Lec 18 Stat 135

Quiz 4 ANOVA + Bonferroni test

Today Simple Linear Regression Sec 14.2

Note
There is no take home exam portion of the final since I was able to reserve our room Wed 9-12 for the final.
More about the final coming soon.

Chap 14

Often we assume there is a linear relationship between a normal response variable, y, and $K \geq 1$ independent predictor normal RVs, $X_1, \ldots, X_K$.
We have n observations of $y, X_1, \ldots, X_K$.
In the case where we have a single predictor x and we make a scatter diagram of our n data pts, the pts will be separated from a line by an error term.

Let $e_1, \ldots, e_n \overset{iid}{\sim} N(0, \sigma^2)$ be n independent error terms. We assume our error terms are homoscedastic meaning that the variance, σ^2, doesn't depend on x.

consider matrices:

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \quad \underbrace{x_0 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n\times 1} \quad x_1 = \begin{bmatrix} x_{12} \\ \vdots \\ x_{n2} \end{bmatrix}_{n\times 1} \cdots x_k = \begin{bmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{bmatrix}_{n\times 1}}_{X_{n\times(k+1)}} \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}_{n\times 1}$$

$$X = [x_0, \dots, x_k]_{n\times(k+1)}$$

design matrix

where $n \geq k+1$

X **full rank**

$x_0, x_1, \dots, x_k$ column vectors are indep. so $X'X$ is a $(k+1)\times(k+1)$ invertible matrix.

The standard statistical model is

$y = X\beta + e$ where e is **homoscedastic**.

this means that the variance of e_i doesn't depend on X.

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix}_{n\times(k+1)} \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}_{n\times 1}$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} \beta_0 + \beta_1 x_{11} + \cdots + \beta_k x_{1k} + e_1 \\ \vdots \\ \beta_0 + \beta_1 x_{n1} + \cdots + \beta_k x_{nk} + e_1 \end{bmatrix}_{n\times 1}$$

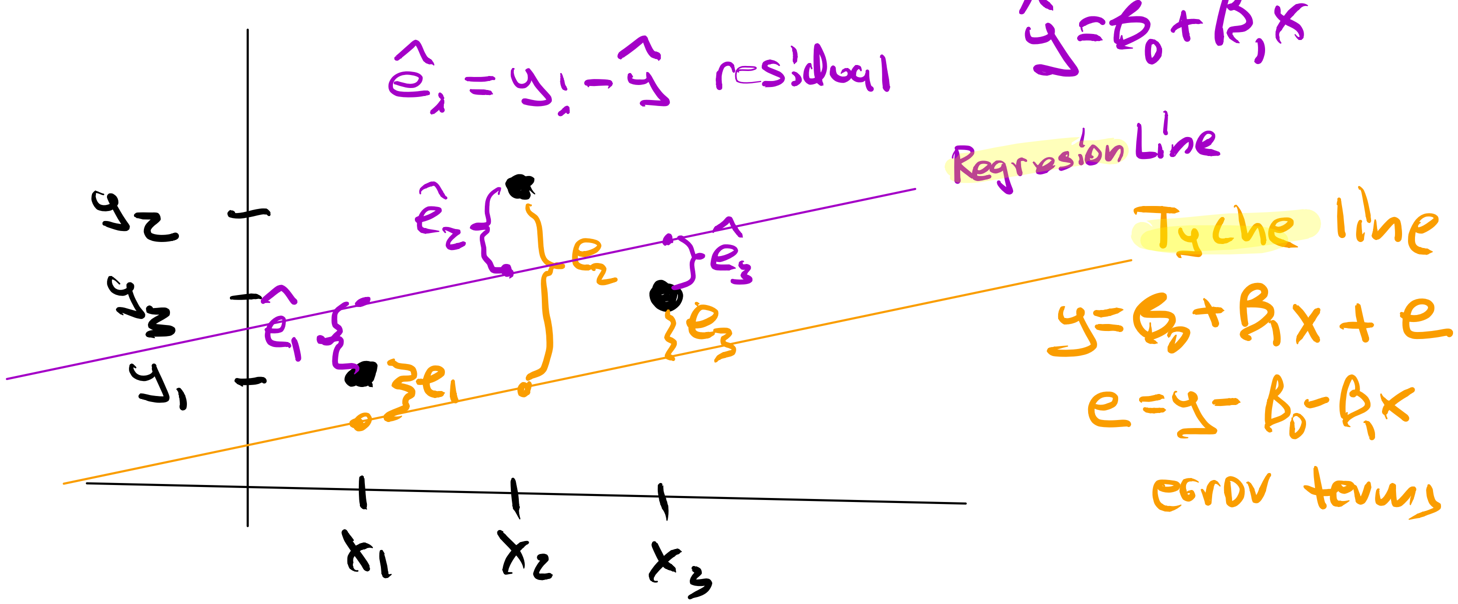
ex (Simple standard statistical model)

$K=1$

$n=3$

Lets assume there is a linear relationship between x = father's height and y = son's height

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



we wish to find the best fitting line through our data (reg line).

$$\hat{e}_i = y_i - \hat{y}_i \text{ are residuals,}$$
$$= y_i - \hat{\beta}_0 - \hat{\beta}_1 x$$

Since e_i are homoscedastic the residuals should also be approximately homoscedastic,

You should always plot $\hat{e}_i$, called a residual plot, and the vertical spread should be constant

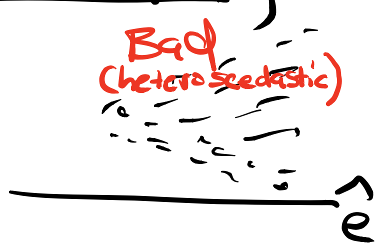
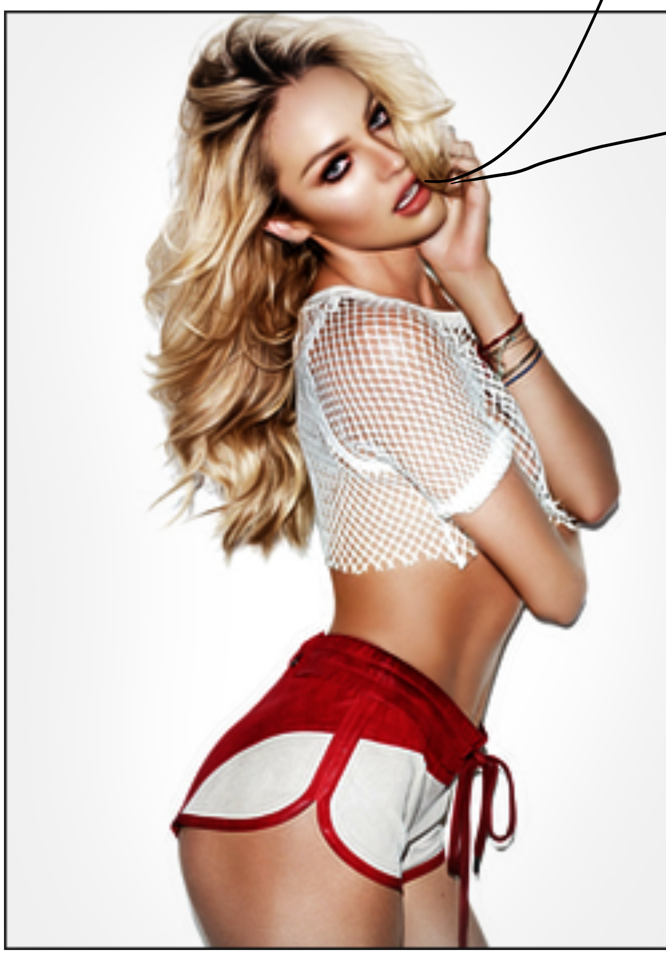


Image of Tyche the goddess of fortune.

I know B_0, B_1 !

An image of a modern day Tyche from sin city.

I know β_0, β_1!

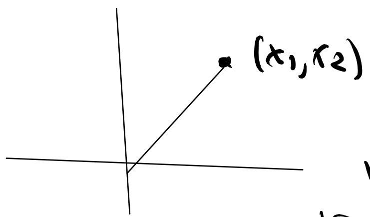


We want to minimize the sum of the squares of the residuals $\sum_{i=1} \hat{e}_i^2$ ⟵ Euclidean norm of the vector $\hat{e} = \begin{bmatrix} \hat{e}_1 \\ \vdots \\ \hat{e}_n \end{bmatrix}$

Background:

<u>Euclidean norm</u>

let $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be a vector in the plane



$$\|X\|^2 = X \cdot X = X'X$$
$$= [x_1\ x_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2$$

more generally

if $x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix}$, $\|x\|^2 = \sum_{i=1}^{k} x_i^2$

Two ways to find $\hat{B}$ (our estimates for B):

① Calculus: – method of least squares

We have a minimization problem with respect to $B_0, B_1, \ldots, B_k$.

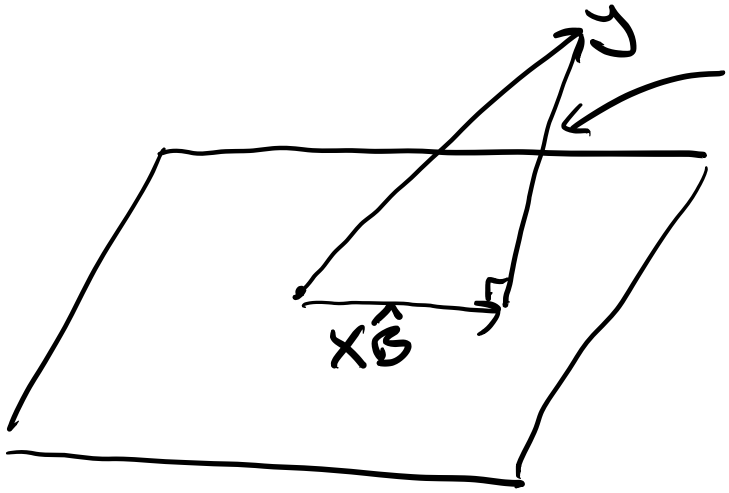
We have n simultaneous equations in $k+1$ variables w/ $n \geq k+1$.

② Geometrically (not in book)

<u>Geometric approach</u>

$y \in \mathbb{R}^n$

The columns of X span a $k+1$ dimensional subspace since X has full rank,



$y - X\hat{\beta} = \hat{e}$

column space of X is dimension $k+1 \le n$

vector of residuals

Find $\hat{\beta}$ such that $\|y - X\hat{\beta}\|$ is as small as possible.

perpendicular

To do this make $y - X\hat{\beta} \perp$ to the column space of X (i.e the dot product of every column of X with the vector $y - X\hat{\beta}$ is zero).

Notice that every pt in the column space of X can be written as

$$[x_0, \dots, x_k] \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix} = \beta_0 x_0 + \dots + \beta_k x_k .$$

so $$\underset{(k+1)\times n}{X'} \cdot \underset{n\times 1}{(y - X\hat{\beta})} = 0$$

$$X' = \begin{array}{c} x_0 \\ x_1 \\ \vdots \\ x_k \end{array} \begin{pmatrix} \text{———} \\ \text{———} \\ \vdots \\ \text{———} \end{pmatrix}_{(k+1)\times n}$$

this vector is perpendicular to all the rows of X' since it is perpendicular to all the columns of X.

$$\Rightarrow X'y - X'X\hat{\beta} = 0$$

$$\Rightarrow X'y = X'X\hat{\beta}$$

$$\Rightarrow \boxed{\hat{\beta} = (X'X)^{-1}X'y}$$

If $n = k+1$ so y lies in the column space of X then $\|y - X\hat{\beta}\| = 0$ and $y = X\hat{\beta}$ which says that the residuals are zero and the regression plane fits the data perfectly.

we show below that for $\underline{k=1}$ (simple case):

$$\hat{\beta}_1 = \frac{\mathrm{cov}(x,y)}{\mathrm{var}(x)}$$

$$\hat{\beta}_0 = E(y) - \hat{\beta}_1 E(x)$$

x_{11} $\quad$ x_{n1}

here $x = \mathrm{unif}(x_1, \dots, x_n)$

$y = \mathrm{unif}(y_1, \dots, y_n)$

so $E(x) = \bar{x}$

$E(y) = \bar{y}$

$\mathrm{var}(x) = \frac{1}{n}\sum(x_i - \bar{x})^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2$

$\mathrm{cov}(x,y) = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})$

EXAMPLE B Returning to Example A on fitting a straight line, we have

$\hat{\beta} = (X'X)^{-1}XY$

where

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix}$$

$n^2 var(x) \rightarrow$

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

Thus,

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

$$= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

$$= \frac{1}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

$n^2 Var(x) = \frac{\sum x^2}{n} - \left(\frac{\sum x_i}{n}\right)^2$

$= n\sum x^2 - (\sum x_i)^2$ ✓

$$= \frac{1}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \begin{bmatrix} \left(\sum_{i=1}^n y_i\right)\left(\sum_{i=1}^n x_i^2\right) - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n x_i y_i\right) \\ n\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right) \end{bmatrix}$$

$n^2 var(x) \rightarrow$

n

$n^2 cov(x,y)$

which agrees with the earlier calculation. ■

so $\boxed{\hat{\beta}_1 = \frac{cov(x,y)}{var(x)}}$ ✓

We show that $\hat{\beta}_0 = E(y) - \hat{\beta}_1 E(x)$.

$$\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i = \sum y_i \sum x_i^2 - \frac{\sum y_i (\sum x_i)^2}{n} - \sum x_i \sum x_i y_i + \frac{\sum y_i (\sum x_i)^2}{n}$$

$$= \sum y_i \left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right] - \left[\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}\right] \sum x_i$$

$$= n^2 E(y)\, var(x) - n^2 cov(x,y) E(x)$$

Hence

$$\hat{\beta}_0 = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{n^2 E(y) var(x) - n^2 cov(x,y) E(x)}{n^2 var(x)}$$

$$= E(y) - \frac{cov(x,y)}{var(x)} E(x)$$ $= \hat{\beta}_1$

so $\boxed{\hat{\beta}_0 = E(y) - \hat{\beta}_1 E(x)}$ ✓

break

Calculus approach (we will explore this only for k=1 case):

Using calculus for k=1 case we have,

$S(\beta_0, \beta_1) = \sum (y_i - \beta_0 - \beta_1 x_i)$

To find β_0 and β_1, we calculate

$$\left.\begin{aligned}
\frac{\partial S}{\partial \beta_0} &= -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) = 0 \implies \sum \hat{e}_i = 0 \\
\frac{\partial S}{\partial \beta_1} &= -2 \sum_{i=1}^{n} x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \implies \sum x_i \hat{e}_i = 0
\end{aligned}\right\} \text{called normal eqns.}$$

Setting these partial derivatives equal to zero, we have that the minimizers $\hat{\beta}_0$ and $\hat{\beta}_1$ satisfy

$$\sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i y_i = \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2$$

Solving for $\hat{\beta}_0$ and $\hat{\beta}_1$, we obtain

$$\hat{\beta}_0 = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Problem 10 at the end of the chapter asks you to derive the following useful equivalent expressions:

$E(y) - \hat{\beta}_1 E(x)$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

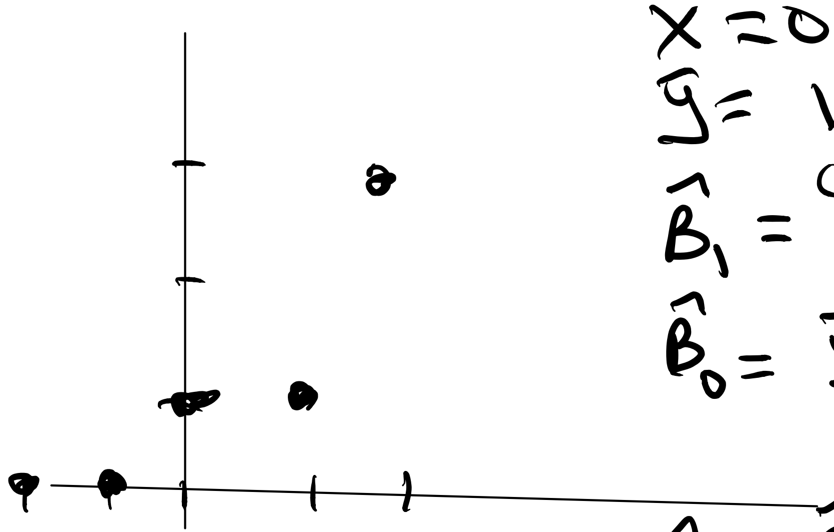
$\frac{\mathrm{cov}(x,y)}{\mathrm{var}(x)}$

Ex use the method of least squares to fit $n=5$ data pts

$(-2,0)$, $(-1,0)$, $(0,1)$, $(1,1)$, $(2,3)$

Plot pts

Compute $\hat{\beta}_0$, $\hat{\beta}_1$ and find regression line.



$$\bar{x} = 0$$

$$\bar{y} = 1$$

$$\hat{\beta}_1 = \frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{7}{10}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 1 - \frac{7}{10} \cdot 0 = 1$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\boxed{\hat{y} = 1 + \frac{7}{10} x}$$

on HW6 you will show reg line can be written as

$$\frac{\hat{y} - \bar{y}}{\sigma_y} = r \left(\frac{x - \bar{x}}{\sigma_x} \right)$$

correlation coeff of x and y.

Here is how to do this in R:

Stat135 lecture 18

Simple Linear Regression

Start Over

in-class exercise lec 18 on bcourses.

Simple Linear Regression

Here is a basic example of how to draw a linear regression line in R

```
x <- -2:2
y <- c(0,0,1,1,3)
df <- data.frame(x,y)
df
```

x <int>	y <dbl>
-2	0
-1	0
0	1
1	1
2	3

5 rows

```
reg <- lm(formula=y~x)
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##          1          2          3          4          5
##  4.000e-01 -3.000e-01 -2.776e-16 -7.000e-01  6.000e-01
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.0000     0.2708   3.693   0.0345 *
## x             0.7000     0.1915   3.656   0.0354 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6055 on 3 degrees of freedom
## Multiple R-squared:  0.8167, Adjusted R-squared:  0.7556
## F-statistic: 13.36 on 1 and 3 DF,  p-value: 0.03535
```

```
reg$coefficients
```

```
## (Intercept)           x
##         1.0         0.7
```

Lets find the expected y value for an x value of 3:

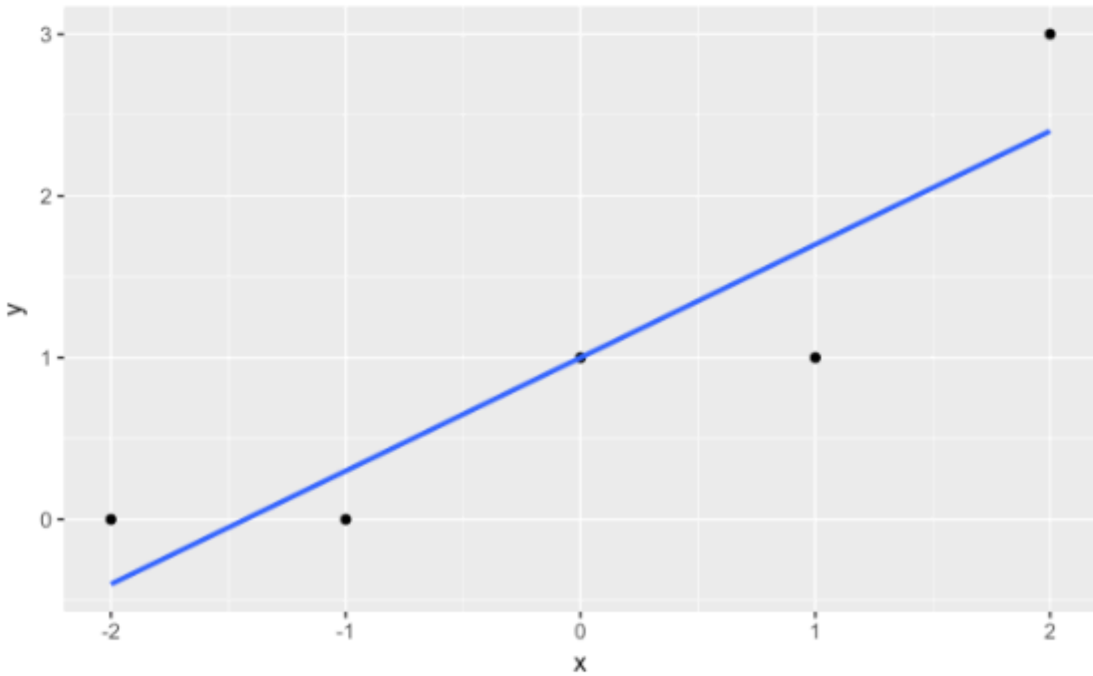
"1" "·7"

```
reg$coefficients[1]+ 3*reg$coefficients[2]
```

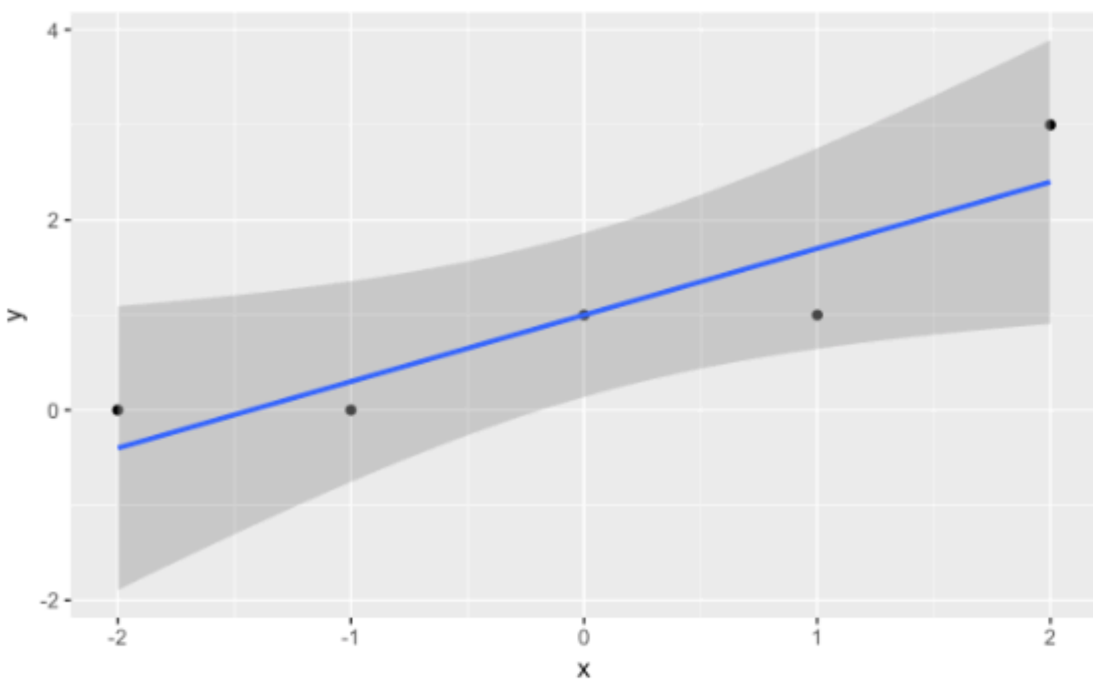
$= 1+3(.7)$

```
## (Intercept)
##         3.1
```

```
df %>% ggplot(aes(x=x,y=y)) + geom_point() + geom_smooth(method=lm,se=FALSE)
```




```
df %>% ggplot(aes(x=x,y=y)) + geom_point() + geom_smooth(method=lm)
```



example of drawing a regression line

There are two variables age and height. ages: 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29
heights: 76.1, 77, 78.1, 78.2, 78.8, 79.7, 79.9, 81.1, 81.2, 81.8, 82.8, 83.5

Find the equation of the regression line predicting height for different ages (i.e. x=ages, y=heights).
Predict the height of a 1 year old. Draw the regression line and the points (no error bars).

Code | Start Over | Solution | Run Code

```
1
2
3
```

we need to derive some machinery:

We need to develop some machinery.

A vector valued RV is called a <u>random vector</u>

ex $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

we can generalize covariance of a random variable to covariance of a random vector.

let $\hat{B} = \begin{bmatrix} \hat{B}_0 \\ \vdots \\ \hat{B}_k \end{bmatrix}$ be a random vector.

then $E(\hat{B}) = \begin{bmatrix} E(\hat{B}_0) \\ \vdots \\ E(\hat{B}_k) \end{bmatrix}$

called <u>covariance matrix</u>

$\mathrm{Var}(\hat{B}) = \begin{bmatrix} & & \\ & \mathrm{Cov}(\hat{B}_i, \hat{B}_j) & \\ & & \end{bmatrix}$ $\begin{matrix} i = 0, \dots, k \\ j = 0, \dots, k \end{matrix}$

$\Sigma_{\hat{B}\hat{B}}$ in Rice.

<u>Fact</u> Thm A, B p 573-574 — proof is straight-forward matrix algebra

$E(\hat{B}) = B$ ($\hat{B}$ is unbiased)

$E(e) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ and $\mathrm{Var}(e) = \begin{bmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{bmatrix}$

$\mathrm{Var}(\hat{B}) = \underbrace{\sigma^2 (X'X)^{-1}}_{(k+1) \times (k+1)}$ covariance matrix

We will focus now on the $k=1$ simple case.

Lets see what $\mathrm{var}(\hat{B})$ is in this case:

P574 Rice

EXAMPLE A We return to the case of fitting a straight line. From the computation of $(\mathbf{X}^T\mathbf{X})^{-1}$ in Example B in Section 14.3, we have

$$\Sigma_{\hat{\beta}\hat{\beta}} = \frac{\sigma^2}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \begin{bmatrix} \sum_{i=1}^n x_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i & n \end{bmatrix}$$

Therefore,

algebra see below*

$$\operatorname{Var}(\hat{\beta}_0) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{\underbrace{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}_{n^2 \operatorname{var}(x)}} = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{n\operatorname{var}(x)}\right]$$

$$\operatorname{Var}(\hat{\beta}_1) = \frac{n\sigma^2}{\underbrace{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}_{n^2 \operatorname{var}(x)}} = \frac{\sigma^2}{n\operatorname{var}(x)}$$

$$\operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2 \sum_{i=1}^n x_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{-\sigma^2\bar{x}}{\operatorname{var}(x)} \quad \blacksquare$$

* algebra:

$$\mathrm{var}(x) = \frac{1}{n}\sum x_i^2 - \bar{x}^2$$

$$\Rightarrow \sum x_i^2 = n\,\mathrm{var}(x) + n\bar{x}^2$$

$$\Rightarrow \mathrm{var}(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n^2\,\mathrm{var}(x)} = \sigma^2\left(\frac{n\,\mathrm{var}(x) + n\bar{x}^2}{n^2\,\mathrm{var}(x)}\right)$$

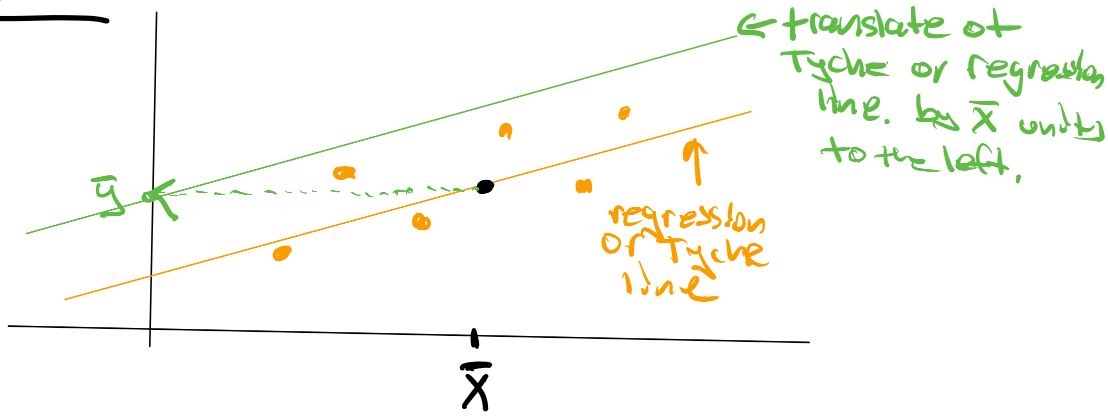
$$= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{n\,\mathrm{var}(x)}\right)$$

<u>Estimating σ^2</u> (not shown in book for $k=1$ case):

Define $RSS = \sum_{i=1}^{n} \hat{e}_i^2$ residual sum of squares

Lets normalize our data so that $\bar{x}=0$. This amounts to translating our scatter diagram horizontally by $\bar{x}$ units so the center of our data is now 0. This doesnt change σ^2 or RSS since the Tyche and Regression line move with the pts.

<u>Picture</u>



we will show now that $\frac{\sum \hat{e}_i^2}{\sigma^2} \sim \chi^2_{n-2}$

by first showing that $\frac{\sum e_i^2}{\sigma^2} \sim \chi^2_n$:

$$\sum e_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

we add and subtract $\hat{\beta}_0, \hat{\beta}_1 x_i$

$$= \sum_{i=1}^{n} \Big[\underbrace{(\hat{\beta}_0 - \beta_0)}_{A} + \underbrace{(\hat{\beta}_1 - \beta_1) x_i}_{B} + \underbrace{(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)}_{C = \hat{e}_i} \Big]^2$$

$$= \sum_{i=1}^{n} \left[A^2 + B^2 + C^2 + 2AB + 2AC + 2BC \right]$$

Note that $\sum_{i=1}^{n} 2AB = 0$ since $\bar{x} = 0$

$\sum_{i=1}^{n} 2AC = 0$ since $\sum_{i=1}^{n} \hat{e}_i = 0$

$\sum_{i=1}^{n} 2BC = 0$ since $\sum_{i=1}^{n} x_i \hat{e}_i = 0$

so

$$\frac{\sum_{i=1}^{n} (e_i)^2}{\sigma^2} = \frac{n(\hat{\beta}_0 - \beta_0)^2}{\sigma^2} + \frac{(\hat{\beta}_1 - \beta_1)^2 \overbrace{\sum x_i^2}^{n\,\mathrm{var}(x)}}{\sigma^2} + \frac{\overbrace{\sum \hat{e}_i^2}^{RSS}}{\sigma^2}$$

with $\bar{x} = 0$, $\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2}{n}\right)$

$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{n\,\mathrm{var}(x)}\right)$

Note $\hat{\beta} = (X'X)^{-1} X'Y$ is multivariate normal since it is a linear combination of normals (Y is normal and X is constant).

Hence

$$\underbrace{\frac{\sum_{i=1}^{n} e_i^2}{\sigma^2}}_{\chi^2_n} = \underbrace{\frac{(\hat{B}_0 - B_0)^2}{\frac{\sigma^2}{n}}}_{\chi^2_1} + \underbrace{\frac{\hat{B}_1 - B_1}{\frac{\sigma^2}{n\,var(x)}}}_{\chi^2_1} + \frac{RSS}{\sigma^2}$$

We next show that $\hat{B}_1$ and $\hat{B}_0$ are independent when $\bar{X}=0$.

$\bar{X}=0$

Recall that $Cov(\hat{B}_0, \hat{B}_1) = \dfrac{-\sigma^2 \bar{X}}{var(x)} = 0$

so $(\hat{B}_0, \hat{B}_1)$ are uncorrelated bivariate normal $\Rightarrow$ $\hat{B}_0$ and $\hat{B}_1$ are independent.

It follow that

$$\underbrace{\frac{(\hat{B}_0 - B_0)^2}{\frac{\sigma^2}{n}}}_{\chi^2_1} + \underbrace{\frac{\hat{B}_1 - B_1}{\frac{\sigma^2}{n\,var(x)}}}_{\chi^2_1} \sim \chi^2_2 \quad \text{and} \quad \frac{RSS}{\sigma^2} \sim \chi^2_{n-2}$$

we say RSS has $n-2$ degrees of freedom.

we know $E(\chi^2_{n-2}) = n-2$

$$\Rightarrow E\left(\frac{RSS}{\sigma^2}\right) = n-2$$

$$\Rightarrow E\left(\frac{RSS}{n-2}\right) = \sigma^2 \quad \text{so}$$

analogous to $\sum(x_i-\bar{x})^2$ which has $n-1$ d.f.

$S^2 = \frac{RSS}{n-2}$ is an unbiased estimator of σ^2.

Confidence interval for β_0, β_1

Recall from chapter 6

$Z \sim N(0,1)$

$U \sim \chi^2_n$

Z, U indep then $Z/\sqrt{U/n} \sim t_n$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{n\,\mathrm{var}(x)}\right)$$

let $Z = \dfrac{\hat{\beta}_1 - \beta_1}{\sqrt{\dfrac{\sigma^2}{n\,\mathrm{var}(x)}}}$ and $U = \dfrac{RSS}{\sigma^2} \sim \chi^2_{n-2}$

$$\Rightarrow t_{n-2} = \frac{Z}{\sqrt{\frac{U}{n-2}}} = \frac{\hat{B}_1 - B}{\sqrt{\frac{\cancel{\sigma^2}}{n\,var(x)}}} \cdot \frac{\sqrt{n-2}}{\sqrt{\frac{RSS}{\cancel{\sigma^2}}}}$$

$$= \frac{(\hat{B}_1 - B_1)\sqrt{n-2}}{\sqrt{\frac{RSS}{n\,var(x)}}} = \frac{(\hat{B}_1 - B_1)}{\frac{\sqrt{\frac{RSS}{n\,var(x)}}}{\sqrt{n-2}}}$$

$S_{\hat{B}_1}$

Define $S_{\hat{B}_1} = \sqrt{\frac{RSS}{n(n-2)\,var(x)}}$ to be ~~the~~ SE of $\hat{B}_1$

This allows us to find 95% CI for B_1 is

$$\hat{B}_1 \pm t_{n-2}(.025)\, S_{\hat{B}_1}$$

You can also find 95% CI for B_0

$$Var(\hat{B}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{n\,var(x)}\right)$$

$$\Rightarrow \hat{B}_0 \sim N\left(B_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{n\,var(x)}\right)\right)$$

$$\Rightarrow t_{n-2} = \frac{\hat{B}_0 - B_0}{\sqrt{s^2\left(\frac{1}{n} + \frac{\bar{x}^2}{n\,var(x)}\right)}} \text{ where } s^2 = \frac{RSS}{n-2}$$

$$\sqrt{s^2\left(\frac{1}{n} + \frac{\bar{x}^2}{n\,var(x)}\right)} = S_{\hat{B}_0}$$

and a 95% CI for B_0 is

$$\hat{B}_0 \pm t_{n-2}(.025)\, S_{\hat{B}_0} \quad \text{where} \quad S_{\hat{B}_0} = \sqrt{\frac{RSS}{n-2}\left(\frac{1}{n} + \frac{\bar{x}^2}{n\,var(x)}\right)}$$