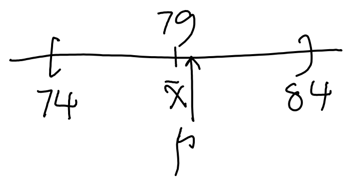
2013-11-21-C02

Thursday, November 21, 2013

10:21 AM

CI for exam μ



Assumptions ✓ for ANOVA

1. Pop. variances are equal
2. " normal
3. samples are independent

Ex. Farmer

✓ 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.xls>

c) Conf. Intervals for $\mu_i - \mu_j$

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

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

$\alpha = .05$

$\alpha = .05$

L-M $(\mu_1 - \mu_2)$

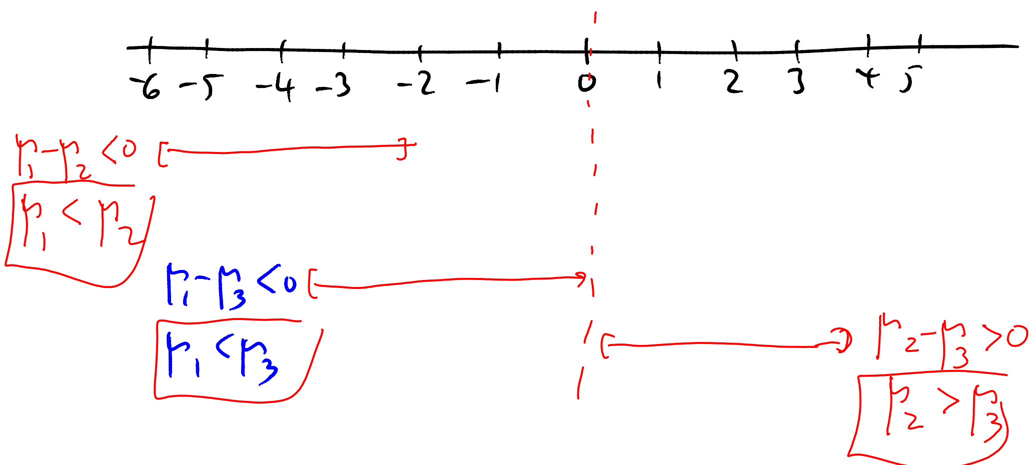
$$\left[(3-7) \mp 2.13\sqrt{2.4\left(\frac{1}{6}+\frac{1}{6}\right)}\right] = [-5.90, -2.09]$$

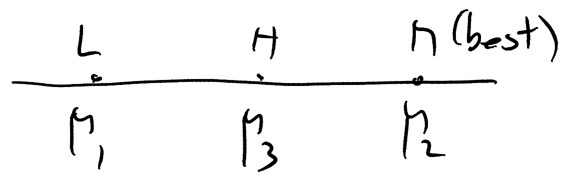
L-H $(\mu_1 - \mu_3)$

$$\left[(3-5) \mp 2.13\sqrt{\phantom{2.4\left(\frac{1}{6}+\frac{1}{6}\right)}}\right] = [-3.90, -0.09]$$

M-H $(\mu_2 - \mu_3)$

$$\left[(7-5) \mp 2.13\sqrt{\phantom{2.4\left(\frac{1}{6}+\frac{1}{6}\right)}}\right] = [0.09, 3.90]$$





Ch. 11 Correlation Coefficient & Simple Linear Regression

Two variables x y

$x \downarrow$ ind't

$y \downarrow$ dep't

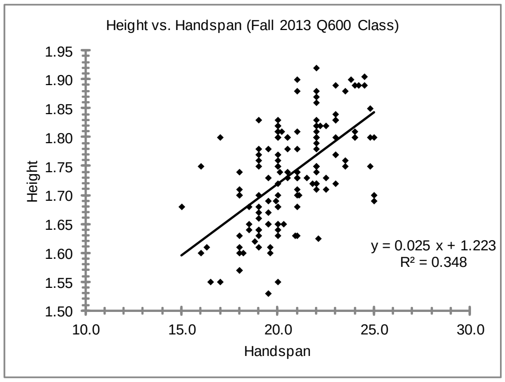
a) Covariance & Correlation Coefficient

a) Covariance & Correlation Coefficient

Ex. Handspan vs. height

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



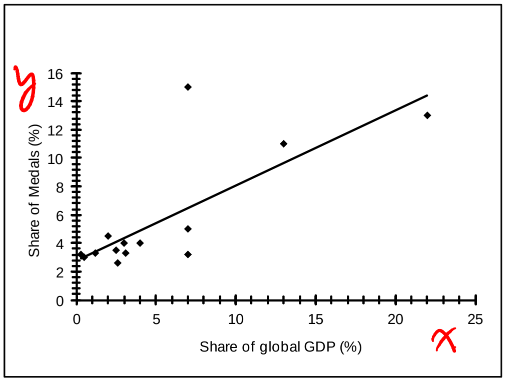
Ex. Olympic medals vs. GDP

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.xls>

Descriptive statistics		
	Share of global GDP (%)	*Share of Medals (%)*
count	14	14
mean	$\bar{x}$ 5.371	5.614 $\bar{y}$
sample variance	34.575	17.018
sample standard deviation	S_x 5.880	S_y 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.xls>



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

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

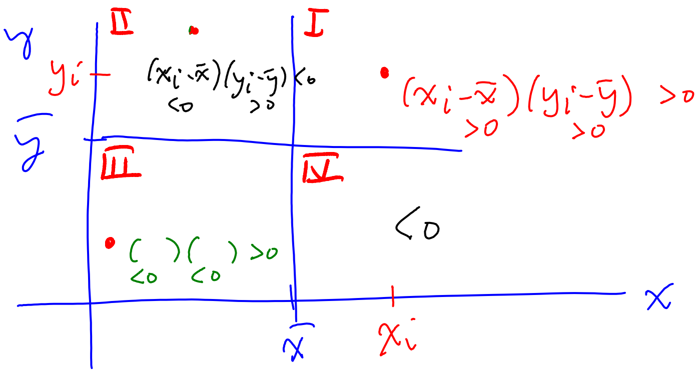
Co-variance (two ") $\quad S_{xy} = \dfrac{\sum (x_i - \bar{x})\cdot(y_i - \bar{y})}{n-1}$

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

$n = 14$

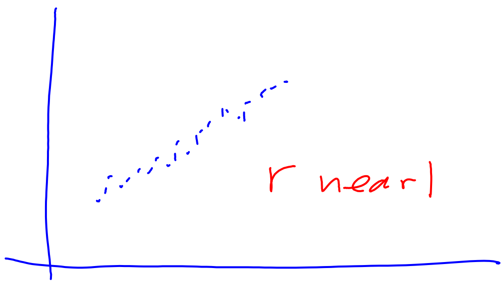
$$S_{xy} = \frac{238.36}{13} = 18.33 > 0$$

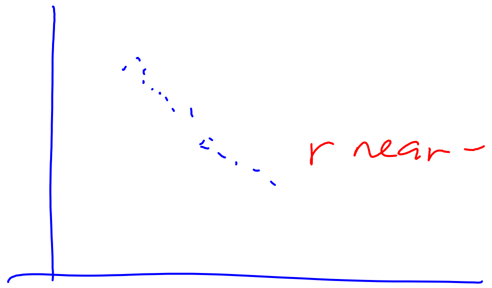
Sign of Covariance



A better measure : Correlation coefficient (r)

$$r = \frac{S_{xy}}{S_x S_y}$$ Always in $[-1, \; 0, \; +1]$





Ex. Medals

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

$$r = \frac{18.33}{(5.88)(4.12)} = 0.75$$ somewhat strong

[Correlation does not imply causation!]

b) Simple Linear Regression

Ex. Naïve model

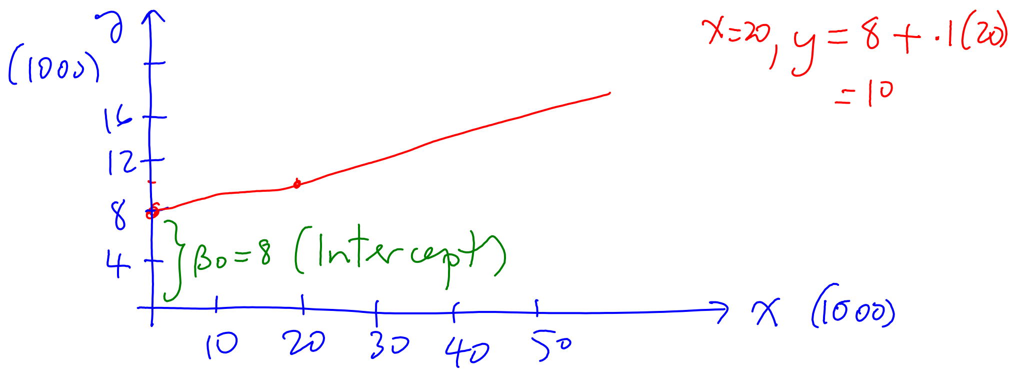
x: (gross) Sales at local Esso station

y: payment made to parent comp

Agreement: ① \$8,000/month

② 10% of gross

$y = 8000 + 0.10x$ Linear model



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

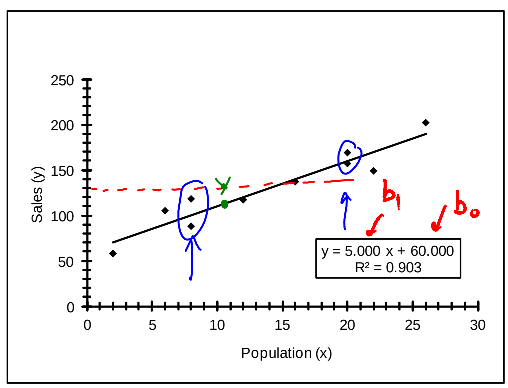
intercept slope

Ex. Statistical model - Harvey's restaurant

$\bar{y} = 130$

Student Population (x)	Monthly Sales (y)
2	58
6	105
8	88
8	118
12	117
16	137
20	157
20	169
22	149
26	202

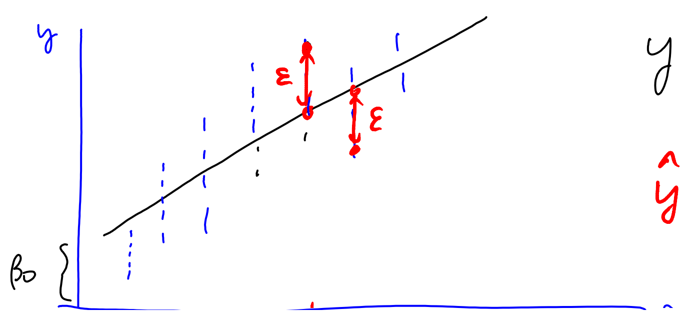
Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\Harveys.xls>



x=10, y=?

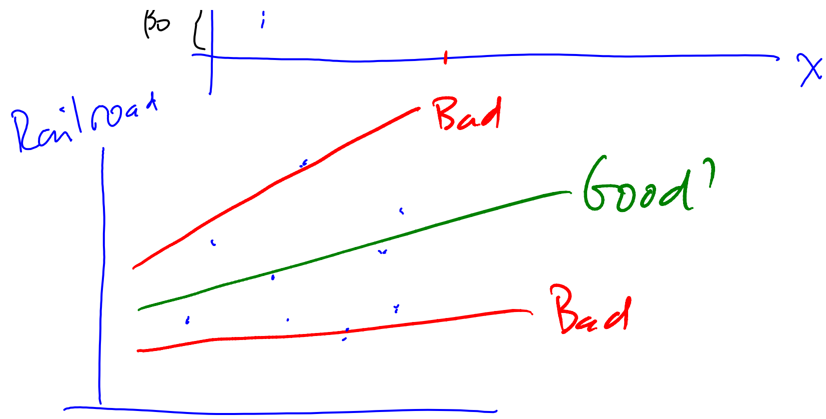
Need a statistical model!

Pop'n data available (300+)

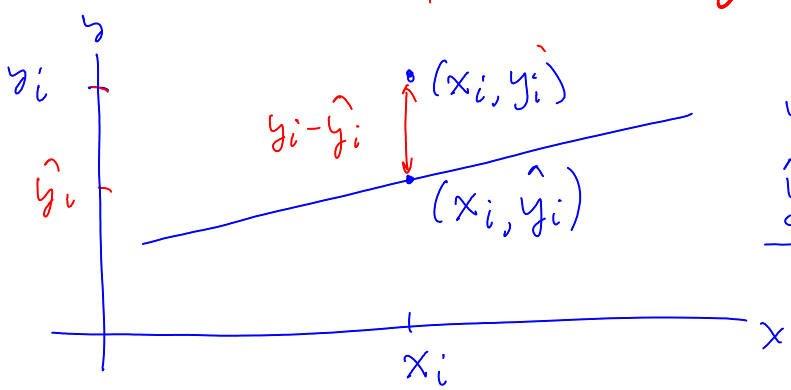


$y = \beta_0 + \beta_1 x + \varepsilon$

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



c) Best method to find the regression line



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

Consider

(x_i, y_i) : Actual y_i

− Estimate $\hat{y}_i$

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

$$e_i = y_i - \hat{y}_i = y_i - (b_0 + b_1 x_i)$$

min

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

? ?

Sol'n

1 $\quad b_1 = \dfrac{SS_{xy}}{SS_{xx}}$

$$SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum y_i\right)$$

$$SS_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2$$

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

$$\bar{y} = \frac{1}{n}\sum y_i, \quad \bar{x} = \frac{1}{n}\sum x_i$$