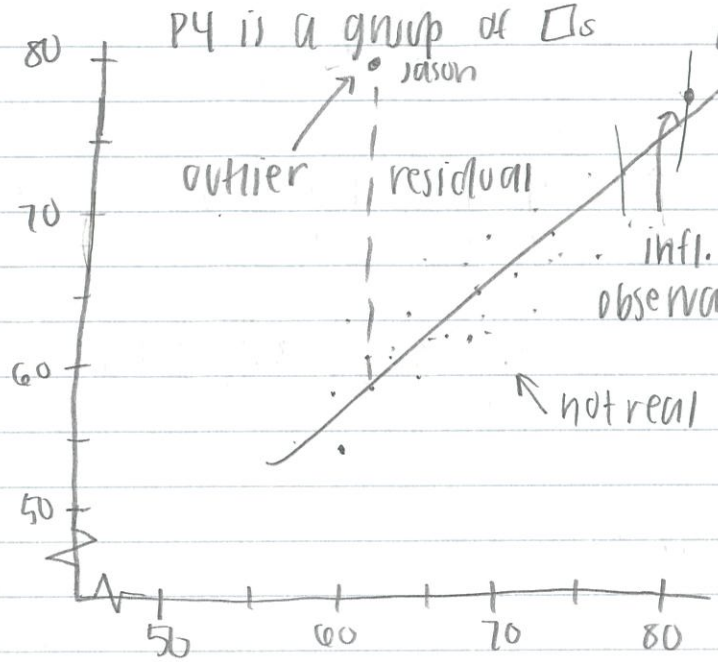
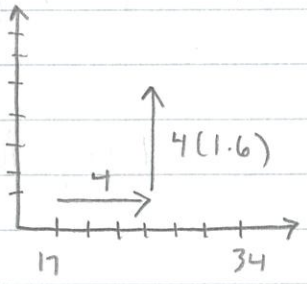
WINGSPAN SCATTERPLOT 10/23/23



ANA:
+ linear association
str: wingspan and height are proportional (sp have sq)
ins: (81, 80)
outs: (60, 80)

9.2 pg 555 CARIBOU PROBLEM 10/27

$b = \frac{1.6}{1}$



559 CAR PROBLEM

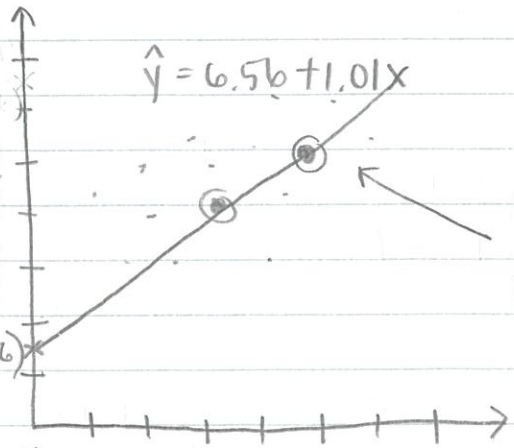
"Famous" cars

$\hat{y} = 6.56 + 1.01x$

$\hat{y} = 6.56 + 1.01x$

$b = \frac{1.01}{1}$

$a = 6.56$



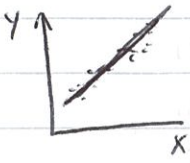
if $X = 12$, $y_p = 18.68$

$y_p = 6.56 + 1.01(12)$

CORRELATION COEFFICIENT 10/31

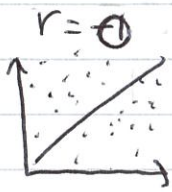
Karl Pearson (*not quizzed)

$-1 \leq r \leq 1$ $r = +1$ "perfect linear correlation"

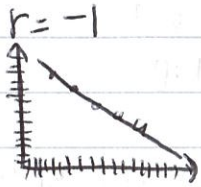


slope doesn't matter

dots perfectly on line



not tight hug



negative

"perfect linear correlation" but negative slope

not linear

$$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}}$$

r^2

pg 540 EXAMPLES

SAND DUNES —

$r = 0.949$ "high/strong positive correlation"

+association

as the wind speed increases, the drift rate increases (trend)

park closure/warning

BOSTON — baseball / ppl smw else

- association

trend: as police officers increase, muggings decrease

$r = -0.969$ "super strong negative linear correlation"

r^2 = coefficient of determination = 0.9390

6.1% unexplained (conjectures ~ weather/time)

variation

r^2 = 93.9% of (V) in the y (# of muggings) is explained by by least square regression line and the variation in x (# of police)

9.1 #15 HW PROBLEM CYCLONES * draw scatterplots

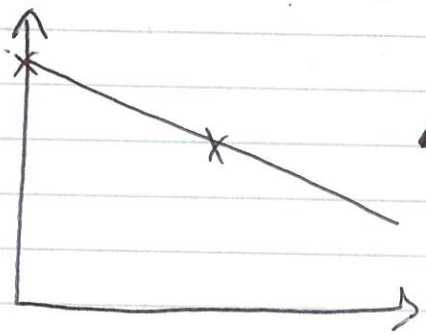
strong negative linear correlation
trend: a) the lowest pressure increases, the maximum wind speed decreases

$\bar{x} = 970.5$
$\bar{y} = 96.67$

CALC 8
$a = 1496.78$
$b = -1.4427$

$y = a + bx$
$y = 1496.78 + (-1.4427)x$

x	y
970.5	96.67
0	1496



$r^2 = 0.9796$
$r = -0.9897$

y= vars
stats EQ
} sort like that
2nd trace value

spurriance correlation - seemingly related but isn't (lemons and facilities)

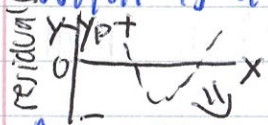
NO causation

STUDY GUIDE 11/8

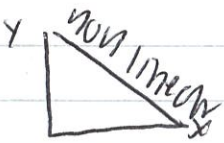
a residual is a value's distance away from the least squares regression line ✱ $y - y_p$ diff b/w observed and predicted

an outlier is a data point that has a high residual from the least squares regression line

what is a residual plot?



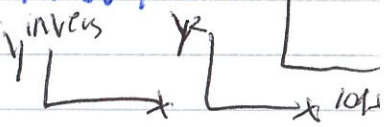
if you see a pattern

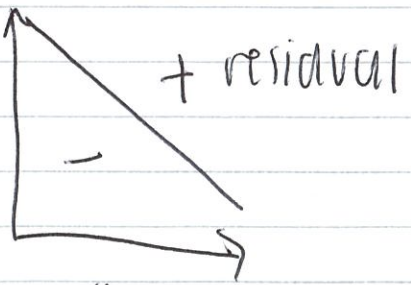


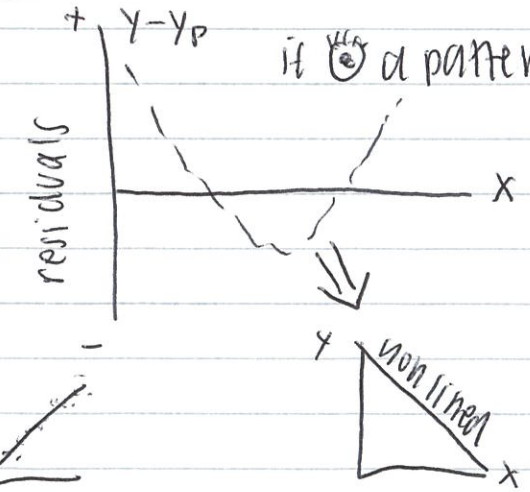
An influential observation is a data point / value that is has a relatively small residual but the x is far off from majority of the data ✱ exact answers in GC residual packet

why is (24, 76000) influential (Ex: 4.10) the x value is far from the other values even though the point is close to the LSRL. 4 years is a lot

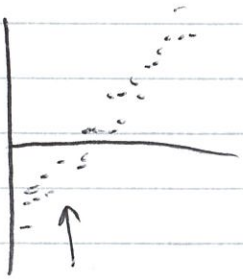
if the Residual Plot shows a pattern, then the LSRL is NOT good representation of the relationship for x and y. so, one type of transformation could be done with $y, \frac{1}{y}$ inverse y^2 $\log y$





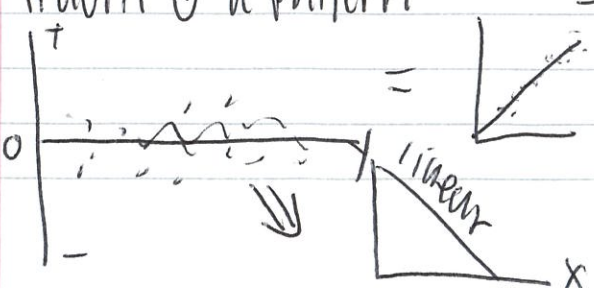


if 😐 a pattern



still a pattern

if 😐 a pattern



COLLEGE BOARD VIDEO NOTES 2.7 RESIDUALS 11/10/2023

Residual Plot – $(y-\hat{y})$ plot showing the residuals and fit of the least squares regression line

GOOD fit
- apparent randomness
- center at 0
- no clear patterns

BAD fit
- curve pattern
  - not best fit line
- accentuate possible trends

2.7B RESIDUALS

Residual – the difference b/w the actual response value and the model's predicted response value

observed y − predicted y

$y - \hat{y}$

+: model underestimated actual response value

−: model overestimated actual response value

2.9A ANALYZING DEPARTURES FROM LINEARITY

$\bar{x}$ and $\bar{y}$ will be on the LSRL

Low leverage point
- LSRL shifts down

High leverage points
- LSRL shifts up/steeper
- points with large/small x values (far from $\bar{x}$)
- affect slope/y-int

Outlier
- unusually high magnitude residual
- big affect on strength of relationship

Influential points
- outliers (correlation)
- high-leverage (slope/y-int)