PLSC 503 – Spring 2017 Stupid Regression Tricks

January 31, 2017

```
> africa<-read.dta("africa2001.dta")
> attach(africa)

> summary(africa)
     ccode          cabbr            country           population           popthou
 Min.   :404.0   Length:43        Length:43        Min.   :   470000   Min.   :   470
 1st Qu.:451.5   Class :character Class :character 1st Qu.: 3446000   1st Qu.: 3446
 Median :510.0   Mode  :character Mode  :character Median : 9662000   Median : 9662
 Mean   :509.5                                     Mean   : 17390256   Mean   : 17390
 3rd Qu.:556.5                                     3rd Qu.: 19189000   3rd Qu.: 19189
 Max.   :651.0                                     Max.   :116929000   Max.   :116929
     popden            polity            gdppppd         tradegdp            war
 Min.   :0.002228   Min.   :-9.0000   Min.   : 0.500   Min.   :  4.030   Min.   :0.0000
 1st Qu.:0.013442   1st Qu.:-4.5000   1st Qu.: 0.855   1st Qu.:  7.645   1st Qu.:0.0000
 Median :0.035707   Median : 0.0000   Median : 1.200   Median : 13.560   Median :0.0000
 Mean   :0.064279   Mean   : 0.5116   Mean   : 2.159   Mean   : 30.493   Mean   :0.1163
 3rd Qu.:0.068302   3rd Qu.: 5.5000   3rd Qu.: 2.040   3rd Qu.: 30.010   3rd Qu.:0.0000
 Max.   :0.574020   Max.   :10.0000   Max.   :10.800   Max.   :272.690   Max.   :1.0000
     adrate           healthexp              subsaharan    muslperc         literacy
 Min.   : 0.100   Min.   :2.000   Not Sub-Saharan: 6   Min.   :  0.00   Min.   :17.00
 1st Qu.: 2.700   1st Qu.:3.450   Sub-Saharan    :37   1st Qu.: 10.00   1st Qu.:43.00
 Median : 6.000   Median :4.400                        Median : 20.00   Median :61.00
 Mean   : 9.365   Mean   :4.598                        Mean   : 35.96   Mean   :60.07
 3rd Qu.:12.900   3rd Qu.:5.800                        3rd Qu.: 55.50   3rd Qu.:78.50
 Max.   :38.800   Max.   :8.600                        Max.   :100.00   Max.   :89.00
   internalwar       intensity
 Min.   :0.0000   Min.   :0.0000
 1st Qu.:0.0000   1st Qu.:0.0000
 Median :0.0000   Median :0.0000
 Mean   :0.3023   Mean   :0.5814
 3rd Qu.:1.0000   3rd Qu.:1.0000
 Max.   :1.0000   Max.   :3.0000
```

A Simple Regression

```
> fit<-lm(adrate~muslperc)
> summary.lm(fit)

Call:
lm(formula = adrate ~ muslperc)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.2787     1.8322    8.34  2.3e-10 ***
muslperc     -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```
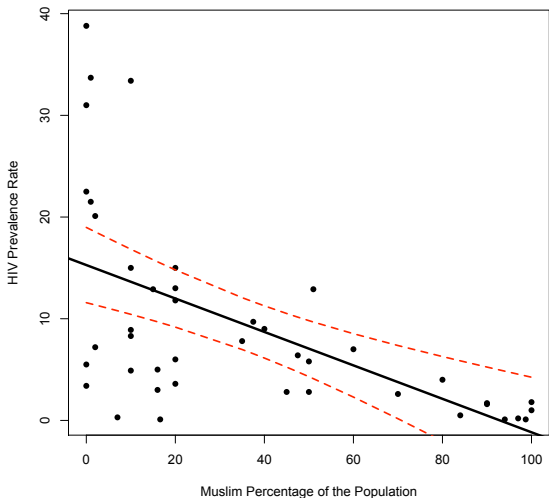
Scatterplot of HIV/AIDS Rates on Muslim Population Percentage, 2001

Adding a Constant to X

```
> africa$muslplusten<-muslperc+10
> fit2<-lm(adrate~muslplusten,data=africa)
> summary(fit2)

Call:
lm(formula = adrate ~ muslplusten, data = africa)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.9232     2.1152    8.00  6.6e-10 ***
muslplusten -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1    1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```
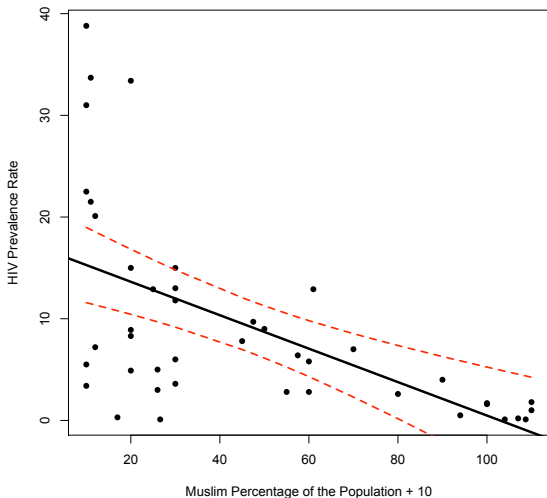
Scatterplot of HIV/AIDS Rates on Rescaled Muslim Population Percentage

HIV Prevalence Rate

Muslim Percentage of the Population + 10

Multiplying Y by a Constant

```
> africa$screwyrate<-adrate*(-314)
> fit3<-lm(screwyrate~muslperc,data=africa)
> summary(fit3)

Call:
lm(formula = screwyrate ~ muslperc, data = africa)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4797.5      575.3   -8.34  2.3e-10 ***
muslperc         51.6       11.6    4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2600 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```
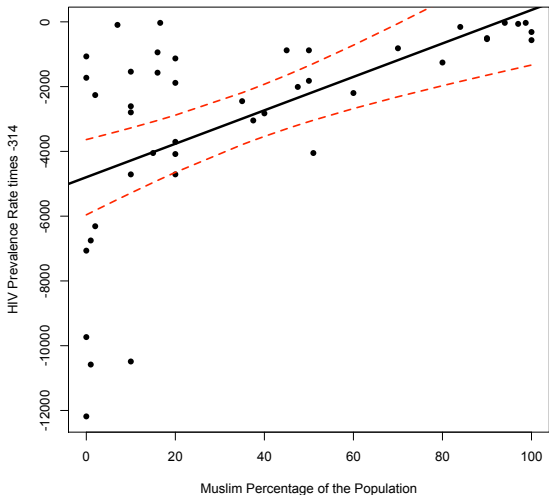
Scatterplot of Rescaled HIV/AIDS Rates on Muslim Population Percentage

Reversing the scales of X and Y

```
> africa$nonmuslimpct <- 100 - muslperc
> africa$noninfected <- 100 - adrate
> fit4<-lm(noninfected~nonmuslimpct,data=africa)
> summary(fit4)

Call:
lm(formula = noninfected ~ nonmuslimpct, data = africa)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 101.1660     2.6808   37.74  < 2e-16 ***
nonmuslimpct -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```
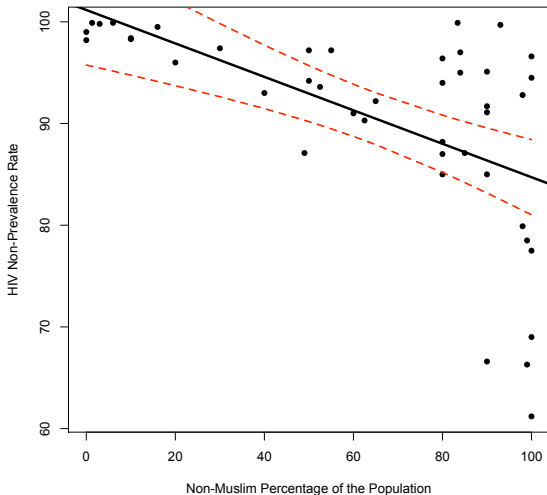
Scatterplot of HIV/AIDS Non-Infection Rates on Non-Muslim Population Percentage

Centering X

```
> africa$muslcenter<-muslperc - 35.96047
> fit5<-lm(adrate~muslcenter,data=africa)
> summary(fit5)

Call:
lm(formula = adrate ~ muslcenter, data = africa)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  9.3651     1.2622    7.42  4.2e-09 ***
muslcenter  -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```
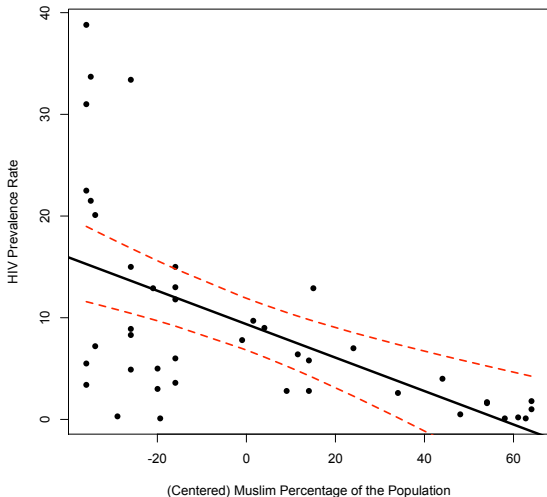
Scatterplot of HIV/AIDS Infection Rates on (Centered) Muslim Population Percentage

Rescaling X for Interpretability

```
> fit6<-lm(adrate~population,data=africa)
> summary(fit6)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.06e+01   1.91e+00    5.53   2e-06 ***
population  -7.05e-05   6.71e-05   -1.05     0.3
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.95 on 41 degrees of freedom
Multiple R-Squared: 0.0262,Adjusted R-squared: 0.00241
F-statistic: 1.1 on 1 and 41 DF,  p-value: 0.3

> africa$popmil<-africa$population / 1000000
> fit7<-lm(adrate~popmil,data=africa)
> summary(fit7)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.59047    1.91435    5.532   2e-06 ***
popmil      -0.07046    0.06714   -1.050     0.3
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.948 on 41 degrees of freedom
Multiple R-squared: 0.02616,Adjusted R-squared: 0.00241
F-statistic: 1.101 on 1 and 41 DF,  p-value: 0.3001
```

Dichotomous Xs: Bivariate Regression $\equiv$ t-test

```
> fit8<-lm(adrate~subsaharan,data=africa)
> summary(fit8)

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)              1.27       3.88    0.33     0.75
subsaharanSub-Saharan    9.41       4.19    2.25     0.03 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.51 on 41 degrees of freedom
Multiple R-Squared: 0.11,Adjusted R-squared: 0.088
F-statistic: 5.05 on 1 and 41 DF,  p-value: 0.03


> t.test(adrate~subsaharan,var.equal=TRUE)

Two Sample t-test

data:  adrate by subsaharan
t = -2.248, df = 41, p-value = 0.03
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -17.8659  -0.9576
sample estimates:
mean in group Not Sub-Saharan     mean in group Sub-Saharan
                      1.267                        10.678
```
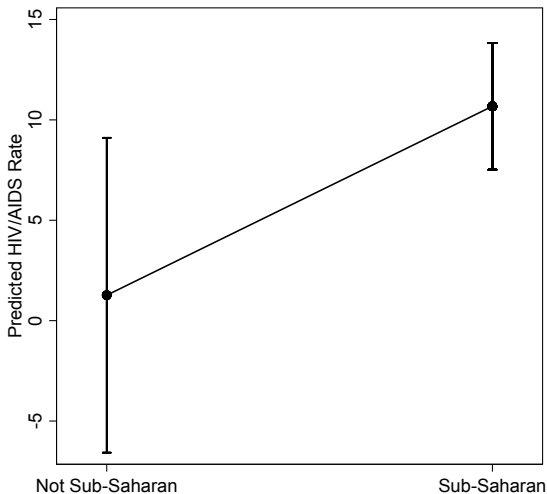
Expected Values of HIV/AIDS Infection Rates in Saharan and Sub-Saharan Africa

The results:

```
> fit<-lm(adrate~muslperc)
> summary.lm(fit)

Call:
lm(formula = adrate ~ muslperc)

Residuals:
    Min      1Q  Median      3Q     Max
-13.828  -5.206   0.279   2.022  23.521

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.2787     1.8322    8.34  2.3e-10 ***
muslperc    -0.1644     0.0369   -4.45  6.4e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.28 on 41 degrees of freedom
Multiple R-Squared: 0.326,Adjusted R-squared: 0.31
F-statistic: 19.8 on 1 and 41 DF,  p-value: 6.39e-05
```

The table:

Table: OLS Regression Model of HIV/AIDS Rates in Africa, 2001

Variables	Model I
(Constant)	15.28
	(1.83)
Muslim Percentage of the Population	-0.164*
	(0.037)
Adjusted R^2	0.31

Note: $N = 43$. Cell entries are coefficient estimates; numbers in parentheses are estimated standard errors. Asterisks indicate $p < .05$ (one-tailed). See text for details.