Bayesian Inference for Normal Distribution

Peter Büchel

HSLU W

SA: Week 13

Assumption: Normal Distribution

- Until now: Assumption of normal distribution in hypothesis testing
- How to check whether data is normally distributed?
- There are several methods: Here graphical method

Example data set: concrete compressive strength

- Measurement: concrete compressive strength of $n = 20$ different samples:

  24.4, 27.6, 27.8, 27.9, 28.5, 30.1, 30.3, 31.7, 32.2, 32.8, 33.3, 33.5, 34.1, 34.6, 35.8, 35.9, 36.8, 37.1, 39.2, 39.7

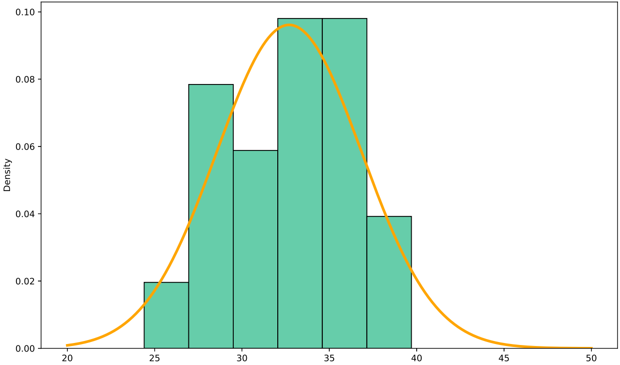
- How well can data be described by a normal distribution?

- Mean and standard deviation:

```
data <- c(24.4, 27.6, 27.8, 27.9, 28.5, 30.1, 30.3, 31.7, 32.2, 32.8,
            33.3, 33.5, 34.1, 34.6, 35.8, 35.9, 36.8, 37.1, 39.2, 39.7)

mean(data)
[1] 32.665

sd(data)
[1] 4.149734
```

- *If* data follow normal distribution: $\mathcal{N}(32.665, 4.15^2)$

- Normalised histogram of data with normal distribution curve of $\mathcal{N}(32.7, 4.15^2)$:



- Should data really be normally distributed: Histogram will match normal distribution curve “well”
- Criterion for “well”: Quantiles match
- What does that mean?
- Here: Slightly simplified approach to that used „officially”
- Principle remains the same
- 20 data points and choose quantiles in steps of 0.05:
- Quantile $q_{0.05}$:

```
quantile(data, probs=0.05)
```

```
 5%
27.44
```

- Value where 5 % of are less than or equal to this value
- Compare quantile with $q^*_{0.05}$ of the normal distribution:

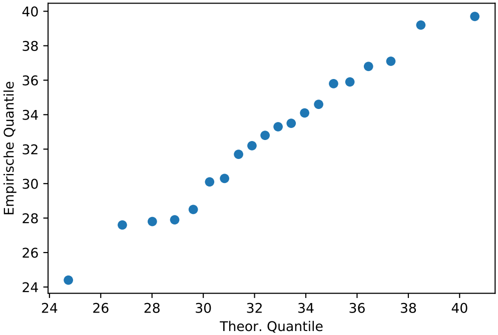
```
qnorm(p=0.05, mean=32.7, sd=4.15)

[1] 25.87386
```

- Values are slightly different
- Now select 10 % quantile:

```
quantile(data, probs=0.1)

10%
27.78

qnorm(p=0.1, mean=32.7, sd=4.15)

[1] 27.38156
```

- Almost identical

- Idea: Determine and plot all quantiles from 0.05 to 1 as above

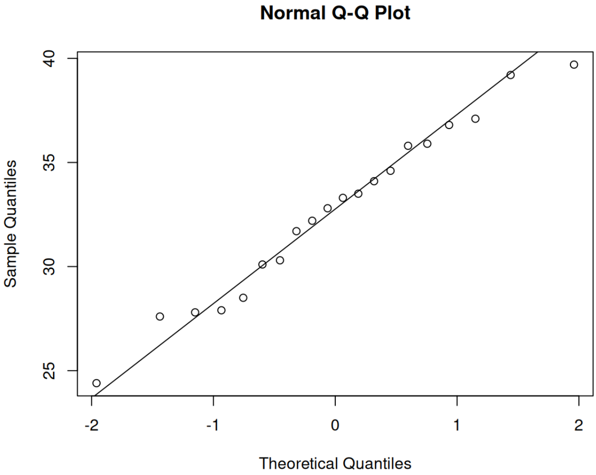


- *If* data follow normal distribution: Empirical quantiles (of data) and quantiles of normal distribution should be approximately equal in size
- Lie on bisector $y = x$

R

- Procedure implemented in R with commands:

```r
qqnorm(data)
qqline(data)
```

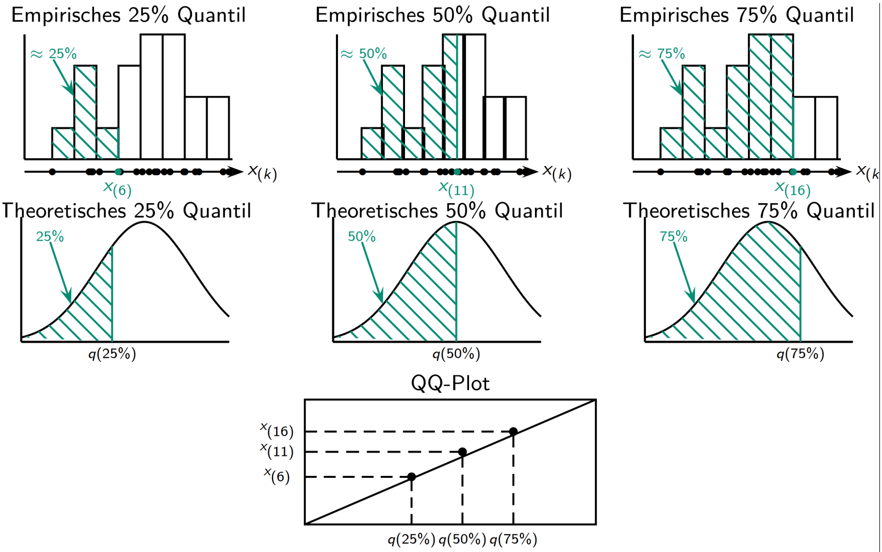


- Unlike previous plot: x- and y-axis scaled to standard normal distribution
- In addition, a line is drawn
- The more points on this line, the more likely it is that data follow normal distribution:

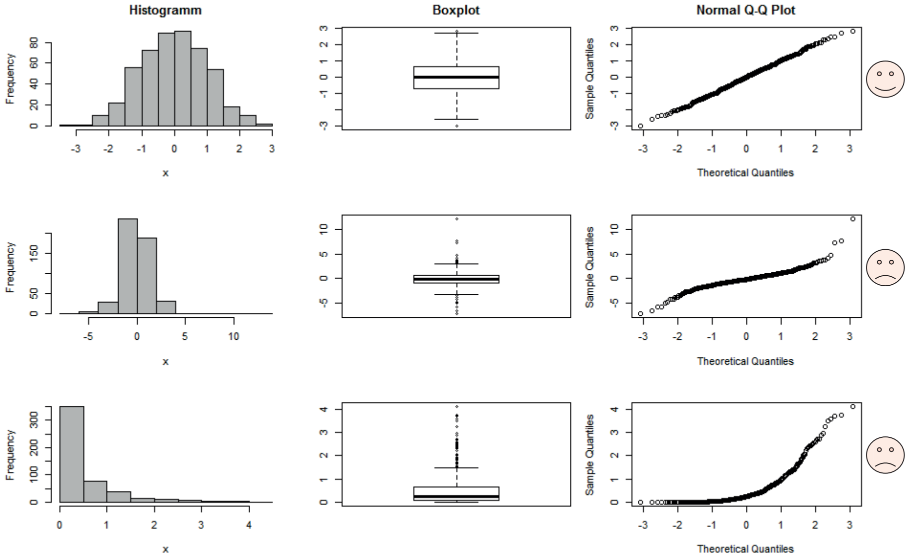
$$\mathcal{N}(32.7, 4.15^2)$$

- Smaller deviations from line are the rule

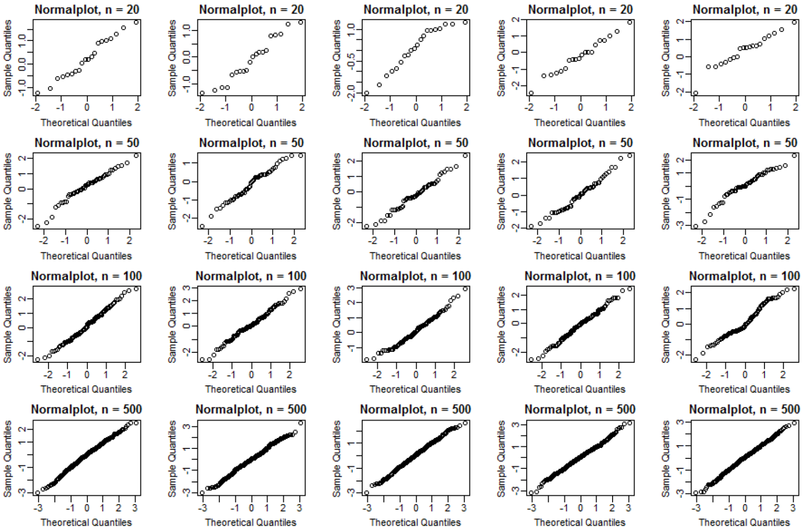
QQ-Plot: Graphical



Examples of normal plots for 3 data sets with $n = 500$



Normal plots of simulated standard normal distributions



MCMC for Normally Distributed Data and Unpaired Samples

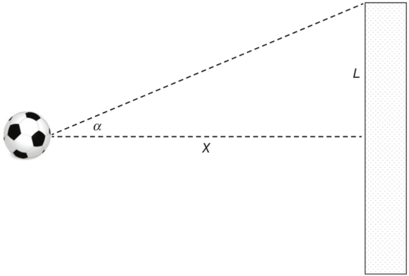
- Example in Jupyter Notebook: `students_heights.ipynb`

MCMC for Normal and Paired Samples

- Example in Jupyter Notebook `husband_wife_age.ipynb`

Reasonable choice of prior distribution

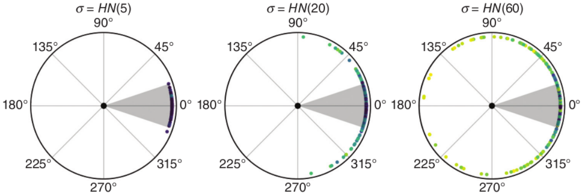
- Random penalty shootout:



- Shooter kicks randomly at goal (without goalkeeper)
- Shooter shoots with angle normally distributed $\mathcal{N}(0, \sigma^2)$

- Describe "accuracy" of the shooter with different half-normal distributions as prior distributions for σ

- Figure:



- In the grey area, the shooter scores goals

- If σ_σ for $\mathcal{HN}(\sigma_\sigma)$ is small, the shooter scores almost every time (left)

- If σ_σ for $\mathcal{HN}(\sigma_\sigma)$ is medium, then the shooter scores occasionally (centre).

- If $\sigma_\sigma$for $\mathcal{HN}(\sigma_\sigma)$ is large, then the shot can literally backfire (right)