Exercises for Day 1 – Solution
Applied Statistics & Statistical methods for SCIENCE

Anton Rask Lundborg

November 2023

Exercise 1.1 Significance vs. importance

Problem

In the lecture we proposed the following four main questions[1] to be answered by the statistical analysis of a dataset:

1. Is there an effect?

2. Where is the effect?

3. What is the effect?

4. Can the conclusions be trusted?

The founder of modern statistics R. A. Fisher once wrote:

> "It is the magnitude of treatment differences that is of primary importance, not their statistical significance"

Which of the four questions listed above are concerned with *significance* and with *magnitude* respectively? Do you agree with Fisher?

Solution

Questions 1 is clearly concerned with significance and the same can be said to an extent about Question 2. Question 3 concerns magnitude while Question 4 is again about significance.

One could argue that large treatment differences that are insignificant are also not of primary importance. It seems that one needs to care about magnitude and significance simultaneously.

Exercise 1.2 Datasets, variables and observations

Problem

Often data is organized in tables in a laboratory diary or in an Excel sheet. Below you see four examples from the biosciences. For each of the following four examples, discuss these questions, and summarize your conclusions in a *Table-of-Variables*:

(a) How many observations have been made?

(b) What are the variables in the experiment?

(c) What are the variable types (nominal, ordinal, interval, ratio)?

(d) What do you think is the relevant question to be answered by the statistical analysis?

---

[1]In some situations the word *"effect"* should be replaced by *"association"* in these questions.

(e) Which variable would you use as the response?

Data example 1: In an experiment concerning the effect of antibiotic and vitamin additives on growth 12 rats were given two different levels of antibiotics and two different levels of vitamins in their diet, and their growth was measured over some time period. The following table shows the measurements for all 12 rats.

Level of antibiotic	Level of vitamin 0	Level of vitamin 5
0	1.30 1.19 1.08	1.26 1.21 1.19
40	1.05 1.00 1.05	1.52 1.56 1.55

Hint: There are three variables in this example.

Data example 2: An experiment made by Anders Juel Møller (KVL) compared two chilling methods (tunnel-chilling and fast-chilling) of pork. 24 pigs were sampled from two pH groups (high and low pH). After slaughtering the 24 pigs were divided into two halves. One half was tunnel-chilled, the other fast-chilled. After some time the tenderness of the 48 pieces of meat was measured. The measurements are displayed in the following table.

Pig	pH	Tunnel	Fast
1	low	7.22	5.56
2	low	3.11	3.33
3	low	7.44	7.00
4	low	4.33	4.89
5	low	6.78	6.56
6	low	5.56	5.67
7	low	7.33	6.33
8	low	4.22	5.67
9	low	3.89	4.00
10	low	5.78	5.56
11	low	6.44	5.67
12	low	8.00	5.33
13	high	8.44	8.44
14	high	7.11	6.00
15	high	6.00	5.78
16	high	7.56	7.67
17	high	5.11	4.56
18	high	8.67	8.00
19	high	5.78	7.67
20	high	6.11	5.67
21	high	7.44	7.56
22	high	7.67	6.11
23	high	8.00	8.22
24	high	8.78	8.44

Data example 3: 20 people participated in an experiment comparing the difference between two different diets. By randomization 10 people were assigned to each diet and every week a weight gain or weight loss was observed. The observations are the number of weeks where the diet resulted in a weight loss for each of the 20 people in the experiment. The table below displays the results for a period of eight weeks showing the number of people for each combination of diet and weeks with weight loss.

	Weeks with weight loss 0	1	2	3	4	5	6	7	8
Diet 1	1	0	2	0	1	1	2	0	3
Diet 2	2	1	0	1	2	1	2	1	0

Hint: The observations are perhaps not what they seem at first sight. How many observations are there here?

Data example 4: In an experiment concerning the influence of stress on metabolism in rats the regulation of 96 genes were measured using the qPCR method. A total of 47 rats were allocated to 8 groups as shown in the following table.

Group	1	2	3	4	5	6	7	8
Number of rats	6	5	6	6	6	6	6	6
Sex	male	male	male	male	female	female	female	female
Stabling	single	single	group	group	single	single	group	group
Food additive	no	yes	no	yes	no	yes	no	yes

In each group the average gene regulation was measured on a logarithmic scale. The following table shows the measurements for 8 genes.

	Group							
Gene	1	2	3	4	5	6	7	8
Abcb1b	5.554	4.49	4.85	5.076	7.416	6.684	7.524	6.894
Abcb1	5.334	5.55	5.53	4.656	3.456	3.134	3.894	3.004
Abcb4	1.134	1.19	1.51	1.406	1.916	1.454	2.054	1.684
Abcc1	8.114	8.01	8.86	8.466	8.316	7.104	7.884	6.644
Abp1	8.224	8.68	9.24	8.676	11.406	8.504	10.604	8.214
Adh1	−2.996	−3.38	−2.92	−3.214	−3.964	−4.216	−3.766	−4.416
Adh4	2.944	3.10	3.24	3.786	2.456	2.474	2.154	2.494
Ahr	3.624	3.62	4.56	4.976	3.136	3.334	3.014	3.294

The experiment was conducted by Tina Vicky Alstrup Hansen (UCPH-LIFE).

Solution

1. We have $N = 12$ and have the following table of variables:

Variable	**Type**	**Range**	**Usage**
`antibiotic`	Nominal	0, 40	fixed effect
`vitamin`	Nominal	0, 5	fixed effect
`growth`	Continuous	[1.00 ; 1.56]	response

   The most important question of this dataset is how the level of vitamin and/or antibiotic changes the growth of the rats.

2. There are two possible options here. If we consider modelling the difference between the two methods, we have $N = 24$ and the following table:

Variable	**Type**	**Range**	**Usage**
`pH.group`	Nominal	low, high	fixed effect
`Tunnel`	Continuous	[3.11 ; 8.78]	response
`Fast`	Continuous	[3.33 ; 8.44]	response

   Alternatively, we can include the method as a predictor but then correct for the pig using a random effect (we will discuss this in detail later in the course). In this case, we have $N = 48$ and the following table:

Variable	**Type**	**Range**	**Usage**
`Pig`	Nominal	24 levels	random effect
`pH.group`	Nominal	low, high	fixed effect
`method`	Nominal	tunnel, fast	fixed effect
`tenderness`	Continuous	[3.11 ; 8.44]	response

   The primary scientific question is to determine which of the two chilling methods produce the most tender meat and perhaps how this is affected by the pH of the pig.

3. We again have two possible tables here. If we consider the weight loss as a binary response variable, then we have $N = 160$ observations for different combinations of person, diet and week. We have the following table:

Variable	Type	Range	Usage
diet	Nominal	2 levels	fixed effect
week	Ordinal	$1 < \ldots < 8$	fixed effect
person	Nominal	20 levels	random effect
weight.loss	Binary	no, yes	response

Alternatively, we can model the $N = 20$ counts with just the diet as a fixed effect. This results in the table below:

Variable	Type	Range	Usage
diet	Nominal	2 levels	fixed effect
weeks.with.weight.loss	Count	$0 < \ldots < 8$	response

The primary scientific question is to determine which of the diets are most effective at weight loss.

4. We have observations of 8 groups for 96 genes resulting in $N = 8 \cdot 96 = 768$. We then have the following table:

Variable	Type	Range	Usage
number.of.rats	Continuous	5,6	weight
group	Nominal	8 levels	random effect
sex	Nominal	male, female	fixed effect
stabling	Nominal	no, yes	fixed effect
food.additive	Nominal	no, yes	fixed effect
gene	Nominal	96 levels	fixed effect
expression	Continuous	[–5.7060 ; 13.2240]	response

Note that the variable number.of.rats is used as weight to correct for the fact that we can be more certain of the average gene regulation for groups with 6 rats than groups of 5. The primary scientific question is whether the gene expression is affected by stabling or food additive perhaps with the possibility that this is different across sexes or genes.

Exercise 1.5   Hypertension in diabetic patients

Problem

Before commencing on the statistical methods we introduce yet another R technicality. So far we have seen data encoded in text-files, Excel sheets, and R scripts. But of course R also has a format for saving data, namely in RData-files[2]. If RStudio is open, then you may read RData files using the "open file'' icon in the *Environment* window, or by using the load() function from the *Console*. If RStudio is not open, then you may open RStudio together with the RData file by double clicking on the file (in Windows).

The data for this exercise is available in the file hypertension.RData, and also in an Excel sheet (just in case you need it, which you should not).

An experiment on 19 diabetic patients was conducted in order to compare the effects of two drugs called *Drug E* and *Drug N* on the treatment of high blood pressure. The experiment is a cross-over study. This means that all patients try both drugs in two different study periods. Both study periods lasted for 14 days. In between the two study periods was a wash-out period, which also lasted for 14 days. The patients were randomly assigned to two groups called *E/N* and *N/E*. The patients in the E/N-group received drug E in the first study period and drug N in the second study period. The patients in the N/E-group received drug N in the first study period and drug E in the second study period.

The systolic and the diastolic blood pressure was measured for all the patients at the beginning and the end of both study periods. In this exercise we will only analyse the observations of the systolic blood pressure. These observations are shown in the table on the next page. %The observations for the diastolic blood pressure may be found on the internet (search on the reference given at the end of this exercise).

---

[2]We have already worked with RData-files in Exercise 1.3.

Patient id	Treatment order	Systolic blood pressure			
		Baseline 1	End 1	Baseline 2	End 2
9	Drug E, Drug N	124	136	120	145
21	Drug E, Drug N	120	132	138	126
8	Drug E, Drug N	115	96	111	91
12	Drug E, Drug N	134	118	123	123
16	Drug E, Drug N	131	106	111	123
19	Drug E, Drug N	119	108	113	112
20	Drug E, Drug N	124	112	108	112
24	Drug E, Drug N	127	113	121	143
13	Drug N, Drug E	113	113	107	97
17	Drug N, Drug E	132	109	122	119
18	Drug N, Drug E	129	133	139	130
23	Drug N, Drug E	124	120	127	118
25	Drug N, Drug E	112	103	112	121
10	Drug N, Drug E	124	112	128	122
11	Drug N, Drug E	144	154	156	137
14	Drug N, Drug E	134	118	122	109
15	Drug N, Drug E	119	118	115	114
22	Drug N, Drug E	123	123	114	108
26	Drug N, Drug E	122	123	124	120

The R dataset `hypertension.RData` contains the dataset. Beside the raw observations encoded in the variables `patient`, `order`, `baseline1`, `end1`, `baseline2` and `end2` five new variables called `change1`, `change2`, `average`, `diff` and `E_diff_N` have been defined.

- The variable `change1` contains the change of blood pressure over study period 1.

- The variable `change2` contains the change of blood pressure over study period 2.

- The variable `average` contains the average change of the blood pressure over both study periods.

- The variable `diff` contains the difference of the changes of blood pressure between study period 1 and study period 2.

- The variable `E_diff_N` contains the difference of the changes of the blood pressure between the study periods given drug E and drug N.

To analyze the dataset for the cross-over study the following four t-tests may be performed:

- Two sample t-test comparing `E_diff_N` in the E/N-group against the N/E-group.

- Two sample t-test comparing `average` in the E/N-group against the N/E-group.

- Two sample t-test comparing `diff` in the E/N-group against the N/E-group.

- One sample t-test comparing `E_diff_N` against the mean value 0.

Two of these t-tests do the actual comparison between the effects of drug E and drug N. These tests, however, are only valid when the following two problems do not occur:

(i) A spill-over (also called a carry-over) from study period 1 to study period 2. A possible explanation for such an effect is that the drug given in study period 1 still has an effect in study period 2.

(ii) An interaction between the effects of the drugs and the study periods. For instance that the effect of drug E for some strange reason is larger in study period 1 than in study period 2.

The two remaining t-tests are done to validate that these two problems have not occurred.

(a) Which of the four t-tests listed above do the drug comparison, and which t-tests validates against problem 1 and 2?

Help to get started: If the drugs have different effects and if there is a spill-over from period 1 to period 2, then the difference between the changes in the E- and the N-period will depend on the order the

drugs were given.

(b) Perform the relevant t-tests. Remember to validate the underlying normality assumption before you make the t-tests. What is the conclusion from these tests?

Remark: Using all these t-tests for the statistical analysis would be uncommon. Instead, the analysis is usually done using a random effect model. We will return to this on course day 5.

Reference: Bradstreet, T.E. (1994) "Favorite Data Sets from Early Phases of Drug Research - Part 3.'' *Proceedings of the Section on Statistical Education of the American Statistical Association.*

Solution

Usually a cross-over experiment would be analyzed using a random effects model (Day 5 of the course). However, it is also possible to analyze this dataset using t-tests. In the exercise text 4 t-tests are proposed.
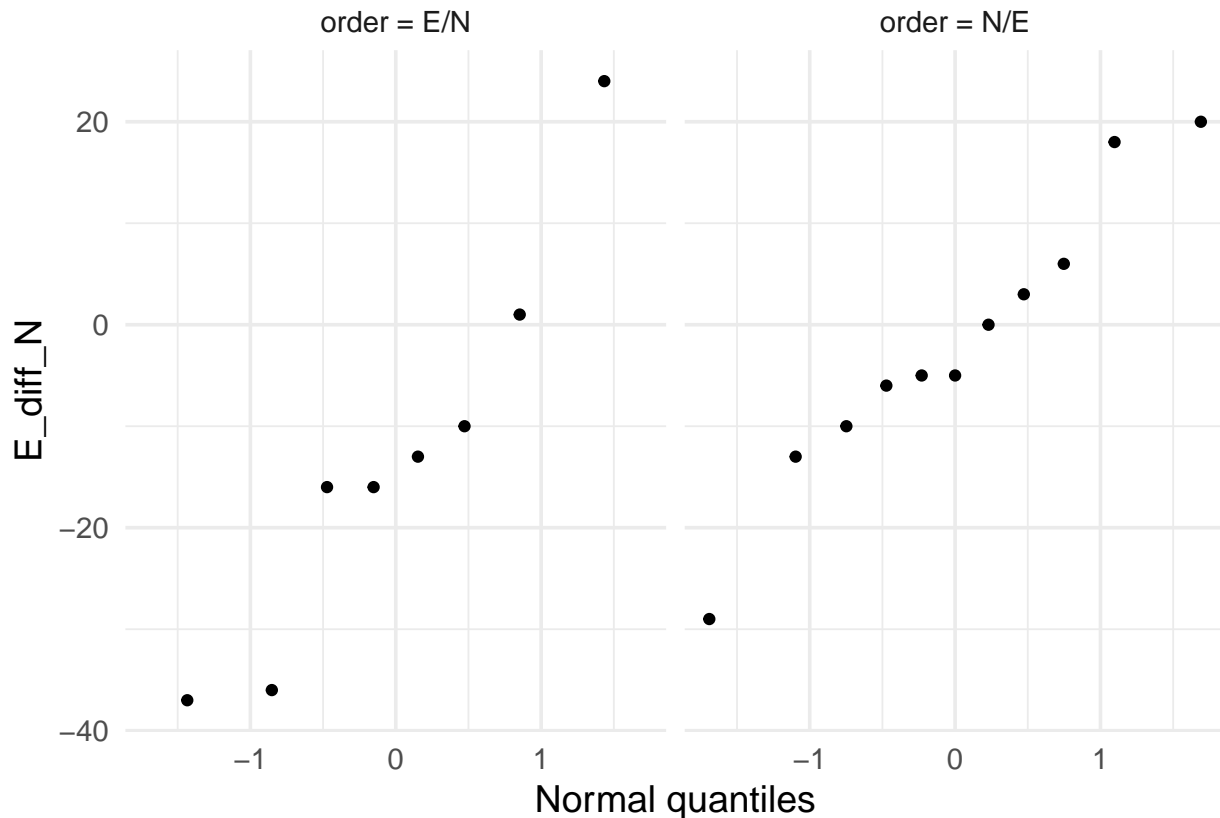
First we read the data (available in an R dataset) and quickly summarize it:

```
load("hypertension.RData")
summary(hypertension)
```

```
##     patient      order      baseline1          end1          baseline2
##  Min.   : 8.0   E/N: 8   Min.   :112.0   Min.   : 96.0   Min.   :107.0
##  1st Qu.:12.5   N/E:11   1st Qu.:119.5   1st Qu.:110.5   1st Qu.:112.5
##  Median :17.0            Median :124.0   Median :118.0   Median :121.0
##  Mean   :17.0            Mean   :124.7   Mean   :118.3   Mean   :121.6
##  3rd Qu.:21.5            3rd Qu.:130.0   3rd Qu.:123.0   3rd Qu.:125.5
##  Max.   :26.0            Max.   :144.0   Max.   :154.0   Max.   :156.0
##      end2           change1           change2          average
##  Min.   : 91.0   Min.   :-25.000   Min.   :-20.000   Min.   :-19.500
##  1st Qu.:112.0   1st Qu.:-15.000   1st Qu.: -9.500   1st Qu.: -7.250
##  Median :120.0   Median : -9.000   Median : -4.000   Median : -4.500
##  Mean   :119.5   Mean   : -6.474   Mean   : -2.158   Mean   : -4.316
##  3rd Qu.:124.5   3rd Qu.:  0.500   3rd Qu.:  2.000   3rd Qu.: -1.250
##  Max.   :145.0   Max.   : 12.000   Max.   : 25.000   Max.   : 18.500
##      diff            E_diff_N
##  Min.   :-37.000   Min.   :-37.000
##  1st Qu.:-16.000   1st Qu.:-14.500
##  Median : -3.000   Median : -6.000
##  Mean   : -4.316   Mean   : -6.526
##  3rd Qu.:  5.500   3rd Qu.:  2.000
##  Max.   : 29.000   Max.   : 24.000
```

The t-test comparing `E_diff_N` in the two groups will test Problem (i), that is, it investigates whether there is a spill-over effect. If the drugs have different effects and if there is a spill-over from period 1 to period 2, then the differences between the groups will depend on the order of the groups. Therefore, when there is no spill-over, we would expect the order to be constant, which we test with the t-test. To investigate whether it is reasonable to perform the t-test, we do a normal QQ-plot for both orders:

```
library(ggplot2)
ggplot(hypertension, aes(sample = E_diff_N)) +
  stat_qq() + scale_y_continuous(name = "E_diff_N") +
  scale_x_continuous("Normal quantiles") +
  facet_grid(. ~ order, labeller = as_labeller(c(
    "E/N" = "order = E/N",
    "N/E" = "order = N/E"
  )))
```

In both plots, it looks like the points are roughly on a line so the normality assumption is okay-ish. We now perform a Welch t-test (which does not assume equal variances):

```
t.test(E_diff_N ~ order, data = hypertension)
```
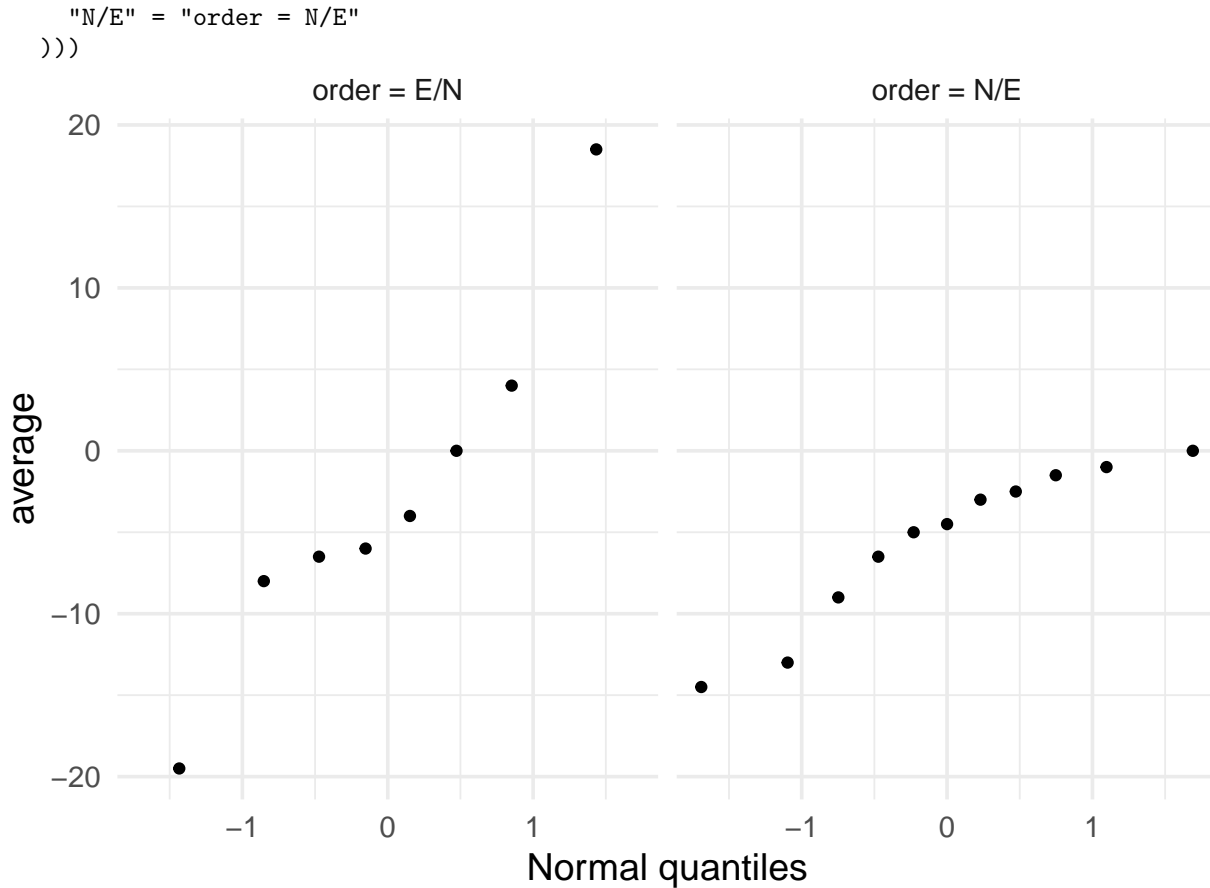
```
##
##  Welch Two Sample t-test
##
## data:  E_diff_N by order
## t = -1.3533, df = 11.911, p-value = 0.2011
## alternative hypothesis: true difference in means between group E/N and group N/E is not equal to 0
## 95 percent confidence interval:
##  -28.63542    6.70360
## sample estimates:
## mean in group E/N mean in group N/E
##          -12.875000         -1.909091
```

The test is non-significant ($p = 0.2011$) so we cannot reject the null hypothesis of no spill-over.

which yields a p-value of 0.1067. The test remains non-significant.

The t-test comparing `average` in the two gorups will test Problem (ii), that is, it investigates whether there is an interaction between study period and drug type. If the drugs have different effects for different study periods, then the average response will depend on the order. There there is no interaction, we would expect the average to be the same for both orders, which test with the t-test. We can again check normality:

```
ggplot(hypertension, aes(sample = average)) +
  stat_qq() + scale_y_continuous(name = "average") +
  scale_x_continuous("Normal quantiles") +
  facet_grid(. ~ order, labeller = as_labeller(c(
    "E/N" = "order = E/N",
```

```
   "N/E" = "order = N/E"
)))
```


The normality assumption is perhaps not entirely reasonable here. We still perform a Welch t-test (which does not assume equal variances):

```
t.test(average ~ order, data = hypertension)
```
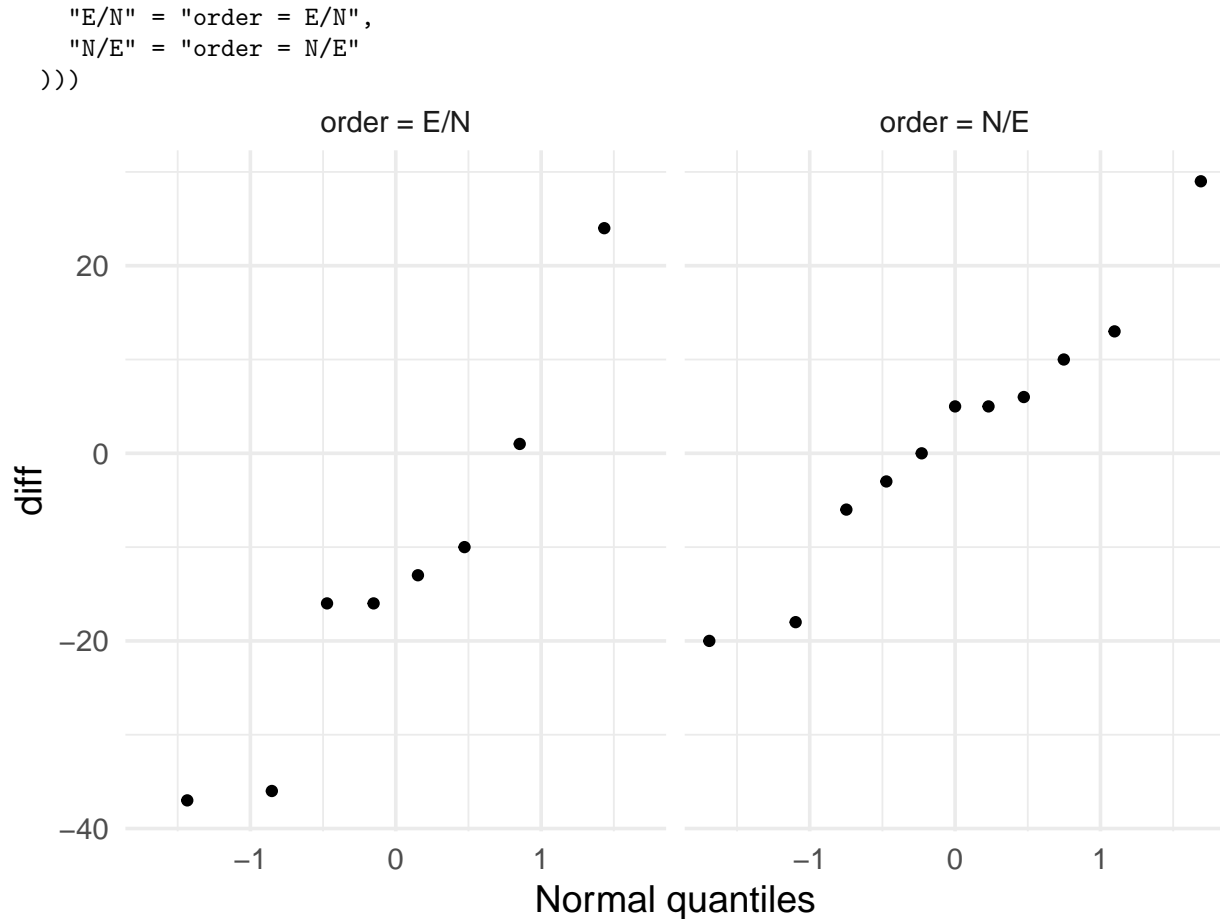
```
##
##  Welch Two Sample t-test
##
## data:  average by order
## t = 0.68005, df = 8.9977, p-value = 0.5136
## alternative hypothesis: true difference in means between group E/N and group N/E is not equal to 0
## 95 percent confidence interval:
##   -6.543544 12.168544
## sample estimates:
## mean in group E/N mean in group N/E
##           -2.6875            -5.5000
```

The test is non-significant ($p = 0.5136$) so we cannot reject the null hypothesis of no interaction. We therefore accept this hypothesis.

Finally, we can test the difference between the effects of the two drugs with either of the remaining t-tests. We can check the difference between the study periods for each of the two orders. If this difference depends on the order, then there is a difference between using drugs E and N. We first check normality:

```
ggplot(hypertension, aes(sample = diff)) +
  stat_qq() + scale_y_continuous(name = "diff") +
  scale_x_continuous("Normal quantiles") +
  facet_grid(. ~ order, labeller = as_labeller(c(
```

```
    "E/N" = "order = E/N",
    "N/E" = "order = N/E"
)))
```


Again the normality assumption is perhaps not entirely reasonable here. We perform a *t*-test:

```
t.test(diff ~ order, data = hypertension)

##
##  Welch Two Sample t-test
##
## data:  diff by order
## t = -1.8245, df = 11.911, p-value = 0.09324
## alternative hypothesis: true difference in means between group E/N and group N/E is not equal to 0
## 95 percent confidence interval:
##  -32.453600   2.885418
## sample estimates:
## mean in group E/N mean in group N/E
##        -12.875000          1.909091

wilcox.test(diff ~ order, data = hypertension)

## Warning in wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...): cannot
## compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  diff by order
## W = 23, p-value = 0.09022
## alternative hypothesis: true location shift is not equal to 0
```
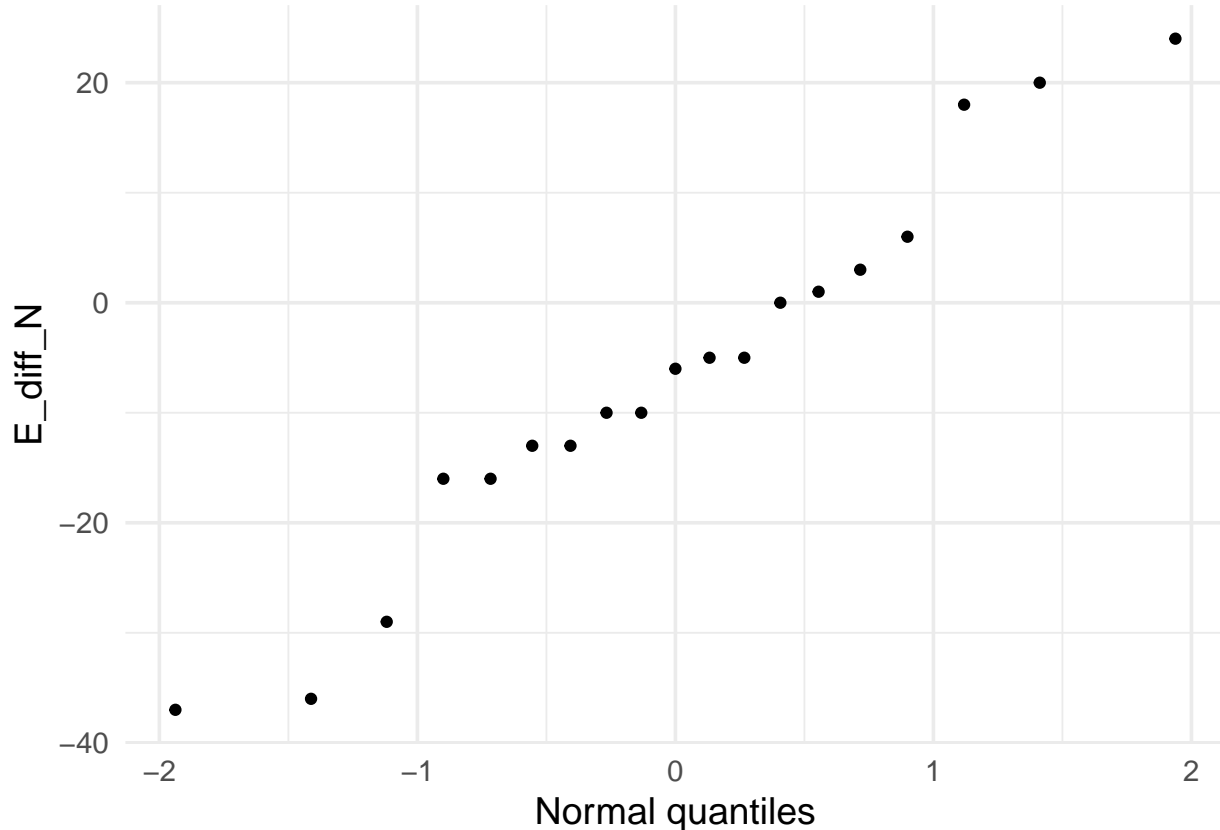
9

The test is non-significant ($p = 0.09324$) so we cannot reject the null hypothesis of no effect using this test. Alternatively, we could have done the one-sample t-test. We check normality:

```
ggplot(hypertension, aes(sample = E_diff_N)) +
  stat_qq() + scale_y_continuous(name = "E_diff_N") +
  scale_x_continuous("Normal quantiles")
```


It looks like the points are roughly on a line so the normality assumption is reasonable. We now perform a one-sample t-test:

```
t.test(E_diff_N ~ 1, data = hypertension)
```

```
##
##   One Sample t-test
##
## data:  E_diff_N
## t = -1.6771, df = 18, p-value = 0.1108
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -14.70171    1.64908
## sample estimates:
## mean of x
## -6.526316
```

The test is non-significant ($p = 0.1108$) so we cannot reject the null hypothesis of no effect using this test either. Note that the p-values differ in the two tests but the conclusions are roughly similar.