4/24 **Correlation**

"parametric" "r" Pearson's product moment correlation coefficient

"non-parametric" - order statistic
kendall's tau or spearman's rho

μ - t

p - Z

σ^2 - X^2

$\{X : X_1, X_2, X_3, X_{\dots}, X_n\}$ $\bar{x}$ s

$\{X : X_{(1)}, X_{(2)}, X_{(3)}, X_{(\dots)}, X_{(n)}\}$

$X_{(1)}$ ↓ min

$X_{(n)}$ ↓ max

1st order statistic

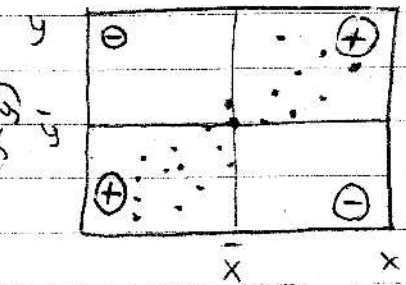
survey Rank

1 Best

|

10 Worst

How could you measure linear correlation?



x ↑ and y ↑ - + 'r'

x ↑ and y ↓ - − 'r'

$(x-\bar{x})(y-\bar{y})$ What if $\Sigma(x-\bar{x})(y-\bar{y})$

If $\Sigma(x-\bar{x})(y-\bar{y}) = 0$ no linear correlation

km	
100	2
100,000	2,000 → m

$$\frac{\Sigma x}{n} = \bar{x}$$

$$\frac{1}{(n-1)}\Sigma(x-\bar{x})(y-\bar{y}) = \underline{\underline{\text{covariance of } x \neq y}}$$

$$\frac{\Sigma(x-\bar{x})(y-\bar{y})}{n-1}$$

$$\frac{\Sigma(x-\bar{x})(x-\bar{x})}{n-1} = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$\frac{\left(\frac{1}{n-1}\right)\Sigma(x-\bar{x})(y-\bar{y})}{S_x \cdot S_y} \rightarrow \text{unit less}$$

$$-1 \leq \frac{\left(\frac{1}{n-1}\right)\Sigma(x-\bar{x})(y-\bar{y})}{S_x \cdot S_y} \leq 1$$

$$\downarrow$$

$$r$$

p.497.

$$\frac{\Sigma\left[\frac{x-\bar{x}}{S_x}\right]\left[\frac{y-\bar{y}}{S_y}\right]}{n-1} = r$$

Quiz 17(3)

2nd → Data → 2-VAR → Enter →

→ 2nd → Data → CLR DATA → Enter

Data → $x_1 =$ $y_1 =$

$x_2 =$ $y_2 =$

$x_n =$ $y_n =$

→ STATVAR

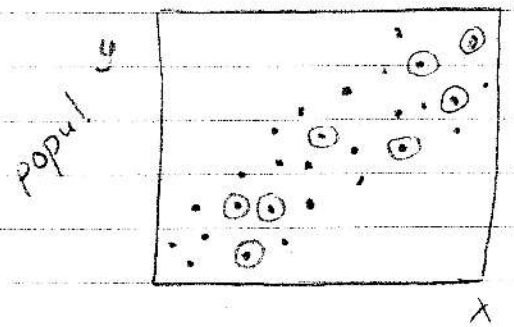
$n, \bar{x}, s_x, \sigma_x, \bar{y}, s_y, \sigma_y, \Sigma x, \Sigma x^2, \Sigma y, \Sigma y^2, \Sigma xy, a, b, r, x', y'$

-0.81

$+0.806$

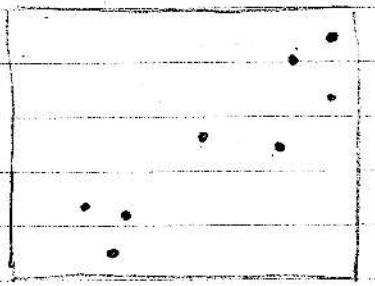
-0.8056

p. 488 1. random sample (x, y)



observation units

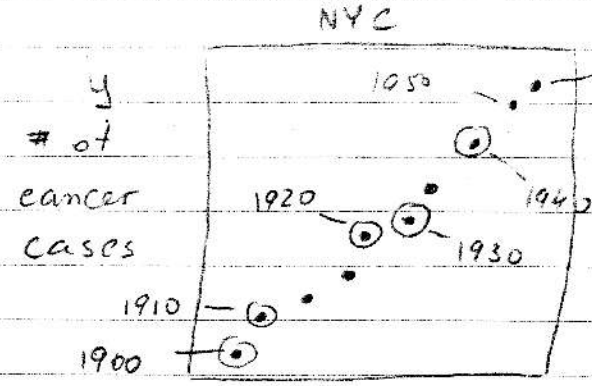
person (ht, wt)

2.  scatterplot

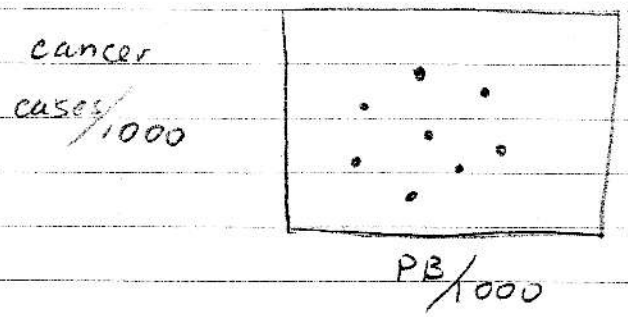
3. outliers — make no sence

you know something went bad or wrong

p. 493 Properties of the Linear Correlation Coefficient r

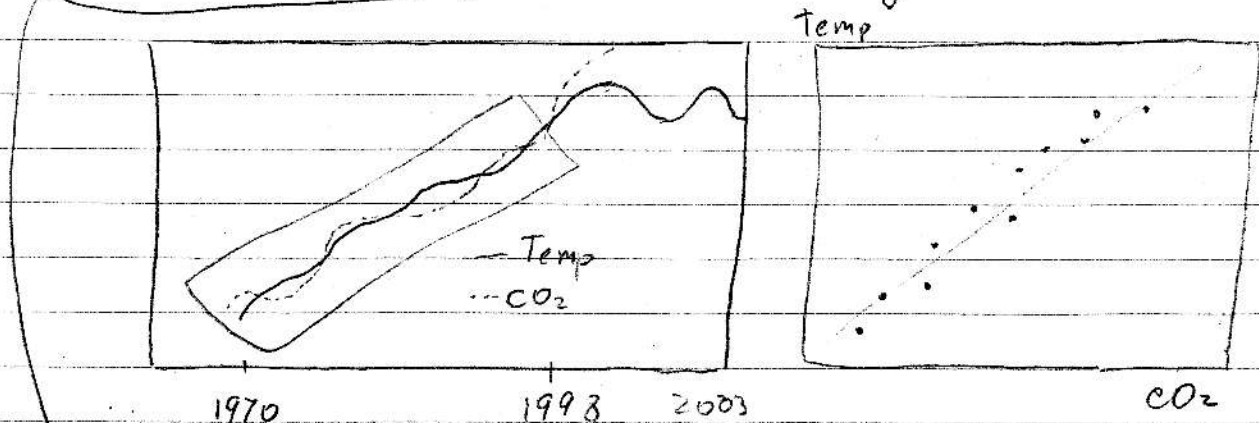


$r = 0.992$



no correlation

man-made ⊕
catastrophic ⊕ global warming



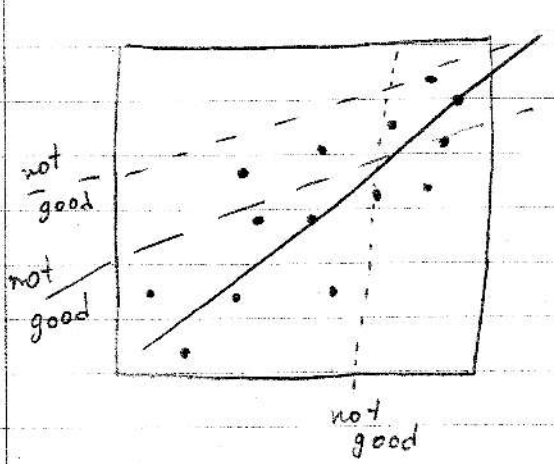
computer models

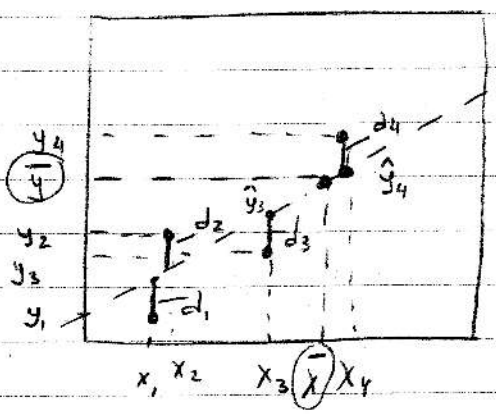
p.494 Hypothesis test for ρ

Ex. Exam # 3 set p 10(3)

Quiz 17(2)

Quiz 17(3)





$\hat{y} = mx + b$

I want the line that makes $\Sigma (y - \hat{y})^2$ as small as possible

"least squares"

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

"b" "a"

x - independent variable "predictor"

y - dependent variable "predicted"

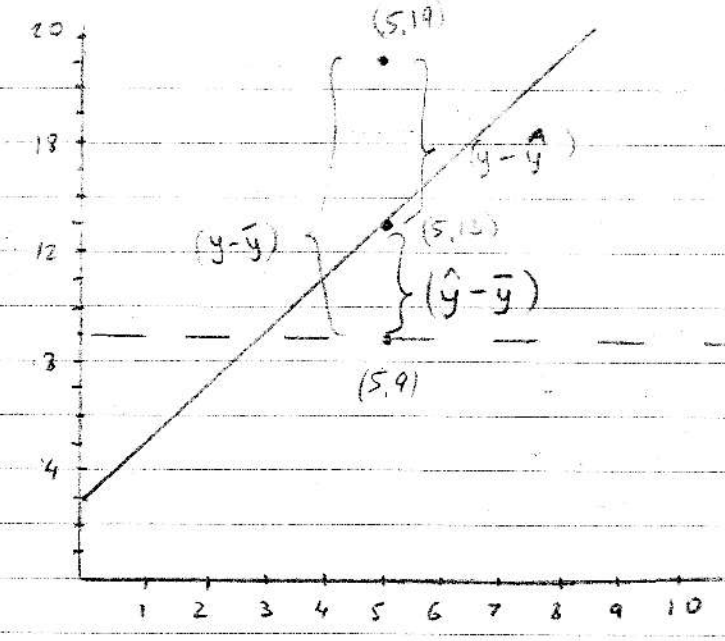
a - slop and b - inters. calculator

y'_1 → y_2 → Enter

p.525 Ch 10-4

Variation and Prediction Intervals

p. 526

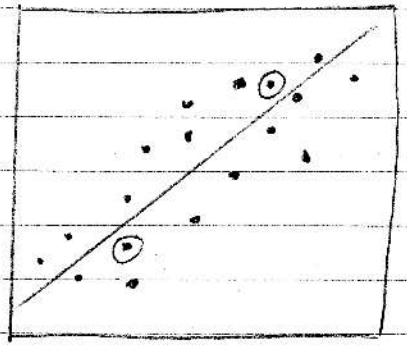


$(y-\bar{y})$ = "total" deviation

$(y-\hat{y})$ = "unexplained" deviation

$(\hat{y}-\bar{y})$ = "explained" deviation

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



$x_1 = 15$

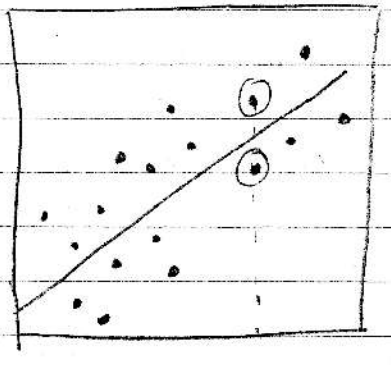
$x_2 = 14$

$x_3 = 24$

$x_4 = 58$

$x_5 = 70$

$x_5 > x_1$ because it is random relationship is explained.



"unexplained" deviation

p. 527

$\sum_{i=1}^{n} (y_i - \bar{y})^2$ = Total variation [in y]

$\sum (\hat{y} - \bar{y})^2$ = explained variation [in y]

$\sum (y - \hat{y})^2$ = unexplained variation [in y]

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

$c = a + b$

$c^2 = a^2 + b^2$

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

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

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

$S_y \rightarrow [x^2] \rightarrow [(] \rightarrow [\text{STATVAR}] \; [-] \; [1] \; [2]$

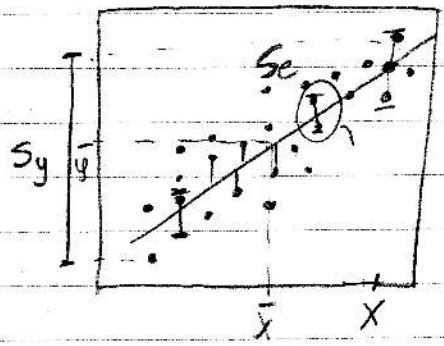
860 people 1/2 of them are men

1/2 of 860 $= \frac{1}{2} \cdot 860 = 430$

(Total) r^2 = explained

(Total)$(1-r^2)$ = unexplained

Standart error of estimate



$(y-\hat{y})$

d_1

d_2

d_n

$$S_e = \sqrt{\frac{\text{unexplained}}{n-2}} =$$

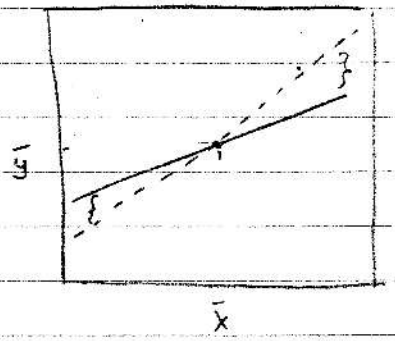
$$= \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-1}}$$

mean slop

p. 529 $CI(y|x) = \hat{y} \pm t_{\alpha/2} \cdot Se \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x-\bar{x})^2}}$

$n-2 = d.f.$

total variation of x



Quiz. 17 (make up) 95% CI