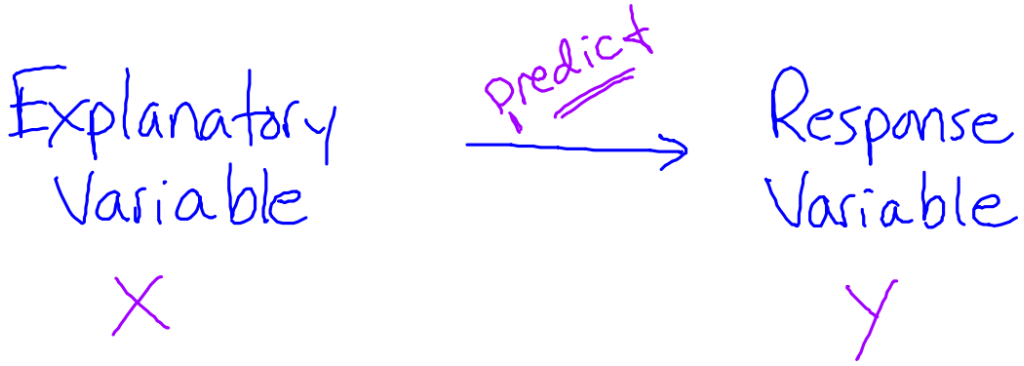
Sec 3.1


Explanatory Variable
X
predict
Response Variable
Y

Scatterplots (P. 124)

Response (y)

- Direction (Pos / Neg)
- Form (Linear, Curved, Clustered)
- Strength (Linear: Weak, Moderate, Strong)
- Outliers?

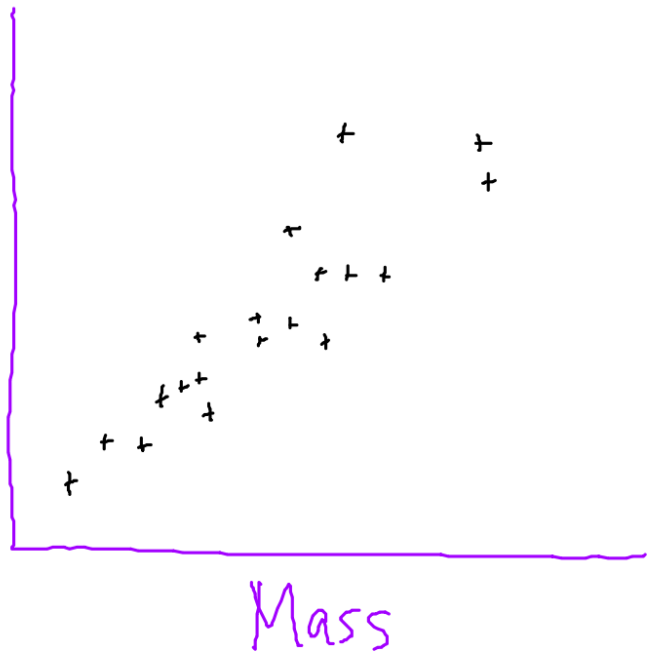
Explanatory (x)

Ex Do Heavier People Burn More Energy (P. 132)

L_1 (x) — Mass

L_2 (y) — Rate

predict



Sec 3.2

Pearson Product Moment Correlation Coefficient

- Measures the direction and "strength" of the linear relationship between 2 quantitative variables
- Correlation ≠ Causation (Ice Cream / Drownings)

Formula

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$n-1$: # pairs

$\frac{x_i - \bar{x}}{s_x}$, $\frac{y_i - \bar{y}}{s_y}$: Z-scores

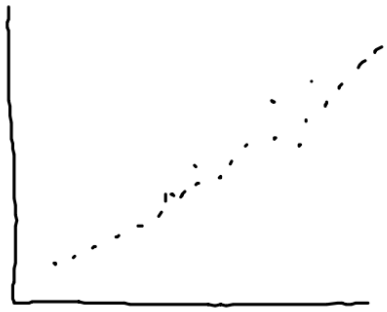
$\frac{X - M}{\sigma}$

$-1 \leq r \leq 1$

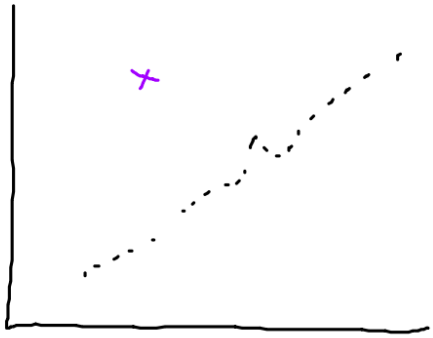
Significant (Soc Sci)	$\lvert .90 - 1.00 \rvert$	Strong Linear Relationship
	$\lvert .50 - .89 \rvert$	Moderate
	$\lvert .25 - .49 \rvert$	Weak
	$\lvert 0 - .25 \rvert$	None

P. 145

Correlation Is Not Resistant



$r = .9$



$r = .7$

Calculating Correlation

1) Use Formula.

2) Use Calculator (GPA → Math SAT)

$r = .3498$

x y

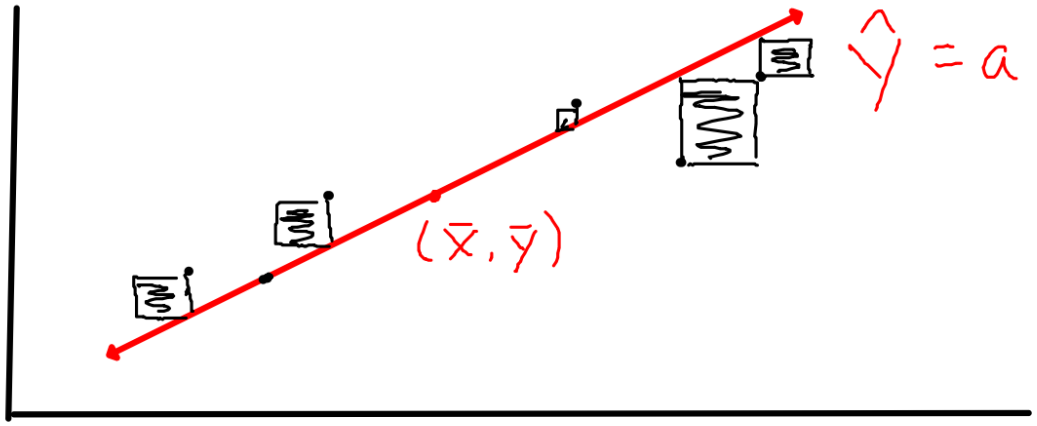
a) Data in Lists → Scatterplot

★ b) STAT → CALC → LinReg (a + bx) L_1, L_2

Sec 3.3

Least Squares Regression Line (LSRL)

- Line ($\hat{y} = a + bx$) which minimizes the sum of the squares of the vertical distances of the observed points from the line
- LSRL is a model used to make predictions



$$\text{Slope}(b) = r\frac{S_y}{S_x}$$

$$\hat{y} = a + bx$$

$$\bar{y} = a + b\bar{x}$$

$$\text{Intercept}(a) = \bar{y} - b\bar{x}$$

Finding LSRL (P.127, Ex 3.4)

L_1 (x)	L_2 (y)
Degree Days	Gas Consumption

1) Using Formulas

a) Calculate Statistics

$\bar{x} = 22.31$ $\quad \bar{y} = 5.31$ } 1 or 2 Var Stats

$Sx = 17.74$ $\quad Sy = 3.37$ } 1 or 2 Var Stats

$r = .995$ } LinReg

b) Calculate Slope and y-Intercept

$$b = r\frac{s_y}{s_x} = (.995)\frac{3.37}{17.74} \approx .189$$

$$a = \bar{y} - b\bar{x} = 5.31 - (.189)(22.31) \approx 1.09$$

c) Write Equation (In Words)

$$\hat{y} = 1.09 + .189x$$

Gas Consumption $= 1.09 + .189$ (Degree Days)

2) Using Calculator

- See Handout

Making Predictions

1) Plug 'n Chug

$$\hat{Gas} = 1.09 + .189(\text{Degree Days})$$

50

$$\hat{Gas} = 10.54 \text{ or } 1054 \text{ ft}^3$$

2) Use Calculator (If Stored)

- See Handout

Coefficient of Determination (r^2)

- Tells what proportion/percent the variation/change in y is determined by x based on a linear model

LSRL

$\hat{y} = a + bx$

$\hat{y} = \bar{y}$

% Seniors Taking SAT → SAT Scores (Pp 70-71)

$r = -.89$

Moderately strong negative linear relationship

$r^2 = .78$

The percent of seniors taking the SAT explains 78% of the variation in SAT scores

$\hat{y} = \bar{y} = 1066$

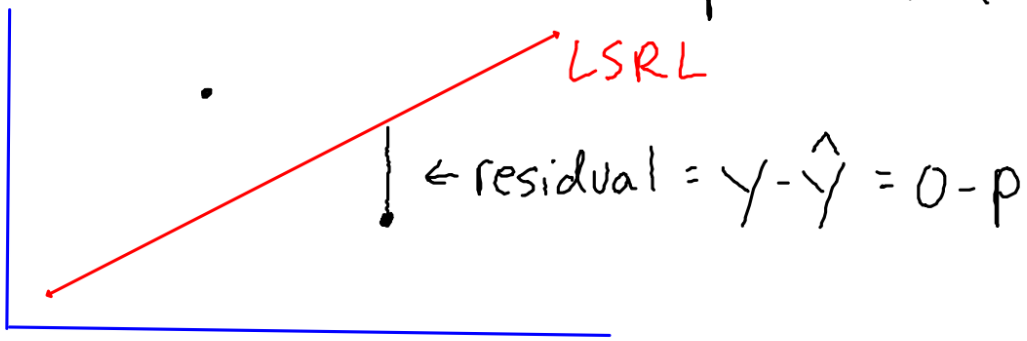
Regardless of % taking SAT → Score = 1066

$\widehat{\text{SAT Score}} = 1147 - 2.18(\%\text{ Taking SAT})$

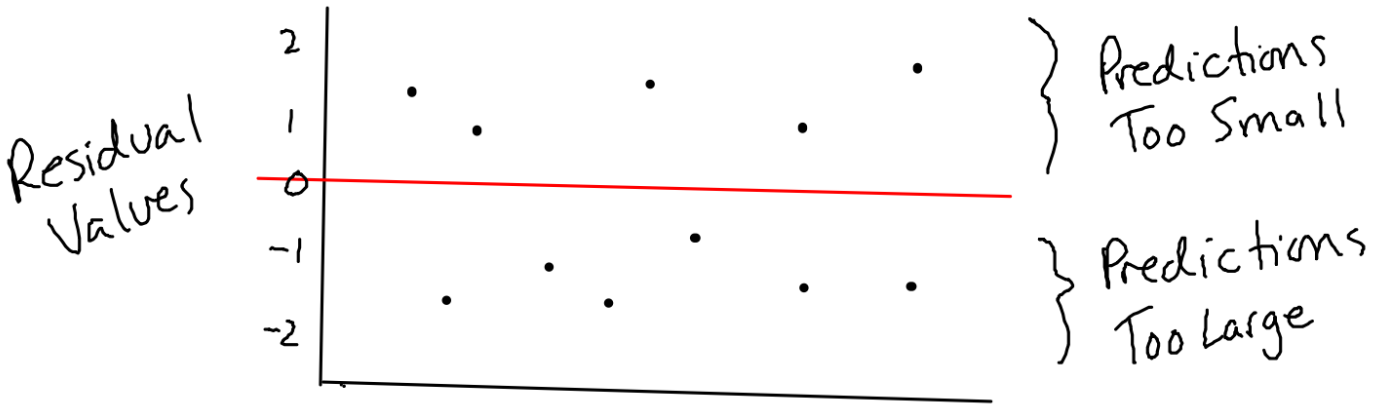
100% = 929 10% = 1125

Residuals (Errors)

Difference between observed and predicted



Residual Plot (x, residuals)

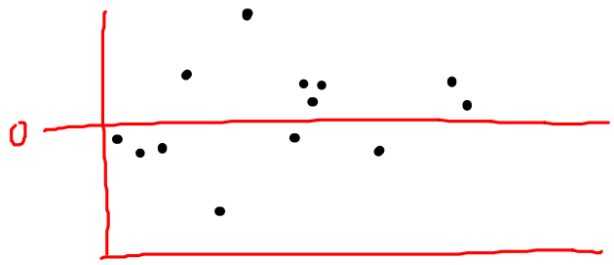


Pattern → Not Linear
No Pattern → Linear

} P. 170

Making Residual Plots (P.127)

L_1 (x)	L_2 (y)	L_3 (Residuals)
Degree Days	Gas	0 - P
		$L_2 - Y_1(L_1)$



Notes

Residual Sum = 0 .0000000000002

Mean = 0 .00000000000001

Minitab Output For Linear Regression

Variable		Coefficient ...
Constant	a	2939.93
Years	b	233.517

Chapter Summary (Pp 181-183)

Given 2 variables (x, y), is there a linear association?

↓

Scatterplot

↓

Numerical Summaries (r, r^2, LSRL)

↓

Residual Plot

- No Pattern → Linear
- Pattern → Not Linear (Ch 4)