# STAT 210
# Applied Statistics and Data Analysis
# First Exam

October 25, 2025

**You are not allowed to use AI tools to solve this exam**

**You are reminded to adhere to the academic integrity code established at KAUST.**

**This exam is open notes and open book but not open internet. You are not allowed to use the internet except for downloading the exam and uploading the solution.**

**Show complete solutions to get full credit. Writing code is not enough to answer a question. Your comments are more important than the code. Do not write comments inside chunks. Label your graphs appropriately. Please identify the files you submit with your surname**

**For all tests in this exam use a significance level of $\alpha = 0.01$ unless otherwise specified**

## Question 1 (50 points)

An industrial engineer is testing whether a new and cheaper machine calibration method (Method B) preserves production performance compared to the current method (Method A). The main quality measure is the tensile strength of metal rods (in MPa) produced by the machines.

a) Historically, rods produced using Method A have an average tensile strength of 150 MPa. After implementing Method B, a random sample of 18 rods is tested, yielding the following strengths:

```
methodB <- c(152.4, 155.2, 149.6, 153.8, 156.1, 150.9, 154.5, 157.0, 151.8,
             155.6, 150.3, 153.0, 154.2, 156.7, 149.8, 152.9, 155.1, 154.0)
```

What parametric test would you use to compare the new calibration method with 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.

the only measure of quality we have is the tensile strength, and we know MethodA had an average strength of 150 MPa. Thus we want to test if the new method can keep up with this quality control

We would use a one-sample left tailed t-test with the hypotheses:

H0: mu = 150 (quality is maintained)

H1: mu != 150 (quality has changed)

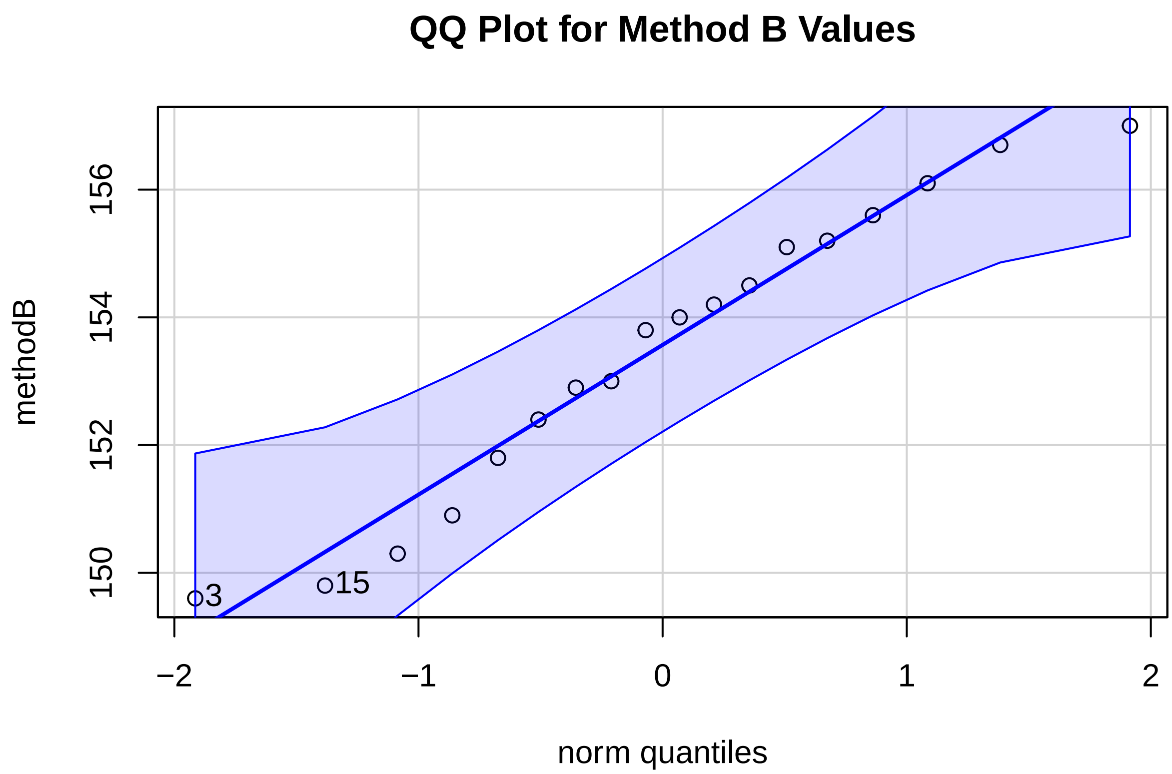
Assumptions of the t-test are: Independance: samples (rods) should not affect each other. Identical conditions(low variance): assumed since all factors in the factory remain unchanged except for the method itself. approximately normal distribution: We test that in the following snippet of code.

```r
library(car)
```

```
## Loading required package: carData
```

```r
methodB <- c(152.4, 155.2, 149.6, 153.8, 156.1, 150.9, 154.5, 157.0, 151.8,
             155.6, 150.3, 153.0, 154.2, 156.7, 149.8, 152.9, 155.1, 154.0)

qqPlot(methodB, line = "r", main = "QQ Plot for Method B Values")
```




```
## [1]  3 15
```

```r
shapiro.test(methodB) #shapiro-wlk test of normality.
```

```
##
##  Shapiro-Wilk normality test
##
## data:  methodB
## W = 0.955, p-value = 0.52
```

We can see through the QQ Plot and the shapiro test that we have strong evidence that the data is normally distributed, and we know that the t-test would work.

the test statistic formula is t = (mean- mu) / (sd / sqrt(n)) with mu being our hypothesis (150).

```r
(t_stat <- (mean(methodB) - 150) / (sd(methodB) / sqrt(25)))
```

```
## [1] 7.5291
```

Type I Error (False Positive): concluding that the system did not preserve the quality even if mu = 150 in other words, falsely rejecting the hypothesis.

Type II Error (False Negative): concluding that the system did preserve quality even though mu != 150 in other words, false failure to reject hypothesis.

```r
#I did the test on multiple alternatives depending on what the engineer wants to test.

#If we wanted to check that we at least have 150 MPa,
# H1: mu < 150
#t.test(methodB, mu = 150, alternative = "less", conf.level = 0.99)

#If we wanted to check that we hold that value and it doesn't change much which is my understanding of
# H1: mu != 150
t.test(methodB, mu = 150, alternative = "two.sided", conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  methodB
## t = 6.39, df = 17, p-value = 6.7e-06
## alternative hypothesis: true mean is not equal to 150
## 99 percent confidence interval:
##  151.91 155.08
## sample estimates:
## mean of x
##    153.49
```

```r
#If we wanted to check that our tensile has at most 150 MPa,
# H1: mu > 150
#t.test(methodB, mu = 150, alternative = "greater", conf.level = 0.99)


#If the wanted test is for the tensile strength to have a minimum of 150, then we have very strong evid

#If the objective was to hold a maximum tensile strength at 150 then we reject the null hypothesis base
#and we have evidence pointing that the mean tensile strength increased
```

If the objective was to test if the new method would hold the tensile strength value, then we reject the null hypothesis based on the p-value 6.735e-6 « 0.01 and we have evidence to point out that the mean tensile strength shifted.

b) To compare the two methods directly, two machines are run for one day each: Machine A uses the standard calibration, Machine B uses the new calibration method. The tensile strengths (in MPa) of 12 rods from each machine are recorded:

```
machineA <- c(149.2, 151.0, 148.7, 150.1, 149.8, 152.3, 150.5, 151.2,
              148.9, 149.7, 150.9, 149.4)
machineB <- c(153.4, 155.1, 152.8, 154.3, 156.0, 155.6, 153.9, 154.8,
              152.5, 155.2, 154.0, 153.7)
```

The engineer now wants to specifically test whether Machine B produces stronger rods than Machine A. What parametric test would be adequate for comparing the tensile strengths corresponding to the two methods? State clearly what hypothesis you are testing and which assumptions are needed for the test. Explain why you think they are satisfied. Carry out this test and discuss the results.

Since we have two data sets, we can use the paired t-test with the following hypotheses:

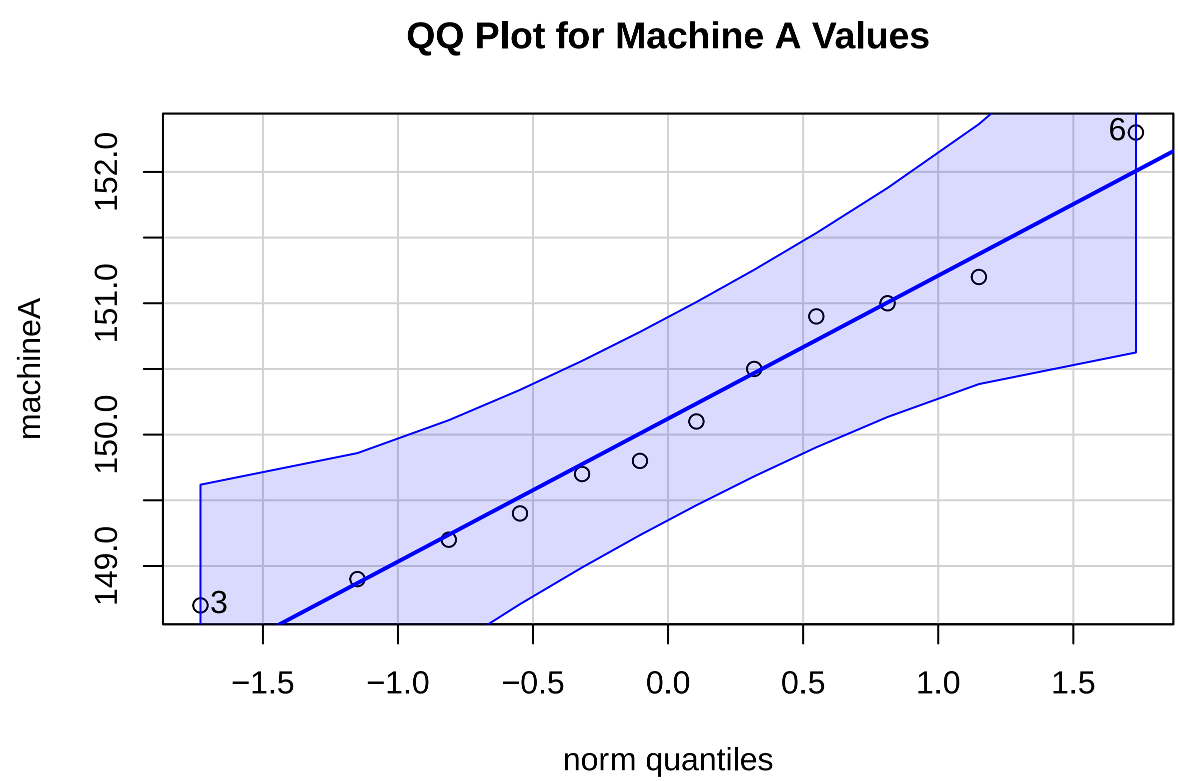
H0: mu_B = mu_A (same strength)

H1: mu_B > mu_A (methodB produces stronger rods)

Assumptions of the paired t-test are similar to the one-sample t-test:

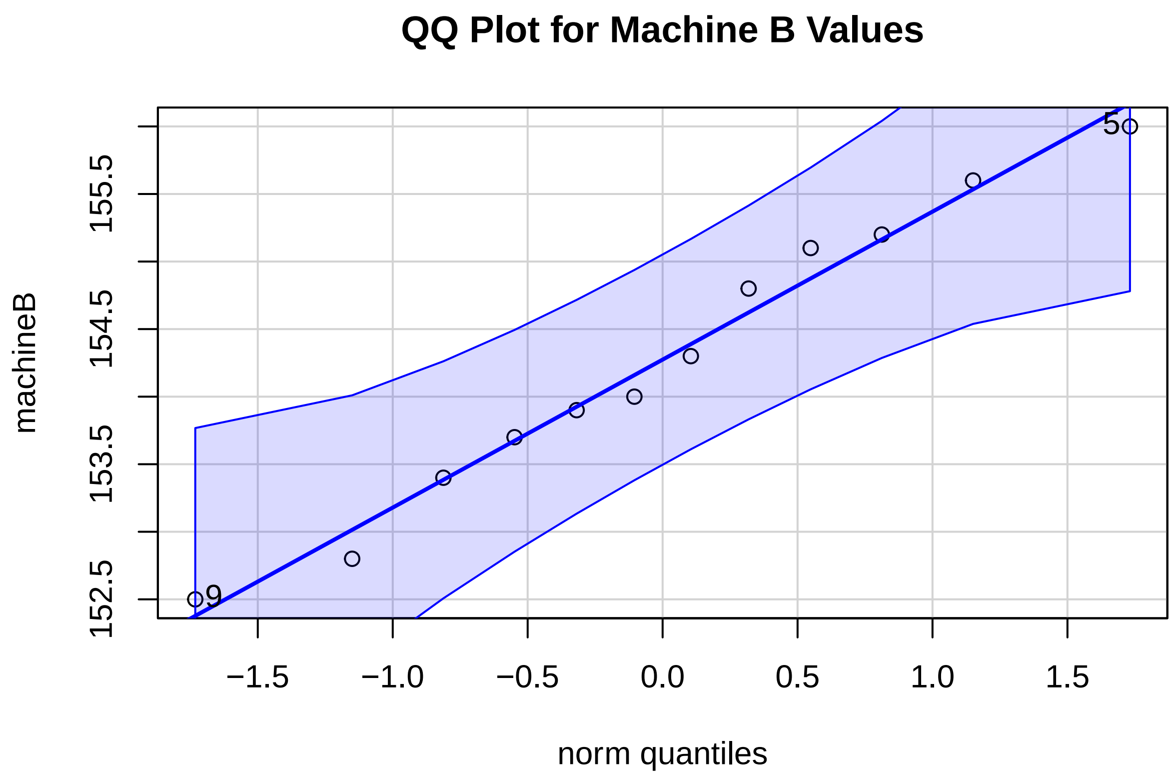
Independance: samples (rods) should not affect each other. Identical conditions(low variance): assumed since all factors in the factory remain unchanged except for the method itself, and they all run for one day each. approximately normal distribution: We test that in the following snippet of code.

```
machineA <- c(149.2, 151.0, 148.7, 150.1, 149.8, 152.3, 150.5, 151.2,
              148.9, 149.7, 150.9, 149.4)
machineB <- c(153.4, 155.1, 152.8, 154.3, 156.0, 155.6, 153.9, 154.8,
              152.5, 155.2, 154.0, 153.7)
qqPlot(machineA, line = "r", main = "QQ Plot for Machine A Values")
```



```
## [1] 6 3
```

```r
qqPlot(machineB, line = "r", main = "QQ Plot for Machine B Values")
```




```
## [1] 9 5
```

```r
shapiro.test(machineA) #shapiro-wlk test of normality.
```

```
##
##  Shapiro-Wilk normality test
##
## data:  machineA
## W = 0.962, p-value = 0.81
```

```r
shapiro.test(machineB) #shapiro-wlk test of normality.
```

```
##
##  Shapiro-Wilk normality test
##
## data:  machineB
## W = 0.974, p-value = 0.95
```

From the QQ plots of the data and the shapiro test pvalues of 0.9514 and 0.8054 » 0.01, we can safely assume the data are normally distributed.

Type I Error (False Positive): conclude that machine B does not produce stronger rods even though mu_B > mu_A

Type II Error (False Negative): conclude that machine B produces stronger rods when mu_B = mu_A.

```
t.test(machineB, machineA, paired = TRUE, alternative = "greater", conf.level = 0.99)
```

```
##
##  Paired t-test
##
## data:  machineB and machineA
## t = 15.8, df = 11, p-value = 3.3e-09
## alternative hypothesis: true mean difference is greater than 0
## 99 percent confidence interval:
##  3.4218    Inf
## sample estimates:
## mean difference
##          4.1333
```

With a p-value of 3.319e-09 « 0.01, we reject the null hypothesis and we have strong evidence suggesting machine B does in fact produce stronger rods than machine A

c) In a controlled laboratory experiment, 10 batches of raw material were processed under both calibration methods (A and B). The same batch of material was used for both methods to eliminate raw material variation. The measured tensile strength results (MPa) are:

```
mtdA <- c(148.5, 149.7, 150.3, 151.1, 149.0, 150.8, 149.5, 150.0, 151.2, 149.8)
mtdB <- c(149.6, 150.5, 151.1, 152.0, 149.2, 151.6, 150.1, 151.0, 152.3, 150.4)
```

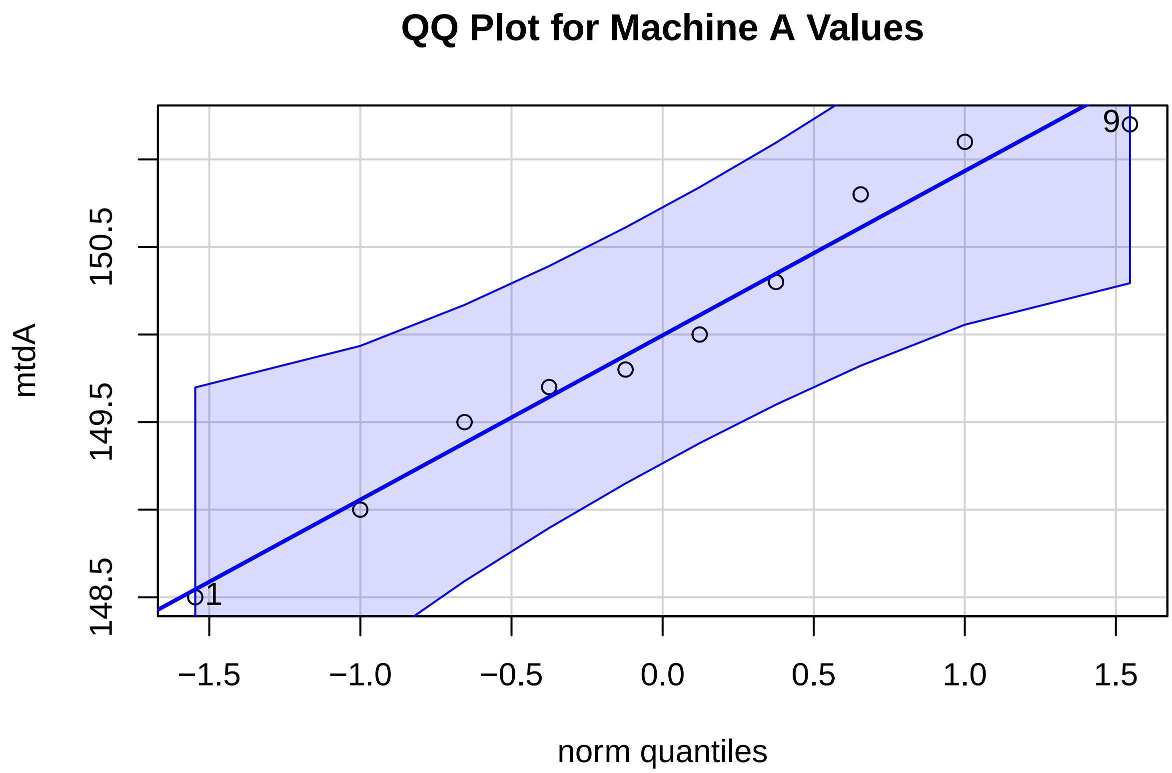
The engineer wants to determine if Method B (`mtdB`) produces stronger rods than Method A (`mtdA`). What parametric test would you use in this case? State clearly what hypothesis you are testing and which assumptions are needed for the test, and explain why you think they are satisfied. Identify the type I and type II errors. Carry out this test and discuss the results.

Similar to part B, We can do a paired t-test to analyze this data with hypotheses: H0: mu_B = mu_A H1: mu_B > mu_A

Assumptions:

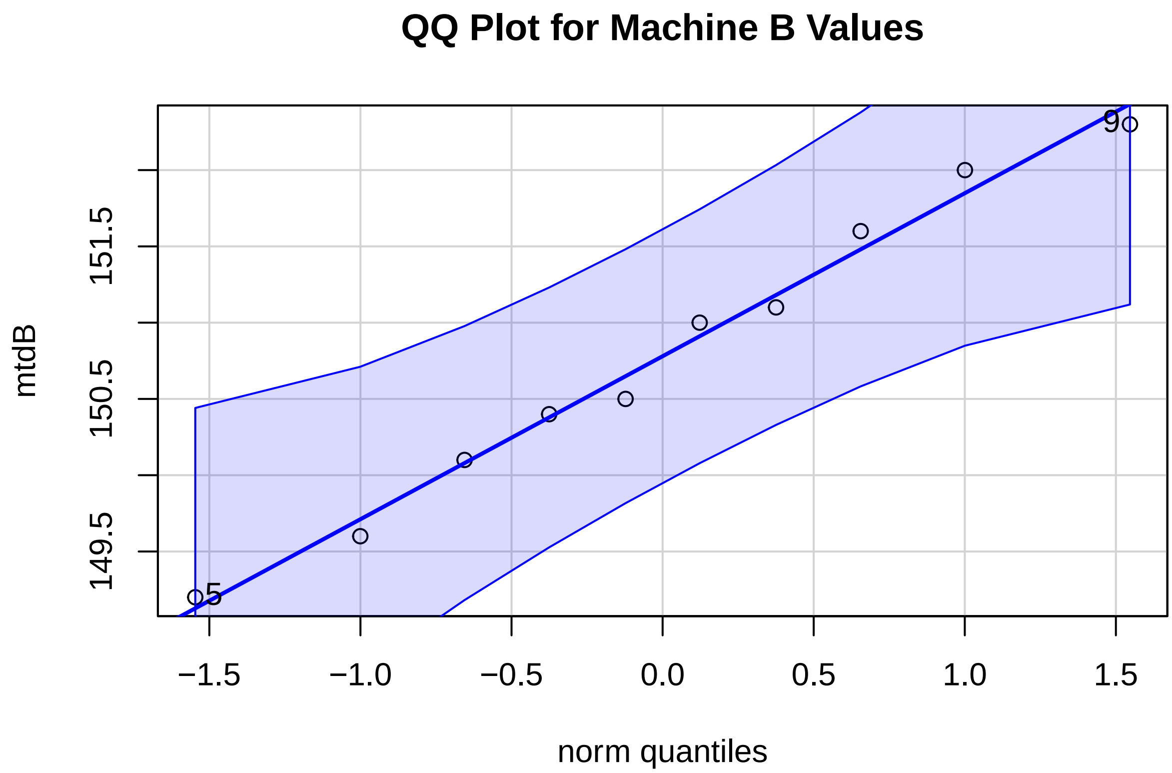
Independance: samples (rods) should not affect each other. further emphasized since this is a controlled lab environment. Identical conditions(low variance): assumed since the same raw materials were used and the sampling was done in a controlled environment. approximately normal distribution: We test that in the following snippet of code.

```
mtdA <- c(148.5, 149.7, 150.3, 151.1, 149.0, 150.8, 149.5, 150.0, 151.2, 149.8)
mtdB <- c(149.6, 150.5, 151.1, 152.0, 149.2, 151.6, 150.1, 151.0, 152.3, 150.4)
qqPlot(mtdA, line = "r", main = "QQ Plot for Machine A Values")
```




```
## [1] 1 9
```

```
qqPlot(mtdB, line = "r", main = "QQ Plot for Machine B Values")
```




```
## [1] 5 9
```

```r
shapiro.test(mtdA) #shapiro-wlk test of normality.
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mtdA
## W = 0.965, p-value = 0.84
```

```r
shapiro.test(mtdB) #shapiro-wlk test of normality.
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mtdB
## W = 0.976, p-value = 0.94
```

from the QQ plots and the shapiro tests p-values of 0.8371 and 0.9388 » 0.01, normal distribution assumption is valid.

Type I Error (False Positive): conclude that mtdB does not produce stronger rods even though mu_B > mu_A

Type II Error (False Negative): conclude that mtdB produces stronger rods when mu_B = mu_A.

```r
t.test(mtdB, mtdA, paired = TRUE, alternative = "greater", conf.level = 0.99)
```

```
##
##  Paired t-test
##
## data:  mtdB and mtdA
## t = 9.16, df = 9, p-value = 3.7e-06
## alternative hypothesis: true mean difference is greater than 0
## 99 percent confidence interval:
##  0.54674     Inf
## sample estimates:
## mean difference
##            0.79
```

given that the p-value is 3.685e-6 « 0.01, we reject the null hypothesis and have evidence pointing to the fact that mtdB does in fact produce stronger rods.

(d) What non-parametric tests will be adequate for the problems in (a), (b), and (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.

A non-parametric alternative to the t-test would be the Wilcoxon sign-ranked test.

The Wilcoxon signed rank test could be thought of as the non-parametric version of the t-test. it is used to check if the median of the population differs from a specified value

Assumptions: Continuous Data: correct assumption given that we found the normality assumption to be true Symmetry: Data points are symmetric around the median, which can be seen in the QQ plots. also reasonable since we found normality assumption to be true

```r
#Non-parametric test for data in part A
#Wilcoxon Signed rank test

#H0: median of methodB = 150
#H1: median of methodB != 150
wilcox.test(methodB,
            mu = 150,
            alternative = "two.sided",
            conf.level = 0.99)
```

```
##
##  Wilcoxon signed rank exact test
##
## data:  methodB
## V = 167, p-value = 5.3e-05
## alternative hypothesis: true location is not equal to 150
```

```r
#Non-parametric test for data in part B
#Paired Wilcoxon Signed rank test

#H0: median of machineB = median of machineA
#H1: median of machineB > median of machineA
```

```r
wilcox.test(machineB, machineA,
            paired = TRUE,
            alternative = "greater",
            conf.level = 0.99)
```

```
## Warning in wilcox.test.default(machineB, machineA, paired = TRUE,
## alternative = "greater", : cannot compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  machineB and machineA
## V = 78, p-value = 0.0013
## alternative hypothesis: true location shift is greater than 0
```

```r
#Non-parametric test for data in part C
#Paired Wilcoxon Signed rank test

#H0: median of mtdB = median of mtdA
#H1: median of mtdB > median of mtdA
wilcox.test(mtdB, mtdA,
            paired = TRUE,
            alternative = "greater",
            conf.level = 0.99)
```

```
## Warning in wilcox.test.default(mtdB, mtdA, paired = TRUE, alternative
## = "greater", : cannot compute exact p-value with ties
```

```
##
##  Wilcoxon signed rank test with continuity correction
##
## data:  mtdB and mtdA
## V = 55, p-value = 0.0029
## alternative hypothesis: true location shift is greater than 0
```

For part A using the wilcox test, we get a p-value of 5.341e-05 « 0.01 which means we reject the null hypothesis and have strong evidence that the medians are different; matching our previous parametric test.

For part B using the wlcox test, we get a p-value of $0.001 < 0.01$ which means we reject the null hypothesis, and have evidence that machine B does produce stronger rods; matching our previous parametric test

For part C using the wlcox test, we get a p-value of $0.002929 < 0.01$ which means we reject the null hypothesis, and have evidence that mtdB does produce stronger rods; matching our previous parametric test

---

## Question 2 (50 points)

An industrial engineer wants to compare five different lubricants used in a manufacturing process to determine their effect on the mean friction coefficient of machine parts. The lubricants are coded `A`, `B`, `C`, `D`, and `E`. Each lubricant is tested on five identical machines under the same operating conditions, and the friction coefficient is measured (lower values are better). The data is stored in the file `XM125F_Q2.csv`.

Do a complete analysis of variance model including the following: Plot the data. Determine whether the lubricants have an effect on the mean friction coefficient through a hypothesis test and state explicitly the null and alternative hypotheses. Obtain the estimated values for the cell (marginal) means and the effects. Write the equation for the model and state explicitly the assumptions on which the model is based. Plot the diagnostic charts and comment on them. Use Levene's and Shapiro-Wilk's tests also. Use Tukey's HSD procedure to make pairwise comparisons and comment on the results. Use a non-parametric alternative to the analysis of variance and compare the results. Give your comments on every step that you take. If the objective is a lower friction coefficient, which design would you select and why?

```
data <- read.csv("XM125F_Q2.csv", header = TRUE) # read the file
str(data)
```

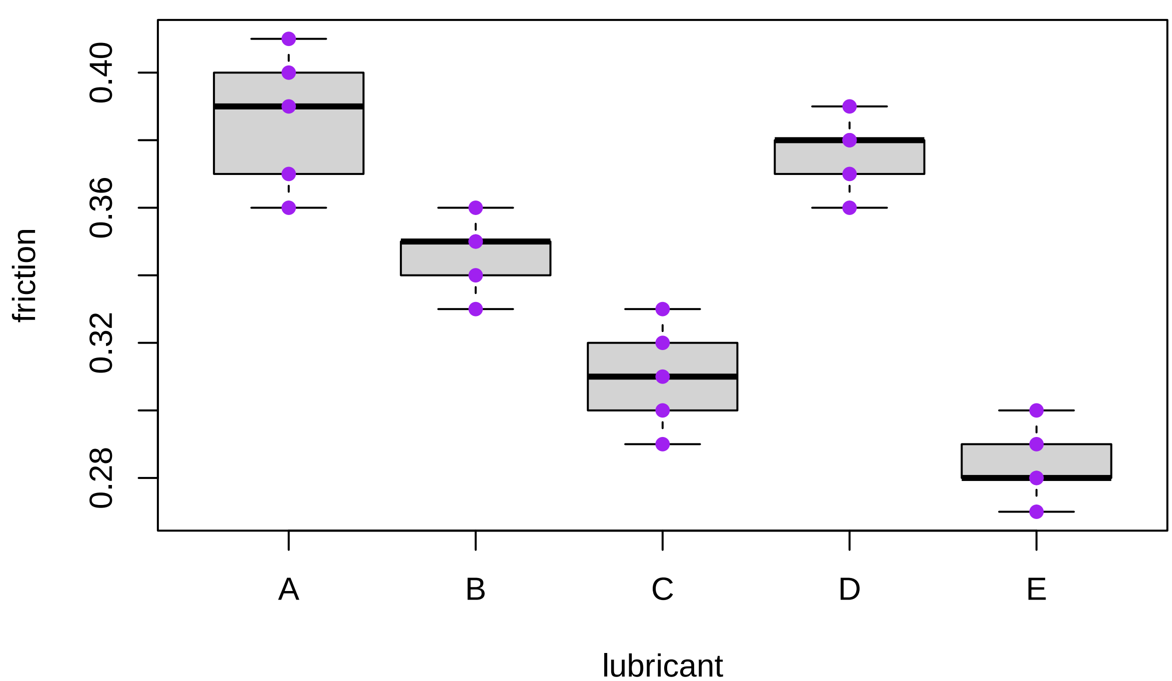
```
## 'data.frame':    25 obs. of  2 variables:
##  $ friction : num  0.41 0.39 0.37 0.36 0.4 0.35 0.33 0.34 0.36 0.35 ...
##  $ lubricant: chr  "A" "A" "A" "A" ...
```

```
data$lubricant <- as.factor(data$lubricant) #make the lubricant a factor instead of character
str(data)
```

```
## 'data.frame':    25 obs. of  2 variables:
##  $ friction : num  0.41 0.39 0.37 0.36 0.4 0.35 0.33 0.34 0.36 0.35 ...
##  $ lubricant: Factor w/ 5 levels "A","B","C","D",..: 1 1 1 1 1 2 2 2 2 2 ...
```

```
boxplot(friction ~ lubricant, data = data) #boxplot of the friction for each lubricant

points(friction ~ lubricant,data = data, col = "purple", pch = 16) #plot the points on the boxplot
```



For the parametric test, we will use an anova with the hypotheses as follows: H0: mu_A = mu_B = mu_C = mu_D = mu_E H1: at least one mean is different.

H0 basically means that all lubricant types are the same and have no effect.

The anova model assumptions are independence, normality, and equal variance.

```
mod1<- aov(friction ~ lubricant, data = data)
summary(mod1)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## lubricant    4 0.0374 0.00935    43.7 1.3e-09 ***
## Residuals   20 0.0043 0.00021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value of the anova (1.3e-9) is lower than our signficance level of 0.01 we reject the null hypothesis and we have evidence to say that at least one lubricant is different.

we then find the cell means and standard error.

```
model.tables(mod1, 'mean', se = T)
```

```
## Tables of means
## Grand mean
##
## 0.3404
##
##  lubricant
## lubricant
##     A     B     C     D     E
## 0.386 0.346 0.310 0.376 0.284
##
## Standard errors for differences of means
##         lubricant
##          0.009252
## replic.         5
```

```
model.tables(mod1, se = T)
```

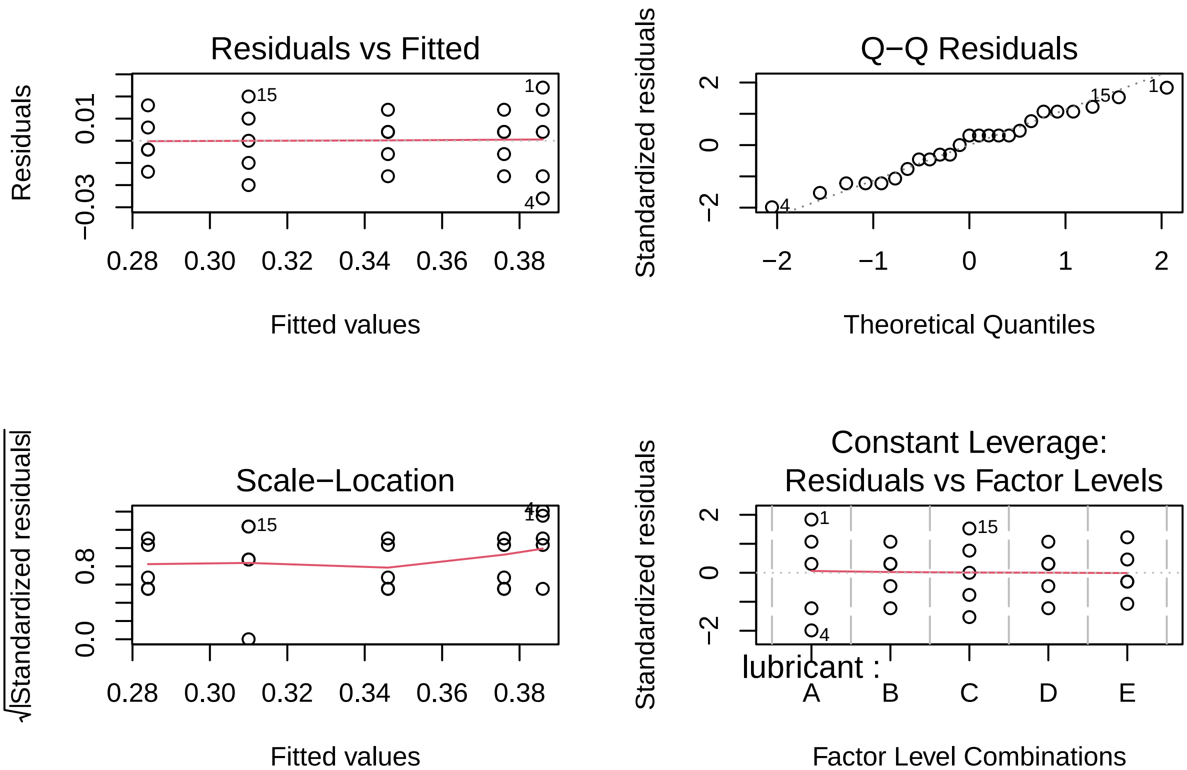
```
## Tables of effects
##
##  lubricant
## lubricant
##       A       B       C       D       E
##  0.0456  0.0056 -0.0304  0.0356 -0.0564
##
## Standard errors of effects
##         lubricant
##          0.006542
## replic.         5
```

Equation of the anova model: Y_ij = mu + tau_i + epsilon_ij

Where: Y_ij = friction for the j-th observation in the i-th lubricant mu = overall mean friction strength tau_i = effect of the i-th lubricant (deviation from overall mean) epsilon_ij = random error term i = 1, 2, 3, 4, 5 (lubricant types : A, B, C, D, E) j = 1, 2, . . . , n_i (observations within each group)

we then print out the diagnostic plots:

```
par(mfrow=c(2,2))
plot(mod1)
```




From the plot we see the following: -Residuals vs fitted: the line is orizantal at or very near 0, which indicates linearity -QQ Residuals: The data is aligned with the normal line, which indicates normality. -Scale-Location: the line is mostly horizantal with some tails, further testing may be needed to decide if variance is equal (levene test) -Constant lev: plot shows no influential outliers

Testing for equal variance:

```
leveneTest(mod1)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value Pr(>F)
## group  4    0.78   0.55
##       20
```

With a p-value of 0.551 » 0.01, we fail to reject the null hypothesis and evidence suggests that we have equal variance.

Testing for normality:

```r
shapiro.test(rstandard(mod1))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  rstandard(mod1)
## W = 0.971, p-value = 0.66
```

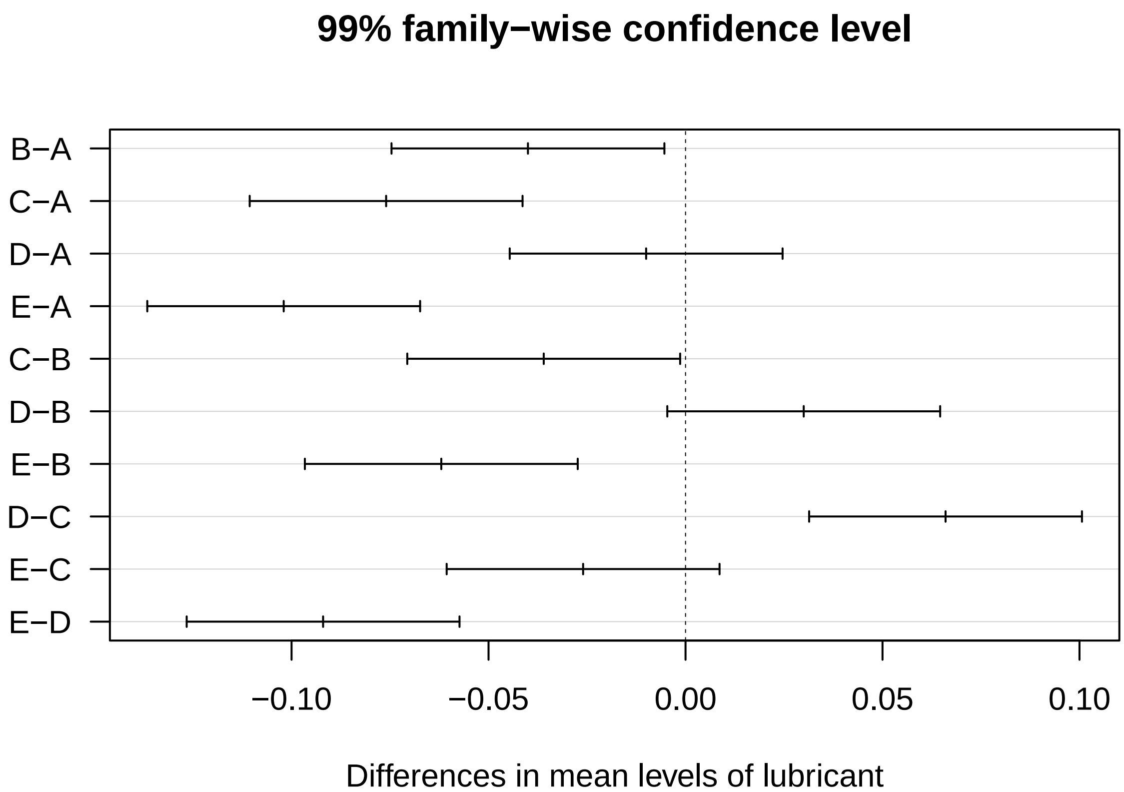
with a p-value of 0.6644 » 0.01, we fali to reject the null hypothesis and evidence suggests that data is normal

Next we use TukeyHSD with confidence level = 0.99 to further analyze the data:

```r
(mod1.tky <- TukeyHSD(mod1, conf.level = 0.99))
```

```
##   Tukey multiple comparisons of means
##     99% family-wise confidence level
##
## Fit: aov(formula = friction ~ lubricant, data = data)
##
## $lubricant
##       diff        lwr        upr   p adj
## B-A -0.040 -0.0746294 -0.0053706 0.00272
## C-A -0.076 -0.1106294 -0.0413706 0.00000
## D-A -0.010 -0.0446294  0.0246294 0.81410
## E-A -0.102 -0.1366294 -0.0673706 0.00000
## C-B -0.036 -0.0706294 -0.0013706 0.00719
## D-B  0.030 -0.0046294  0.0646294 0.02968
## E-B -0.062 -0.0966294 -0.0273706 0.00001
## D-C  0.066  0.0313706  0.1006294 0.00001
## E-C -0.026 -0.0606294  0.0086294 0.07215
## E-D -0.092 -0.1266294 -0.0573706 0.00000
```

```r
par(mfrow = c(1,1))
plot(mod1.tky, las = 1)
```




from the printed data, we see that the following pairs have p values > 0.01, therefore A and D are indistinguishable B and D are indistinguishable C and E are indistinguishable

from the difference we see that C has lower friction than A, B, and D and its indistinguishable from E.

therefore if we want lower friction we go with C or E, and for higher friction we go for lubricants A, B, or D

For a non-parametric alternative to the anova we use the Kruskal-Wallis rank sum test it has the same assumptions of the anova except for the normality requirement.

```
kruskal.test(friction ~ lubricant, data = data)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  friction by lubricant
## Kruskal-Wallis chi-squared = 21.4, df = 4, p-value = 0.00026
```

with a p-value of 0.0002605 « 0.01, it leads to the same conclusion as the anova model we did.

Since the objective is to get the lowest amount of friction, we have three major groupings.

-C,E are indistinguishable and they have the lowest friction values.

-A,D are indistinguishable and they have medium friction values.

-B,D are indistinguishable and they have high friction values.

This interaction happens because although D is indistinguishable from A and B, A and B ARE distinguishable from each other.

Thankfully since the objective is to find the lubricants with the lowest friction, we can recommend C,E for best results.