Stat 330X Final Exam

May 4, 2000 Prof. Vardeman

1. Attached to this exam is a Mintab printout summarizing some regression work on a small data set from *The Real Estate Appraiser and Analyst*, giving 1986 selling prices y (in thousands of dollars), of some homes of size x_1 (in 100 ft^2), and of "condition" x_2 ($1 =$ poor, $10 =$ excellent). Use it to answer the following questions.

Consider first treating y as a function of $x_1$alone.

(a) What fraction of the raw variability in y is accounted for by an equation linear in x_1?

(b) What do you estimate as a mean 1986 selling price in this region for a home of 2000 ft^2?

(c) What do you estimate to be the standard deviation of selling prices for homes of fixed size in 1986?

(d) Give a 90% two-sided confidence interval for the increase in mean selling price that accompanied a 200 square foot increase in home size in 1986.

(e) Real estate records for this region are to be checked for 1986 and the price of a single additional 2000 square foot home obtained. Give two-sided limits that you are "90% sure" will contain this additional price.

Now consider treating y as a function of both x_1 and x_2.

(f) On the basis of fraction of variability in selling price "explained," do you believe that x_2 adds in an important way to one's ability to predict/explain selling price? Explain in terms of some figures from the printout.

(g) What do you estimate to be the standard deviation of selling prices for homes of fixed size *and condition* in 1986?

(h) What do you estimate to be the difference in selling prices of a "condition $=$ 1" home and a "condition $=$ 10" home of a given size?

(i) There is information on the printout that will let you give a 95% confidence interval for the mean 1986 selling price of a 2000 ft^2 condition $=$ 5 home. Give this interval.

2. A corporate executive is responsible for the purchase and maintenance of standard desktop computer hardware across his company. The vendor of the current hardware guarantees units to have 3 year mean time to catastrophic failure. Suppose that in addition, the executive believe lifetimes of these units to be exponentially distributed.

(a) Evaluate the probability that a single one of these units experiences a time to failure of .5 year or less.

(b) Under the executive's model for the life of one of these units, the probability of failure in 1 year or less is about .28. In one corporate office, 40 out of 100 of these units fail in the first year of operation. Assess the strength of the evidence in this data to the effect that the one year failure rate for these machines exceeds 28%. (Show all 5 steps.)

(c) Use the mean and variance of the exponential distribution and make an appropriate approximation to the probability that (under the executive's model for unit life) the sample mean time to failure for $n = 50$ of these machines is less than 2.5 years.

3. Suppose that for X exponentially distributed with mean 1, the behavior of $V = \sin(X)$ is of interest.

(a) Write out (but do not attempt to evaluate) a definite integral giving the mean of V, E$V = \mathrm{E}\sin(X)$.

Attached to this exam is printout useful in studying V. It summarizes the simulation of 1000 values of V. Use it in the rest of this question.

(b) Give approximate 95% two-sided confidence limits for $P[V < -.05]$. (No need to simplify.)

(c) Give approximate 95% two-sided confidence limits for $\mathrm{E}V$. (No need to simplify.)

4. In the inspection of newly produced 500 MHz Power PC chips, a test method 1) correctly detects 90% of faulty chips and 2) incorrectly rejects 5% of OK chips. Suppose that, in fact 20% of newly produced chips are faulty.

a) What fraction of chips are faulty and correctly identified as faulty?

b) What fraction of all chips are identified as faulty by this test method?

c) What fraction of those chips identified as faulty actually are faulty?

5. A small but busy convenience store has room in its building for only 4 customers. It has 2 employees who work (interchangeably) waiting on customers and performing other tasks (such as stocking shelves). The work rule the employee follow is this: When the are 3 or 4 customers in the store, both wait on customers. When there are 2 or fewer customers in the store, only 1 employee waits on customers. Customers arrive at the store according to a Poisson process with rate $\lambda = 1$ (and go away if there is no room in the store). We will assume that they spend no time finding their items and that either employee can service them at a rate $.8$. THIS IS NOT A STANDARD QUEUEING PROBLEM. NO FORMULAS IN THE BACK OF THE BOOK ARE RELEVANT. WORK FROM 1st PRINCIPLES.

(a) Carefully draw an appropriate birth and death process transition rate diagram for this scenario.

(b) Find the long run probability that there is no one in the store.

(c) What fraction of potential customers are turned away from this store?

(d) Find the long run mean number of employees at work serving customers. (If no one is in the store, both work at other tasks.)

Printout for Problem 1

Data Display

Row	size	condition	saleprice
1	23	5	60.0
2	11	2	32.7
3	20	9	57.7
4	17	3	45.5
5	15	8	47.0
6	21	4	55.3
7	24	7	64.5
8	13	6	42.6
9	19	7	54.5
10	25	2	57.5

Plot

```
                       Scatterplot of Price Versus Size
         -
         -                                                     *
         -
S      60+                                                 *
a        -                                     *                   *
l        -                                 *       *
e        -
         -
P      48+                 *
r        -                         *
i        -         *
c        -
e        -
       36+
         - *
         -
         -
           ------+---------+---------+---------+---------+---------+
              12.5      15.0      17.5      20.0      22.5      25.0
                                   Home Size
```

Regression Analysis

The regression equation is
saleprice = 16.0 + 1.90 size

Predictor	Coef	StDev	T	P
Constant	16.008	4.805	3.33	0.010
size	1.9001	0.2486	7.64	0.000

S = 3.529   R-Sq = 88.0%   R-Sq(adj) = 86.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	727.85	727.85	58.43	0.000
Residual Error	8	99.65	12.46		
Total	9	827.50			

Unusual Observations

Obs	size	salepric	Fit	StDev Fit	Residual	St Resid
10	25.0	57.50	63.51	1.90	-6.01	-2.02R

R denotes an observation with a large standardized residual

```
MTB > Regress 'saleprice' 2 'size' 'condition';
SUBC>   Constant;
SUBC>   Predict 20 5;
SUBC>   Brief 2.
```

Regression Analysis

The regression equation is
saleprice = 9.78 + 1.87 size + 1.28 condition

Predictor	Coef	StDev	T	P
Constant	9.782	1.630	6.00	0.001
size	1.87094	0.07617	24.56	0.000
conditio	1.2781	0.1444	8.85	0.000

S = 1.081 R-Sq = 99.0% R-Sq(adj) = 98.7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	819.33	409.66	350.87	0.000
Residual Error	7	8.17	1.17		
Total	9	827.50			

Source	DF	Seq SS
size	1	727.85
conditio	1	91.48

Unusual Observations

Obs	size	salepric	Fit	StDev Fit	Residual	St Resid
10	25.0	57.500	59.112	0.766	-1.612	-2.11R

R denotes an observation with a large standardized residual

Predicted Values

Fit	StDev Fit	95.0% CI	95.0% PI
53.592	0.357	(52.747, 54.436)	(50.899, 56.284)

Printout for Problem 3

```
MTB > Random 1000 c1;
SUBC>   Exponential 1.0.
MTB > let c2=sin(c1)
MTB > Describe C2.

Descriptive Statistics


Variable             N       Mean     Median     TrMean      StDev    SE Mean
C2                1000     0.4778     0.5017     0.5061     0.4060     0.0128

Variable       Minimum    Maximum         Q1         Q3
C2             -0.9983     1.0000     0.1937     0.8297

MTB > GStd.
* NOTE  * Character graphs are obsolete.

MTB > Histogram C2;
SUBC>   Start -1 1;
SUBC>   Increment .1.

Histogram


Histogram of C2   N = 1000
Each * represents 5 observation(s)

Midpoint        Count
 -1.0000            8  **
 -0.9000            3  *
 -0.8000            7  **
 -0.7000            4  *
 -0.6000            3  *
 -0.5000            4  *
 -0.4000            4  *
 -0.3000            5  *
 -0.2000            6  **
 -0.1000            3  *
  0.0000           81  *****************
  0.1000           92  *******************
  0.2000           86  ******************
  0.3000           90  ******************
  0.4000           72  ***************
  0.5000           66  **************
  0.6000           71  ***************
  0.7000           74  ***************
  0.8000           86  ******************
  0.9000           91  *******************
  1.0000          144  *****************************
```