Assignment template

Oleg Litvinov, Oguzhan Yetkin, group 4

28 February 2022

Exercise 1. Waiting time.

A researcher measured (in minutes) how long patients have to wait in the waiting room of a doctor's office: 15.4, 17.9, 19.0, 0.5, 15.9, 2.7, 6.2, 2.5, 4.7, 6.9, 10.8, 24.3, 5.6, 23.0, 10.7. Denote the mean waiting time by μ.

```
x <- as.numeric(list(15.4, 17.9, 19.0, 0.5, 15.9, 2.7, 6.2, 2.5,
                     4.7, 6.9, 10.8, 24.3, 5.6, 23.0, 10.7))
```

a) Check normality of the data. Assuming normality (irrespective of your conclusion about normality of the data), construct a 97%-CI for μ. Evaluate the sample size needed to provide that the length of the 97%-CI is at most 2. Compute a bootstrap 97%-CI for μ and compare it to the above CI.

Let's check the normality using Shapiro-Wilk test. H_0 is that sample x came from normally distributed population.

```
shapiro.test(x)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  x
## W = 0.93473, p-value = 0.3207
```

From the output, the p-value > 0.05 implying that the distribution of the data are not significantly different from normal distribution, i.e. the null hypothesis can not be rejected. In other words, we can assume the normality.

Estimated mean value:

```
mu = mean(x)
mu
```

```
## [1] 11.07333
```

Next, we are going to construct a 97%-CI for μ. The standard deviation σ is unknown, therefore, we estimate it by s.

```
s = sd(x)
s
```

```
## [1] 7.727545
```

The confidence interval in such a case is based on a t-distribution and the upper t-quantile.

```
alpha <- 1 - 0.97
n <- length(x)
ta <- qt(1-alpha/2, df=n-1)
ta
```

```
## [1] 2.414898
```

t-confidence interval of level 97% for μ:

```
CI_97 <- c(mu - ta*s/sqrt(n), mu + ta*s/sqrt(n))
CI_97
```

```
## [1]  6.255024 15.891642
```

Next, we evaluate the sample size needed to provide that the length of the 97%-CI is at most 2. For this, we have to solve $t_{\alpha/2}\frac{s}{\sqrt{n}} \leq E$ for n.

```
E <- 2
n_min <- (ta*s/E)^2
n_min
```

```
## [1] 87.06039
```

To provide the length of the 97%-CI less than 2, we have to collect the sample of at lest 88 objects.

Let's compute a bootstrap 97%-CI for μ using 1000 samples.

```
B = 1000
Tstar = numeric(B)

for(i in 1:B) {
  Xstar = sample(x, replace=TRUE)
  Tstar[i] = mean(Xstar)
}

TstarLower = quantile(Tstar, alpha/2)
TstarUpper = quantile(Tstar, 1-alpha/2)

bootstrap_CI_97 <- c(2*mu - TstarUpper, 2*mu - TstarLower)
bootstrap_CI_97
```

```
##    98.5%     1.5%
##  6.63860 15.12727
```

The confidence intervals look very close to each other. The one, calculated with a bootstrapping, is stochastic and therefore differs from launch to launch.

b) The doctor claims that the mean waiting time is less than 15 minutes. Under an assumption, verify this claim by a relevant t-test, explain the meaning of the CI in the R-output for this test. Propose and perform a suitable sign tests for this problem. Can we use yet another test based on ranks?

One-sided t-test with H_0: mean waiting time ≥ 15; H_1: mean waiting time < 15:

```
t.test(x, mu=15, alt='l')
```

```
##
##  One Sample t-test
##
## data:  x
## t = -1.968, df = 14, p-value = 0.0346
## alternative hypothesis: true mean is less than 15
## 95 percent confidence interval:
##      -Inf 14.58758
## sample estimates:
## mean of x
##  11.07333
```

H_0 is rejected. The doctor's claim (alternative hypothesis) is accepted. The confidence interval is also one-sided (left-sided). The given value of 15 is outside CI and this also tells about rejecting H_0.

A sign test for median of a single sample may be applied if we state the claim as "the median waiting time is less than 15 minutes":

```
binom.test(sum(x<15), length(x), p = 0.5, alternative = "less", conf.level = 0.95)
```

```
##
##  Exact binomial test
##
## data:  sum(x < 15) and length(x)
## number of successes = 9, number of trials = 15, p-value = 0.8491
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
##  0.0000000 0.8091353
## sample estimates:
## probability of success
##                    0.6
```

The calculated p-value is 0.85. Since this is not less than 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to say that median waiting time is less than 15 minutes.

In the same manner one-sample Wilcoxon signed rank test may be applied.

c) Propose a way to compute the powers of the t-test and sign test from b) at $\mu = 14$ and $\mu = 13$, comment.

The powers may be computed during a simulation as a probability of rejecting H_0 when H_1 is true. For this, we have to generate samples from H_1. For both tests we can generate from normal distribution with the mean of 15, 14, 13.

```
B <- 1000

for(m in 13:15){
  ttest <- numeric(B)
  sign <- numeric(B)
  for(i in 1L:B){
    h1_sample = rnorm(n, mean=m, sd=s)

    ttest[i] <- t.test(h1_sample, mu=mu, alt='l')[[3]]
    sign[i] <- binom.test(sum(h1_sample<mu), length(h1_sample), p = 0.5,
                          alternative = "less", conf.level = 0.95)[[3]]
  }
  print(paste0("H1 mu=", m))
  print(paste0("t-test power ", sum(ttest < 0.05)/B))
  print(paste0("sign test power ", sum(sign < 0.05)/B))
}
```

```
## [1] "H1 mu=13"
## [1] "t-test power 0.006"
## [1] "sign test power 0.086"
## [1] "H1 mu=14"
## [1] "t-test power 0.002"
## [1] "sign test power 0.183"
## [1] "H1 mu=15"
## [1] "t-test power 0"
## [1] "sign test power 0.294"
```

d) Let p be the probability that a patient has to wait longer than 15.5 minutes. Using asymptotic normality, the researcher computed the right end $\hat{p}_r$= 0.53 of the confidence interval $[\hat{p}_l, \hat{p}_r]$ for p. Recover the whole confidence interval and its confidence level.

Let's estimate a proportion of patients to wait longer than 15.5 minutes. *p_hat* is a point estimate for p.

```
p_hat = mean(x > 15.5)
p_hat
```

```
## [1] 0.3333333
```

(1-α)-confidence interval for p is $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

```
p_hat_r <- 0.53
margin_error = p_hat_r - p_hat
p_hat_l <- p_hat - margin_error
p_hat_l
```

```
## [1] 0.1366667
```

Let's calculate Z alpha/2 quantile:

```r
se <- sqrt((p_hat * (1 - p_hat)) / n)
z_alpha_by_2 <- margin_error / se
z_alpha_by_2
```

```
## [1] 1.615782
```

```r
alpha = (1 - pnorm(z_alpha_by_2))*2
1-alpha
```

```
## [1] 0.8938584
```

It was a 0.89-confidence interval for p

e) The researcher also reported that there were 3 men and 2 women among 5 patients who had to wait more than 15.5 minutes, 4 men and 6 women among the remaining 10 patients. The researcher claims that the waiting time is different for men and women. Verify this claim by an appropriate test.

Here we test whether the proportions of men and women in two groups waiting more and less than 15.5 minutes are significantly different. We apply the approximate proportion test:

```r
prop.test(c(2, 6), c(5, 10))
```

```
## Warning in prop.test(c(2, 6), c(5, 10)): Chi-squared approximation may be
## incorrect
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(2, 6) out of c(5, 10)
## X-squared = 0.033482, df = 1, p-value = 0.8548
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.8759135  0.4759135
## sample estimates:
## prop 1 prop 2
##    0.4    0.6
```

There is no significant evidence that the waiting time is different for men and women.

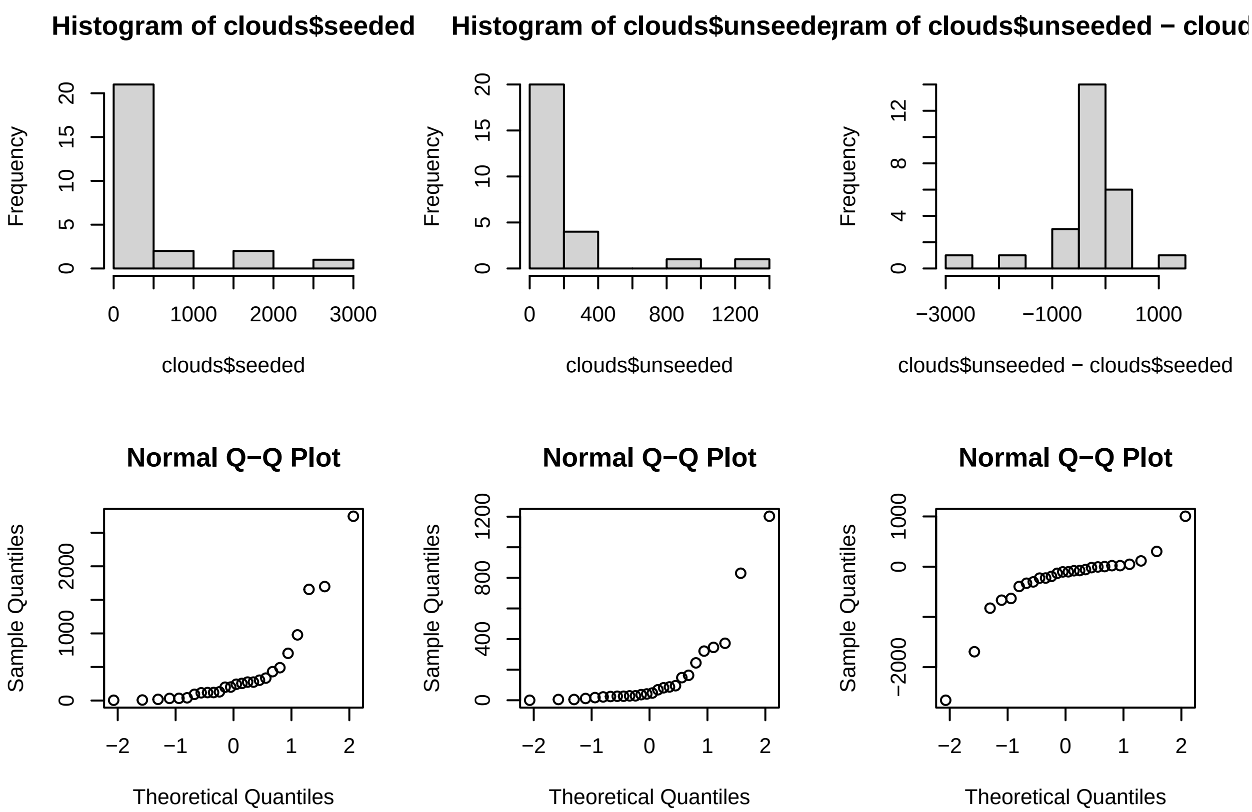
Exercise 2. Seeded clouds.

To improve rain fall in dry areas, an experiment was carried out with 52 clouds. Scientists investigated whether the addition of silver nitrate leads to more rainfall. They chose 26 out of a sample of 52 clouds and seeded it with silver nitrate. The file clouds.txt contains the precipitation values (records the rainfall in feet per acre) of seeded and unseeded clouds.

```r
clouds <- read.table("data/clouds.txt", header=TRUE)

par(mfrow=c(2, 3))
```

```
hist(clouds$seeded)
hist(clouds$unseeded)
hist(clouds$unseeded - clouds$seeded)
qqnorm(clouds$seeded)
qqnorm(clouds$unseeded)
qqnorm(clouds$unseeded - clouds$seeded)
```




```
shapiro.test(clouds$unseeded - clouds$seeded)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  clouds$unseeded - clouds$seeded
## W = 0.75882, p-value = 3.791e-05
```

From the histograms we can easily notice that data is distributed not normally. The distributions look closer to exponential. The difference also is not distributed normally according to Shapiro-Wilk test.

a) Test whether silver nitrate has an effect by performing three tests: the two samples t-test (argue whether the data are paired or not), the Mann-Whitney test and the Kolmogorov-Smirnov test. Indicate whether these tests are actually applicable for our research question. Comment on your findings.

The data might be counted as paired if the data was collected in the following way: two more or less similar clouds are found not far from each other and only one of them is seeded. In the target experiment, the half of clouds was selected without any requirements so we are not assuming that the samples are paired.

```
t.test(clouds$unseeded, clouds$seeded, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  clouds$unseeded and clouds$seeded
## t = -1.9984, df = 33.856, p-value = 0.05375
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -559.585876    4.740491
## sample estimates:
## mean of x mean of y
##  164.5619  441.9846
```

According to two not paired samples t-test, the H_0 states that the means are equal is not rejected. T-test actually may not be performed on our data as the columns are even approximately not distributed normally as well as their difference.

Mann-Whitney test doesn't assume normality and, therefore may be applied. The data is continuous and we can limit the alternative to a shift in location.

```
wilcox.test(clouds$unseeded, clouds$seeded)
```

```
## Warning in wilcox.test.default(clouds$unseeded, clouds$seeded): cannot compute
## exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  clouds$unseeded and clouds$seeded
## W = 203, p-value = 0.01383
## alternative hypothesis: true location shift is not equal to 0
```

```
median(clouds$unseeded); median(clouds$seeded)
```

```
## [1] 44.2
```

```
## [1] 221.6
```

According to Mann-Whitney test, H_0 of equal means is rejected. The underlying distribution of precipitation for seeded clouds is shifted to the right from that of unseeded ones.

Kolmogorov-Smirnov test also doesn't assume normality. H_0: equality of continuous distributions.

```
ks.test(clouds$unseeded, clouds$seeded)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  clouds$unseeded and clouds$seeded
## D = 0.42308, p-value = 0.01905
```

```
## alternative hypothesis: two-sided
```

```
mean(clouds$unseeded); mean(clouds$seeded)
```

```
## [1] 164.5619
```

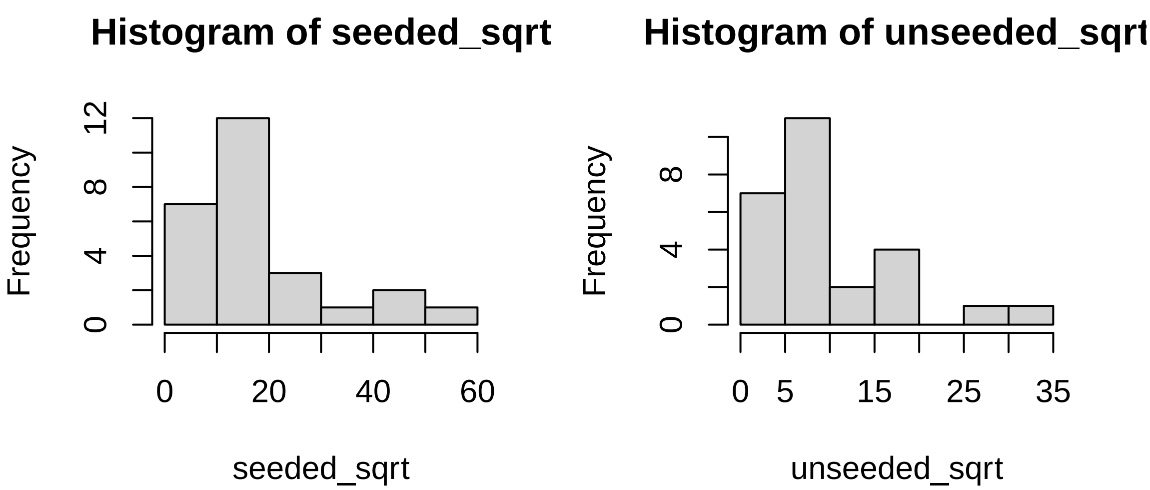
```
## [1] 441.9846
```

Kolmogorov-Smirnov test also rejects H_0. The mean amount of precipitation is larger for seeded clouds than for unseeded.

b) Repeat the procedures from a) first on the square root of the values in *clouds.txt*, then on the square root of the square root of the values in *clouds.txt*. Comment on your findings.

```
unseeded_sqrt <- sqrt(clouds$unseeded)
seeded_sqrt <- sqrt(clouds$seeded)

par(mfrow=c(1, 2))
hist(seeded_sqrt)
hist(unseeded_sqrt)
```



Not the data looks more normal. Let's check it for normality once again.

```
shapiro.test(unseeded_sqrt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  unseeded_sqrt
## W = 0.83744, p-value = 0.0008196
```

```
shapiro.test(seeded_sqrt)
```

```
##
##  Shapiro-Wilk normality test
##
```

```
## data:  seeded_sqrt
## W = 0.87394, p-value = 0.004298
```

The p-value < 0.05 for both columns. This implies that the distributions of the data are significantly different from normal distribution. This means that t-test may not be performed on our data and applied just for interest.

```
t.test(unseeded_sqrt, seeded_sqrt, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  unseeded_sqrt and seeded_sqrt
## t = -2.4246, df = 43.363, p-value = 0.01956
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.071300  -1.202087
## sample estimates:
## mean of x mean of y
##  9.931321 17.068014
```

```
wilcox.test(unseeded_sqrt, seeded_sqrt)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  unseeded_sqrt and seeded_sqrt
## W = 203, p-value = 0.01383
## alternative hypothesis: true location shift is not equal to 0
```

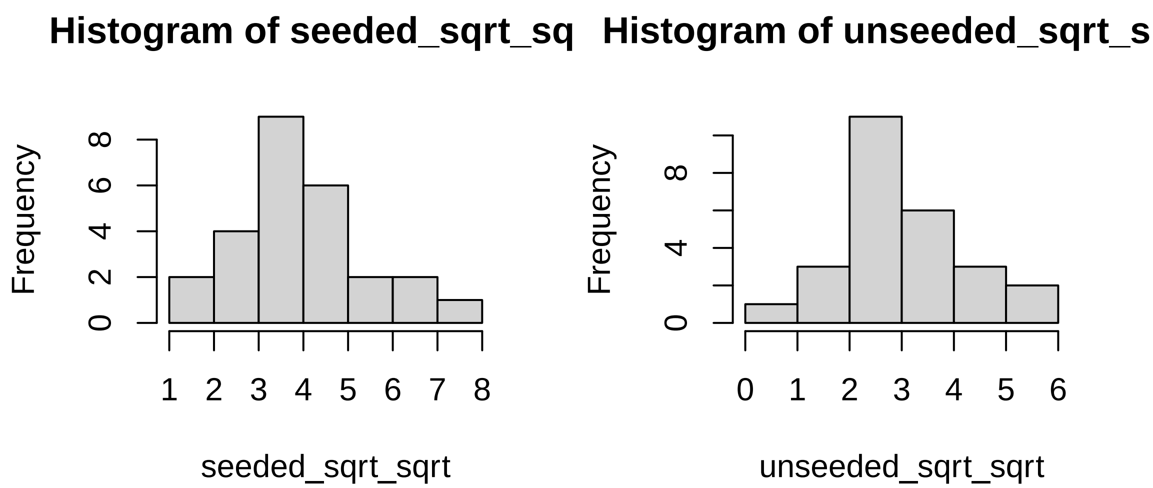
```
ks.test(unseeded_sqrt, seeded_sqrt)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  unseeded_sqrt and seeded_sqrt
## D = 0.42308, p-value = 0.01905
## alternative hypothesis: two-sided
```

In this case t-test rejects the H_0, so means of squared values a significantly different. Interestingly, Wilcoxon and Kolmogorov-Smirnov tests remained completely the same as they are both based on ranks. The ranks remain the same because square root function increases monotonically.

```
unseeded_sqrt_sqrt <- sqrt(unseeded_sqrt)
seeded_sqrt_sqrt <- sqrt(seeded_sqrt)

par(mfrow=c(1, 2))
hist(seeded_sqrt_sqrt)
hist(unseeded_sqrt_sqrt)
```



Not the data looks normal. Let's check it for normality once again.

```
shapiro.test(unseeded_sqrt_sqrt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  unseeded_sqrt_sqrt
## W = 0.95778, p-value = 0.3497
```

```
shapiro.test(seeded_sqrt_sqrt)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  seeded_sqrt_sqrt
## W = 0.96504, p-value = 0.5004
```

From the output, the p-value > 0.05 for both columns implying that the distributions of the data are not significantly different from normal distribution. Only now, for 4th roots of columns we can apply t-test.

```
t.test(unseeded_sqrt_sqrt, seeded_sqrt_sqrt, paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  unseeded_sqrt_sqrt and seeded_sqrt_sqrt
## t = -2.5968, df = 48.826, p-value = 0.0124
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.7236468 -0.2196477
## sample estimates:
## mean of x mean of y
##  2.907340  3.878988
```

```r
wilcox.test(unseeded_sqrt_sqrt, seeded_sqrt_sqrt)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  unseeded_sqrt_sqrt and seeded_sqrt_sqrt
## W = 203, p-value = 0.01383
## alternative hypothesis: true location shift is not equal to 0
```

```r
ks.test(unseeded_sqrt_sqrt, seeded_sqrt_sqrt)
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  unseeded_sqrt_sqrt and seeded_sqrt_sqrt
## D = 0.42308, p-value = 0.01905
## alternative hypothesis: two-sided
```

Wilcoxon and Kolmogorov-Smirnov tests didn't change for the same reason as before. But now all three tests reject H_0 and we can conclude that for 4th roots of measurements, the columns are distributed differently.

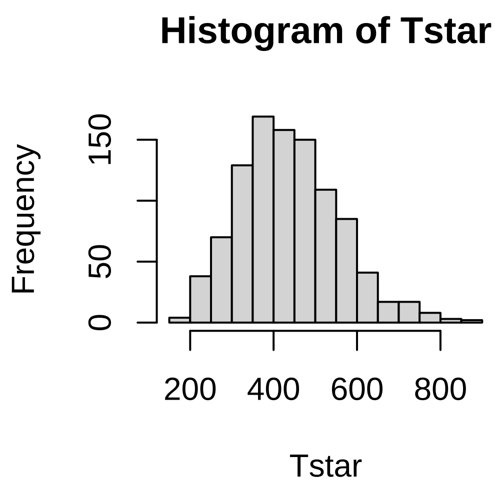
c) Let X1,...,X26 be the sample for seeded clouds (column *seeded*). Assuming X1,...,X26$\sim$Exp(λ) and using the central limit theorem, find an estimate $\hat{\lambda}$ of λ and construct a 95%-CI for λ. By using a bootstrap test with the test statistic T=median(X1,...,X26), test the hypothesis H_0:X1,...,X26$\sim$Exp(λ_0) with the parameter λ_0=$\hat{\lambda}$. Test this also by the Kolmogorov-Smirnov test.

```r
seeded <- clouds$seeded

B <- 1000
Tstar <- numeric(B)
for(i in 1:B){
  Xstar <- sample(seeded, replace=TRUE)
  Tstar[i] <- mean(Xstar)
}
lambda_hat <- 1/mean(Tstar)
lambda_hat
```

```
## [1] 0.002283315
```

```r
hist(Tstar)
```




```
alpha <- 1 - 0.95
deltastar <- 1/Tstar - lambda_hat
d <- quantile(deltastar, c(alpha/2, 1-alpha/2))
CI95 = lambda_hat - c(d[2], d[1])
lambda_hat; CI95
```

```
## [1] 0.002283315
```

```
##       97.5%        2.5%
## 0.000365829 0.003155424
```

Next, we check H_0: X1,...,X26$\sim$Exp(λ_0) with the parameter λ_0=$\hat{\lambda}$ using a bootstrap test.

```
B <- 1000
t <- median(seeded)
tstar <- numeric(B)
n <- length(seeded)
for(i in 1:B){
  xstar <- rexp(n, lambda_hat)
  tstar[i] <- median(xstar)
}
pl <- sum(tstar<t)/B
pr <- sum(tstar>t)/B
p <- 2*min(pl, pr)
pl;pr;p
```

```
## [1] 0.15
```

```
## [1] 0.85
```

```
## [1] 0.3
```

There is no evidence against H_0. Let's test the same hypothesis with Kolmogorov-Smirnov test:

```
ks.test(seeded, rexp(n, lambda_hat))
```

```
## Warning in ks.test(seeded, rexp(n, lambda_hat)): cannot compute exact p-value
## with ties
```

```
##
##  Two-sample Kolmogorov-Smirnov test
##
## data:  seeded and rexp(n, lambda_hat)
## D = 0.26923, p-value = 0.3027
## alternative hypothesis: two-sided
```

This test also doesn't reject the null hypothesis.

d) Using an appropriate test, verify whether the median precipitation for seeded clouds is less than 300. Next, design and perform a test to check whether the fraction of the seeded clouds with the precipitation less than 30 is at most 25%.

To check whether the median precipitation for seeded clouds is less than 300 (H_1), we will use binomial test for a proportion. The test is non–parametric, so we do not assume that the data is normally distributed. As the theoretical probabilities are equal, the binomial test becomes its special case - sign test.

```
binom.test(sum(seeded<300), length(seeded), p = 0.5, alternative = "less", conf.level = 0.95)
```

```
##
##  Exact binomial test
##
## data:  sum(seeded < 300) and length(seeded)
## number of successes = 17, number of trials = 26, p-value = 0.9622
## alternative hypothesis: true probability of success is less than 0.5
## 95 percent confidence interval:
##  0.0000000 0.8060396
## sample estimates:
## probability of success
##              0.6538462
```

Since this is not less than 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to say that median precipitation for seeded clouds is less than 300.

Similarly, we check whether the fraction of the seeded clouds with the precipitation less than 30 is at most 25%.

```
binom.test(sum(seeded<30), length(seeded), p = 0.25, alternative = "less", conf.level = 0.95)
```

```
##
##  Exact binomial test
##
## data:  sum(seeded < 30) and length(seeded)
## number of successes = 3, number of trials = 26, p-value = 0.08019
## alternative hypothesis: true probability of success is less than 0.25
## 95 percent confidence interval:
##  0.000000 0.271902
## sample estimates:
## probability of success
##              0.1153846
```

Again, we do not have sufficient evidence to say that the fraction of the seeded clouds with the precipitation less than 30 is at most 25%.

Exercise 3. Concentrations of epinephrine.

The concentrations (in nanograms per millimeter) of plasma epinephrine were measured for 10 dogs under *isofluorane*, *halothane*, and *cyclopropane* anesthesia, represented as three columns in data frame dogs.txt. We are interested in differences in the concentration for the different drugs.

```
dogs <- read.table("data/dogs.txt", header=TRUE)
```

a) Is it reasonable to assume that the three columns of dogs.txt were taken from normal populations?

```
shapiro.test(dogs$isofluorane)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dogs$isofluorane
## W = 0.83093, p-value = 0.03434
```

```
shapiro.test(dogs$cyclopropane)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dogs$cyclopropane
## W = 0.93334, p-value = 0.4815
```

```
shapiro.test(dogs$halothane)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dogs$halothane
## W = 0.9234, p-value = 0.3862
```

Only the data for isofluorane shows non-normal distribution with a p-value of 0.03434. However the concentrations of plasma epinephrine under cyclopropane and halothane are normally distributed. We conclude that these dogs are not from a normal population.

b) Investigate whether the columns isofluorane and halothane are correlated. Apply relevant tests to verify whether the distributions of these columns are different. Is a permutation test applicable?

As isofluorane column is not normally distributed, we use non-parametric correlation test.

```
cor.test(dogs$isofluorane, dogs$halothane, method="spearman")
```

```
##
```

```
##  Spearman's rank correlation rho
##
## data:  dogs$isofluorane and dogs$halothane
## S = 128.89, p-value = 0.5436
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.218846
```

The result shows small correlation according to Cohen rho = 0.218846. Therefore we conclude that the correlation is small and not significant.

To check whether the distributions of these columns are different, we apply a permutation test as normality is not assumed.

```r
mystat <- function(x, y) {mean(x-y)}
B <- 1000; tstar <- numeric(B)
for (i in 1:B){
  dogstar <- t(apply(cbind(dogs[,1], dogs[,2]), 1, sample))
  tstar[i] <- mystat(dogstar[,1], dogstar[,2])
}
myt <- mystat(dogs[,1], dogs[,2])
pl <- sum(tstar<myt)/B
pr <- sum(tstar>myt)/B
p <- 2*min(pl, pr)
pl;pr;p
```

```
## [1] 0.354
```

```
## [1] 0.644
```

```
## [1] 0.708
```

A permutation test with mean statistic didn't reject the H_0 that there is no difference between the distributions of isofluorane and halothane columns.

c) Conduct a one-way ANOVA to determine whether the type of drug has an effect on the concentration of plasma epinephrine. Give the estimated concentrations for each of the three anesthesia drugs.

```r
dogframe <- data.frame(concentration=as.vector(as.matrix(dogs)),
                       variety=factor(rep(1:3, each=10)))
aov <- lm(concentration~variety, data=dogframe)
anova(aov)
```

```
## Analysis of Variance Table
##
## Response: concentration
##           Df Sum Sq Mean Sq F value Pr(>F)
## variety    2 1.0808 0.54040   5.355  0.011 *
## Residuals 27 2.7247 0.10092
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

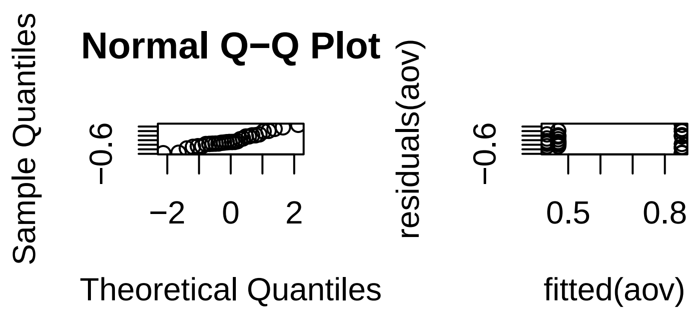
First of all the data has been conducted to one column of data set and then a one way anova has been conducted. The result of this one way anova is significant (p = 0,011), therefore there is an effect of the drug type on the concentration of plasma epinephrine.

We also have to check normality of errors.

```
shapiro.test(residuals(aov))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(aov)
## W = 0.96362, p-value = 0.3819
```

```
par(mfrow=c(1,2)); qqnorm(residuals(aov)); plot(fitted(aov),residuals(aov))
```



Residuals look normal and the fitted values show no pattern against them.

```
summary(aov)$coefficients
```

```
##             Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)    0.434  0.1004571 4.3202520 0.0001888078
## variety2       0.035  0.1420678 0.2463612 0.8072659880
## variety3       0.419  0.1420678 2.9492961 0.0065037447
```

The estimated concentrations are 0.434, 0.035, 0.419 for isofluorane, halothane, and cyclopropane respectively. For halothane t-test doesn't reveal a significant difference from 0.

d) Does the Kruskal-Wallis test arrive at the same conclusion about the effect of drug as the test in c)? Explain possible differences between conclusions of the Kruskal-Wallis and ANOVA tests.

```
kruskal.test(concentration ~ variety, data = dogframe)[[3]]
```

```
## [1] 0.05948078
```

H_0 is not rejected. The Kruskal-Wallis test did not arrive at the same conclusion as the one way ANOVA. Compared to the ANOVA, the Kruskal-Wallis test is a non-parametric counterpart of ANOVA which does not rely on normality but on ranks thereby a bit less powerful results than 1-way ANOVA.

Exercise 4. Hemoglobin in trout.

Hemoglobin is measured (g/100 ml.) in the blood of brown trout after 35 days of treatment with four rates of sulfamerazine: the daily rates of 0, 5, 10 and 15 g of sulfamerazine per 100 pounds of fish, denoted as rates 1, 2, 3 and 4, respectively. (Beware that the levels of the factor rate are coded by numbers.) Two methods (denoted as A and B) of administering the sulfamerazine were used. The data is collected in data set hemoglobin.txt.

a) Present an R-code for the randomization process to distribute 80 fishes over all combinations of levels of factors rate and method.

```
blood <- read.table("data/hemoglobin.txt", header=TRUE)
set.seed(42)
rows <- sample(nrow(blood))
randomized <- blood[rows, ]
```

b) Perform the two-way ANOVA to test for effects of factors rate, method and their interaction on the response variable hemoglobin. Comment on your findings.

```
res.aov3 <- aov(hemoglobin ~ rate * method, data = randomized)
summary(res.aov3)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## rate         1  27.93  27.931  11.933 0.000905 ***
## method       1   2.42   2.415   1.032 0.312963    
## rate:method  1   1.24   1.243   0.531 0.468373    
## Residuals   76 177.90   2.341                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results of the Two Way Anova analysis show that Rate has a significant effect F =11.933, p =0.000905, but method has no significant effect F = 1.032, p= 0.312963. Furthermore, there is no interaction effect between rate and method F = 0.531, p = 0.468373. The results show that only rate is a significant factor in influencing the Hemoglobin levels. The method that is used is not important because it does not influence the result, accordingly, the method does neither decrease nor increase the influence of the rate

c) Which of the two factors has the greatest influence? Is this a good question? Consider the additive model. Which combination of rate and method yield the highest hemoglobin? Estimate the mean hemoglobin value for rate 3 by using method A. What rate leads to the highest mean hemoglobin?

d) Test the null hypothesis that the hemoglobin is the same for all rates by a one-way ANOVA test, ignoring the variable method. Is it right/wrong or useful/not useful to perform this test on this dataset?

```
res.aov <- aov(hemoglobin ~ rate, data = randomized)
summary(res.aov)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## rate          1  27.93  27.931      12 0.000867 ***
## Residuals    78 181.55   2.328
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yes it is usefull because we already have shown that method has no influence at all (HOWEVER NOT SURE ABOUT THAT)

Exercise 5. Sour cream.

The file cream.txt contains data on an experiment to produce sour cream. Yogurt was placed in sweet cream, and yogurt bacteria were allowed to develop. Bacteria produce lactic acid, and as a surrogate for the number of yogurt bacteria, the acidity of the cream was measured. Interest was in the effect of the type of yogurt (denoted as *starter*) on *acidity*. The mixtures of yogurt and sweet cream were kept at constant temperature in a yogurt maker, in which five different positions could be used. The experiment was carried out with five batches of sweet cream, which were meant to have the same composition. With each batch each of five types of starter was used, with the yogurt placed in one of the five positions. The combinations of levels of three factors form a three-dimensional latin square. (You may need to install the R-package *lme4*, which is not included in the standard distribution of R.)

a) Analyze the data in a three-way experiment without interactions with acidity as response and starter, batch and position as factors.

By using summary command, can you tell whether there is a significant difference between the effects of starter 1 and starter 2 on acidity?

```r
cream <- read.table("data/cream.txt", header=TRUE)

cream$starter <- factor(cream$starter)
cream$position <- factor(cream$position)
cream$batch <- factor(cream$batch)

aovcream <- lm(acidity~batch+position+starter,data=cream)
anova(aovcream)
```

```
## Analysis of Variance Table
##
## Response: acidity
##           Df Sum Sq Mean Sq F value    Pr(>F)
## batch      4 18.778  4.6944  8.5975  0.001632 **
## position   4  2.348  0.5870  1.0750  0.411191
## starter    4 44.136 11.0340 20.2080 2.904e-05 ***
## Residuals 12  6.552  0.5460
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Repeated Anova is used to analyze these data. The results of the analysis show that batch has a significant effect p =0.00163 and starter has a significant effect p = 2.904e-05. However position does not show a significant effect p = 0.411.

b) Recall that the main interest is in the effect of starter on the acidity; factors *positions* and *batches* represent the block variables. Remove insignificant block variable(s) if there are such, and perform an ANOVA for the resulting "fixed effects" model. Which starter(s) lead to significantly different acidity?

c) For the model from b), can we also apply the Friedman test to test whether there is an effect of starter on acidity? Motivate your answer.

d) Repeat b) by performing a mixed effects analysis, modeling the block variable(s) (if there are any) as a random effect by using the function *lmer*. Compare your results to the results found by using the fixed effects model in b).