

## STA302/STA1001, Weeks 9-10

Mark Ebden, 14 & 16 November 2017

With grateful acknowledgment to Alison Gibbs

# This week's lectures

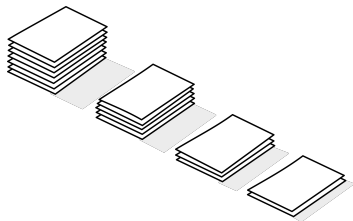
- ▶ Poll regarding the Exam Jam on 8 December
- ▶ Chapter 5:
  - ▶ Matrix version of SLR
  - ▶ Multiple linear regression (MLR)



## If you haven't retrieved your midterm

To pick up your test at SS 6027 CLTA:

1. Book an office-hours slot online for any Wednesday until 6 December, or
2. Drop in on Tuesday 14 November, 2-2:30 pm, or
3. Drop in on Thursday 16 November, 1-1:30 pm



## To appeal/discuss a recent TA decision on regrading

Please express your appeal via your section's regrading email address.



If the TA who had examined your work doesn't reply within a week then please notify me.

## Recap of our recent studies

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

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

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

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

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

$$\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\beta} = \mathbf{Y} - \mathbf{H}\mathbf{Y} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

We find  $E(\hat{\mathbf{e}}) = \mathbf{0}$  and were about to try  $\text{var}(\hat{\mathbf{e}})$ , requiring the notion of a covariance matrix:  $\text{var}(\mathbf{X}) = E[(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))']$ .

## Five facts about idempotent matrices (*Weeks 8–9, slide 18*)

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  $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B}) + \text{rank}(\mathbf{C})$

## Variance of the residuals, in matrix form

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

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

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

## Variance of the residuals, in matrix form

The middle factor is  $E(\mathbf{Y}\mathbf{Y}') = E\{(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})'\}$

$$\begin{aligned} &= E\{(\mathbf{X}\boldsymbol{\beta} + \mathbf{e})(\boldsymbol{\beta}'\mathbf{X}' + \mathbf{e}')\} \\ &= E\{\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \mathbf{X}\boldsymbol{\beta}\mathbf{e}' + \mathbf{e}\boldsymbol{\beta}'\mathbf{X}' + \mathbf{e}\mathbf{e}'\} \\ &= E\{\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\} + \mathbf{0} + \mathbf{0} + E(\mathbf{e}\mathbf{e}') \\ E(\mathbf{Y}\mathbf{Y}') &= \mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \sigma^2\mathbf{I} \end{aligned}$$

Inserting the above into  $\text{var}(\hat{\mathbf{e}})$  gives

$$\begin{aligned} \text{var}(\hat{\mathbf{e}}) &= (\mathbf{I} - \mathbf{H})(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \sigma^2\mathbf{I})(\mathbf{I} - \mathbf{H}) \\ &= [\mathbf{I}(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \sigma^2\mathbf{I}) - \mathbf{H}(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \sigma^2\mathbf{I})](\mathbf{I} - \mathbf{H}) \\ &= [\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}' + \sigma^2\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}') - \sigma^2\mathbf{H}](\mathbf{I} - \mathbf{H}) \\ &= \sigma^2(\mathbf{I} - \mathbf{H})(\mathbf{I} - \mathbf{H}) \\ \text{var}(\hat{\mathbf{e}}) &= \sigma^2(\mathbf{I} - \mathbf{H}) \end{aligned}$$



What's the rank of  $\text{var}(\hat{\mathbf{e}})$ ?



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

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

Put another way,  $\text{rank}(\mathbf{B}) = \text{rank}(\mathbf{A}) - \text{rank}(\mathbf{C})$ .

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

## ANOVA in matrix terms

Recall from Week 3 that

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

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

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

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

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

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

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* (second-degree polynomial).



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  $\text{rank}(\mathbf{I}) - \text{rank}(\frac{1}{n}\mathbf{J}) = n - 1$ .

- This is the number of degrees of freedom for SST

## Decomposing SST

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

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

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,

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

## 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  $\hat{\beta} = \mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ ,

$$\begin{aligned}\text{SSReg} &= \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{Y}'\frac{1}{n}\mathbf{J}\mathbf{Y} \quad \text{from slide 12} \\ &= \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\frac{1}{n}\mathbf{J}\mathbf{Y} \\ &= \mathbf{Y}'\underbrace{\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{H}}\mathbf{Y} - \mathbf{Y}'\frac{1}{n}\mathbf{J}\mathbf{Y} \\ &= \mathbf{Y}'\left(\mathbf{H} - \frac{1}{n}\mathbf{J}\right)\mathbf{Y}\end{aligned}$$

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

## Using RSS to estimate $\sigma^2$

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

$$\begin{aligned} E(\text{RSS}) &= E(\hat{\mathbf{e}}' \hat{\mathbf{e}}) \\ &= E\{\mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}\} && \text{from slide 13} \\ &= E\{\text{trace}[\mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}]\} && \text{since } \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y} \text{ is a scalar} \\ &= E\{\text{trace}[(\mathbf{I} - \mathbf{H})\mathbf{Y}\mathbf{Y}']\} && \text{since } \text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA}) \\ &= \text{trace}[(\mathbf{I} - \mathbf{H})E(\mathbf{Y}\mathbf{Y}')] \\ &= \text{trace}[(\mathbf{I} - \mathbf{H})(\sigma^2\mathbf{I} + \mathbf{X}\beta\beta'\mathbf{X}')] && \text{from slide 8} \\ &= \text{trace}[(\mathbf{I} - \mathbf{H})\sigma^2 + \mathbf{X}\beta\beta'\mathbf{X}' - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta\beta'\mathbf{X}'] \\ &= \text{trace}(\mathbf{I} - \mathbf{H})\sigma^2 \end{aligned}$$

$$E(\text{RSS}) = (n-2)\sigma^2$$

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

## The big picture

We have expressed our ANOVA identity in matrix form:

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



What does R have to say about this?



## ANOVA in R

The `anova` command is one way in R to produce an ANOVA table (see Week 3, slide 42), in addition to analysing it. For example, for the 654-point SLR problem in Assignment 2, question 1:

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

```
## 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:

```
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
```

# This week's lectures

- ▶ Poll regarding the Exam Jam on 8 December
- ▶ Chapter 5:
  - ▶ Matrix version of SLR
  - ▶ **Multiple linear regression (MLR)**



## Multiple regression

Multiple regression is used when we have more than one explanatory variable. Multiple  $x$ 's can arise naturally. In addition, sometimes we want to:

- ▶ Control for some  $x$ 's to consider the effect on  $y$  of other  $x$ 's over and above the control variables
- ▶ Fit a polynomial
- ▶ Compare the regression line for two or more groups

In multiple linear regression (MLR), generally we let  $p$  represent the number of explanatory variables in the model, i.e.

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i$$

for  $i \in \{1, \dots, n\}$ . How many parameters do we need to estimate?

And therefore, how many observations do we need at a minimum?

## Matrix version of MLR

Our main equation is unchanged:  $\mathbf{Y} = \mathbf{X}\beta + \mathbf{e}$

However, the **design matrix**  $\mathbf{X}$  and  $\beta$  are bigger:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & & X_{2p} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

A design matrix gives the explanatory variables (often without the column of 1's). Each row is an observation and each column corresponds to a different kind of variable.

## Gauss-Markov assumptions for MLR

The key equations are unchanged:

$$E(\mathbf{e}) = \mathbf{0} \quad \text{and} \quad \text{var}(\mathbf{e}) = \sigma^2 \mathbf{I}$$

For our inference methods (CIs etc), we need  $\mathbf{e}$  to have a multivariate normal distribution as before.

The expression for residuals is still  $\hat{\mathbf{e}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\mathbf{b} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ , where now we have

$$\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{pmatrix}$$

## Estimating $\sigma^2$ in MLR

Recall that

$$S^2 = \text{MSE} = \frac{\sum_{i=1}^n \hat{e}_i^2}{\text{d.f. of error}} = \frac{\hat{\mathbf{e}}' \hat{\mathbf{e}}}{\text{d.f. of error}}$$

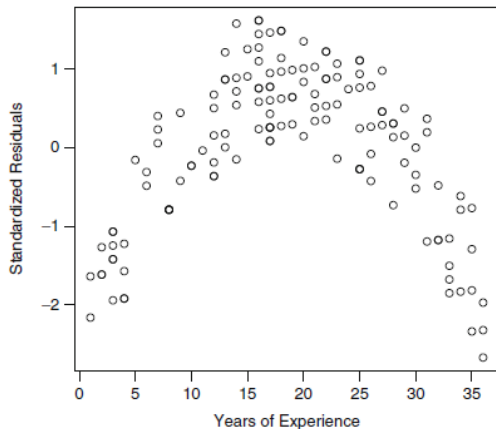
The degrees of freedom was  $n - 2$  in SLR, and is  $n - p - 1$  in MLR. To see this, recall that  $\text{RSS} = \hat{\mathbf{e}}' \hat{\mathbf{e}} = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$ . Using our five key properties of idempotent matrices again,  $\text{rank}(\mathbf{I} - \mathbf{H}) = \text{rank}(\mathbf{I}) - \text{rank}(\mathbf{H}) = n - (p + 1)$  assuming that the columns of  $\mathbf{X}$  are linearly independent.

To show that  $S^2$  is unbiased in MLR, similar to before we can show  $E(\text{RSS}) = (n - p - 1)\sigma^2$ . The proof is akin to the SLR proof except that:

$$\begin{aligned}\text{trace}(\mathbf{I} - \mathbf{H}) &= \text{trace}(\mathbf{I}) - \text{trace}(\mathbf{H}) \\ &= n - \text{trace}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= n - \text{trace}[\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= n - \text{trace}(\mathbf{I}_2) \\ &= n - (p + 1)\end{aligned}$$

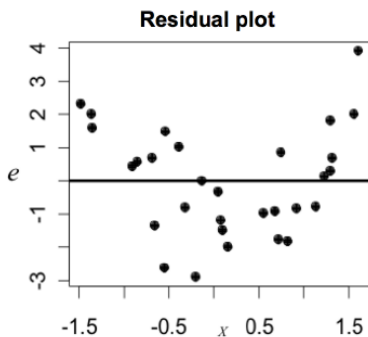
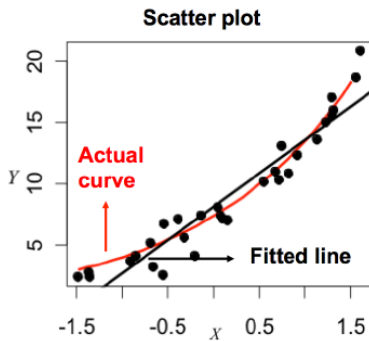
## Example of MLR: Fitting a polynomial

A professional-salary database contains 143 ordered pairs: (years of experience, salary). Generally, but not monotonically, salary increases with years of experience. Using SLR, our model is  $Y_i = \beta_0 + \beta_1 x_i + e_i$ . After fitting a straight line, this is the plot of standardized residuals:





## Example of a nonlinear relationship (*Weeks 4–5, Slide 41*)

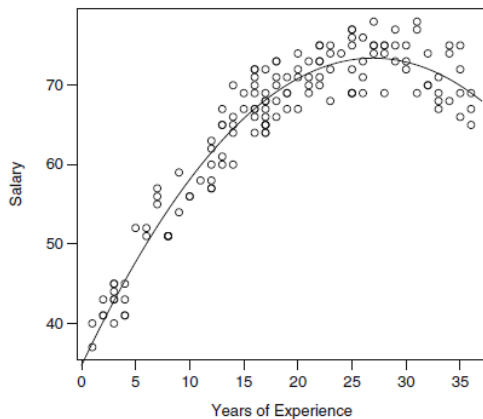


Remedial measure: If the regression function isn't linear,

- ▶ In some cases, a variable transformation can make the data “more linear”
- ▶ Otherwise, a different (e.g. nonlinear) model might be better

## Back to our salary database

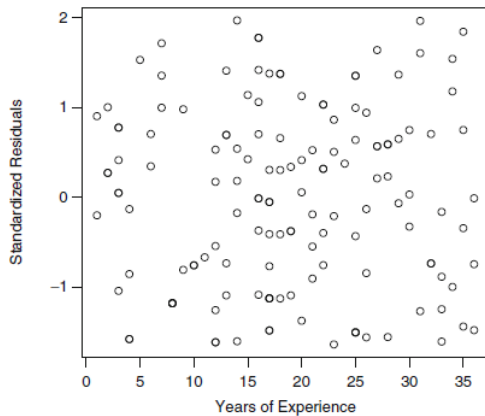
A simple nonlinear model is MLR in which we fit a parabola, i.e. incorporate  $x$  and  $x^2$ . The model is  $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i$  and the plot is:



Here,  $\beta_0 \approx 35$ ,  $\beta_1 \approx 2.87$ , and  $\beta_2 \approx -0.053$ , each with  $p < 2 \times 10^{-16}$ .

## MLR example: fitting a polynomial

The residuals no longer have a pattern:



## R code for MLR

```
X <- read.csv("profsalary.txt",sep="\t")
mod1 <- lm(Salary ~ Experience + I(Experience^2), data=X)
summary(mod1)
```



Typing `I(.)` is a way to express formulae within a call to `lm`.

The `+` sign indicates that more than one explanatory variable is being used. To have four variables, use e.g. `y ~ x1 + x2 + x3 + x4`

## R output for MLR

```
##
## Call:
## lm(formula = Salary ~ Experience + I(Experience^2), data = X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5786 -2.3573  0.0957  2.0171  5.5176
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   34.720498   0.828724   41.90  <2e-16 ***
## Experience     2.872275   0.095697   30.01  <2e-16 ***
## I(Experience^2) -0.053316   0.002477  -21.53  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.817 on 140 degrees of freedom
## Multiple R-squared:  0.9247, Adjusted R-squared:  0.9236
## F-statistic: 859.3 on 2 and 140 DF,  p-value: < 2.2e-16
```

## Using the R model

Interpolate at 5 years of experience:

```
e <- 5; mod1$coefficients%*%c(1,e,e^2)
```

```
##           [,1]  
## [1,] 47.74897
```

Alternatively, use the predict command:

```
predict(mod1,data.frame(Experience=5))
```

```
##           1  
## 47.74897
```

The data frame passed to predict names and initializes all of the information used towards making the predictor variables. Another example would be:

```
predict(lm(y~x1+x2),data.frame(x1=5,x2=3))
```

## Interpreting MLR coefficients

How should we interpret  $\beta_j$ , or similarly their estimates  $b_j$  — i.e. what's the meaning of the coefficients of MLR predictor variables?

In general,  $\beta_j$  is the change in the mean value of  $Y$  associated with a one-unit change in the predictor variable  $x_j$ , with *all other variables held constant*.

For our salary database example, this is impossible. The closest interpretations we can make are of this sort:

- ▶ If Experience increases from 5 years to 6 years, the estimated change in mean Salary is  $2.87 - 0.053(36 - 25) \approx 2.3$
- ▶ If Experience increases from 35 years to 36 years, the estimated change in mean Salary is  $2.87 - 0.053(36^2 - 35^2) \approx -0.9$

## Do we need a polynomial fit?

We can quantify whether the quadratic term is 0 or not using familiar hypothesis testing:

$$H_0 : \beta_2 = 0 \quad \text{vs} \quad H_a : \beta_2 \neq 0$$

**Exercise:** Try this on the salary database. What do you find?





Do we need the  $j$ th predictor?



In general, a test of  $H_0 : \beta_j = 0$  gives an indication of whether or not the  $j$ th predictor variable statistically significantly contributes to the estimation/prediction of  $Y$  *over and above* the other predictor variables.

That is, the test assumes that the other variables are in the model.

## Next steps

- ▶ Solutions to **HW2** were posted during the Study Break
- ▶ Complete Chapter 5's **question 1**
- ▶ The lecture on Tuesday **21 November** will start at 11:10 am (not 10:10 am) and finish at the usual time. Sections 1 and 2 will then be resynchronized
- ▶ **Further ahead:**
  - ▶ In Chapter 6, we won't cover Marginal Model Plots, Inverse Response Plots, or Box-Cox transformations
  - ▶ In Chapter 7, mainly we'll cover §7.2.3 and p. 252

