POLS 6481. Research Design and Quantitative Methods II
Lecture 3. Properties of Simple Regression Estimates and Residuals
Readings: Wooldridge, *Introductory Econometrics 5e*, 2.3 + 2.5 + Appx.2.A

Outline:

1. Review of Lecture 2
2. Unbiasedness
3. Variances and Standard Errors of Important Regression Parameters
4. Example: *three data points*
5. The 'Coefficient of Determination'

POLS 6481. Research Design and Quantitative Methods II
Lecture 3. Properties of Simple Regression Estimates and Residuals
Readings: Wooldridge, *Introductory Econometrics 5e*, 2.3 + 2.5 + Appx.2.A

1. Review of Lecture 2

Define fitted values $\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} \cdot x_i$ for $i = 1, \ldots, n$ [2.20]

where $\widehat{\beta_1} = \dfrac{\sum(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$ [2.19]

and $\widehat{\beta_0} = \bar{y} - \beta_1 \bar{x}$ [2.17]

Define residuals $\hat{u}_i = y_i - \hat{y}_i$

$= y_i - \widehat{\beta_0} - \widehat{\beta_1} \cdot x_i$ [2.21]

Residuals satisfy three properties: $\sum \hat{u}_i = 0$ [2.30 , 2.60]

$\sum \hat{u}_i x_i = 0$ [2.31 , 2.60]

$\bar{y} = E(y|x = \bar{x}) = \widehat{\beta_0} + \widehat{\beta_1} \cdot \bar{x}$ (*line bisects* $(\bar{x}, \bar{y})$)

To show another property — that $\sum \hat{u}_i{}^2$ [2.22] is *minimized* — requires calculus [Appx 2A].

2. Unbiasedness

Three key conditions: Linearity in Parameters $y = \beta_0 + \beta_1 x + u$ [SLR.1]

Simple Random Sampling $\{(x_i, y_i): i = 1,\ldots, n\}$ [SLR.2]

Zero *Conditional* Mean Disturbance $E(u \mid x) = 0$ [SLR.4]

Combine SLR.1 and SLR.2: $y_i = \beta_0 + \beta_1 x_i + u_i$, $i = 1, \ldots, n$

Then add SLR.4: $E(y|x) = \widehat{\beta_0} + \widehat{\beta_1} \cdot x$

Theorem 2.1 $E\left(\widehat{\beta_1}\right) = \beta_1$ (see p. 48 for proof of $\widehat{\beta_1} = \beta_1 + \dfrac{cov(x,u)}{var(x)}$) [2.52]

$E\left(\widehat{\beta_0}\right) = \beta_0$ (see p. 49) [2.53]

POLS 6481. Research Design and Quantitative Methods II
Lecture 3. Properties of Simple Regression Estimates and Residuals
Readings: Wooldridge, *Introductory Econometrics 5e*, 2.3 + 2.5 + Appx.2.A

3. Variances and Standard Errors of Important Regression Parameters

Fourth key condition: Homoskedasticity (constant error variance): $var(u \mid x) = \sigma^2$ [SLR.5]

Estimated error variance $\hat{\sigma}^2 = \frac{1}{n-2} \sum \hat{u}_i{}^2$ a.k.a. Mean Squared Error [2.61]

Standard error of regression $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{\sum \widehat{u_i}^2}{n-2}}$ a.k.a. 'root MSE' [2.62]

Sampling variance of slope (β_1) $\text{var}(\widehat{\beta_1}) = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}$ [2.57]

Standard error of slope $\text{se}(\widehat{\beta_1}) = \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \bar{x})^2}}$ p. 56

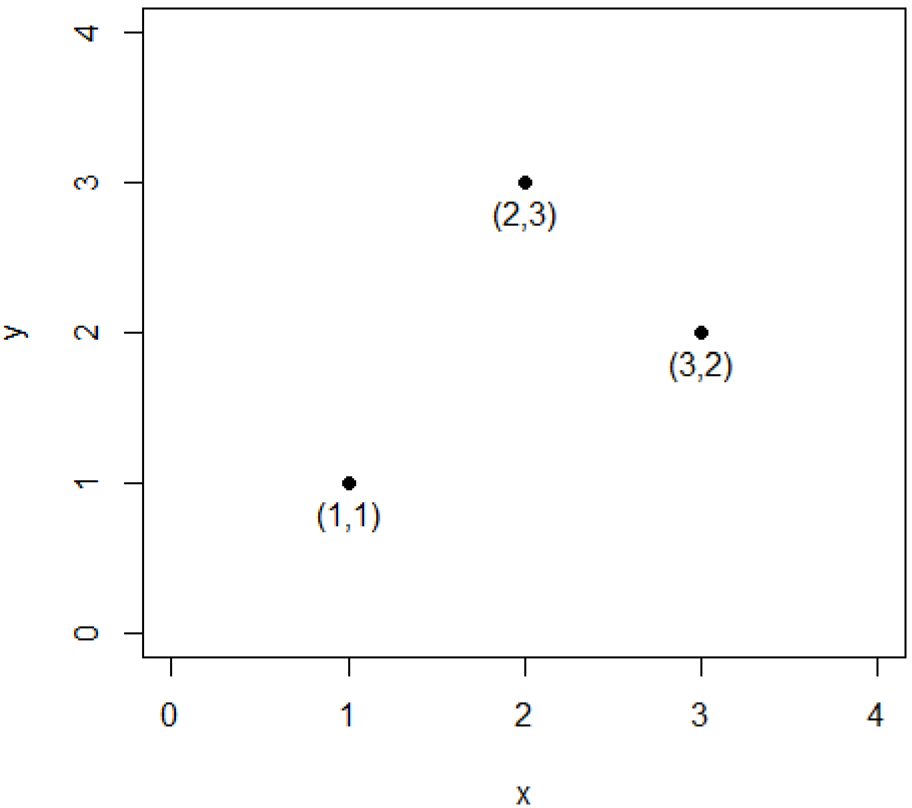
Sampling variance of intercept (β_0) $\text{var}(\widehat{\beta_0}) = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} \cdot \frac{\sum x_i^2}{n}$ [2.58]

Standard error of intercept $\text{se}(\widehat{\beta_0}) = \frac{\hat{\sigma}}{\sqrt{\sum (x_i - \bar{x})^2}} \cdot \sqrt{\frac{\sum x_i^2}{n}}$

4. Example: *three data points…*

Cartesian coordinates: {(1,1), (2,3), (3,2)} = $\{(x_1,y_1), (x_2,y_2), (x_3,y_3)\}$

```
x <- c(1,2,3)
y <- c(1,3,2)
cartesian <- c("(1,1)", "(2,3)", "(3,2)")
lim <- c(0,4)
plot(x,y, pch=16, xlim = lim, ylim = lim)
text(x,y, labels=cartesian, pos=1, xpd=TRUE)
```



4. Example: *three data points…*

Fill in the following table:

i	x_i	y_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	1	1	–1	1	–1	1	1
2	2	3	0	0	1	1	0
3	3	2	1	1	0	0	0
Σ	6	6		$SST_x = 2$		$SST_y = 2$	
Σ/n	$\bar{x} = 2$	$\bar{y} = 2$					
$\Sigma/(n\text{-}1)$				$var(x)=1$		$var(y)=1$	$cov(x,y)= ½$

Fill in the following equations:

Slope: $\widehat{\beta_1} = \frac{cov(x,y)}{var(x)} = \frac{1/2}{1} = 1/2$

Intercept: $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x} = 2 - 1/2 \cdot 2 = 1$

POLS 6481. Research Design and Quantitative Methods II
Lecture 3. Properties of Simple Regression Estimates and Residuals
Readings: Wooldridge, *Introductory Econometrics 3e*, 2.3 + 2.5 + Appx.2.A

4. Example: *three data points…*

Fill in the following table:

i	x_i	y_i	$\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i$	$\hat{u}_i = (y_i - \hat{y}_i)$	$\hat{u}_i \cdot x_i$	$\hat{u}_i^2 = (y_i - \hat{y}_i)^2$
1	1	1	1+(½)×1= 1½	1−1½ = −½	−½	¼
2	2	3	1+(½)×2= 2	3−2 = 1	2	1
3	3	2	1+(½)×3= 2½	2−2½ = −½	−1½	¼
Σ	6	6	$\Sigma\hat{y}_i = 6$	$\Sigma\hat{u}_i = 0$	$\Sigma\hat{u}_i \cdot x_i = 0$	SSR = 1½
Σ/n	$\bar{x} = 2$	$\bar{y} = 2$	$\bar{\hat{y}} = 2$			

Fill in the following equations:

sigma: $\sqrt{\widehat{\sigma^2}} = \sqrt{\frac{SSR}{n-2}} = \sqrt{(1½)/1} \cong 1.225$

$$\mathrm{se}(\widehat{\beta_1}) = \frac{\hat{\sigma}}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{\sqrt{1½}}{\sqrt{2}} = .866$$

$$\mathrm{se}(\widehat{\beta_0}) = \frac{\hat{\sigma}}{\sqrt{\Sigma(x_i - \bar{x})^2}} \cdot \sqrt{\frac{\Sigma x_i^2}{n}} = 1.871$$

R^2: $$\frac{SST_y - SSR}{SST_y} = \frac{2 - 1½}{2} = .25$$

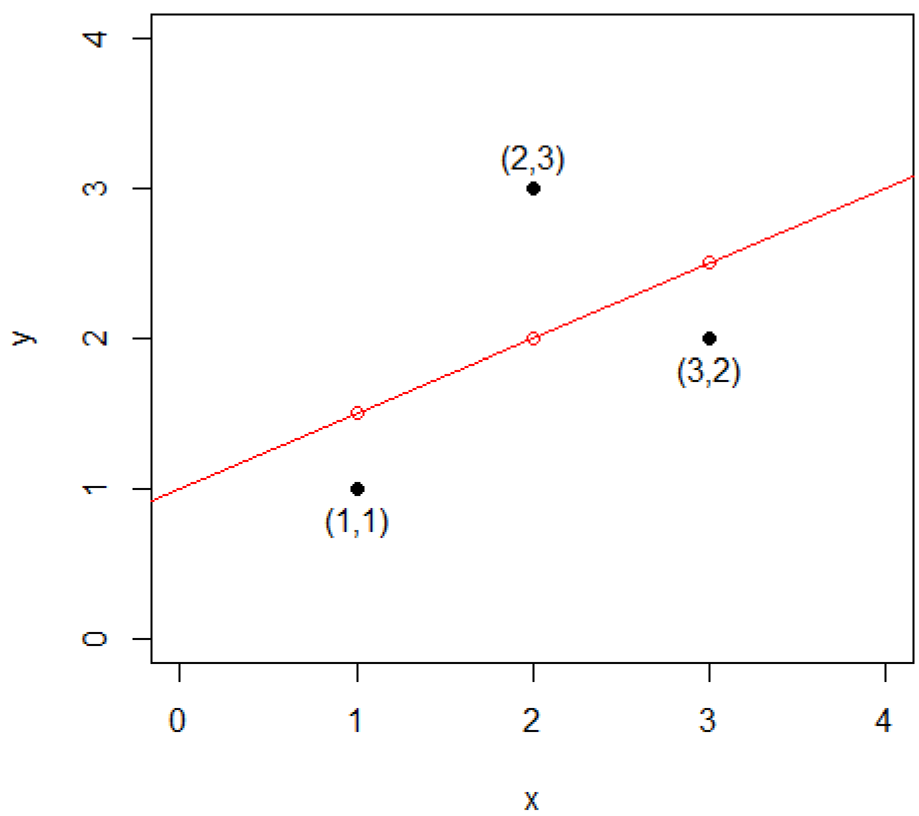
4. Example: *three data points…*

Use R to check results – estimated slope and intercept; graph regression line and plot fitted values; verify properties of residuals; verify standard errors; verify R^2

```
model <- lm(y~x)
summary(model)


plot(x,y, pch=16, xlim = c(0,4), ylim = c(0,4))
text(x,y, labels=cartesian, pos=c(1,3,1), xpd=TRUE)
abline(a=model$coefficients[1],b=model$coefficients[2], col="red")
points(x,model$fitted, col="red")


cbind(y, model$fitted, model$residuals, model$residuals*x)

round(sum(model$residuals), digits=2)
round(cor(model$residuals,x) , digits=2)
round(sum(model$residuals*x), digits=2)
```

4. Example: *three little data points…*

Check R results – estimated slope, intercept, residuals, and fitted values; sigma, standard errors, R^2



Readings: Wooldridge, *Introductory Econometrics 5e*, 2.3 + 2.5 + Appx.2.A

5. The 'Coefficient of Determination'

Total Sum of Squares	$\text{SST} = \sum(y_i - \bar{y})^2$	$= (n-1)\cdot var(y)$	[2.33]
Residual Sum of Squares	$\text{SSR} = \sum(y_i - \hat{y}_i)^2 = \sum \hat{u}_i{}^2$	$= (n-2)\cdot$M.S.E.	[2.35]
Explained Sum of Squares	$\text{SSE} = \sum(\hat{y}_i - \bar{y})^2$		[2.34]

Relationship: SST = SSR + SSE [2.36]

Coefficient of Determination $R^2 = 1 - \frac{SSR}{SST} = \frac{SSE}{SST}$ [2.38]

Note: $\text{SST}_x = \sum(x_i - \bar{x})^2$ (p. 48)