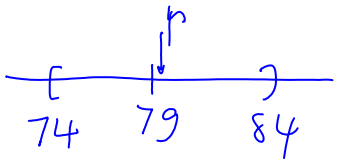
2013-11-20-C01

Wednesday, November 20, 2013
1:20 PM

CI



Assumptions for ANOVA

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

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

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. Int. for $\mu_i - \mu_j$

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

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

L-M (4-5)

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

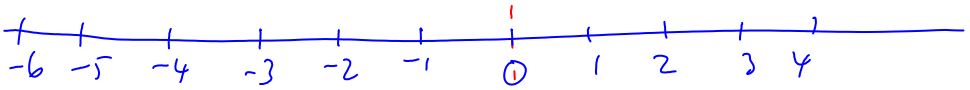
$\left[(3-7) \mp 2.13\sqrt{2.4\left(\frac{1}{6}+\frac{1}{6}\right)}\right] = [-5.90, -2.09)$

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

$[(3-5) \mp \quad \cdot \quad - \qquad] = [-3.90, -0.09)$

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

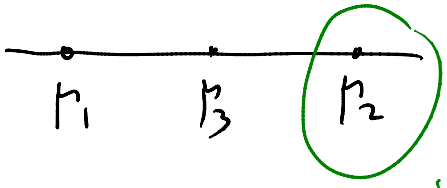
$[(7.5) \mp \quad \cdot \quad \cdot \qquad) = [0.09, 3.90)$



$\mu_1-\mu_2<0$ ~~$\mu_1<\mu_2$~~ ✓

$\mu_1-\mu_3<0$ ~~$\mu_1<\mu_3$~~ ✓

~~$\mu_2>\mu_3$~~ ✓ $\mu_2-\mu_3>0$



M is best!

Megastat $H_0: \mu_i = \mu_j$

$H_a: \mu_i \neq \mu_j$

Post hoc analysis					
p-values for pairwise t-tests		L			
		Group 1	Group 3	Group 2	
		3.0	5.0	7.0	
Group 1	3.0				
Group 3	5.0	.0410			
M Group 2	7.0	.0004	.0410		

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

$\mu_1 \neq \mu_2$, etc.

Ch.11 Correlation Coefficient & Simple linear regression

Two variables & how they relate

a) Covariance & correlation coefficient

Ex, Height vs. handspan

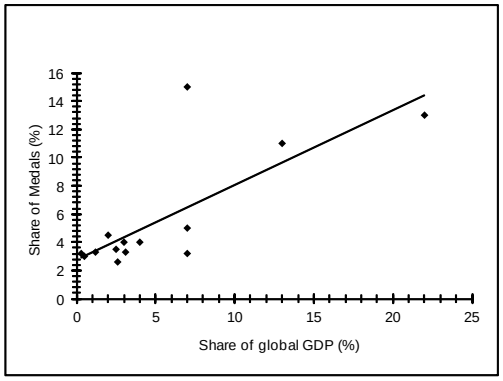
http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/Q600-2013-Scanned-Height-Gender-Handspan.pdf

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

Ex. Olympic medals vs GDP

2008-08-08 8:08 pm

http://profs.degroote.mcmaster.ca/ads/parlar/courses/q600/ChapterComments/documents/MedalsAndEconomy_000.xls



	x	y
Country	Share of global GDP (%)	Share of Medals (%)
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>

Covariance

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

Co-variance (two vars) $S_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

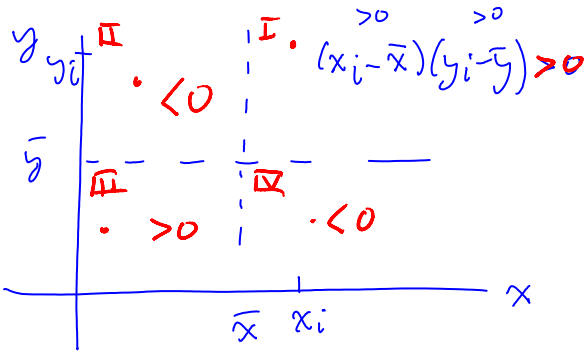
	Share of global GDP (%)	Share of Medals (%)
count	14	14
mean	$\bar{x}$ 5.371	$\bar{y}$ 5.614
sample variance	34.575	17.018
sample standard deviation	5.880	4.125
minimum	0.3	2.6
maximum	22	15
range	21.7	12.4

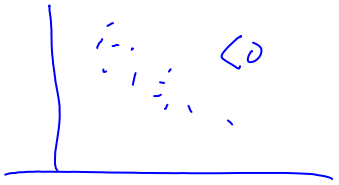
Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\MedalsAndEconomy_000.xls>

$n = 14$

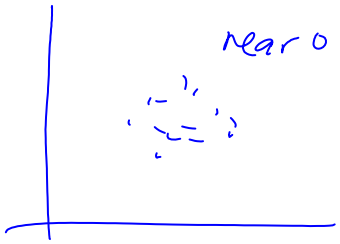
x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$
22	13	16.62	7.38	122.81
⋮	⋮	⋮	⋮	⋮
2.6	2.6	-2.77	-3.01	8.35
				238.36

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





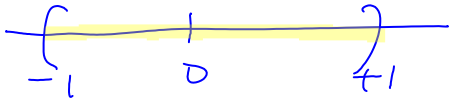
Better measure: Correlation Coefficient (r)



$r = \frac{S_{XY}}{S_X S_Y}$

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

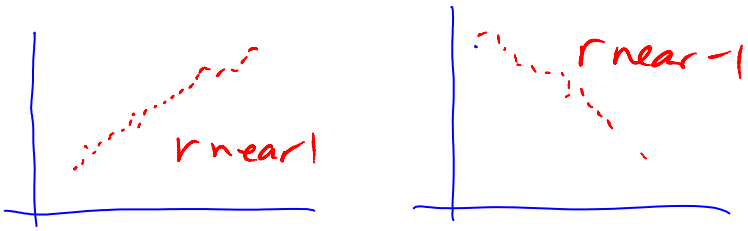
$$S_x = \sqrt{\frac{1}{n-1}\sum (x_i - \bar{x})^2}, \quad S_y = \sqrt{\frac{1}{n-1}\sum (y_i - \bar{y})^2}$$



Ex. Medals

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

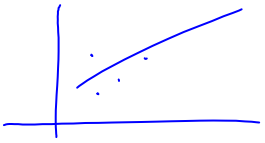
$$r = \frac{18.33}{(5.88)(4.12)} = 0.75$$ somewhat strong



Correlation doesn't imply causation

b) Simple Linear Regression

x indt var
y depdt var



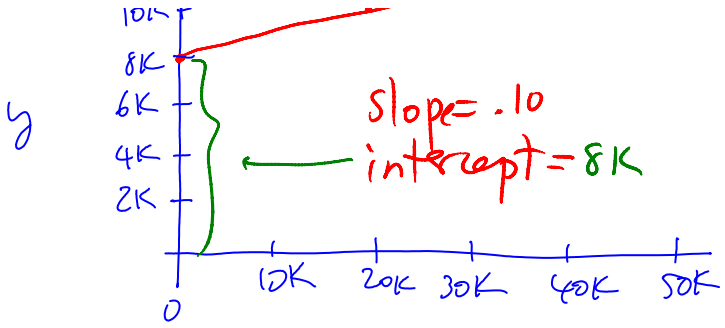
Ex. Naive model (no uncertainty)

x: gross sales at local Esso (monthly)
y: payments to parent comp. (")

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

$y = 8000 + 0.10x$ Linear model

10% of



$x=0$, $y=8$

$x=20$, $y = 8 + .1(20)$

$= 8 + 2 = 10$

$y = \beta_0 + \beta_1 x$

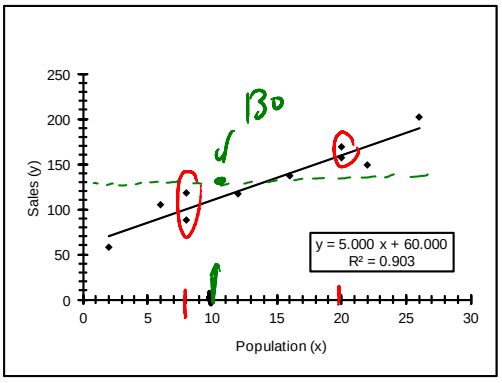
β_0 ↓ intercept ; β_1 ↑ slope

Ex. Statistical model – Harvey's restaurant

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

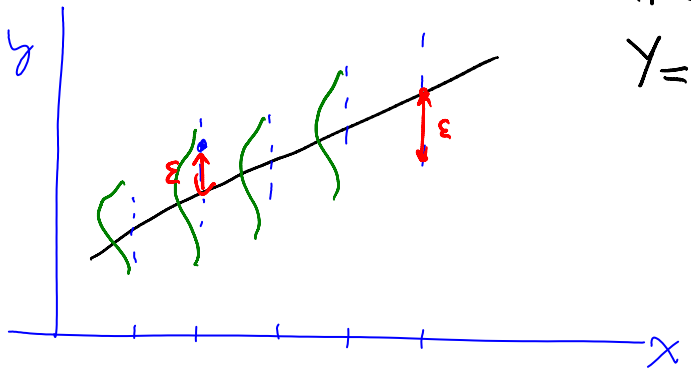
Pasted from <file:///C:\DOCUME~1\parlar\LOCALS~1\Temp\Harveys-1.xls>

$\bar{y} = 130$



Statistical model needed

Popn of 300+ restaurants

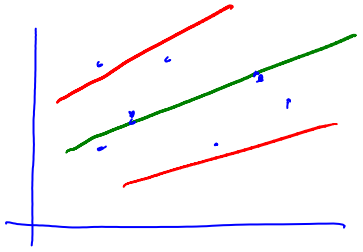


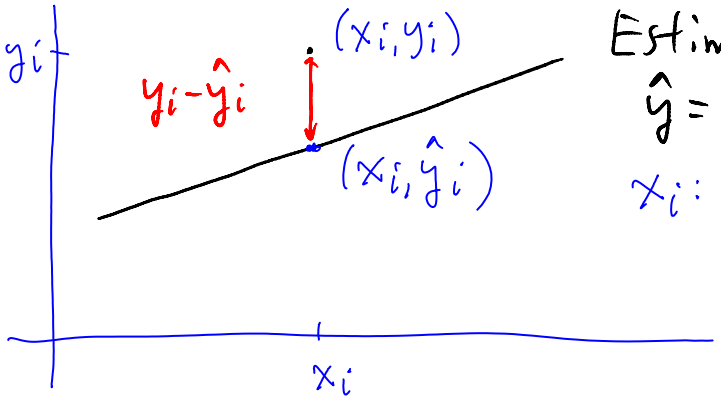
True model

$Y = \beta_0 + \beta_1 X + \varepsilon$

b_0 b_1

c) Best method to find regression line





Estimated line

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

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

Consider (x_i, y_i) : Actual y_i

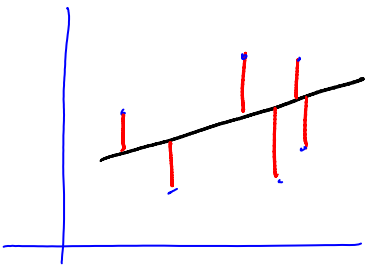
Estimate $\hat{y}_i$

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

$= y_i - (b_0 + b_1 x_i)$

Minimize

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



Solution for finding b_0 & b_1 to min. SSE

① $b_1 = \dfrac{SS_{xy}}{SS_{xx}}$, $SS_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$

$= \sum x_i y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum y_i\right)$

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

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

② $b_0 = \bar{y} - b_1 \bar{x}$, $\bar{y} = \frac{1}{n}\sum y_i$

$\bar{x} = \frac{1}{n}\sum x_i$