

**FINAL EXAM**  
**STAT 572**  
**Spring 1996**

Name \_\_\_\_\_

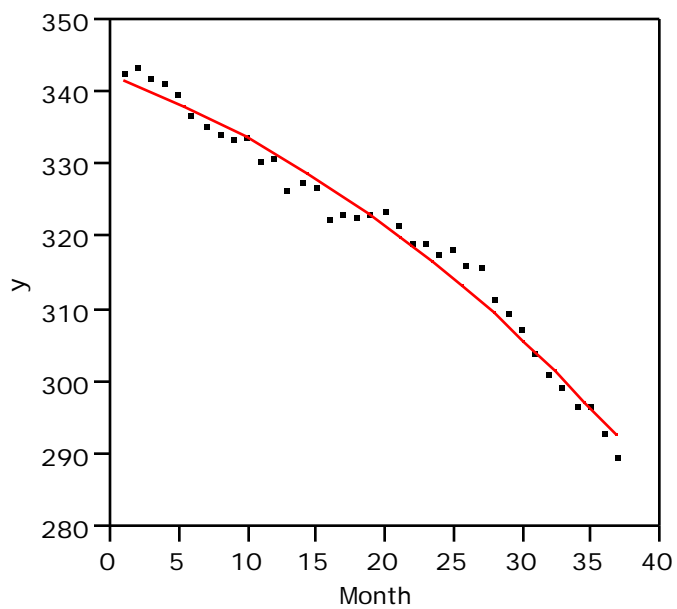
Instructions: Do not spend a lot of time on one problem. If you get stuck go to the next one. Show all your work for partial credit.

1. (Data: Frees, p. 432) Health care analysts are concerned with levels of utilization of nursing home facilities. To investigate this problem, monthly data were gathered from June, 1990 to June 1993, of a nursing home facility. The measure of utilization was daily patients days. The objective was to develop a model for short term forecasts of utilization. The data appears below.

OBS	MONTH	Y	PREDAUTO
1	1	342.870	341.780
2	2	343.490	341.894
3	3	341.990	342.110
4	4	341.180	340.791
5	5	339.715	339.956
6	6	336.835	338.639
7	7	335.350	336.294
8	8	334.310	334.939
9	9	333.595	333.892
10	10	334.090	333.068
11	11	330.755	333.101
12	12	331.105	330.370
13	13	326.740	330.276
14	14	327.595	326.782
15	15	327.025	327.027
16	16	322.730	326.237
17	17	323.335	322.759
18	18	322.935	322.789
19	19	323.350	322.086
20	20	323.605	321.957
21	21	321.970	321.702
22	22	319.400	320.077
23	23	319.360	317.768
24	24	317.725	317.266
25	25	318.575	315.605
26	26	316.205	315.719
27	27	315.905	313.508
28	28	311.425	312.772
29	29	309.540	309.021
30	30	307.585	307.124
31	31	304.125	305.164
32	32	301.135	302.111
33	33	299.425	299.384
34	34	296.870	297.566
35	35	296.835	295.128
36	36	292.995	294.490
37	37	289.800	291.106

An analysis of the data appears in the following page.

## y By Month



— Polynomial Fit degree=2

## Polynomial Fit degree=2

$$y = 342.193 - 0.64022 \text{ Month} - 0.01875 \text{ Month}^2$$

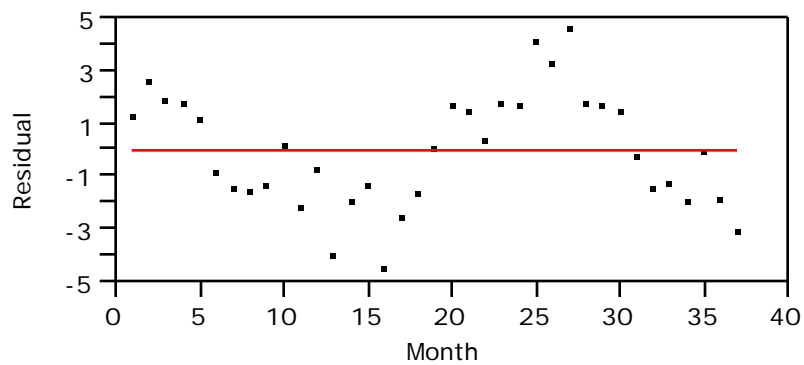
## Summary of Fit

## Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	7854.8006	3927.40	791.7726
Error	34	168.6489	4.96	Prob>F
C Total	36	8023.4495		<.0001

## Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	342.19299	1.160601	294.84	<.0001
Month	-0.640217	0.140842	-4.55	<.0001
Month^2	-0.018754	0.003595	-5.22	<.0001



a) Comment on the results of this analysis (including residuals and  $R^2$ ). (10 p)

b) Use this model to forecast the utilization for the following two months. (5 p)

Using PROC AUTOREG it was found that only the first order autocorrelation term was significant. Results are shown in the following page.

## Autoreg Procedure

Dependent Variable = Y

## Ordinary Least Squares Estimates

SSE	168.6489	DFE	34
MSE	4.960263	Root MSE	2.227165
SBC	171.9596	AIC	167.1268
Reg Rsq	0.9790	Total Rsq	0.9790
Durbin-Watson	0.4979		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	342.192992	1.1606	294.841	0.0001
MONTH	1	-0.640217	0.1408	-4.546	0.0001
MONTH2	1	-0.018753524	0.0036	-5.217	0.0001

## Estimates of Autocorrelations

Lag	Covariance	Correlation	-1	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	1	
0	4.558079	1.000000																						*****
1	3.275149	0.718537																						*****

Preliminary MSE = 2.204764

## Estimates of the Autoregressive Parameters

Lag	Coefficient	Std Error	t Ratio
1	-0.71853704	0.12106904	-5.934936

## Yule-Walker Estimates

SSE	74.93206	DFE	33
MSE	2.270669	Root MSE	1.506874
SBC	146.281	AIC	139.8373
Reg Rsq	0.9287	Total Rsq	0.9907
Durbin-Watson	2.1301		

Variable	DF	B Value	Std Error	t Ratio	Approx Prob
Intercept	1	342.407474	2.0728	165.190	0.0001
MONTH	1	-0.607065	0.2488	-2.440	0.0202
MONTH2	1	-0.0206461	0.0063	-3.278	0.0025

c) Write out the estimated model that incorporates the autocorrelation (5 p)

d) Use the structural and autoregressive parts of the model to forecast the utilization for the next two months. Compare these forecasts with the ones in b) (10 p) [Hint: The last column in page one contains the forecasts from the model that incorporates the autoregressive part]

2. (Data Frees p. 367). It is of interest to determine what influences a taxpayer to have a professional prepare his/her tax return. A sample of 192 individual returns were selected and a number of variables were recorded. Of those, only three were studied:

f1040: form 1040 used? (0 if no, 1 if yes)  
 tpi: total personal income (in dollars)  
 emp: self-employed? (0 if no, 1 if yes)

The response was prep (1 if professional help, 0 otherwise). The results of the analysis appear below.

**Important:** The two dummy regressors, f1040 and emp, were coded as shown above and declared continuous in the analysis.

Response: prep

### Iteration History

Converged by Gradient

### Whole-Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	114.10604	3	228.2121	<.0001
Full	18.13323			
Reduced	132.23927			
RSquare (U)		0.8629		
Observations (or Sum Wgts)		192		

### Lack of Fit

### Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	41.4130512	7.9807282	26.93	<.0001
f1040	-3.631583	1.1729213	9.59	0.0020
tpi	-0.0011048	0.0002153	26.33	<.0001
emp	0.98399456	1.3417191	0.54	0.4633

### Effect Test

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
f1040	1	1	9.586371	0.0020
tpi	1	1	26.334237	0.0000
emp	1	1	0.537851	0.4633

### Effect Likelihood-Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
f1040	1	1	14.05440	0.0002
tpi	1	1	182.27675	0.0000
emp	1	1	0.55137	0.4578

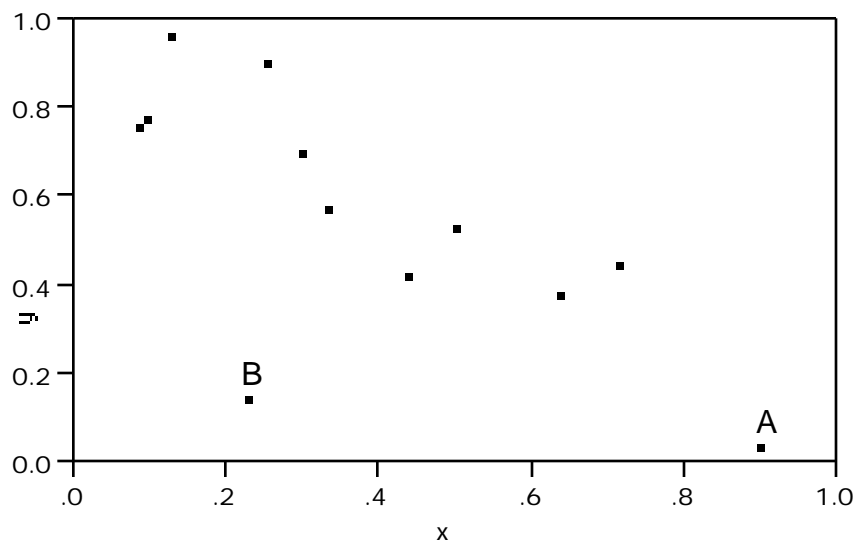
a) Comment on the fit of the model. Be thorough. (10 p)

b) Write out the estimated model (5 p)

c) Predict the probability that a taxpayer hires a professional to prepare her tax form if, she files a 1040 form, has a total personal income of \$50,000, and is self-employed .(5 p)

d) Predict the total personal income of a taxpayer that seeks professional help 80% of the time and who files a 1040 form and is not self-employed. (5 p)

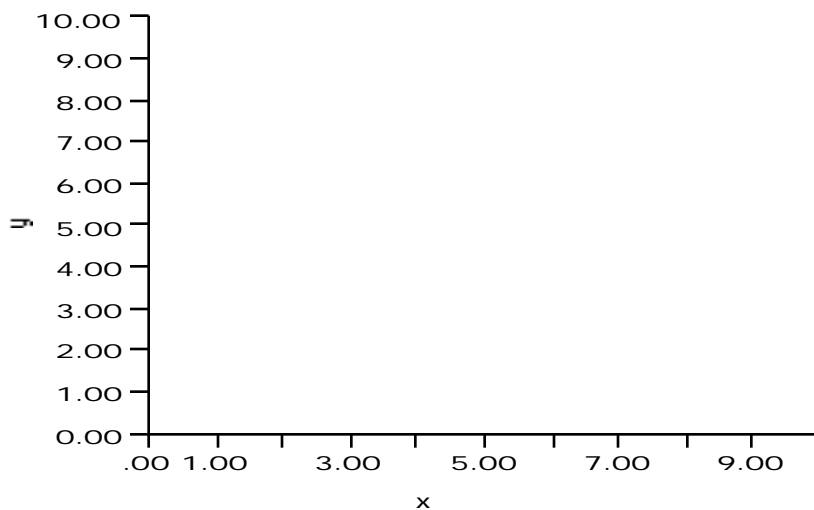
3. Consider the following scatterplot. (10 p)



For points A and B, determine whether each of the following measures are high, medium, or low and whether they are positive or negative. Example: High (-)

Point	Residual	Leverage	DFBETAS	COVRATIO
A				
B				

4. Draw a scatterplot below showing an observation with high leverage and large negative DFBETAS.



**BONUS**

You are analyzing a data set with  $n = 100$  using SLR. The  $i$ th observation has a residual of  $e_i = 5s$ , where  $s^2$  is the mean square error. What percentage of the error sum of squares, SSE, is due to this observation? (5 p)