Chapter 3: Examining Relationships

3.1 Scatter Plots - Day 1

OBJ. You will classify vars. + describe the relationship between two quantitative vars.

When Examining the relationship between 2 variables Ask:
- What individuals do the data describe?
- What are the variables? How are they measured?
- Are the variables all quant. or is there at least one cat?
- Do you want to explore the relationship of the vars, or do you think some of the vars. explain a cause Δ in others?

Explanatory var - Attempts to explain the obs. outcome.
Response var - Measures the outcome of a study.

Example 3.1
Example 3.2 + TP After!! :)

3.1-3.3

Scatter Plot - Shows the relationship between two quant. vars.

* If there is an explanatory variable, plot it on the horizontal axis. (Explanatory = x, Response = y)

Go to Pg 70 SAT M - NY (502) - 76% } Conclusion?
SAT M - AL (565) - 9% } Look @ Plot on Pg 124

Examining a Scatter Plot
Look for: Overall Pattern and striking deviations from that pattern.
- Form - clusters / gaps
- Direction - See Association Below
- Strength - How close the pts follow a clear form.

Association
Positive Assoc - When above AVG values of one tend to accompany above average values of the other and below avg. values also occur together (Pos. slope)

Negative Assoc - when above AVG values of one tend to accompany below AVG values of the other and vice versa.

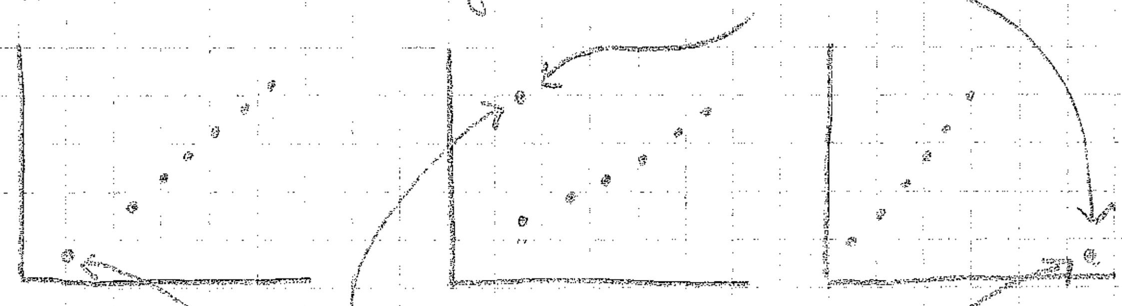
$\boxed{\text{End of Day 1}}$ - No HW!! :)

3.1 Scatter Plots - Day 2

3.6
3.9
3.11 - show $\underbrace{SEQ}_{\text{List/OPS}}$(10x, x, 1, 15) for 10→150 in L1   ~~NOT NEEDED~~

__HW:__ 3.16, 3.19, 3.21 (Sketch b+d by hand)   Quiz Tomorrow

Talk About Regression Outliers



vs
Outliers

Regression outliers Δ the LSRL

outliers are obs. that fall outside the overall pattern of the data.

Least Sqr Reg Line - Stat → Calc #8 Lin Reg (a+bx)

Diagnostics on

3.2 Correlation

<u>OBJ</u>: You will calculate the correlation (r) between two <u>quant.</u> variables

<u>Correlation</u> - Measures the direction (+/-) and strength (#) of the linear relationship between two quant. vars. (Symbol = r)

$$r = \frac{1}{n-1}\sum\left(\frac{x-\bar{x}}{S_x}\right)\left(\frac{y-\bar{y}}{S_y}\right)$$

$\bar{x} + S_x$: Mean + St. Dev for x vars.
$\bar{y} + S_y$: " " " " y ".

Figure 3.8 - pg 141

Facts about correlation - see pg 143 + 144.

** Correlation is <u>not</u> a complete description of two-variable data. You <u>should</u> also include the means + st. Dev of both x and y. (Since they were used to calculate r)

** Read Example 3.7 (Pg 145) - <u>If</u> you +/-/x/÷ the same # to <u>All</u> values of x or y or both, r does not Δ.

Instead - Do an ex w/ +100 to x values.

Figure 3.9 - classify each from very weak → very strong.

Find r for:

① (10, 10)
(10, 20)
(20, 10)
(20, 20)

② Add in (100, 100)

③ Add 5 sets of Pts.
$10 \le x \le 20$
$10 \le y \le 20$

④ Add in (95, 10)

⑤ Out (95, 10) in (10, 95)

⑥ Out misc. to see inc, dec r value.

Activity: r-value + scatter plots.

<u>HW</u>: 3.31, 3.34, 3.37

3.3 Least-Squares Regression - Day 1

OBJ: You will find and interpret a least-squares regression line. (LSRL)

Correlation - measures the direction (+/-) and strength (#) of the Lin. rel. between two quant. vars.

Least-Squares Regression - a method for finding a line that summarizes the relationship between 2 vars.

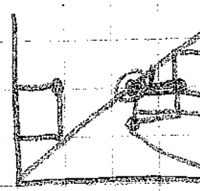
Regression Line - A straight line that describes how a response var. y Δ's as an expl. var. x Δ's.

- used to predict the value of y given a value of x.
- Must have an exp. + resp. var.

Example 3.8 - pg 150

Prediction Errors are errors in y ∴ we want a line that is as close as possible to the pts. in the vertical direction. (Error = observed - Predicted)

(Residuals = $y - \hat{y}$)



The LSRL makes the total Area of the squares as small (Least) as possible. ($y \neq \hat{y}$ usually)

Equation of LSRL: $\hat{y} = a + bx$ slope = $\underbrace{\boxed{b = r\frac{s_y}{s_x}}}_{\text{The Amt of } \Delta \text{ in y when x inc. by 1.}}$ * on FS

Intercept = $\boxed{a = \bar{y} - b\bar{x}}$ * on FS

LSRL passes through $(\bar{x}, \bar{y})$ - Always!

$\hat{y} = a + bx$
$\bar{y} = a + b\bar{x}$
$a = \bar{y} - b\bar{x}$

Example 3.9

$b = r\frac{s_y}{s_x} = 1.6945$

$a = \bar{y} - b\bar{x} = 131.9283$

$\hat{y} = 131.9283 + 1.6945x$

* To Plot by hand, select 2 extreme x-values, calculate $\hat{y}$ values and then draw line - Don't "wing it".

3.40

HW: 3.41 (Tomorrow to pg 165)

3.3 Least-Squares Regression - Day 2

<u>Coefficient of Determination</u> (r^2) - The % of variation in the values of y that is explained by the Least-Squares Regression of y on x.

-OR-

?% of the variation in y (response) is accounted for by the LIN. REL. w/ x (Explanatory)

-OR-

Give "very loose" explanation.

Facts about Least-square Regression (Pg 162-164 BOLD)

1. The distinction between exp. + response vars is essential in regression. (See Figure 3.15)
2. $b = r\frac{s_y}{s_x}$; a Δ of one st. Dev. of x corresponds to A Δ of r st. Dev in y.
3. LSRL <u>ALWAYS</u> passes through $(\bar{x}, \bar{y})$
4. r^2 is the fraction of the variation in the values of y that is explained by the least-squares regression of y on x.

r^2 Measures how successful the regression was in explaining the response.

$r = \pm .7$ $r^2 = .49$ ⟶ About 1/2 the variation is accounted for by the Lin. Rel.
$r = \pm .8$ $r^2 = .64$
$r = \pm .9$ $r^2 = .81$
$r = \pm .95$ $r^2 = .9025$

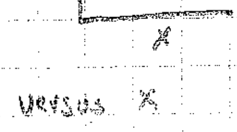
If $r^2 = .7$ $r = .837$ <u>BUT</u> is r + or -? Depends on the <u>Association!</u>

Example 3.12

3.45

<u>HW:</u> 3.42-3.44 (Tomorrow to pg 172)

<u>PS:</u>



Y versus X
Y on X

3.3 Least Squares Regression - Day 3

Residual - The difference between an observed value (response var.) and the value predicted by the regression line.

Residual = observed y - Predicted y
= $y - \hat{y}$

Do Example 3.14 - pg 167

Residual Plot - A scatterplot of the regression residuals against the explanatory variable.
(See Pg 174 - Resid. Plots by calculator)

Residuals: Things to Look For: (Look @ Plots on Pg 170 + 171 Also)

1) a) No Pattern, uniform scatter of pts. indicates linear relationship.
   b) Curved or linear pattern indicates relationship is NOT linear.
2) Inc. or Dec. spread about the line as x inc. indicates the prediction of y will be less accurate for larger or smaller x (fan like pattern)
3) Indiv. w/ Large residuals may be regression outliers - Child 19
4) Indiv. pts. that are extreme in the x direction maybe influential pts. (Removing this pt. markedly Δ's the result of the calculation + position of the LSRL - Child 18.

ALL Pts	DEL. 18	18 In, 19 OUT
$r = -.6403$	$r = .3349$	$r = -.7561$
$r^2 = .4100$	$r^2 = .1122$	$r^2 = .5716$

Show 1-Var Calc resid
$\sum x = 0$
$\bar{x} = 0$
ALWAYS (A.P. MC)

- outliers in the y-direction have less influence on the LSRL b/c there are other pts. w/ similar values of x that "Anchor" the line well below the pt.

- Influential Pts. Pull the line towards themselves (Small residuals)

3.46

HW: 3.52, 3.57