PART II (Take Home) Eco 400 MidTerm

Please answer the following 4 questions as you analyze this problem using STATA.

1. Run 5 different linear regression equations using only some selected variables of those provided. One regression can contain all of the variables, if you wish. Determine which one is best and give the criteria used.
2. Run 3 different specifications of the functional form of the regression equation (aside from linear). Select the one that is best and give the criteria used.
3. Then select your final form and best fit equation and explain the meaning of all coefficients, their significance and the overall significance of the equation.
4. Are there any other variables that you might have liked to include in the analysis of the problem? Why?

Building a Model of the Demand for Water in Los Angeles

This model is an extension of the short example presented in Section 2.5. The goal is to build a model of the demand for water in Los Angeles County as a function of population, rainfall, and other variables. The dependent variable is:

W_t = Consumption of water in Los Angeles county in year t (hundreds of millions of gallons).

A few comments about the data are in order. Population data for Los Angeles county were unavailable for the entire sample when the data were collected, so population in Glendale was used as a proxy. Second, conservation efforts took place only during the last two years in the sample, so if conservation efforts have an effect but with a lag, we'd be forced to use a "one-time dummy." Thus there are reasons to be concerned that population and conservation are not accurately measured in the sample. The means, standard deviations, and simple correlation coefficients of the variables, plus the raw data themselves, are contained in tables on the next two pages.

The available independent variables are:

POP_t = Population in Glendale, California in year t (thousands of people).

T_t = Average annual temperature at the Los Angeles Civic Center in year t (degrees F).

R_t = Inches of rainfall at the Los Angeles Civic Center in year t.

PR_t = Average price of a gallon of water in Los Angeles in year t (dollars).

CO_t = A dummy variable to measure the conservation efforts undertaken in the last two years in the sample, when there was a severe drought in Northern California (the source of much of L.A.'s water) even though no such drought took place in Southern California.

Y_t = Total personal income for Los Angeles County in year t (in billions of dollars).

Means and Variances for the Demand for Water Data

Variable	Mean	Standard Deviation
W	71.01	7.97
POP	126.83	9.78
T	65.15	1.24
R	13.75	6.20
PR	0.33	0.13
CO	0.08	0.27
Y	26.91	13.55

Simple Correlation Coefficients for the Data

Y, PR	0.954	CO, PR	0.764
Y, T	0.108	CO, T	0.096
Y, CO	0.639	CO, R	–0.036
Y, POP	0.828	CO, POP	0.294
Y, R	0.009	T, POP	0.158
PR, POP	0.630	T, PR	0.079
PR, R	–0.024	R, T	–0.063
		R, POP	0.124

Data for the Demand for Water Interactive Exercise

YEAR	W	POP	T	R	PR	Y
1	53.38500	108.700	63.6	26.21	0.23	9.5345
2	60.31100	110.780	64.6	9.46	0.22	10.7830
3	60.06400	113.340	64.9	11.99	0.22	11.9910
4	58.93900	114.790	63.7	11.94	0.23	13.0190
5	62.33000	115.780	64.8	16.00	0.22	14.2880
6	62.36100	117.110	65.7	9.54	0.23	15.4280
7	65.40900	118.260	67.1	21.13	0.25	15.8760
8	71.91400	118.380	67.3	5.58	0.25	17.3270
9	70.41700	119.440	65.7	8.18	0.26	18.0020
10	73.54900	121.880	65.3	4.83	0.26	18.8880
11	69.09300	123.120	62.3	18.79	0.27	20.3420
12	67.47900	126.770	65.1	8.38	0.26	21.5810
13	73.27700	129.830	63.2	7.93	0.26	22.9790
14	68.94700	132.440	64.3	13.69	0.26	24.2280
15	73.43700	135.790	66.0	20.44	0.26	26.0990
16	72.34000	137.050	66.5	22.00	0.28	27.9240
17	76.04400	137.270	66.2	16.58	0.30	29.9080
18	75.15500	137.920	65.4	27.47	0.34	32.4070
19	81.66400	132.660	66.1	7.77	0.37	34.0980
20	80.54400	132.750	65.1	12.32	0.39	34.3140
21	81.41800	133.110	66.5	6.54	0.43	37.1160
22	76.70700	132.700	65.0	17.45	0.44	39.1910
23	80.61500	136.510	65.3	16.69	0.47	43.7840
24	80.06200	138.010	63.0	10.70	0.51	47.8220
25	81.46700	139.260	65.2	11.01	0.60	53.3340
26	69.31500	133.920	65.9	14.97	0.75	59.2670

Table B-1 Critical Values of the *t*-Distribution

Degrees of Freedom	Level of Significance: One-Sided: 10% / Two-Sided: 20%	5% / 10%	2.5% / 5%	1% / 2%	0.5% / 1%
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
(Normal) ∞	1.282	1.645	1.960	2.326	2.576

Source: Reprinted from Table IV in Sir Ronald A. Fisher, *Statistical Methods for Research Workers*, 14th ed. (copyright © 1970, University of Adelaide) with permission of Hafner, a division of the Macmillan Publishing Company, Inc.