Chapter 9

9-2	CI $(p_1 - p_2)$	
	HT $(p_1 - p_2)$	
9-3	CI $(\mu_1 - \mu_2)$ HT $(\mu_1 - \mu_2)$	Independent Samples
9-4	CI $(\mu_1 - \mu_2)$	Dependent Samples > because matched pairs

- observational unit
- Subject unit
- experimental unit

	Treatment A	B	differences
1	A_1	B_1	$A_1 - B_1$
2	A_2	B_2	$A_2 - B_2$
3	A_3	B_3	$A_3 - B_3$
⋮			
n	A_n	B_n	$A_n - B_n$
	$\bar{X}_A$	$-\ \bar{X}_B$	$= \bar{X}_{A-B}$

μ_d

$$CI(\mu_A - \mu_B)$$

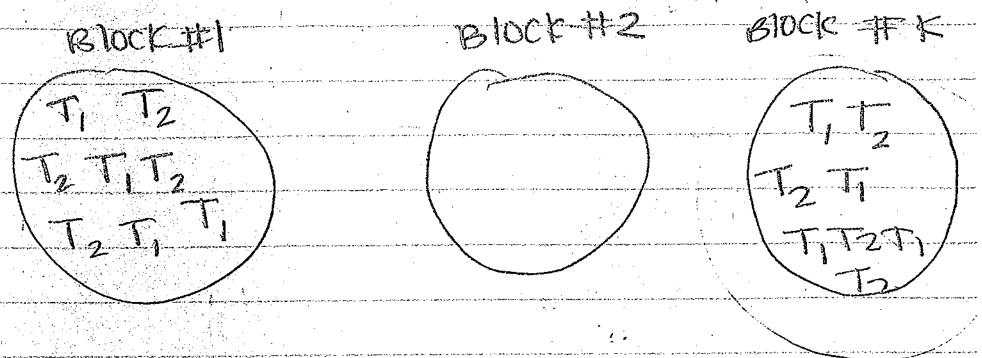
$\bar{X}_{A-B}$

(old one) $CI(\mu) = \bar{x} \pm t\left(\frac{s}{\sqrt{n}}\right)$

$$= \bar{d} \pm t\left(\frac{s_d}{\sqrt{n}}\right)$$

Blocks

- Groups of Similar exp. units
- in each group, all treatments were applied



- Type of control

Treatments that destroys or alters

Exp unit

notes 6/25

(pg. 462) $CI(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \, S(\bar{X}_1 - \bar{X}_2)$

$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

The 2 samples are independent and $\sigma_1 = \sigma_2$

$CI(\mu) = \bar{x} \pm t_{\alpha/2} \left(\frac{s}{\sqrt{n}}\right)$

$S_{\bar{x}}$

$\bar{X}_1 = \text{"}x\text{"}$ $\quad \bar{X}_2 = \text{"}y\text{"}$

If x and y are independent random variables, then variance of $(x-y) = var(x) + var(y)$

$$\sigma^2_{x-y} = \sigma^2_x + \sigma^2_y$$

$$\sigma_{x+y} = \sigma^2_x + \sigma^2_y$$

var = "variance of"

$$var(\bar{x}_1 - \bar{x}_2) = var(\bar{x}_1) + var(\bar{x}_2)$$

$$(\sigma_{\bar{x}})^2 = \frac{\sigma^2}{n} \qquad = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad \bigg| \qquad \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

what about <u>degrees of freedom</u>?

The smaller of $(n_1 - 1)$ and $(n_2 - 1)$.

~~VARIATION!~~

when $\sigma_1 \neq \sigma_2$;

Quiz #16 #4 (Not same)

$$95\% \ CI(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(Same) When $\sigma_1^2 = \sigma_2^2$ variation is same or similar for both population then pool variance.

$n_1 = 100$
$n_2 = 4$

$$\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} = s^2_{pool} \rightarrow \sigma^2$$

* $df = df_1 + df_2$

Quiz #16
#4
The way
it's written

variation is the same

pool variation

$$s^2_{pool} = \frac{(n_2-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$$

101.06

$$\frac{(7)(3.80)^2 + (7)(3.73)^2}{7+7}$$

$$s^2_{pool} = 14.18$$

$$CI(\mu_1-\mu_2) = (\bar{x}_1-\bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s^2_{pool}}{n_1}+\frac{s^2_{pool}}{n_2}}$$

$$= (18.71-17.86) \pm 2.145\sqrt{\frac{14.18}{8}+\frac{14.18}{8}}$$

$$= .85 \pm (2.145)(1.883)$$

$$= .85 \pm 4.04$$

$$\boxed{3.19 \quad 4.89}$$

$M_H > M_C + 100$

$(M_H - M_C) > 100 : H_1$

Unit #3 more problems

(Summer 2014)

#9

Claim: $M_{LR} < M_{Car} + 10$

$H_0: MLR - MCar \geq 10$

$H_1: MLR - Mcar < 10$

#1 $M_A > M_B + 5$

Proportions are related to Binomial

$X \sim B(n, p)$

$\sigma_X \sqrt{npq}$

$\sigma^2_{X} = npq$

$\sigma^2_{\hat{p}} = \left(\frac{1}{n \cdot \cancel{n}}\right) \cancel{n} pq$

$= \frac{pq}{n}$

$\hat{p} = \frac{x}{n}$ let $a = \frac{1}{n}$

MATH truth

X is a random variable with variance $= \sigma^2$

what is var(ax)?

$\Rightarrow a^2 \sigma^2$

Proportions

only proportion formula used for CI

$$CI(p_1 - p_2) = (\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$CI(p) = \hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\mathrm{Var}(\hat{p}_1 - \hat{p}_2) = \mathrm{var}(\hat{p}_1) + \mathrm{var}(\hat{p}_2)$$

$$\text{variance} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

9-2 Hypothesis testing "for proportions" $(p_1 - p_2)$

$H_0: (p_1 - p_2) \; \leq \; = \; \geq \; \underline{\underline{0}}$ or $\underline{\underline{C}}$ ← non-zero

$H_1: (p_1 - p_2) \; > \; \neq \; < \; 0$ or C

Claim: $p_1 < p_2$

$(p_1 - p_2) < 0$

Claim: p_1 is 3% pts less than p_2

$\cancel{p_1} + 0.03 = p_2 - p_1$

$(p_2 - p_1) = 0.03$

If H_0: has non-zero "C" then the test statistic

(Not on pg. 451)

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\dfrac{\hat{p}_1 \hat{q}_1}{n_1} + \dfrac{\hat{p}_2 \hat{q}_2}{n_2}}}$$

If H_0: has "0" then:

$\hat{p}_1 \xrightarrow{\text{estimates}} p_1$

$\hat{p}_2 \xrightarrow{\text{estimates}} p_2$

$\Big\} \; p \qquad \bar{p}$

all the successes ÷ all the trials

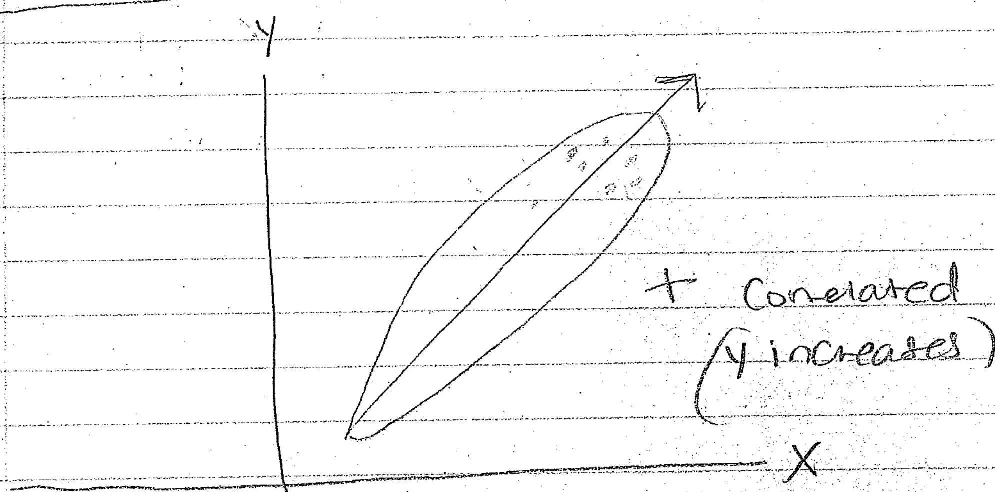
$$\bar{p} = \frac{X_1 + X_2}{N_1 + N_2} \qquad \bar{q} = 1 - \bar{p}$$

Test statistic

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\sqrt{\frac{\bar{p}\bar{q}}{N_1} + \frac{\bar{p}\bar{q}}{N_2}}}$$

Quiz #15

[illegible]



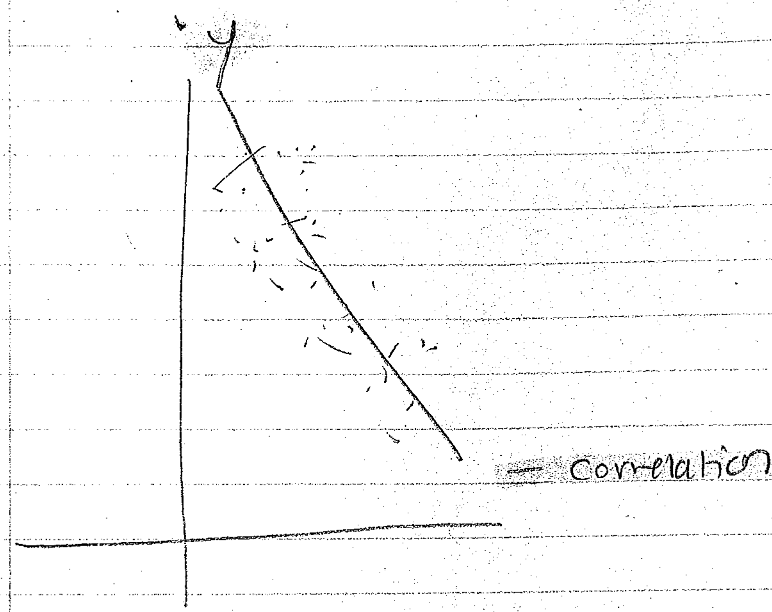
If y gets bigger as x gets bigger then x & y are positively correlated

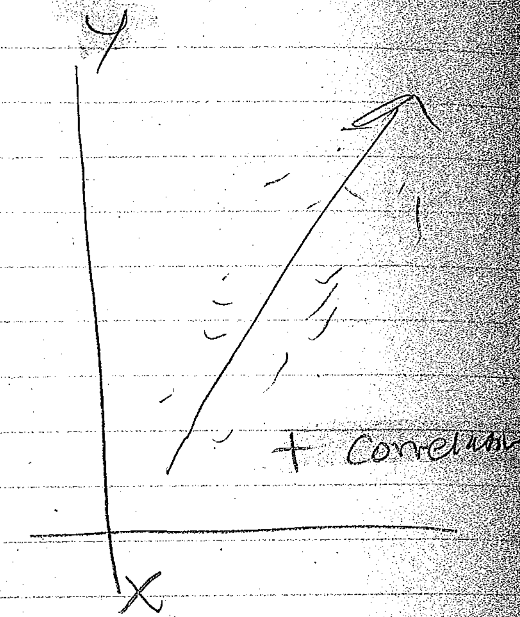
How does y relate to x?

Correlated?

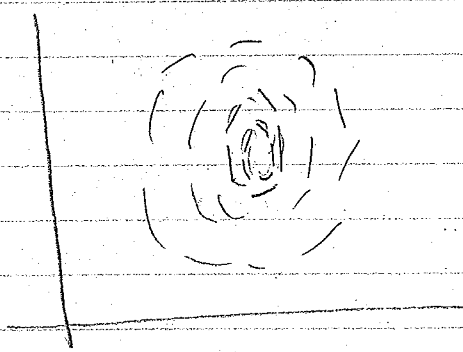
Linearly Correlated

Chp-10 Linear Correlation

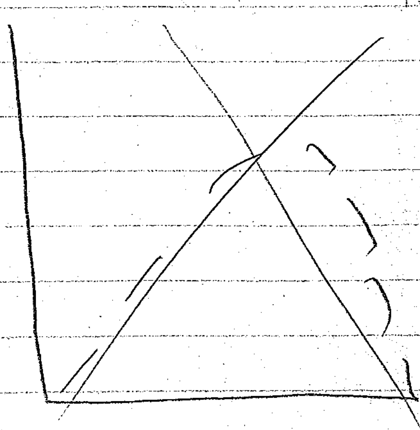




If y gets bigger as x gets bigger then x & y are positively correlated.



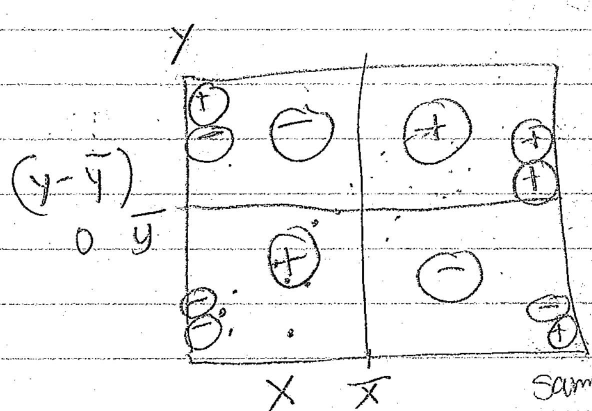
zero correlation
$r = 0$



Zero correlation because ups cancels down.
Quadratic correlation

* The closer the points get to line the bigger the correlation (absolute value)

① Correlation does not imply cause & effect necessarily



Sample Average for x coordinates

$(x-\bar{x})$

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1} \Rightarrow (+) \quad \text{covariance}$$

n points

prior knowledge

$$S^2 = \frac{\sum(x-\bar{x})(x-\bar{x})}{n-1} \quad \text{Variance}$$

distance in meters

$Cov = m^2 = 100{,}000\ m^2$

distance in km $= 0.1\ m^2$

Don't use

$$-1 \le \frac{\frac{1}{(n-1)} \sum (x-\bar{x})(y-\bar{y})}{s_x \cdot s_y} = r \le 1$$

$$\frac{\sum z_x \cdot z_y}{(n-1)}$$

7/28/14 notes — clear data — variance 2

Chp. 10-2

r = correlation coefficient

Never use TABLE A-5

ex: Quiz# 17 #2

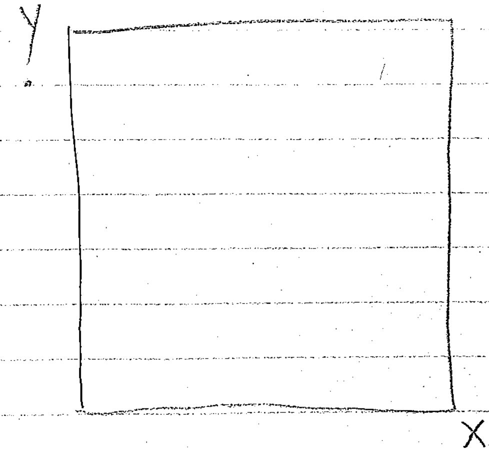
r^2 = coefficient of determination

Hypothesis test "for correlation"

t table

$r = 0.66$

$$\frac{r}{\sqrt{\frac{1-(r)^2}{n-2}}} = \frac{.66}{\sqrt{\frac{1-(.66)^2}{6-2}}} = \frac{0.66}{.3756} = 1.757$$



Claim: $\rho > 0$

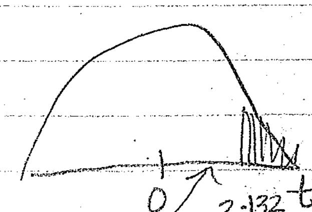
(row)

$H_0: \rho \le 0$

$H_1: \rho > 0$

$\alpha = 0.05$ right tail

$n = 6$



$df = n-2 = 6-2$
$= 4$

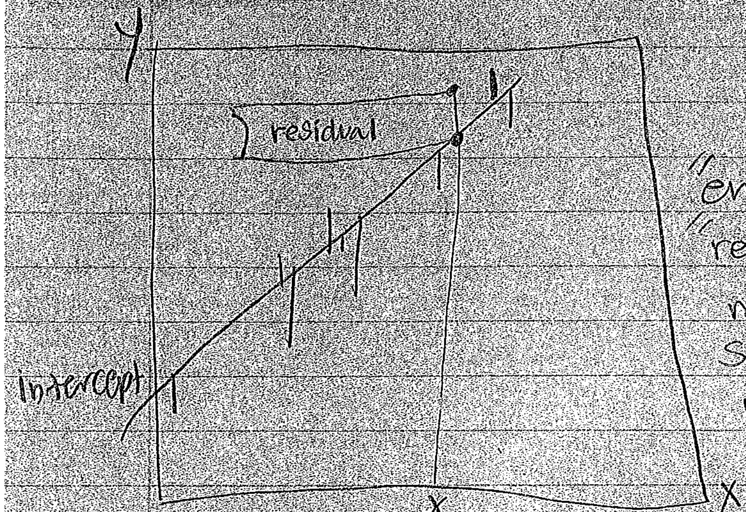
DO not reject Ho.

Most H.T (ρ.) involve a test against $\rho = 0$

is r good enough?

If I need $r > 0.95$

10-3 Simple linear regression



"errors"
"residuals"
make these as small as possible

$$y = \overset{\text{slope}}{mx} + \underset{\text{Intercept}}{b}$$
(a) (b)
$$= b_1 x + b_0$$

$$\hat{y} = b + ax$$

$$\text{residual} = y - \hat{y}$$

$$y - (b + ax)$$

Σ residuals as small as possible

$a = \text{slope} = (r)\left(\frac{s_y}{s_x}\right)$

$r = .6418 \qquad = .6418\left(\frac{538.1}{50}\right)$

$= 6.907$

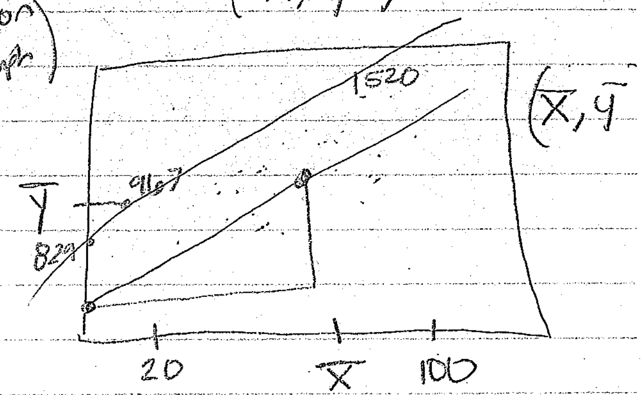
once the slope (a) is known, then

$b = \bar{y} - \underline{a}\bar{x}$

int =

The regression line goes through $(\bar{x}, \bar{y})$

plot line on graph



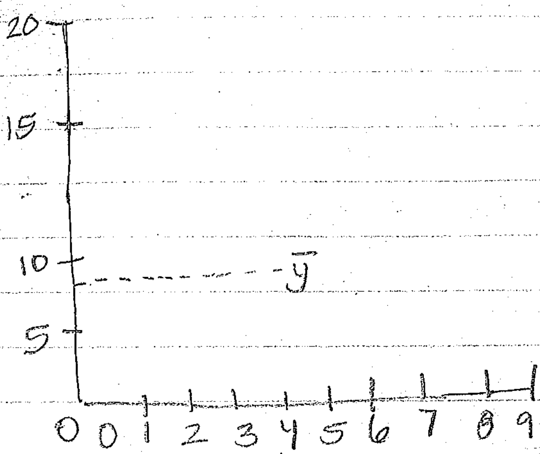
$y' \overset{\text{(or calc)}}{=} \hat{y}$

10-3

- How to calculate slope and intercept for the best-fitting straight line.
- using a regression (best fit) line to predict y for a known x

7/30/14 notes

10-4 notes Variation in the "regression context



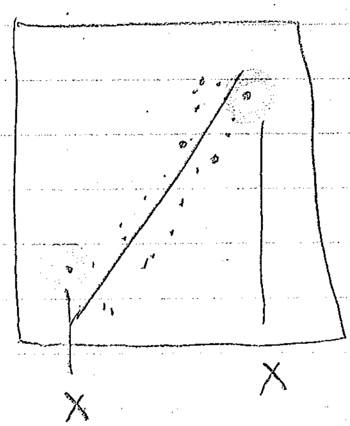
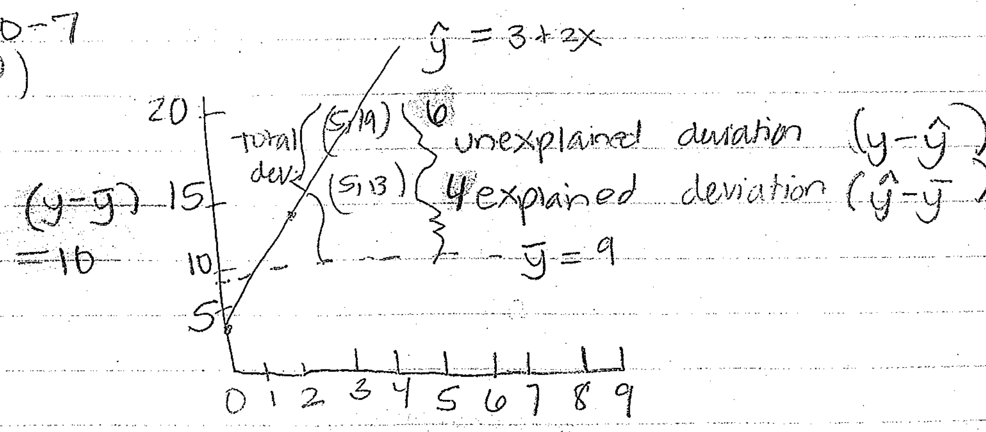


Figure 10-7
(pg. 530)



explained deviation + unexplained deviation = Total deviation

$\{$pairs (x, y) : $(x_1 y_1)$ (x_2, y_2)

*

$$\Sigma(y-\bar{y})^2 = \Sigma(\hat{y}-\bar{y})^2 + \Sigma(y-\hat{y})^2$$

same name (Total variation / total sum of squares) — explained variation sum of squares + unexplained variation sum of squares

calculator

$s_y^2(n-1)$

Coefficient

$$r^2 = \frac{\Sigma(\hat{y}-\bar{y})^2}{\Sigma(y-\bar{y})^2} = \frac{\text{explained}}{\text{total}}$$

proportion of the total variation that is explained by the line.

$\Sigma(y-\bar{y})^2 = 400$ $\qquad$ $\Sigma(\hat{y}-\bar{y})^2$ $200 = 400(.50)$

$r^2 = 0.5$

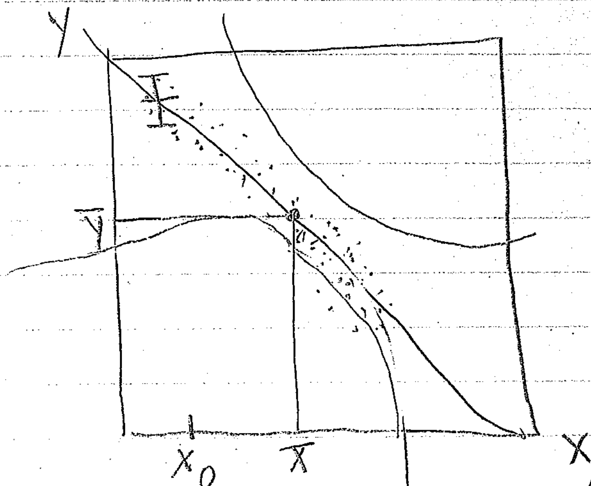
$$\Sigma(\hat{y}-\bar{y})^2 = s_y^2(n-1) \times r^2$$

total * r^2

$Sy^2 (n-1) = 1158101.2 = \Sigma(y-\bar{y})^2$

Standard error of estimate

$$S_e \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}} = \sqrt{\frac{\text{unexplained}}{n-2}}$$



$\hat{y} = b + ax = \cancel{b_0 + b_1 X}$

$$S_e = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}}$$

$$\hat{y} \pm t_{\alpha/2} \cdot S_e \sqrt{1 + \frac{1}{n} + \frac{(X_0-\bar{X}}{\Sigma(x-\bar{x}}}$$

Quiz # 17
CI

$CI(y/X_0)$

$90\%\ CI\ (y/X=80) = \hat{y}$

$n = 5$

$n-2 = 3$ d.f

$\hat{y} = 1382$

$1382 \pm$

80 —

$t = 2.353$

$$1382 \pm (2.353)(4765)\underbrace{\sqrt{1 + \frac{1}{5} + \frac{(80-50}{10{,}000}}}_{1.136}$$

$S_e = 476.5$

$\bar{X} = 50$

$\pm\ 1274$

Correlation

7/30/12 notes

Pvalue approach to Hypothesis Testing

a bunch of proportions

- 11-2 Goodness-of-fit
- 11-3 contingency tables
- 11-4 Analysis of variance

a bunch of means

$H_0: \mu \leq 60$

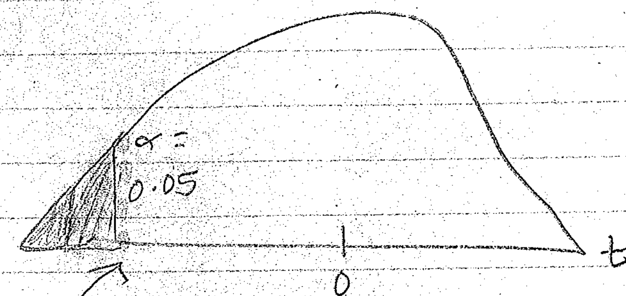
$H_1: \mu < 60$

$\alpha = 0.05$ left tail test

$n = 24$

$df = 23$

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \bigcirc$$



p-value is area under distribution curve to the left of the test statistic

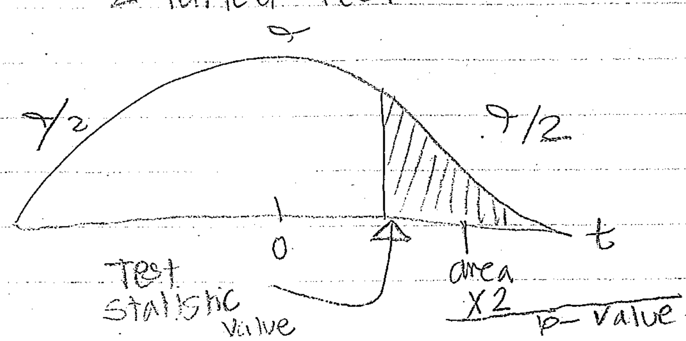
Traditional: do not reject

P-value method (p.394)

- If p-value $\geq \alpha$ then do not reject Ho:
- If p-value $< \alpha$ then reject Ho:

* p-value bigger than alpha (sig. level) so do not reject

2 tailed test



11-2 Goodness-of-fit [multinomial]

Ex:
Quiz #18
(prob #2)

Categories	(observed count) Frequency	Hypothetical Expected count	$\frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}}$
1			
2			
⋮			
K			
	sum = N		$\Sigma =$

Test statistic

$$\sum_{i=1}^{K} \left[\frac{obs_i - Exp_i}{Exp_i}\right]^2$$

(No decimals) Count

(can have decimals) expected

10-5 not Responsible

11-3 contingency tables (pg.568)

- Log-linear model
- Cross-classified binary data

observed counts   Expected counts

$$\sum \left[\frac{(obs - Exp)^2}{Exp} \right]$$

- Expected counts are differently from multi-nomi
- D.F different
  (# of rows-1)(# of columns-1)
  $(r-1)(c-1)$

117		
90		
82		

289

(AOV)

Analysis of Variance

H_0: $\mu_1 = \mu_2$

H_1: Not H_0:

(means atleast one is different fromother)

Calculations with "unequal sample sizes"

Treatments

1	2	3
$X_{1,1}$	$X_{2,1}$	
$X_{1,2}$	$X_{2,2}$	
	e.t.c	

$\bar{X}_1 \longrightarrow \mu_1$

$\bar{X}_2 \longrightarrow \mu_2$

Regression Total variation

AOV Total Sum of Squares $n-1$

Reg Explained variation $k-1$

AOV Treatment Sum of Sq. (Between groups)

Reg: unexplained variation $n-k$

AOV: Error Sum of squares (within groups)

Aov table

$$SS / d.f = MS$$

$$d.f(M.S) = SS$$

source	SS	d.F	MS	F
Treatments	18.9944	8	2.3743	MS(Treat) ÷ MS(error) = 1.6204
error	26.3746	18	1.46525	8 d.F numerator / 18 deno
	45.369	26	X	

F Distribution

(chart @ home in red file)