Chapter 10: Statistics

10.1 Two variable distributions

Contingency Tables

A contingency table illustrates a two-variable distribution.
It is another way to display statistical data, like a scatter pl

Ex: Here is some data about the **number of hours of sleep** (
that a student gets before an exam and their **grade on t**
exam (y).

(2, 50) (3, 50) (4, 45) (4, 35) (5, 50) (5, 53) (6, 60) (6, 45) (7, 70) (7, 64) (8
80) (8, 55) (8, 75) (8, 78) (8, 83) (9, 79) (9, 30) (9, 89) (9, 62) (9, 93) (9, 8
(10, 85) (10, 71)
(10, 92) (10, 88)

Do you think there is a correlation between the two variab
Fill in the table:

Grade % (y) \ Hours Slept (x)	2	3	4	5	6	7	8	9	10
[30, 40[	0	0	1	0	0	0	0	1	0
[40, 50[	0	0	1	0	1	0	0	0	0
[50, 60[	1	1	0	2	0	0	1	0	0
[60. 70[	0	0	0	0	1	1	0	1	0
[70, 80[	0	0	0	0	0	1	2	1	1
[80, 90[	0	0	0	0	0	0	2	2	2
[90, 100[	0	0	0	0	0	0	0	1	1

HHT
5

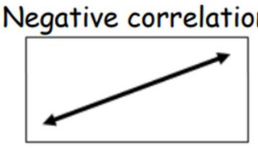
- Correlation exists if...

large numbers on the diagonal, small numbers (look for zeros) in the corners

- the more obvious the pattern = stronger correlation

Positive correlation:

Negative correlation



when in a contingency table

Grade % (y) \ Hours Slept (x)	2	3	4	5	6	7	8	9	10	Total
[30, 40[			1					1		2
[40, 50[			1		1					2
[50, 60[	1	1		2			1			5
[60, 70[					1			1		3
[70, 80[						1	2	1	1	5
[80, 90[							2	2	2	6
[90, 100[								1	1	2
Total	1	1	2	2	2	2	5	6	4	25

Contingency Tables

1. What kind of correlation exists, given the table below?

X \ Y	[0,1[	[1,2[	[2,3[	[3,4[	[4,5[
2	0	0	0	0	4
4	0	0	0	3	0
6	0	0	3	0	0
8	0	2	0	0	0
10	3	0	0	0	0

neg
Strong

2. What kind of correlation exists, given the table below?

X \ Y	[0,20[	[20,40[	[40,60[	[60,80[	[80,100[
10	6	0	0	0	0
20	0	7	0	0	0
30	0	0	9	0	0
40	0	0	0	8	0
50	0	0	0	0	9

+ Strong

3. What kind of correlation exists, given the table below?

X \ Y	[0,1[	[1,2[	[2,3[	[3,4[	[4,5[
1	3	1	5	2	3
2	2	4	2	4	3
3	5	2	4	1	4
4	3	3	3	3	3
5	4	4	1	6	5

Strength

- [x] Zero
- [ ] Weak
- [ ] Strong

Direction

- [ ] Positive
- [ ] Negative

4. What kind of correlation exists, given the table below?

X \ Y	[0,2[	[2,4[	[4,6[	[6,8[	[8,10[
10	0	0	0	2	2
20	0	0	3	4	2
30	1	1	4	1	0
40	2	3	3	2	0
50	4	2	0	0	0

Strength | Direction

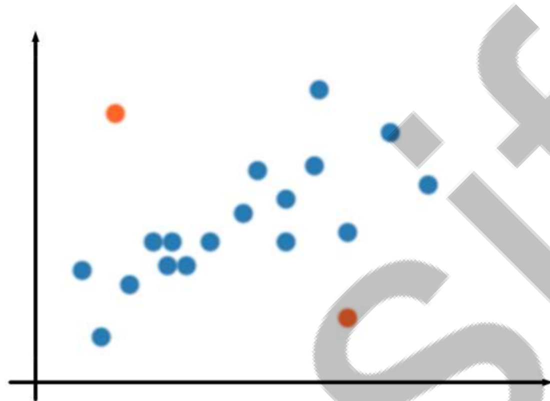
- [ ] Zero
- [x] Weak — Positive [ ]
- [ ] Strong — Negative [x]

Handout

10.2 Linear correlation

Scatterplots

A scatterplot is a graph with a bunch of dots on it. We do not connect the dots. The dot that is far away from the rest is called the OUTLIER.



Linear Correlation

A linear correlation is a relationship between the x and y variables.

Ex: smoking and cancer

Ex: studying for a test and passing

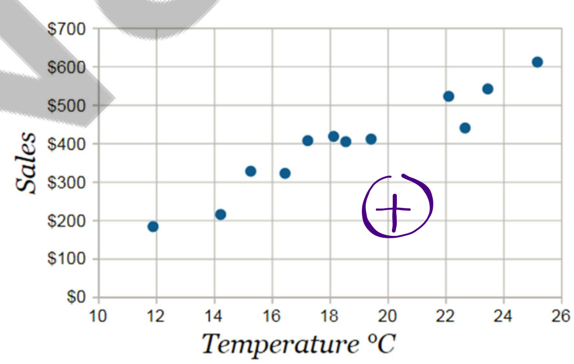
There is a correlation when the points on the graph tend to make a line.

Example: Ice CreamSales

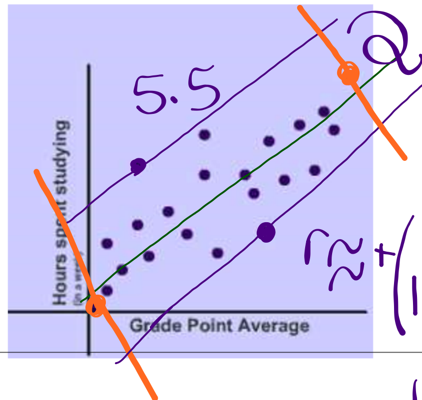
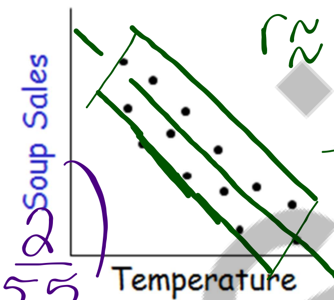
The local ice cream shop keeps track of how much ice cream they sell versus the temperature on that day, here are their figures for the last 12 days:

Ice Cream Sales vs Temperature	
Temperature °C	**Ice Cream Sales**
14.2°	$215
16.4°	$325
11.9°	$185
15.2°	$332
18.5°	$406
22.1°	$522
19.4°	$412
25.1°	$614
23.4°	$544
18.1°	$421
22.6°	$445
17.2°	$408

And here is the same data as a Scatter Plot:



We can easily see that warmer weather and higher sales go together. The relationship is good but not perfect. It is **0.9575**.

POSITIVE CORRELATION	NEGATIVE CORRELATION
Both values increase together 	As one value increases, the other decreases 

$r \approx +\left(1-\frac{2}{5.5}\right)$

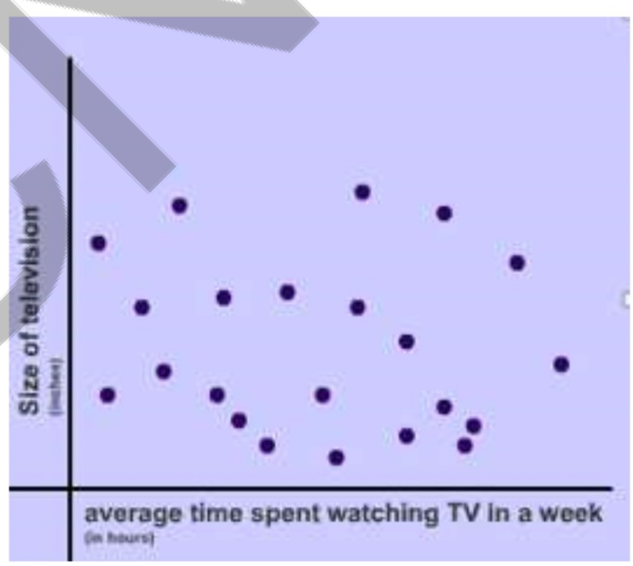
$1-0.36$

0.64

$r \approx -\left(1-\frac{1}{4}\right)$

$-(1-0.25)$

-0.75



Correlation coefficient (r)

approx

$$r \approx \pm\left(1 - \frac{short\ side}{long\ side}\right)$$

It ranges from -1 to 1.

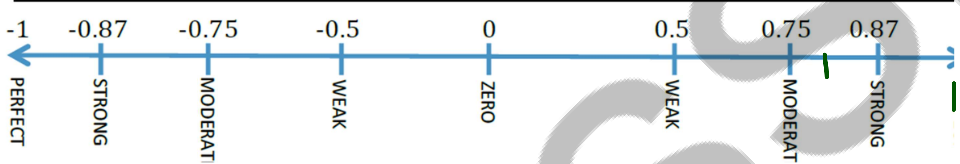
Use $+$ if the line is going up

Use $-$ if the line is going down

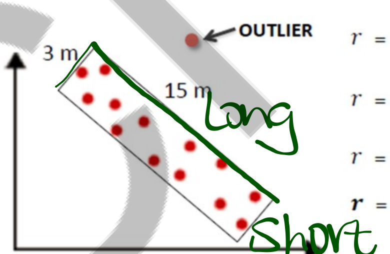
you pick

$\sim$ $\cong$ $\approx$

INTERPRETING CORRELATION COEFFICIENT



CORRELATION (ESTIMATING COEFFICIENT)



$r = \pm\left(1 - \frac{short\ side}{long\ side}\right)$

$r = -\left(1 - \frac{3}{15}\right)$

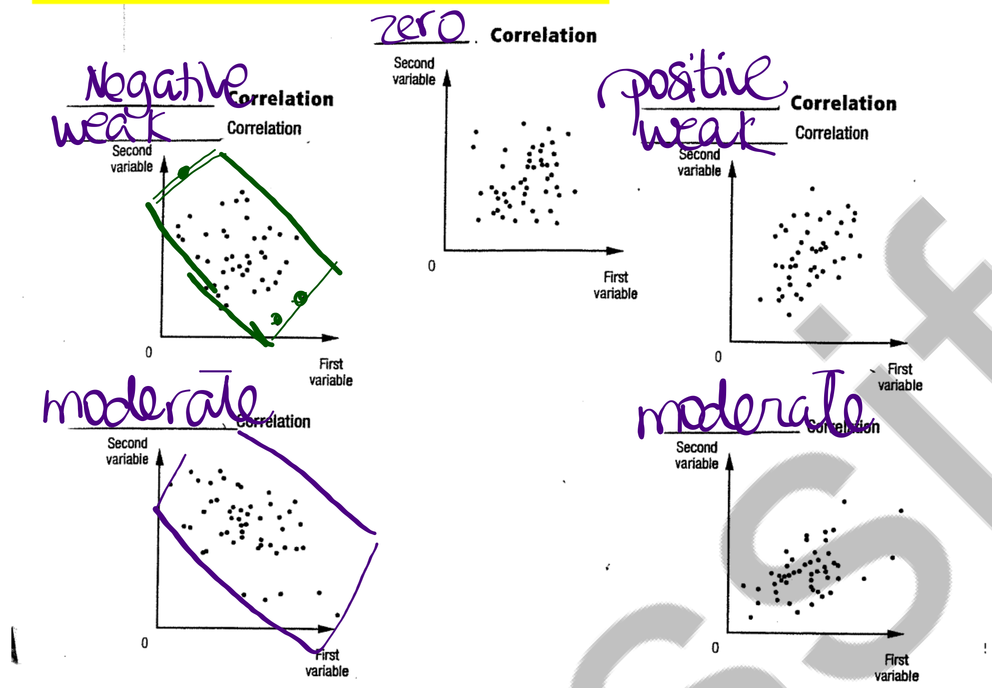
$r = -(1 - 0.2)$

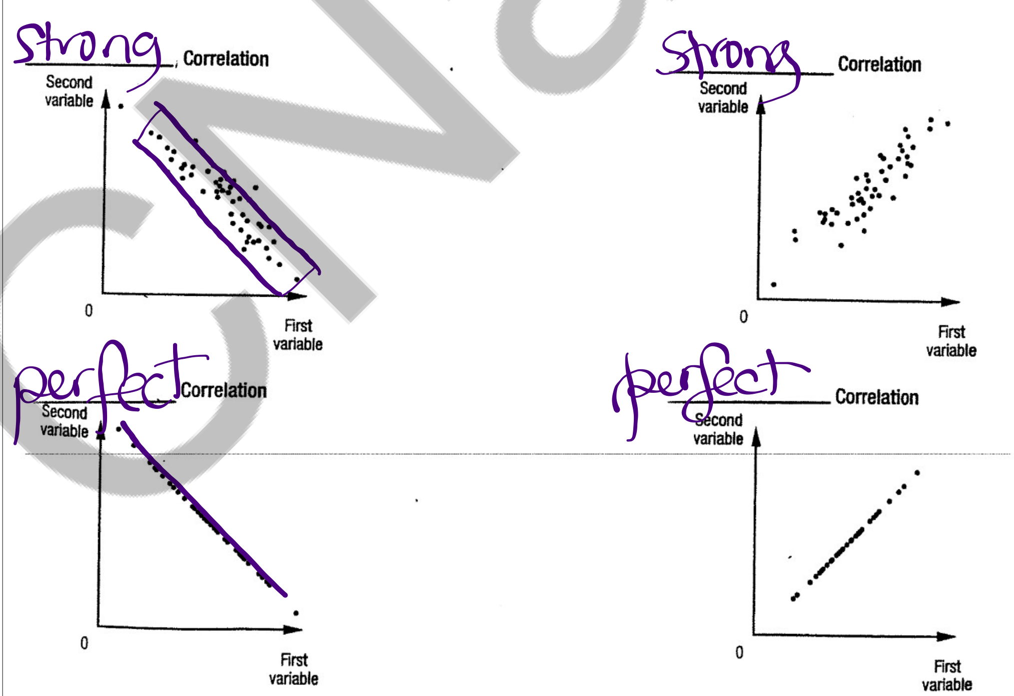
$\boldsymbol{r = -0.8}$

STEPS TO CALCULATE "r"

1. DRAW A RECTANGLE (MEASURE SIDE STO BE SURE)
2. LEAVE OUT OUTLIERS
3. CHECK IF THE TREND IS GOING "UP" OR "DOWN" (POSITIVE OR NEGATIVE SLOPE)
4. PUT MEASUREMENTS INTO THE FORMULA

Linear Correlation Coefficient



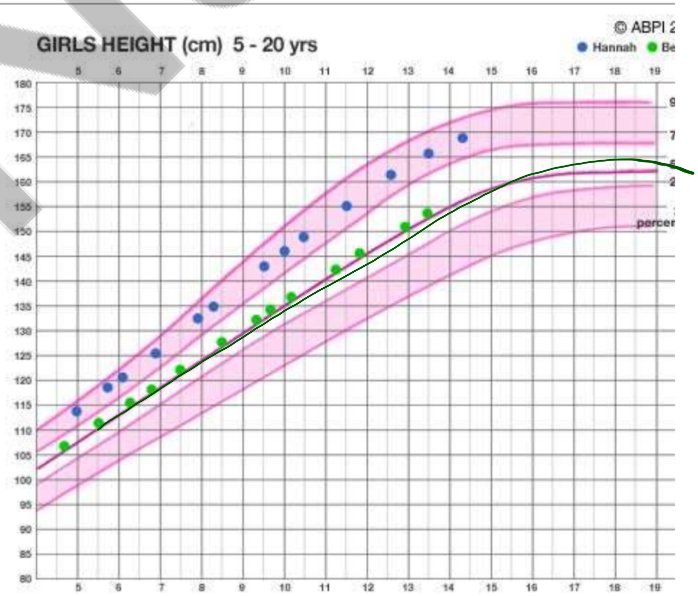


Booklet page 282 #6

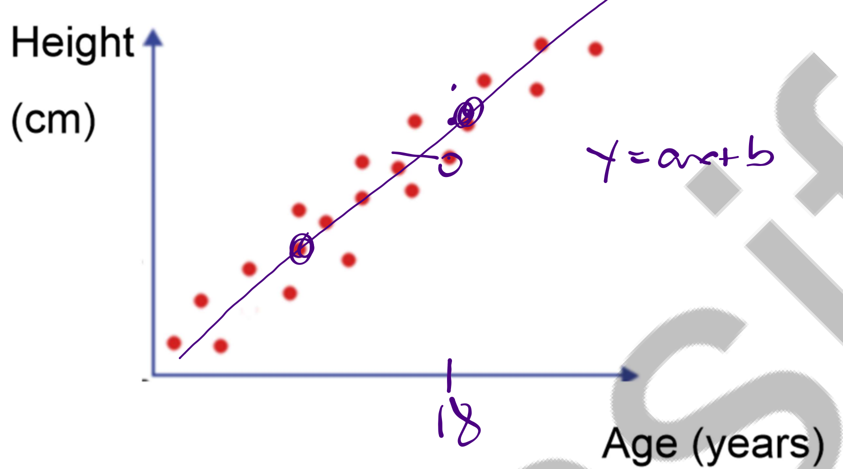
Booklet page 285 #1-6

10.3 Regression Line

When a correlation is strong, we can predict values that occur in the future. **Ex: doctor predicts how tall you will get as an adult.**



How can we figure out your height at 18 years old given the following scatterplot which represents the age and height of 20 kids?

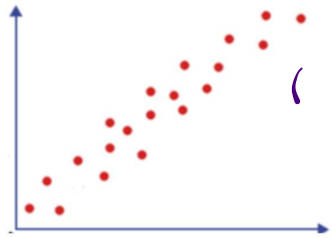


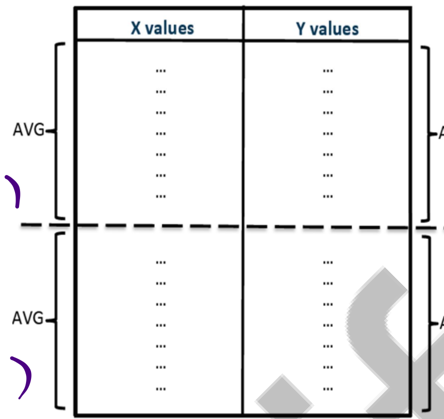
REGRESSION LINE

In a scatterplot relating two variables, the line that best represents all of the points is called a **regression line.**

To find the regression line, we use the Mayer Line Method.

Mayer Line Method





Steps
1) Put the x values in order.
2) Cut the table in half.
3) Find the average of each section.
4) Find the equation $y = ax + b$

Ex: The following points compares x: the age of a tree (years) in relation to y: circumference of a tree (cm):

(10, 39), (15, 55), (23, 68), (18, 50), (7, 26),
(25, 72), (13, 44), (19, 65), (6, 23), (14, 48)

1. Find the regression line.
2. Find the circumference of a 15 year old tree
3. Find the age for a tree with a 100 cm circumference.

1) Mayer

(6,23)
(7,26)
(10,39)
(13,44)
(14,48)

(15,55)
(18,50)
(19,65)
(23,68)
(25,72)

1.) Find the regression li
2.) Find the circumference of a 15 year old tr
3.) Find the age for a tree with a 100 cm circumferen

(10,36)
x_1, y_1

(20,62)
x_2 y_2

$$a = \frac{62-36}{20-10} = 2.6$$

$y = 2.6x + b$

$62 = 2.6(20) + b$

$b = 10$

$y = 2.6x + 10$

2) $y = 2.6x + 10$

$y = 2.6(15) + 10$

$y = 49$ cm

3) $y = 2.6x + 10$

$100 = 2.6x + 10$

$x = 34.5$

year

No Solution Key Available

Write on looseleaf and hand in

Book Page 300 #4-5

Book page 301 #2-5