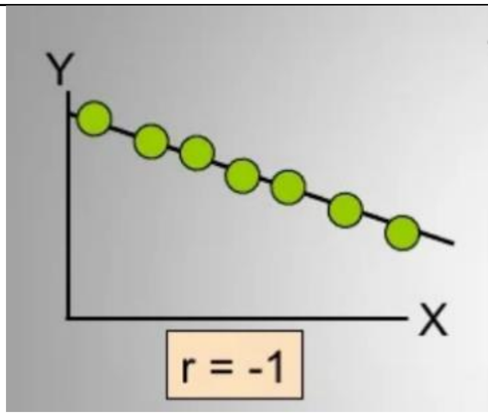
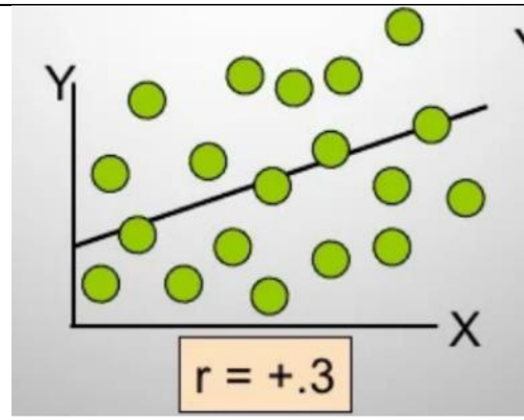
Exam 4 Practice

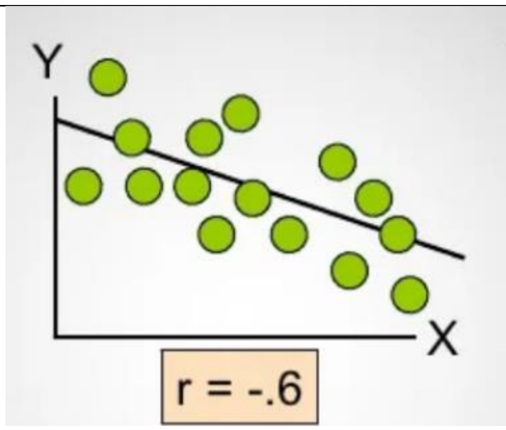
1. Look at the graphs below and interpret them.



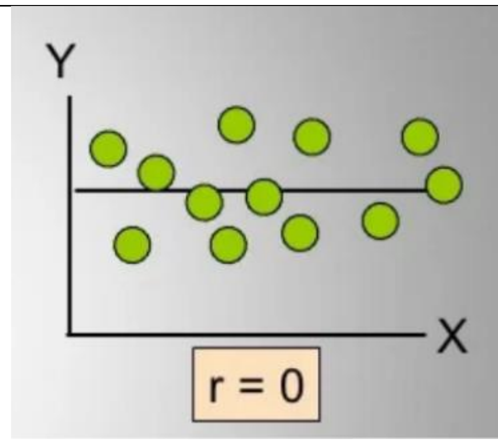
Perfect Negative Linear Relationship



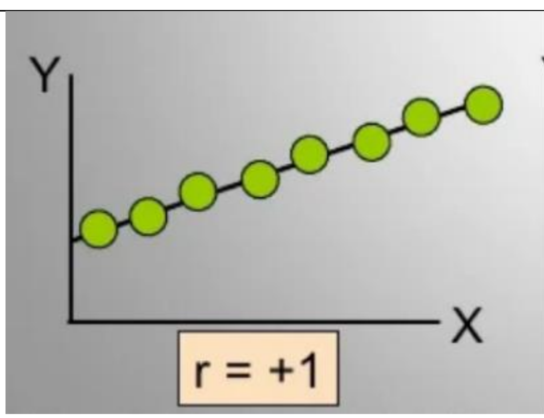
Weak Positive Linear Relationship



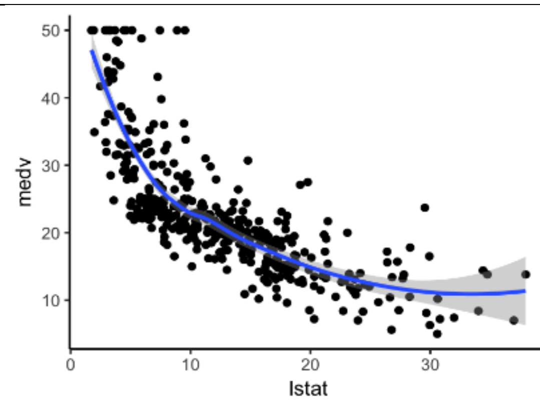
Strong Negative Linear Relationship



No Relationship



Perfect Positive Linear Relationship

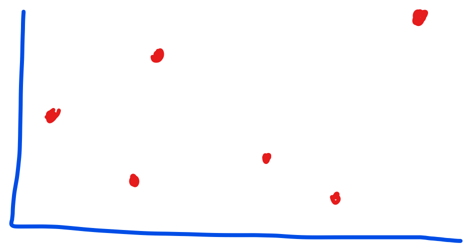


Negative Non-Linear Relationship

2. Trevor wants to examine the relationship between two field surveys on prey abundance. Answer the following questions using the data below. **<u>(By Hand)</u>**

	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6
Trial 1	13	17	15	14	12	16
Trial 2	10	19	12	17	14	11

a. Create a scatterplot.



b. Find the values below.

$\sum(x_i) = 13+17+\ldots+16 = 87$

$\sum(y_i) = 10+19+\ldots+11 = 83$

$\sum(x_i^2) = 13^2+17^2+\ldots+16^2 = 1279$

$\sum(y_i^2) = 10^2+19^2+\ldots+11^2 = 1211$

$\sum(x_i y_i) = (13\times10)+(17\times19)+\ldots+(16\times11) = 1215$

c. What is the slope?

$$\hat{\beta}_1 = \frac{(6)(1215)-(87)(83)}{(6)(1279)-(87)^2} = 0.66$$

d. What is the vertical intercept? $\bar{y} = \frac{83}{6} = 13.83$; $\bar{x} = \frac{87}{6} = 14.5$

$\hat{\beta}_0 = (13.83)-(0.66)(14.5) = 4.26$

e. Write the linear regression model.

$$\hat{y} = 4.26 + 0.66x$$

f. What is the coefficient of correlation?

$$R = \frac{(6)(1215) - (87)(83)}{\left(\sqrt{(6)(1279) - (87)^2}\right)\left(\sqrt{(6)(1211) - (83)^2}\right)} = 0.35$$

g. What is the coefficient of determination?

$$R^2 = (0.35)^2 = 0.1225 = 12.25\%$$

h. Interpret parts f and g.

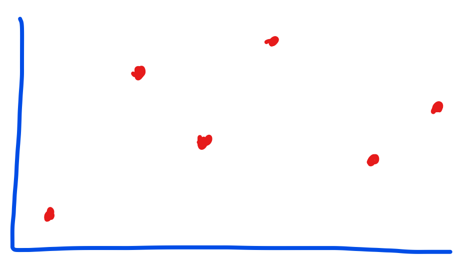
R: Weak Positive Linear Relationship

R^2: 12.25% of variation in y can be explained by x.

3. Trevor saw that there was data on prey abundance from several decades ago. Analyze that data so Trevor can compare it to his. **(By Hand)**

	Sector 1	Sector 2	Sector 3	Sector 4	Sector 5	Sector 6
Trial 1	12	14	15	13	17	16
Trial 2	9	13	17	16	14	12

a. Create a scatterplot.



b. Find the values below.

$$\sum(x_i) = 12 + 14 + \ldots + 16 = 87$$

$$\sum(y_i) = 9+13+\ldots+12 = 81$$

$$\sum(x_i^2) = 12^2+14^2+\ldots+16^2 = 1279$$

$$\sum(y_i^2) = 9^2+13^2+\ldots+12^2 = 1135$$

$$\sum(x_iy_i) = (12\times 9)+(14\times 13)+\ldots+(16\times 12) = 1183$$

c. What is the slope?

$$\hat{\beta}_1 = \frac{(6)(1183)-(87)(81)}{(6)(1279)-(87)^2} = 0.49$$

d. What is the vertical intercept? $\bar{y} = \frac{81}{6} = 13.5$; $\bar{x} = \frac{87}{6} = 14.5$

$$\hat{\beta}_0 = (13.5)-(0.49)(14.5) = 6.40$$

e. Write the linear regression model.

$$\hat{y} = 6.40 + 0.49x$$

f. What is the coefficient of correlation?

$$R = \frac{(6)(1183)-(87)(81)}{\left(\sqrt{(6)(1279)-(87)^2}\right)\left(\sqrt{(6)(1135)-(81)^2}\right)} = 0.32$$

g. What is the coefficient of determination?

$$R^2 = (0.32) = 0.1024 = 10.24\%$$

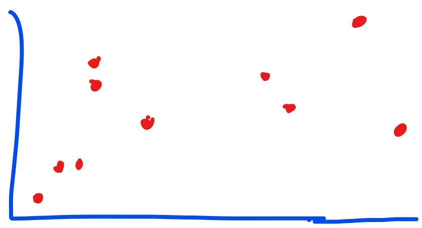
h. Interpret parts f and g. R: Weak Positive Linear Relationship

R^2: 10.24% of variation in y can be explained by x.

4. A study considering whether hours spent studying a week has a relationship with the number of complaints for headaches/migraines (of any duration) in college students was conducted. Using the information below, solve the following questions. **(By Calculator)**

$(X)L_1 \rightarrow$	Hrs	24	43	33	22	21	18	35	63	22	55	47	33
$(Y)L_2 \rightarrow$	#oC	2	5	3	6	2	1	0	3	5	7	4	0

a. Create a scatterplot.



b. Find the values below. (Via 2-Var Stat)

$\sum(x_i) = 416$

$\sum(y_i^2) = 178$

$\sum(y_i) = 38$

$\sum(x_i y_i) = 1426$

$\sum(x_i^2) = 16764$

Via LinRegTTest

c. What is the slope?

$\hat{\beta}_1 = b = 0.05$

d. What is the vertical intercept?

$\hat{\beta}_0 = a = 1.56$

e. Write the linear regression model.

$\hat{y} = 1.56 + 0.05x$

Still via LinRegTTest

f. What is the coefficient of correlation?

$R = r = 0.30$

g. What is the coefficient of determination?

$R^2 = r^2 = 0.09 = 9\%$

h. Interpret parts f and g.

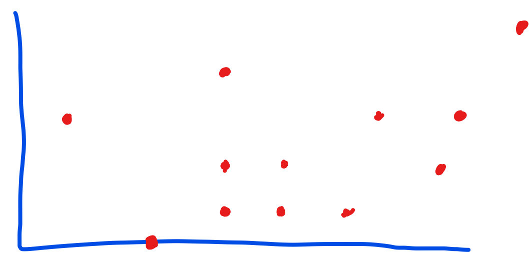
R: Weak Positive Linear Relationship

R^2: 9% of variation in y can be explained by X.

5. In a similar study to question 4, it was exploring whether the average number of hours of sleep each night has a relationship with the number of complaints for headache/migraines (of any duration) each day. Using the information below, solve the following questions. **(By Calculator)**

(x) L_3 → Hrs	3	4	2	6	6	7	5	4	8	7	4	5
(y) L_4 → #oC	0	2	3	1	3	2	2	1	5	3	4	1

a. Create a scatterplot.



b. Find the values below. (Via 2-Var Stat)

$\sum(x_i) = 61$

$\sum(y_i^2) = 83$

$\sum(y_i) = 27$

$\sum(x_i y_i) = 148$

$\sum(x_i^2) = 345$

Via LinRegTTest

c. What is the slope?

$\hat{\beta}_1 = b = 0.31$

d. What is the vertical intercept?

$\hat{\beta}_0 = a = 0.68$

e. Write the linear regression model.

$\hat{y} = 0.68 + 0.31x$

Still via LinRegTTest

f. What is the coefficient of correlation?

$R = r = 0.39$

g. What is the coefficient of determination?

$R^2 = r^2 = 0.15 = 15\%$

h. Interpret parts f and g.

R: Weak Positive Linear Relationship

R^2: 15% of variation in y can be explained by x.

6. Use the R output below to answer the questions below.

```
Call:
lm(formula = myResponse ~ myFactor)

Residuals:
     1      2      3      4      5      6      7
 1.239 -9.249 -5.164  4.104 -1.030  5.617  4.483

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 102.4925     5.1381  19.948 5.85e-06 ***
myFactor     -3.6219     0.5649  -6.411  0.00137 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.055 on 5 degrees of freedom
Multiple R-squared:  0.8915,    Adjusted R-squared:  0.8699
F-statistic:  41.1 on 1 and 5 DF,  p-value: 0.00137
```

$\hat{\beta}_0$ $\hat{\beta}_1$ R^2

a. Write the Linear Regression Model for this question.

$\hat{y} = 102.4925 - 3.6219x$

b. What is the coefficient of determination?

$R^2 = 0.8915 = 89.15\%$

c. What is the coefficient of correlation? What is sign of $\hat{\beta}_1$?

→ thats the sign for R!

$R = \sqrt{0.8915} = -0.94$

i. What is the strength?

Very Strong

ii. Positive or Negative?

Negative

d. Interpret the values found for part b and c.

R: Very Strong Negative Linear Relationship

R^2: 89.15% of variation in y can be explained by x.

7. Use the R output below to answer the questions below.

```
Call:
lm(formula = y ~ x, data = oilAndRepairsDF)

Residuals:
     Min       1Q   Median       3Q      Max
-108.824  -63.971   -2.941   61.029  114.706

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   673.53      51.99   12.95 1.19e-06 ***
x             -88.24      14.71   -6.00 0.000323 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 79.06 on 8 degrees of freedom
Multiple R-squared:  0.8182,	Adjusted R-squared:  0.7955
F-statistic:    36 on 1 and 8 DF,  p-value: 0.0003234
```

$\hat{\beta}_0$ 673.53

$\hat{\beta}_1$ -88.24

R^2 0.8182

a. Write the Linear Regression Model for this question.

$$\hat{y} = 673.53 - 88.24x$$

b. What is the coefficient of determination?

$$R^2 = 0.8182 = 81.82\%$$

c. What is the coefficient of correlation? $\hat{\beta}_1$ is $(-)$

$$R = \sqrt{0.8182} = -0.90$$

i. What is the strength?

Very Strong

ii. Positive or Negative?

Negative

d. Interpret the values found for part b and c.

R: Very Strong Negative Linear Relationship

R^2: 81.82% of variation in y can be explained by x.

8. For this question, we are going to solve this example from the notes. We will be following along with putting the data in our calculators!

Example 11.20. consider the following situations.

Situation 3

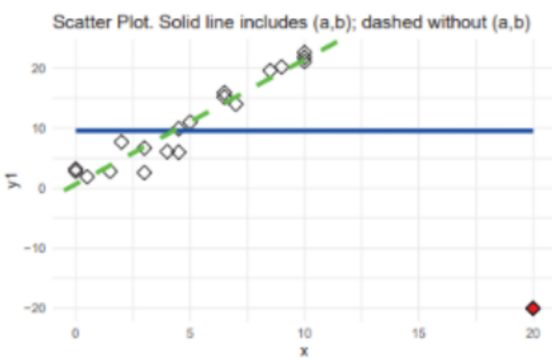
Consider the data, which includes the point $(a, b) = (20, -20)$ at the end and figures shown below. Note that the least squares regression line is

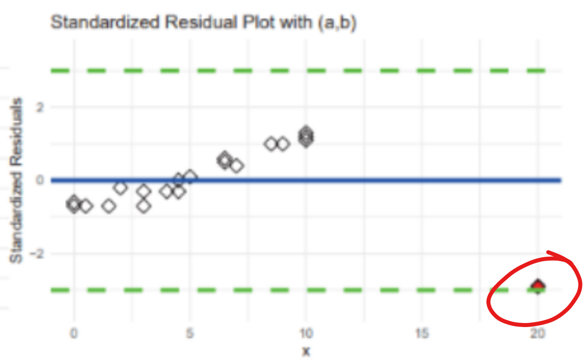
$$\hat{y} = 0x + 9.56.$$

L1 — x_i
L2 — y_i
L3 — $\hat{y}_i$
$L_2 - L_3 = L_4$ — Residual
$\frac{L_4}{S_e} = L_5$ — Std Resid

	1	2	3	4	5	6	7	8	9	10
x_i	1.50	3.00	0.00	0.50	5.00	6.50	8.50	9.00	0.00	10.00
y_i	2.80	2.60	3.20	1.90	11.00	15.90	19.60	20.20	2.80	22.60
$\hat{y}_i$	9.56	9.56	9.56	9.56	9.56	9.55	9.55	9.55	9.56	9.55
Residual $e_i = y_i - \hat{y}_i$	-6.76	-6.96	-6.36	-7.66	1.44	6.35	10.05	10.65	-6.76	13.05
Std Resid e_i/s_e	-0.66	-0.68	-0.62	-0.74	0.14	0.62	0.98	1.03	-0.66	1.27

	11	12	13	14	15	16	17	18	19	20
x_i	10.00	4.50	6.50	4.50	3.00	4.00	7.00	10.00	2.00	**20.00**
y_i	21.10	6.00	15.20	9.90	6.70	6.10	14.00	21.80	7.70	**-20.00**
$\hat{y}_i$	9.55	9.56	9.55	9.56	9.56	9.56	9.55	9.55	9.56	**9.53**
Residual $e_i = y_i - \hat{y}_i$	11.55	-3.56	5.65	0.34	-2.86	-3.46	4.45	12.25	-1.86	**-29.53**
Std Resid e_i/s_e	1.12	-0.35	0.55	0.03	-0.28	-0.34	0.43	1.19	-0.18	**-2.87**

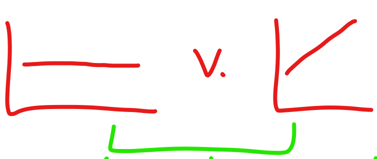




A. Is the point $(20, -20)$ an outlier?

No b/c it is <|3|

B. Is the point $(20, -20)$ influential?



Very different

Yes

9. We will be doing the same thing as question 7 for this question as well.

Example 11.19. consider the following situations.

Situation 2

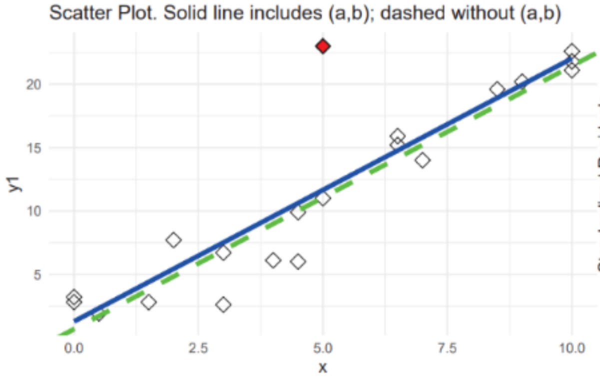
Consider the data, which includes the point $(a, b) = (5, 23)$ at the end and figures shown below. Note that the least squares regression line is

$$\hat{y} = 2.08x + 1.27.$$

	1	2	3	4	5	6	7	8	9	10
L_1 — x_i	1.50	3.00	0.00	0.50	5.00	6.50	8.50	9.00	0.00	10.00
L_2 — y_i	2.80	2.60	3.20	1.90	11.00	15.90	19.60	20.20	2.80	22.60
L_3 — $\hat{y}_i$	4.39	7.50	1.27	2.31	11.65	14.77	18.92	19.96	1.27	22.03
$L_2 - L_3 = L_4$ — Residual $e_i = y_i - \hat{y}_i$	-1.59	-4.90	1.93	-0.41	-0.65	1.13	0.68	0.24	1.53	0.57
$\frac{L_4}{S_e} = L_5$ — Std Resid e_i/s_e	-0.47	-1.44	0.57	-0.12	-0.19	0.33	0.20	0.07	0.45	0.17

	11	12	13	14	15	16	17	18	19	20
x_i	10.00	4.50	6.50	4.50	3.00	4.00	7.00	10.00	2.00	**5.00**
y_i	21.10	6.00	15.20	9.90	6.70	6.10	14.00	21.80	7.70	**23.00**
$\hat{y}_i$	22.03	10.62	14.77	10.62	7.50	9.58	15.80	22.03	5.43	**11.65**
Residual $e_i = y_i - \hat{y}_i$	-0.93	-4.62	0.43	-0.72	-0.80	-3.48	-1.80	-0.23	2.27	**11.35**
Std Resid e_i/s_e	-0.27	-1.36	0.13	-0.21	-0.24	-1.02	-0.53	-0.07	0.67	**3.34**

Scatter Plot. Solid line includes (a,b); dashed without (a,b)



Standardized Residual Plot with (a,b)

A. Is the point $(5, 23)$ an outlier?

Yes b/c it is $> |3|$

B. Is the point $(5, 23)$ influential?

v. No

Very Similar