Quantitative Methoden 1
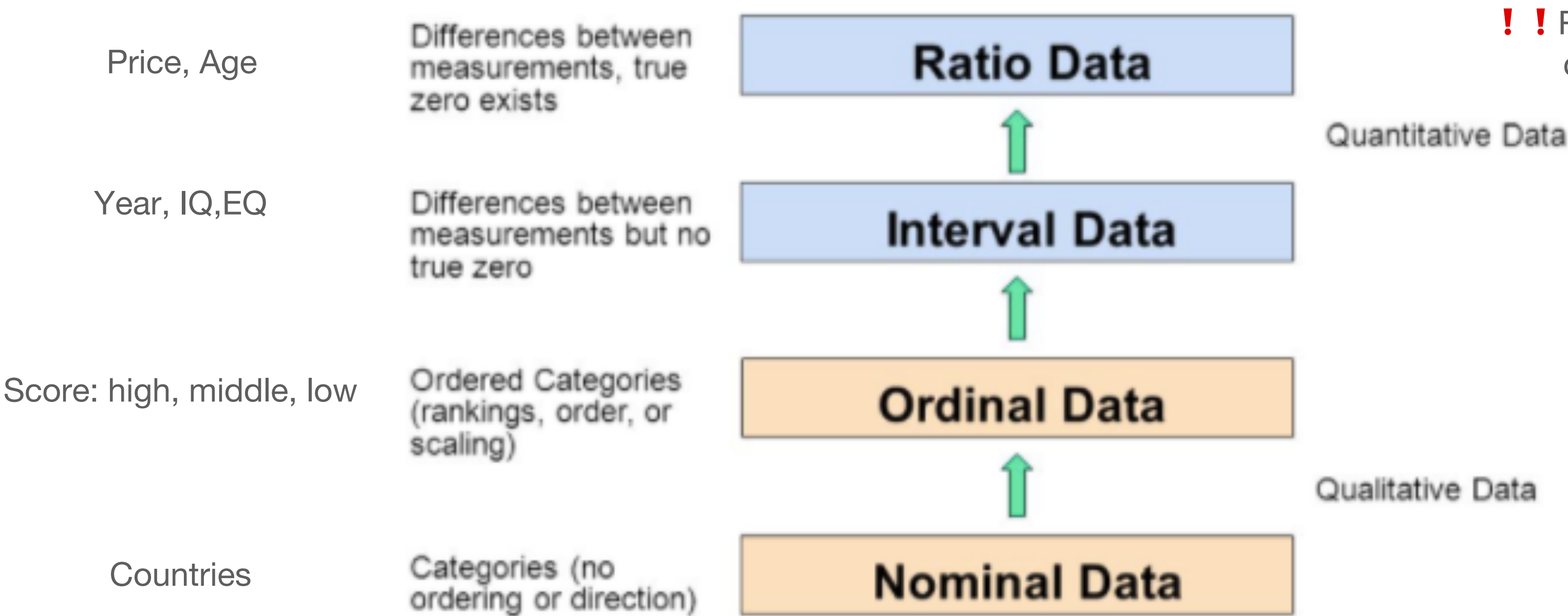
Tutorium 26/03/2021

- Skalenniveau / Type of Variables

- Normalverteilung / Normal Distribution

- R-Basic-System-Package

Skalenniveau / Type of Variables

Price, Age

Differences between measurements, true zero exists

Ratio Data

Year, IQ,EQ

Differences between measurements but no true zero

Interval Data

Quantitative Data

❗️❗️Ratio data and Interval data can be also discrete or continues, these are just two ways of sayings.
More details : book: Ims 1.2.2

Score: high, middle, low

Ordered Categories (rankings, order, or scaling)

Ordinal Data

Countries

Categories (no ordering or direction)

Nominal Data

Qualitative Data

Online Quiz time **?**

- Click the quiz link and enter with the pin code sent in the chat, and start~

- Tips:
- Make sure that you have a stable internet connection.
- If you for any reason leave the live quiz, you can rejoin using the PIN code.
- Please don't leave your browser or app while the quiz is running.

Normal Distribution
Density curve/Frequency distribution

STANDARD NORMAL DISTRIBUTION

▸ $\mu = 0$

▸ $\sigma = 1$

Unimodel, symmetric

Describe the tendency for data to cluster around the central value, in fact the central value is the mean, and always located at the centrer of the curve

The SD determines how spred out the distribution will be.
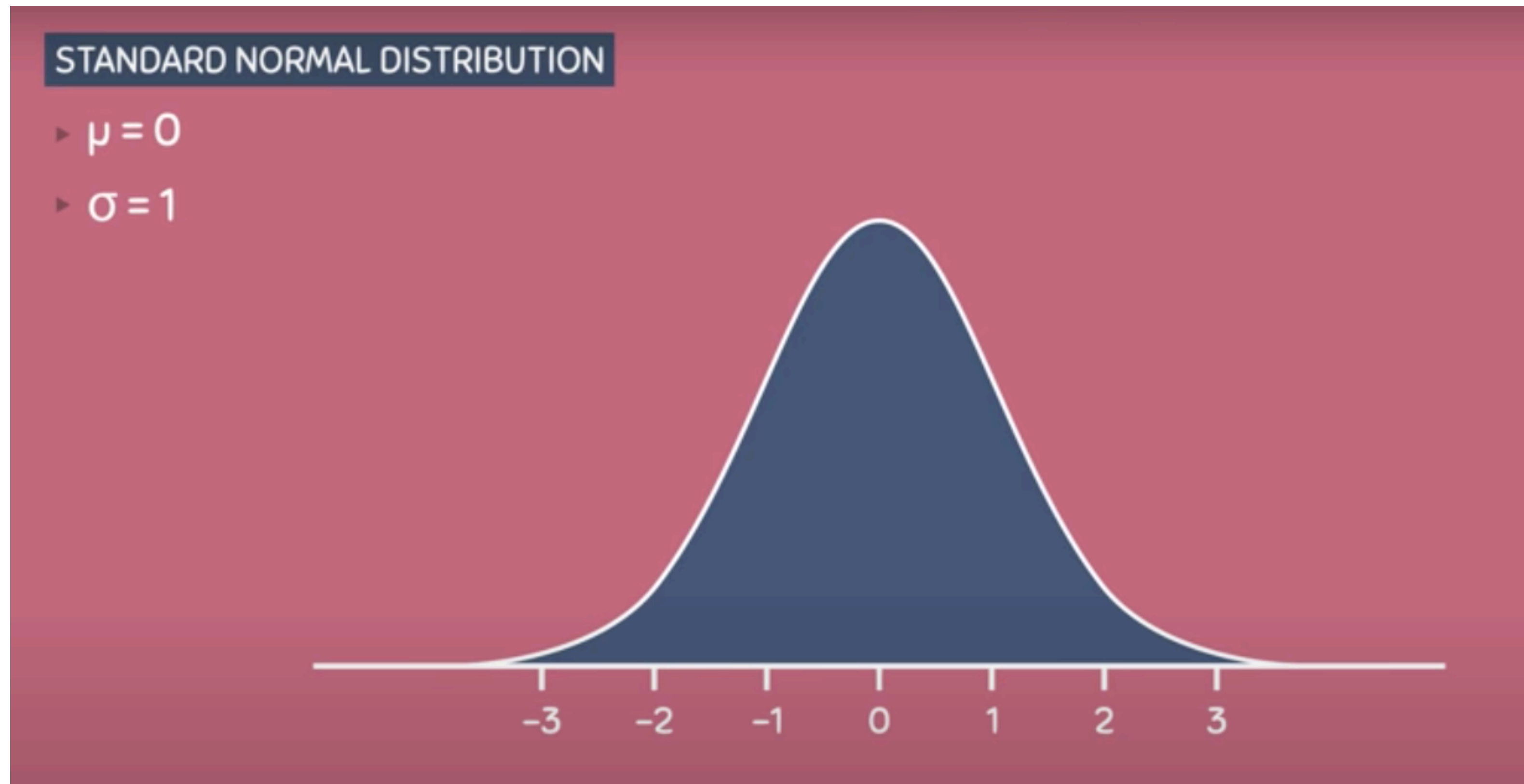
❗❗❗Some useful videos :
https://www.youtube.com/watch?v=mtbJbDwqWLE
https://www.youtube.com/watch?v=CjF_yQ2N638
https://www.youtube.com/watch?v=2tuBREK_mgE

Normal Distribution
68-95-99.7 Rule, total area=total probability=1(always)

About **68**% of values fall within one standard deviation of the mean. About **95**% of the values fall within two standard deviations from the mean. Almost all of the values—about **99.7**%—fall within three standard deviations from the mean

❗❗❗Some useful videos :
https://www.youtube.com/watch?v=mtbJbDwqWLE
https://www.youtube.com/watch?v=CjF_yQ2N638
https://www.youtube.com/watch?v=2tuBREK_mgE
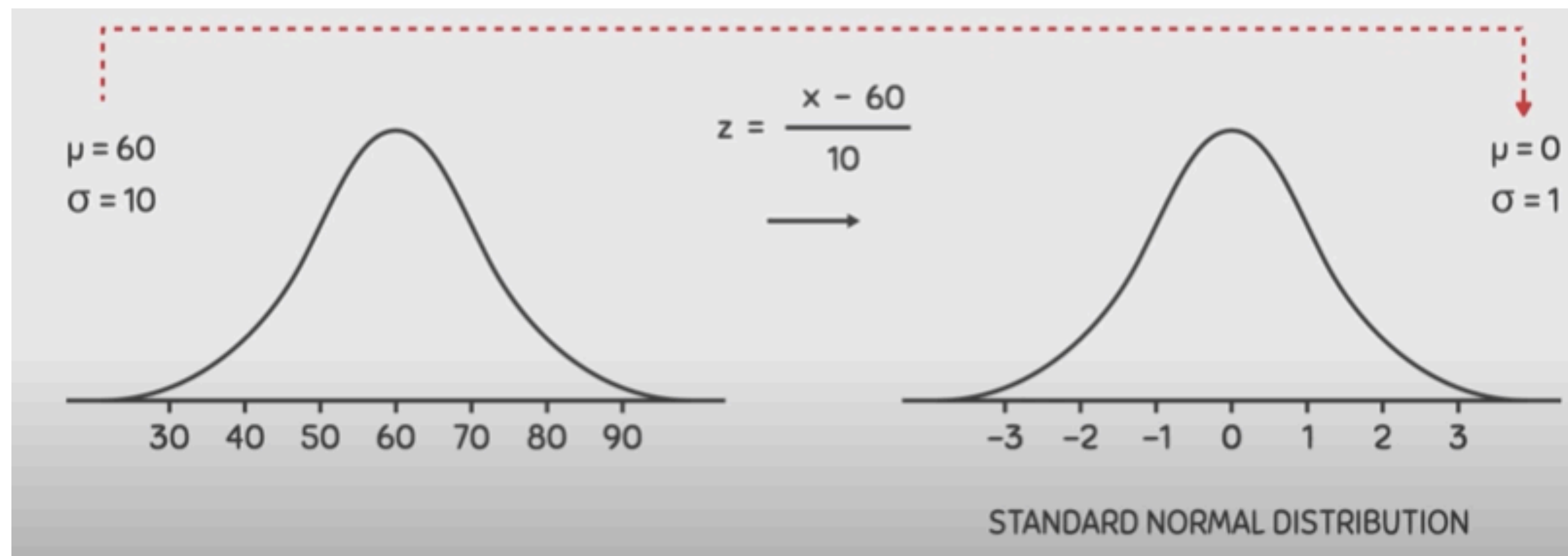https://www.youtube.com/watch?v=mtbJbDwqWLE&t=209s

5

Normal Distribution
Z-Value
A important criteria for tidying data

The Z-score is defined as the number of standard deviations it falls above or below the mean.
If the observation is one standard deviation above the mean, its Z-score is 1. If it is 1.5 standard deviations below the mean, then its Z-score is -1.5.

❗❗❗Some useful videos :
https://www.youtube.com/watch?v=mtbJbDwqWLE
https://www.youtube.com/watch?v=CjF_yQ2N638
https://www.youtube.com/watch?v=2tuBREK_mgE

https://www.openintro.org/book/os/

Online Quiz time ?

- Tips:
- Make sure that you have a stable internet connection.
- If you for any reason leave the live quiz, you can rejoin using the PIN code.
- Please don't leave your browser or app while the quiz is running.

R-Basic-Package

❗ Nicht vergessen:

- Installieren Pakete muss man eine Software nur einmal.
- Starten (laden)muss man die R-Pakete jedes Mal, wenn man sie vorher geschlossen hat und wieder nutzen möchte.

install.packages("ggplot2") DO THIS AT THE FIRST TIME!
library(ggplot2) DO THIS WHENEVER YOU USE IT!

Sauer, S. (2019). Moderne Datenanalyse mit R. Springer
https://www.springer.com/de/book/9783658215866

R-Basic-Syntax <- and ==

<-: allocate the value

==: the values are exactly indentical

temperatur <- 9

Mit dem Operator <- (Zuweisungspfeil) weisen wir dem Behälter (der Variablen) mit dem Namen temperatur den Wert 9 zu;

„X gleich Y" hat in R drei Gesichter:

1. x <- y (oder x = y) weist x den Wert von y zu.
2. x == y prüft, ob x und y identisch sind.
3. fun (x = y) weist innerhalb der Funktion fun() dem Argument x den Wert der Variablen y zu.

Sauer, S. (2019). Moderne Datenanalyse mit R. Springer
https://www.springer.com/de/book/9783658215866

Online Quiz time **?**

- Tips:
- Make sure that you have a stable internet connection.
- If you for any reason leave the live quiz, you can rejoin using the PIN code.
- Please don't leave your browser or app while the quiz is running.

A good website to recommend: post questions on there and wait…

https://stackoverflow.com/

4.4 Exercises

1. Why does this code not work?

<div style="border:1px solid #ccc; padding:8px;">

```
my_variable <- 10
my_variable
#> Error in eval(expr, envir, enclos): object 'my_variable' not found
```

[Copy]

</div>

Look carefully! (This may seem like an exercise in pointlessness, but training your brain to notice even the tiniest difference will pay off when programming.)

2. Tweak each of the following R commands so that they run correctly:

<div style="border:1px solid #ccc; padding:8px;">

```
library(tidyverse)

ggplot(dota = mpg) +
  geom_point(mapping = aes(x = displ, y = hwy))

fliter(mpg, cyl = 8)
filter(diamond, carat > 3)
```

[Copy]

</div>

3. Press Alt + Shift + K. What happens? How can you get to the same place using the menus?

https://r4ds.had.co.nz/workflow-basics.html#exercises-7

```
> library(tidyverse)
>
> library(tidyverse)
>
> ggplot(dota = mpg) + geom_point(mapping = aes(x = displ, y = hwy))
Error in FUN(X[[i]], ...) : object 'displ' not found
> ggplot(data = mpg) + geom_point(mapping = aes(x = displ, y = hwy))
> filter(mpg,cyl=8)
Error: Problem with `filter()` input `..1`.
x Input `..1` is named.
i This usually means that you've used `=` instead of `==`.
i Did you mean `cyl == 8`?
Run `rlang::last_error()` to see where the error occurred.
> filter(mpg,cyl==8)
# A tibble: 70 x 11
   manufacturer model         displ year cyl trans      drv   cty  hwy fl   class
   <chr>        <chr>         <dbl> <int> <int> <chr>     <chr> <int> <int> <chr> <chr>
 1 audi         a6 quattro      4.2  2008   8 auto(s6)   4      16   23 p    midsize
 2 chevrolet    c1500 suburban 2wd  5.3  2008   8 auto(l4)   r      14   20 r    suv
 3 chevrolet    c1500 suburban 2wd  5.3  2008   8 auto(l4)   r      11   15 e    suv
 4 chevrolet    c1500 suburban 2wd  5.3  2008   8 auto(l4)   r      14   20 r    suv
 5 chevrolet    c1500 suburban 2wd  5.7  1999   8 auto(l4)   r      13   17 r    suv
 6 chevrolet    c1500 suburban 2wd  6    2008   8 auto(l4)   r      12   17 r    suv
 7 chevrolet    corvette        5.7  1999   8 manual(m6) r      16   26 p    2seater
 8 chevrolet    corvette        5.7  1999   8 auto(l4)   r      15   23 p    2seater
 9 chevrolet    corvette        6.2  2008   8 manual(m6) r      16   26 p    2seater
10 chevrolet    corvette        6.2  2008   8 auto(s6)   r      15   25 p    2seater
# … with 60 more rows
> filter(diamond, carat > 3)
Error in filter(diamond, carat > 3) : object 'diamond' not found
> filter(diamonds, carat > 3)
# A tibble: 32 x 10
   carat cut     color clarity depth table price    x    y    z
   <dbl> <ord>   <ord> <ord>   <dbl> <dbl> <int> <dbl> <dbl> <dbl>
 1  3.01 Premium I     I1       62.7    58  8040  9.1  8.97 5.67
 2  3.11 Fair    J     I1       65.9    57  9823  9.15 9.02 5.98
 3  3.01 Premium F     I1       62.2    56  9925  9.24 9.13 5.73
 4  3.05 Premium E     I1       60.9    58 10453  9.26 9.25 5.66
 5  3.02 Fair    I     I1       65.2    56 10577  9.11 9.02 5.91
 6  3.01 Fair    H     I1       56.1    62 10761  9.54 9.38 5.31
 7  3.65 Fair    H     I1       67.1    53 11668  9.53 9.48 6.38
 8  3.24 Premium H     I1       62.1    58 12300  9.44 9.4  5.85
 9  3.22 Ideal   I     I1       62.6    55 12545  9.49 9.42 5.92
10  3.5  Ideal   H     I1       62.8    57 12587  9.65 9.59 6.03
# … with 22 more rows
```