

Form of the relationship

Strength of the relationship

Direction of the relationship

We have data, from an experiment or a survey

First job:

Which is the explanatory which is the response variable?

- Choose by logic
- If it is an experiment the description will tell you.

Second job:

You will make scatter plots.

EXPLANATORY IS X, RESPONSE IS Y

Third job:

Describe what you see

NEXT:

If it is a line....

Positive is/here < relationship is
F, S, D, O

Babies.idf

Cats.idf my dino Dinosaurs.id

Explanatory relationships show a pattern we can use to make predictions.



Vocabular for describing patterns

Form: linear, curved, clusters

forming

Strength: Strong, weak

- 2 When one variable increases, the other variable is very likely, or not, to increase a predictable amount.
- 1 How closely the points fit to the form you described ("Yeah, I see a line" or "Well, if you squint..")

Direction: Positive or negative no relation

As one increase the other increases

As one increase the other decreases

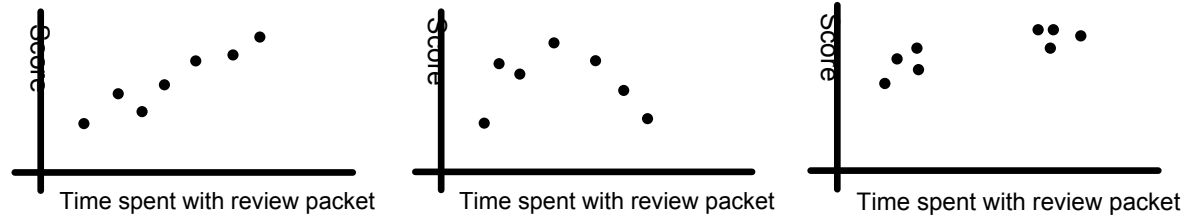
Outliers X or Y

X far from rest of data left to right, a gap left to right

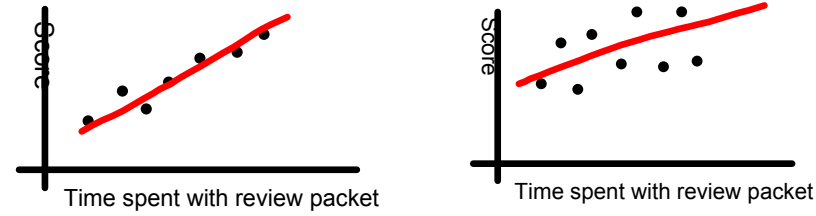
off the pattern

Y far above or below the data

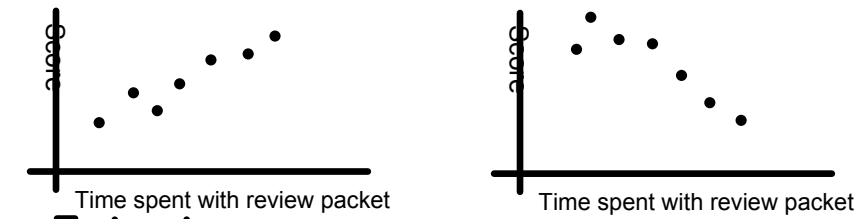
Form:



Strength:

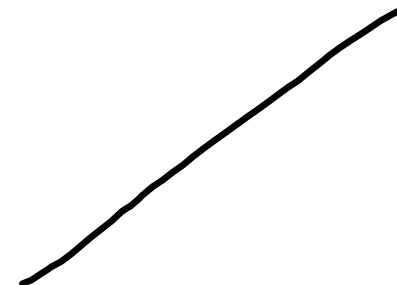
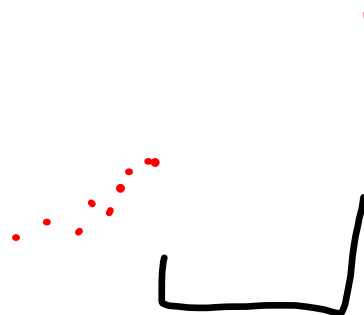
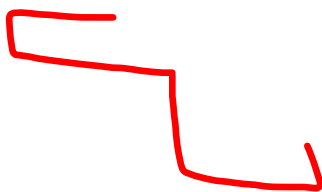
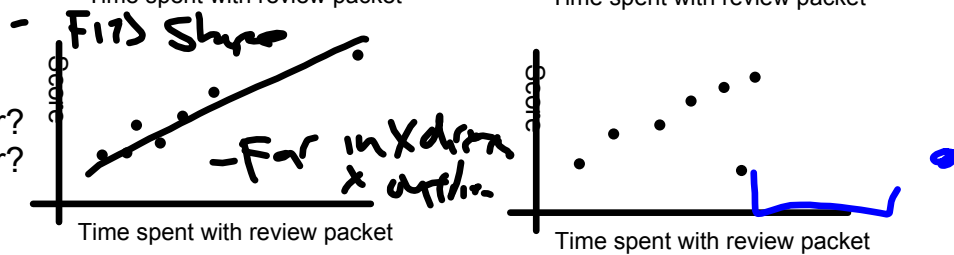


Direction:



Outliers:

Which is x-outlier?
Which is y-outlier?

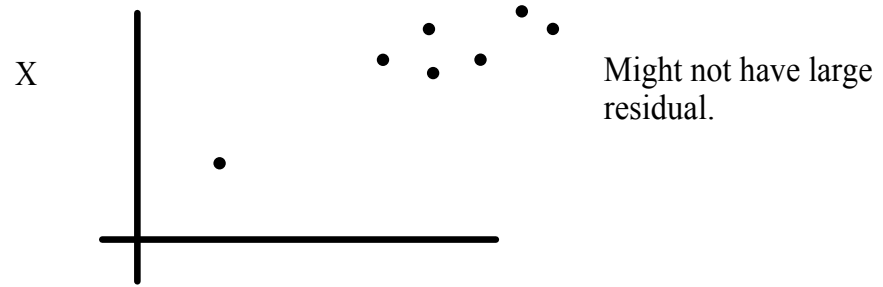
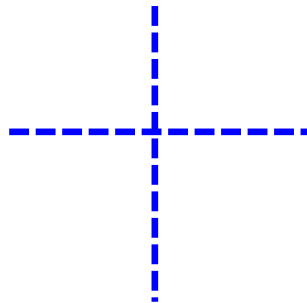


Outliers: Breaks the pattern

Two kinds
x and y

Influential:

Changes your results.
slope, correlation, intercept
x's outliers can be more potent



Interpreting the LEAST SQUARES REGRESSION LINE

Slope: The value of slope states the number of units of change in the response variable for each unit of change in the explanatory variable ON AVERAGE

Intercept: The value of the intercept states the PREDICTED value of the response variable when the explanatory variable is 0.



Using the line:

Plug in the correct units, write your answer with the correct units.

$$\hat{y} = -0.13 + .018x$$

$x = \underline{\hspace{2cm}}$



$$\hat{y} = \text{predicted } \underline{\hspace{2cm}}$$

How analysis of two variables allows us to make predictions:

Make scatter plot

Analyze scatter plot

- o When there is a relation state FSD
- o When there is no relation says and stop

When Form = Linear, create LSRL

- o Interpret the LSRL
- o Interpret the correlation

→ Use LSRL to make predictions

Interpreting the LEAST SQUARES REGRESSION LINE

Slope: The value of slope states the number of units of change in the response variable for each unit of change in the explanatory variable ON AVERAGE

Intercept: The value of the intercept states the PREDICTED value of the response variable when the explanatory variable is 0.

$$\hat{y} = 0.355x + .4473$$

- For each increase of 1 (units) in (explanatory variable) there is a (slope #) increase in the (response variable) ON AVERAGE.

- When the (explanatory variable) is 0 the (response variable) IS PREDICTED to be (intercept #).

where x=the height of the dinosaur in meters

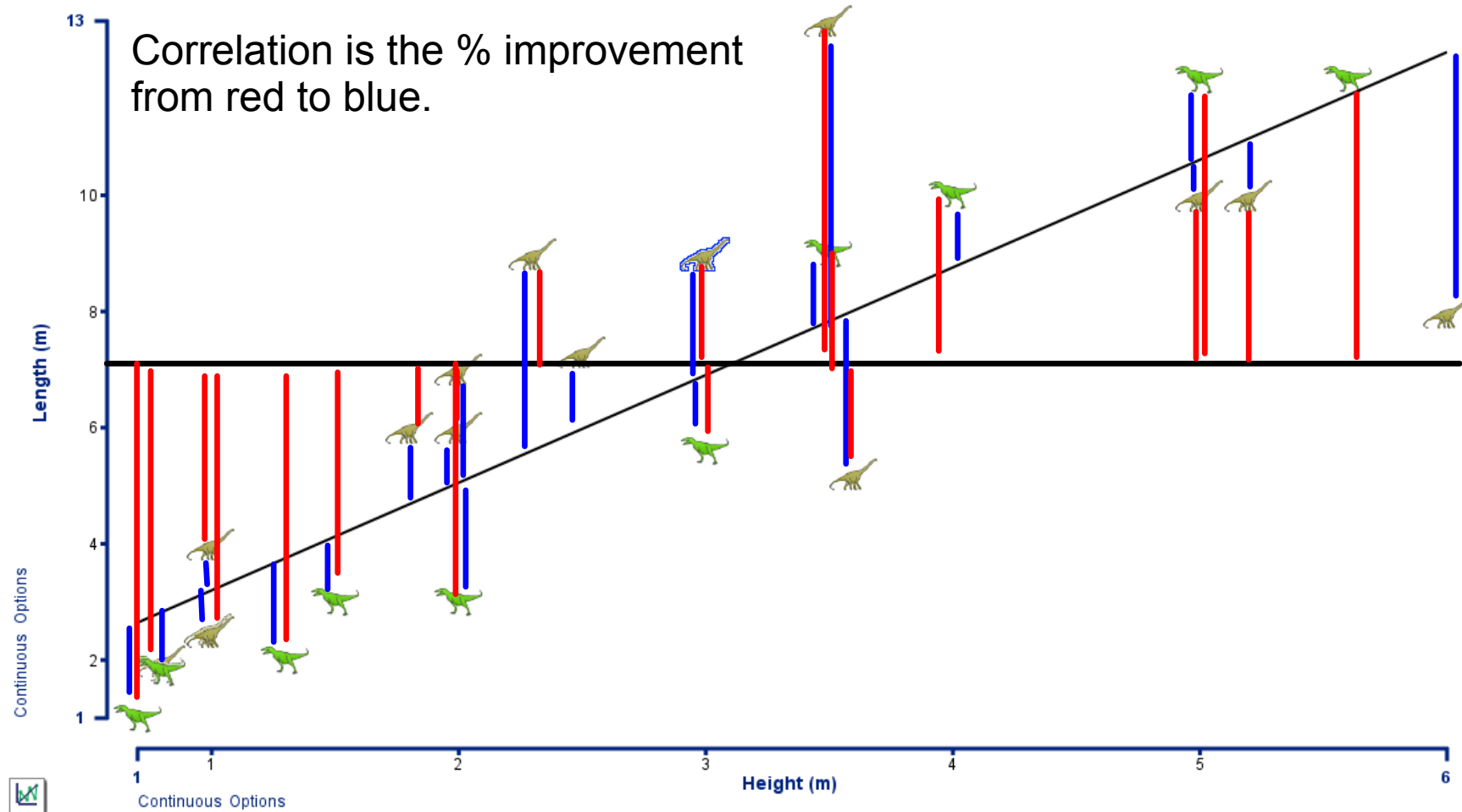
\hat{y} = the predicted length in meters

my dino Dinosaurs.idf





Correlation Coefficient: 0.84
View: 26 records



Interpreting the LEAST SQUARES REGRESSION LINE

$$\hat{y} = 0.355x + .4473$$

where x=the height of the dinosaur in meters

\hat{y} = the predicted length in meters

Key items:

- on average
- predicted
- using hats
- units

Interpreting r^2 : (coefficient of determination)

- r^2 states the percentage of the variation in the response variable that is explained by a linear regression of the response variable on the explanatory variable.
- what is strong depends on context.

Strong correlations are above .85. Moderate are between .65 and .85. Weak .35 to .65

No relation $<.35$

Interpreting r : (correlation coefficient)

- r is closely related to slope.
- r tells us the direction of slope
- To discuss strength, square it.

Interpreting the LEAST SQUARES REGRESSION LINE

Interpreting r^2 :

- r^2 states the percentage of the variation in the response variable that is explained by a linear regression of the response variable on the explanatory variable.
- what is strong depends on context.

Strong correlations are above .85. Moderate are between .65 and .85. Weak .35 to .65.
 No relation $< .35$

___ (r^2 number as a %) ___ of the variation in ___ (response variable) ___
 is explained by a linear regression on
 ___ (explanatory variable).

$$r = -.18$$

$$r^2 = .0324$$

6.25-.27, 6.29, .31, ~~...~~

Interpreting the LEAST SQUARES REGRESSION LINE

Using the line:

Plug in the correct units, write your answer with the correct units.

But only within the range of validity

outside is extrapolation.

DO NOT EXTRAPOLATE

Range of validity; The range of the x values in the original data.

Calculator skills

Correlation, correlation coefficient = r

Direction and strength

coefficient of determination = r^2
strength

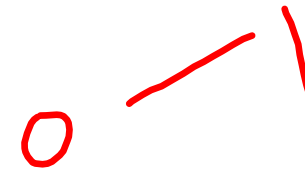
- Correlation describes
- the strength and
 - direction of
 - the **straight line** relation

Objectives today:

- Recap features of correlation
- Execute a complete analysis

Correlation describes

- the strength and
- direction of
- the **straight line** relation



- Positive is positive, negative is negative
- -1 to 1, closer to 1s is stronger **-1 to 1**
- Does not care about units
- Correlation is the same if you switch x and y
- Straight line only
- Really affected by outliers

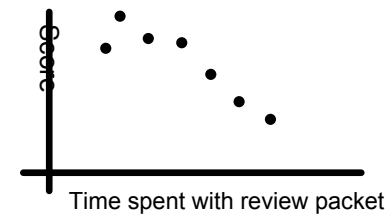
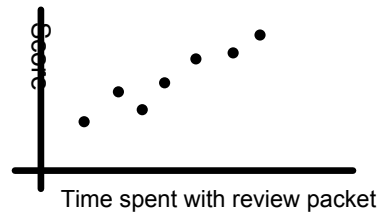
Correlation describes

- the strength and
- direction of
- the **straight line** relation

Interpret

- relationship*
- Positive is positive, negative is negative
 - -1 to 1, closer to 1s is stronger
 - Does not care about units
 - Correlation is the same if you switch x and y
 - Straight line only
 - Really affected by outliers

% of variation
0 - 1



Correlation describes

- the strength and
- direction of
- the **straight line** relation

- ✓ - Positive is positive, negative is negative
- ✓ - **-1 to 1, closer to 1s is stronger**
- ⌋ - Does not care about units
- ⌋ - Correlation is the same if you switch x and y
- ✓ - Straight line only
- ✓ - Really affected by outliers

RANT

Do not get into using "r" for strength.
Square it and then the numbers make sense.

"Interpret an $r = -.8$ "

- Negative relation
- r^2 of .64, indicating a weak linear relation

Strong $r^2 = .85$ to 1

Moderately strong $r^2 = .65$ to .85

Weak $r^2 = .45$ to .65

Very weak to none $r^2 < .45$

What does correlation say?

The strength and the direction of the relationship.

Ask me for the strength and I will square it and look at that number.

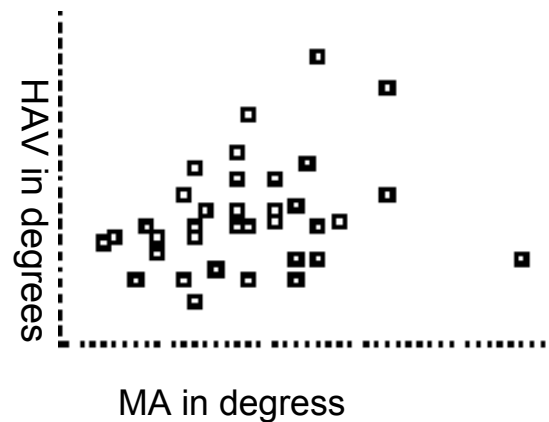
TABLE 6.4 The Severity of Metatarsus Adductus and Hallux Abducto Valgus

HAV angle	MA angle	HAV angle	MA angle	HAV angle	MA angle
28	18	21	15	16	10
32	16	17	16	30	12
25	22	16	10	30	10
34	17	21	7	20	10
38	33	23	11	50	12
26	10	14	15	25	25
25	18	32	12	26	30
18	13	25	16	28	22
30	19	21	16	31	24
26	10	22	18	38	20
28	17	20	10	32	37
13	14	18	15	21	23
20	20	26	16		

MA makes the front of your foot turn. HAV is big toe problem requiring surgery.
 - make scatterplot - describe the relation - calculate correlation - do you think there is relation?

Relationship of the HAV impact and MA impact

- make scatterplot
- describe the relation
- calculate correlation
- do you think there is relation?



```

LinReg
y=ax+b
a=.2694129029
b=9.96702989
r2=.0912866377
r=.3021367864

```

The points spread out in a fan shape.
 There is no strong line or curve.
 The relationship is positive.

There is one apparent x and y outlier at 50° MA and 12° HAV.

The weak PATTERN shown in the graph is verified by the very low r^2 of .09.

There appears to be no LINEAR relation between MA and HAV, but I am curious about the influence of the outlier.

Correlation measures the strength of a line.

Lets talk about the line

Calculator skills

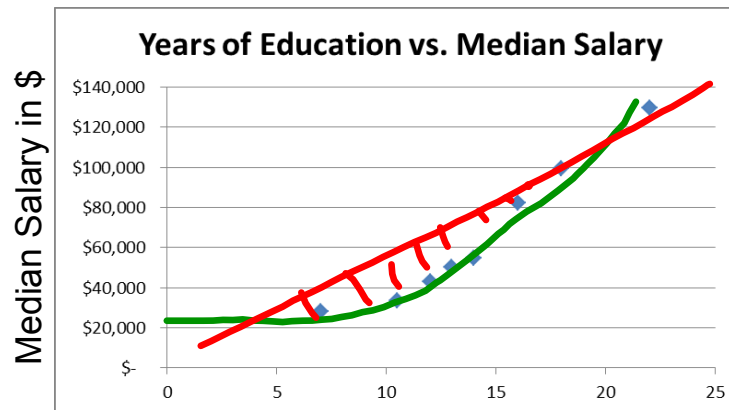
- Diagnostics on
- Running a line
- Piping it
- Plotting scatter plot and line

$$Y = mx + b, y = ax + b$$

a = slope

b = intercept

Years of School	Median salary
7	\$ 27,964
11	\$ 33,435
12	\$ 43,165
13	\$ 50,359
14	\$ 54,861
16	\$ 82,197
18	\$ 99,516
22	\$ 129,773



16.8 Years of Education

Correlation = ~~0.970466~~ $r^2 = 0.941805$

~~$\hat{y} = \$7,448.49x - \$39,557.58$~~ ~~73.6~~ $7,508.03$ $54,0897.11$

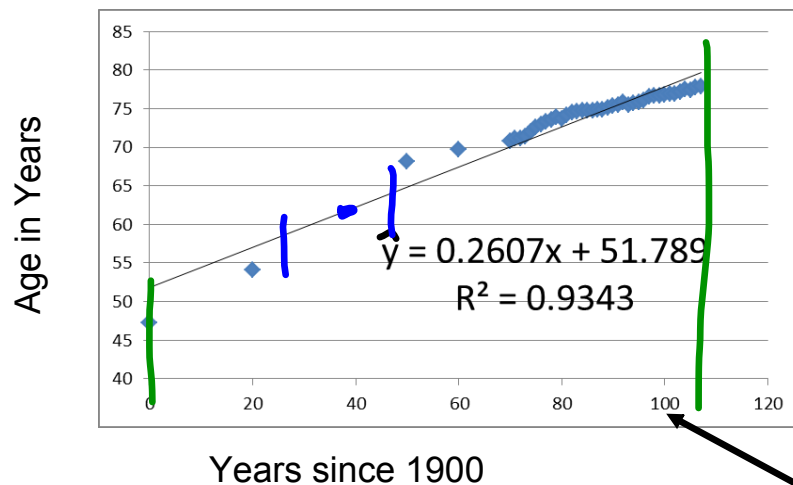
where x =years of school and \hat{y} = predicted median salary

- 1) Describe the relationship and comment on the validity of the linear relation?
- 2) What should you earn with 20 years of education?
- 3) What should you earn with 40 years of education?

Comment on this result

- 4) Interpret the intercept
- 5) Interpret the slope
- 6) Interpret r^2

Life Expectancy at Birth



where x = years since 1900
and
 \hat{y} = predicted life expectancy at birth

2000

When will life exp be 100 years?

- 1) Describe the relation and comment on the validity of the linear relation?
- 2) What is the estimated life expectancy in 1940?
- 3) What life expectancy be in 2050?
Comment on this result
- 4) Interpret the intercept
- 5) Interpret the slope
- 6) Interpret r^2

___(r² number as a %)___ of the variation in ___(response variable)___
is explained by a linear regression on
___(explanatory variable).

- When the ___(explanatory variable)___
is 0 the ___(response variable)___ IS PREDICTED to be ___(intercept #).

- For each increase of 1 ___(units)___ in ___(explanatory variable)___
there is a ___(slope #)___ increase in the ___(response variable)___
ON AVERAGE.

- Make a scatter plot
- Figure out explanatory vs response
- F,S,D
 - x and y outliers
- Run regression (write it properly)
- Know six facts about r
 - Positive is positive, negative is negative
 - -1 to 1, closer to 1s is stronger
 - Does not care about units
 - Correlation is the same if you switch x and y
 - Straight line only
 - Really affected by outliers
- Interpret r
 - o Direction
 - o Strength
- Interpret slope
- Interpret intercept
- Interpret r^2
- Recognize to range of validity/extrapolation problem
- Use the line to predict, use the line "backwards"

Correlation describes

- the strength and
- direction of
- the **straight line** relation

- Positive is positive, negative is negative
- -1 to 1, closer to 1s is stronger
- Does not care about units
- Correlation is the same if you switch x and y
- Straight line only
- Really affected by outliers

Attachments

my dino Dinosaurs.idf

Babies.idf

Cats.idf

regressdemline.gsp