PLSC 503 – Spring 2024

Multivariate Regression

January 22, 2024

"Multivariate" linear regression:

$$\underset{N\times 1}{\mathbf{Y}} = \underset{N\times K}{\mathbf{X}}\underset{K\times 1}{\boldsymbol{\beta}} + \underset{N\times 1}{\mathbf{u}}$$

or:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_K X_{Ki} + u_i$$

or:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{K1} \\ 1 & X_{12} & X_{22} & \cdots & X_{K2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1N} & X_{2N} & \cdots & X_{KN} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}.$$

Estimating $\boldsymbol{\beta}$

Residuals:

$$\mathbf{u} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$$

The inner product of **u**:

$$\begin{aligned}
\mathbf{u}'\mathbf{u} &= \begin{bmatrix} u_1 & u_2 & \cdots & u_N \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \\
&= u_1^2 + u_2^2 + ... + u_N^2 \\
&= \sum_{i=1}^{N} u_i^2
\end{aligned}$$

Estimating $\boldsymbol{\beta}$

Start with:

$$\begin{aligned}\mathbf{u}'\mathbf{u} &= (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\end{aligned}$$

Now get:

$$\frac{\partial \mathbf{u}'\mathbf{u}}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

Solve:

$$\begin{aligned}-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} &= 0 \\ -\mathbf{X}'\mathbf{Y} + \mathbf{X}'\mathbf{X}\boldsymbol{\beta} &= 0 \\ \mathbf{X}'\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}'\mathbf{Y} \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \boldsymbol{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\end{aligned}$$

Estimation Issues

> *"Do not compute the least squares estimates using $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$!"*
>
> – *Weisberg (2013, p. 61)*

Most software uses a QR decomposition of $\mathbf{X}$:

$$\mathbf{X} = \mathbf{QR}$$

where $\mathbf{Q}$ is orthogonal ($\mathbf{Q}'\mathbf{Q} = \mathbf{QQ}' = \mathbf{I}$) and $\mathbf{R}$ is upper-triangular.

Why? See e.g. here, here, or section 3.19 of this.

By decomposing $\mathbf{X}$ into $\mathbf{QR}$, we can have:

$$\begin{aligned}
\mathbf{X}'\mathbf{X}\beta &= \mathbf{X}'\mathbf{Y} \\
(\mathbf{QR})'(\mathbf{QR})\beta &= (\mathbf{QR})'\mathbf{Y} \\
\mathbf{R}'\mathbf{Q}'\mathbf{QR}\beta &= \mathbf{R}'\mathbf{Q}'\mathbf{Y} \\
\mathbf{R}'\mathbf{R}\beta &= \mathbf{R}'\mathbf{Q}'\mathbf{Y} \\
\mathbf{R}\beta &= \mathbf{Q}'\mathbf{Y}
\end{aligned}$$

Here, $\mathbf{Q}'\mathbf{Y}$ is a vector, so we might write:

$$\mathbf{R}\beta = \mathbf{V}$$

...which is an upper-triangular system of equations that can be easily solved by backward substitution. Doing this avoids inverting $(\mathbf{X}'\mathbf{X})$ entirely, which is good for a few reasons.

OLS Assumptions

1. Expectation-Zero Disturbances

$$\mathrm{E}(\mathbf{u}) = \mathbf{0}$$

2. Homoscedasticity / No Error Correlation

$$\begin{aligned}
\mathbf{u}\mathbf{u}' &= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_N \end{bmatrix} \\
&= \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots & u_1 u_N \\ u_2 u_1 & u_2^2 & \cdots & u_2 u_N \\ \vdots & \vdots & \ddots & \vdots \\ u_N u_1 & u_N u_2 & \cdots & u_N^2 \end{bmatrix}
\end{aligned}$$

Assumption: Expectation must be:

$$\mathrm{E}(\mathbf{u}\mathbf{u}') = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \underset{N \times N}{\mathbf{I}}$$

3. "Fixed" $\mathbf{X}$...

- No *measurement error* in the $\mathbf{X}$s, and
- No *endogeneity*, and
- $\text{Cov}(\mathbf{X}, \mathbf{u}) = \mathbf{0}$.

4. $\mathbf{X}$ is full column rank.

Means:

- No exact linear relationship among $\mathbf{X}$, and
- $K < N$.

5. Normal Disturbances

$$\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

OLS: Unbaisedness (again)

Start with:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Substitute OLS $\hat{\boldsymbol{\beta}}$:

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
&= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}
\end{aligned}$$

and so:

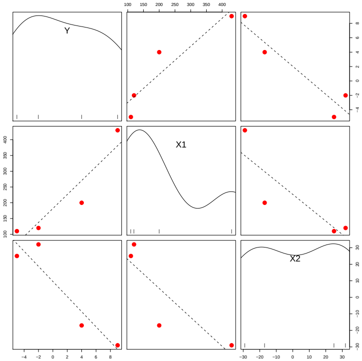
$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}.$$

By $\mathrm{Cov}(\mathbf{X}, \mathbf{u}) = \mathbf{0}$, we have $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.

A Toy Example

$$\mathbf{Y} = \begin{bmatrix} 4 \\ -2 \\ 9 \\ -5 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 200 & -17 \\ 1 & 120 & 32 \\ 1 & 430 & -29 \\ 1 & 110 & 25 \end{bmatrix}$$



So:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 4 & 860 & 11 \\ 860 & 251400 & -9280 \\ 11 & -9280 & 2779 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 3.2453 & -0.0132 & -0.05694 \\ -0.0132 & 0.000058 & 0.0002468 \\ -0.0569 & 0.000247 & 0.001409 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} 6 \\ 3880 \\ 518 \end{bmatrix}$$

...which means that:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \begin{bmatrix} 3.2453 & -0.0132 & -0.05694 \\ -0.0132 & 0.000058 & 0.0002468 \\ -0.0569 & 0.000247 & 0.001409 \end{bmatrix} \begin{bmatrix} 6 \\ 3880 \\ 518 \end{bmatrix} \\ &= \begin{bmatrix} -2.264 \\ 0.0190 \\ -0.1141 \end{bmatrix} \end{aligned}$$

Minimal Example: Correlation

```
Y<-c(4,-2,9,-5)
X1<-c(200,120,430,110)
X2<-c(-17,32,-29,25)
data<-cbind(Y,X1,X2)
```

```
cor(data)
```

```
         Y      X1      X2
Y   1.0000  0.9285 -0.9425
X1  0.9285  1.0000 -0.8613
X2 -0.9425 -0.8613  1.0000
```

OLS via `lm`:

```
> fit<-lm(Y~X1+X2)

> summary(fit)

Call:
lm(formula = Y ~ X1 + X2)

Residuals:
     1      2      3      4 
 0.531  1.639 -0.201 -1.970 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.2643     4.7284   -0.48     0.72
X1            0.0190     0.0200    0.95     0.52
X2           -0.1141     0.0985   -1.16     0.45

Residual standard error: 2.62 on 1 degrees of freedom
Multiple R-Squared: 0.941,Adjusted R-squared: 0.823 
F-statistic: 7.99 on 2 and 1 DF,  p-value: 0.243
```

Inference, In General

One approach is "testing":

- Pick some $\mathbf{H}_A : \boldsymbol{\beta} = \boldsymbol{\beta}_A$
- Estimate $\hat{\boldsymbol{\beta}}$
- Determine distribution of $\hat{\boldsymbol{\beta}}$ under $\mathbf{H}_A$
- Form a *test statistic* $\hat{\mathbf{S}} = h(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}})$
- Assess $\Pr(\hat{\mathbf{S}} | \mathbf{H}_A)$

The Importance of $\mathbf{V}(\hat{\beta})$

Start with:

$$\begin{aligned}\mathbf{V}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})]^2 \\ &= \mathrm{E}\{[\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})][\hat{\boldsymbol{\beta}} - \mathrm{E}(\hat{\boldsymbol{\beta}})]'\}\end{aligned}$$

Rewrite:

$$\begin{aligned}\mathbf{V}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \\ &= \mathrm{E}\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]'\} \\ &= \mathrm{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]\end{aligned}$$

Taking expectations:

$$\begin{aligned}\mathbf{V}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Estimating $\mathbf{V}(\hat{\beta})$

Empirical estimate:

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N - K}$$

Yields:

$$\widehat{\mathbf{V}(\hat{\beta})} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$$

Single Coefficient Hypothesis Tests

We know that:

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}].$$

In practice, using $\hat{\sigma}^2$ means

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim t_{N-K}$$

Procedure:

- Choose a value of β_k that you want to test (say, $\beta_k = 0$),
- Calculate the t-statistic for the coefficient associated with X_k, which is:

$$\hat{t}_k = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\widehat{\mathbf{V}(\hat{\beta}_k)}}}$$

- Compare $\hat{t}_k$ to a t distribution with $N - K$ degrees of freedom.

E.g.: $\quad H_0 : \beta_1 = \beta_2 = ... = \beta_K = 0$

or: $\quad H_0 : \beta_3 = \beta_6 = 0$

Generally: *Linear restrictions*:

$$\underset{q\times k}{\mathbf{R}}\,\underset{k\times 1}{\boldsymbol{\beta}} = \underset{q\times 1}{\mathbf{r}}$$

E.g.:

$$\beta_2 = -2 \iff (0\ 1\ 0)\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = -2$$

Recall:

$$\textbf{TSS} \quad = \quad \textbf{MSS} \quad + \quad \textbf{RSS}$$

Consider:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{Ui}$$

and the restriction:

$$\mathrm{H}_a : \beta_2 = \beta_4 = 0.$$

Restricted model:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + 0X_{2i} + \beta_3 X_{3i} + 0X_{4i} + u_i \\ &= \beta_0 + \beta_1 X_{1i} + \beta_3 X_{3i} + u_{Ri} \end{aligned}$$

F-tests: Sums of Squared Residuals

"Unrestricted":

$$\text{RSS}_U \equiv \hat{\mathbf{u}}_U' \hat{\mathbf{u}}_U = \sum_{i=1}^{N} \hat{u}_{Ui}^2$$

"Restricted":

$$\text{RSS}_R \equiv \hat{\mathbf{u}}_R' \hat{\mathbf{u}}_R = \sum_{i=1}^{N} \hat{u}_{Ri}^2$$

F-statistic:

$$\mathbf{F} = \frac{(\mathrm{RSS}_R - \mathrm{RSS}_U)/q}{\mathrm{RSS}_U/(N-K)}$$

$$= \frac{(R_U^2 - R_R^2)/q}{(1-R_U^2)/N-K}$$

Testing:

$$\mathbf{F} \sim F_{q,N-K}$$

F-Test: Example

Consider:

$$\begin{aligned} \mathrm{H}_b: \qquad & \beta_1 + \beta_4 = 1 \\ & \beta_1 = 1 - \beta_4 \end{aligned}$$

Implies:

$$\begin{aligned} Y_i &= \beta_0 + (1 - \beta_4) X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{R'i} \\ &= \beta_0 + X_{1i} - \beta_4 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_{R'i} \\ &= \beta_0 + X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 (X_{4i} - X_{1i}) + u_{R'i} \end{aligned}$$

...which further implies the restricted model:

$$Y_i - X_{1i} = \beta_0 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 (X_{4i} - X_{1i}) + u_{R'i}$$

Note that:

$$F = \frac{(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q^H)'\hat{\mathbf{V}}_q^{-1}(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q^H)}{q\hat{\sigma}^2}$$

...which implies that, for some confidence level α:

$$\Pr\left[\frac{(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q^H)'\hat{\mathbf{V}}_q^{-1}(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q^H)}{q\hat{\sigma}^2} \leq F_{q,N-K}\right] = 1 - \alpha.$$

$\rightarrow$ "confidence region" of all points satisfying:

$$(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q^H)'\hat{\mathbf{V}}_q^{-1}(\hat{\boldsymbol{\beta}}_q - \boldsymbol{\beta}_q^H) \leq q\hat{\sigma}^2 F_{q,N-K}.$$

The linear prediction:

$$\hat{Y}_j = \mathbf{X}_j \hat{\boldsymbol{\beta}}$$

...has variance:

$$\widehat{\mathbf{V}(\hat{Y}_j)} = \hat{\sigma}^2[1 + \mathbf{X}_j(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_j']$$

...$\rightarrow$ standard error:

$$\widehat{\text{s.e.}(\hat{Y}_j)} = \sqrt{\hat{\sigma}^2[1 + \mathbf{X}_j(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_j']}$$

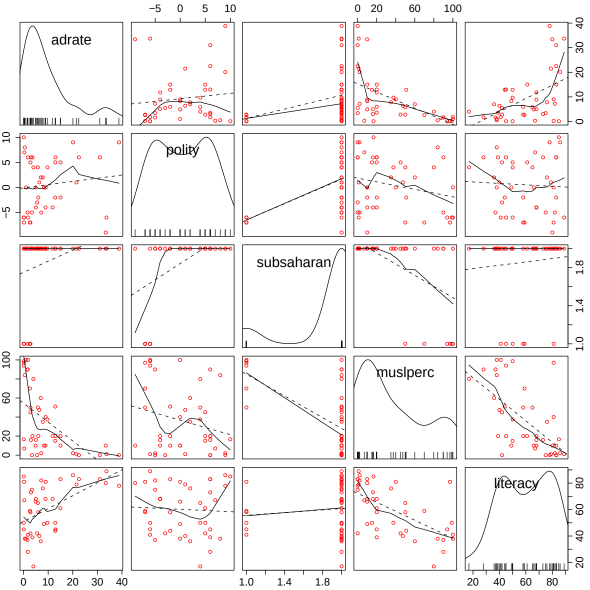
Example: Africa Data

```
> Data<-read_csv("https://raw.githubusercontent.com/PrisonRodeo/PLSC503-2024-git/master/Data/africa2001.csv")
> Data<-with(Data, data.frame(adrate,polity,
+           subsaharan=as.numeric(as.factor(subsaharan))-1,
+           muslperc,literacy))
```

```
> describe(Data)

           vars  n  mean    sd median trimmed   mad  min   max range  skew kurtosis   se
adrate        1 43  9.37  9.96      6    7.58  6.38  0.1  38.8  38.7  1.44     1.23 1.52
polity        2 43  0.51  5.41      0    0.46  7.41 -9.0  10.0  19.0  0.01    -1.38 0.82
subsaharan    3 43  0.86  0.35      1    0.94  0.00  0.0   1.0   1.0 -2.01     2.08 0.05
muslperc      4 43 35.96 34.58     20   32.87 29.65  0.0 100.0 100.0  0.68    -1.04 5.27
literacy      5 43 60.07 18.94     61   60.63 26.69 17.0  89.0  72.0 -0.20    -1.18 2.89
```

```
> cor(Data)
            adrate   polity subsaharan muslperc literacy
adrate      1.0000  0.11794    0.33129  -0.5709  0.51489
polity      0.1179  1.00000    0.52820  -0.2392 -0.05079
subsaharan  0.3313  0.52820    1.00000  -0.5773  0.09473
muslperc   -0.5709 -0.23917   -0.57725   1.0000 -0.61960
literacy    0.5149 -0.05079    0.09473  -0.6196  1.00000
```

Africa Data



A Regression

```
> model<-lm(adrate~polity+subsaharan+muslperc+literacy,data=Data)
> summary(model)

Call:
lm(formula = adrate ~ polity + subsaharan + muslperc + literacy,
    data = Data)

Residuals:
     Min       1Q   Median       3Q      Max
-15.4681  -4.3947  -0.5251   3.4246  22.9358

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.39843   14.94744  -0.294   0.7702
polity      -0.01390    0.27969  -0.050   0.9606
subsaharan   3.72969    5.43093   0.687   0.4964
muslperc    -0.08689    0.06282  -1.383   0.1747
literacy     0.16575    0.09433   1.757   0.0869 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.264 on 38 degrees of freedom
Multiple R-squared:  0.3771,Adjusted R-squared:  0.3115
F-statistic: 5.751 on 4 and 38 DF,  p-value: 0.001013
```

A Regression, Annotated

```
> model<-lm(adrate~polity+subsaharan+muslperc+literacy,data=Data)
> summary(model)

Call:
lm(formula = adrate ~ polity + subsaharan + muslperc + literacy,
    data = Data)

Residuals:
     Min       1Q   Median       3Q      Max
-15.4681  -4.3947  -0.5251   3.4246  22.9358

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.39843   14.94744  -0.294   0.7702
polity      -0.01390    0.27969  -0.050   0.9606
subsaharan   3.72969    5.43093   0.687   0.4964
muslperc    -0.08689    0.06282  -1.383   0.1747
literacy     0.16575    0.09433   1.757   0.0869 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 8.264 on 38 degrees of freedom
Multiple R-squared:  0.3771,Adjusted R-squared:  0.3115
F-statistic: 5.751 on 4 and 38 DF,  p-value: 0.001013
```

Summarizing the model object...

Showing you what model you fit...

Median, minimum, maximum, and 25th/75th percentiles of the model residuals (u-hats); ideally the median is close to zero, and the other numbers symmetrical around it

t-statistics associated with the test of the null hypothesis that that particular beta equals zero

Beta-hats

Standard errors [square roots of the diagonal of Var(Beta)]

Two-tailed P-values associated with the t-statistics

Indicators of what the little "stars" (asterisks) next to the P-values mean

The R-squared and adjusted R-squared values of the model

RSE: the size of a "typical" residual in the model (in units of Y)

The F-statistic and associated P-value for the joint null hypothesis that all the betas except the intercept are zero

Variance-Covariance Matrix of $\hat{\beta}$

Matrix:

```
> options(digits=4)
> vcov(model)

            (Intercept)    polity subsaharan  muslperc  literacy
(Intercept)    223.4259  1.088030   -72.2628 -0.771309 -1.002421
polity           1.0880  0.078229    -0.6642 -0.000293  0.001968
subsaharan     -72.2628 -0.664212    29.4950  0.206067  0.171765
muslperc        -0.7713 -0.000293     0.2061  0.003946  0.004098
literacy        -1.0024  0.001968     0.1718  0.004098  0.008898
```

$\rightarrow$ standard error estimates:

```
> sqrt(diag(vcov(model)))

(Intercept)      polity  subsaharan    muslperc    literacy
10.41131167  0.27969425  5.43093195  0.06281584  0.09432686

> summary(model)[[4]][,2]

(Intercept)      polity  subsaharan    muslperc    literacy
10.41131167  0.27969425  5.43093195  0.06281584  0.09432686
```

Diversion: "Added Variable Plots"

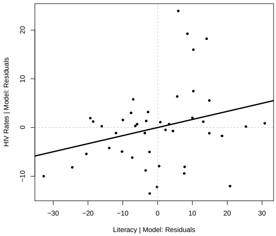
"Added variable plot": If X_J is the (potential) added variable:

- Regress Y on $X_1, X_2, \ldots$ and save the residuals $\hat{u}_i$,
- Regress X_J on $X_1, X_2, \ldots$ and save the residuals (call these $\hat{e}_i$),
- Plot $\hat{u}_i$ (conventionally on the y-axis) vs. $\hat{e}_i$ (conventionally on the x-axis).

Example: Estimate the regression:

$$\text{HIV Rate} = \beta_0 + \beta_1 \text{POLITY} + \beta_2 \text{Subsaharan} + \beta_3 \text{Muslim Pct.} + u$$

and consider the added variable `literacy`:



Tests...

Test $H_0 : \beta_{\texttt{polity}} = \beta_{\texttt{subsaharan}} = 0$:

```
> library(lmtest)
> modelsmall<-lm(adrate~muslperc+literacy,data=Data)
> waldtest(model,modelsmall)

Wald test

Model 1: adrate ~ polity + subsaharan + muslperc + literacy
Model 2: adrate ~ muslperc + literacy
  Res.Df Df    F Pr(>F)
1     38
2     40 -2 0.27   0.76
```

Test $H_0 : \beta_{\texttt{muslperc}} = 0.1$:

```
> library(car)
> linearHypothesis(model,"muslperc=0.1")

Linear hypothesis test

Hypothesis:
muslperc = 0.1

Model 1: restricted model
Model 2: adrate ~ polity + subsaharan + muslperc + literacy

  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     39 3200
2     38 2595  1       605 8.85 0.0051 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

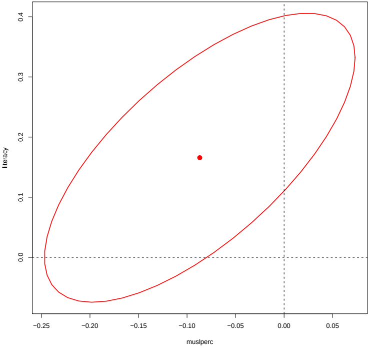
More tests...

Test $H_0 : \beta_{\texttt{literacy}} = \beta_{\texttt{muslperc}}$:

```
> linearHypothesis(model,"literacy=muslperc")

Linear hypothesis test

Hypothesis:
- muslperc  + literacy = 0

Model 1: restricted model
Model 2: adrate ~ polity + subsaharan + muslperc + literacy

  Res.Df  RSS Df Sum of Sq    F  Pr(>F)
1     39 3534
2     38 2595  1       938 13.7 0.00067 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Confidence Regions / Ellipses

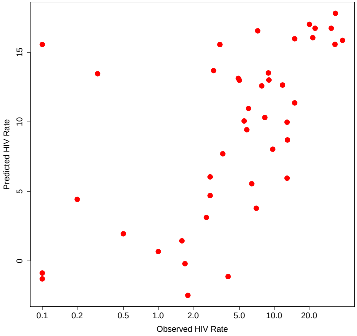
```
> confidenceEllipse(model=model,which.coef=c(4,5),
                    xlab="Muslim Percentage",ylab="Literacy")
> abline(h=0,v=0,lty=2)
```



Predicted Values

```
> hats<-fitted(model)
> # Or, alternatively:
> fitted<-predict(model,se.fit=TRUE, interval=c("confidence"))
> scatterplot(model$fitted~adrate,log="x",smooth=FALSE,boxplots=FALSE,
  reg.line=FALSE,xlab="Observed HIV Rate",ylab="Predicted HIV Rate",
  pch=16,cex=2)
```

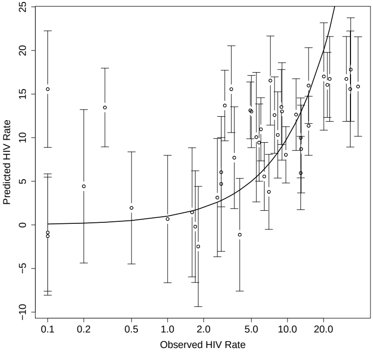
Predicted and Actual HIV/AIDS Rates (X-Axis Logged)



An Even More Useful Plot

```
> library(plotrix)
> plotCI(Data$adrate,model$fitted,uiw=(1.96*(fitted$se.fit)),
        log="x",xlab="Observed HIV Rate",ylab="Predicted HIV Rate")
> lines(lowess(Data$adrate,Data$adrate),lwd=2)
```

Predicted and Actual HIV/AIDS Rates, with 95% C.I.s

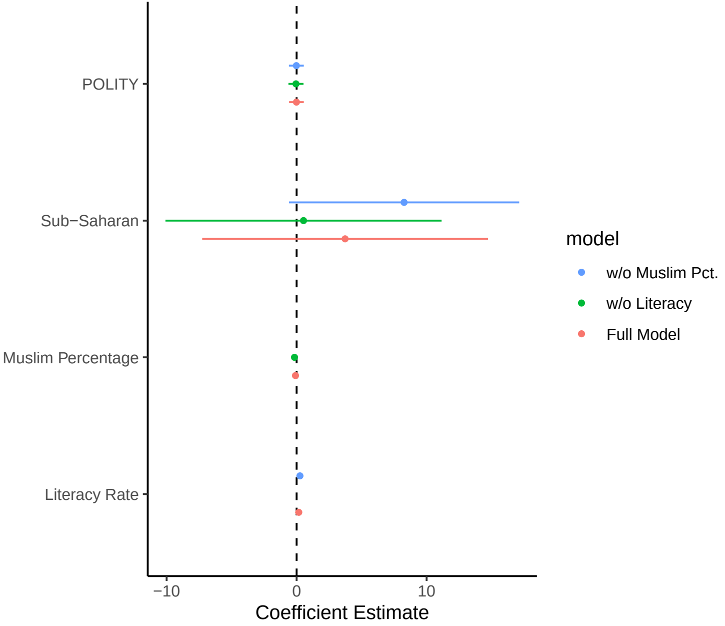


Presentation: A (De)Fault-y Table

```
> M1<-lm(adrate~polity+subsaharan+muslperc+literacy,data=Data)
> M2<-lm(adrate~polity+subsaharan+muslperc,data=Data)
> M3<-lm(adrate~polity+subsaharan+literacy,data=Data)
>
> stargazer(M1,M2,M3)
```

	Dependent variable:		
	adrate		
	(1)	(2)	(3)
polity	−0.014	−0.051	−0.020
	(0.280)	(0.286)	(0.283)
subsaharan	3.730	0.530	8.268^{*}
	(5.431)	(5.252)	(4.379)
muslperc	−0.087	-0.163^{***}	
	(0.063)	(0.047)	
literacy	0.166^{*}		0.256^{***}
	(0.094)		(0.069)
Constant	−0.669	14.800^{**}	-13.120^{**}
	(10.410)	(5.701)	(5.298)
Observations	43	43	43
R^2	0.377	0.326	0.346
Adjusted R^2	0.312	0.275	0.295
Residual Std. Error	8.264 (df = 38)	8.483 (df = 39)	8.361 (df = 39)
F Statistic	5.751^{***} (df = 4; 38)	6.302^{***} (df = 3; 39)	6.870^{***} (df = 3; 39)
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

A Dot-Whisker Plot



Gelman (2008 *Statistics in Medicine*)

Suggestion: Rescale *all* non-binary predictors by **dividing them by two times their standard deviation**.

- Creates a "common scale" for every predictor.
- More specifically: Scales continuous predictors to be comparable to binary (0/1) ones.
- $\hat{\beta}_X$ now represents the change in $E(Y)$ associated with a change in X of two standard deviations (for example, from one s.d. below the mean to one s.d. above the mean).

Note that:

- People don't *routinely* (or even generally) do this. But...
- ...it can be very useful when you have predictor variables that are measured on very different "natural" scales.

A (Better?) Dot-Whisker Plot