Samkedemia
EDUCATION

ULTIMATE SOLUTION BOOK:
ALL EXERCISES & PAST EXAMS

Research Methods 3

KU LEUVEN • BACHELOR OF BUSINESS ADMINISTRATION

2027 EDITION

LET ME INTRODUCE MYSELF.

About Me

I'll make sure you pass this exam. I know exactly what it takes to get there.

BACKGROUND

- My name is **Jovan Samardzic**, I am 26 years old
- Got my **MSc in Statistics at KU Leuven**, graduated **Magna Cum Laude**
- Recipient of the **Science@Leuven Scholarship**

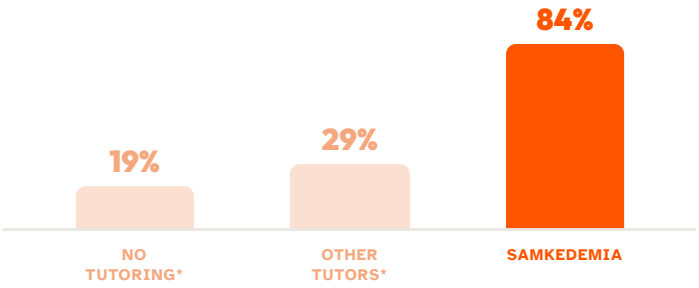
STYLE

- Founder of **Samkedemia**, helping students pass their exams
- Worked with **200+ university students** from all kinds of backgrounds
- **Plain language, not textbook theory** - just what actually makes sense

PROOF THAT IT WORKS.

Results

PASSING RATES



* BASED ON INTERNAL RESEARCH CONDUCTED IN 2026.

BY THE NUMBERS

1300+	200+	+82%	99%
TUTORING HOURS	PASSED THEIR EXAMS	AVG. GRADE IMPROVEMENT	SATISFACTION SCORE

CONTACT.

Let's Pass That Exam!

+32 490 10 08 21 · samkedemia · jovan@samkedemia.com · samkedemia.com

PLEASE READ BEFORE USING THIS GUIDE.

Copyright & Usage Policy

This document is a proprietary educational resource developed by **Samkedemia**.

It is intended for use as a supplemental learning aid for registered students and customers who have purchased it directly from the official source.

UNAUTHORIZED USE IS STRICTLY PROHIBITED

- Uploading this file on any public or private digital platform or cloud drive.
- Sharing via WhatsApp, Discord servers, or social media.
- Reproducing, copying, or reselling the material without explicit consent.

Violation of these terms constitutes an infringement of copyright law and will result in legal action.

OFFICIAL DISTRIBUTOR

JOVAN SAMARDZIC PR BEOGRAD (STARI GRAD)

PIB: 115342864 MB: 68292042

+32490100821 samkedemia jovan@samkedemia.com samkedemia.com

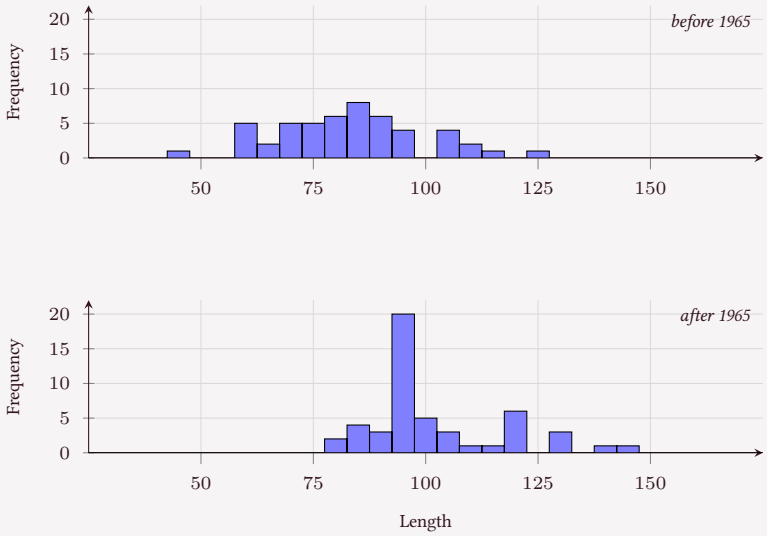
Contents

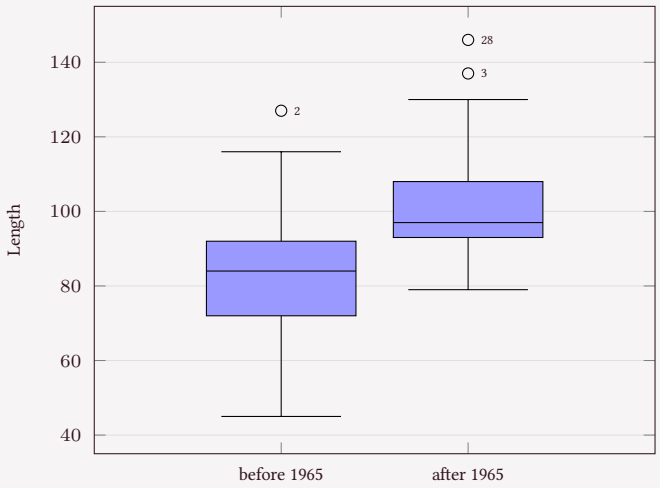
Chapter 1 - Hypothesis Testing

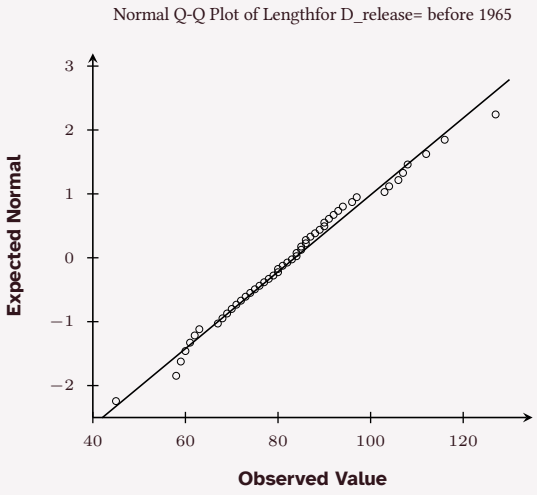
Question 1

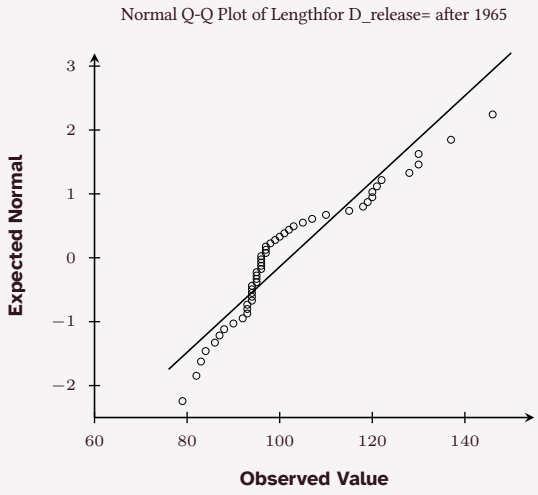
We consider data on a simple random sample of 100 movies (see file **movies.sav**). In a first analysis we want to test whether movies released before 1965 are on average longer than movies released after 1965. The variable **length** measures the length of a movie in minutes. The variable **D_release** indicates whether a movie was released before or after 1965.

Analysis with SPSS yields the following results:









Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Length	before 1965	50	83.64	16.624	2.351
	after 1965	50	102.10	14.941	2.113

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Length	Equal variances assumed	.351	.555	-5.840	98	.000	-18.460	3.161	-24.733	-12.187
	Equal variances not assumed			-5.840	96.904	.000	-18.460	3.161	-24.734	-12.186

a. Test whether movies released after 1965 are longer than movies released before 1965.

b. Discuss whether the assumptions of the independent-samples t-test are satisfied.

c. Explain what you can conclude from the "95% confidence interval of the difference" in the output of the independent samples t-test.

d. Indicate which test could be used to test whether the median length of movies released after 1965 is larger than the median length of movies released before 1965. Discuss whether the assumptions of this test are satisfied.

Tip: Independent-Samples t-test

The **independent-samples *t*-test** is a parametric test used to compare the population means of two independent samples.

$$H_0 : \mu_1 = \mu_2 \qquad H_A : \mu_1 \neq \mu_2$$

It comes with two assumptions:

- The two samples are independent (different subjects in each group, not the same subjects measured twice).
- $n_1, n_2 > 30$, so the CLT ensures the test statistic has an approximate t-distribution in each group. Otherwise, the variable must be Normally distributed within each group for the test statistic to have an exact t-distribution.

This course does not derive the test statistic by hand - it is only ever read off SPSS output.

Tip: Levene's Test

Levene's test checks whether the population variances of two or more groups are equal - it is used to check an assumption of the independent-samples t-test (and of ANOVA), not as a test of interest on its own.

$$H_0 : \sigma_1^2 = \sigma_2^2 \qquad H_A : \sigma_1^2 \neq \sigma_2^2$$

Like the t-test, this course only ever reads F and $Sig.$ off SPSS output: if $Sig. \geq \alpha$, we do not reject equal variances; if $Sig. < \alpha$, we reject equal variances.

Solution 1: a. Testing whether movies released after 1965 are longer

We define the two variables as follows:

$$X_1 = \text{length (in minutes) of a movie released before 1965}$$

$$X_2 = \text{length (in minutes) of a movie released after 1965}$$

Step 1: Define the hypotheses.

We test whether the average length is larger for movies released after 1965:

$$H_0 : \mu_1 \geq \mu_2$$

$$H_A : \mu_1 < \mu_2$$

$$\alpha = 0.05$$

Step 2: State the estimator.

We use the independent-samples t-test. Before reading the test value, we first check Levene's test for equality of variances, reported in the same Independent Samples Test table:

$$F = 0.351$$

$$Sig. = 0.555$$

Since $Sig. \geq \alpha$, hence we do not reject equal variances, and read the "Equal variances assumed"

row in the Independent Samples Test table.

Step 3: Find the test value.
As mentioned above, we look at the "Equal variances assumed" row of the Independent Samples Test table. In the columns t and df, we see:

$$t = -5.840$$

$$df = 98$$

Step 4: Find the p-value.
SPSS reports a two-tailed significance in the same row:

$$Sig.\ (\text{2-tailed}) = 0.000$$

The mean difference between group 1 and group 2 is negative, so the test value t is negative as well, which matches the direction that H_A claims. We therefore divide the reported significance by two to obtain the one-tailed p-value:

$$p = \frac{0.000}{2} \approx 0.000$$

Step 5: Make the decision.
Since $p < \alpha$, we reject H_0.

Answer: We conclude with 95% confidence that movies released after 1965 are, on average, longer than movies released before 1965.

Solution 1: b. Discuss whether the assumptions of the test are satisfied

The independent-samples t-test makes two assumptions.

Step 1: Check independence.
The before-1965 and after-1965 groups consist of different movies, not the same movies measured twice, so the two samples are independent. This assumption is satisfied.

Step 2: Check Normality/CLT.
Both groups have $n = 50 > 30$, so the CLT ensures the test statistic has an approximate t-distribution regardless of the shape of the population distribution. This assumption is satisfied. The boxplot does show a few outliers in each group (case 2 before 1965; cases 3 and 28 after 1965), but with $n = 50$ in each group the CLT still applies.

Answer: Both assumptions of the independent-samples t-test are satisfied.

Solution 1: c. Interpreting the 95% confidence interval of the difference

We read the 95% confidence interval of the difference from the same "Equal variances assumed" row of the Independent Samples Test table, in the "Lower" and "Upper" columns under 95% Confidence Interval of the Difference:

$$[-24.733,\ -12.187]$$

We are 95% confident that the true population mean length of movies released before 1965 is, on average, between 24.733 and 12.187 minutes *shorter* than movies released after 1965. Since the entire interval lies below 0, this is consistent with the conclusion in part a: movies released after 1965 are on average longer.

Answer: With 95% confidence, movies released before 1965 are on average between 12.187 and 24.733 minutes shorter than movies released after 1965.

Tip: Mann-Whitney-Wilcoxon Test

The **Mann-Whitney-Wilcoxon test** is a non-parametric test used to compare the population medians of two independent samples, when the data is ordinal or non-normal quantitative.

$$H_0 : \text{median}_1 = \text{median}_2 \qquad H_A : \text{median}_1 > / < / \neq \text{median}_2$$

It comes with four assumptions:

- The variable is quantitative or ordinal.
- The two samples are independent.
- The two groups' distributions have the same shape (they can only differ by a shift in location).
- $n_1 + n_2 \geq 20$.

We prefer it over the independent-samples t-test when Normality/CLT fails, or the median is more meaningful than the mean.

Solution 1: d. Testing the median, and its assumptions

Step 1: Choose the test.
Movies released before 1965 and movies released after 1965 are two independent samples of different movies, not the same movies measured twice, and length is a quantitative variable. We are asked about the median, not the mean, so we use the non-parametric Mann-Whitney-Wilcoxon test to compare the two groups' medians, instead of the independent-samples t-test.

Step 2: Check the assumptions.

The Mann-Whitney-Wilcoxon test makes four assumptions:

Step 2.1: Variable is quantitative.

We are measuring length, which is a quantitative variable, so this assumption is satisfied.

Step 2.2: Samples are independent.

Movies released before 1965 and movies released after 1965 are different movies, not the same movies measured twice, so the two samples are independent. This assumption is satisfied.

Step 2.3: Same shape.

Looking at the two histograms, both distributions are unimodal and roughly bell-shaped, though the after-1965 group has a taller, narrower peak. Both boxplots also show outliers (case 2 before 1965; cases 3 and 28 after 1965). The two distributions look reasonably similar in shape, so this assumption is roughly satisfied.

Step 2.4: Sample size.

$n_1 + n_2 = 50 + 50 = 100 \geq 20$, so this assumption is satisfied.

Answer: We would use the Mann-Whitney-Wilcoxon test; all of its assumptions are satisfied (or roughly satisfied).

Question 2

The file **movies.sav** also contains the rating (ranging from 1=very bad to 4=very good) of each movie from a certain critic. Some movie directors claim that this critic is unfair in that he assigns a negative rating (=rating lower than or equal to 2.5) to more than 65% of the movies.

The frequency distribution of movie ratings of the critic equals:

Rating

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.0	8	8.0	8.0	8.0
	1.5	14	14.0	14.0	22.0
	2.0	19	19.0	19.0	41.0
	2.5	27	27.0	27.0	68.0
	3.0	25	25.0	25.0	93.0
	3.5	6	6.0	6.0	99.0
	4.0	1	1.0	1.0	100.0
Total		100	100.0	100.0	

a. Test whether more than 65% of the movies in the population receive a rating from the critic lower than or equal to 2.5.

b. Discuss whether the assumptions of the test are satisfied.

Tip: Binomial Test

The **binomial test** is a non-parametric test used to test a population proportion π, for one category of a two-category variable in a single sample, against a hypothesized value π_0.

$$H_0 : \pi = \pi_0 \qquad H_A : \pi \neq \pi_0$$

It comes with three assumptions:

- Each observation is a Bernoulli trial: a yes/no outcome with the same probability π every

time.

- The number of trials n is fixed in advance.
- The observations are independent of each other.

The estimator we use has an exact binomial distribution under H_0:

$$X \sim \text{Binom}(n,\ \pi_0) = \text{number of observations (out of } n\text{) in the category of interest}$$

Solution 2: a. Testing whether more than 65% of movies receive a negative rating

Step 1: Define the hypotheses.

Let π denote the population proportion of movies that receive a rating of 2.5 or lower from this critic. The movie directors claim this is more than 65%, so we test:

$$H_0 : \pi \leq 0.65$$

$$H_A : \pi > 0.65$$

$$\alpha = 0.05$$

Step 2: State the estimator.

We use the binomial test, with the estimator and its distribution under H_0:

$$X \sim \text{Binom}(n,\ \pi_0) = \text{number of movies (out of } n\text{) that receive a rating of 2.5 or lower}$$

Step 3: Find the test value.

Here $n = 100$ movies were sampled, and $\pi_0 = 0.65$ from H_0. In the Rating table, we read the Cumulative Percent column at the row for a rating of 2.5, which shows that 68.0% of the 100 movies received a rating of 2.5 or lower:

$$x = 68$$

Step 4: Find the p-value.

Since H_A says the proportion is higher than 0.65, we need the probability of a value at least as large as x:

$$p = P(X \geq 68 \mid H_0) = 1 - \texttt{binomcdf}(100,\ 0.65,\ 67) \approx 0.3029$$

Step 5: Make the decision.

Since $p \geq \alpha$, we cannot reject H_0.

Answer: We cannot conclude, with 95% confidence, that more than 65% of the movies in the population receive a rating of 2.5 or lower from this critic - the population proportion could be 65% or lower.

Solution 2: b. Discuss whether the assumptions of the test are satisfied

The binomial test makes three assumptions.

Step 1: Fixed probability.

Each movie's rating is a yes/no outcome (rating $\leq$ 2.5 or not), with the same probability π across movies, since every movie is rated by the same critic on the same scale. This assumption is satisfied.

Step 2: Fixed number of trials.

The number of movies sampled, $n = 100$, was fixed in advance, before observing the ratings. This assumption is satisfied.

Step 3: Independent observations.

As stated at the start of this exercise, the 100 movies are a simple random sample, so the observations are independent of each other. This assumption is satisfied.

Answer: All three assumptions of the binomial test are satisfied.

Question 3

The following cross-table shows for a random sample ($n = 183$) how many males and females would spend more than 500 euro when they buy a new smartphone.

gender * D_budget Crosstabulation

Count

		D_budget		
		less than 500 euro	more than 500 euro	Total
gender	male	46	29	75
	female	57	51	108
Total		103	80	183

a. Test whether the proportion of males who spend more than 500 euro on a new smartphone differs from the proportion of females who spend more than 500 euro.

b. Discuss whether the assumptions of the test are satisfied.

Question 4

The following cross-table shows for a random sample of 130 persons the relation between net household income and how much they would spend when they buy a new smartphone.

net household income * budget Crosstabulation

		budget					
		between 0 and 200 euro	between 200 and 500 euro	between 500 and 700 euro	between 700 and 900 euro	more than 900 euro	Total
less than 1000	Count	3	14	10	8	0	35
	Expected Count	7.8	12.1	6.5	5.7	3.0	35.0
between 1000 and 1500	Count	7	5	2	4	3	21
	Expected Count	4.7	7.3	3.9	3.4	1.8	21.0
between 1500 and 2000	Count	7	12	5	5	3	32
	Expected Count	7.1	11.1	5.9	5.2	2.7	32.0
between 2000 and 3000	Count	10	10	6	2	1	29
	Expected Count	6.5	10.0	5.4	4.7	2.5	29.0
more than 3000	Count	2	4	1	2	4	13
	Expected Count	2.9	4.5	2.4	2.1	1.1	13.0
Total	Count	29	45	24	21	11	130
	Expected Count	29.0	45.0	24.0	21.0	11.0	130.0

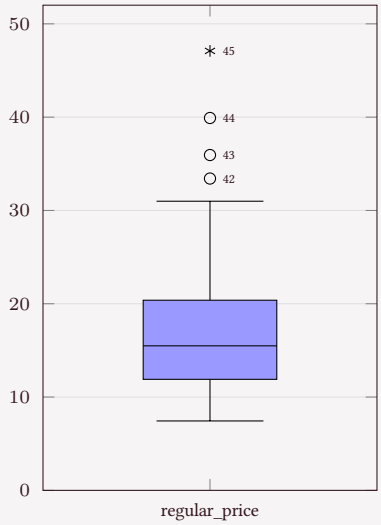
a. Test whether the variables **net household income** and **budget** are statistically dependent.

b. Discuss whether the assumptions of the test are satisfied, and indicate how you can proceed when the assumptions of the test are not satisfied.

Question 5

The file **price_wine.sav** contains the regular price and the promotion price for a random sample of 45 wines in a specialized wine store.

The following graph shows the distribution of the regular price in the sample:



We also compute descriptive statistics and the frequency distribution for the variable **regular_price**:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
regular_price	45	7.44	47.10	17.7298	8.97385
Valid N (listwise)	45				

regular_price

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	7.44	1	2.2	2.2	2.2
	7.81	1	2.2	2.2	4.4
	7.98	1	2.2	2.2	6.7
	8.18	1	2.2	2.2	8.9
	9.30	2	4.4	4.4	13.3
	9.92	1	2.2	2.2	15.6
	10.04	1	2.2	2.2	17.8
	10.29	1	2.2	2.2	20.0
	11.03	1	2.2	2.2	22.2
	11.53	1	2.2	2.2	24.4
	11.90	3	6.7	6.7	31.1
	12.39	2	4.4	4.4	35.6
	13.01	1	2.2	2.2	37.8
	13.14	1	2.2	2.2	40.0
	14.25	1	2.2	2.2	42.2
	14.58	1	2.2	2.2	44.4
	14.87	1	2.2	2.2	46.7
	15.49	2	4.4	4.4	51.1
	16.11	1	2.2	2.2	53.3
	16.36	1	2.2	2.2	55.6
	16.86	1	2.2	2.2	57.8
	17.10	1	2.2	2.2	60.0
	17.35	1	2.2	2.2	62.2
	17.48	1	2.2	2.2	64.4
	17.72	1	2.2	2.2	66.7
	17.97	1	2.2	2.2	68.9
	18.59	1	2.2	2.2	71.1
	20.20	1	2.2	2.2	73.3
	20.38	1	2.2	2.2	75.6
	22.19	1	2.2	2.2	77.8
	22.68	1	2.2	2.2	80.0
	22.93	1	2.2	2.2	82.2
	23.55	1	2.2	2.2	84.4
	28.51	1	2.2	2.2	86.7
	30.37	1	2.2	2.2	88.9
	30.99	1	2.2	2.2	91.1
	33.42	1	2.2	2.2	93.3
	35.94	1	2.2	2.2	95.6
	39.91	1	2.2	2.2	97.8
	47.10	1	2.2	2.2	100.0
Total		45	100.0	100.0	

a. Use a one-sample t-test to test whether the average regular price of wines in the store is larger than 14 euro. Discuss whether the assumptions of the test are satisfied.

b. Use a sign test to test whether the median price of regular wines in the store is larger than 14 euro. Discuss whether the assumptions of the test are satisfied.

c. Use a signed-rank test to test whether the median price of regular wines in the store is larger than 14 euro. Make the necessary computations in excel or SPSS. Discuss whether the assumptions of the test are satisfied.

Question 6

The file **price_wine.sav** contains both the regular price and the promotion price for a sample of 45 wines from a specialized wine store. The frequency distribution of the regular price minus the promotion price is displayed in the table below:

regular price minus promotion price

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0.73	1	2.2	2.2	2.2
	0.86	1	2.2	2.2	4.4
	0.95	1	2.2	2.2	6.7
	1.00	2	4.4	4.4	11.1
	1.03	1	2.2	2.2	13.3
	1.14	1	2.2	2.2	15.6
	1.15	2	4.4	4.4	20.0
	1.18	1	2.2	2.2	22.2
	1.19	1	2.2	2.2	24.4
	1.24	1	2.2	2.2	26.7
	1.30	1	2.2	2.2	28.9
	1.31	1	2.2	2.2	31.1
	1.34	1	2.2	2.2	33.3
	1.36	1	2.2	2.2	35.6
	1.37	1	2.2	2.2	37.8
	1.39	1	2.2	2.2	40.0
	1.40	1	2.2	2.2	42.2
	1.44	1	2.2	2.2	44.4
	1.48	1	2.2	2.2	46.7
	1.50	2	4.4	4.4	51.1
	1.53	1	2.2	2.2	53.3
	1.56	1	2.2	2.2	55.6
	1.59	1	2.2	2.2	57.8
	1.60	1	2.2	2.2	60.0
	1.64	1	2.2	2.2	62.2
	1.64	1	2.2	2.2	64.4
	1.65	1	2.2	2.2	66.7
	1.66	2	4.4	4.4	71.1
	1.68	1	2.2	2.2	73.3
	1.70	1	2.2	2.2	75.6
	1.73	1	2.2	2.2	77.8
	1.81	1	2.2	2.2	80.0
	1.86	1	2.2	2.2	82.2
	1.98	1	2.2	2.2	84.4
	2.39	1	2.2	2.2	86.7
	2.43	1	2.2	2.2	88.9
	2.44	1	2.2	2.2	91.1
	2.49	1	2.2	2.2	93.3
	2.93	1	2.2	2.2	95.6
	3.87	1	2.2	2.2	97.8
	3.94	1	2.2	2.2	100.0
Total		45	100.0	100.0	

Conduct a sign test to test whether the median price reduction in the store for a wine in promo-

tion is larger than 1 euro. Compute the test statistic of the sign test from the information in the frequency distribution of regular price minus promotion price. Discuss whether the assumptions of the test are satisfied.

Question 7

The file **price_wine.sav** contains both the regular price and the promotion price for a sample of 45 wines from a specialized wine store. Let the variable **X=regular_price-promotion_price** be the price reduction for a wine due to a promotion.

We conduct a signed-rank test in SPSS to test whether the median price reduction in the store for a wine in promotion is larger than 1 euro. The output of the analysis shows that two observations are discarded from the analysis because they have X=1. Applying the signed-rank test to the remaining 43 observations indicates that the sum of negative ranks is equal to 15.5.

a. Compute the sum of positive ranks and use the signed rank test to test whether the median price reduction in the store for a wine in promotion is larger than 1 euro.

b. Discuss whether the assumptions of the test are satisfied.

Question 8

20 students take an intensive course in Spanish. To assess the effect of the course, students take a test in Spanish before and after the course.

Let **X** be the test score after taking the course minus the test score before taking the course. The table below shows descriptive statistics for the variable X.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
score after minus score before course	20	-5.00	7.00	1.4500	3.20321
Valid N (listwise)	20				

The following table shows the frequency distribution of X:

score after minus score before course

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	-5.00	1	5.0	5.0	5.0
	-4.00	1	5.0	5.0	10.0
	-3.00	1	5.0	5.0	15.0
	-2.00	1	5.0	5.0	20.0
	-1.00	2	10.0	10.0	30.0
	1.00	2	10.0	10.0	40.0
	2.00	3	15.0	15.0	55.0
	3.00	4	20.0	20.0	75.0
	4.00	2	10.0	10.0	85.0
	5.00	2	10.0	10.0	95.0
	7.00	1	5.0	5.0	100.0
Total		20	100.0	100.0	

a. Use a parametric test to test whether the average test score of students is higher after the course than before the course. Discuss whether the assumptions of the test are satisfied.

b. Use a sign-test to test whether the median change in the test score (due to taking the course) is positive. Discuss whether the assumptions of the test are satisfied.

c. Use a signed-rank test to test whether the median change in the test score (due to taking

the course) is positive (use SPSS or excel to compute the test statistic). Discuss whether the assumptions of the test are satisfied.

Question 9

We want to test whether 14 boys and 13 girls differ in the time they need to solve the exam of statistical modelling. The data below show the order in which boys (M) and girls (F) handed in the exam.

gender	rank
F	1
F	2
M	3
F	4
F	5
M	6
M	7
F	8
F	9
F	10
M	11
F	12
M	13
F	14
F	15
F	16
M	17
M	18
F	19
M	20
F	21
M	22
M	23
M	24
M	25
M	26
M	27

a. Suppose analysis of the data indicates that the distribution of the solution time has the same shape for boys and girls. Use a MWW test to test whether the median time to solve the exam is larger for boys than for girls.

b. Discuss whether the assumptions of the test are satisfied.

Question 10

On 1/1/2017 the proportion of inhabitants of Flanders with a specific civil status is as follows:

Civil status	Population Proportion
Not married	.476
Married	.374
widowed	.060
Divorced	.090

In a random sample of 400 inhabitants of the population we obtain the following counts:

Civil status	Observed count in sample
Not married	175
Married	150
widowed	33
Divorced	42

a. Use an appropriate statistical test to investigate whether the sample is representative with respect to civil status.

b. Discuss whether the assumptions of the test are satisfied.

Question 11

A sample of $n = 12$ male NBA basketball players have the following lengths (in cm):

183, 188, 192, 193, 200, 204, 206, 206, 206, 187, 194, 202

a. Conduct an appropriate statistical test to investigate whether the median length of NBA basketball players is larger than 190 cm. Make sure the assumptions of the test are satisfied.

b. Which test would you use if the sample size was $n = 30$? Justify your answer.

Question 12

In a survey on transportation a random sample of 661 respondents indicate to what extent they have experience with biking (1=no experience, 2=little experience, 3=very experienced). The following cross-table shows the relation between **gender** and the categorical variable **biking_experience**.

gender * biking_experience Crosstabulation

		biking_experience			
		no experience	little experience	very experienced	Total
male	Count	51	30	193	274
	Expected Count	66.7	43.1	A	274.0
female	Count	110	74	203	387
	Expected Count	94.3	60.9	231.8	387.0
Total	Count	161	104	396	661
	Expected Count	161.0	104.0	396.0	661.0

a. Compute the value A in the cross-table.

b. Use an appropriate statistical test to investigate whether the distribution of the categorical variable **biking_experience** is different for males and females.

c. Discuss whether the assumptions of the test are satisfied.

d. Suppose in the sample, the age of the respondents (measured in years) ranges from 16 to 81. Which test can you conduct to test the null hypothesis that, in the population, the average age is the same at the three biking experience levels? Discuss the assumptions that need to be satisfied for this test to be valid.

Question 13

In a survey about ecotourism 110 respondents answered the following question:

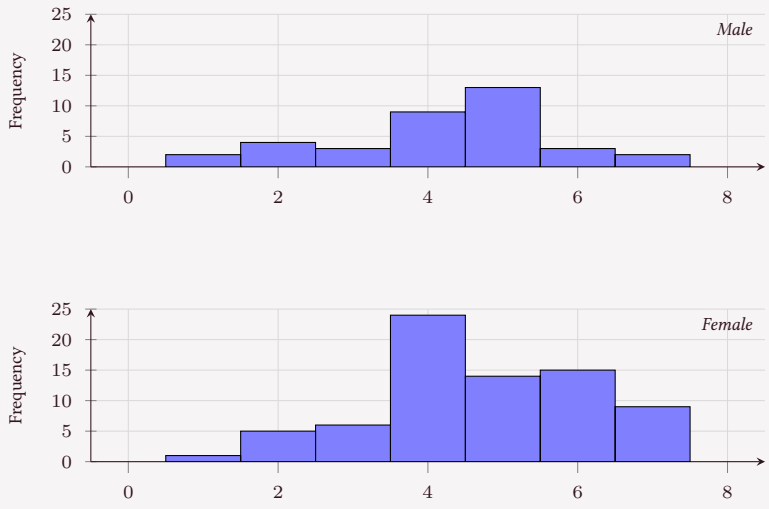
> It is important for me to make conscious consumption choices with respect to sustainability (1=strongly disagree, ..., 7=strongly agree).

We use a Mann-Whitney-Wilcoxon test to compare the median scores of males and females on the question. Analysis with SPSS yields the following results:

Report

It is important for me to make conscious consumption choices with regard to sustainability.

What is your gender?	Mean	N	Std. Deviation	Median
Male	4.22	36	1.495	4.50
Female	4.70	74	1.450	5.00
Total	4.55	110	1.475	5.00



It is important for me to make conscious consumption choices with regard to sustainability.

Ranks

		N	Mean Rank	Sum of Ranks
It is important for me to make conscious consumption choices with regard to sustainability.	Male	36	49.85	1794.50
	Female	74	58.25	4310.50
	Total	110		

a. Test whether the median score on the question is higher for females than for males.

b. Discuss whether the assumptions of the test are satisfied.

Question 14

A factory wants to investigate whether the number of products produced by a machine in a period of an hour is Normally distributed. The variable **production** in the table below indicates the number of products produced for a random sample of 11 hours. The sample mean and sample standard deviation of the variable production equal $\bar{x} = 98.182$ and $s_x = 14.11253$. The table below also shows part of the computations needed for the Kolmogorov-Smirnov (KS) test statistic.

obs	production	EDF	CDF	EDF-CDF	EDF-CDF+
1	80	0.0909	0.0988	-0.0079	-0.1232
2	87	0.2727	0.2141	0.0586	0.0586
3	87	0.2727	0.2141	0.0586	-0.0327
4	91	0.3636	0.3054	0.0582	0.0330
5	92	0.4545	0.3307	0.1239	0.0711
6	94	0.5455	0.3835	0.1620	0.0506
7	98	0.7273	0.4949	0.2324	0.2324
8	98	A	B	C	D
9	108	E	F	G	H
10	119	0.9091	0.9299	-0.0208	-0.0666
11	126	1.0000	0.9756	0.0244	1.0000

a. Compute the numbers in the cells A, B, C, D, E, F, G and H.

b. Compute the value of the KS test statistic.

Question 15

12 students obtain the following grade for the exam of statistical modelling in September.

$$6,\ 7,\ 8,\ 9,\ 9,\ 11,\ 11,\ 11,\ 14,\ 16,\ 17,\ 20$$

Compute the value of the Kolmogorov-Smirnov test statistic for testing whether the grades come from a Normal distribution. Use for instance Excel to perform the calculations.

Question 16

A simple random sample of 250 lecturers who teach statistics indicate which type of software package they prefer, S1, S2, S3 or S4.

100 lecturers prefer S1, 80 lecturers prefer S2, 35 lecturers prefer S3 and 35 lecturers prefer S4.

Can you conclude from the data that lecturers prefer a certain software package more? Use $\alpha = 0.05$. Indicate whether the assumptions of the test are satisfied.

Question 17

A firm aims to evaluate whether the type of sales training affects the sales performance of managers. They randomly assign 10 sales managers to three conditions: 20 managers receive no sales training (=control condition), 10 managers receive sales training A and 10 managers receive sales training B. 6 weeks after the training the sales turnover generated by each manager is gathered.

a. Which kind of test can you conduct to test the global null hypothesis that the average turnover is equal across the three conditions?

b. Which kind of test can you use to test the difference in average sales performance for each pair of conditions?

Question 18

A researcher is conducting an experiment to evaluate the impact of market context on strategic performance in a competitive setting. The researcher recruits 60 subjects from the entering class of MBA students at a large university. Each subject is put in charge of a firm in a market simulation game and is given the objective of strengthening the competitive position of his or her firm.

After 2 hours of play, the subject is given advice about the changing market context in the industry. Each subject is randomly assigned to one of three conditions in a balanced design. In the first control condition subjects receive no advice. In the second condition, the advice suggests to increase competition. In the third condition, the advice suggests to increase cooperation. Each subject then continues to play the game for 2 more hours.

The researcher uses the following measure of strategic performance:

Score_marketshare =

100 $\times$ market share after 4 hours of play / market share after 2 hours of play.

We conduct a one-way ANOVA with as dependent variable **score_marketshare** and as factor **condition**. Part of the SPSS output of the analysis is shown below:

ANOVA

score_marketshare

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10290.700	2			
Within Groups	22982.900	57			
Total	33273.600	59			

Test of Homogeneity of Variances

score_marketshare

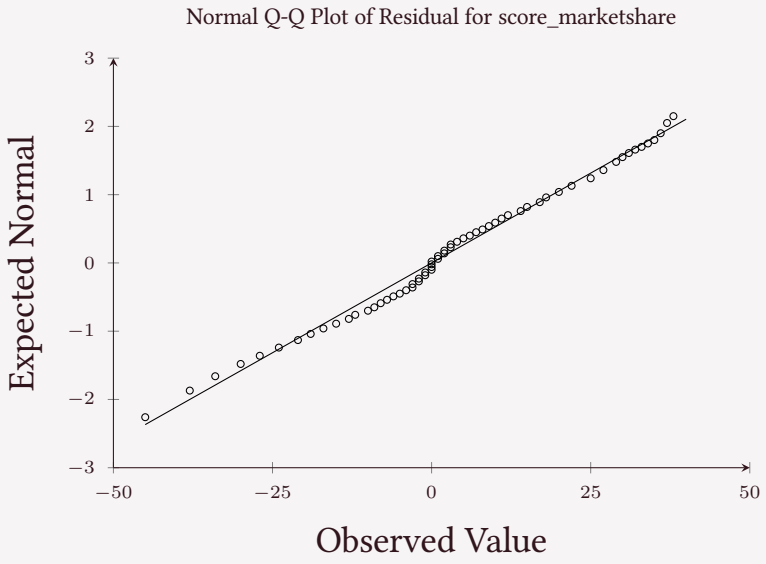
Levene Statistic	df1	df2	Sig.
2.215	2	57	.119

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual for score_marketshare	.075	60	.200*	.974	60	.235

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

We also inspect the Q-Q plot of the residuals:



a. Complete the ANOVA table and test whether that the average market share score differs across the three conditions.

b. Discuss whether the assumptions of the F-test are satisfied.

c. Compute the proportion of the variation in the dependent variable that is due to differences between the three conditions.

Question 19

In a random sample ($n = 1000$), the sample Pearson correlation equals .07. Test whether the population correlation is positive with 99% reliability. Discuss whether the assumptions of the test are satisfied.

Question 20

A HR manager wants to evaluate the effect of the number of training days on the performance of sales managers. The manager has sales performance scores for three groups: employees with 1, 2 or 3 days of training.

We run one-way anova with as dependent variable **sales_performance** and with the **number of training days** as factor. Part of the SPSS output is shown below:

Report

sales performance

number of training days	Mean	N	Std. Deviation
1	63.5798	20	13.50858
2	73.5677	20	10.60901
3	79.2792	20	4.40754
Total	72.1422	60	12.00312

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Residual for score	.123	60	.025	.957	60	.032

a. Lilliefors Significance Correction

Levene's Test of Equality of Error Variances

Dependent Variable: sales performance

F	df1	df2	Sig.
4.637	2	57	.014

a. Design: Intercept + number_of_trainingdays

Multiple Comparisons

Dependent Variable: sales performance Tukey HSD

(I) number of training days	(J) number of training days	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-9.98789*	3.23759	.009	-17.7789	-2.1969
	3	-15.69947*	3.23759	.000	-23.4905	-7.9085
2	1	9.98789*	3.23759	.009	2.1969	17.7789
	3	-5.71158	3.23759	.191	-13.5026	2.0794
3	1	15.69947*	3.23759	.000	7.9085	23.4905
	2	5.71158	3.23759	.191	-2.0794	13.5026

*. The mean difference is significant at the 0.05 level.

a. Use an F-test to investigate whether the number of training days affects the average sales performance of sales managers.

b. Discuss whether the assumptions of the F-test are violated at the 1% significance level.

c. How many training days would you recommend for a future sales training?

Exam: 2022 January

Exam: 2022 August

Exam: 2024 Example 1

Exam: 2024 Example 2

Exam: 2024 January

Exam: 2024 August

Exam: 2025 January

Exam: 2025 August

Exam: 2026 Example

Exam: 2026 January

Exam: Unknown 1

Exam: Unknown 2

Samkedemia
EDUCATION

Did you like this solution guide?

Reach out to get the other ones as well!

Math A **Math B** **Stats 1**
Stats 2 **IMOR** **RM 3**

Let's maximize your results!

Go beyond the solutions with premium tutoring!

samkedemia

KU LEUVEN • BACHELOR OF BUSINESS ADMINISTRATION

2027 EDITION