$.

Solution: Since

$$E(X^2) = \sum_x x^2 p(x) = 0^2 \times 0.6 + 1^2 \times 0.2 + 2^2 \times 0.1 + 4^2 \times 0.1 = 2.2$$

we have

$$E(Y) = 3E(X^2) + 12 = 3 \times 2.2 + 12$$

∎

2. The tensile strength of a metal alloy produced by a certain manufacturing process is assumed to follow a normal distribution with mean 60 and variance 16. A specimen is acceptable only if its tensile strength exceeds 52. In a random sample of 300 specimen what is the expected number of acceptable specimen?

Solution: Let p be the probability a specimen is acceptable. Then

$$p = P\{N(60, 16) > 52\} = P\{N(0, 1) > \frac{52 - 60}{4}\} = 0.9772$$

and so the expected number of acceptable specimen in a sample of 300 is $300 \times 0.9772 \doteq 293$.

∎

3. Suppose on any given day the probability that Sam receives at least one junk phone call, i.e. a call from some one trying to sell him something, is 0.5 and the probability that he receives at least one junk email is 0.4. Assume that on any given day these events are independent and independent from day to day.

i) What is the probability that Sam will receive at least one junk phone call and at least one junk email today?

ii) In the next ten days how many days should Sam expect to receive at least one junk phone call and at least one junk email during the day?

Solution:

i) Let A be the event that Sam receives at least one junk phone call and B be the even that he receives at least one junk email. Then by independence

$$P(A \cap B) = P(A)P(B) = 0.5 \times 0.4 = 0.20$$

ii) $10 \times 0.2 = 2$

∎

4. 95% of the widgets produced by supplier I pass inspection while 90% of those produced by supplier II pass inspection. Suppose 60% of the shipments come from supplier I while the remaining 40% come from supplier II.

i) Find an expression for the probability that in a shipment containing 9 widgets at least 7 of them will pass inspection.

ii) Given that in a shipment of 9 widgets at least 7 of them passed inspection find an expression for the conditional probability that the shipment came from supplier I.

Solution:

i) Let A be the event that at least 7 in a shipment of 9 pass inspection then

$$P(A) = P(I)P(A|I) + P(II)P(A|II) = 0.60p_1 + 0.40p_2$$

where

$$p_1 = P(A|I) = \sum_{x=7}^{9} \binom{9}{x} (.95)^x (.05)^{9-x} \quad \text{and} \quad p_2 = P(A|II) = \sum_{x=7}^{9} \binom{9}{x} (.9)^x (.1)^{9-x}$$

ii) By Bayes theorem

$$P(I|A) = \frac{P(I \cap A)}{P(A)} = \frac{P(I)P(A|I)}{P(A)} = \frac{0.60p_1}{0.60p_1 + 0.40p_2}$$

∎

5. A experiment was conducted to study the effect of diet on the time it takes for blood to coagulate. For four different diets four observations of Y, the "blood coagulation time" was collected. The results were

diet	1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4
Y	62	60	63	59	63	67	71	67	68	66	71	67	57	62	60	61

The oneway analysis of variance was done on Y using the model $Y_{ij} = \mu + \alpha_i + Z_{ij}$. Recall that $\mu + \alpha_i = \mu_i$ is the average coagulation time for the ith diet. The results were

	Df	Sum of Sq	Mean Sq
diet	3	200	66.66667
Residuals	12	70	5.83333

i) At level $\alpha = .05$ test H: $\alpha_1 = \alpha_2 = \alpha_3 = 0$ against K: At least one is not zero.

ii) Find a 95% confidence interval for $\mu_1 - \mu_3$.

Solution:

i) Since

$$\frac{66.67}{5.83} > 3.49 = f_{3,12,.05}$$

we reject H.

ii) Since $\hat{\mu}_1 = (62 + 60 + 63 + 59)/4 = 61$ and $\hat{\mu}_3 = 68$ and $t_{12,.025} = 2.179$ the interval is

$$61 - 68 \quad \mp \quad \sqrt{5.83}\sqrt{1/4 + 1/4}\ 2.179$$

∎

6. The following data represents smoking status by level of education for residents of a certain city for adults 30 years or older from a random sample of 200.

Education	Smoking Status		
	Current	Former	Never
< College Degree	40	40	60
≥ College Degree	10	10	40

Test whether smoking status and level of education are independent at the $\alpha = 0.05$ level.

Solution:

Under the hypothesis of independence the expected number in the first cell is $(140 \times 50)/200 = 35$. After similar calculations for the other cells we have

$$\begin{aligned}\sum (obs - exp)^2/exp =& \frac{(40-35)^2}{35} + \frac{(40-35)^2}{35} + \frac{(60-70)^2}{70} \\ &+ \frac{(10-15)^2}{15} + \frac{(10-15)^2}{15} + \frac{(40-30)^2}{30} \\ >& \ 5.99\end{aligned}$$

and so we reject the hypothesis of independence since $P(\chi_2^2 > 5.99) = 0.05$.

∎

7. Data were collected to study the relationship between the variable Y, level of Nitric Oxide (or NOx), and a predictor variable E. A total of 88 observations were collected. It was believed that Y was roughly a quadratic function of E. So the model

$$Y = \beta_0 + \beta_1 E + \beta_2 Esq + Z \qquad (1)$$

was fit to the data where $Esq = E^2$ and the following results were obtained.

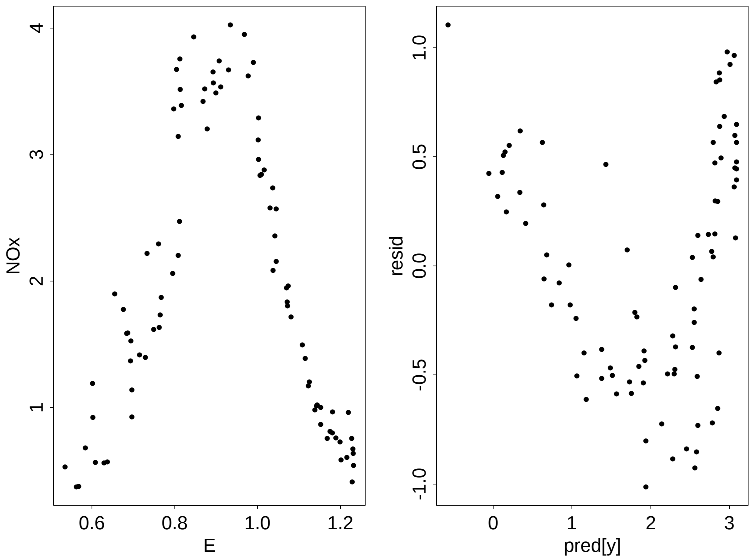
```
Coefficients:
               Value Std. Error  t value
(Intercept) -19.0939   1.2847   -14.8623
          E  49.2177   2.9124    16.8994
        Esq -27.2957   1.5892   -17.1756

n = 88   MSE = 0.29071  DF: 85  R-Squared: 0.7786
F-statistic: 149.5 on 2 and 85 degrees of freedom
```

For the first three data points the observed y-values, the predicted y-values and the standard error of the predicted y-values were as follows.

```
observed y  pred[y]    SE[predy]
3.741       3.091798   0.08724758
2.295       2.553287   0.07979082
1.498       1.929377   0.07732621
```

Below are the plots of y against E and the residuals against the pred[y].



Use the output to answer the following questions. If a question cannot be answered with the output in hand, explain what additional output is needed.

i) In the model of equation (1) find a 99% confidence interval for β_2.

ii) In the model of equation (1) find a 99% confidence interval for the average level of NOx for the third data point with an observed y-value of 1.498.

Next the model

$$Y = \beta_0 + \beta_1 E + \beta_2 Esq + \beta_3 Ecb + \beta_4 Eqt + Z \tag{2}$$

where $Ecb = E^3$ and $Eqt = E^4$ was fit to the data and two analysis of variance tables were found. Remember the summaries are sequential.

variable	Df	Sum of Sq
constant	1	337.1559
E	1	1.15491
Ecb	1	85.97290
Esq	1	0.02387
Eqt	1	13.20688
Residuals	83	11.26525

variable	Df	Sum of Sq
constant	1	337.1559
Esq	1	3.89061
Eqt	1	80.58945
E	1	2.53301
Ecb	1	13.34550
Residuals	83	11.26525

iii) In the model of equation (2) at level $\alpha = .01$ test H: $\beta_2 = \beta_4 = 0$ against K: At least one is not zero.

iv) In the model $Y = \beta_0 + \beta_1 E + \beta_2 Esq + \beta_4 Eqt + Z$ at level $\alpha = .01$ test H: $\beta_1 = 0$ against K: $\beta_1 \neq 0$.

v) In the model $Y = \beta_0 + \beta_1 E + \beta_2 Esq + \beta_4 Eqt + Z$ find the value of the least squares estimate of β_1.

vi) Is there anything in the computer output on page 4 which suggests that the quadratic model in equation 1 should not be used for the prediction of NOx using some function of E. Which of the two models given in equations 1 and 2 would you recommend. Briefly justify your answer.

Solution:

i) Since $t_{85,.005} \doteq 2.65$ we have

$$-27.30 \quad \mp \quad 1.59 \times 2.65$$

ii)

$$1.929 \quad \mp \quad 0.0773 \times 2.65$$

iii) Since $f_{(2,83),.01} \doteq 4.92$ and

$$\frac{(SSReg(\beta_2|\beta_0,\beta_1,\beta_3) + SSReg(\beta_4|\beta_0,\beta_1,\beta_3,\beta_2))/2}{MSE} = \frac{(0.02387 + 13.20688)/2}{11.26525/83} > 4.92$$

we reject H. Note here we are using the information from the first anova table.

iv) Now we use the information from the second anova table and the fact that $f_{(1,84),.01} \doteq 7.00$. Since

$$\frac{SSReg(\beta_1|\beta_0,\beta_2,\beta_4)/1}{MSE} = \frac{2.533}{(13.345 + 11.265)/84} = 8.646 > 7.00$$

we reject H.

v) Cannot answer. You need to fit this model to the data to find the least squares estimates.

vi) The plot of the residuals versus the predicted y for model 1 shows a pattern which indicates that model 1 is not adequate. You might not guess this just looking at the plot of NOx versus E. From either of the two anova tables we can find R^2 for model 2. For example from the first table

$$R^2 = \frac{1.15 + 85.97 + 0.02 + 13.21}{1.15 + 85.97 + 0.02 + 13.21 + 11.26} = 0.90$$

Since 0.90 is larger than 0.78, the R^2 value for model 1, we prefer model 2.

∎