STAT 210
Applied Statistics and Data Analysis
Problem List 3
(due on week 4)

Fall 2025

Problem 1

The data for this question is stored in the file `PL3_P1.csv` and comes from a series of experiments testing the bending strength of corner plastic joints, measured in Newton-meter per centimeter (N-m/cm). On a routine quality control exercise, a sample of 20 joints was taken, and the bending strength was measured. The sample is stored as `bs1`. The reference value for the factory's production is 2.1 N-m/cm, and it is important that the joints do not have a smaller bending strength than the reference value.

(a) Graphical exploratory analysis:

- Create a boxplot of the data and add a horizontal red line on the reference value.
- Use also the function `stripchart` to create a plot of the observed values and add a vertical red bar on the reference value.
- Finally, use the function `sort` to order the values of `bs1` and feed the result to the function `dotchart`. Add a red line on the reference value.

Comment on what you observe. Do you think the engineers have cause for concern?

(b) The engineers in the plant assume that the sample comes from a normal distribution. Use graphical tools to verify if this assumption seems reasonable for this data set.

(c) Assuming that the sample comes from a normal distribution, describe explicitly the sampling distribution for the sample mean. Find a lower one-sided confidence interval for the mean at the 98% level. This gives a lower bound for the average bending strength at the 98% confidence level. Verify whether the reference value falls inside this interval and interpret the result.

(d) What parametric test would you use to check that the bending strength is not below the reference value? State clearly what hypotheses you are testing and which assumptions are needed for the test. Explain why you think they are satisfied. Describe the test statistic and calculate its value. Describe the sampling distribution and explicitly identify the errors of types I and II. Carry out this test and discuss the results.

(e) What non-parametric tests will be adequate for the problem in (d)? What assumptions are needed, and why do you think they are satisfied? Perform this test, discuss the results, and compare them with your previous results.

Solution

Read the data

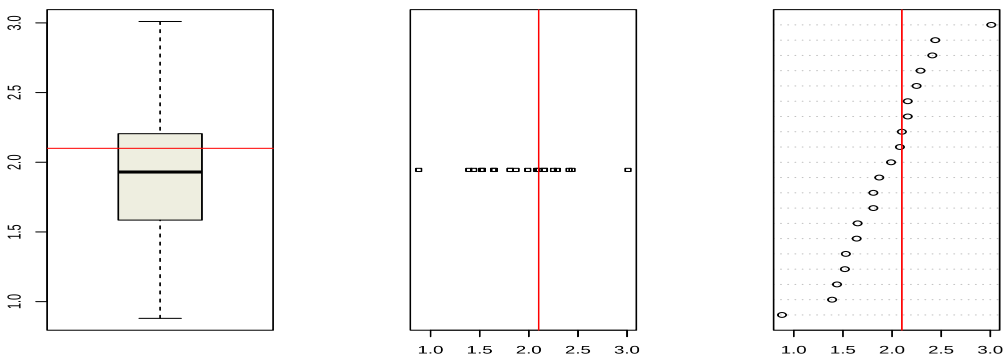
```
dat1 <- read.csv('PL3_P1.csv')
str(dat1)
```

```
## 'data.frame':    20 obs. of  5 variables:
```

```
## $ bs1   : num  1.99 2.25 2.29 3.01 0.88 2.16 2.16 2.44 1.81 2.1 ...
## $ bs_ref: num  1.63 1.93 2.7 2.21 2.16 2.17 2.59 2.48 2.2 1.76 ...
## $ bs_new: num  1.47 1.69 2.52 2.34 1.98 2.05 2.61 2.3 2.22 1.8 ...
## $ bs_A  : num  2.75 2.68 2.66 2.27 2.68 1.97 2.27 1.9 2.07 1.97 ...
## $ bs_B  : num  2.24 2.28 2.71 2.46 1.96 2.1 1.65 2.29 2.56 1.37 ...
```

(a) Graphical exploratory analysis

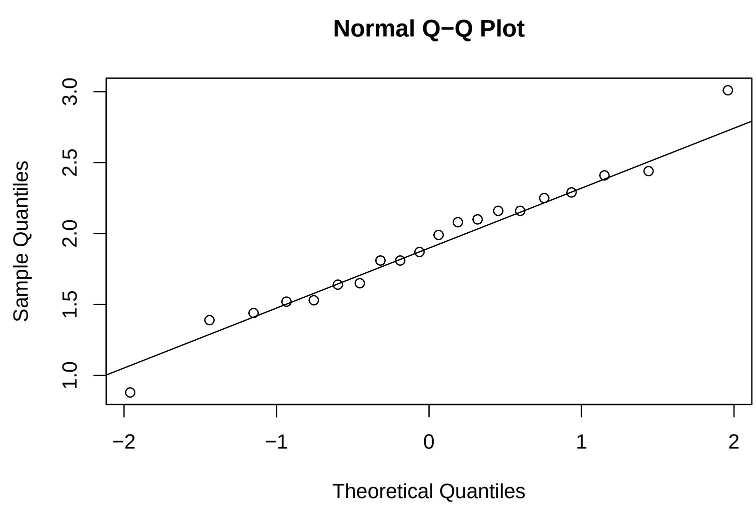
```
par(mfrow = c(1,3))
boxplot(dat1$bs1, col = 'ivory2')
abline(h= 2.1, col = 'red')
stripchart(dat1$bs1)
abline(v= 2.1, col = 'red')
dotchart(sort(dat1$bs1))
abline(v= 2.1, col = 'red')
```



The graphs show that the majority of the values fall below the reference value, so they should have cause for concern.

(b) We use a quantile plot for this

```
qqnorm(dat1$bs1)
qqline(dat1$bs1)
```



The plot is very good and supports the assumption of normality.

To verify this conclusion we can use the Shapiro-Wilk test, which is a goodness-of-fit test for the normal distribution.

```r
shapiro.test(dat1$bs1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dat1$bs1
## W = 0.98077, p-value = 0.9437
```

The p value is large and there is no evidence to reject the null hypothesis of normality.

(c) If we have a sample $X_1, X_2, \dots X_n$ from a normal distribution with mean μ and variance σ^2, the sampling distribution for $\bar{X}_n = (X_1 + \cdots + X_n)/n$ is also normal with parameters $N(\mu, \sigma^2/n)$, when the value of σ^2 is known. In this problem, this is not the case, but we know that when we need to estimate the variance by the sample variance s_n^2, we have that

$$\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \sim t_{n-1}$$

where t_{n-1} denotes a t-distribution with $n-1$ degrees of freedom. In our case we have

$$\frac{\bar{X}_n - 2.1}{s_n/\sqrt{20}} \sim t_{n-1}$$

Define $t_{n,\alpha}$ to be the α-quantile for the t distribution with n degrees of freedom, i.e., if T_n is a random variable with t_n distribution,

$$P(T_n \leq t_{n,\alpha}) = \alpha.$$

Using this, we have that

$$P\Big(\frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} \leq t_{19,0.02}\Big) = 0.02$$

and therefore

$$P\Big(\frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} > t_{19,0.02}\Big) = 0.98$$

By the symmetry of the t-distribution, we have that

$$P\Big(\frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} < -t_{19,0.02}\Big) = 0.98$$

Multiplying by $s_n/\sqrt{n}$ inside the probability, we get

$$P\Big(\hat{\mu}_n - \mu < -\frac{s_n t_{19,0.02}}{\sqrt{n}}\Big) = 0.98$$

And from this, we get that

$$P\Big(\mu > \hat{\mu}_n + \frac{s_n t_{19,0.02}}{\sqrt{n}}\Big) = 0.98$$

We now proceed to evaluate the terms in the formula

```r
n <- 20
error <- qt(0.02,df=n-1)*sd(dat1$bs1,)/sqrt(n)
mean(dat1$bs1,)-error
```

```
## [1] 2.153434
```

The confidence interval is $(-\infty, 2.153434)$ and we observe that it includes the reference value of 2.1. This implies that, based on this sample, a hypothesis test of the null hypothesis that the mean strength is 2.1 against the alternative that it is lower would not lead to the rejection of the null hypothesis at the 98% confidence level.

(d) The adequate parametric test is a one-sample t-test with one-sided alternative hypothesis, since we want to check that the bending strength is not below the reference value, i.e., the hypotheses are

$$H_0 \mu = 2.1 \qquad \text{vs.} \qquad H_1 : \mu < 2.1$$

The test is based on the assumption that the data come from a normal distribution, which is supported by the quantile plot above. The test statistic is

$$\hat{t} = \frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}}$$

which has a t_{19} (sampling) distribution. The value for this statistic is

```
(tstat <- (mean(dat1$bs1) -2.1)/(sd(dat1$bs1)/sqrt(20)))
```

```
## [1] -1.696772
```

A type I error would be to reject the null hypothesis that the average strength of the joints is 2.1 N-m/cm when, in fact it is. This may imply costs for the manufacturer, because it may imply a revision of the production process to determine where the problems are. A type II error would be to claim that the breaking strength of the joints is consistent with the reference value when, in fact, it is lower. This would produce joints that break more frequently than desired and may cause problems with the customers and loss of business.

We use `R` to do the test:

```
t.test(dat1$bs1, mu = 2.1, alternative = 'l', conf.level = 0.98)
```

```
##
##  One Sample t-test
##
## data:  dat1$bs1
## t = -1.6968, df = 19, p-value = 0.05303
## alternative hypothesis: true mean is less than 2.1
## 98 percent confidence interval:
##      -Inf 2.153434
## sample estimates:
## mean of x
##    1.9215
```

The p-value is small but not small enough to reject the null hypothesis at the 98% confidence level. We conclude that the sample does not give evidence that the strength of the joints produced in the factory is below the reference level.

Observe that the value of the test statistic coincides with the one we calculated above and also the confidence interval is the same as we obtained in (c).

(e) We use the Wilcoxon signed ranks test, which assumes that the sample comes from a symmetric distribution. Since we accepted previously that the sample comes from a normal distribution, we can assume that this condition is satisfied.

```
wilcox.test(dat1$bs1, mu = 2.1, alternative = 'l')
```

```
## Warning in wilcox.test.default(dat1$bs1, mu = 2.1, alternative = "l"): cannot
## compute exact p-value with ties

## Warning in wilcox.test.default(dat1$bs1, mu = 2.1, alternative = "l"): cannot
## compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
##
```

```
## data:  dat1$bs1
## V = 55, p-value = 0.05593
## alternative hypothesis: true location is less than 2.1
```

The p-value is similar to what we obtained with the t-test and we reach the same conclusion.

Problem 2

A chocolate manufacturer claims that the medium size chocolate bars they produce have a caloric content of 190 calories. The variable `barA` in the file `calories.csv` has the result of calorie measurements in 25 randomly chosen bars produced by the manufacturer.

(a) Graphical exploratory analysis:

- Create a boxplot and add a red line on the reference value of 190 calories.
- Plot a histogram for the data and add a graph for the estimated density. Add also a red line on the reference value.
- Finally, use the function `sort` to order the values of `barA` and feed the result to the function `dotchart`. Add a red line on the reference value.

Comment on what you observe. Do you think the manufacturer has cause for concern?

(b) The manufacturer assumes that the sample comes from a normal distribution. Use graphical tools to verify if this assumption seems reasonable for this data set.

(c) Assuming that the sample comes from a normal distribution, describe explicitly the sampling distribution for the sample mean. Find an upper one-sided confidence interval for the mean at the 99% level. This gives an upper bound for the average bending strength at the 99% confidence level. Verify whether the reference value falls inside this interval and interpret the result.

(d) What parametric test would you use to check that the caloric content is not above the reference value? State clearly what hypotheses you are testing and which assumptions are needed for the test. Explain why you think they are satisfied. Describe the test statistic and calculate its value. Describe the sampling distribution and explicitly identify the errors of types I and II. Carry out this test and discuss the results.

(e) What non-parametric tests will be adequate for the problem in (d)? What assumptions are needed, and why do you think they are satisfied? Perform this test, discuss the results, and compare them with your previous results.

Solution:

Read the data

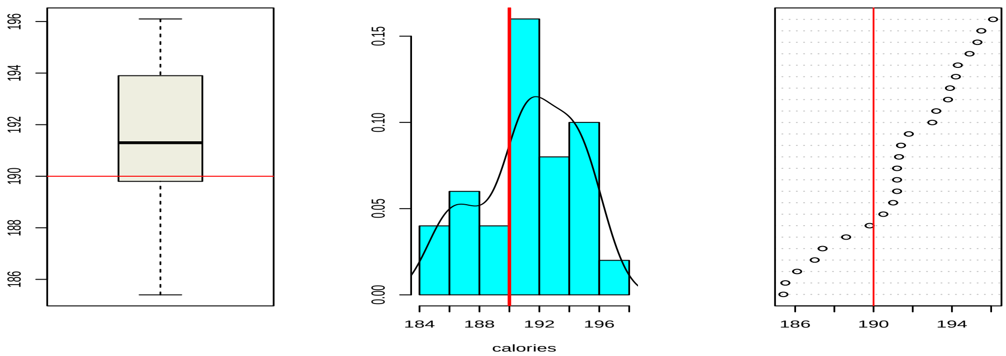
```
dat2 <- read.csv('calories.csv')
str(dat2)
```

```
## 'data.frame':    25 obs. of  4 variables:
##  $ barA: num  186 187 194 191 187 ...
##  $ barB: num  186 187 194 191 187 ...
##  $ barC: num  101 100 110 102 102 ...
##  $ barD: num  102.2 98.5 106 104.7 101 ...
```

(a) Graphical exploratory analysis

```
library(MASS)
par(mfrow = c(1,3))
boxplot(dat2$barA, col = 'ivory2')
```

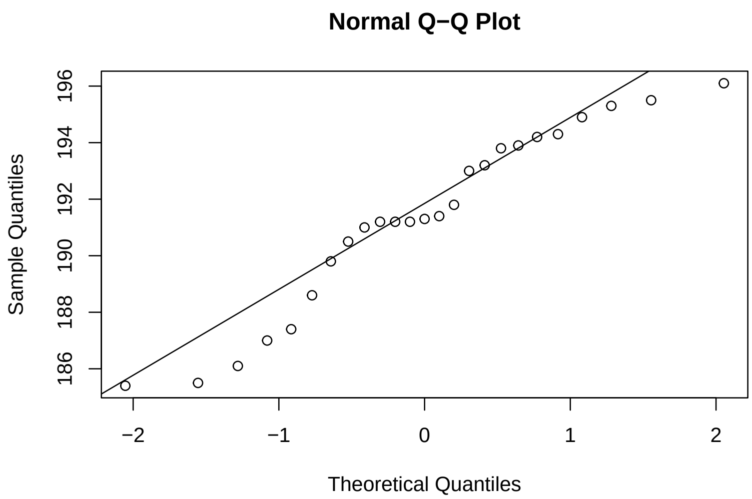
```r
abline(h= 190, col = 'red')
truehist(dat2$barA, prob = T, xlab = 'calories', main ='', h=2)
lines(density(dat2$barA))
abline(v= 190, col = 'red', lwd = 2)
dotchart(sort(dat2$barA))
abline(v= 190, col = 'red')
```



The plots show that the majority of values are above the reference value of 190 calories, giving the manufacturer cause for concern.

(b) We check this with a quantile plot:

```r
qqnorm(dat2$barA)
qqline(dat2$barA)
```



The fit does not look good, and we do a Shapiro-Wilk goodness-of-fit test for normality:

```r
shapiro.test(dat2$barA)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dat2$barA
## W = 0.93714, p-value = 0.1272
```

The p-value is above the usual levels and we do not reject the null hypothesis of normality.

(c) If we have a sample $X_1, X_2, \dots X_n$ from a normal distribution with mean μ and variance σ^2, the sampling distribution for $\bar{X}_n = (X_1 + \cdots + X_n)/n$ is also normal with parameters $N(\mu, \sigma^2/n)$, when the value of σ^2 is known. In this problem, this is not the case, but we know that when we need to estimate the variance by the sample variance s_n^2, we have that

$$\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \sim t_{n-1}$$

where t_{n-1} denotes a t-distribution with $n-1$ degrees of freedom. In our case we have

$$\frac{\bar{X}_n - 190}{s_n/\sqrt{25}} \sim t_{n-1}$$

Define $t_{n,\alpha}$ to be the α-quantile for the t distribution with n degrees of freedom, i.e., if T_n is a random variable with t_n distribution,

$$P(T_n \leq t_{n,\alpha}) = \alpha.$$

Using this, we have that

$$P\Big(\frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} \leq t_{24,0.01}\Big) = 0.01$$

and therefore

$$P\Big(\frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} > t_{24,0.01}\Big) = 0.99$$

Multiplying by $s_n/\sqrt{n}$ inside the probability, we get

$$P\Big(\hat{\mu}_n - \mu > \frac{s_n t_{24,0.01}}{\sqrt{n}}\Big) = 0.99$$

And from this, we get that

$$P\Big(\mu < \hat{\mu}_n - \frac{s_n t_{24,0.01}}{\sqrt{n}}\Big) = 0.99$$

We now proceed to evaluate the terms in the formula

```
n <- 25
error <- qt(0.01,df=n-1)*sd(dat2$barA,)/sqrt(n)
mean(dat2$barA,) + error
```

```
## [1] 189.7474
```

The confidence interval is $(189.7474, \infty)$ and we observe that it includes the reference value of 190. This implies that, based on this sample, a hypothesis test of the null hypothesis that the mean caloric content is 190 against the alternative that it is higher would not lead to the rejection of the null hypothesis at the 99% confidence level.

(d) The adequate parametric test is a one-sample t-test with one-sided alternative hypothesis, since we want to check that the caloric content is not above the reference value, i.e., the hypotheses are

$$H_0\mu = 190 \qquad \text{vs.} \qquad H_1 : \mu > 190$$

The test is based on the assumption that the data come from a normal distribution, which is supported by the Shapiro-Wilk test above. The test statistic is

$$\hat{t} = \frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}}$$

which has a t_{24} (sampling) distribution. The value for this statistic is

```
(tstat <- (mean(dat2$barA) -190)/(sd(dat2$barA)/sqrt(25)))
```

```
## [1] 2.09786
```

A type I error would be to reject the null hypothesis that the average caloric content of the bars is 190 calories when, in fact it is. This may imply costs for the manufacturer, because it may imply a revision of the production process to determine where the problems are. A type II error would be to claim that the caloric content of the bars is consistent with the reference value when, in fact, it is higher. This would produce chocolate bars that do not satisfy the conditions atated by the manufacturer and may cause problems with the customers and loss of business.

We use `R` to do the test:

```
t.test(dat2$barA, mu = 190, alternative = 'g', conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  dat2$barA
## t = 2.0979, df = 24, p-value = 0.02331
## alternative hypothesis: true mean is greater than 190
## 99 percent confidence interval:
##  189.7474      Inf
## sample estimates:
## mean of x
##   191.344
```

The p-value is small but not small enough to reject the null hypothesis at the 99% confidence level. We conclude that the sample does not give evidence that the caloric content of the joints produced by the manufactirer is above the reference level.

Observe that the value of the test statistic coincides with the one we calculated above and also the confidence interval is the same as we obtained in (c).

(e) We use the Wilcoxon signed ranks test, which assumes that the sample comes from a symmetric distribution. Since we accepted previously that the sample comes from a normal distribution, we can assume that this condition is satisfied.

```
wilcox.test(dat2$barA, mu = 190, alternative = 'g')
```

```
## Warning in wilcox.test.default(dat2$barA, mu = 190, alternative = "g"): cannot
## compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  dat2$barA
## V = 234.5, p-value = 0.02715
## alternative hypothesis: true location is greater than 190
```

The p-value is similar to what we obtained with the t-test and we reach the same conclusion.

Problem 3

This is a continuation of problem 2. The quality control department for the chocolate manufacturer wants to introduce a new and cheaper method for measuring the caloric content of the bars. Samples from the same chocolate bars as in problem 2 were used to test the new process. The results are stored in the variable `barB` in the file `calories.csv`, in the same order as the original measurements in variable `barA`.

(a) Draw boxplots to compare the results for the two measurement methods and comment on the results.

(b) Create a new variable `diff` which is the difference between the caloric content measured with the standard procedure minus the caloric content measured with the new procedure. If the two procedures are equivalent, we would expect these values to be close to zero. Do an exploratory graphical analysis to check if this looks likely. Check also whether the sample comes from a normal distribution using graphical tools.

(c) Assuming that the sample of differences comes from a normal distribution, describe explicitly the sampling distribution for the sample mean. Find an two-sided confidence interval for the mean at the 95% level. Verify whether zero falls inside this interval and interpret the result.

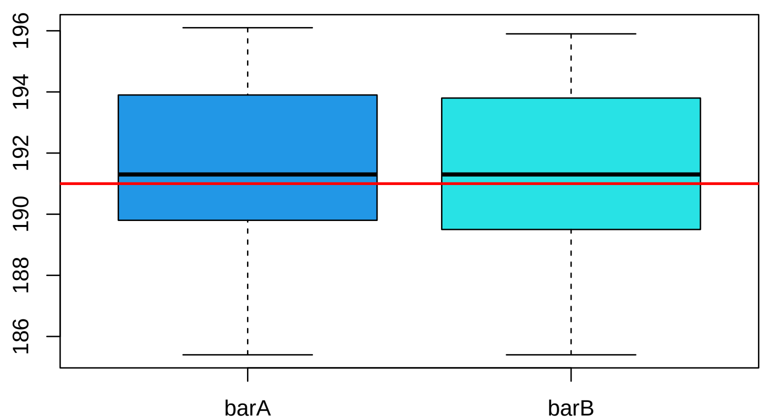
(d) What parametric test would be adequate for testing whether the average difference is zero? State clearly what hypothesis you are testing and which assumptions are needed for the test, and explain why they are satisfied. Identify the type I and type II errors. Carry out this test and discuss the results.

(e) What non-parametric tests will be adequate for the problem in (c)? What assumptions are needed, and why do you think they are satisfied? Perform these tests, discuss the results, and compare them with your previous results.

Solution:

(a) Boxplots:

```
boxplot(dat2[,1:2], col = c(4,5))
abline(h = 191, col = 'red', lwd = 2)
```



The boxplots are almost identical and they do not suggest differences in the methods for measuring the caloric content of the chocolate bars.

(b) Define the new variable and look at a numerical summary:

```
diff <- dat2$barB - dat2$barA
summary(diff)
```

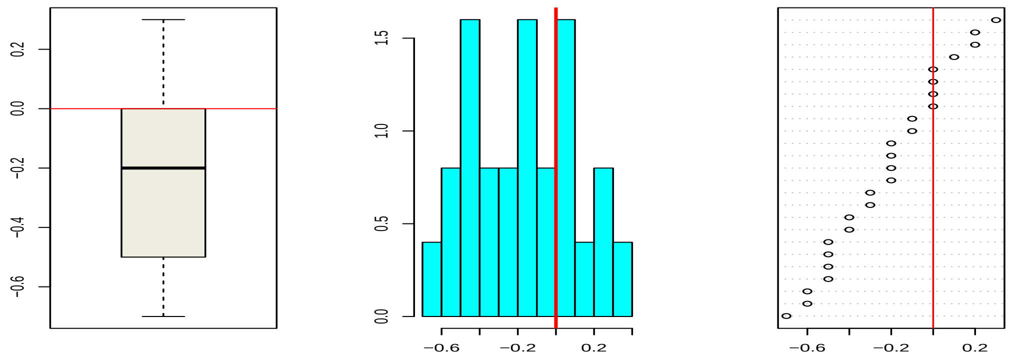
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   -0.70   -0.50   -0.20   -0.22    0.00    0.30
```

We see that the values are small but the third quartile is 0, which means that 3/4 of the data are below or equal to 0. This points to more negative than positive values.

Graphical exploratory analysis

```
library(MASS)
par(mfrow = c(1,3))
boxplot(diff, col = 'ivory2')
abline(h= 0, col = 'red')
truehist(diff, prob = T, xlab = '', main ='', h = 0.1)
```

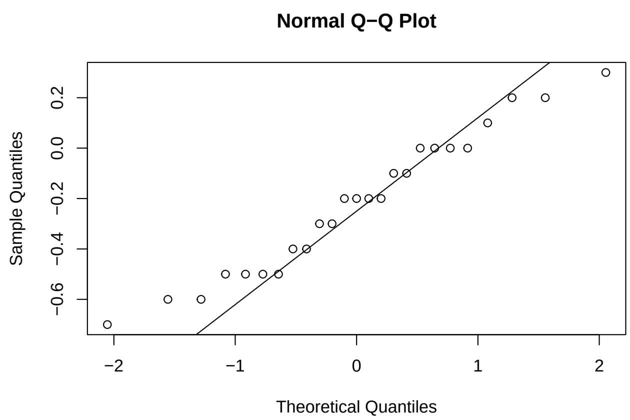
```r
lines(density(dat2$barA))
abline(v= 0, col = 'red', lwd = 2)
dotchart(sort(diff))
abline(v= 0, col = 'red')
```



We see that the sample values are not symmetric about zero, and most of the values are negative. This points to a possible difference in the measuring pocedures.

To check for normality we use the quantile plot:

```r
qqnorm(diff)
qqline(diff)
```



Since the precision of the measuring procedures is limited and there are repeated values, we see a 'staircase' pattern in the plot. We also look at the Shapiro-Wilk test for normality

```r
shapiro.test(diff)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  diff
## W = 0.96344, p-value = 0.4873
```

The p-value is large and we do not reject the null hypothesis that the data come from a normal distribution.

(c) If we have a sample $X_1, X_2, \ldots X_n$ from a normal distribution with mean μ and variance σ^2, the sampling distribution for $\bar{X}_n = (X_1 + \cdots + X_n)/n$ is also normal with parameters $N(\mu, \sigma^2/n)$, when the value of σ^2 is known. In this problem, this is not the case, but we know that when we need to estimate the variance by the sample variance s_n^2, we have that

$$\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \sim t_{n-1}$$

where t_{n-1} denotes a t-distribution with $n-1$ degrees of freedom. In our case we have

$$\frac{\bar{X}_n - 0}{s_n/\sqrt{25}} \sim t_{24}$$

Define $t_{n,\alpha}$ to be the α-quantile for the t distribution with n degrees of freedom, i.e., if T_n is a random variable with t_n distribution,

$$P(T_n \leq t_{n,\alpha}) = \alpha.$$

Using this we have that

$$P\Big(\frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} \leq t_{24,0.025}\Big) = 0.025$$

and therefore, using the symmetry of the t distribution,

$$P\Big(t_{24,0.025} < \frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} < -t_{24,0.025}\Big) = 0.95$$

Multiplying by $s_n/\sqrt{n}$ inside the probability, we get

$$P\Big(\frac{s_n t_{24,0.025}}{\sqrt{n}} < \hat{\mu}_n - \mu < -\frac{s_n t_{24,0.025}}{\sqrt{n}}\Big) = 0.95$$

And from this, we get that

$$P\Big(\hat{\mu}_n + \frac{s_n t_{24,0.025}}{\sqrt{n}} < \mu < \hat{\mu}_n - \frac{s_n t_{24,0.025}}{\sqrt{n}}\Big) = 0.95$$

We now proceed to evaluate the terms in the formula

```
n <- 25
error <- qt(0.025,df=n-1)*sd(diff)/sqrt(n)
c(mean(diff) + error , mean(diff) - error)
```

```
## [1] -0.334913 -0.105087
```

The confidence interval is $(-0.335, -0.105)$ and we observe that it does includes zero. This implies that, based on this sample, a hypothesis test of the null hypothesis that the two methods are equivalent against the alternative that they are not would not to the rejection of the null hypothesis at the 95% confidence level.

(d) The adequate parametric test is a one-sample t-test with two-sided alternative hypothesis, since we want to check that the two methods are equivalent, i.e., the hypotheses are

$$H_0 \mu = 0 \qquad \text{vs.} \qquad H_1 : \mu \neq 0$$

The test is based on the assumption that the data come from a normal distribution, which is supported by the Shapiro-Wilk test above. The test statistic is

$$\hat{t} = \frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}}$$

which has a t_{24} (sampling) distribution. The value for this statistic is

```
(tstat <- mean(diff)/(sd(diff)/sqrt(25)))
```

```
## [1] -3.951317
```

A type I error would be to reject the null hypothesis that the two measuring procedures produce the same result when, in fact they do. This may imply costs for the manufacturer, because it may imply not switching to a cheaper measuring process. A type II error would be to claim that the two measuring processes produce the same result when, in fact, they do not. This would produce inaccurate mesurements of the caloric content of the bars.

We use `R` to do the test:

```r
t.test(diff)
```

```
##
##  One Sample t-test
##
## data:  diff
## t = -3.9513, df = 24, p-value = 0.0005958
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.334913 -0.105087
## sample estimates:
## mean of x
##     -0.22
```

The p-value is small enough to reject the null hypothesis at the 95% confidence level. We conclude that the sample gives evidence that the two measuring procedures are not equivalent.

Observe that the value of the test statistic coincides with the one we calculated above and also the confidence interval is the same as we obtained in (c).

(e) We use the Wilcoxon signed ranks test, which assumes that the sample comes from a symmetric distribution. Since we accepted previously that the sample comes from a normal distribution, we can assume that this condition is satisfied.

```r
wilcox.test(diff)
```

```
## Warning in wilcox.test.default(diff): cannot compute exact p-value with ties

## Warning in wilcox.test.default(diff): cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  diff
## V = 24, p-value = 0.001516
## alternative hypothesis: true location is not equal to 0
```

The p-value is also small and we reach the same conclusion.

Problem 4

The Quality Control Engineer at an electronics plant wants to test high precision fuses produced in the factory. The fuses are supposed to take 100 milliseconds to cut off the current when there is an overload. From previous experience, the engineer knows that the normal distribution is a good approximation to the distribution of the breaking time of the fuses.

The engineer draws a sample of size 10 from the factory's production and obtains the values

```r
breaktimes <- c(98.808, 99.814, 99.406, 100.558, 99.114,
                99.618, 100.08, 98.36, 98.868, 99.73)
```

1) Using this sample, estimate mean, variance, and standard deviation. Which is the correct sampling distribution for the sample mean in this situation? Using this distribution, find a two-sided confidence interval at the 98% level for the mean. Determine whether the reference value of 100 ms falls inside this interval or not and interpret this.

Solution:

For this we use the functions `mean`, `var`, and `sd`.

```
(mean.bkt <- mean(breaktimes))
```

```
## [1] 99.4356
```

```
var(breaktimes)
```

```
## [1] 0.4350478
```

```
(sd.bkt <- sd(breaktimes))
```

```
## [1] 0.6595816
```

Since we have to estimate the variance from the sample, the correct distribution is the t-distribution with $n-1=9$ degrees of freedom. Recall that

$$\frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} \sim t_{n-1} \tag{1}$$

This is equation (5) in V15-IntervalEstimation.

We have from (1) that

$$P(t_{9,0.01} < \frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}} < t_{9,0.99}) = 0.98$$

where $t_{9,\alpha}$ is the α quantile for the t_9 distribution. Multiplying by $s_n/\sqrt{n}$ inside the probability, we get

$$P\Big(\frac{s_n t_{9,0.01}}{\sqrt{n}} < \hat{\mu}_n - \mu < \frac{s_n t_{9,0.99}}{\sqrt{n}}\Big) = 0.98$$

And from this, we get that

$$P\Big(\hat{\mu}_n - \frac{s_n t_{9,0.99}}{\sqrt{n}} < \mu < \hat{\mu}_n - \frac{s_n t_{9,0.01}}{\sqrt{n}}\Big) = 0.98$$

By symmetry, $t_{9,0.99} = -t_{9,0.01}$ and we finally get that

$$P\Big(\hat{\mu}_n - \frac{s_n t_{9,0.99}}{\sqrt{n}} < \mu < \hat{\mu}_n + \frac{s_n t_{9,0.99}}{\sqrt{n}}\Big) = 0.98$$

You can find this formula in slide 22 of V15-IntervalEstimation. We now proceed to evaluate the terms in the formula

```
n <- 10
error <- qt(0.99,df=n-1)*sd.bkt/sqrt(n)
c(mean.bkt-error,mean.bkt+error)
```

```
## [1]  98.84711 100.02409
```

2) Using this sample, the engineer wants to test whether the fuses have the expected break time. What hypothesis test should she carry out? What is the test statistic? Calculate the value for this statistic using the sample above. Using a confidence level of 98%, find a rejection region for the test. Is the value of the test statistic inside or outside the rejection region? What is your conclusion?

Solution

To test whether the break times are as expected, she should test

$$H_0 : \mu = 100 \qquad \text{vs.} \qquad \mu \neq 100$$

The test statistic is

$$\hat{t} = \frac{\hat{\mu}_n - \mu}{s_n/\sqrt{n}}$$

which by equation (1) has a t_9 distribution. The value for this statistic is

```
(t.hat <- (mean.bkt - 100)/(sd.bkt/sqrt(10)))
```

```
## [1] -2.705942
```

3) Carry out the test you selected in (2) using a command in `R` and look at the p-value that you obtain. What is your conclusion?

Solution:

The command is `t.test`:

```
t.test(breaktimes, mu = 100, conf.level = .98)
```

```
##
##  One Sample t-test
##
## data:  breaktimes
## t = -2.7059, df = 9, p-value = 0.02416
## alternative hypothesis: true mean is not equal to 100
## 98 percent confidence interval:
##   98.84711 100.02409
## sample estimates:
## mean of x
##   99.4356
```

The p-value is 0.02416, which is slightly bigger than the level set by the engineer. Therefore, we do not have enough evidence to reject the null hypothesis at this level.

Looking at the confidence interval in the result for the test, we see that it coincides with the interval calculated in (3).

4) Since the p-value is close to the confidence level that she set, the engineer decides to take a new sample of size 20 and obtains the following values

```
breaktimes2 <- c(99.448,99.486,100.128,98.118,99.248,99.488,100.338,
                 97.996,98.766,99.120,100.866,99.022,99.588,99.006,
                 98.150,99.632,98.374,98.588,98.824,100.016)
```

Using this new sample, repeat the test you carried out in (6) and comment on the results you obtain.

Solution:

```
t.test(breaktimes2, mu = 100, conf.level = .98)
```

```
##
##  One Sample t-test
##
## data:  breaktimes2
## t = -4.5712, df = 19, p-value = 0.0002086
## alternative hypothesis: true mean is not equal to 100
## 98 percent confidence interval:
##  98.77128 99.64892
## sample estimates:
```

```
## mean of x
##   99.2101
```

The p-value now is 0.0002086, which is small enough to reject the null hypothesis. The conclusion is that the break times are not as expected.

5) For the clients that buy the fuses, it is very important that the fuse does not take longer than expected to break the circuit, since this would put at risk the circuits they produce and would affect their guarantee, but it is not very important if the break occurs sooner than expected since this would not have an adverse effect. What would be a reasonable test of hypothesis in this context? Carry out this test and comment on the results.

Solution:

In this case the engineer should carry out a one-sided test:

$$H_0 : \mu = 100 \qquad \text{vs.} \qquad \mu > 100.$$

In `R`, we carry out this test with the command

```
t.test(breaktimes2, mu = 100, conf.level = .98, alternative = 'greater')
```

```
##
##  One Sample t-test
##
## data:  breaktimes2
## t = -4.5712, df = 19, p-value = 0.9999
## alternative hypothesis: true mean is greater than 100
## 98 percent confidence interval:
##  98.82913      Inf
## sample estimates:
## mean of x
##   99.2101
```

The p-value is now 0.9999, and we cannot reject the null hypothesis in this test. The break times for the fuses are not above the expected value.

By the way, observe the confidence interval that we get in the one-sided case: $(98.83, \infty)$.

6) In the lectures, we worked out an example about the power of a hypothesis test for the normal distribution. Assuming that the variance for the population was known, we calculated and plotted the power function. However, when the variance of the population is not known, calculating the power is not so easy since the sampling distribution under the alternative distribution changes. Fortunately, there is a function in `R` for doing this: `power.t.test`. Look at the help for this function to get familiar with the required arguments.

   Using `power.t.test`, calculate the power of the test

   $$H_0 : \mu = 100 \qquad \text{vs.} \qquad \mu = 100.4$$

   using a confidence level of 98% for the two sample sizes that we have considered before.

Solution:

```
power.t.test(n = 10, delta = 0.4, sd = sd(breaktimes),
             type = 'one.sample', sig.level = 0.02,
             alternative = 'one.sided')
```

```
##
##      One-sample t test power calculation
##
##               n = 10
##           delta = 0.4
##              sd = 0.6595816
##       sig.level = 0.02
##           power = 0.3600212
##     alternative = one.sided
```

```r
power.t.test(n = 20, delta = 0.4, sd = sd(breaktimes2),
             type = 'one.sample', sig.level = 0.02,
             alternative = 'one.sided')
```

```
##
##      One-sample t test power calculation
##
##               n = 20
##           delta = 0.4
##              sd = 0.7727745
##       sig.level = 0.02
##           power = 0.5524943
##     alternative = one.sided
```

7) We can use the `power.t.test` function to determine the sample size to obtain a given power for a fixed alternative and significance level for the test. Suppose the engineer wants to detect when the fuses are taking 2.5 milliseconds more than expected with a probability of at least 0.7. Calculate the sample size.

Solution:

```r
power.t.test(delta = 0.5, sd = sd(breaktimes2), power = 2/3, sig.level = 0.02,
             type = 'one.sample', alternative = 'one.sided')
```

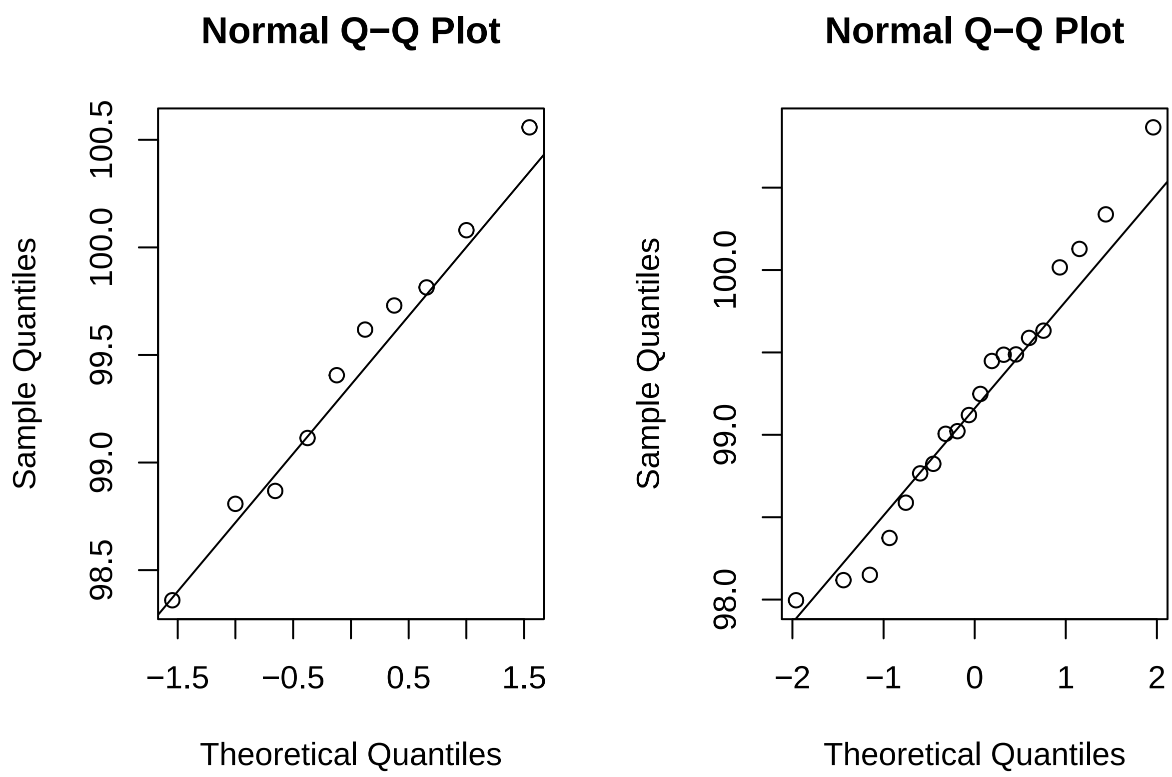
```
##
##      One-sample t test power calculation
##
##               n = 16.93739
##           delta = 0.5
##              sd = 0.7727745
##       sig.level = 0.02
##           power = 0.6666667
##     alternative = one.sided
```

Therefore, a sample of size 17 would be enough.

8) All the calculations we have made rely on the fact that the engineer 'knows' by experience that the break times of the fuses follows a normal distribution. How would you verify this assumption for the two samples considered?

Solution:

```r
par(mfrow=c(1,2))
qqnorm(breaktimes); qqline(breaktimes)
qqnorm(breaktimes2); qqline(breaktimes2)
```




```r
shapiro.test(breaktimes)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  breaktimes
## W = 0.98755, p-value = 0.9928
```

```r
shapiro.test(breaktimes2)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  breaktimes2
## W = 0.97604, p-value = 0.8736
```

9) Use a non-parametric test as an alternative to the tests carried out in Exercise 2 and compare your results.

Solution:

```r
wilcox.test(breaktimes, mu = 100)
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  breaktimes
## V = 6, p-value = 0.02734
## alternative hypothesis: true location is not equal to 100
```

```r
wilcox.test(breaktimes2, mu = 100)
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  breaktimes2
## V = 16, p-value = 0.0003223
## alternative hypothesis: true location is not equal to 100
```

Although the *p*-values are different, we would reach the same conclusions for both samples.

–>