Name: Bhargav Gopalratnam

UID: 904562676

Course- Stats 202-A

> #Preliminary Examination of the Data

> state.x77 # output not shown

Population Income Illiteracy Life Exp Murder HS Grad

Alabama 3615 3624 2.1 69.05 15.1 41.3

Alaska 365 6315 1.5 69.31 11.3 66.7

Arizona 2212 4530 1.8 70.55 7.8 58.1

Arkansas 2110 3378 1.9 70.66 10.1 39.9

California 21198 5114 1.1 71.71 10.3 62.6

Colorado 2541 4884 0.7 72.06 6.8 63.9

Connecticut 3100 5348 1.1 72.48 3.1 56.0

Delaware 579 4809 0.9 70.06 6.2 54.6

Florida 8277 4815 1.3 70.66 10.7 52.6

Georgia 4931 4091 2.0 68.54 13.9 40.6

Hawaii 868 4963 1.9 73.60 6.2 61.9

Idaho 813 4119 0.6 71.87 5.3 59.5

Illinois 11197 5107 0.9 70.14 10.3 52.6

Indiana 5313 4458 0.7 70.88 7.1 52.9

Iowa 2861 4628 0.5 72.56 2.3 59.0

Kansas 2280 4669 0.6 72.58 4.5 59.9

Kentucky 3387 3712 1.6 70.10 10.6 38.5

Louisiana 3806 3545 2.8 68.76 13.2 42.2

Maine 1058 3694 0.7 70.39 2.7 54.7

Maryland 4122 5299 0.9 70.22 8.5 52.3

Massachusetts 5814 4755 1.1 71.83 3.3 58.5

Michigan 9111 4751 0.9 70.63 11.1 52.8

Minnesota 3921 4675 0.6 72.96 2.3 57.6

Mississippi 2341 3098 2.4 68.09 12.5 41.0

Missouri 4767 4254 0.8 70.69 9.3 48.8

Montana 746 4347 0.6 70.56 5.0 59.2

Nebraska 1544 4508 0.6 72.60 2.9 59.3

Nevada 590 5149 0.5 69.03 11.5 65.2

New Hampshire 812 4281 0.7 71.23 3.3 57.6

New Jersey 7333 5237 1.1 70.93 5.2 52.5

New Mexico 1144 3601 2.2 70.32 9.7 55.2

New York 18076 4903 1.4 70.55 10.9 52.7

North Carolina 5441 3875 1.8 69.21 11.1 38.5

North Dakota 637 5087 0.8 72.78 1.4 50.3

Ohio 10735 4561 0.8 70.82 7.4 53.2

Oklahoma 2715 3983 1.1 71.42 6.4 51.6

Oregon 2284 4660 0.6 72.13 4.2 60.0

Pennsylvania 11860 4449 1.0 70.43 6.1 50.2

Rhode Island 931 4558 1.3 71.90 2.4 46.4

South Carolina 2816 3635 2.3 67.96 11.6 37.8

South Dakota 681 4167 0.5 72.08 1.7 53.3

Tennessee 4173 3821 1.7 70.11 11.0 41.8

Texas 12237 4188 2.2 70.90 12.2 47.4

Utah 1203 4022 0.6 72.90 4.5 67.3

Vermont 472 3907 0.6 71.64 5.5 57.1

Virginia 4981 4701 1.4 70.08 9.5 47.8

Washington 3559 4864 0.6 71.72 4.3 63.5

West Virginia 1799 3617 1.4 69.48 6.7 41.6

Wisconsin 4589 4468 0.7 72.48 3.0 54.5

Wyoming 376 4566 0.6 70.29 6.9 62.9

Frost Area

Alabama 20 50708

Alaska 152 566432

Arizona 15 113417

Arkansas 65 51945

California 20 156361

Colorado 166 103766

Connecticut 139 4862

Delaware 103 1982

Florida 11 54090

Georgia 60 58073

Hawaii 0 6425

Idaho 126 82677

Illinois 127 55748

Indiana 122 36097

Iowa 140 55941

Kansas 114 81787

Kentucky 95 39650

Louisiana 12 44930

Maine 161 30920

Maryland 101 9891

Massachusetts 103 7826

Michigan 125 56817

Minnesota 160 79289

Mississippi 50 47296

Missouri 108 68995

Montana 155 145587

Nebraska 139 76483

Nevada 188 109889

New Hampshire 174 9027

New Jersey 115 7521

New Mexico 120 121412

New York 82 47831

North Carolina 80 48798

North Dakota 186 69273

Ohio 124 40975

Oklahoma 82 68782

Oregon 44 96184

Pennsylvania 126 44966

Rhode Island 127 1049

South Carolina 65 30225

South Dakota 172 75955

Tennessee 70 41328

Texas 35 262134

Utah 137 82096

Vermont 168 9267

Virginia 85 39780

Washington 32 66570

West Virginia 100 24070

Wisconsin 149 54464

Wyoming 173 97203

> str(state.x77) # clearly not a data frame!

num [1:50, 1:8] 3615 365 2212 2110 21198 ...

- attr(\*, "dimnames")=List of 2

..$ : chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...

..$ : chr [1:8] "Population" "Income" "Illiteracy" "Life Exp" ...

> st = as.data.frame(state.x77) # so we'll make it one

> str(st)

'data.frame': 50 obs. of 8 variables:

$ Population: num 3615 365 2212 2110 21198 ...

$ Income : num 3624 6315 4530 3378 5114 ...

$ Illiteracy: num 2.1 1.5 1.8 1.9 1.1 0.7 1.1 0.9 1.3 2 ...

$ Life Exp : num 69 69.3 70.5 70.7 71.7 ...

$ Murder : num 15.1 11.3 7.8 10.1 10.3 6.8 3.1 6.2 10.7 13.9 ...

$ HS Grad : num 41.3 66.7 58.1 39.9 62.6 63.9 56 54.6 52.6 40.6 ...

$ Frost : num 20 152 15 65 20 166 139 103 11 60 ...

$ Area : num 50708 566432 113417 51945 156361 ...

> colnames(st)[4] = "Life.Exp" # no spaces in variable names, please

> colnames(st)[6] = "HS.Grad"

> st[,9] = st$Population \* 1000 / st$Area

> colnames(st)[9] = "Density" # create and name a new column

> str(st)

'data.frame': 50 obs. of 9 variables:

$ Population: num 3615 365 2212 2110 21198 ...

$ Income : num 3624 6315 4530 3378 5114 ...

$ Illiteracy: num 2.1 1.5 1.8 1.9 1.1 0.7 1.1 0.9 1.3 2 ...

$ Life.Exp : num 69 69.3 70.5 70.7 71.7 ...

$ Murder : num 15.1 11.3 7.8 10.1 10.3 6.8 3.1 6.2 10.7 13.9 ...

$ HS.Grad : num 41.3 66.7 58.1 39.9 62.6 63.9 56 54.6 52.6 40.6 ...

$ Frost : num 20 152 15 65 20 166 139 103 11 60 ...

$ Area : num 50708 566432 113417 51945 156361 ...

$ Density : num 71.291 0.644 19.503 40.62 135.571 ...

> summary(st)

Population Income Illiteracy Life.Exp

Min. : 365 Min. :3098 Min. :0.500 Min. :67.96

1st Qu.: 1080 1st Qu.:3993 1st Qu.:0.625 1st Qu.:70.12

Median : 2838 Median :4519 Median :0.950 Median :70.67

Mean : 4246 Mean :4436 Mean :1.170 Mean :70.88

3rd Qu.: 4968 3rd Qu.:4814 3rd Qu.:1.575 3rd Qu.:71.89

Max. :21198 Max. :6315 Max. :2.800 Max. :73.60

Murder HS.Grad Frost Area

Min. : 1.400 Min. :37.80 Min. : 0.00 Min. : 1049

1st Qu.: 4.350 1st Qu.:48.05 1st Qu.: 66.25 1st Qu.: 36985

Median : 6.850 Median :53.25 Median :114.50 Median : 54277

Mean : 7.378 Mean :53.11 Mean :104.46 Mean : 70736

3rd Qu.:10.675 3rd Qu.:59.15 3rd Qu.:139.75 3rd Qu.: 81162

Max. :15.100 Max. :67.30 Max. :188.00 Max. :566432

Density

Min. : 0.6444

1st Qu.: 25.3352

Median : 73.0154

Mean :149.2245

3rd Qu.:144.2828

Max. :975.0033

> cor(st) # correlation matrix

Population Income Illiteracy Life.Exp Murder

Population 1.00000000 0.2082276 0.107622373 -0.06805195 0.3436428

Income 0.20822756 1.0000000 -0.437075186 0.34025534 -0.2300776

Illiteracy 0.10762237 -0.4370752 1.000000000 -0.58847793 0.7029752

Life.Exp -0.06805195 0.3402553 -0.588477926 1.00000000 -0.7808458

Murder 0.34364275 -0.2300776 0.702975199 -0.78084575 1.0000000

HS.Grad -0.09848975 0.6199323 -0.657188609 0.58221620 -0.4879710

Frost -0.33215245 0.2262822 -0.671946968 0.26206801 -0.5388834

Area 0.02254384 0.3633154 0.077261132 -0.10733194 0.2283902

Density 0.24622789 0.3299683 0.009274348 0.09106176 -0.1850352

HS.Grad Frost Area Density

Population -0.09848975 -0.332152454 0.02254384 0.246227888

Income 0.61993232 0.226282179 0.36331544 0.329968277

Illiteracy -0.65718861 -0.671946968 0.07726113 0.009274348

Life.Exp 0.58221620 0.262068011 -0.10733194 0.091061763

Murder -0.48797102 -0.538883437 0.22839021 -0.185035233

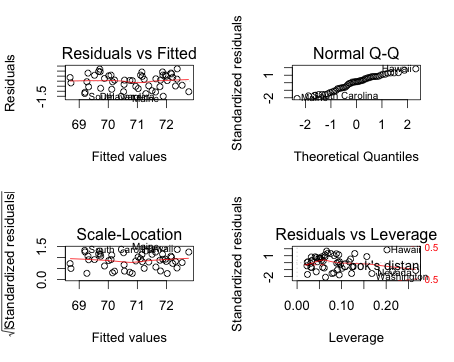
HS.Grad 1.00000000 0.366779702 0.33354187 -0.088367214

Frost 0.36677970 1.000000000 0.05922910 0.002276734

Area 0.33354187 0.059229102 1.00000000 -0.341388515

Density -0.08836721 0.002276734 -0.34138851 1.000000000

> pairs(st) # scatterplot matrix



> #The Maximal Model (Sans Interactions)

> options(show.signif.stars=F) # I don't like significance stars!

> names(st) # for handy reference

[1] "Population" "Income" "Illiteracy" "Life.Exp" "Murder"

[6] "HS.Grad" "Frost" "Area" "Density"

> model1 = lm(Life.Exp ~ Population + Income + Illiteracy + Murder +

+ + HS.Grad + Frost + Area + Density, data=st)

> summary(model1)

Call:

lm(formula = Life.Exp ~ Population + Income + Illiteracy + Murder +

+HS.Grad + Frost + Area + Density, data = st)

Residuals:

Min 1Q Median 3Q Max

-1.47514 -0.45887 -0.06352 0.59362 1.21823

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.995e+01 1.843e+00 37.956 < 2e-16

Population 6.480e-05 3.001e-05 2.159 0.0367

Income 2.701e-04 3.087e-04 0.875 0.3867

Illiteracy 3.029e-01 4.024e-01 0.753 0.4559

Murder -3.286e-01 4.941e-02 -6.652 5.12e-08

HS.Grad 4.291e-02 2.332e-02 1.840 0.0730

Frost -4.580e-03 3.189e-03 -1.436 0.1585

Area -1.558e-06 1.914e-06 -0.814 0.4205

Density -1.105e-03 7.312e-04 -1.511 0.1385

Residual standard error: 0.7337 on 41 degrees of freedom

Multiple R-squared: 0.7501, Adjusted R-squared: 0.7013

F-statistic: 15.38 on 8 and 41 DF, p-value: 3.787e-10

> lm(formula = Life.Exp ~ Population + Income + Illiteracy + Murder +

+ HS.Grad + Frost + Area + Density, data = st)

Call:

lm(formula = Life.Exp ~ Population + Income + Illiteracy + Murder +

HS.Grad + Frost + Area + Density, data = st)

Coefficients:

(Intercept) Population Income Illiteracy Murder

6.995e+01 6.480e-05 2.701e-04 3.029e-01 -3.286e-01

HS.Grad Frost Area Density

4.291e-02 -4.580e-03 -1.558e-06 -1.105e-03

> summary.aov(model1)

Df Sum Sq Mean Sq F value Pr(>F)

Population 1 0.409 0.409 0.760 0.38849

Income 1 11.595 11.595 21.541 3.53e-05

Illiteracy 1 19.421 19.421 36.081 4.23e-07

Murder 1 27.429 27.429 50.959 1.05e-08

HS.Grad 1 4.099 4.099 7.615 0.00861

Frost 1 2.049 2.049 3.806 0.05792

Area 1 0.001 0.001 0.002 0.96438

Density 1 1.229 1.229 2.283 0.13847

Residuals 41 22.068 0.538

> #The Minimal Adequate Model

> model2 = update(model1, .~.-Area)

> summary(model2)

Call:

lm(formula = Life.Exp ~ Population + Income + Illiteracy + Murder +

HS.Grad + Frost + Density, data = st)

Residuals:

Min 1Q Median 3Q Max

-1.50252 -0.40471 -0.06079 0.58682 1.43862

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.094e+01 1.378e+00 51.488 < 2e-16

Population 6.249e-05 2.976e-05 2.100 0.0418

Income 1.485e-04 2.690e-04 0.552 0.5840

Illiteracy 1.452e-01 3.512e-01 0.413 0.6814

Murder -3.319e-01 4.904e-02 -6.768 3.12e-08

HS.Grad 3.746e-02 2.225e-02 1.684 0.0996

Frost -5.533e-03 2.955e-03 -1.873 0.0681

Density -7.995e-04 6.251e-04 -1.279 0.2079

Residual standard error: 0.7307 on 42 degrees of freedom

Multiple R-squared: 0.746, Adjusted R-squared: 0.7037

F-statistic: 17.63 on 7 and 42 DF, p-value: 1.173e-10

> lm(formula = Life.Exp ~ Population + Income + Illiteracy + Murder +

+ HS.Grad + Frost + Density, data = st)

Call:

lm(formula = Life.Exp ~ Population + Income + Illiteracy + Murder +

HS.Grad + Frost + Density, data = st)

Coefficients:

(Intercept) Population Income Illiteracy Murder

7.094e+01 6.249e-05 1.485e-04 1.452e-01 -3.319e-01

HS.Grad Frost Density

3.746e-02 -5.533e-03 -7.995e-04

> anova(model1, model2)

Analysis of Variance Table

Model 1: Life.Exp ~ Population + Income + Illiteracy + Murder + +HS.Grad +

Frost + Area + Density

Model 2: Life.Exp ~ Population + Income + Illiteracy + Murder + HS.Grad +

Frost + Density

Res.Df RSS Df Sum of Sq F Pr(>F)

1 41 22.068

2 42 22.425 -1 -0.35639 0.6621 0.4205

> model3 = update(model2, .~.-Illiteracy)

> summary(model3)

Call:

lm(formula = Life.Exp ~ Population + Income + Murder + HS.Grad +

Frost + Density, data = st)

Residuals:

Min 1Q Median 3Q Max

-1.49555 -0.41246 -0.05336 0.58399 1.50535

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.131e+01 1.042e+00 68.420 < 2e-16

Population 5.811e-05 2.753e-05 2.110 0.0407

Income 1.324e-04 2.636e-04 0.502 0.6181

Murder -3.208e-01 4.054e-02 -7.912 6.32e-10

HS.Grad 3.499e-02 2.122e-02 1.649 0.1065

Frost -6.191e-03 2.465e-03 -2.512 0.0158

Density -7.324e-04 5.978e-04 -1.225 0.2272

Residual standard error: 0.7236 on 43 degrees of freedom

Multiple R-squared: 0.745, Adjusted R-squared: 0.7094

F-statistic: 20.94 on 6 and 43 DF, p-value: 2.632e-11

> lm(formula = Life.Exp ~ Population + Income + Murder + HS.Grad +

+ Frost + Density, data = st)

Call:

lm(formula = Life.Exp ~ Population + Income + Murder + HS.Grad +

Frost + Density, data = st)

Coefficients:

(Intercept) Population Income Murder HS.Grad

7.131e+01 5.811e-05 1.324e-04 -3.208e-01 3.499e-02

Frost Density

-6.191e-03 -7.324e-04

> model4 = update(model3, .~.-Income)

> summary(model4)

Call:

lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost +

Density, data = st)

Residuals:

Min 1Q Median 3Q Max

-1.56877 -0.40951 -0.04554 0.57362 1.54752

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.142e+01 1.011e+00 70.665 < 2e-16

Population 6.083e-05 2.676e-05 2.273 0.02796

Murder -3.160e-01 3.910e-02 -8.083 3.07e-10

HS.Grad 4.233e-02 1.525e-02 2.776 0.00805

Frost -5.999e-03 2.414e-03 -2.485 0.01682

Density -5.864e-04 5.178e-04 -1.132 0.26360

Residual standard error: 0.7174 on 44 degrees of freedom

Multiple R-squared: 0.7435, Adjusted R-squared: 0.7144

F-statistic: 25.51 on 5 and 44 DF, p-value: 5.524e-12

> lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost +

+ Density, data = st)

Call:

lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost +

Density, data = st)

Coefficients:

(Intercept) Population Murder HS.Grad Frost

7.142e+01 6.083e-05 -3.160e-01 4.233e-02 -5.999e-03

Density

-5.864e-04

> model5 = update(model4, .~.-Density)

> summary(model5)

Call:

lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,

data = st)

Residuals:

Min 1Q Median 3Q Max

-1.47095 -0.53464 -0.03701 0.57621 1.50683

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.103e+01 9.529e-01 74.542 < 2e-16

Population 5.014e-05 2.512e-05 1.996 0.05201

Murder -3.001e-01 3.661e-02 -8.199 1.77e-10

HS.Grad 4.658e-02 1.483e-02 3.142 0.00297

Frost -5.943e-03 2.421e-03 -2.455 0.01802

Residual standard error: 0.7197 on 45 degrees of freedom

Multiple R-squared: 0.736, Adjusted R-squared: 0.7126

F-statistic: 31.37 on 4 and 45 DF, p-value: 1.696e-12

> lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,

+ data = st)

Call:

lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,

data = st)

Coefficients:

(Intercept) Population Murder HS.Grad Frost

7.103e+01 5.014e-05 -3.001e-01 4.658e-02 -5.943e-03

> anova(model5, model4)

Analysis of Variance Table

Model 1: Life.Exp ~ Population + Murder + HS.Grad + Frost

Model 2: Life.Exp ~ Population + Murder + HS.Grad + Frost + Density

Res.Df RSS Df Sum of Sq F Pr(>F)

1 45 23.308

2 44 22.648 1 0.66005 1.2823 0.2636

> model6 = update(model5, .~.-Population)

> summary(model6)

Call:

lm(formula = Life.Exp ~ Murder + HS.Grad + Frost, data = st)

Residuals:

Min 1Q Median 3Q Max

-1.5015 -0.5391 0.1014 0.5921 1.2268

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 71.036379 0.983262 72.246 < 2e-16

Murder -0.283065 0.036731 -7.706 8.04e-10

HS.Grad 0.049949 0.015201 3.286 0.00195

Frost -0.006912 0.002447 -2.824 0.00699

Residual standard error: 0.7427 on 46 degrees of freedom

Multiple R-squared: 0.7127, Adjusted R-squared: 0.6939

F-statistic: 38.03 on 3 and 46 DF, p-value: 1.634e-12

> lm(formula = Life.Exp ~ Murder + HS.Grad + Frost, data = st)

Call:

lm(formula = Life.Exp ~ Murder + HS.Grad + Frost, data = st)

Coefficients:

(Intercept) Murder HS.Grad Frost

71.036379 -0.283065 0.049949 -0.006912

> #Stepwise Regression

> step(model1, direction="backward")

Start: AIC=-22.89

Life.Exp ~ Population + Income + Illiteracy + Murder + +HS.Grad +

Frost + Area + Density

Df Sum of Sq RSS AIC

- Illiteracy 1 0.3050 22.373 -24.208

- Area 1 0.3564 22.425 -24.093

- Income 1 0.4120 22.480 -23.969

<none> 22.068 -22.894

- Frost 1 1.1102 23.178 -22.440

- Density 1 1.2288 23.297 -22.185

- HS.Grad 1 1.8225 23.891 -20.926

- Population 1 2.5095 24.578 -19.509

- Murder 1 23.8173 45.886 11.707

Step: AIC=-24.21

Life.Exp ~ Population + Income + Murder + HS.Grad + Frost + Area +

Density

Df Sum of Sq RSS AIC

- Area 1 0.1427 22.516 -25.890

- Income 1 0.2316 22.605 -25.693

<none> 22.373 -24.208

- Density 1 0.9286 23.302 -24.174

- HS.Grad 1 1.5218 23.895 -22.918

- Population 1 2.2047 24.578 -21.509

- Frost 1 3.1324 25.506 -19.656

- Murder 1 26.7071 49.080 13.072

Step: AIC=-25.89

Life.Exp ~ Population + Income + Murder + HS.Grad + Frost + Density

Df Sum of Sq RSS AIC

- Income 1 0.132 22.648 -27.598

- Density 1 0.786 23.302 -26.174

<none> 22.516 -25.890

- HS.Grad 1 1.424 23.940 -24.824

- Population 1 2.332 24.848 -22.962

- Frost 1 3.304 25.820 -21.043

- Murder 1 32.779 55.295 17.033

Step: AIC=-27.6

Life.Exp ~ Population + Murder + HS.Grad + Frost + Density

Df Sum of Sq RSS AIC

- Density 1 0.660 23.308 -28.161

<none> 22.648 -27.598

- Population 1 2.659 25.307 -24.046

- Frost 1 3.179 25.827 -23.030

- HS.Grad 1 3.966 26.614 -21.529

- Murder 1 33.626 56.274 15.910

Step: AIC=-28.16

Life.Exp ~ Population + Murder + HS.Grad + Frost

Df Sum of Sq RSS AIC

<none> 23.308 -28.161

- Population 1 2.064 25.372 -25.920

- Frost 1 3.122 26.430 -23.877

- HS.Grad 1 5.112 28.420 -20.246

- Murder 1 34.816 58.124 15.528

Call:

lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,

data = st)

Coefficients:

(Intercept) Population Murder HS.Grad Frost

7.103e+01 5.014e-05 -3.001e-01 4.658e-02 -5.943e-03

> lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost, data = st)

Call:

lm(formula = Life.Exp ~ Population + Murder + HS.Grad + Frost,

data = st)

Coefficients:

(Intercept) Population Murder HS.Grad Frost

7.103e+01 5.014e-05 -3.001e-01 4.658e-02 -5.943e-03

> #Confidence Limits on the Estimated Coefficients

> confint(model6)

2.5 % 97.5 %

(Intercept) 69.05717472 73.015582905

Murder -0.35700149 -0.209128849

HS.Grad 0.01935043 0.080546980

Frost -0.01183825 -0.001985219

> #Predictions

> predict(model6, list(Murder=10.5, HS.Grad=48, Frost=100))

1

69.77056

> #Regression Diagnostics

> par(mfrow=c(2,2)) # visualize four graphs at once

> plot(model6)

> par(mfrow=c(1,1)) # reset the graphics defaults

> #Extracting Elements of the Model Object

> model6[[1]] # extract list item 1: coefficients

(Intercept) Murder HS.Grad Frost

71.036378813 -0.283065168 0.049948704 -0.006911735

> model6[[2]]

Alabama Alaska Arizona Arkansas

0.36325842 -0.80873734 -1.07681421 0.93888883

California Colorado Connecticut Delaware

0.60063821 0.90409006 0.48472687 -1.23666537

Florida Georgia Hawaii Idaho

0.10114571 -0.17498630 1.22680042 0.23279723

Illinois Indiana Iowa Kansas

0.26968086 0.05432904 0.19534036 0.61342480

Kentucky Louisiana Maine Maryland

0.79770164 -0.56481311 -1.50150772 -0.32455693

Massachusetts Michigan Minnesota Mississippi

-0.48235430 0.96231978 0.80350324 -1.11037437

Missouri Montana Nebraska Nevada

0.59509781 -0.94669741 0.38328311 -0.70837880

New Hampshire New Jersey New Mexico New York

-0.54666731 -0.46189744 0.10159299 0.53349703

North Carolina North Dakota Ohio Oklahoma

-0.05444180 0.91307523 0.07808745 0.18464735

Oregon Pennsylvania Rhode Island South Carolina

-0.41031105 -0.51622769 0.10314800 -1.23162114

South Dakota Tennessee Texas Utah

0.05138438 0.58330361 1.19135836 0.72277428

Vermont Virginia Washington West Virginia

0.46958000 -0.06731035 -1.04976581 -1.04653483

Wisconsin Wyoming

0.60046076 -0.73927257

> sort(model6$resid) # extract residuals and sort them

Maine Delaware South Carolina Mississippi

-1.50150772 -1.23666537 -1.23162114 -1.11037437

Arizona Washington West Virginia Montana

-1.07681421 -1.04976581 -1.04653483 -0.94669741

Alaska Wyoming Nevada Louisiana

-0.80873734 -0.73927257 -0.70837880 -0.56481311

New Hampshire Pennsylvania Massachusetts New Jersey

-0.54666731 -0.51622769 -0.48235430 -0.46189744

Oregon Maryland Georgia Virginia

-0.41031105 -0.32455693 -0.17498630 -0.06731035

North Carolina South Dakota Indiana Ohio

-0.05444180 0.05138438 0.05432904 0.07808745

Florida New Mexico Rhode Island Oklahoma

0.10114571 0.10159299 0.10314800 0.18464735

Iowa Idaho Illinois Alabama

0.19534036 0.23279723 0.26968086 0.36325842

Nebraska Vermont Connecticut New York

0.38328311 0.46958000 0.48472687 0.53349703

Tennessee Missouri Wisconsin California

0.58330361 0.59509781 0.60046076 0.60063821

Kansas Utah Kentucky Minnesota

0.61342480 0.72277428 0.79770164 0.80350324

Colorado North Dakota Arkansas Michigan

0.90409006 0.91307523 0.93888883 0.96231978

Texas Hawaii

1.19135836 1.22680042

> #Beta Coeffieicents

> model7 = lm(scale(Life.Exp) ~ scale(Murder) + scale(HS.Grad) + scale(Frost), data=st)

> summary(model7)

Call:

lm(formula = scale(Life.Exp) ~ scale(Murder) + scale(HS.Grad) +

scale(Frost), data = st)

Residuals:

Min 1Q Median 3Q Max

-1.11853 -0.40156 0.07551 0.44111 0.91389

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -4.604e-15 7.824e-02 0.000 1.00000

scale(Murder) -7.784e-01 1.010e-01 -7.706 8.04e-10

scale(HS.Grad) 3.005e-01 9.146e-02 3.286 0.00195

scale(Frost) -2.676e-01 9.477e-02 -2.824 0.00699

Residual standard error: 0.5532 on 46 degrees of freedom

Multiple R-squared: 0.7127, Adjusted R-squared: 0.6939

F-statistic: 38.03 on 3 and 46 DF, p-value: 1.634e-12

> lm(formula = scale(Life.Exp) ~ scale(Murder) + scale(HS.Grad) +

+ scale(Frost), data = st)

Call:

lm(formula = scale(Life.Exp) ~ scale(Murder) + scale(HS.Grad) +

scale(Frost), data = st)

Coefficients:

(Intercept) scale(Murder) scale(HS.Grad) scale(Frost)

-4.604e-15 -7.784e-01 3.005e-01 -2.676e-01

> -0.283 \* sd(st$Murder) / sd(st$Life.Exp)

[1] -0.778241

> #Partial Correlations

> ### Partial correlation coefficient

> ### From formulas in Sheskin, 3e

> ### a,b=variables to be correlated, c=variable to be partialled out of both

> pcor = function(a,b,c)

+ {

+ (cor(a,b)-cor(a,c)\*cor(b,c))/sqrt((1-cor(a,c)^2)\*(1-cor(b,c)^2))

+ }

> ### end of script

> pcor(st$Life.Exp, st$Murder, st$HS.Grad)

[1] -0.6999659

> #Making Predictions From a Model

> rm(list=ls()) # clean up (WARNING! this will wipe your workspace!)

> data(airquality) # see ?airquality for details on these data

> na.omit(airquality) -> airquality # toss cases with missing values

> lm(Ozone ~ Solar.R + Wind + Temp + Month,

+ data=airquality) -> model

> coef(model)

(Intercept) Solar.R Wind Temp Month

-58.05383883 0.04959683 -3.31650940 1.87087379 -2.99162786

> (prediction <- c(1,200,11,80,6) \* coef(model))

(Intercept) Solar.R Wind Temp Month

-58.053839 9.919365 -36.481603 149.669903 -17.949767

> ### Note: putting the whole statement in parentheses not only stores the values but also prints them to the Console.

> sum(prediction)

[1] 47.10406

> ### Prediction of mean response for cases like this...

> predict(model, list(Solar.R=200,Wind=11,Temp=80,Month=6), interval="conf")

fit lwr upr

1 47.10406 41.10419 53.10393

> ### Prediction for a single new case...

> predict(model, list(Solar.R=200,Wind=11,Temp=80,Month=6), interval="pred")

fit lwr upr

1 47.10406 5.235759 88.97236