Biostatistics: Exercise 03

Beate Sick, Lisa Herzog

29.09.2020

Exercise 01: Binomial distribution

Imagine that there is a new treatment on the market, which offers a relief of symptoms for 80% of patients with Rheumatoid Arthritis (RA). This treatment is given to 100 new RA patients.

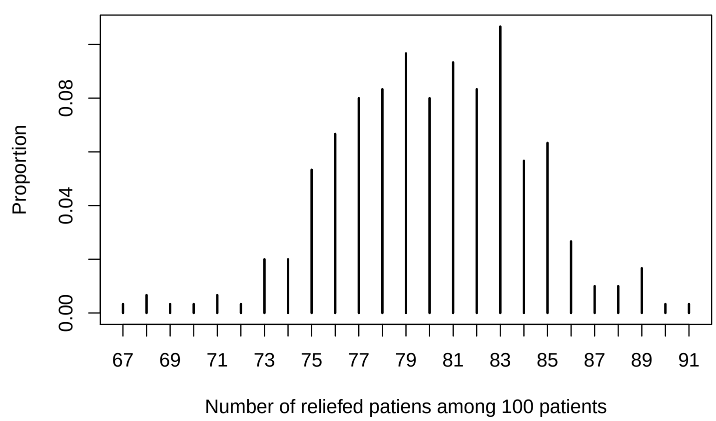
- What is the probability that the symptoms disappear in exactly 70 of the RA patients (R-Hint: `dbinom()`)?

```
dbinom(x = 70, size = 100, prob = 0.8)
```

```
## [1] 0.005189643
```

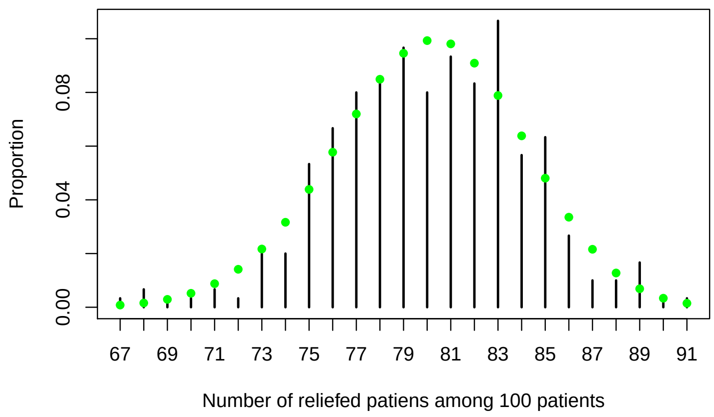
- Simulate 300 data points for the situation described above. That is, simulate the results that you would get, if you repeat the experiment 300 times, each time with 100 patients and with a probability of 0.8 for relief. Visualize the resulting distribution. (R-Hint: `rbinom()`).

```
set.seed(3004)
sim <- rbinom(n = 300, size = 100, prob = 0.8)
plot(table(sim)/length(sim), xlab="Number of reliefed patiens among 100 patients",
     ylab = "Proportion")
```



- Visualize both the simulated data and the underlying theoretical distribution to compare the distributions.

```
plot(table(sim)/length(sim),
     xlab = "Number of reliefed patiens among 100 patients", ylab = "Proportion")
x <- seq(min(sim), max(sim))
points(x, dbinom(x, 100, 0.8), pch = 16, col="green")
```



Exercise 02: Normal distribution

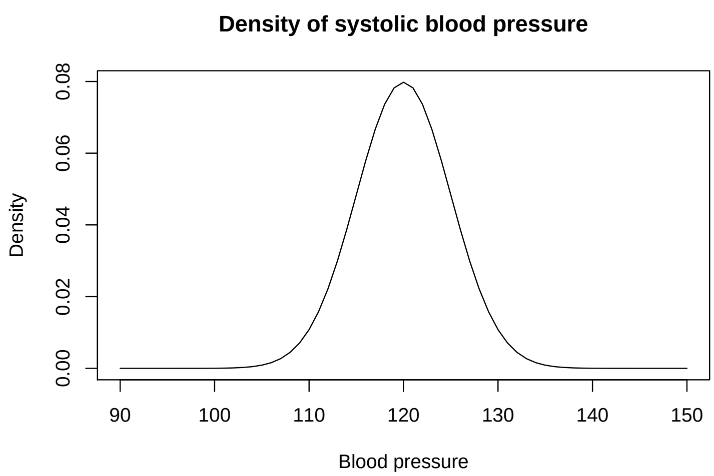
Assume that the systolic blood pressure of a patient cohort is normally distributed with mean 120 and standard deviation 5.

- What is the variance of systolic blood pressure in that patient cohort?

```
# The variance is the squared standard deviation:
5^2
## [1] 25
```

- Plot the corresponding normal distribution for blood pressure values between 90 and 150 (R-Hint: `dnorm()`)

```
bp <- seq(90, 150, 1)
plot(bp, dnorm(x = bp, mean = 120, sd = 5), main = "Density of systolic blood pressure",
     ylab = "Density", xlab = "Blood pressure", type = "l")
```



- What is the probability for a patient to have a blood pressure value between 130 and 140? (R-Hint: `pnorm()`)

```r
# The value can be calculated by integrating the density
# function in the respective interval
pnorm(140, mean = 120, sd = 5) - pnorm(130, mean = 120, sd = 5)
## [1] 0.02271846
```

- Simulate 50 patients of the cohort and calculate the exact 95% CI for the expected value.

```r
set.seed(3004)

n <- 50
sim <- rnorm(n, mean = 120, sd = 5)

mu_hat <- mean(sim)
sigma_hat <- sd(sim)

# upper bound
ci_upper <- mu_hat + qt(0.975, df=n-1) * sigma_hat / sqrt(n)

# lower bound
ci_lower <- mu_hat - qt(0.975, df=n-1) * sigma_hat / sqrt(n)

# CI
ci_lower; ci_upper
## [1] 117.7439
## [1] 120.651

# the 95%CI for the expected value covers the true value of 120
```

Exercise 03: Interpretation of a Confidence Interval

In a prospective study, researchers compare the mean weight loss (in kg) in adults under diet with additional medication (treatment) and diet only (placebo). They report a mean difference (μ) of 1.5kg in weight loss and a corresponding 95% CI of [0.1, 1.8]. Which of the following results are true, which are wrong?

- 95% of the study participants have a weight loss between 0.1kg and 1.8kg.

```r
# Wrong, a CI of 95% doesn't mean that 95% of the
# data lies within that interval.
```

- The confidence interval covers the true difference in weight loss between treatment and placebo with a probability of 95%.

```r
# True
```

- The treatment effect is significantly different from 0 at the 5% level.

```r
# True, 0 is not covered in the CI.
```

- We have no evidence against $H_0 : \mu = 0$ at the 5% level.

```r
# Wrong, there is evidence against the null.
```