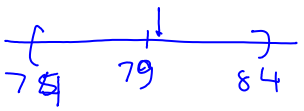
2013-11-19-EC01

Tuesday, November 19, 2013
6:16 PM



b) One-way ANOVA

✓ print

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/ANOVA-Calcs-Scan-Colour.pdf

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/Fertilizer.xls

One factor ANOVA						
	Mean	*n*	*Std. Dev*			
	3.0	6	1.90	Group 1		
	7.0	6	1.26	Group 2		
	5.0	6	1.41	Group 3		
	5.0	18	2.22	Total		
ANOVA table						
Source	*SS*	*df*	*MS*	*F*	*p-value*	
Treatment	48.00	2	24.000	10.00	.0017	
Error	36.00	15	2.400			
Total	84.00	17				
Post hoc analysis						
p-values for pairwise t-tests						
		Group 1	Group 3	Group 2		
		3.0	5.0	7.0		
Group 1	3.0					
Group 3	5.0	.0410				
Group 2	7.0	.0004	.0410			

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\Fertilizer-1.xls>

$H_0: \mu_i = \mu_j$

$H_a: \mu_i \neq \mu_j$

c) CI for $\mu_i - \mu_j$

Result: $100(1-\alpha)\%$ CI for $\mu_i - \mu_j$

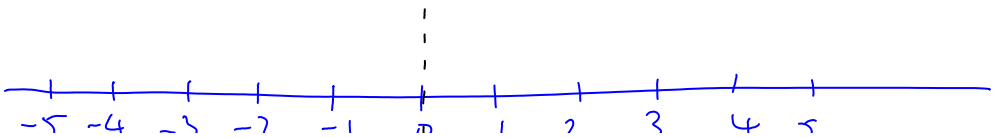
$$\left[(\bar{x}_i - \bar{x}_j) \mp t_{\alpha/2}\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}\right], \quad df = n-p$$

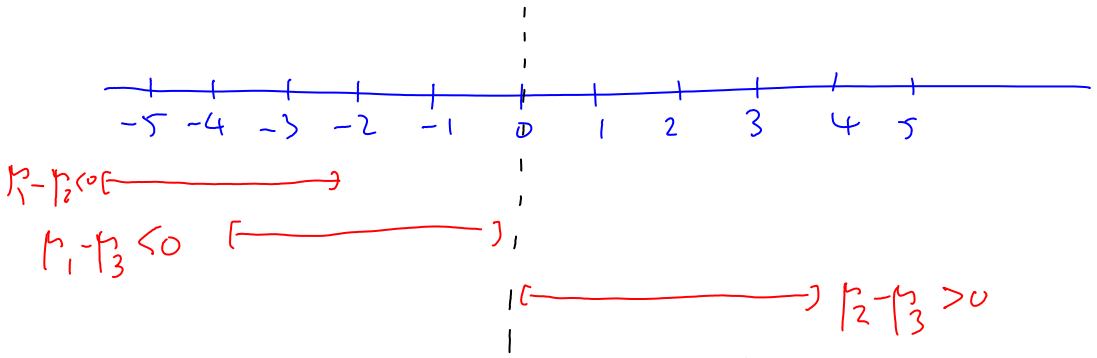
$\alpha = .05$

L-M: $\mu_1 - \mu_2$: $\left[(3-7) \mp 2.13\sqrt{2.4\left(\frac{1}{6}+\frac{1}{6}\right)}\right] = [-5.90, -2.09]$

L-H: $\mu_1 - \mu_3$: $[(3-5) \mp 2.13\sqrt{\quad}] = [-3.90, -0.09]$

M-H: $\mu_2 - \mu_3$: $[(7-5) \mp 2.13\sqrt{\quad}] = (0.09, 3.90)$

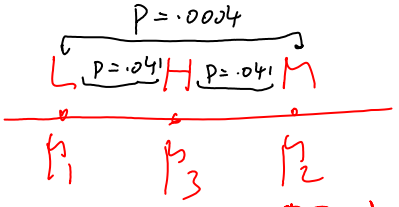




$\mu_1 - \mu_2 < 0 \Rightarrow \mu_1 < \mu_2$

$\mu_1 - \mu_3 < 0 \Rightarrow \mu_1 < \mu_3$

$\mu_2 - \mu_3 > 0 \Rightarrow \mu_2 > \mu_3$



TBS+ level!

42

Ch. 11 Correlation Coefficient & Simple linear regression

Two variables & how they relate

a) Covariance & correlation coefficient

Ex. Height vs. handspan

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/Q600-2013-Scanned-Height-Gender-Handspan.pdf

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/Q600-2013-Height-Gender-Handspan-Regression.xlsx

Ex. Olympic medals vs. econ. power (GDP)

11:8 2008-08-08 8:08 pm

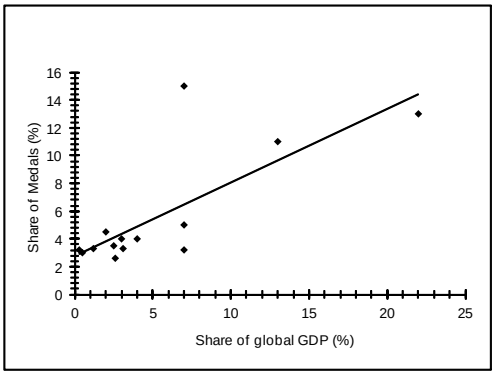
http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/MedalsAndEconomy.pdf

Country	Share of global GDP (%) (x)	Share of Medals (%) (y)
USA	22	13
China	13	11
Russia	7	15
Great Britain	7	5
Australia	2	4.5
Germany	4	4

France	3	4
Korea	2.5	3.5
Italy	3.1	3.3
Ukraine	1.2	3.3
Japan	7	3.2
Cuba	0.3	3.2
Belarus	0.5	3
Canada	2.6	2.6

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\MedalsAndEconomy_000-2.xls>

$\bar{x} = 5.37 \quad \bar{y} = 5.61$



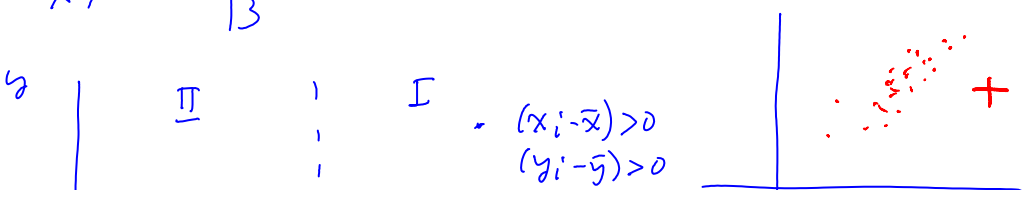
Variance $\quad S_x^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

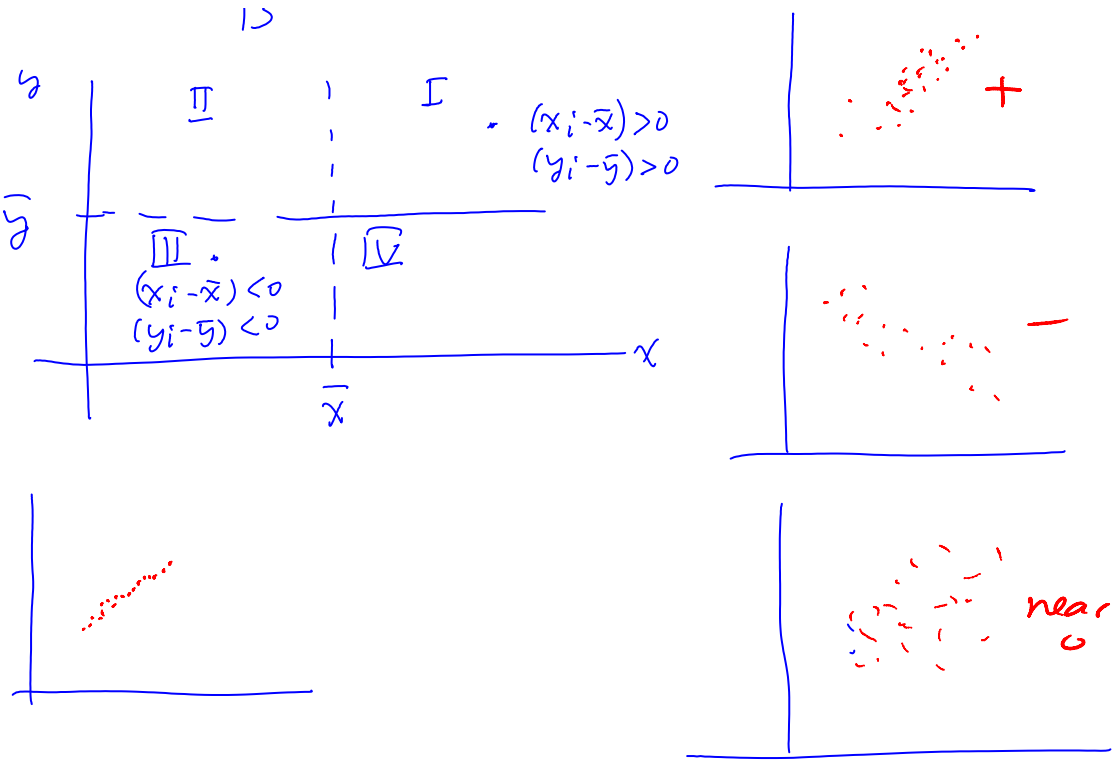
Covariance $\quad S_{xy} = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

$n = 14, \quad \bar{x} = 5.37, \quad \bar{y} = 5.61$

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
22	13	16.62	7.38	122.81
⋮	⋮	⋮	⋮	⋮
2.6	2.6	−2.77	−3.01	8.35
				238.36

$$S_{xy} = \frac{238.36}{13} = 18.33 > 0$$

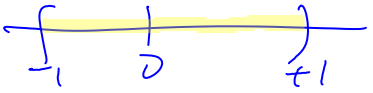




A better measure: Correlation coefficient r

$$r = \frac{S_{xy}}{S_x S_y}$$

Always in $[-1, 1]$



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

$$S_y = \sqrt{\frac{1}{n-1}\sum (y_i-\bar{y})^2}$$

Ex. Medals

$S_{xy} = 18.33$, $S_x = 5.88$, $S_y = 4.12$

Descriptive statistics	x	y
	Share of global GDP (%)	*Share of Medals (%)*
count	14	14
mean	5.371	5.614
sample variance	34.575	17.018
sample standard deviation	5.880	4.125
minimum	0.3	2.6
maximum	22	15
range	21.7	12.4

Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\MedalsAndEconomy_000-2.xls>

$$r = \frac{S_{xy}}{S_x S_y} = \frac{18.33}{5.88 \cdot 4.12} = 0.75$$

HT: $H_0: \rho = 0$ ⊥ $r\sqrt{n-2}$

HT: $H_0: \rho = 0$ $\quad H_a: \rho \neq 0$ $\quad t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

h3

b) Simple linear regression

↳ x ind't var } linear
y dep't "

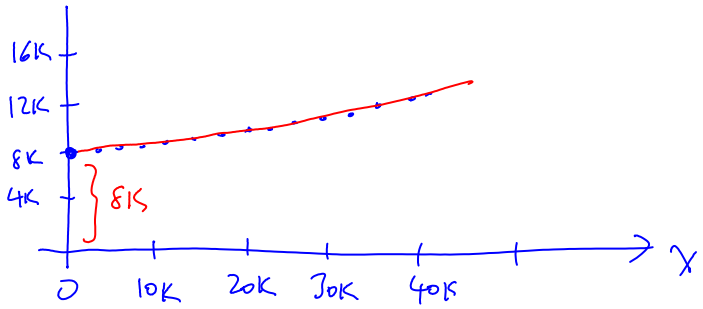
Ex. Naive model (no uncertainty)

x: sales at local Esso station (monthly)
gross

y: payment to parent comp. (")

Agreement: $8,000/month
+ 10% of gross

$y = 8000 + 0.10x$



Intercept = 8,000
slope = 0.10

In general $\quad y = \beta_0 + \beta_1 x$

true model ↗ β_0 ↓ intercept $\quad \beta_1$ ↘ slope

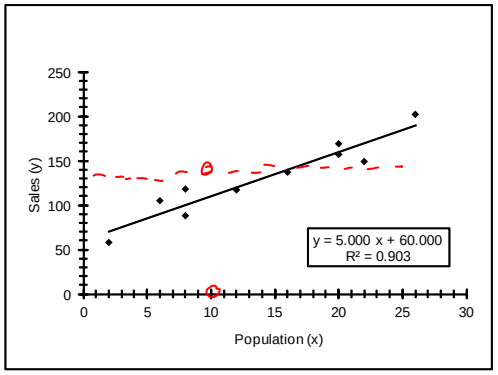
Ex. Statistical model – Harvey's restaurant

Estimate monthly revenue

Stud't pop $\longrightarrow$ sales revenue
x y

Data for 10 restaurants, only

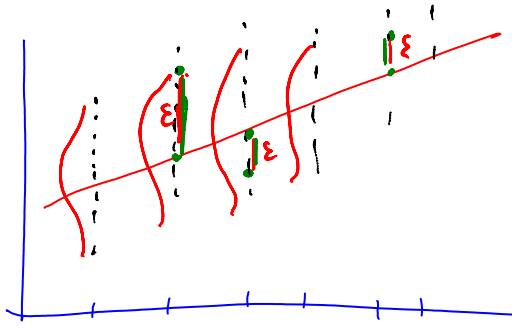
	Stu't pop	Monthly Sals ($1,000)
i	x_i (1000)	y_i
1	2	58
2	6	105
3	8	88
4	8	118
5	12	117
6	16	137
7	20	157
8	20	169
9	22	149
10	26	202
		$\bar{y} = 130$



$x = 10$
$\Rightarrow \bar{y} = 130$??

Statistical method needed!

Suppose we had data on all 300 Harvey's

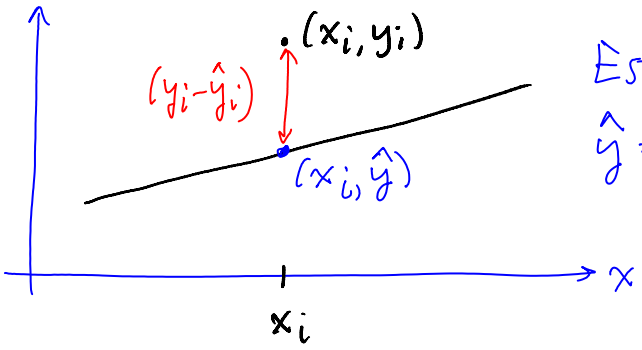


True model (error)
$Y = \beta_0 + \beta_1 X + \varepsilon$

Q: How do we estimate β_0, β_1

b_0 b_1

c) Best method to find regression line



Estimated line

$\hat{y} = b_0 + b_1 x$

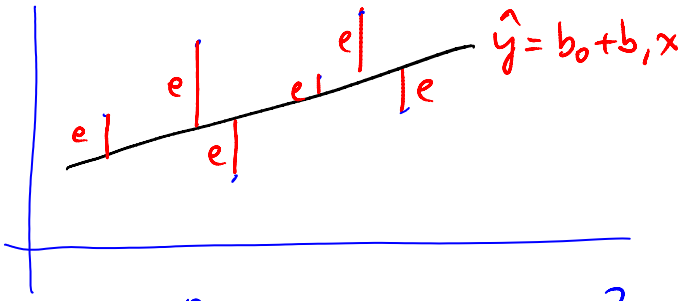
β_0 β_1

Consider (x_i, y_i): Actual y_i

Estimate $\hat{y}_i$

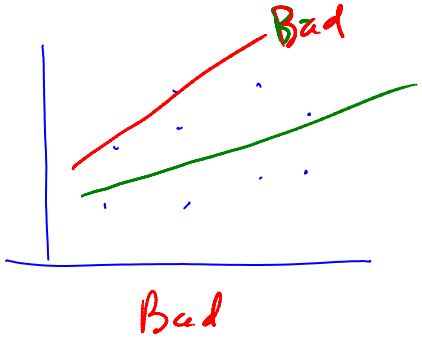
Residual (error) $e_i = y_i - \hat{y}_i$

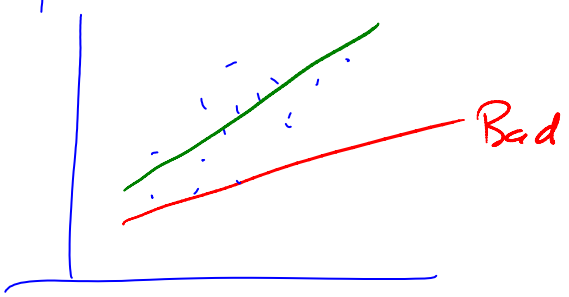
$= y_i - (b_0 + b_1 x_i)$



$$SSE = \sum_{i=1}^{n} [y_i - (b_0 + b_1 x_i)]^2 \leftarrow \text{minimize}$$

Problem: Find b_0 & b_1 that make SSE as small as possible

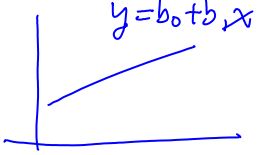




Solution ① $\boxed{b_1 = \frac{SS_{xy}}{SS_{xx}}}$

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

$$SS_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$



② $\boxed{b_0 = \bar{y} - b_1 \bar{x}}$

$$\bar{y} = \frac{\sum y_i}{n}, \quad \bar{x} = \frac{\sum x_i}{n}$$