Course of Study Bachelor Computer Science	Exercises Statistics WS 2023/24
Sheet VIII - Solutions	

Hypothesis Testing

1. A tire manufacturer claims that its tires will last no less than an average of 50,000 km before they need to be replaced. A consumer group wishes to challenge this claim.

   (a) Clearly define the parameter of interest in this problem.

   (b) State H_0 and H_1 in terms of this parameter.

   (c) In the context of the problem, state what it means to make a type I and type II error.

   (d) Suppose we set the significance level of the test at 10%, what does this number mean?

   Answer:

   (a) Parameter: μ = the average life span (in km) of this manufacturer's tires.

   (b) $H_o : \mu >= 50, H_1 : \mu < 50$

   (c) Type I error: rejecting the manufacturer's claim when in fact it is true.
   Type II error: not rejecting the manufacturer's claim when in fact it is false.

   (d) A ten percent significance level means that we are setting an upper limit of 10% for the probability of making type I error.

2. Discus the following statement:
   "When test results are significant at the 5-percent level, this means that there is at least a 95% chance of being correct if you reject the null hypothesis."

   Answer: The statement is false. The statement "...there is at least a 95% chance of being correct if you reject the null hypothesis" is equivalent to saying "... there is at least a 95% chance that the null hypothesis is true". The significance level does not tell us the precise

probability that the hypothesis is true: That probability depends on more than just the data - it depends for example on the reputation of the person making the statement and what our general background knowledge suggests.
The significance level (p-value) of the data tells us how likely it would be for us to see such data, if we live in a world where the null hypothesis is defintely true. But there is no direct correspondens between the p-value and the probability that the hypothesis is true.
From Bayes' Rule we get:

> P(statement is true | we see this data) =
> P(we see this data | statement is true) x P(statement is true)
> / P(we see this data)

The last two probabilities both depend on our prior beliefs,

3. A sample of lightbulbs is studied, to test the hypothesis that the mean lifetime of the bulbs is 200 hours. The sample data has a significance level of 1%, i.e. the hypothesis is rejected with significance level of 1%.

   Is the following statement true or false: If the mean lifetime in the population is indeed 200 hours, then a second sample (of the same size, analyzed similarly) has only one chance in a hundred of yielding a sample mean as far from 200 as the first sample mean.

   Answer: True. The significance level of the data (with respect to the null hypothesis) is the probability that, in a world where the null hypothesis is true, if we were to carry out the procedure we just carried out, we would see data at least as contradictory to the null hypothesis as the data we are, in fact, seeing.

4. A vaccine that is currently used to immunize people against a certain infection has an 80% success rate. That is, 80% of individuals who receive this vaccine will develop immunity against the infection. A manufacturer of a new vaccine claims that its vaccine has a higher success rate.

   (a) Define the parameter of interest.

   (b) Suppose in a clinical trial, 200 people received the new vaccine. Of these, 172 became immune to the infection. Based on this, can we say that the new vaccine is indeed more effective than the current one? What is the corresponding Null-Hypothesis? Test at a 5% significance level and state your conclusion in the context of the problem.

(c) In making the above conclusion, which type of error are you risking, type I or type II?

(d) What is the probability of a type II error if the true success rate is 82%? What should be the minimal sample size if the probabiility of the type II error should be less than 5%?

Find the answers by using a normal approximation resp. without a normal approximation.

Answer:

(a) Parameter: p = the proportion of individuals who will develop immunity against the infection after being inoculated with the new vaccine.

(b) Data: $\begin{array}{lcl} n & = & 200 \\ 172 \text{ immune} & \Rightarrow & \hat{p} = \frac{172}{200} = 0.86 \\ \alpha & = & 5\% \end{array}$, $\begin{array}{ll} H_0 : & p \leq 0.8 \\ H_1 : & p > 0.8 \end{array}$

Test statistic: $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$ is approximatively $N(0,1)$-distributed for large n.

From the sample: $z = \frac{0.86-0.8}{\sqrt{\frac{0.8(1-0.8)}{200}}} \approx 2.121$

For 2.121 the value of standard normal distribution function is 0.9830, i.e. for a one-sided test the p-value is $p = 1 - 0.9830 = 0.0170$

Decision: Since the p-value 0.0170 is smaller than the given significance level of 0.05, the null hypothesis is rejected. We can conclude that the new vaccine is indeed more effective than the current one.

(c) Since we reject H_0, we are risking making type I error.

(d) If $p_1 = 0.85$ the random variable $X = n \cdot \hat{p} \sim B(n, p_1)$ and $\hat{p} \sim N(p_1, \frac{p_1(1-p_1)}{n})$ approximately. Thus we get

$$\beta_{exact} = P(n\hat{p} \leq 95\%\text{quantile of } B(n, p_0)) \approx 0.451$$

Using a normal approximation for the test we get

$$\beta_{approx} = P(\hat{p} \leq u_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} + p_0) \approx 0.451$$

$$\beta_{approx} \leq 0.05 \quad \Leftrightarrow \quad \Phi\left(\frac{u_{1-\alpha}\cdot\sqrt{\frac{p_0(1-p_0)}{n}}+p_0-p_1}{\sqrt{\frac{p1(1-p_1)}{n}}}\right) \leq 0.05$$

$$\Leftrightarrow \quad \frac{u_{1-\alpha}\cdot\sqrt{\frac{p_0(1-p_0)}{n}}+p_0-p_1}{\sqrt{\frac{p1(1-p_1)}{n}}} \leq u_\alpha.$$

By $u_\alpha = -u_{1-\alpha}$ we get

$$n \geq \left(\frac{u_{1-\alpha}(\sqrt{p_0(1-p_0)}+\sqrt{p1(1-p_1)})}{p_1-p_0}\right)^2 \approx 620.28$$

```
#################################################################
# A vaccine that is currently used to immunize people against a
# certain infection has an 80% success rate. That is, 80% of
# individuals who receive this vaccine will develop immunity
# against the infection. A manufacturer of a new vaccine claims
# that its vaccine has a higher success rate.
# file: infstat_testing_vaccine.R
#################################################################

# a) parameter of interest: success rate p

# b) Suppose in a clinical trial, 200 people received the new
# vaccine. Of these, 172 became immune to the infection. Based on
# this, can we say that the new vaccine is indeed more effective
# than the current one?
# What is the corresponding Null-Hypothesis?
# Null Hypothesis H0: p<=p0   against H1: p>p0
p0 <- 0.8
# Test at a 5% significance level and state your conclusion in the
# context of the problem.
n <- 200
p <- 172/n
alpha <- 0.05
# normal approximation
test_statistic <- (p-p0)/sqrt(p0*(1-p0)/n)
# reject if test_statistic > qnorm(1-alpha)
test_statistic > qnorm(1-alpha) # 2.12132 > 1.644854
# reject H0, thus the new vaccine seems to be better than the old one
p_value_app <- 1-pnorm(test_statistic) # 0.01694743
# exact test
binom.test(172,p=0.8,n,alternative = "greater", conf.level = 1-alpha)
p_value <- 1-pbinom(n*p-1,n,p0)
p_value # = 0.01792922 < 0.05, i.e. rejection of H0

# c) In making the above conclusion, which type of error are you
# risking, type I or type II?
# Since we reject H0, we are risking making type I error.

# d) What is the probability of a type II error if the true success
# rate is 82%?
p1 <- 0.82
# beta.approx = P(test_statistic <= qnorm(1-alpa)) if p=p1 <=>
# beta.approx = P(p <= qnorm(1-alpa)*(p0(1-p0)/n)^0.5+p0)
# with p ~ N(p1,p1(1-p1)/n) approx.
beta.approx <- pnorm(qnorm(1-alpha)*(p0*(1-p0)/n)^0.5+p0,
                     mean = p1, sd = (p1*(1-p1)/n)^0.5)
beta.approx
# alternative beta.approx =
pnorm(qnorm(0.95, mean = 0.8, sd = (0.8*0.2/n)^0.5),
       mean = p1, sd = (p1*(1-p1)/n)^0.5)

# beta.exact = P(np <= 95% quantile of B(n,p0) with n*p ~ B(n,p1)
beta_exact <- pbinom(qbinom(1-alpha,size = n, prob = p0),
                     size = n, prob = p1)
beta_exact

# e) What should be the minimal sample size if the probability
# of the type II error should be less than 5%?
```

```
library(tidyverse)
tibble(
  n = 200:5000,
  b.ex = pbinom(qbinom(1-alpha,size = n, prob = p0),
                size = n, prob = p1),
  b.approx = pnorm(qnorm(1-alpha)*(p0*(1-p0)/n)^0.5+p0,
                   mean = p1, sd = (p1*(1-p1)/n)^0.5)
) %>% filter(b.approx <= 0.05) %>% filter(n == min(n))

# direct determination of n using a normal approximation
(qnorm(0.95)*(sqrt(p0*(1-p0))+sqrt(p1*(1-p1)))/(p1-p0))^2
```

5. A magician uses a coin. You believe that the coin is biased, but you are not sure if it will come up heads or tails more often. You watch the magician flip the coin and record what percentage of the time the coin comes up heads.

   (a) Is this a one-tailed or two-tailed test?

   (b) Assuming that the coin is fair, what is the probability that out of 30 flips, it would come up one side 23 or more times?

   (c) Can you reject the null hypothesis at the 0.05 level? What about at the 0.01 level?

   Answer: a) two tailed, b) 0.005222879, c) rejection at both levels

   (a) Coin, which we believe is biased. We make a two-sided test with $\begin{cases} H_0: & p = \frac{1}{2} \\ H_1: & p \neq \frac{1}{2} \end{cases}$

   (b) Assumption: Coin is not biased with $p = \frac{1}{2}$
   Wanted: The probability that for 30 flips, 23 times or more a special side occurs. Let X = Number of heads for 30 flips. Then we have $X \sim B(n = 30, p = \frac{1}{2})$
   We want $P(X \geq 23$ oder $X \leq 7)$ and have
   $P(X = i) = \binom{30}{i}\left(\frac{1}{2}\right)^i\left(\frac{1}{2}\right)^{30-i} = \binom{30}{i}\left(\frac{1}{2}\right)^{30}$ for $i = 0, 1, 2, \ldots, 30$.
   We become
   $P(X = 0) + \ldots + P(X = 7) + P(X = 23) + \ldots + P(X = 30) \approx 0.005223$
   Reject H_0, if $X \geq o$ or $X \leq u$. Since the p-value $0.005223 \leq \alpha = 0.05\ ;\ 0.01$, we have rejection at both levels.

6. A bag of potato chips of a certain brand has an advertised weight of 250 grams. Actually, the weight (in grams) is a random variable. Suppose that a sample of 81 bags has mean 248 and standard deviation 5. At the 0.05 significance level, conduct the following tests and calculate the p-values.

(a) $H_0 : \mu \geq 250$ versus $H_1 : \mu < 250$

(b) $H_0 : \sigma \geq 7$ versus $H_1 : \sigma < 7$

Hint: Assume that the data is approximately normally distributed.

Answer: Data:
$$\begin{aligned} n &= 81 \\ \bar{x} &= 248 \\ s &= 5 \\ \alpha &= 0.05 \\ X &= \text{Weight} \end{aligned}$$

(a) $\begin{array}{ll} H_0: & \mu \geq 250 \\ H_1: & \mu < 250 \end{array}$, Test statistic: $t = \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$

From the sample: $t_{beob} = \frac{248-250}{\frac{5}{\sqrt{81}}} = -3.6$

With $\alpha = 0.05$ and a one-sided test, we become the following rejection region from the normal distribution table: $(-\infty, -t_{n-1,1-\alpha}) = (-\infty, -1.6441)$

We have $-3.6 < -1.6441$ i.e. H_0 is rejected. The p-value is given by $P_{H_0}(t \leq t_{beob}) \approx 0.0002750739$.

(b) $\begin{array}{ll} H_0: & \sigma \leq 7 \\ H_1: & \sigma > 7 \end{array}$, Test statistic: $t = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$

From the sample: $t_{beob} = \frac{80 \cdot 5^2}{7^2} = 40.81633$

With $\alpha = 0.05$ and a one-sided test, we become the following rejection region from the χ^2-distribution table:
$(0, \chi^2_{n-1,\,\alpha}) = (0,\ \chi^2_{80,\,0.05}) = (0,\ 60.39148)$

We have $40.81633 < 60.39148$ i.e. H_0 is rejected. The p-value is given by $P_{H_0}(t \leq t_{beob}) \approx 8.081861e - 05$.

```
######################################################################
# A bag of potato chips of a certain brand has an advertised weight
# of 250 grams. Actually, the weight (in grams) is a random variable.
# Suppose that a sample of 81 bags has mean 248 and standard
# deviation 5. At the 0.05 significance level, conduct the following
# tests and calculate the p-values.
# a) H_0: mu >= 250 versus H_1: mu < 250
# b) H_0: sigma >= 7 versus H_1: sigma < 7
# file: infstat_testing_potatoe_chips.R
#######################################################################

n <- 81
alpha <- 0.05
mean.sample <- 248
sd.sample <- 5
mean.0 <- 250
sd.0 <- 7

# a) t-test
tstat_a <- (mean.sample-mean.0)*sqrt(n)/sd.sample
tstat_a # -3.6
qt(1-alpha,n-1) # 1.664125
pvalue_a <- pt(tstat_a, df = n-1)
pvalue_a # 0.0002750739

# b)
tstat_b <- (n-1)*sd.sample^2/sd.0^2
```

```
tstat_b # 40.81633
qchisq(alpha,n-1) # 60.39148
pvalue_b <- pchisq(tstat_b, df = n-1)
pvalue_b # 8.081861e-05
```

7. The length of a certain machined part is supposed to be 10 centimeters. In fact, due to imperfections in the manufacturing process, the actual length is a random variable. The standard deviation is due to inherent factors in the process, which remain fairly stable over time. From historical data, the standard deviation is known with a high degree of accuracy to be 0.3. The mean, on the other hand, may be set by adjusting various parameters in the process and hence may change to an unknown value fairly frequently. We are interested in testing

$$H_0 : \mu = 10 \quad \text{versus} \quad H_1 : \mu \neq 10$$

(a) Suppose that a sample of 100 parts has mean 10.1. Perform the test at the 0.1 level of significance.

(b) Compute the p-value for the data.

(c) Compute the probability of a type II error β of the test at $\mu = 10.05$.

(d) Compute the approximate sample size needed for significance level 0.1 and beta = 0.2 when mu = 10.05.

(e) Plot the probability of a type II error depending on the value of μ for different values of the sample size n = 50, 100, 150, 200, 250.

(f) Show that H_0 will be not rejected if the sample mean is 10.1. Determine the smallest n that the p-value for a sample with sample mean = 10.01 is less than 0.001. In general by increasing the sample size every small sample mean will become "highly significant - p value < 0.001".

Hint: Assume that the data is approximately normally distributed. **Answer:**

(a) Data: $\begin{array}{rcl} n & = & 100 \\ \bar{x} & = & 10.1 \\ \sigma & = & 0.3 \\ \alpha & = & 0.10 \end{array}$, $\begin{array}{rl} H_0 : & \mu = 10 \\ H_1 : & \mu \neq 10 \end{array}$

Test statistic: $T = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

From the sample: $t_{beob} = \frac{10.1-10}{\frac{0.3}{\sqrt{100}}} = 3.33$

With $\alpha = 0.10$ and a two-sided test, we become the following

rejection region from the normal distribution table:
$(-\infty, -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty) = (-\infty, -1.6449) \cup (1.6449, \infty)$
We have $3.33 > 1.6449$ i.e. H_0 is rejected.

(b) $p = P(|T| \geq 3.33) = 2(1 - \Phi(3.33)) = 2(1 - 0.99957) = 0.00086$ with the value $\Phi(3.33) = 0.99957$ from the normal distribution table.

(c) $P_{\mu=10.05}\left(\left|\frac{\bar{X}_n - \mu_0}{\sigma_0/\sqrt{n}}\right| \leq u_{1-\alpha/2}\right) =$
$= P_{\mu=10.05}\left(\mu_0 - \frac{u_{1-\alpha/2}\sigma_0}{\sqrt{n}} \leq \bar{X}_n \leq \mu_0 + \frac{u_{1-\alpha/2}\sigma_0}{\sqrt{n}}\right) =$
$= \Phi\left(\frac{10 + \frac{u_{1-\alpha/2} \cdot 0.3}{10} - 10.05}{0.3/10}\right) - \Phi\left(\frac{10 - \frac{u_{1-\alpha/2} \cdot 0.3}{10} - 10.05}{0.3/10}\right) \approx 0.491$ with $\mu_0 = 10, \sigma_0 = 0.3$

```
#######################################################################
# The length of a certain machined part is supposed to be 10
# centimeters. In fact, due to imperfections in the manufacturing
# process, the actual length is a random variable. The standard
# deviation is due to inherent factors in the process, which remain
# fairly stable over time. From historical data, the standard
# deviation is known with a high degree of accuracy to be 0.3. The
# mean, on the other hand, may be set by adjusting various parameters
# in the process and hence may change to an unknown value fairly
# frequently. We are interested in testing
# H_0: mu = 10 versus H_1: mu <> 10
# resp.
# H_0: mu <= 10 versus H_1: mu > 10
#
# Hint: Assume that the data is approximately normally distributed.
# file: infstat_testing_length.R
#######################################################################
library(tidyverse)
# H_0: mu = 10 versus H_1: mu <> 10
sigma <- 0.3
hyp.mean <- 10

# a) Suppose that a sample of 100 parts has mean 10.1. Perform the
# test at the 0.1 level of significance.
sample.size <- 100
sample.mean <- 10.1
alpha <- 0.1
# Test statistic
T <- (sample.mean - hyp.mean)*sqrt(sample.size)/sigma
T
# rejection region: T < lb or T > ub
qnorm(c(alpha/2,1-alpha/2)) # -1.64, 1.64 -> reject H_0

# b) Compute the p-value for the data.
1-pnorm(abs(T)) + pnorm(-abs(T))

# c) Compute the probability of a type II error beta of the
# test at mu = 10.05.
beta <- function(mu1,mu0,n,alpha,sigma) {
  # rejection bounds
  lb <- mu0-sigma*qnorm(1-alpha/2)/sqrt(n)
  ub <- mu0+sigma*qnorm(1-alpha/2)/sqrt(n)
  # prob. of type II error
  beta <- pnorm(ub, mean = mu1, sd = sigma/sqrt(n)) -
    pnorm(lb, mean = mu1, sd = sigma/sqrt(n))
  return(beta)
}
beta(10.05, hyp.mean,sample.size,alpha, sigma)

# d) Compute the approximate sample size needed for significance
# level 0.1 and beta = 0.2 when mu = 10.05.
tibble(
  n = 100:500,
  p.II = beta(10.05, hyp.mean,n,alpha, sigma)
) %>% filter(p.II <= 0.2)

# approximated value
```

```
((qnorm(1-alpha/2)-qnorm(0.2))*sigma/(10.05-hyp.mean))^2

# e) plot the prob. of a type II error depending on the value of mu for different
# values of the sample size n = 50, 100, 150, 200, 250
plot(x=seq(9.8,10.2,by=0.005),
     y=beta(seq(9.8,10.2,by=0.005), hyp.mean,sample.size,alpha, sigma),
     type="l",
     main = "probability of a type II error",
     sub = "blue n=50, black n=100, red n=150,200,250",
     xlab = "mu", ylab = "beta")
lines(x=seq(9.8,10.2,by=0.005),
      y=beta(seq(9.8,10.2,by=0.005), hyp.mean,50,alpha, sigma),
      type="l", col="blue")
lines(x=seq(9.8,10.2,by=0.005),
      y=beta(seq(9.8,10.2,by=0.005), hyp.mean,150,alpha, sigma),
           type="l", col="red")
lines(x=seq(9.8,10.2,by=0.005),
      y=beta(seq(9.8,10.2,by=0.005), hyp.mean,200,alpha, sigma),
      type="l", col="red")
lines(x=seq(9.8,10.2,by=0.005),
      y=beta(seq(9.8,10.2,by=0.005), hyp.mean,250,alpha, sigma),
      type="l", col="red")

# f) Show that the H_0 will be not rejected if the sample mean is 10.01. Determine
# the smallest n for the p-value for a sample mean 10.01 is less than 0.001.
# In general by increasing the sample size every small sample mean will become "highly
# significant - p value < 0.001"
pvalue <- function(n,s.mean,mean.0) {
  # value of the test statistic
  TS <- (s.mean - mean.0)*sqrt(n)/sigma
  return(1-pnorm(abs(TS))+pnorm(-abs(TS)))
}
# p values for the sample means 10.1 and 10.01 are 0.0008581207 and 0.7388827
pvalue(sample.size,10.1,hyp.mean)
pvalue(sample.size,10.01,hyp.mean)
df.pvalue <- tibble(
  n = sample.size:10000,
  p = pvalue(n,10.01,hyp.mean)
)
df.pvalue %>% filter(p <= 0.001) # n >= 9745

# exact
(30*qnorm(0.9995))^2
```

8. A coin is tossed 500 times and results in 302 heads. At the 0.05 level, test to see if the coin is unfair.

Answer: Data: $\begin{array}{rcl} n & = & 500 \\ \hat{p} & = & \frac{302}{500} \\ \alpha & = & 5\% \end{array}$, $\begin{array}{ll} H_0: & p = 0.50 \\ H_1: & p \neq 0.50 \end{array}$

Test statistic: $Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$ is approximatively $N(0,1)$-distributed for large n.

From the sample: $z = \frac{\frac{302}{500}-0.5}{\sqrt{\frac{0.5(1-0.5)}{500}}} = 4.6510$

With $\alpha = 0.05$ and a two-sided test, we become the following rejection region from the normal distribution table: $(-\infty, -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}, \infty) = (-\infty, -1.96) \cup (1.96, \infty)$

We have $4.65 > 1.96$ i.e. H_0 is rejected.

```
###################################################################
# A coin is tossed 500 times and results in 302
# heads. At the 0.05 level, test to see if the
# coin is unfair.
#
# file: infstat_testing_coin_tosses.R
####################################################################
n <- 500
h <- 302
p0 <- 0.5
```

```
p <- h/n
alpha <- 0.05

test_statistic <- (p-p0)/(p0*(1-p0)/n)^0.5
test_statistic # 4.651021

lb_rejection_region <- -qnorm(1-alpha/2,0,1)
ub_rejection_region <- qnorm(1-alpha/2,0,1)
lb_rejection_region; ub_rejection_region
# -1.959964; 1.959964

test_statistic < lb_rejection_region ||
  test_statistic > ub_rejection_region
# TRUE -> reject H0

# pvalue: approximate test
2*pnorm(-abs(test_statistic))

# exact test
binom.test(x = c(h,n-h), alternative = "two.sided", conf.level = 1-alpha)

# pvalue: exact test
pbinom(n-h, size = n, prob = 0.5) + (1-pbinom(h-1, size = n, prob = 0.5))
```

9. Weapons and Aggression

Compare: Case Study: https://onlinestatbook.com/case_studies_rvls/guns/index.html

"Weapons effect" is the finding that the presence of a weapon or even a picture of a weapon can cause people to behave more aggressively. One explanation of the weapons effect is that because guns have been associated with aggression, seeing a gun increases the accessibility of associated aggressive thoughts which in turn facilitate aggressive behavior. The idea that activation of a concept in semantic memory increases activation and therefore accessibility of related concepts is called spreading activation.If this spreading activation explanation of the weapons effect is correct, then the presence of a weapon word (such as "dagger" or "bullet) should increase the accessibility of an aggressive word (such as "destroy" or "wound"). The accessibility of a word can be measured by the time it takes to name a word presented on computer screen.

The hypothesis is that a person can name an aggressive word more quickly if it is preceded by a weapon word than if it word is preceded by a neutral word. Each subject named both aggressive and non-agressive words following both weapon and non-weapon "primes." The subjects were undergraduate students ranging in age between 18 and 24 years. They were told that the purpose of this study was to test reading ability of various words. On each of the 192 trials, a computer presented a priming stimulus word (either a weapon or non-weapon word) for 1.25 seconds, a blank screen for 0.5 seconds, and then a target word (agressive or nonagressive word). The experimenter instructed the subjects to read the first word to themselves and then to read the second word out loud as quickly as they could. The computer recorded reaction times. The means of the times in each of the four conditions

- AN: Aggressive target word, nonweapon prime
- AW: Aggressive target word, weapon prime
- CN: Control target word (nonaggressive) and nonweapon prime
- CW: Control target word (nonaggressive) and weapon prime

were used as the dependent variables. Therefore, each subject provided four scores to the analysis.

(a) Load the csv-file weapons.aggression.csv and make the data tidy if necessary.

(b) Generate side by side boxplots to visualize the results for the different experiments. Evaluate the number of observations, min, max, mean and sd for all experimental settings. Comparing the boxplots and the characteristic numbers what is the effect of a preceded weapon?

(c) The hypothesis is that a person can name an aggressive word more quickly if it is preceded by a weapon word than if the aggressive word is preceded by a neutral word. Compute the difference between the mean naming time of aggressive words when preceded by a neutral word and the mean naming time of aggressive words when preceded by a weapon word separately for each subject. This difference score (prime_agg) will be called the "aggressive-word priming effect." The hypothesis would be supported if: the mean aggressive-word priming effect were positive. Evaluate prime_agg and the min, max, mean, Q1, Q2, Q3 and standard deviation of prime_agg. What do you conclude from these characteristic numbers? Use an approbriate statistical test to check wether the mean aggressive-word priming effect differs significantly different from 0 or not.

(d) Check now the possibility that "weapon words" prime non-aggressive as well as aggressive words. To control for this, compute the difference between

   i. the mean naming time of non-aggressive words when preceded by a non-weapon word
   ii. the mean naming time of non-aggressive words when preceded by a weapon word

   separately for each subject. This difference represents the how much preceding a non-aggressive word by a weapon word decreases the time it takes to name the nonaggressive word and is called the “non-aggressive word priming effect.” Finally, for each subject, subtract this “non-aggressive word priming effect” from the “aggressive-word priming effect” described above. This difference will be called “prime difference” and labeled “prime_diff”. Evaluate prime_diff and the min, max, mean, Q1, Q2, Q3 and standard deviation of prime_diff. What do you conclued? Use an approbriate statistical test to check wether the mean prome difference score differs significantly different from 0 or not.

(e) What is your final conclusion?

```
##################################################################################
# Weapons and Aggression: t-test
# Case Study: https://onlinestatbook.com/case_studies_rvls/guns/index.html
#
#
# file: case_study_weapons_aggression_sol.R
##################################################################################

library(tidyverse)

# Background: The "weapons effect" is the finding that the presence of a weapon
# or even a picture of a weapon can cause people to behave more aggressively.
# One explanation of the weapons effect is that because guns have been associated
# with aggression, seeing a gun increases the accessibility of associated aggressive
# thoughts which in turn facilitate aggressive behavior. The idea that activation
# of a concept in semantic memory increases activation and therefore accessibility
# of related concepts is called spreading activation.If this spreading activation
# explanation of the weapons effect is correct, then the presence of a weapon
# word (such as "dagger" or "bullet") should increase the accessibility of an
# aggressive word (such as "destroy" or "wound"). The accessibility of a word
# can be measured by the time it takes to name a word presented on computer screen.

# Experimental Design: The hypothesis is that a person can name an aggressive word
# more quickly if it is preceded by a weapon word than if it word is preceded by
# a neutral word. Each subject named both aggressive and non-agressive words
# following both weapon and non-weapon "primes."
# The subjects were undergraduate students ranging in age between 18 and 24 years.
# They were told that the purpose of this study was to test reading ability of
# various words. On each of the 192 trials, a computer presented a priming stimulus
# word (either a weapon or non-weapon word) for 1.25 seconds, a blank screen for
# 0.5 seconds, and then a target word (agressive or nonagressive word). The
# experimenter instructed the subjects to read the first word to themselves and
# then to read the second word out loud as quickly as they could. The computer
# recorded reaction times. The means of the times in each of the four conditions
# AN: Aggressive target word, nonweapon prime
# AW: Aggressive target word, weapon prime
# CN: Control target word (nonaggressive) and nonweapon prime
# CW: Control target word (nonaggressive) and weapon prime
# were used as the dependent variables. Therefore, each subject provided four
# scores to the analysis.

# Load the raw data and make the data tidy if necessary.
raw.data <- read_delim(
  "C:/Users/Egbert Falkenberg/Nextcloud/NextCloud Lehre/aktuelle_LV/Statistik/Exercises/R_Solutions/Sheet_
  delim = ";", escape_double = FALSE, trim_ws = TRUE)
raw.data

# make the data tidy
raw.data %>%
  gather(aw:cxen, key = "exp.setting", value = "time") -> tidy.data
tidy.data %>% str()

# Generate side by side boxplots to visualize the results for the different experiments.
boxplot(tidy.data$time ~ tidy.data$exp.setting,
        names = c("AN", "AW", "CN", "CW"),
        ylab = "", xlab = "")
# Comparing the boxplots what is the effect of a preceded weapon?
# Answer: Notice that it took less time to name an aggressive target word when it was
# preceded by a weapon prime (AW) then when it was preceded by a non-weapon
# prime (AN). A comparison of the CW and CN conditions reveals no evidence of a
# weapon prime for the nonaggressive target words.

# Evaluate the number of observations, min, max, mean and sd for all experimental settings.
tidy.data %>%
```

```
  group_by(exp.setting) %>%
  summarise(N = n(), Min = min(time), Max = max(time), Mean = mean(time),
            Q1 = quantile(time, probs = 0.25), Q2 = quantile(time, probs = 0.5),
            Q3 = quantile(time, probs = 0.75), SD = sd(time))

# The hypothesis is that a person can name an aggressive word more quickly if it
# is preceded by a weapon word than if the aggressive word is preceded by a
# neutral word.
# Compute the difference between the mean naming time of aggressive words when
# preceded by a neutral word and the mean naming time of aggressive words when
# preceded by a weapon word separately for each subject. This difference score
# (prime_agg) will be called the "aggressive-word priming effect." The hypothesis
# would be supported if: the mean aggressive-word priming effect were positive.
# Evaluate prime_agg and the min, max, mean, Q1, Q2, Q3 and standard deviation of
# prime_agg
prime_agg <- raw.data$an - raw.data$aw
tibble(diff = prime_agg) %>%
  summarise(N = n(), Min = min(diff), Max = max(diff), Mean = mean(diff),
            Q1 = quantile(diff, probs = 0.25), Q2 = quantile(diff, probs = 0.5),
            Q3 = quantile(diff, probs = 0.75), SD = sd(diff))

# As hypothesized, the mean aggressive-word priming effect is positive.

# Regard now the possibility that "weapon words" prime non-aggressive as well as
# aggressive words. To control for this, compute the difference between
# (a) the mean naming time of non-aggressive words when preceded by a non-weapon word
# and
# (b) the mean naming time of non-aggressive words when preceded by a weapon word
# separately for each subject. This difference represents the how much preceding
# a non-aggressive word by a weapon word decreases the time it takes to name the
# nonaggressive word and is called the "non-aggressive word priming effect."
# Finally, for each subject, subtract this "non-aggressive word priming effect"
# from the "aggressive-word priming effect" described above. This difference
# will be called "prime difference" and labeled "prime_diff".
# Evaluate prime_diff and the min, max, mean, Q1, Q2, Q3 and standard deviation of
# prime_diff

prime_diff <-
  # aggressive-word priming effect
  prime_agg -
  # non-aggressive word priming effect
  (raw.data$cxen - # mean naming time of non-aggressive words when preceded
                   # by a non-weapon word
  raw.data$cxew)   # mean naming time of non-aggressive words when preceded by
                   # a weapon word

tibble(diff = prime_diff) %>%
  summarise(N = n(), Min = min(diff), Max = max(diff), Mean = mean(diff),
            Q1 = quantile(diff, probs = 0.25), Q2 = quantile(diff, probs = 0.5),
            Q3 = quantile(diff, probs = 0.75), SD = sd(diff))

# The hypothesis is that the mean prime difference will be positive.

# As shown before, the mean aggressive-word priming effect was 0.721. Check if
# this value differs significantly different from 0.

t.test(x = prime_agg, alternative = "two.sided", mu = 0)

# Since p-value = 0.0323 we can support the hypothesis that preceding an
# aggressive word with a weapon prime decreases the time it takes to name the
# aggressive word.

# More important is the question that the"prime difference" score mean of 0.8432
# differs significantly from zero.

res.t.test.primed.diff <- t.test(x = prime_diff, alternative = "two.sided", mu = 0)
res.t.test.primed.diff

# Therefore, it can be concluded that weapon words prime aggressive words more
# than they prime non-aggressive words.
```