STA302 weeks 8–9

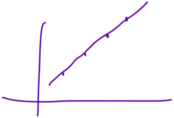
Mark Ebden 2018. Continuing Ch 5

With grateful acknowledgment to Alison Gibbs

What we're doing

- Midterms discussed on Thursday (not Tuesday)
- §5.2 (Estimation and Inference in MLR) is what we've begun, via the RMA
- Chapter 5's matrices

Fitted values ($\widehat{\mathbf{Y}}$) in matrix form



Recall from Weeks 6–7 that our model is:

$$\boxed{\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}}$$

Recalling that $\hat{\beta}$ is unbiased and that $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, we have $\widehat{\mathbf{Y}} = \mathbf{X}\hat{\beta}$. So:

$$\widehat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the **hat matrix**, comprised of the h_{ij} values.

$$\begin{bmatrix} h_{11} & h_{12} & \cdots & \\ & h_{22} & & \\ \vdots & & \ddots & \\ & & & h_{nn} \end{bmatrix}$$

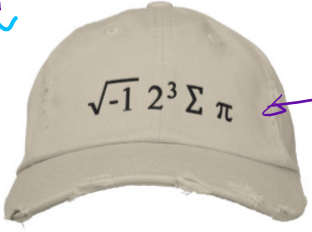
Re-"cap"

Recall from weeks 4-5 that h in $h_{ij} = \frac{1}{n} + \frac{(x_i - \bar{x})(x_j - \bar{x})}{S_{xx}}$ stands for "hat". This is because, considering $\hat{y}_i = \sum_{j=1}^{n} h_{ij}\, y_j$, the h values show how to get from y_i's to $\hat{y}_i$'s.

This is even more apparent in the matrix notation.

$\underset{\sim}{\hat{Y}} = \underset{\sim}{H}\,\underset{\sim}{Y}$

$\hat{y}$ h y



Properties of **H**

$A^{-1}A = I$

$H = H^2 = HH$

same

$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is an example of an *idempotent* matrix. **Exercise:** Show this.

$HH = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = H$

Remember

$h_{ij} = h_{ji}$?

H is symmetric. **Exercise:** Show this.

$H' = [\underbrace{X(X'X)^{-1}}_{A}\underbrace{X'}_{B}]'$

$= X((X'X)^{-1})'X' = H$

B'

Five facts about idempotent matrices

$$J = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad (n=3)$$

1. A square matrix $\mathbf{A}$ is idempotent iff $\mathbf{A}^2 = \mathbf{A}$
2. If $\mathbf{A}$ is idempotent then $\text{trace}(\mathbf{A}) = \text{rank}(\mathbf{A})$
3. $\mathbf{A}$ is idempotent iff $\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{I} - \mathbf{A}) = n$ where the dimensions of $\mathbf{A}$ are $n \times n$ and $\mathbf{I}$ is the $n \times n$ identity matrix
4. For hat matrix $\mathbf{H}$ and matrix of all 1's $\mathbf{J}$, the following matrices are idempotent:

$$\mathbf{H} \qquad \mathbf{I} - \mathbf{H} \qquad \frac{1}{n}\mathbf{J} \qquad \mathbf{H} - \frac{1}{n}\mathbf{J}$$

5. If $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are idempotent and $\mathbf{A} = \mathbf{B} + \mathbf{C}$, then $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B}) + \text{rank}(\mathbf{C})$

iff = "if and only if"

#4 is the only one we prove here

Residuals ($\widehat{\mathbf{e}}$) in matrix form

$$e_i = y_i - \hat{y}_i$$

The residuals are given by

$$\widehat{\mathbf{e}} = \begin{pmatrix} \hat{e}_1 \\ \vdots \\ \hat{e}_n \end{pmatrix} = \mathbf{Y} - \underbrace{\widehat{\mathbf{Y}}}_{HY} = \mathbf{Y} - \mathbf{X}\hat{\beta}$$

Donning our new hat matrix, this can be rewritten as $\widehat{\mathbf{e}} = \mathbf{Y} - \mathbf{HY}$ before determining $\mathrm{E}(\widehat{\mathbf{e}})$ and $\mathrm{var}(\widehat{\mathbf{e}})$.

To begin, how could we factorize $\widehat{\mathbf{e}} = \mathbf{Y} - \mathbf{HY}$?

$$= \hat{I}Y - \hat{H}Y$$

$$= (I - H)Y$$

Properties of $\mathbf{I} - \mathbf{H}$

Is $\mathbf{I} - \mathbf{H}$ idempotent?

$(I-H)(I-H) = II - IH - HI + HH$

$= I - H - \cancel{H} + \cancel{H}$

$= I - H$

Is $\mathbf{I} - \mathbf{H}$ symmetric?

$(I-H)' = I' - H'$

$= I - H$

Prev-class list:

$\hat{e}_i = y_i - \hat{y}_i$

$\hat{y}_i = \ldots$

$b_0 = \ldots$

$b_1 = \ldots$

$S_{xy} = \,, S_{xx}$

$\Sigma \hat{e}_i = 0$

$\Sigma \hat{e}_i x_i = 0$

$\Sigma \hat{e}_i \hat{y}_i = 0$

vars

cov's

etc

Matrices:

$\underset{\sim}{\hat{e}} = (\underset{\sim}{I} - \underset{\sim}{H})\underset{\sim}{Y}$

$\underset{\sim}{\hat{Y}} = \underset{\sim}{H}\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\hat{\beta}}$

$\underset{\sim}{H} = \underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'$

$\underset{\sim}{\hat{\beta}} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{Y}$

$\underset{1\times n}{\underset{\sim}{\hat{e}}'}\ \underset{n\times 2}{\underset{\sim}{X}} = \underset{1\times 2}{\underset{\sim}{0}}$

$\underset{\sim}{\hat{e}}'\underset{\sim}{\hat{Y}} = \underset{1\times 1}{0}$

Continuing:

$\mathrm{E}(\widehat{\mathbf{e}})=$

$$= E(Y - X\hat{\beta}) = E(Y) - E(X\hat{\beta})$$

$$= X\beta - XE(\hat{\beta})$$

$$= X\beta - X\beta$$

$$= 0$$

$\mathrm{var}(\widehat{\mathbf{e}})=$

$$\mathrm{var}\left[(I-H)Y\right] = (I-H)\,\mathrm{var}(Y)\,(I-H)'$$

$$= \underbrace{(I-H)}\;\sigma^2 I\;\underbrace{(I-H)}$$

$$= \sigma^2 (I-H)(I-H)$$

$$= \sigma^2 (I-H)$$

Recap of our recent studies

The SLR model in matrix form is $\boxed{\mathbf{Y} = \mathbf{X}\beta + \mathbf{e},}$ in which:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \qquad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

Setting the derivative of RSS(β) to zero yielded $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ when rank($\mathbf{X}$) $= 2$. This plus the fact that $\mathrm{E}(\mathbf{e}) = \mathbf{0}$ gives

$$\widehat{\mathbf{Y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$$

We can write the residuals in terms of idempotent matrix $\mathbf{I} - \mathbf{H}$ as

$$\widehat{\mathbf{e}} = \mathbf{Y} - \widehat{\mathbf{Y}} = \mathbf{Y} - \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\,\mathbf{Y}$$

We found $\mathrm{E}(\widehat{\mathbf{e}}) = \mathbf{0}$ and found ~~were about to try~~ var($\widehat{\mathbf{e}}$), requiring the notion of a covariance matrix: $\mathrm{var}(\mathbf{X}) = \mathrm{E}\left[(\mathbf{X} - \mathrm{E}(\mathbf{X}))\,(\mathbf{X} - \mathrm{E}(\mathbf{X}))'\right]$.

Variance of the residuals, in matrix form

$$
\begin{aligned}
\mathrm{var}(\hat{\mathbf{e}}) &= \mathrm{E}\left\{[\hat{\mathbf{e}} - \mathrm{E}(\hat{\mathbf{e}})]\,[\hat{\mathbf{e}} - \mathrm{E}(\hat{\mathbf{e}})]'\right\} \\
&= \mathrm{E}\left\{(\mathbf{I} - \mathbf{H})\,\mathbf{Y}\mathbf{Y}'\,(\mathbf{I} - \mathbf{H})\right\} \\
&= (\mathbf{I} - \mathbf{H})\,\mathrm{E}(\mathbf{Y}\mathbf{Y}')\,(\mathbf{I} - \mathbf{H})
\end{aligned}
$$

Compare to our previous work: $\mathrm{var}(\hat{e}_i) = \sigma^2(1 - h_{ii})$. Does the above match?

NB: As before, the "$|\mathbf{X}$" is implicit — e.g. $\mathrm{var}(\hat{\mathbf{e}}|X)$ is abbreviated as $\mathrm{var}(\hat{\mathbf{e}})$.

Variance of the residuals, in matrix form

$$
\begin{aligned}
\text{The middle factor is } \mathrm{E}(\mathbf{YY}') &= \mathrm{E}\left\{(\mathbf{X}\beta + \mathbf{e})(\mathbf{X}\beta + \mathbf{e})'\right\} \\
&= \mathrm{E}\left\{(\mathbf{X}\beta + \mathbf{e})(\beta'\mathbf{X}' + \mathbf{e}')\right\} \\
&= \mathrm{E}\left\{\mathbf{X}\beta\beta'\mathbf{X}' + \boxed{\mathbf{X}\beta\mathbf{e}'} + \mathbf{e}\beta'\mathbf{X}' + \mathbf{ee}'\right\} \\
&= \mathrm{E}\left\{\mathbf{X}\beta\beta'\mathbf{X}'\right\} + \mathbf{0} + \mathbf{0} + \mathrm{E}(\mathbf{ee}') \\
\mathrm{E}(\mathbf{YY}') &= \mathbf{X}\beta\beta'\mathbf{X}' + \sigma^2\mathbf{I}
\end{aligned}
$$

$E(X\beta e')$

const

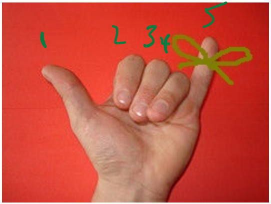
Gauss-Markov cond'n:

$E(e) = 0$

Inserting the above into $\mathrm{var}(\widehat{\mathbf{e}})$ gives

$$
\begin{aligned}
\mathrm{var}(\widehat{\mathbf{e}}) &= (\mathbf{I} - \mathbf{H})(\mathbf{X}\beta\beta'\mathbf{X}' + \sigma^2\mathbf{I})(\mathbf{I} - \mathbf{H}) \\
&= \left[\mathbf{I}\left(\mathbf{X}\beta\beta'\mathbf{X}' + \sigma^2\mathbf{I}\right) - \mathbf{H}\left(\mathbf{X}\beta\beta'\mathbf{X}' + \sigma^2\mathbf{I}\right)\right](\mathbf{I} - \mathbf{H}) \\
&= \left[\mathbf{X}\beta\beta'\mathbf{X}' + \sigma^2\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{X}\beta\beta'\mathbf{X}'\right) - \sigma^2\mathbf{H}\right](\mathbf{I} - \mathbf{H}) \\
&= \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) \\
\mathrm{var}(\widehat{\mathbf{e}}) &= \sigma^2(\mathbf{I} - \mathbf{H})
\end{aligned}
$$

What's the rank of var($\widehat{\mathbf{e}}$)? $= \underbrace{\sigma^2}\ \underbrace{(I - H)}$



Recall the fifth of our *Five facts about idempotent matrices*:

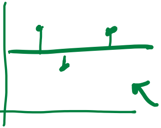
$$\text{If } \mathbf{A} = \mathbf{B} + \mathbf{C}, \text{ then } \operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{B}) + \operatorname{rank}(\mathbf{C}).$$

Put another way, $\operatorname{rank}(\underset{I-H}{\mathbf{B}}) = \operatorname{rank}(\underset{I}{\mathbf{A}}) - \operatorname{rank}(\underset{H}{\mathbf{C}})$.

Therefore, for example $\operatorname{rank}(\mathbf{I} - \mathbf{H}) = \operatorname{rank}(\mathbf{I}) - \operatorname{rank}(\mathbf{H}) = n - 2$. We'll do other similar calculations when considering ANOVA in matrix terms.

ANOVA in matrix terms

Recall from Week 3 that



$$\boxed{\text{SST} = \text{SSReg} + \text{RSS}}$$

where

$$\text{SST} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

Exercise: Show that SST can be re-expressed as

$$\text{SST} = \underbrace{\mathbf{Y}'\mathbf{Y}}_{I} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$$

where $\mathbf{J}$ is an $n \times n$ matrix of 1's. This means we can also write

$$n=3: \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = J$$

$$\boxed{\text{SST} = \mathbf{Y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}}$$

$$\overbrace{\sum (y_i - \bar{y})^2}^{SST} = \overbrace{\sum (y_i - \hat{y}_i)^2}^{RSS} + \overbrace{\sum (\hat{y}_i - \bar{y})^2}^{SSReg}$$

$$\left((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})\right)^2$$

$$b_1^2 \sum (x_i - \bar{x})^2$$

Some properties of $\mathbf{I} - \frac{1}{n}\mathbf{J}$

$\underbrace{Y'(I - \frac{1}{n}J)Y}_{SST}$

1. Note that $\mathbf{I} - \frac{1}{n}\mathbf{J}$ is symmetric. For a vector $\mathbf{Y}$ and symmetric matrix $\mathbf{A}$, you may recall from other courses that $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ is a *quadratic form* (entirely second-degree polynomial).

q. form in 1 variable: $a y_1^2$

2 vars: $a y_1^2 + b y_2^2 + c y_1 y_2$

⋮

n vars: $a_1 y_1^2 + \cdots a_n y_n^2 + b_1 y_1 y_2 + b_2 y_1 y_3 + \cdots$

2. ~~Since~~ $\mathbf{I}$ is idempotent and $\frac{1}{n}\mathbf{J}$ is idempotent (from the five facts), $\mathbf{I} - \frac{1}{n}\mathbf{J}$ is also idempotent.

3. The rank of $\mathbf{I} - \frac{1}{n}\mathbf{J}$ is $\text{rank}(\mathbf{I}) - \text{rank}\left(\frac{1}{n}\mathbf{J}\right) = n - 1$.

- This is the number of degrees of freedom for SST

Decomposing SST $y_i - \hat{y}_i + \hat{y}_i$

Taking the first term of SST $= \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$,

$$\begin{aligned}
\mathbf{Y}'\mathbf{Y} &= (\mathbf{Y} - \mathbf{Xb} + \mathbf{Xb})'\,(\mathbf{Y} - \mathbf{Xb} + \mathbf{Xb}) \\
&= (\mathbf{Y} - \mathbf{Xb})'\,(\mathbf{Y} - \mathbf{Xb}) + (\mathbf{Y} - \mathbf{Xb})'\mathbf{Xb} + (\mathbf{Xb})'\,(\mathbf{Y} - \mathbf{Xb}) + (\mathbf{Xb})'(\mathbf{Xb}) \\
&= \hat{\mathbf{e}}'\hat{\mathbf{e}} + \underbrace{\hat{\mathbf{e}}'\mathbf{Xb} + (\mathbf{Xb})'\hat{\mathbf{e}}}_{\text{equal scalars, both 0}} + \mathbf{b}'\mathbf{X}'\mathbf{Xb} \\
&= \boxed{\hat{\mathbf{e}}'\hat{\mathbf{e}}} + \mathbf{b}'\mathbf{X}'\mathbf{Xb}
\end{aligned}$$

RSS

The middle terms were zero because

$$\mathbf{X}'\hat{\mathbf{e}} = \mathbf{X}'(\mathbf{I} - \mathbf{H})\mathbf{Y} = \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{0}$$

So,

$$\text{SST} = \mathbf{Y}'\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}$$

$$\text{SST} = \underbrace{\hat{\mathbf{e}}'\hat{\mathbf{e}}}_{\text{RSS}} + \boxed{\underbrace{\mathbf{b}'\mathbf{X}'\mathbf{Xb} - \frac{1}{n}\mathbf{Y}'\mathbf{J}\mathbf{Y}}_{\text{SSReg}}}$$

A closer look at RSS

$$\boxed{\text{SST} = \text{SSReg} + \text{RSS}}$$

Making use of our expression $\hat{\mathbf{e}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$, we have

$$\begin{aligned} \text{RSS} &= \hat{\mathbf{e}}'\hat{\mathbf{e}} \\ &= \mathbf{Y}'\,(\mathbf{I} - \mathbf{H})'\,(\mathbf{I} - \mathbf{H})\,\mathbf{Y} \\ &= \mathbf{Y}'\,(\mathbf{I} - \mathbf{H})\,\mathbf{Y} \end{aligned}$$

This is another quadratic form in $\mathbf{Y}$. Also, $\text{rank}(\mathbf{I} - \mathbf{H}) = n - 2$ from earlier, the number of degrees of freedom for the error.

A closer look at SSReg

$$\boxed{\text{SST} = \text{SSReg} + \text{RSS}}$$

Making use of $\hat{\beta} = \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, and slide 16

$$\begin{aligned}
\text{SSReg} &= \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{Y}'\frac{1}{n}\mathbf{J}\mathbf{Y} \\
&= \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\frac{1}{n}\mathbf{J}\mathbf{Y} \\
&= \mathbf{Y}'\underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{H}}\mathbf{Y} - \mathbf{Y}'\frac{1}{n}\mathbf{J}\mathbf{Y} \\
&= \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}
\end{aligned}$$

This is again a quadratic form in $\mathbf{Y}$, since the middle matrix is symmetric. Also, $\text{rank}(\mathbf{H} - \frac{1}{n}\mathbf{J}) = \text{rank}(\mathbf{H}) - \text{rank}\left(\frac{1}{n}\mathbf{J}\right) = 2 - 1 = 1$, the number of degrees of freedom for SSReg.

Recall: Rank(X) is generally 2. And from RMA p 26: rank$(H) = \cdots = 2$

Using RSS to estimate σ^2

In $S^2 = \text{RSS}/(n-2)$, we have an unbiased estimator for σ^2. We can show it's unbiased using matrices by showing that $\mathrm{E}(\text{RSS}) = (n-2)\sigma^2$ as we did without matrices — i.e. when we considered $\mathrm{E}(\text{RSS}) = \mathrm{E}\left(\sum_{i=1}^{n} \hat{e}_i^2\right)$.

trace(scalar) = itself

$$
\begin{aligned}
\mathrm{E}(\text{RSS}) &= \mathrm{E}\left(\hat{\mathbf{e}}'\hat{\mathbf{e}}\right) \\
&= \mathrm{E}\left\{\mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}\right\} && \text{from slide 17} \\
&= \mathrm{E}\left\{\text{trace}\left[\mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}\right]\right\} && \text{since } \mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y} \text{ is a scalar} \\
&= \mathrm{E}\left\{\text{trace}\left[(\mathbf{I}-\mathbf{H})\mathbf{Y}\mathbf{Y}'\right]\right\} && \text{since } \text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA}) \quad \star \\
&= \text{trace}\left[(\mathbf{I}-\mathbf{H})\,\mathrm{E}(\mathbf{Y}\mathbf{Y}')\right] \\
&= \text{trace}\left[(\mathbf{I}-\mathbf{H})\left(\sigma^2\mathbf{I} + \mathbf{X}\beta\beta'\mathbf{X}'\right)\right] && \text{from slide 12} \\
&= \text{trace}\left[(\mathbf{I}-\mathbf{H})\,\sigma^2 + \underbrace{\mathbf{X}\beta\beta'\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta\beta'\mathbf{X}'}_{=0}\right] \\
&= \text{trace}(\mathbf{I}-\mathbf{H})\,\sigma^2 \\
\mathrm{E}(\text{RSS}) &= (n-2)\sigma^2
\end{aligned}
$$

where we have used $\text{trace}(\mathbf{A}+\mathbf{B}) = \text{trace}(\mathbf{A}) + \text{trace}(\mathbf{B})$, $\star$ and thus $\text{trace}(\mathbf{I}-\mathbf{H}) = \text{trace}(\mathbf{I}) - \text{trace}(\mathbf{H}) = n - \underbrace{\sum_{i=1}^{n} h_{ii} = n-2}$. ~ Week 4

The big picture

We have expressed the ANOVA identity in matrix form:

$$\text{SST} = \text{SSReg} + \text{RSS}$$

$$\underbrace{\mathbf{Y}'\left(\mathbf{I} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}}_{\text{SST}} = \underbrace{\mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}}_{\text{SSReg}} + \underbrace{\mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}}_{\text{RSS}}$$

y_i $\bar{y}$ $\hat{y}_i$ $\bar{y}$ y_i $\hat{y}_i$

What does R have to say about this?

ANOVA in R

$Y = \beta_0 + \beta_1 x + e$

PPC

SSReg
RSS
SST

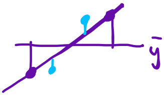
The anova command is one way in R to produce an ANOVA table (of the sort we saw in Week 3), in addition to analysing it. For example, for a 654-point SLR problem:

PPC

```
a2 = read.table("DataA2.txt",sep=" ",header=T) # Load the data set
fev <- a2$fev; age <- a2$age
mod1 = lm(fev~age)
anova(mod1)
```

estimate of σ^2
(if $H_0: \beta_1 = 0$ is true)

```
## Analysis of Variance Table
##
## Response: fev
##            Df Sum Sq Mean Sq F value    Pr(>F)
## age         1 280.92 280.919  872.18 < 2.2e-16 ***
## Residuals 652 210.00   0.322
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



ANOVA in R

The p-value will match that obtained from the `summary(lm...)` command:

```r
summary(mod1)
```

```
##
## Call:
## lm(formula = fev ~ age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.57539 -0.34567 -0.04989  0.32124  2.12786
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.431648   0.077895   5.541 4.36e-08 ***
## age         0.222041   0.007518  29.533  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5675 on 652 degrees of freedom
## Multiple R-squared:  0.5722, Adjusted R-squared:  0.5716
## F-statistic: 872.2 on 1 and 652 DF,  p-value: < 2.2e-16
```

$\hat{\beta}_0$

$\hat{\beta}_1$

We hope you have an enjoyable week

- Remember that for the Study Break of 5-9 November, there will be a pause in lectures, TA office hours, and instructor office hours
- No pause in the New College Stats Aid Centre