22S:30/105, Statistical Methods and Computing
Spring 2015, Instructor: Cowles
Final Exam

Name: _____ Course no. (30 or 105) _____

1. The following question appears on a survey of high risk behavior: "Have you ever used methamphetamine? (yes/no)."

   (a) The responses to this question provide a variable of which data type? (Circle the one best answer.)

      i. binary
      ii. nominal
      iii. ordinal
      iv. quantitative discrete
      v. quantitative continuous

   (b) The data from this question could be represented by which kinds of plots? (Circle all that apply.)

      i. bar chart
      ii. histogram
      iii. line plot
      iv. pie chart
      v. stem-and-leaf plot
      vi. scatterplot

   (c) The data from this question could be used to draw inference about which kind of population parameter? (Circle the one best answer.)

      i. a population mean
      ii. a population proportion
      iii. a population standard deviation
      iv. none of the above

   (d) Which is the most appropriate type of confidence interval for the parameter you selected in the previous question? (Circle the one best answer.)

      i. a chi-square interval
      ii. a t interval
      iii. a z interval

(e) The standard symbol for the *point estimate* of the population parameter that you selected is: (Circle the one best answer.)

i. μ
ii. $\bar{x}$
iii. s
iv. σ
v. p
vi. $\hat{p}$
vii. π

2. This question is based on a dataset described as follows:

```
NAME: Televisions, Physicians, and Life Expectancy
TYPE: Sample
SIZE: 40 observations, 6 variables

DESCRIPTIVE ABSTRACT:
For each of the forty largest countries in the world (according to 1990
population figures), data are given for the country's life expectancy
at birth, number of people per television set, and number of people per
physician.

SOURCE:
_The World Almanac and Book of Facts 1993_ (1993), New York:  Pharos
Books.

VARIABLE DESCRIPTIONS:
Columns
 1 - 15  Country
23 - 26  Life expectancy
29 - 33  People per television
36 - 40  People per physician
43 - 44  Female life expectancy
47 - 48  Male life expectancy

Additional information about these data can be found in the "Datasets
and Stories" article "Televisions, Physicians, and Life Expectancy" in
the _Journal of Statistics Education_ (Rossman 1994).

SUBMITTED BY:
Allan J. Rossman
Dickinson College Department of Mathematics and Computer Science
P.O. Box 1773
Carlisle, PA  17013
rossman@dickinson.edu
```

The variable names that I used were:

- Country – name of country
- life – life expectancy at birth
- pplpertv – number of people per television set
- pplperdr – number of people per doctor
- logtv – log of number of people per tv
- logdr – log of number of people per doctor

(a) We wish to develop a linear regression model to predict life expectancy. Look at the scatterplots of `life` versus `pplpertv` and `life` versus `logtv`. Briefly explain why we should use `logtv` rather than `pplpertv` in a simple linear regression model to predict `life`.

(b) Based on the scatterplot, is the relationship between `life` and `logtv` negative or positive?

(c) Consider the two regression models. Which of the two predictor variables, `logtv` or `logdr` explains more of the variability in `life`? Cite specific SAS output to justify your answer.

(d) The SAS output includes a test of the hypothesis that there is no linear relationship between `life` and `logtv` in the population of all large countries.

i. Write this null hypothesis as a statement about population parameters, using conventional statistical symbols.

ii. At the .05 significance level, should you reject the null hypothesis? (yes or no) Justify your answer by citing numerical values from the SAS output for:

A. a test statistic and p-value

B. a confidence interval for the parameter

iii. Does your answer to the previous question mean that giving more television sets to people in a country will cause life expectancy to go up in the country? Briefly explain.

iv. Based on the regression model, what is the point estimate of the average life expectancy in all countries with `logtv` = 1.386?

A. Calculate this value by hand; show your work.

B. Can this estimated value be found in the SAS output? If so, explain where to find it.

3. Researchers designed an experiment to assess the effects of different antihistamines on drivers' alertness. The experiment was carried out using a driving simulator. The 80 participants were randomly assigned to 4 different groups, each of 20 people. Participants in reatment groups 1, 2, and 3 received antihistamines 1, 2, and 3 respectively, and participants in treatment group 4 received placebo.

   The procedure was that each participant was given a single dose of the type assigned to his group. Thirty minutes later the participant was placed in the driving simulator and told to drive normally in response to the scene shown in the simulators windshield. At some point during the session, the image of a dog running into the street in front of the car would appear. The measured variable was the time between the appearance of the dog and the participant applying the brakes.

   (a) What were the experimental units in this experiment? (Circle the one best answer.)

   i. the 80 people
   ii. the 4 groups
   iii. the 3 antihistamines and placebo
   iv. the time to apply the brakes

   (b) What were the treatments? (Circle the one best answer.)

   i. the 80 people
   ii. the 4 groups
   iii. the 3 antihistamines and placebo
   iv. the time to apply the brakes

   (c) What was the response variable? (Circle the one best answer.)

   i. the 80 people

ii. the 4 groups
iii. the 3 antihistamines and placebo
iv. the time to apply the brakes

(d) Which statistical test would be most appropriate for determining whether there were any differences among the effects on alertness of the three antihistamines and placebo? (Circle the one best answer.)

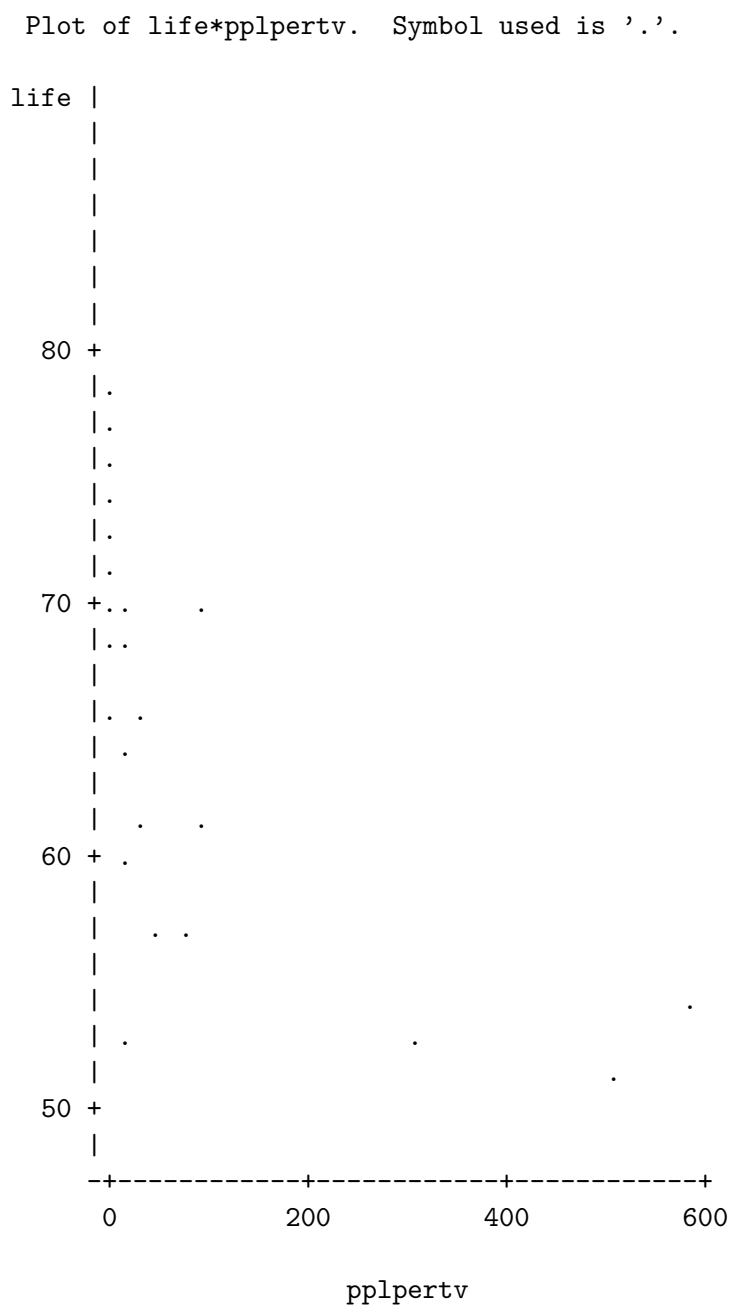
i. ANOVA
ii. Chi square test
iii. paired t-test
iv. two-independent-sample t-test

4. Researchers wish to estimate the mean head circumference in the population of male history professors. They believe that the population is normal with standard deviation equal to 1.5 inches.

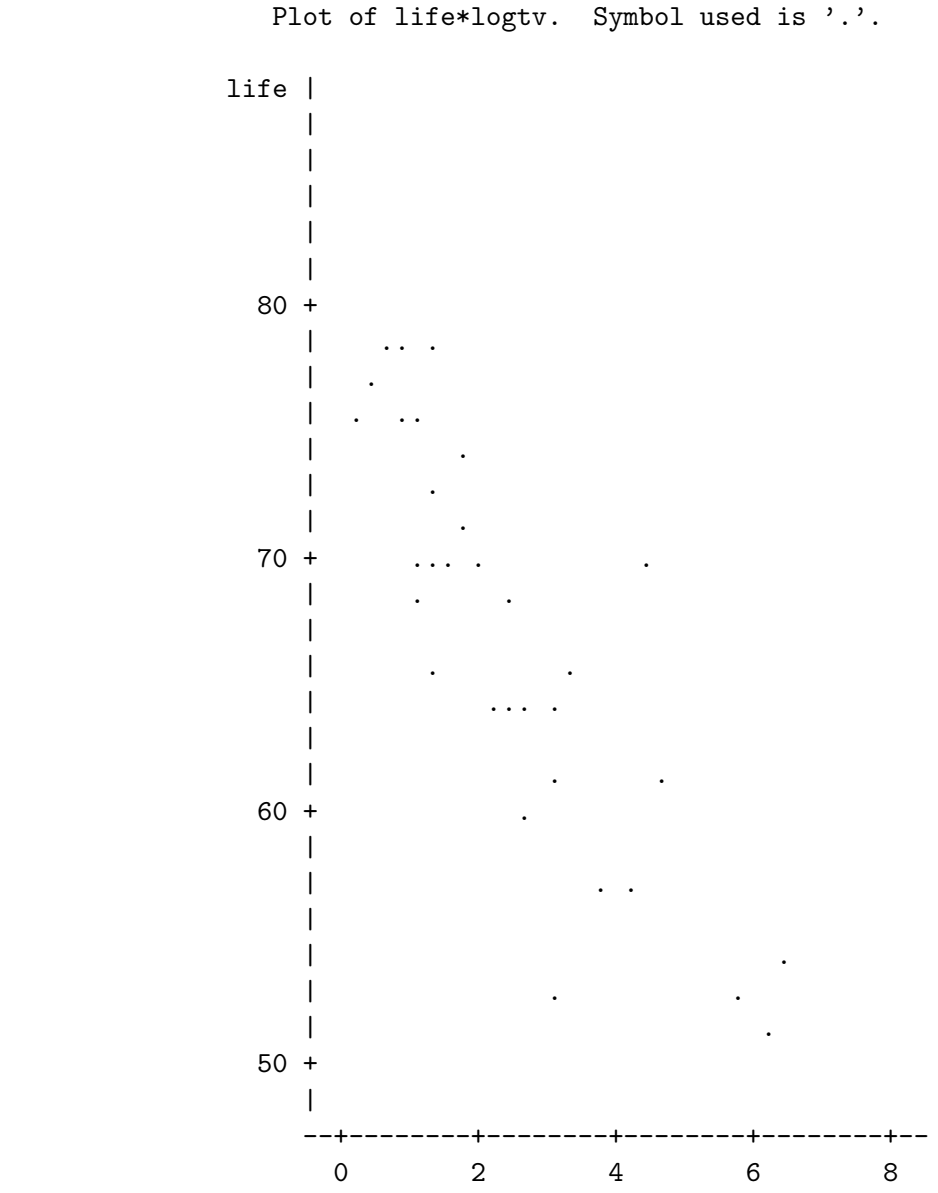
(a) How large a simple random sample of male history professors will they need in order to get a 90% confidence interval of width no greater than 1 inch? (Numeric answer; show your work.)

(b) Do they need to make sure that that their sample is no greater than 1/10 the size of the population? Why or why not?

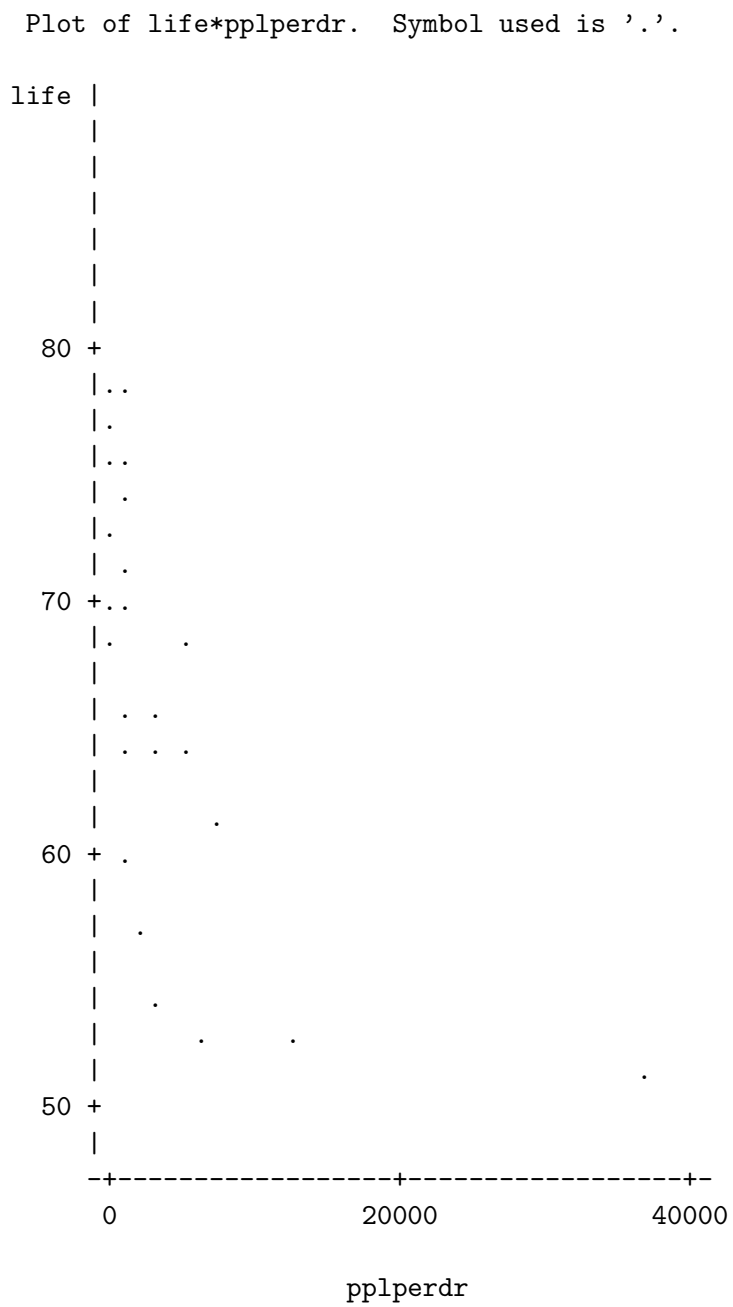
Obs	Country	life	pplpertv	pplperdr	female	male	logtv	logdr
1	Argentina	70.5	4.0	370	74	67	1.38629	5.9135
2	Bangladesh	53.5	315.0	6166	53	54	5.75257	8.7268
3	Brazil	65.0	4.0	684	68	62	1.38629	6.5280
4	Canada	76.5	1.7	449	80	73	0.53063	6.1070
5	China	70.0	8.0	643	72	68	2.07944	6.4661
6	Colombia	71.0	5.6	1551	74	68	1.72277	7.3467
7	Egypt	60.5	15.0	616	61	60	2.70805	6.4232
8	Ethiopia	51.5	503.0	36660	53	50	6.22059	10.5094
9	France	78.0	2.6	403	82	74	0.95551	5.9989
10	Germany	76.0	2.6	346	79	73	0.95551	5.8464
11	India	57.5	44.0	2471	58	57	3.78419	7.8124
12	Indonesia	61.0	24.0	7427	63	59	3.17805	8.9129
13	Iran	64.5	23.0	2992	65	64	3.13549	8.0037
14	Italy	78.5	3.8	233	82	75	1.33500	5.4510
15	Japan	79.0	1.8	609	82	76	0.58779	6.4118
16	Kenya	61.0	96.0	7615	63	59	4.56435	8.9379
17	Korea, North	70.0	90.0	370	73	67	4.49981	5.9135
18	Korea, South	70.0	4.9	1066	73	67	1.58924	6.9717
19	Mexico	72.0	6.6	600	76	68	1.88707	6.3969
20	Morocco	64.5	21.0	4873	66	63	3.04452	8.4915
21	Myanmar (Burma)	54.5	592.0	3485	56	53	6.38351	8.1562
22	Pakistan	56.5	73.0	2364	57	56	4.29046	7.7681
23	Peru	64.5	14.0	1016	67	62	2.63906	6.9236
24	Philippines	64.5	8.8	1062	67	62	2.17475	6.9679
25	Poland	73.0	3.9	480	77	69	1.36098	6.1738
26	Romania	72.0	6.0	559	75	69	1.79176	6.3261
27	Russia	69.0	3.2	259	74	64	1.16315	5.5568
28	South Africa	64.0	11.0	1340	67	61	2.39790	7.2004
29	Spain	78.5	2.6	275	82	75	0.95551	5.6168
30	Sudan	53.0	23.0	12550	54	52	3.13549	9.4375
31	Taiwan	75.0	3.2	965	78	72	1.16315	6.8721
32	Tanzania	52.5	.	.	55	50	.	.
33	Thailand	68.5	11.0	4883	71	66	2.39790	8.4935
34	Turkey	70.0	5.0	1189	72	68	1.60944	7.0809
35	Ukraine	70.5	3.0	226	75	66	1.09861	5.4205
36	United Kingdom	76.0	3.0	611	79	73	1.09861	6.4151
37	United States	75.5	1.3	404	79	72	0.26236	6.0014
38	Venezuela	74.5	5.6	576	78	71	1.72277	6.3561
39	Vietnam	65.0	29.0	3096	67	63	3.36730	8.0379
40	Zaire	54.0	.	.	56	52	.	.



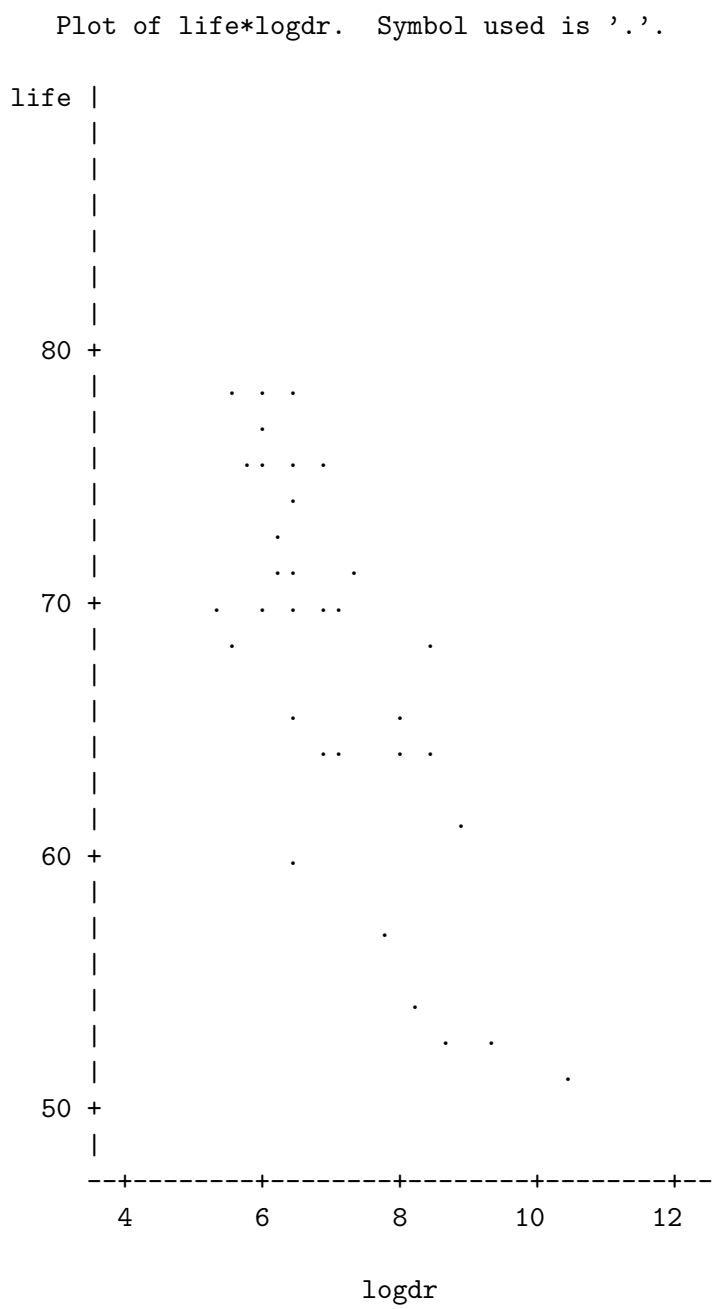
NOTE: 2 obs had missing values. 15 obs hidden.



NOTE: 2 obs had missing values. 6 obs hidden.



NOTE: 2 obs had missing values. 14 obs hidden.



NOTE: 2 obs had missing values. 5 obs hidden.

The REG Procedure
Model: MODEL1
Dependent Variable: life

Number of Observations Read	40
Number of Observations Used	38
Number of Observations with Missing Values	2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1646.97245	1646.97245	97.94	<.0001
Error	36	605.39597	16.81655		
Corrected Total	37	2252.36842			

Root MSE	4.10080	R-Square	0.7312
Dependent Mean	67.76316	Adj R-Sq	0.7238
Coeff Var	6.05166		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	77.88728	1.22029	63.83	<.0001
logtv	1	-4.25968	0.43043	-9.90	<.0001

Parameter Estimates

Variable	DF	95% Confidence Limits	
Intercept	1	75.41242	80.36214
logtv	1	-5.13263	-3.38673

Output Statistics

Obs	Country	logtv	life	Dependent Variable	Predicted Value	Residual
1	Argentina	1.386	70.5	70.5	71.9821	-1.4821
2	Bangladesh	5.753	53.5	53.5	53.3832	0.1168
3	Brazil	1.386	65.0	65.0	71.9821	-6.9821
4	Canada	0.531	76.5	76.5	75.6270	0.8730
5	China	2.079	70.0	70.0	69.0295	0.9705
6	Colombia	1.723	71.0	71.0	70.5488	0.4512
7	Egypt	2.708	60.5	60.5	66.3519	-5.8519
8	Ethiopia	6.221	51.5	51.5	51.3896	0.1104
9	France	0.956	78.0	78.0	73.8171	4.1829
10	Germany	0.956	76.0	76.0	73.8171	2.1829
11	India	3.784	57.5	57.5	61.7678	-4.2678
12	Indonesia	3.178	61.0	61.0	64.3498	-3.3498
13	Iran	3.135	64.5	64.5	64.5311	-0.0311
14	Italy	1.335	78.5	78.5	72.2006	6.2994
15	Japan	0.588	79.0	79.0	75.3835	3.6165
16	Kenya	4.564	61.0	61.0	58.4446	2.5554
17	Korea, North	4.500	70.0	70.0	58.7195	11.2805
18	Korea, South	1.589	70.0	70.0	71.1176	-1.1176
19	Mexico	1.887	72.0	72.0	69.8490	2.1510
20	Morocco	3.045	64.5	64.5	64.9186	-0.4186
21	Myanmar (Burma)	6.384	54.5	54.5	50.6956	3.8044
22	Pakistan	4.290	56.5	56.5	59.6113	-3.1113
23	Peru	2.639	64.5	64.5	66.6457	-2.1457
24	Philippines	2.175	64.5	64.5	68.6235	-4.1235
25	Poland	1.361	73.0	73.0	72.0900	0.9100
26	Romania	1.792	72.0	72.0	70.2550	1.7450
27	Russia	1.163	69.0	69.0	72.9326	-3.9326
28	South Africa	2.398	64.0	64.0	67.6730	-3.6730
29	Spain	0.956	78.5	78.5	73.8171	4.6829
30	Sudan	3.135	53.0	53.0	64.5311	-11.5311
31	Taiwan	1.163	75.0	75.0	72.9326	2.0674
32	Tanzania	.	52.5	52.5	.	.
33	Thailand	2.398	68.5	68.5	67.6730	0.8270
34	Turkey	1.609	70.0	70.0	71.0316	-1.0316
35	Ukraine	1.099	70.5	70.5	73.2075	-2.7075
36	United Kingdom	1.099	76.0	76.0	73.2075	2.7925
37	United States	0.262	75.5	75.5	76.7697	-1.2697
38	Venezuela	1.723	74.5	74.5	70.5488	3.9512
39	Vietnam	3.367	65.0	65.0	63.5437	1.4563
40	Zaire	.	54.0	54.0	.	.

The REG Procedure
Model: MODEL1
Dependent Variable: life

Number of Observations Read	40
Number of Observations Used	38
Number of Observations with Missing Values	2

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1438.07027	1438.07027	63.58	<.0001
Error	36	814.29815	22.61939		
Corrected Total	37	2252.36842			

Root MSE	4.75598	R-Square	0.6385
Dependent Mean	67.76316	Adj R-Sq	0.6284
Coeff Var	7.01854		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	103.28082	4.52078	22.85	<.0001
logdr	1	-5.03657	0.63166	-7.97	<.0001

Parameter Estimates

Variable	DF	95% Confidence Limits	
Intercept	1	94.11226	112.44938
logdr	1	-6.31764	-3.75550

Output Statistics

Obs	Country	logdr	life	Dependent Variable	Predicted Value	Residual
1	Argentina	5.91	70.5	70.5	73.4970	-2.9970
2	Bangladesh	8.73	53.5	53.5	59.3276	-5.8276
3	Brazil	6.53	65.0	65.0	70.4023	-5.4023
4	Canada	6.11	76.5	76.5	72.5224	3.9776
5	China	6.47	70.0	70.0	70.7136	-0.7136
6	Colombia	7.35	71.0	71.0	66.2789	4.7211
7	Egypt	6.42	60.5	60.5	70.9297	-10.4297
8	Ethiopia	10.51	51.5	51.5	50.3493	1.1507
9	France	6.00	78.0	78.0	73.0668	4.9332
10	Germany	5.85	76.0	76.0	73.8348	2.1652
11	India	7.81	57.5	57.5	63.9332	-6.4332
12	Indonesia	8.91	61.0	61.0	58.3905	2.6095
13	Iran	8.00	64.5	64.5	62.9696	1.5304
14	Italy	5.45	78.5	78.5	75.8263	2.6737
15	Japan	6.41	79.0	79.0	70.9872	8.0128
16	Kenya	8.94	61.0	61.0	58.2646	2.7354
17	Korea, North	5.91	70.0	70.0	73.4970	-3.4970
18	Korea, South	6.97	70.0	70.0	68.1675	1.8325
19	Mexico	6.40	72.0	72.0	71.0622	0.9378
20	Morocco	8.49	64.5	64.5	60.5130	3.9870
21	Myanmar (Burma)	8.16	54.5	54.5	62.2014	-7.7014
22	Pakistan	7.77	56.5	56.5	64.1562	-7.6562
23	Peru	6.92	64.5	64.5	68.4095	-3.9095
24	Philippines	6.97	64.5	64.5	68.1865	-3.6865
25	Poland	6.17	73.0	73.0	72.1861	0.8139
26	Romania	6.33	72.0	72.0	71.4187	0.5813
27	Russia	5.56	69.0	69.0	75.2935	-6.2935
28	South Africa	7.20	64.0	64.0	67.0154	-3.0154
29	Spain	5.62	78.5	78.5	74.9916	3.5084
30	Sudan	9.44	53.0	53.0	55.7483	-2.7483
31	Taiwan	6.87	75.0	75.0	68.6689	6.3311
32	Tanzania	.	52.5	52.5	.	.
33	Thailand	8.49	68.5	68.5	60.5026	7.9974
34	Turkey	7.08	70.0	70.0	67.6175	2.3825
35	Ukraine	5.42	70.5	70.5	75.9799	-5.4799
36	United Kingdom	6.42	76.0	76.0	70.9707	5.0293
37	United States	6.00	75.5	75.5	73.0543	2.4457
38	Venezuela	6.36	74.5	74.5	71.2678	3.2322
39	Vietnam	8.04	65.0	65.0	62.7975	2.2025
40	Zaire	.	54.0	54.0	.	.