

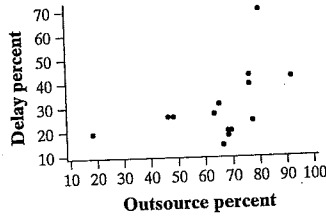
Stats ch 3-1 A 1, 5, 7, 11, 13

- ① explanatory: water temp (quantitative)
 response: weight (quantitative)

⑤ Scatterplot and data:

Airline	Outsource percent	Delay percent
AirTran	66	14
Alaska	92	42
American	46	26
America West	76	39
ATA	18	19
Continental	69	20
Delta	48	26
Frontier	65	31
Hawaiian	80	70
JetBlue	66	18
Northwest	76	43
Southwest	68	20
United	63	27
US Airways	77	24

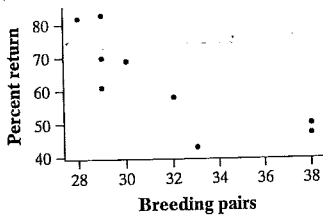
A scatterplot is shown below.



- ⑦ a) direction: positive
 form: somewhat curved
 strength: moderate to weak.
- b) (80, 70) = Hawaiian
 without Hawaiian, it's ~~a straight~~ ^{more} linear fit, but still not strong.

- ⑪ a) Southern States are lower in multiple areas on the graph.
 b) Much Lower than states with same γ_0 taking (x-value)

Breeding pairs:	28	29	29	29	30	32	33	38	38
Percent return:	82	83	70	61	69	58	43	50	47



- 1) State: Number of Pairs \uparrow Males who return \downarrow ?
 2) Plan: Make a scatterplot
 3) Do: Scatter plot show negative association.
 $r = -.7943$
 4) Conclude: this supports the theory

Stats

Sec 3-1 B

14-18, 21, 26, 27-32

14

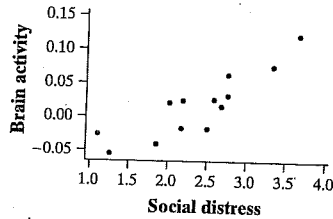
Subject	Social distress	Brain activity
1	1.26	-0.055
2	1.85	-0.040
3	1.10	-0.026
4	2.50	-0.017
5	2.17	-0.017
6	2.67	0.017
7	2.01	0.021
8	2.18	0.025
9	2.58	0.027
10	2.75	0.033
11	2.75	0.064
12	3.33	0.077
13	3.65	0.124

state: rejection \uparrow pain in Brain \uparrow ?

plan: Make scatterplot, find r

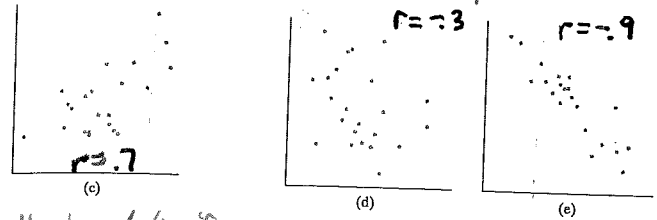
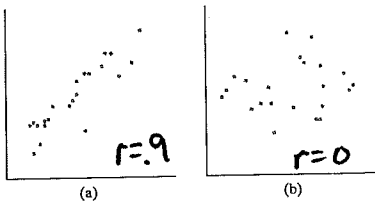
Do: Scatterplot shows strong relationship (positive) No outliers.

$r = .8782$



Social exclusion appears to trigger a pain response. As social distress increases, activity in pain-sensing areas of the brain went up.

15



16

- 1) women at 4 & 18
- 2) Father / son
- 3) Husband / wife

17

- a) gender is categorical, b) r cannot be over 1 c) r has no units

18

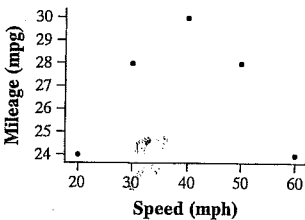
r = 0 means there is no association between research productivity and teaching rating.

21

- a) $r = .87$ means there is a strong, linear association between salt & calories
- b) it would decrease r and the strength of the relationship

26

26. A scatterplot of mileage versus speed is shown below.



r measures a "linear" relationship this plot is non-linear.

27

28

29

30

31

32

- a
- e
- d
- b
- c
- d

Stats Sec. 3-2 C 35, 36, 37, 39, 41

35) $\hat{y} = 80 - 6x$ ← days

↑
estimated weight

36) $\hat{y} = 50 + x$ ← pts above 100 on IQ

↑
predicted reading score

37) $\hat{y} = 4.62 + 1.109x$ ← hw mpg. ← city mpg

a) slope = 1.109 hw mpg goes up 1.109 for each city mpg increase

b) intercept = 4.62 this means the hw mpg = 4.62 if city mpg = 0
Not statistically significant.

c) $x = 16, \hat{y} = 22.36 \text{ mpg}$ $x = 28, \hat{y} = 35.67 \text{ mpg}$

39) $\hat{pH} = 5.43 - .0053x$ $x = \text{weeks}$

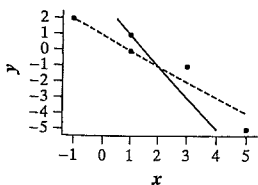
a) slope = .0053 pH decreases .0053 for each 1 week

b) intercept = 5.43 estimate pH at beginning of study

c) $x = 150, \hat{pH} = 4.635$

41) No, this is extrapolation ~~is~~ well outside the observational period
1000 months = 4000 weeks.

43



Dotted: $\hat{y} = 1 - x$

Solid: $\hat{y} = 3 - 2x$

Dotted is closer, so it fits better

45

actual = 5.08

predicted = 5.165

residual = 5.08 - 5.165 = -.085

← predicted pH was .085 too large

47

response
Men

$\bar{y} = 68.5$

$s_y = 2.7$

explanatory
Women

$\bar{x} = 64.5$

$s_x = 2.5$

$r = .5$

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

$b = (.5) \frac{2.7}{2.5} = .54$

intercept calculation

point (\bar{x}, \bar{y})

$\bar{y} = \text{slope} \cdot \bar{x} + \text{intercept}$

$68.5 = .54(64.5) + \text{intercept}$

$33.67 = \text{intercept}$

$\hat{y} = .54x + 33.67$

b) woman $x = 67$

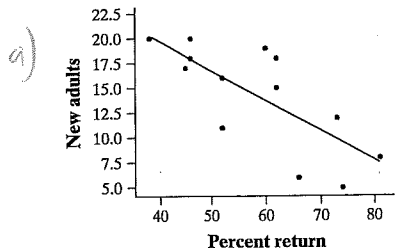
~~$\hat{y} = .54(67) + 33.67$~~

$\hat{y} = 69.85$

67 is 1 st. dev. above mean for women ($64.5 + 2.5$)
For men, this would match with $r \cdot s_y$ added to men's mean
 $.5 \cdot 2.7 = 1.35$
 $68.5 + 1.35 = 69.85$

53 data + scatter

(a) The scatterplot is



b) new = 31.9 - .304(% ret.)

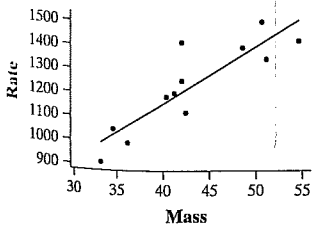
c) the new pop decreases by .304 for each increase of 1 in returning birds

d) $x = 60$ $\hat{y} = 31.9 - .304(60)$
 $\hat{y} = 13.66$

- (49) a) $r = .5$ $r^2 = .25$ This means that the linear line & equation describe 25% of the variation in husbands' height.
- b) $s' = 1.2$ The average error (residual) when using the line for predictions is 1.2 inches

(54)

a) (a) The scatterplot is shown below.



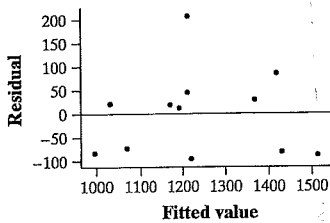
$$\hat{y} = 201.2 + 24.026x$$

b) Rate = 201 + 24 (Mass)

c) slope = 24 Means for each 1kg added to Mass, the rate goes up 24 cal/day

d) $\hat{y} = 201.2 + 24.026(45)$
 $\hat{y} = 1282.4$ cal/day

(56)



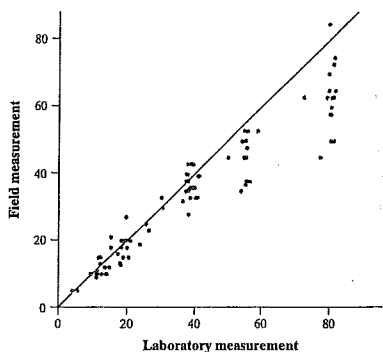
a) Linear Fit is good. center
 1 outlier, it's close to the mean so it does not influence the line much.

b) pt: (42, 207.74) is the greatest residual. resid.
 the line underestimated this value by 207

(58) $r^2 = .768$ Means that 76.8% of the variation in y (Metabolic Rate) can be described by the linear relationship.

$s = 95.08$ Means the average residual size (error) is 95.08 cal. burned per day

(59)



- a) Linear, positive relationship.
 Variation goes up for larger lab. meas.
- b) The larger depths fall below the line so the field measurements are too low
- c) slope would decrease, y -int would increase

(60)

The residuals grow as depth grows

(61) NO - there is a pattern in the residuals (clearly curved) so linear is not appropriate.

63

Predictor	Coef	SE Coef	T	P
Constant	157.68	27.68	5.70	0.001
Pairs	-2.9935	0.8655	-3.46	0.011

S = 9.46334 R-Sq = 63.1% R-Sq (adj) = 75.8%

a) $\hat{y} = 157.68 - 2.99x$
 $x = 30$ $\hat{y} = 68\%$ of males will return

b) $r^2 = .631$ so 63.1% of the variation in percent returning is explained by the number of breeding pairs (x)

c) if $r^2 = .631$ and slope is negative then $r = -.79$

d) $s = 9.46$ the typical/average error when using the line is 9.46%

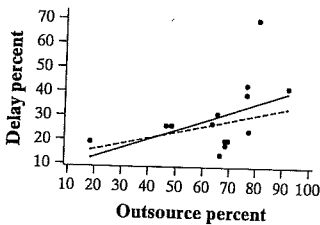
a) ~~the current value is 68%~~

b) with all pts $r = .4765$ without outlier $r = .4838$
 Not a huge difference

c) $\hat{y} = 4.73 + .3868x$ (all pts)
 $x = 76$ $\hat{y} = 34.13\%$
 $\hat{y} = 10.878 + .2495x$ (w/o outlier)
 $x = 76$ $\hat{y} = 29.84$

The outlier is influential.

Airline	Outsource percent	Delay percent
AirTran	66	14
Alaska	92	42
American	46	26
America West	76	39
ATA	18	19
Continental	69	20
Delta	48	26
Frontier	65	31
Hawaiian	80	70
JetBlue	68	18
Northwest	76	43
Southwest	68	20
United	63	27
US Airways	77	24

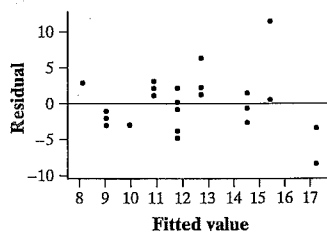
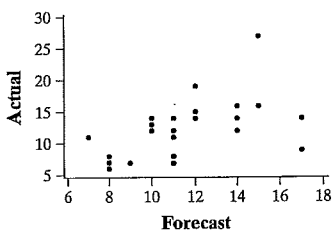


3) examples: weight, gender, food eaten, type of beer, etc.

69

Year	Forecast	Actual	Year	Forecast	Actual
1984	10	12	1997	11	7
1985	11	11	1998	10	14
1986	8	6	1999	14	12
1987	8	7	2000	12	14
1988	11	12	2001	12	15
1989	7	11	2002	11	12
1990	11	14	2003	14	16
1991	8	8	2004	14	14
1992	8	6	2005	15	27
1993	11	8	2006	17	9
1994	9	7	2007	17	14
1995	12	19	2008	15	16
1996	10	13			

State: how accurate are the predictions
 Plan: Make scatter plot make residual plot
~~make~~ Find regression equation
 Find r and r^2 and s
 Do: moderate positive association
 $\hat{y} = 1.688 + .9154x$
 $r^2 = .3$ $r = .55$ $s = 4$
 2005 strengthens the relationship
 2006 & 2007 weaken it.



Conclude: $x = 16$ $\hat{y} = 16.33$
 We do not have much confidence in this prediction since $r^2 = .3$ so only 30% of the variation in storms is explained by the line and the typical error is 4 storms

- (71) B
- (72) C
- (73) B

- (74) A
 - (75) B
 - (76) A
 - (77) D
 - (78) A →
- $\hat{y} = 6.4 + .93(100) = 99.4\text{cm}$

$\hat{y} = 6.4 + .93(60) = 62.2$
 $59 - 62.2 = -3.2$