t-Test & Regression Recap

Statistics Tutorial

Day 13

Prabesh Dhakal
2020 July 09

FEEDBACK: STATS TUTORIAL 2020

● ● ●

Main Link

[t1p.de/py4f](t1p.de/py4f)

Alternative Link

https://forms.gle/bz8EhqCk6ayarCXC7

INFO ON THE EXAM

· · ·

Questions:

will be based on both lecture and tutorial sessions

Do I need a calculator?

not necessary

What about R?:

keep R-Studio ready as there are questions that require you make

some calculations using R commands
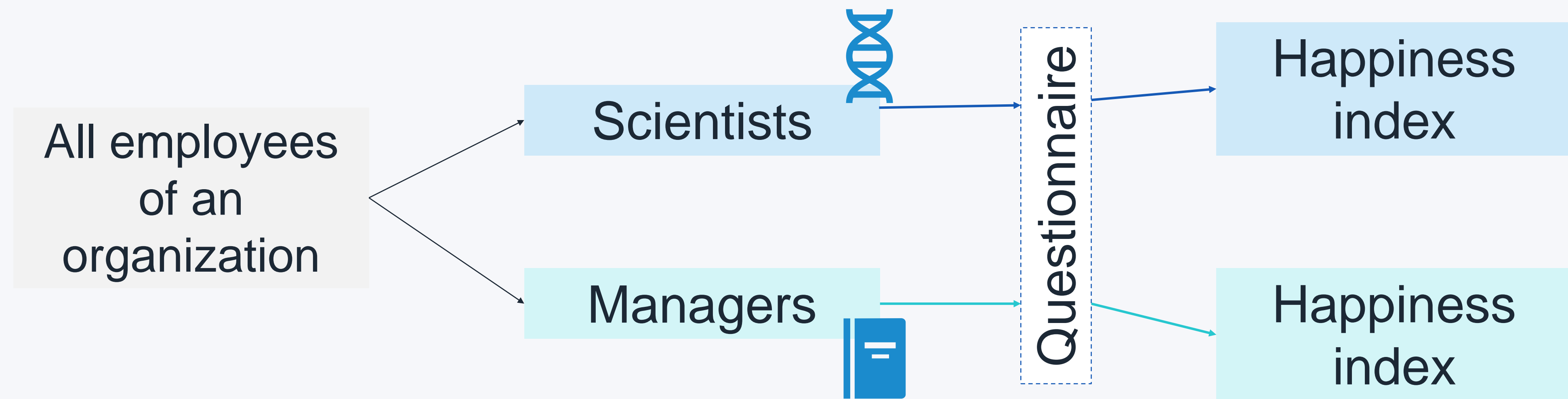
HOW THE QUESTIONS LOOK LIKE

• • •

Question 1:

An important question about a famous statistician that has been highly influential and that everyone that's interested in statistics must know.

X   correct choice 1

X   correct choice 2

☐   wrong choice 3

☐   wrong choice 4

T-TEST

Comparison of sample mean between two groups/levels.

T-TEST

1. **One Sample t-Test**

    - Check if the sample mean differs statistically from a hypothesized population mean

2. **Paired t-Test**

    - Compare means of two samples of same object/category/…

3. **Independent t-Test**

    - Compare means of two independent samples in order to determine whether the associated population means differ significantly

HYPOTHESES IN T-TEST

H_0 : mean value of the two groups are not different

$$H_0: \mu_1 = \mu_2 \quad \text{(equivalently, } \mu_1 - \mu_2 = 0\text{)}$$

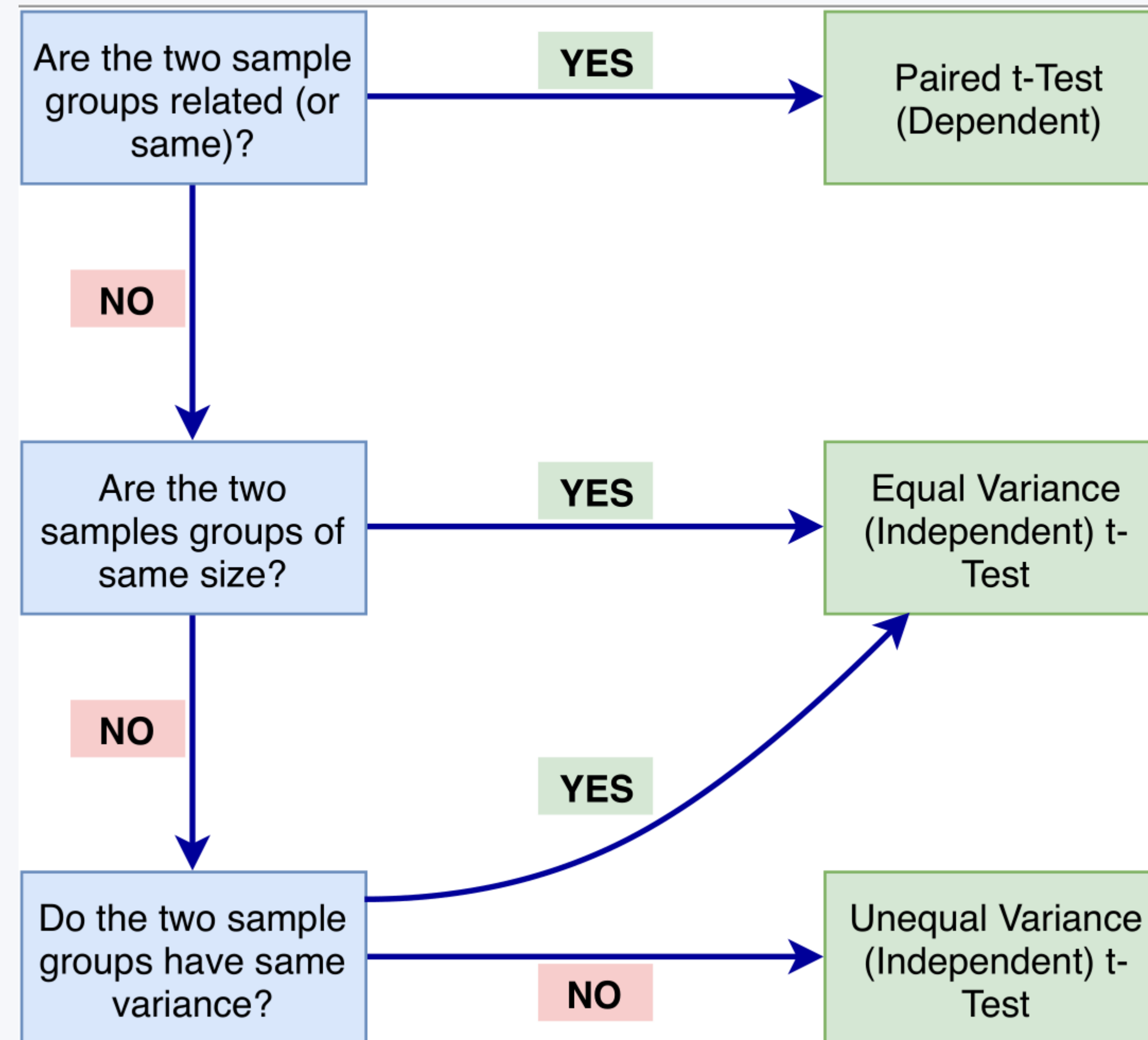H_1 : the 2 groups' mean values are different

$$H_1: \mu_1 \neq \mu_2 \quad \text{(equivalently, } \mu_1 - \mu_2 \neq 0\text{)}$$

$$H_1: \mu_1 > \mu_2 \quad \text{(equivalently, } \mu_1 - \mu_2 > 0\text{)}$$

$$H_1: \mu_1 < \mu_2 \quad \text{(equivalently, } \mu_1 - \mu_2 < 0\text{)}$$

t-Test Decisions

Are the two sample groups related (or same)?	**YES** →	Paired t-Test (Dependent)
	NO ↓	
Are the two samples groups of same size?	**YES** →	Equal Variance (Independent) t-Test
	NO ↓	**YES** ↗
Do the two sample groups have same variance?	**NO** →	Unequal Variance (Independent) t-Test

Assumptions for t-Test

General Assumptions

1. There are no outliers.

2. Test variable is continuous.

3. Samples have been gathered using random sampling.

4. Samples (and populations) are approximately normally distributed.

5. Variances are homogeneous (approximately same in both sample and population).

Paired t-Test

Subjects in each sample (or group) are same:

- subjects in 1st group and 2nd group are same.

Independent t-Test

Subject in samples are independent:

- subjects in 1st group and 2nd groups are different and cannot influence each other.

Questions?

PEARSON'S CORRELATION

$$cor(x, y) = \frac{cov(x, y)}{\sigma_x * \sigma_y}$$

The correlation coefficient

- measures the strength of the linear relationship between two quantitative variables
- value lies between [-1, +1]
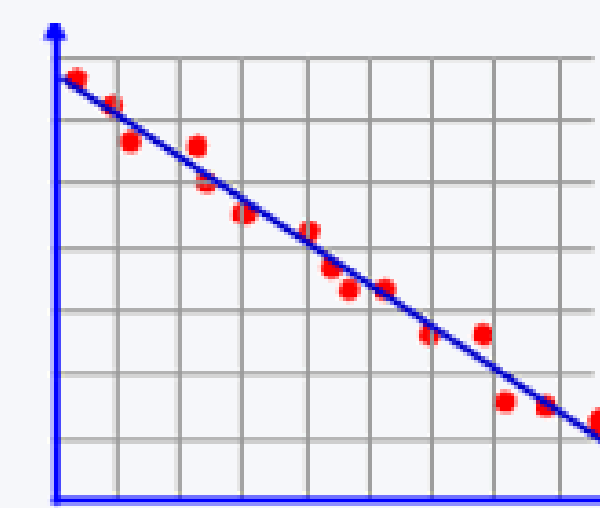 (whereas, covariance can have any value)

Assumptions:

1. Observations are continuous
2. Variables follow a normal distribution
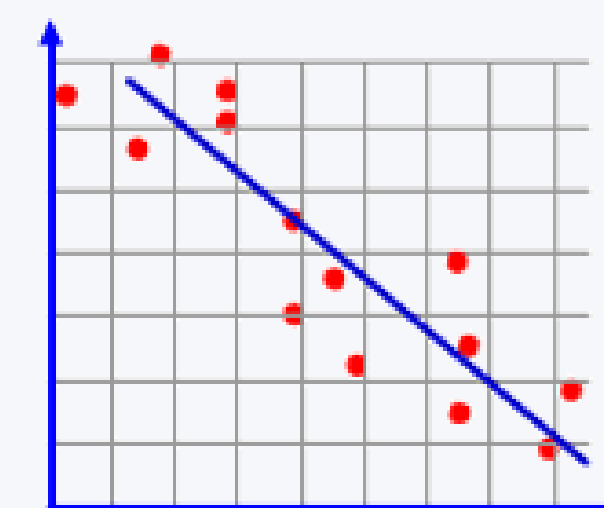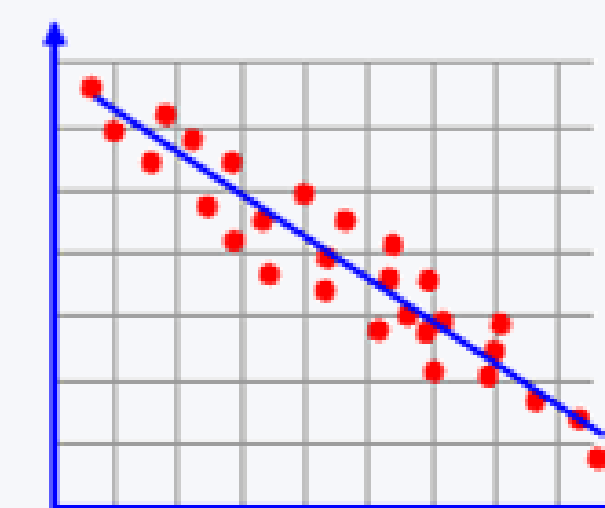3. Variables have a linear relationship
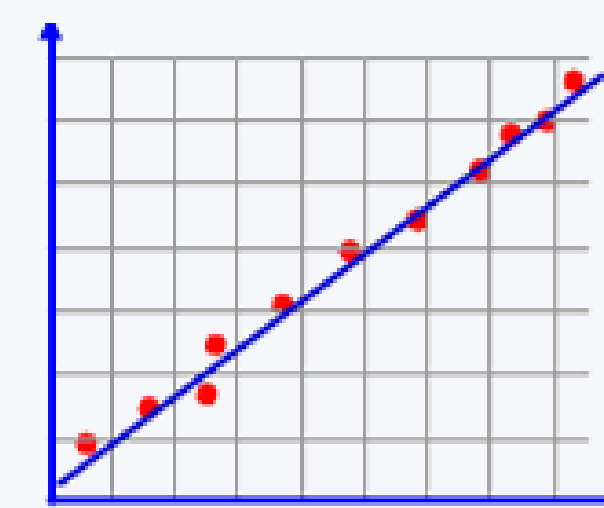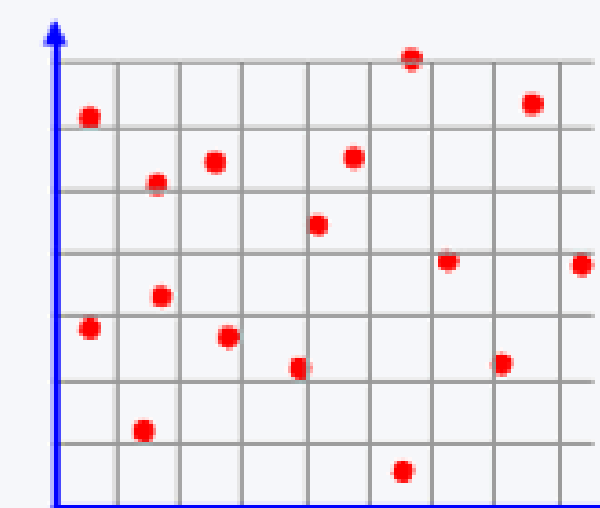

Strong positive correlation
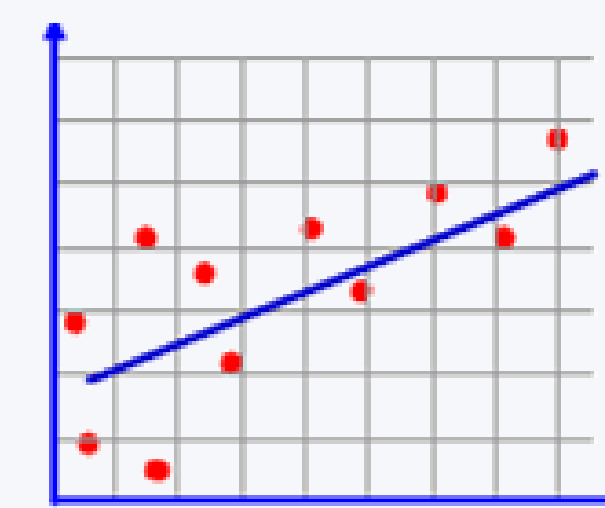
Weak positive correlation
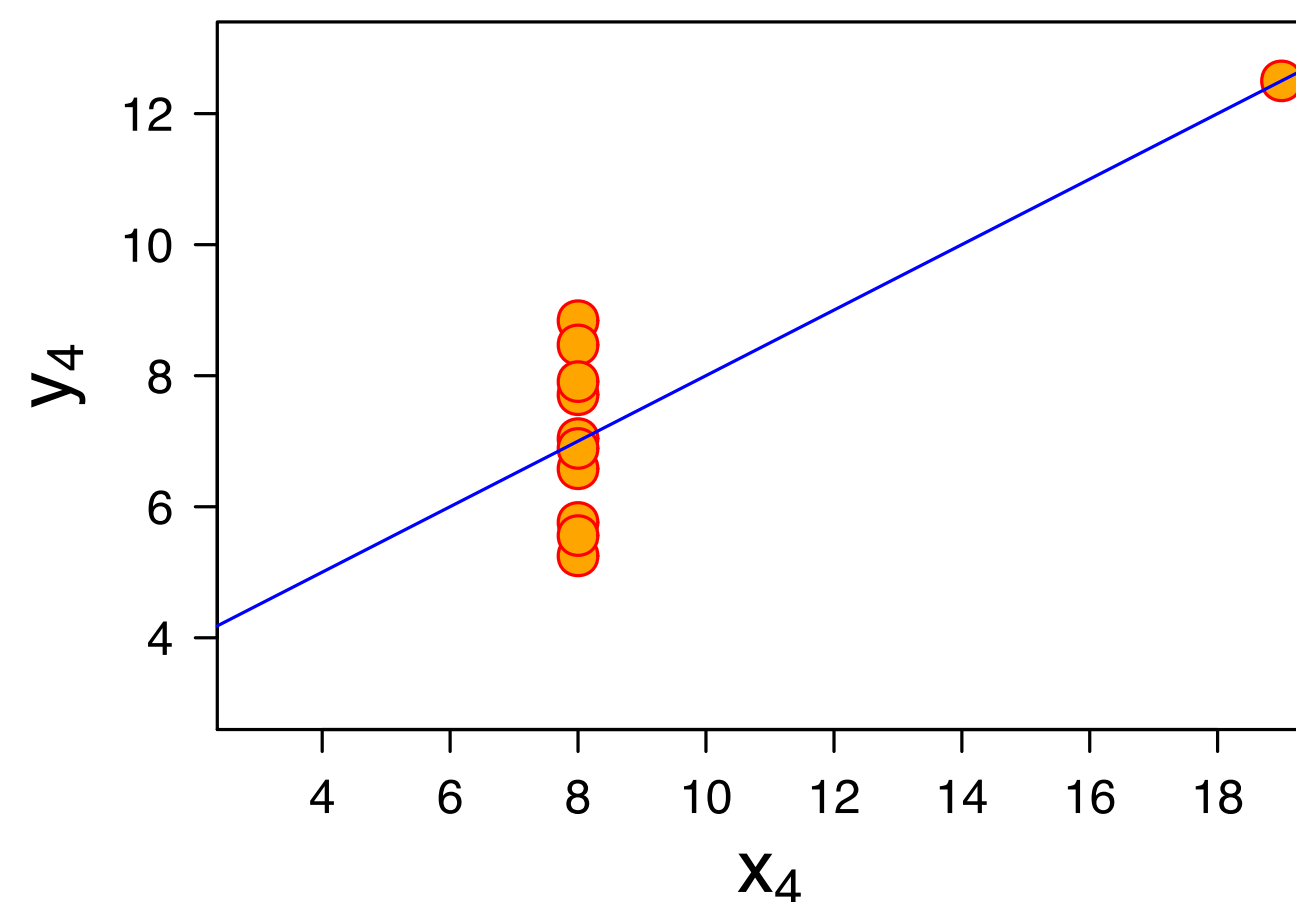
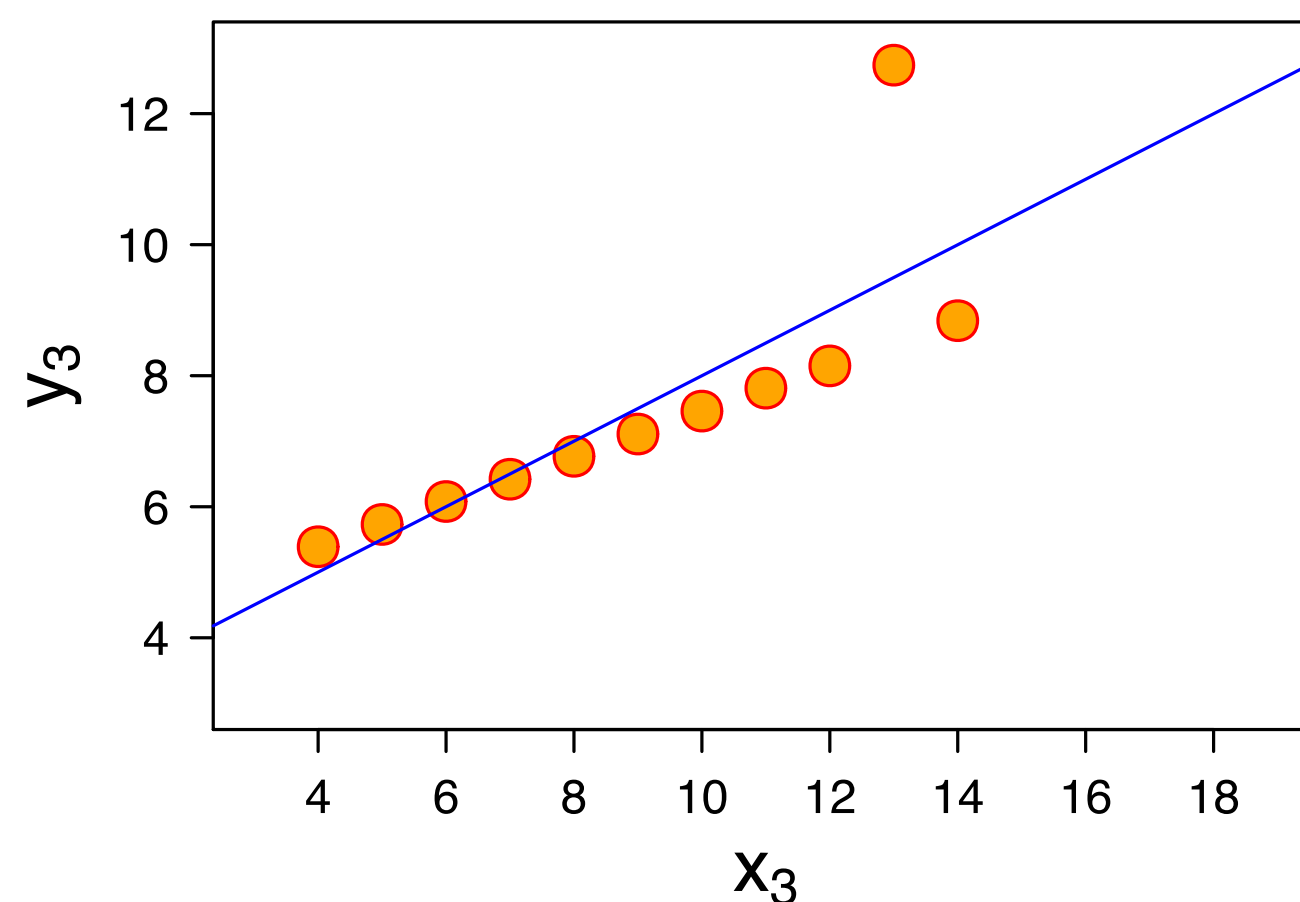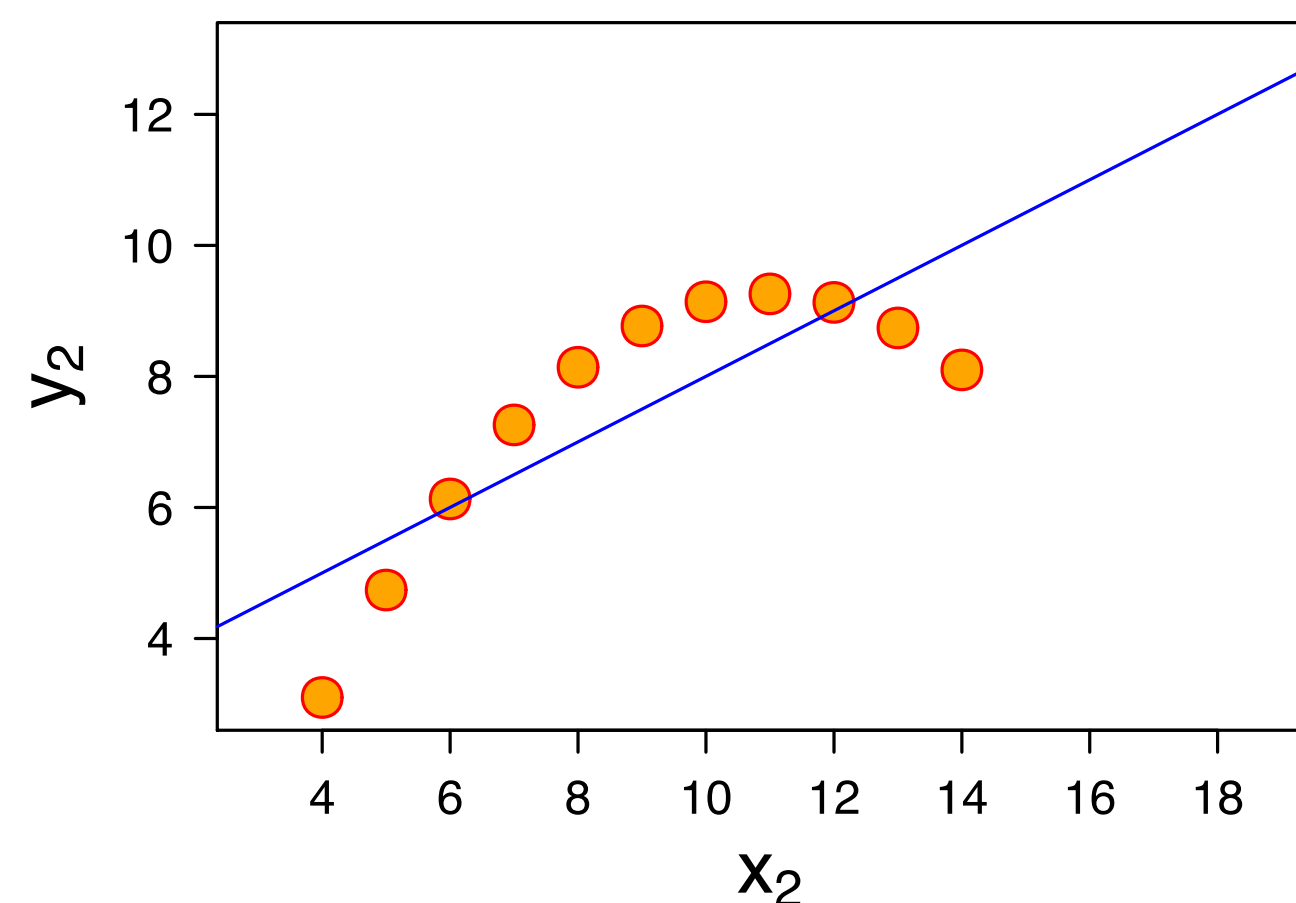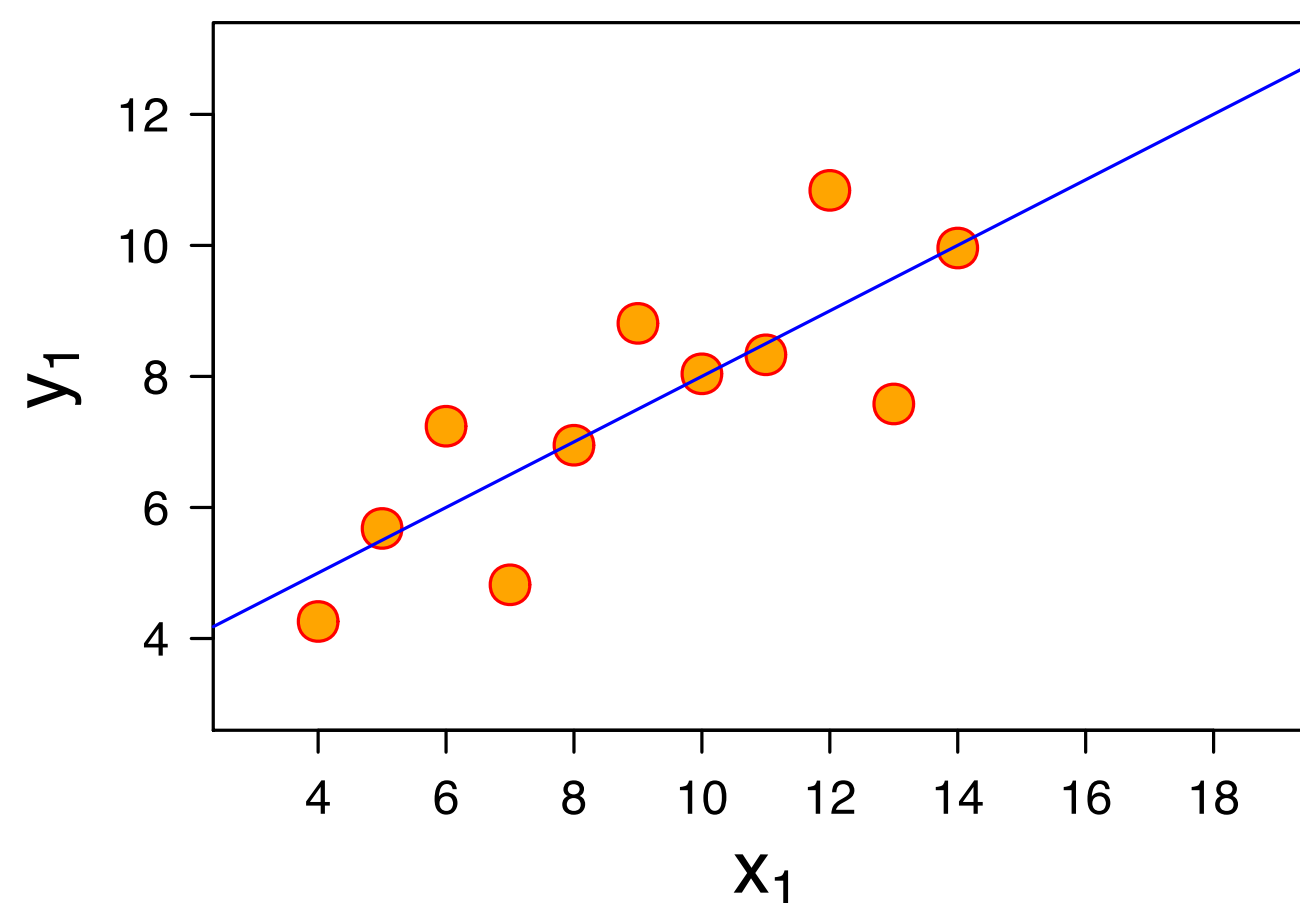Strong negative correlation

Weak negative correlation

Moderate negative correlation

No correlation

ANSCOMBE'S QUARTET

For all 4 datasets:

Mean of x = 9
Var. of x = 11
Mean of y = 7.5
Var. of y = 4.125

$cor(x, y)$ = 0.816

In R:

```
cor(x, y)
cor.test(x, y)
```

REGRESSION

● ● ●

Variables: x and y are continuous, and follow a normal distribution

Objective: we want to predict y based on $x \;\leftrightarrow\; (\,y \sim x\,)$
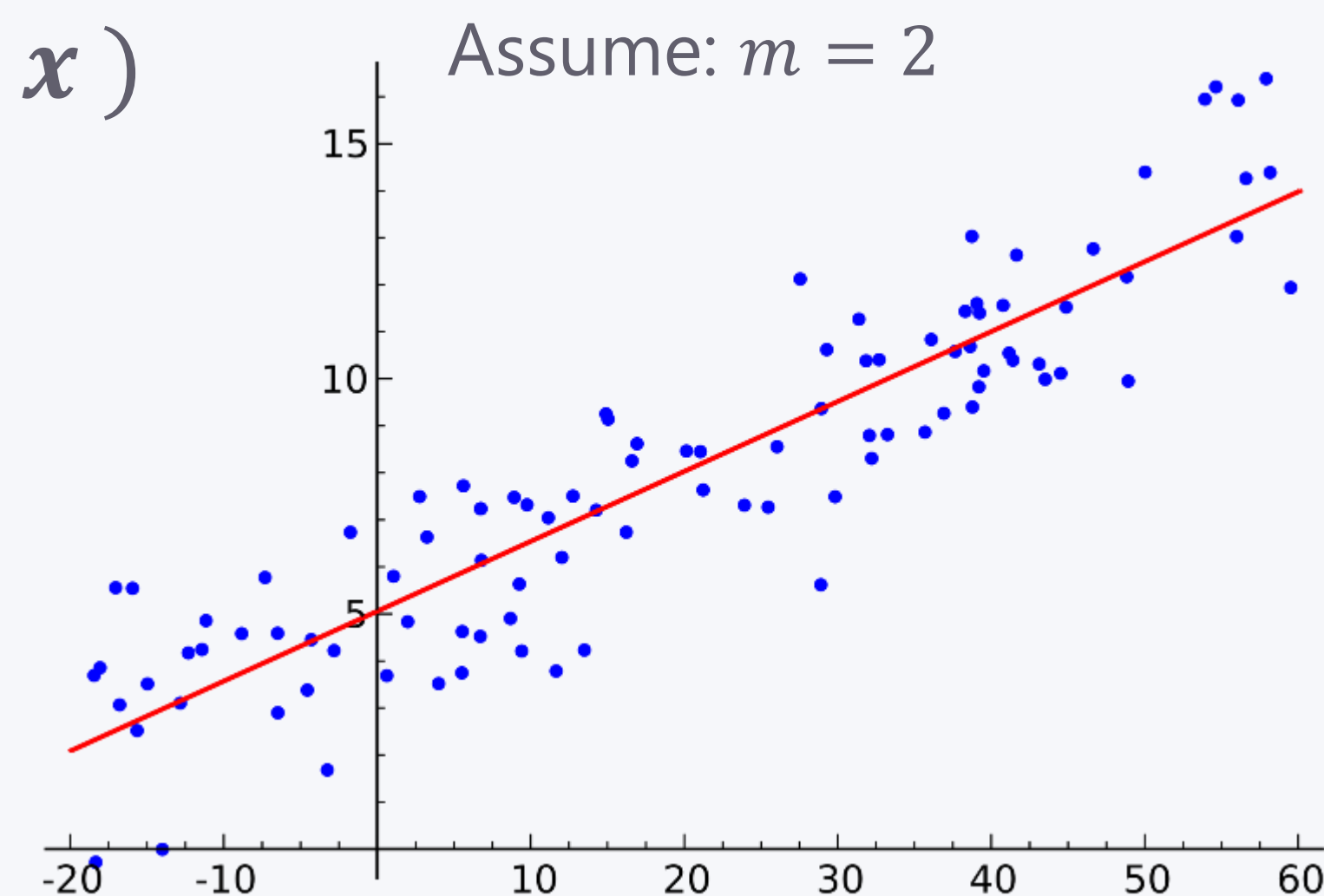
"Simple" Regression Model

$$y = \beta_0 + \beta_1 x$$

Assume: $m = 2$

Dependent variable

Regression coefficients

Independent variable

$$y = mx + b$$

WHY PERFORM REGRESSION?

•••

Interpolate

make predictions for missing values within existing data points
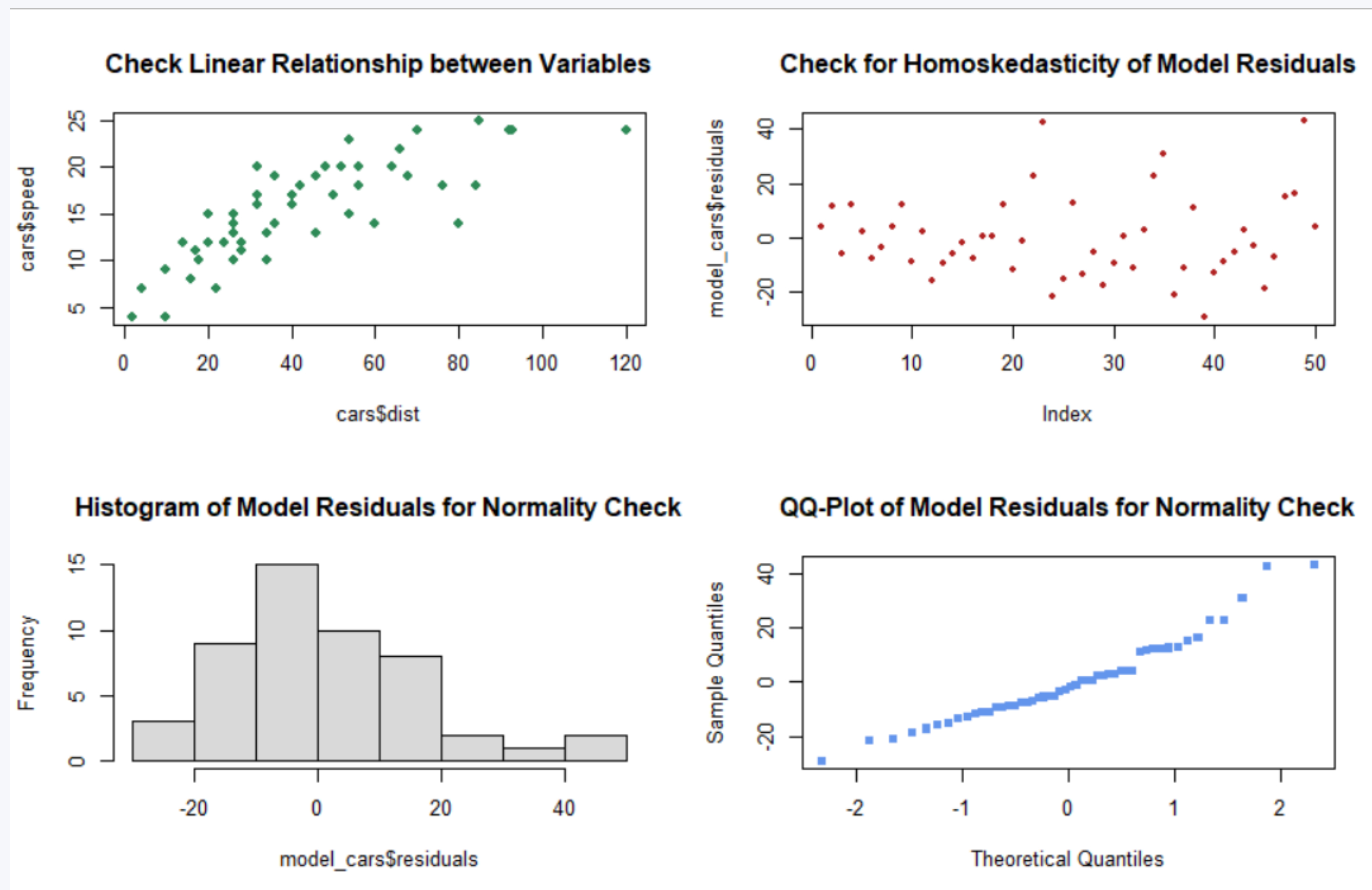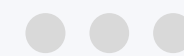
Extrapolate

(could be dangerous)

make predictions for values beyond our dataset

Examine the relationship between 2 variables

check how/how much the independent variable
influences the dependent variable

CHECK THE RESIDUALS

(See R file Day 9)

Questions?

ALL THE BEST FOR YOUR EXAMS!

It was a pleasure to teach you and learn from you.

I really appreciated your enthusiasm throughout the semester.

Thank you for all your questions and suggestions. ☺

If you want to reach me, mail me at:

`prabesh.dhakal@stud.leuphana.de`