Solutions exercises of Chapter 7

Exercise 1

a. These are paired samples: each pair of half plates will have about the same level of corrosion, so the result of polishing by the two brands of polish are most probably dependent. But the series of 23 differences in the grading of the results may be considered independent: a random sample from the distribution of the difference in polishing results.

b. No, the differences in grading are integers (between -4 and +5) and cannot be assumed to be approximately normal since only 10 different numbers occurred.

c. We will apply the sign test on the median, using the following 23 differences in grading $A - B$:

+5, +1, +1, 0, +3, -4, +5, +5, +1, -1, -1, +5, +2, +1, +1, +1, -2, +4, +2, +2, +2, +4, +1

There are 4 negative differences one 0-difference (which we will cancel) and 18 positive differences.
Testing procedure:

1. X = "the number of positive differences among the non-zero differences in grading $A - B$" $\sim B(22, p)$
2. Test $H_0: p = \frac{1}{2}$ against $H_1: p \neq \frac{1}{2}$ with $\alpha_0 = 0.05$
3. Test statistic is X
4. X has under H_0 a $B\left(22, \frac{1}{2}\right)$-distribution (which is approximately $N\left(22 \cdot \frac{1}{2}, 22 \cdot \frac{1}{4}\right)$, since $n \geq 15$).
5. Observed: $X = 18$
6. It is a two-sided test: reject H_0 if the p-value $\leq \alpha_0 = 0.05$ and
   Since X has under H_0 a symmetric distribution the p-value can be computed as follows:

$$p - value = 2 \cdot P(X \geq 18 | H_0) \overset{\text{c.c.}}{=} 2 \cdot P(X \geq 17.5) \approx 2\left[1 - \Phi\left(\frac{17.5 - 11}{\sqrt{5.5}}\right)\right] \approx 2[1 - \Phi(2.77)] = 0.56\%$$

7. The p-value < $\alpha_0 = 0.05$, so reject H_o.
8. At 5% level of significance it is shown that the results of the grading of polish A and B are different.

(*Note that we only answered the research question "is there a difference?" Yes! If an additional question would "Which is the best", the answer would be A, of course since the grades are higher. We did not test this statement. If we would test it, it is a one-sided test!*)

Additional remarks 1c:

1. Exact calculation of the p-value with the binomial distribution results in:

$$\text{p-value} = 2 \cdot P(X \geq 18 | H_0) = 2 \cdot \left[\binom{22}{18} + \binom{22}{19} + \binom{22}{20} + \binom{22}{21} + 1\right] \cdot \left(\frac{1}{2}\right)^{22} \approx 2 \cdot 0.00217 \approx 0.43\%$$

2. Determining the RR = $\{0, 1, \ldots, c\} \cup \{22 - c, .., 21, 22\}$, which is symmetric about $\mu = 11$:

$$P(X \leq c | H_0) \leq \frac{\alpha_0}{2} = 0.025, \text{ so with cont. corr. } \Phi\left(\frac{c + \frac{1}{2} - 11}{\sqrt{5.5}}\right) \leq 0.025, \text{ or } \frac{c + \frac{1}{2} - 11}{\sqrt{5.5}} \leq -1.96,$$

or: $c \leq 10.5 - 1.96 \cdot \sqrt{5.5} \approx 5.90$, so the RR can be defined as $X \leq 5$ and $X \geq 17$.

Ecercise 2

a. A two samples binomial z-test for large samples can be applied since both the samples are large (> 40).
In this case we cannot apply a t-test, since the distributions are not-normal. Another indication that the normal distribution does not apply (approximately) is that in the first sample: $\overline{x} \approx s$: waiting times are always positive but in a normal distribution with $\mu = \sigma$ the probability of negative values is high.
(verify that for a $N(\mu, \sigma^2)$-distribution with $\mu = \sigma$ we have $P(X \leq 0) \approx 16\%$)

b. We will conduct a two independent samples z-test for large samples on the difference of expected values:

1. The waiting times $X_1, \ldots, X_{150}$ (last year) and $Y_1, \ldots, Y_{150}$ (this year) are independent. Both samples are drawn from unknown distributions with expectations $\mu_1$resp. μ_2 and variances σ_1^2 resp. σ_2^2.
2. Test H_0: $\mu_1 = \mu_2$ against H_1: $\mu_1 > \mu_2$ with $\alpha = 5\%$.
3. $Z = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$
4. Z has under H_0 approximately a $N(0, 1)$distribution.
5. Observed: $z = \dfrac{200 - 164}{\sqrt{\frac{1}{150}(180^2 + 100^2)}} \approx 2.14$

6. Reject H_0 if $Z \geq c$, with $c = 1.645$ from the N(0, 1)-table such that $P(Z \geq c) = 0.05$.
7. The observed value 2.14 of Z lies in the rejection region, so reject H_0.
8. At a 5% level of significance the data showed that the waiting times this year are overall shorter than last year.

c. The p-value is $P(Z \geq 2.14) = 1 - \Phi(2.14) = 1.62\%$
Interpretation: at a level of significance $\alpha = 1.62\%$ the null hypothesis will be just rejected. (no rejection if $\alpha < 1.62\%$, rejection if $\alpha \geq 1.62\%$)

Exercise 3

a. We will first use Shapiro-Wilk's table ($n = 20$) to determine the coefficients a_i of the ordered mean temperatures $X_{(1)}, \dots, X_{(20)}$:

i	$X_{(i)}$	a_i	i	$X_{(i)}$	a_i
1	13.7	-0.4734	11	16.3	0.0140
2	14.1	-0.3211	12	17.0	0.0422
3	14.2	-0.2565	13	17.0	0.0711
4	14.9	-0.2085	14	17.2	0.1013
5	15.3	-0.1686	15	17.2	0.1334
6	15.4	-0.1334	16	17.8	0.1686
7	15.4	-0.1013	17	18.1	0.2085
8	15.7	-0.0711	18	18.9	0.2565
9	15.8	-0.0422	19	19.3	0.3211
10	15.9	-0.0140	20	20.1	0.4734

Using a simple calculator we find the summary:
$n = 20$, $\overline{x} = 16.465$ and $s \approx 1.763$ ($s^2 = 3.111$)

$$\sum_{i=1}^{20} a_i X_{(i)} = -0.4734 \cdot 13.7 + \dots + 0.4734 \cdot 20.1 \approx 7.56486$$

So $W = \frac{\left(\sum_{i=1}^{20} a_i X_{(i)}\right)^2}{\sum_{i=1}^{20}(X_i - \overline{X})^2} = \frac{7.56486^2}{19 \cdot 3.111} \approx 0.968$

1. The mean temperatures $X_1, \dots, X_{20}$ are independent and all have the same, but unknown distribution function F.
2. We test $H_0: F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ against $H_1: F(x) \neq \Phi\left(\frac{x-\mu}{\sigma}\right)$ with $\alpha = 5\%$
3. $W = \frac{\left(\sum_{i=1}^{20} a_i X_{(i)}\right)^2}{\sum_{i=1}^{20}(X_i - \overline{X})^2}$, where a_i are taken from the Shapiro-Wilk coefficients table with $n = 20$.
4. W has under H_0 a distribution according the Shapiro-Wilk-table.
5. $W = 0.968$.
6. The Shapiro-Wilk test is left-sided: reject H_0 if $W \leq c$ with $c = 0.905$ (table with $n = 20$ and $\alpha = 5\%$)
7. $W = 0.968$ does not fall in the Rejection Region, so we fail to reject H_0.
8. There is not sufficient statistical evidence to claim that the mean temperatures in July do not have a normal distribution, at a 5% level of significance.
   (*Note, again, that the test* ***does not prove normality****: on the other hand the data provided no evidence that the assumption of a normal distribution is incorrect)*

<u>Check with SPSS</u> (via ***Analyze → Descriptive Statistics → Explore → Plots****: choose "Normality plots with tests" to make SPSS conduct Shapiro Wilk's test):*
The results by SPSS and above agree, since SPSS reports a p-value $P(W \leq 0.968|H_0) = 71.5\% > \alpha$, so we fail to reject H_0.
The skewness coefficient and adapted Kurtosis (Kurtosis -3), that SPSS reports does not deviate much from the normal reference values 0.

Report

Mean temperature in July

Mean	N	Std. Deviation	Kurtosis	Skewness
16,465	20	1,7638	-,460	,409

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Mean temperature in July	,126	20	,200*	,968	20	,715

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

b. 1. Model: the July temperatures $X_1, \ldots, X_{20}$ are independent and $N(\mu, \sigma^2)$-distributed with unknown μ and σ^2

2. Test $H_0: \mu = 17.4$ against $H_1: \mu < 17.4$ with α = 0.05.

3. $T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{\overline{X} - 17.4}{S/\sqrt{20}}$

4. T has a t_{19}-distribution, if H_0 is true.

5. Observed: $\overline{x} = 16.465$ and $s \approx 1.764$, so $t = \frac{16.465 - 17.4}{1.764/\sqrt{20}} \approx -2.370$

6. Lower-tailed test: if $T \leq c$, then reject H_0.
For α = 0.05 we have $c = -1.729$, since $P(T_{19} \geq 1.729) = 0.05 = \alpha$

7. $-2.370 = t < c = -1.729$, so reject H_0.

8. At a 5% significance level we can state that is the expected July temperature in De Bilt is lower than in Maastricht.

c. Now we do not assume a normal distribution for the July temperatures: we will conduct the sign test on the median. The de null hypothesis states that the **median** of the July temperatures is 17.4 °C. If the distribution is symmetrical (which seems a reasonable assumption here) the median equals the expected July temperature.

1. X = “the number of July temperatures larger than 17.4 °C in the random sample of 20 July temperatures”. $X \sim B(20, p)$, where p is the unknown probability of a July temperature > 17.4 °C.
2. Test $H_0: p = \frac{1}{2}$ against $H_1: p < \frac{1}{2}$ with α = 5%
3. Test statistic: X
4. X has under H_0 a $B(20, 0.5)$-distribution (which is given in the binomial table)
5. $X = 5$
6. We will reject H_0 if the p-value ≤ α. The p-value= $P(X \leq 5 | p = 0.5) = 0.021$
7. $0.021 < 0.05$, so we reject H_0.
8. At a 5% significance level we can state that is the expected July temperature in De Bilt is lower than in Maastricht.

Exercise 4

a. We determine first the 25th and the 75th percentiles: $Q_1 = x_{(10)} = 141$ and $Q_3 = x_{(29)}$ = 154,
since 25% (75%) of $n = 38$ is 9.5 (28.5).

The InterQuartile Range IQR = 154 – 141 = 13

Potential outliers can be found outside the interval (Q_1 - 1.5× IQR, Q_3 + 1.5× IQR) = (121.5 , 173.5): there is one (potential) outlier: **121**

b. 1. The numerical summary: the (sample) skewness coefficient = -0.41 which deviates a little from 0, the reference value 0 of the normal distribution. The kurtosis = 3.09 is close to the normal reference value 3.

2. The histogram shows roughly the same shape as the normal distribution, so symmetrical and mound shaped, though there seem to be two peeks.

3. The dots in the **normal** Q-Q plot follow the straight line $y = x$ quite well, confirming a normal distribution.

The overall conclusion is that the assumption of a normal distribution is reasonable: it is not contradicted by any of the three aspects.

c. Se the Shapiro –Wilk`s table of coefficients for $n = 38$ and $i = 3$: $\boldsymbol{a_3 = -0.2391}$

d. 5. Observed value: W = 0.980.

6. Left sided test: reject H_0 (normal distribution) if $\boldsymbol{W} \leq$ **0.947**
(see Shapiro-Wilk`s table for n = 38 and α = 10%, so the left “tail probability” = α = 0.10)

7. W = 0.980 does **not** fall in the Rejection Region (0.980 > 0.947), so we fail to reject H_0.

8. At a 5% significance level a deviation of the normal distribution as a model for the covered distances of fully charged Nissan Leaf’s is not proven.

(*The overall conclusion in b. is confirmed in this test).*

Exerecise 5

a. The two independent samples t-test with equal variances assumed:

[*1. Model assumptions (“statistical assumptions”):*

We have two (small) independent samples at hand: a random sample of $N(\mu_1, \sigma^2)$-distributed crop quantities for variety A and a random sample of $N(\mu_1, \sigma^2)$-distributed crop quantities for variety B (equal σ's!)
Shorter: the crop quantities $X_1, \ldots, X_9, Y_1, \ldots, Y_{11}$ are independent $X_i \sim N(\mu_1, \sigma^2)$ and $Y_j \sim N(\mu_2, \sigma^2)$)
2. We test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ with α = 5%
3. Test statistic $T = \frac{\overline{X}-\overline{Y}}{\sqrt{S^2\left(\frac{1}{9}+\frac{1}{11}\right)}}$, with $S^2 = \frac{8S_1^2+10S_2^2}{9+11-2}$
4. T is under H_0 t-distributed with $df = n_1 + n_2 - 2 = 18$
5. Observed: $s^2 = \frac{8\times 2.598^2 + 10\times 3.286^2}{18} \approx 9.00$, so $t = \frac{35.0-39.0}{\sqrt{9.00\left(\frac{1}{9}+\frac{1}{11}\right)}} = -2.97$.
*6. The test is two-tailed: **reject H_0 if $T \leq -c$ or $T \geq c$***
where $c = 2.101$ from the t_{18}-table
7. $t = -2.97$ lies in the rejection region, so reject H_0.
8. The crop quantities do differ significantly at a 5% level.]

b. For these two relatively small (<40), independent samples we can apply Wilcoxon's rank sum test.

c. First we determine the rank sum of the X-values (variety A) in the ordered sequence of all $9 + 11 = 20$ observations.

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Order	31.6	31.9	33.0	34.1	34.3	35.3	35.6	35.7	36.0	37.2	37.6	37.8	38.4	38.8	39.0	39.4	40.1	42.1	42.8	45.9
$\boldsymbol{R(X_{(i)})}$	**1**	**2**	**3**		**5**	**6**		**8**	**9**	**10**							**17**			

So $W = \sum_{i=1}^{9} R(X_i) = 1 + 2 + 3 + 5 + 6 + 8 + 9 + 10 + 17 = 61$

1. The crop quantities $X_1, \ldots, X_9$ of variety A and $Y_1, \ldots, Y_{12}$ of variety B are independent and are independent random samples of the unknown (continuous) distributions of X and Y.
2. We will test $H_0: f_X(x) = f_Y(x)$ against $H_1: f_X(x) = f_Y(x + a), a \neq 0$, with $\alpha = 1\%$
3. $W = \sum_{i=1}^{9} R(X_i)$
4. W is under H_0 approximately ($n_1 > 5$ and $n_2 > 5$) normal with $\mu_W = 9 \cdot \frac{1+20}{2} = 94.5$
and $\sigma_W = \sqrt{\frac{1}{12} n_1 n_2 (N+1)} = \sqrt{\frac{1}{12} \cdot 9 \cdot 11 \cdot 21} \approx 13.16$
5. Observed: $W = 61$.
6. Two-sided (!) test: if the p-value ≤ 1%, we will reject H_0.
The p-value $= 2 \cdot P(W \leq 61 | H_0) \overset{\text{cont. corr.}}{\approx} 2 \cdot \Phi\left(\frac{61.5 - 94.5}{13.16}\right) \approx 2\Phi(-2.51) = 1.20\%$
(*note that we computed the left-sided p-value $P(W \leq 61|H_0)$* since $W = 61 < \mu_W = 94.5$)
7. The p-value is greater than α = 1%, so we fail to reject H_0.
8. At a 1% level of significance we do not have sufficient statistical evidence to claim that the crop quantities of variety A and variety B differ.

*Check with SPSS (**Analyze → Non-parametric tests → Legacy dialogs → 2 independent samples** and choose Mann-Whitney).*
The Z-score deviates (-2,545 versus -2.51) a little because of the application of continuity correction.
SPSS calculates the exact (not normally approximated) p-value as well: see for this method example 7.5.1 with exact calculation in section 7.5.

Test Statistics[a]

	Crop quantity
Mann-Whitney U	16,000
Wilcoxon W	61,000
Z	-2,545
Asymp. Sig. (2-tailed)	,011
Exact Sig. [2*(1-tailed Sig.)]	,010[b]

a. Grouping Variable: Variety

b. Not corrected for ties.

Ranks

	Varieteit	N	Mean Rank	Sum of Ranks
Crop Quantity	varieteit A	9	6,78	61,00
	varieteit B	11	13,55	149,00
	Total	20		

Exercise 6

First we will compute the sample means and variances for the two sets of observations (*preferably use a simple scientific calculator to find* $\overline{x}$ *and s for both sets*):

i	1	2	3	4	Mean	Stand. dev.	Variance
x_i	3.09	4.67	7.72	6.89	5.5925	2.11	4.44
y_i	4.56	4.44	9.29	8.01	6.575	2.45	6.02

Two independent random samples, drawn from normal distributions:

a. 1. Model fort he observations: $X_1, \ldots, X_4, Y_1, \ldots, Y_4$ are independent and $X_i \sim N(\mu_X, \sigma_X^2)$ and $Y_i \sim N(\mu_Y, \sigma_Y^2)$

2. Test $H_0: \sigma_X^2 = \sigma_Y^2$ against $H_1: \sigma_X^2 \neq \sigma_Y^2$ with $\alpha_0 = 0.10$

3./4. Tets statistic $F = \frac{S_X^2}{S_Y^2}$ has under H_0 a F_3^3-distribution.

5. Observed: $F = \left(\frac{2.11}{2.45}\right)^2 \approx 0.74$

6. Reject H_0 if $F \leq c_1$ or $F \geq c_2$.
From the F_3^3-table with tail probability 5%: $c_2 = 9.28$ and $c_1 = \frac{1}{9.28}$

7. $F = 0.74$ is not included in the Rejection Region, so we fail to reject H_0.

8. At a 10% significance level we were not able to show that the variances are different.

b. 1. Model as in a., but now with extra assumption: $\sigma_X^2 = \sigma_Y^2$ (2 samples t-test on the difference of the μ`s)

2. Test H_0: $\mu_X = \mu_Y$ versus H_1: $\mu_X < \mu_Y$, with $\alpha_0 = 10.$,

3. Toetsingsgrootheid $T = \frac{\overline{X} - \overline{Y}}{\sqrt{S^2\left(\frac{1}{4} + \frac{1}{4}\right)}}$ met $S^2 = \frac{S_X^2 + S_Y^2}{2}$

4. T is under H_0 t_{4+4-2}-distributed.

5. Observed: $t = \frac{5.5925 - 6.575}{\sqrt{\frac{2.11^2 + 2.45^2}{2} \cdot \left(\frac{1}{4} + \frac{1}{4}\right)}} \approx -0.61$ $\quad (s^2 \approx 5.23)$

6. This is a lower-tailed test: reject H_0 if the lower-tailed p-value (linker) <= α_0

7. The lower-tailed p-value $P(T_6 < -0.61) = P(T_6 > 0.61) > 10\% = \alpha_0$,so we cannot reject H_0.

8. The X-scores are not on average less than the Y-scores, at a 10% significance level.

Two **independent random samples, drawn from** not-normal distributions:

c. Wilcoxon`s rank sum test *(below we will choose the rank sum of the* $y_i$*`s. The rank sum of the* x_i*`s would result in the same conclusion, but the test is in that case lower-tailed: we will reject for sufficiently small ranks of the* $x_i$*`s.)*

1. Model for the observations $X_1, \ldots, X_4, Y_1, \ldots, Y_4$ are independent and the X_i`$s$ have an unknown density function $f_X$ and the $Y_i$`s an unknown density f_Y.
2. We will test $H_0: f_Y(x) = f_X(x)$ against $H_1: f_Y(x-a) \leq f_X(x)$ where $a > 0$, with $\alpha_0 = 5\%$
3. W = “the rank sum of the Y-observations” = $\sum_{i=1}^{4} R(Y_i)$
4. Under H_0 the distribution of W is given by:

$$P(W = w) = \frac{\text{\# combinations of 4 ranks with total } w}{\binom{8}{4}}$$

5.

Rank	1	2	3	4	5	6	7	8	
Order statitics	3.09	4.44	4.56	4.67	6.89	7.72	8.01	9.29	Total
$R(Y_i)$		2	3				7	8	W = 20

6. Reject H_0 if $W \geq c$

   Determine the RR: each combination of 4 ranks occurs with probability $\frac{1}{\binom{8}{4}} = \frac{1}{70}$

   What is the minimal value of the total rank sum c, such that $P(W \geq c|H_0) \leq \alpha_0 = 5\%$?

   - One combination jields the largest rank sum: $5 + 6 + 7 + 8 = 26$.
   - $25 = 4 + 6 + 7 + 8$, so $P(W \geq 25|H_0) = \frac{2}{70} = 2.9\%$
   - $24 = 3 + 6 + 7 + 8 = 4 + 5 + 7 + 8$, so $P(W \geq 24|H_0) = \frac{4}{70} \approx 5.7\%$

   Conclusion: the Rejection Region is $W \geq 25$.

7. Since $W = 20$ is not included in the Rejection Region, we fail to reject H_0.
8. At a 5% significance level we could not statistically prove that the values of Y are structurally higher than X.