TABLES, FORMULAS AND EXERCISES WITH KEY

FOR BIOMETRICS

Tadeja KRANER ŠUMENJAK and Vilma SEM

University of Maribor Press

University of Maribor

Faculty of Agriculture
and Life Sciences

TABLES, FORMULAS AND EXERCISES WITH KEY FOR BIOMETRICS

assist. prof. Tadeja Kraner Šumenjak, Ph.D. and assist. Vilma Sem, Ph.D.

Maribor, June 2018

Title: Tables, formulas and exercises with key for biometrics

Authors: assist. prof. Tadeja Kraner Šumenjak, Ph.D.
(University of Maribor, Faculty of Agriculture and Life Sciences)

assist. Vilma Sem, Ph.D.
(University of Maribor, Faculty of Agriculture and Life Sciences)

Review: assoc. prof. Jože Nemec, Ph.D.
(University of Maribor, Faculty of Agriculture and Life Sciences)

assist. prof. Aleksandra Tepeh, Ph.D.
(University of Maribor, Faculty of Electrical Engineering and Computer Science)

Proofreading: Katja Težak (University of Maribor, Faculty of Agriculture and Life Sciences)

Technical editor: sen. lect. Peter Berk, M.S. (University of Maribor, Faculty of Agriculture and Life Sciences) and Jan Perša, M.D.
(University of Maribor Press).

Cover designer: Jan Perša, M.D. (University of Maribor Press)

Other Graphics Authors.

Co-published by / Izdajateljica:
University of Maribor, Faculty of Agriculture and Life Sciences
Pivola 10,2311 Hoče, Slovenia
http://fkbv.um.si, fkbv@um.si

Published by / Založnik:
University of Maribor Press
Slomškov trg 15, 2000 Maribor, Slovenia
http://press.um.si, zalozba@um.si

Edition: 1st

Publication type: e-publication

Available at: http://press.um.si/index.php/ump/catalog/book/336

Published: Maribor, June 2018

© University of Maribor Press
All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publisher.

CIP - Katoložni zapis o publikaciji
Univerzitetna knjižnica Maribor

57.087.1:631.421(075.8)

KRANER Šumenjak, Tadeja
Tables, formulas and exercises with key for biometrics [Elektronski vir] : for biometrics / Tadeja Kraner Šumenjak and Vilma Sem. - 1st ed. - Maribor : University of Maribor Press, 2018

Način dostopa (URL): http://press.um.si/index.php/ump/catalog/book/336

ISBN 978-961-286-166-7

doi: 10.18690/978-961-286-166-7
1. Sem, Vilma
COBISS.SI-ID 94606337

ISBN: 978-961-286-166-7 (PDF)

DOI: https://doi.org/10.18690/978-961-286-166-7

Price: Free copy

For publisher: full prof. dr. Žan Jan Oplotnik, Vice-rector (University of Maribor)

Tables, formulas and exercises with key for biometrics

TADEJA KRANER ŠUMENJAK & VILMA SEM

Abstract This publication includes some materials we use in the one semester bachelor course entitled Biometrics, at the Faculty of Agriculture and Life Sciences.

In the last few years, several foreign students came to study at the University of Maribor. A certain number of them came to study at the Faculty of Agriculture and Life Sciences, and because the foreign students are not able to follow the lectures in Slovene, we organized lectures in English. Since there was no learning material in English that would deal with the basics of statistics through solving real problems from agriculture using the statistical program SPSS, we decided to translate and publish these materials, and thus make them more available.

Keywords: • biometrics • statistical tables • statistical formulas • SPSS • exercises •

CORRESPONDENCE ADDRESS: Tadeja Kraner Šumenjak, Ph.D., Assistant Professor, University of Maribor, Faculty of Agriculture and Life Sciences, Pivola 10,2311 Hoče, Slovenia, e-mail: tadeja.kraner@um.si. Vilma Sem, Ph.D., Assistant, University of Maribor, Faculty of Agriculture and Life Sciences, Pivola 10,2311 Hoče, Slovenia, e-mail: vilma.sem@um.si.

DOI https://doi.org/10.18690/978-961-286-166-7 ISBN 978-961-286-166-7
© 2018 University of Maribor Press
Available at: http://press.um.si.

TABLE OF CONTENTS

1 FORMULAS

1.1 RELATIVE NUMBERS

1.1.1 STRUCTURES

PROPORTION	**PERCENTAGE**	**ANGLE**
of units in the *i^th^* group	of units in the *i^th^* group	which corresponds to the *i^th^* group
$f_i^0 = \frac{f_i}{\sum_{i=1}^{K} f_i}$	$f_i\% = \frac{f_i}{\sum_{i=1}^{K} f_i} \cdot 100$	$\varphi_i = \frac{f_i}{\sum_{i=1}^{K} f_i} \cdot 360°$
where K is the number of groups	where K is the number of groups	where K is the number of groups

1.1.2 INDICES

FIXED BASE INDEX NUMBER	**COMPUTING**
$I_j = \frac{Y_j}{Y_0} \cdot 100$ **CHAIN INDEX NUMBER** $V_j = \frac{Y_j}{Y_{j-1}} \cdot 100$	fixed base index number from the chain index number After base period: $I_j = \frac{V_j \cdot I_{j-1}}{100}$ Before base period: $I_{j-1} = \frac{I_j}{V_j} \cdot 100$
NOTATIONS	**AVERAGE OF RELATIVES**
I_j fixed base index number for the *j^th^* year V_j chain index number for the *j^th^* year Y_j data for the *j^th^* year Y_0 data for the base period Y_{j-1} data for the $(j-1)^{th}$ year I_{j-1} fixed base index number for the $(j-1)^{th}$ year	$I = \sqrt[k]{V_1 V_2 \dots V_k}$, *k* is the number of chain indices or $I = 100 \cdot \sqrt[k-1]{\frac{Y_k}{Y_1}}$, *k* is the number of years in the time series **GROWTH RATE** $S_j = V_j - 100$

1.2 DESCRIPTIVE STATISTICS

1.2.1 FREQUENCY DISTRIBUTIONS

COMULATIVE FREQUENCY	NOTATIONS
$F_{k+1} = F_k + f_k$	$x_{k,min}$ lower bound of the k^{th} class $x_{k,max}$ upper bound of the k^{th} class f_k frequency of the k^{th} class x_k midpoint of the k^{th} class i_k width of the k^{th} class
FREQUENCY DENSITY $g_k = \frac{f_k}{i_k}$	
CLASS WIDTH $i_k = x_{k,max} - x_{k,min}$	

1.2.2 QUANTILES

RELATIVE RANK	RANK
$p = \frac{r-0.5}{n}$	$r = n \cdot p + 0.5$
LINEAR INTERPOLATION $\frac{x-x_0}{r_x-r_0} = \frac{x_1-x_0}{r_1-r_0}$	**NOTATIONS** n number of units x value, for which the rank is computed x_0 value in the array one place before x x_1 value in the array one place after x r_x rank, which corresponds to x r_0 rank, which corresponds to x_0 r_1 rank, which corresponds to x_1
QUANTILE x, WHICH CORRESPONDS TO r_x $x = x_0 + (r_x - r_0)(x_1 - x_0)$	
RANK r_x, WHICH CORRESPONDS TO x $r_x = \frac{x-x_0}{x_1-x_0} + r_0$	

1.2.3 MEASURES OF CENTRAL TENDENCY

SAMPLE MEAN	NOTATIONS
$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$ Grouped data: $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i}$	x_i i^{th} observation of variable X f_i frequency of the i^{th} group n sample size k number of groups
SAMPLE GEOMETRIC MEAN $g = \sqrt[n]{x_1 \cdot x_2 \ldots x_n}$	**SAMPLE MODE (mo)** is the value that occurs most frequently in a data set.
SAMPLE HARMONIC MEAN $h = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}$	**SAMPLE MEDIAN (me)** is a quantile with a relative rank of $\mathbf{0.5}$ or central value in the array.

Note: For computation of population parameters the same expressions can be used, only the mean, geometric mean and harmonic mean are computed over all members of the population.

1.2.4 MEASURES OF VARIABILITY

SAMPLE RANGE	NOTATIONS
$vr = x_{max} - x_{min}$	$x_{\max}$ observed maximum x_{min} observed minimum q_1 first quartile q_3 third quartile $\bar{x}$ sample mean me sample median x_i i^{th} observation of variable X f_i frequency of the i^{th} class N population size M population mean n sample size k number of groups
SAMPLE QUARTILE DEVIATION OR SEMI-INTERQUARTILE RANGE $q_0 = \frac{1}{2}(q_3 - q_1)$	
SAMPLE INTERQUARTILE RANGE $q_r = q_3 - q_1$	
OUTLIERS are observations that are out of interval: $(q_1 - 1.5q_r,\ q_1 + 1.5q_r)$	
VARIANCE Population: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - M^2$ Sample: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}) = $ $= \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{n}{n-1}\bar{x}^2$ Grouped data: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} f_i x_i^2 - M^2$ $s^2 = \frac{1}{n-1}\sum_{i=1}^{n} f_i x_i^2 - \frac{n}{n-1}\bar{x}^2$	**MEAN ABSOLUTE DEVIATION AROUND A CENTRAL POINT** Around mean: $AD_{\bar{x}} = \frac{1}{n}\sum \lvert x_i - \bar{x}\rvert$ Around median: $AD_{me} = \frac{1}{n}\sum \lvert x_i - me\rvert$ **STANDARD DEVIATION** Population: $\sigma = \sqrt{\sigma^2}$ Sample: $s = \sqrt{s^2}$

1.3 INFERENTIAL STATISTICS

1.3.1 ONE POPULATION

CONFIDENCE INTERVAL FOR THE POPULATION MEAN	**NOTATIONS**
$\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \leq M \leq \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ Degrees of freedom for t-distribution (Table 2.2): $df = n - 1$.	M population mean M_H hypothetical mean $\bar{x}$ sample mean s^2 sample variance s sample standard deviation $t_{\frac{\alpha}{2}}$ critical *t*-value for a two-tailed area n sample size
ONE SAMPLE T-TEST a) Hypotheses $H_0: M_H = M$ $H_1: M_H \neq M$ b) Sample test statistic $t = \frac{(\bar{x} - M_H)}{\frac{s}{\sqrt{n}}}$ This test statistic follows a t-distribution (Table 2.2) with degrees of freedom: $df = n - 1$.	

1.3.2 TWO POPULATIONS

INDEPENDENT SAMPLES T-TEST	**PAIRED SAMPLES t-TEST**
a) Hypotheses $H_0: M_1 = M_2$ $H_1: M_1 \neq M_2$ b) Sample test statistic $s^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$ $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2 \cdot \frac{n_1 + n_2}{n_1 n_2}}}$, This statistic follows a t-distribution (Table 2.2) with degrees of freedom: $df = n_1 + n_2 - 2$.	a) Hypotheses $H_0: M_1 = M_2$ or $M_1 - M_2 = 0$ $H_1: M_1 \neq M_2$ or $M_1 - M_2 \neq 0$ b) Sample test statistic $s_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} d_i^2 - \frac{n}{n-1} \bar{d}^2$ $t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$ This statistic follows a *t*-distribution (Table 2.2) with degrees of freedom: $df = n - 1$.
NOTATIONS M_1 mean of population 1 M_2 mean of population 2 $\bar{x}_1$ sample mean from population 1 n_1 sample size from population 1 $\bar{x}_2$ sample mean from population 2 n_2 sample size from population 2 s^2 pooled variance	s pooled standard deviation s_1^2 sample variance from population 1 s_2^2 sample variance from population 2 d_i pair difference $\bar{d}$ mean of the differences d_i s_d^2 sample variance of the differences d_i n number of data pairs

1.3.3 MORE POPULATIONS – ANALYSIS OF VARIANCE

a) HYPOTHESES

$H_0: M_1 = M_2 = \cdots = M_k$

H_1: The means are not all equal.

b) ANOVA TABLE

Source of variation	Sums of squares	Degrees of freedom	Mean square	F-statistic
Between groups	SSB	$df_1 = k - 1$	$MSB = \frac{SSB}{df_1}$	$F = \frac{MSB}{MSW}$
Within groups	SSW	$df_2 = N - k$	$MSW = \frac{SSW}{df_2}$	This statistic follows an *F*-distribution (Tab. 2.3) $F(k-1, N-k)$
Total	SST	$df = N - 1$		

c) NOTATIONS

M_i	mean of the i^{th} population	n_i	number of observations in the i^{th} group
SST	sum of squares total	x_{ij}	j^{th} observation at i^{th} group
SSB	sum of squares between	N	total number of observations
SSW	sum of squares within	MSB	mean square between groups
k	number of groups (treatments)	MSW	mean square within groups

d) COMPUTING SUMS OF SQUARES AND DEGREES OF FREEDOM FOR GROUPS WITH EQUAL SIZES

$$n_1 = n_2 = \cdots = n_k = n$$

$$C = \frac{1}{N}\left(\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}\right)^2$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - C$$

$$SSB = \frac{1}{n}\sum_{i=1}^{k}\left(\sum_{j=1}^{n} x_{ij}\right)^2 - C$$

$$SSW = SST - SSB$$

1.3.4 CORRELATION

<table>
<tr><td>

PEARSON'S CORRELATION COEFFICIENT

a) Hypotheses

$H_0: \rho = 0$ (no linear correlation between X and Y)

$H_1: \rho \neq 0$ (linear correlation between X and Y)

b) Sample Pearson's correlation coefficient

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} =$$

$$= \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n\sum_{i=1}^{n}(x_i)^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n}(y_i)^2 - (\sum_{i=1}^{n} y_i)^2}}$$

c) Compare your obtained correlation coefficient to the critical values in the table 2.5 or use the approximation

$$t = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1 - r_{xy}^2}},$$

that follows a t-distribution (Table 2.2) with the following degrees of freedom:

$df = n - 2.$

</td><td>

NOTATIONS

n number of pairs

x_i i[th] observation of variable X

$\bar{x}$ sample mean of variable X

y_i i[th] observation of variable Y

$\bar{y}$ sample mean of variable Y

</td></tr>
<tr><td>

SPEARMAN'S CORRELATION COEFFICIENT

a) Hypotheses

$H_0: \rho_S = 0$ (no monotone correlation between X and Y)

$H_1: \rho_S \neq 0$ (monotone correlation between X and Y)

b) Sample Spearman's correlation coefficient

Convert the raw data on each variable into ranks.

If there are no ties in the ranks:

$$r_S = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2 - 1)}$$

otherwise:

$$r_S = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Compare your obtained correlation coefficient's r_S against the critical values in the table 2.6 (Critical values of the Spearman's correlation coefficient).

</td><td>

NOTATIONS

n number of pairs

d_i difference in ranks

$\bar{x}$ sample mean of ranked variable X

$\bar{y}$ sample mean of ranked variable Y

</td></tr>
</table>

PEARSON'S CHI-SQUARE TEST

a) Hypotheses

H_0: Variable X and variable Y are independent.

H_1: Variable X and variable Y are not independent.

b) Sample test statistic

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

c) This statistic follows a chi-square distribution (Table 2.4) with the following degrees of freedom: $df = (r-1)(c-1)$.

NOTATIONS

X variable with categories $A_1, A_2, \dots, A_r$

Y variable with categories $B_1, B_2, \dots, B_c$

Contingency table of observed frequencies:

	B_1	B_2	…	B_c	Sum
A_1	O_{11}	O_{12}	…	O_{1c}	$O_{1.}$
A_2	O_{21}	O_{22}	…	O_{2c}	$O_{2.}$
⋮	⋮	⋮	⋱	⋮	⋮
A_r	O_{r1}	O_{r2}	…	O_{rc}	$O_{r.}$
Sum	$O_{.1}$	$O_{.2}$	…	$O_{.c}$	N

O_{ij} observed frequency

$E_{ij} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{N}$ expected frequency

N total number of observations

YATES CORRECTION

a) Hypotheses

H_0: Variable X and variable Y are independent.

H_1: Variable X and variable Y are not independent.

b) Sample test statistic

$$\chi^2 = \frac{n\left(|N_{11}N_{22} - N_{21}N_{12}| - \frac{n}{2}\right)^2}{L_1 L_2 S_1 S_2}$$

c) This statistic follows a chi-square distribution (Table 2.4) with the following degrees of freedom: $df = 1$.

NOTATIONS

Contingency table for variable X (with categories A_1 and A_2) and variable Y (with categories B_1 and B_2).

	B_1	B_2	Sum
A_1	N_{11}	N_{12}	L_1
A_2	N_{21}	N_{22}	L_2
Sum	S_1	S_2	N

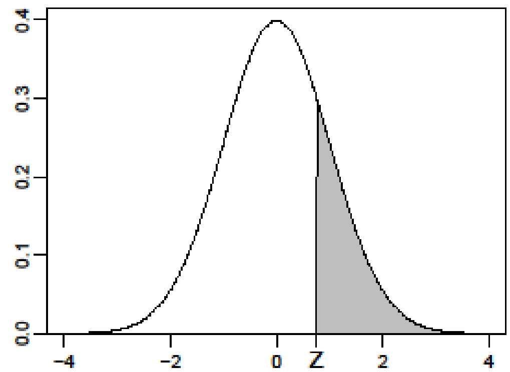
1.3.5 LINEAR REGRESSION

LEAST-SQUARES LINE	**NOTATIONS**
Population: $Y = \alpha + \beta X$ Sample estimate: $\hat{Y} = a + bX$ $b = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$ $a = \bar{y} - b\bar{x}$ $r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n\sum_{i=1}^{n}(x_i)^2 - (\sum_{i=1}^{n} x_i)^2}\sqrt{n\sum_{i=1}^{n}(y_i)^2 - (\sum_{i=1}^{n} y_i)^2}}$ $s_e = \sqrt{\frac{\sum_{i=1}^{n}(y-\hat{y})^2}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i y_i}{n-2}}$	n number of pairs x_i i[th] observation of variable X $\bar{x}$ sample mean of variable X y_i i[th] observation of variable Y $\bar{y}$ sample mean of variable Y b slope of the line (sample) a intercept of the line (sample) β slope of the line (population) α intercept of the line (population) r_{xy} correlation coefficient r_{xy}^2 coefficient of determination s_e standard error of estimate
TESTING $\boldsymbol{\beta}$ a) Hypotheses $H_0: \beta = 0$ (variable X has no effect on Y) $H_1: \beta \neq 0$ (variable X has linear effect on Y) b) Sample test statistic $t = \frac{b}{s_e}\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}$ c) This statistic follows t-distribution (Table 2.2) with degrees of freedom: $df = n - 2$.	**NOTATIONS** β slope of the line (population) s_e standard error of estimate b slope of the line (sample) n number of pairs

2 TABLES

2.1 STANDARD NORMAL DISTRIBUTION TABLE

This table presents the area under the standard normal curve between the z score and infinity: $p = P(Z \geq z)$.



Example:

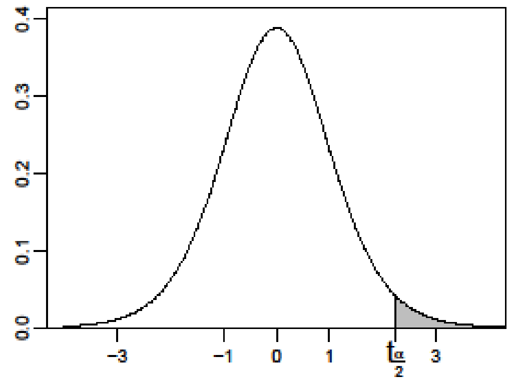
Area under the standard normal curve for $z = 0.75$ is $p = 0.2266$.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

2.2 CRITICAL VALUES OF STUDENTS' T-DISTRIBUTION

The table entry is the value of $t_{\frac{\alpha}{2}}$, having an area to the right of $\frac{\alpha}{2}$ under t-distribution with df degrees of freedom (two-sided test):

$$p = P\left(T \geq t_{\frac{\alpha}{2}}\right) = P\left(T \leq -t_{\frac{\alpha}{2}}\right) = \frac{\alpha}{2}.$$



Example:

$\alpha = 0.05 \implies \frac{\alpha}{2} = 0.025$ and $df = 10$

$t_{\frac{\alpha}{2}}(10) = 2.228$

df	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.660
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

2.3 CRITICAL VALUES OF THE F-DISTRIBUTION (FISHER'S DISTRIBUTION)

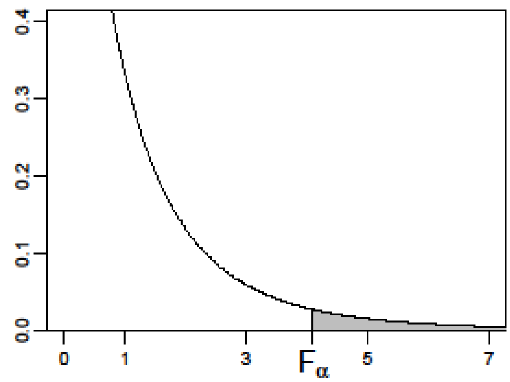
The table entry is the value of F_α, having an area to the right of α under F-distribution with df_1 and df_2 degrees of freedom:

$p = P(F \geq F_\alpha) = \alpha.$

Example:

$\alpha = 0.05, df_1 = 2, df_2 = 10$

$F_{0.05}(2, 10) = 4.103$

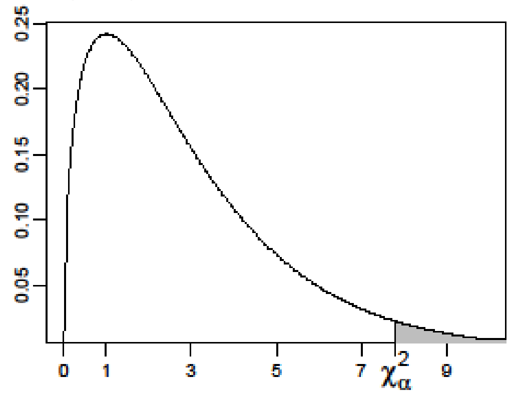


df_2	df_1												
	1	2	3	4	5	6	7	8	10	12	20	25	∞
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.40	19.41	19.45	19.46	19.50
3	10.13	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.785	8.745	8.660	8.634	8.527
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.964	5.912	5.803	5.769	5.628
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.735	4.678	4.558	4.521	4.365
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.060	4.000	3.874	3.835	3.669
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.637	3.575	3.445	3.404	3.230
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.347	3.284	3.150	3.108	2.928
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.137	3.073	2.936	2.893	2.707
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	2.978	2.913	2.774	2.730	2.538
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.854	2.788	2.646	2.601	2.405
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.753	2.687	2.544	2.498	2.297
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.671	2.604	2.459	2.412	2.207
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.602	2.534	2.388	2.341	2.131
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.544	2.475	2.328	2.280	2.066
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.494	2.425	2.276	2.227	2.010
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.450	2.381	2.230	2.181	1.961
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.412	2.342	2.191	2.141	1.917
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.378	2.308	2.155	2.106	1.879
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.348	2.278	2.124	2.074	1.844
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.321	2.250	2.096	2.045	1.812
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.297	2.226	2.071	2.020	1.784
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.275	2.204	2.048	1.996	1.758
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.255	2.183	2.027	1.975	1.734
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.236	2.165	2.007	1.955	1.712
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.220	2.148	1.990	1.938	1.691
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.204	2.132	1.974	1.921	1.672
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.190	2.118	1.959	1.906	1.655
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.177	2.104	1.945	1.891	1.638
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.165	2.092	1.932	1.878	1.623
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.077	2.003	1.839	1.783	1.510
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	1.993	1.917	1.748	1.690	1.390
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.910	1.834	1.659	1.598	1.255
∞	3.842	2.996	2.605	2.372	2.214	2.099	2.010	1.939	1.831	1.752	1.571	1.506	1.025

2.4 CRITICAL VALUES OF THE CHI-SQUARE DISTRIBUTION

The table entry is the value of χ^2_α, having an area to the right of α under chi-square distribution with df degrees of freedom:

$p = P(\chi^2 \geq \chi^2_\alpha) = \alpha.$



Example:

$\alpha = 0.05, \mathrm{df} = 3$

$\chi^2{}_{0,05}(9) = 7.815$

df	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.000	0.001	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.010	0.051	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.072	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620
26	11.160	13.844	31.795	35.563	38.885	41.923	42.856	45.642	48.290	51.627	54.052
27	11.808	14.573	32.912	36.741	40.113	43.195	44.140	46.963	49.645	53.023	55.476
28	12.461	15.308	34.027	37.916	41.337	44.461	45.419	48.278	50.993	54.411	56.892
29	13.121	16.047	35.139	39.087	42.557	45.722	46.693	49.588	52.336	55.792	58.301
30	13.787	16.791	36.250	40.256	43.773	46.979	47.962	50.892	53.672	57.167	59.703
31	14.458	17.539	37.359	41.422	44.985	48.232	49.226	52.191	55.003	58.536	61.098
32	15.134	18.291	38.466	42.585	46.194	49.480	50.487	53.486	56.328	59.899	62.487
33	15.815	19.047	39.572	43.745	47.400	50.725	51.743	54.776	57.648	61.256	63.870
34	16.501	19.806	40.676	44.903	48.602	51.966	52.995	56.061	58.964	62.608	65.247
35	17.192	20.569	41.778	46.059	49.802	53.203	54.244	57.342	60.275	63.955	66.619

2.5 CRITICAL VALUES OF PEARSON'S CORRELATION COEFFICIENT

The results are significant if the calculated value of r_{xy} is greater or equal than the table value.

n	0.1	0.05	0.01
3	0.988	0.997	0.999
4	0.900	0.950	0.990
5	0.805	0.878	0.959
6	0.729	0.811	0.917
7	0.669	0.754	0.875
8	0.621	0.707	0.834
9	0.582	0.666	0.798
10	0.549	0.632	0.765
11	0.521	0.602	0.735
12	0.497	0.576	0.708
13	0.476	0.553	0.684
14	0.458	0.532	0.661
15	0.441	0.514	0.641
16	0.426	0.497	0.623
17	0.412	0.482	0.606
18	0.400	0.468	0.590
19	0.389	0.456	0.575
20	0.378	0.444	0.561
21	0.369	0.433	0.549
22	0.360	0.423	0.537
23	0.352	0.413	0.526
24	0.344	0.404	0.515
25	0.337	0.396	0.505
26	0.330	0.388	0.496
27	0.323	0.381	0.487
28	0.317	0.374	0.479
29	0.311	0.367	0.471
30	0.306	0.361	0.463
40	0.264	0.312	0.403
50	0.235	0.279	0.361
60	0.214	0.254	0.330
70	0.198	0.235	0.306
80	0.185	0.220	0.286
90	0.174	0.207	0.270
100	0.165	0.197	0.256

2.6 CRITICAL VALUES OF SPEARMAN'S CORRELATION COEFFICIENT

The results are significant if the calculated value of r_s is greater or equal than the table value.

n	0.1	0.05	0.01
1			
2			
3			
4	1.000		
5	0.900	1.000	
6	0.829	0.886	1.000
7	0.714	0.786	0.929
8	0.643	0.738	0.881
9	0.600	0.700	0.833
10	0.564	0.648	0.794
11	0.536	0.618	0.755
12	0.503	0.587	0.727
13	0.484	0.560	0.703
14	0.464	0.538	0.679
15	0.446	0.521	0.654
16	0.429	0.503	0.635
17	0.414	0.485	0.615
18	0.401	0.472	0.600
19	0.391	0.460	0.584
20	0.380	0.447	0.570
21	0.370	0.435	0.556
22	0.361	0.425	0.544
23	0.353	0.415	0.532
24	0.344	0.406	0.521
25	0.337	0.398	0.511
26	0.331	0.390	0.501
27	0.324	0.382	0.491
28	0.317	0.375	0.483
29	0.312	0.368	0.475
30	0.306	0.362	0.467
31	0.301	0.356	0.459
32	0.296	0.350	0.452
33	0.291	0.345	0.446
34	0.287	0.340	0.439
35	0.283	0.335	0.433
36	0.279	0.330	0.427
37	0.275	0.325	0.421
38	0.271	0.321	0.415
39	0.267	0.317	0.410
40	0.264	0.313	0.405

3 EXERCISES

3.1 DESCRIPTIVE STATISTICS

1) The given frequency table describes the weights of bulls (hypothetical data):

Weights in kg	Number of bulls
Above 170 to 190	18
Above 190 to 210	38
Above 210 to 230	68
Above 230 to 250	76
Above 250 to 270	30
Above 270 to 290	16
Above 290 to 310	4
Total	250

a) Express this frequency distribution with a histogram.

b) Compute the sample mean, sample variance and sample standard deviation.

2) The average hourly earnings of production workers for selected periods are given below (hypothetical data).

Year	2007	2008	2009	2010	2011	2012	2013
Average hourly earnings (EUR)	6.32	6.57	6.83	7.43	8.28	8.74	9.34

a) From the following data, compute the fixed base index number for 2013 taking 2010 as the base year and interpret the result.

b) Calculate the chain index numbers and the average of relatives. Interpret the chain index number for 2010.

3) The table shows chained index numbers for production of wine (hypothetical data).

Year	2011	2012	2013	2014	2015
Chained index number	–	117.2	113.6	78.8	146.6

Calculate the fixed base index numbers taking 2014 as the base year.

4) A man travels from city A to city B. The first half of the way, he drove at the constant speed of 90 km per hour. Then he (instantaneously) increased his speed and travelled the remaining distance at 130 km per hour. Find the average speed of the motion.

5) A man was travelling from city A to city B. For the first hour, he drove at the constant speed of 90 km per hour. Then he (instantaneously) increased his speed and, for the next hour, kept it at 130 km per hour. Find the average speed of the motion.

6) Lentil seed yields (kg/plot) were as shown in the following table:

52.2	34.9	29.7	28.4	28.0	22.8	18.5	19.4	18.8	19.5
13.1	10.1	41.5	36.3	31.7	31.0	28.2	33.0	26.0	30.6

a) From the data compute the sample mean, the sample standard deviation.
b) Compute the median, the first quartile, the third quartile and the interquartile range.
c) Make a box plot of the data.

7) The following data are the average annual temperatures from 2001-2010 for meteorological station Maribor (Source: ARSO, 2018).

Year	Average annual temperatures
2001	10.983
2002	11.799
2003	11.249
2004	10.406
2005	10.099
2006	10.781
2007	11.711
2008	11.510
2009	11.277
2010	10.470

a) From the data compute the sample mean, the sample standard deviation and the median.
b) Construct a box-plot of the data.
c) Find the 35th percentiles.

3.2 NORMAL DISTRIBUTION

8) Assume X is normally distributed with a mean of 10 and a standard deviation of 2. Determine the following:
a) $P(X \leq 13)$,
b) $P(-2 \leq X \leq 8)$,
c) $P(X \geq 11.5)$.

9) The heights of students (variable X) at the Faculty of Agriculture and Life Sciences in Hoče are normally distributed with a mean of 174 cm and a standard deviation of 8 cm (hypothetical data).
a) What percentage of the students are between 165 and 185 cm tall?
b) Find $P(X \leq 150)$.
c) What is the value that separates the top 1 % of heights from the rest of the population?
d) Find x_1 and x_2 for the middle 95 % of the area under the standard normal curve.

10) The plant height in a rice crop is normally distributed with a mean of 75 cm and standard deviation 5 cm. Find the probability that a random sample of $n = 25$ rice plants will have sample mean

a) less than 73.5 cm.
b) between 74 and 75.4 cm.
c) more than 77 cm.

3.3 CONFIDENCE INTERVAL

11) A study aims to estimate the mean systolic blood pressure of Slovenian adults by randomly sampling and measuring the blood pressure of 100 adults from the population. From their sample, they estimate the sample mean to be 70mmHg and the sample standard deviation to be 8mmHg (hypothetical data). Find the 95% (99 %) confidence interval.

12) Over the past several months, an adult patient has been treated for tetany. This condition is associated with an average total calcium level. Recently, the patients total calcium tests gave the following readings (Source: Brase and Brase, 2006):

9.3	8.8	10.1	8.9	9.4	9.8	10.0	9.9	11.2	12.1

Assume that the population of values has an approximately normal distribution. Find a 99% confidence interval for the population mean of the total calcium value in the patient’s blood.

3.4 ONE SAMPLE T-TEST

13) A sample of 12 eggs are randomly selected and their weights (in g) recorded, which are as follows:

60.2	53.8	67.2	56.9	58.6	60.0
66.3	50.7	56.0	63.3	58.2	64.1

Can we say that the mean egg weight in the population is different from 62.0 g (α=0.05)? Assume the variable is normally distributed.

14) A grain yield of standard variety of wheat is around 150 kg/plot. A new variety is planted on twenty-five randomly selected plots. The observed sample average for the new variety is 159.4 kg of grain per plot with a standard deviation of 14.1 kg. Should the new variety be used instead of the standard one, by using a 5% level of significance by assuming normal distribution of selected variable?

15) The mean length of apple tree roots of apple tree varieties obtained from previous researches is 57.62 cm. Interpret the SPSS output for observed sample which contains 18 trees.

a) Write H_0.

b) Write the test statistic.

c) Determine the level of significance.

d) Write the p-value.

e) Compare p-value with level of significance and interpret the result.

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Length	18	54.8333	19.30331	4.54983

One-Sample Test						
	Test Value = 57.62					
	t	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Length	-.612	17	.548	-2.78667	-12.3860	6.8126

3.5 INDEPENDENT SAMPLES T-TEST

16) The experiment was made for comparing grain yields (kg/parcel) for two varieties of corn. The grain yields were as follows:

Parcel	Variety A	Variety B
1	33	27
2	25	43
3	20	36
4	19	20
5	42	22

Test whether these two varieties differ significantly at a 5 percent level of probability with respect to the grain yield (assume that population variances are equal and variables are normally distributed).

17) The difference between two types of fertilizers (organic and chemical) are being observed in order to compare the yield of grapevine. The organic fertilizer was applied on 44 vines with an average grape yield of 3.5 kg per vine and a standard deviation of 1.5 kg. The chemical fertilizer was used on 47 vines with an average grape yield of 3.9 kg per vine and standard deviation of 1.1 kg. Can we say that the chemical fertilizer produces a higher yield than the organic ($\alpha = 0.05$)?

18) A newly developed variety of alfalfa was investigated in two different climatic zones to test the significance of the difference in the yield (kg/plot). According to the SPSS output, do the following tasks:

a) Write H_0.
b) Write the test statistic.
c) Determine the level of significance.
d) Write the p-value.
e) Compare the p-value with the level of significance and interpret the result.

Group Statistics					
		N	**Mean**	**Std. Deviation**	**Std. Error Mean**
Yield	1	10	24.7000	4.27005	1.35031
	2	10	28.5000	2.75882	0.87242

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
Yield	Equal variances assumed	1.271	0.274	-2.364	18	0.030	-3.800	1.608	-7.177	-0.423
	Equal variances not assumed			-2.364	15.399	0.032	-3.800	1.608	-7.219	-0.381

3.6 PAIRED SAMPLES T-TEST

19) Two laboratories carry out independent estimates of protein content in rice. Eight estimates were made by each laboratory. Each time one sample was taken. Half the quantity was sent to Lab 1 and the other half to Lab 2. We assume that the differences follow from a normally distributed population. The results are as follows ($\alpha = 0.05$).

Lab 1	8	10	8	10	10	9	11	9
Lab 2	9	11	9	11	10	8	10	8

Do the two laboratories report the same results?

20) Consider the following study in which standing and supine systolic blood pressures were compared. This study was performed on 8 patients. Their blood pressures were measured in both positions. According to the SPSS output, do the following tasks:

a) Write H_0.

b) Write the test statistic.

c) Determine the level of significance.

d) Write the p-value.

e) Compare the p-value with level of significance and interpret the result.

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Standing - Supine	-3.25000	3.49489	1.23563	-6.17180	-0.32820	-2.630	7	0.034

3.7 ANALYSIS OF VARIANCE

21) The plot grain yields (kg) of three varieties of oats in an experiment are:

Plot	Variety A	Variety B	Variety C
1	22	24	19
2	18	39	22
3	35	19	24
4	17	25	17
5	42	23	28

Assume that ANOVA assumptions are met. Is there any difference in the grain yield produced by these varieties (0.05)?

22) Four diets were formulated, and their effect on weight loss was investigated on people of the same age, sex, and other activities. After a certain period, the weight loss (kg) was recorded and results were analyzed via SPSS.

a) Write H_0.

b) Write the test statistic.

c) Determine the level of significance.

d) Write the p-value.

e) Compare the p-value with the level of significance and interpret the result.

ANOVA					
weightloss					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	70.245	3	23.415	14.218	0.000
Within Groups	44.464	27	1.647		
Total	114.710	30			

3.8 PEARSON' S CORRELATION COEFFICIENT

23) Consider the following study in which systolic and diastolic blood pressures were compared. This study was performed on 7 patients. Make a scatter diagram. Can we say, at a 0.05 level of significance, that there is a linear correlation between systolic and diastolic blood pressures? Assumed that the assumptions are met.

Systolic blood pressure	Diastolic blood pressure
210	130
169	122
187	124
160	104
167	112
176	101
185	121

24) The calculated Pearson's correlation coefficient from a sample of 28 observations between two variables equals 0,5. Is the correlation significant at level 0,01?

25) The ear length and number of grains per ear of maize cultivar MASTER have been investigated in order to study the relationship between variables. On a base of the SPSS output, do the following tasks:

a) Write H_0.
b) Write the test statistic.
c) Determine the level of significance.
d) Write the p-value.
e) Compare p-value with level of significance and interpret the result.

Correlations			
		length	number
length	Pearson Correlation	1	0.896**
	Sig. (2-tailed)		0.000
	N	12	12
number	Pearson Correlation	0.896**	1
	Sig. (2-tailed)	0.000	
	N	12	12
**. Correlation is significant at the 0.01 level (2-tailed).			

3.9 SPEARMAN' S CORRELATION COEFFICIENT

26) The table shows a maximum monthly averaged temperature and monthly averaged sunshine hours for the climate station Oxford, covering the period from 1981 to 2010.
(Source: Metoffice. gov.uk, 2018). Find out if there is a correlation between both variables.

Month	Max. temp. (°C)	Sunshine (hours)
Jan	7.60	62.50
Feb	8.00	78.90
Mar	10.90	111.20
Apr	13.60	160.90
May	17.10	192.90
Jun	20.30	191.00
Jul	22.70	207.00
Aug	22.30	196.50
Sep	19.10	141.20
Oct	14.80	111.30
Nov	10.50	70.70
Dec	7.70	53.80

27) Ten commercial samples (ten different brands) of yoghurt were analyzed. In order to determine whether the two evaluators agree with one another in their evaluation of ten samples (1: dislike extremely to 10: like extremely). Measure the agreement between the evaluators via the Spearman rank correlation coefficient.

Commercial samples	Evaluator 1	Evaluator 2
1	6	5
2	4	6
3	9	10
4	1	2
5	2	3
6	7	8
7	3	1
8	8	7
9	5	4
10	10	9

28) With a data base for year the 2011 provided by the network of the Farm Accountancy Data (Eng. Abbreviation FADN), researchers studied the correlation between variables of the gross value added and agricultural land in use (Source: Trpin Švikart, 2016). According to SPPS Output, do the following activities:

a) Write H_0.
b) Write the test statistic.
c) Determine the level of significance.
d) Write the p-value.
e) Compare the p-value with level of significance and interpret the result.

Correlations[a]				
			land in use	gross value added
Spearman's rho	land in use	Correlation Coefficient	1.000	0.672**
		Sig. (2-tailed)	.	0.000
		N	2818	2818
	gross value added	Correlation Coefficient	0.672**	1.000
		Sig. (2-tailed)	0.000	.
		N	2818	2821
**Correlation is significant at the 0.01 level (2-tailed).				
Year = 2011				

3.10 CHI-SQUARED TEST

29) Researchers interviewed consumers of different varieties of apples, asking them why they buy a specific variety of apples (random sample). The following results were obtained.

	Variety A	Variety B	Variety C	Total
Colour	320	18	62	400
Taste	42	13	45	100
Size	110	57	83	250
Durability	28	12	10	50
Total	500	100	200	800

Find if the sale of different apple varieties and different apple properties are independent.

30) Entomologists have investigated yellow, short-leaved and spruce pines in a certain forest to see how many were being seriously attacked by insects. An investigation of 250 randomly selected trees of each species gave the following results (Source: Palaniswamy and Palaniswamy, 2006):

Species	Seriously damaged	Not damaged	Total
Yellow	58	192	250
Short leaved	80	170	250
Spruce	78	172	250

Are insects attacking one of the species more than the others?

31) In an experiment, the following contingency table was obtained. We want to test the association between row and column classification using Yates' correction for continuity.

Variable 1 \ Variable 2	Column 1	Column 2
Row 1	7	34
Row 2	16	30

3.11 LINEAR REGRESSION

32) In an investigation into the interdependence of water uptake and food intake in egg production, the following records were obtained from a ten day period of observation of 12 birds (hypothetical data). Assume that regression analysis assumptions are met.

Water uptake and food intake in (g) per bird/day	Egg production (eggs/10days)
265	2
492	5
366	6
514	9
406	6
336	7
343	4
377	6
143	0
204	0
360	2
292	0

a) Express this relation in a scatter diagram and add the best fitting line.
b) What is the equation of the least-squares line?
c) Calculate the predicted egg production when x=450.
d) Compute the coefficient of determination.
e) What percentage of the variation in y is explained by the regression line?
f) What percentage of the variation in y is not explained by the regression line?
g) Use a 1 % level of significance to test the claim that : $\rho \neq 0$.
h) Use a 1 % level of significance to test the claim that : $\beta \neq 0$.

33) In an investigation of the relationship between the amount of addition and chickens' weight, the following results were obtained using SPSS.

a) Express the equation of the line of regression.
b) Find the estimated y for $x = 21$.
c) Test the hypothesis $\beta = 0$ against : $\beta \neq 0$, use ; $\alpha = 0.01$.
d) What percentage of the variation in y is explained by the regression line?

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.937[a]	0.878	0.848	9.12871
a. Predictors: (Constant), food amount				

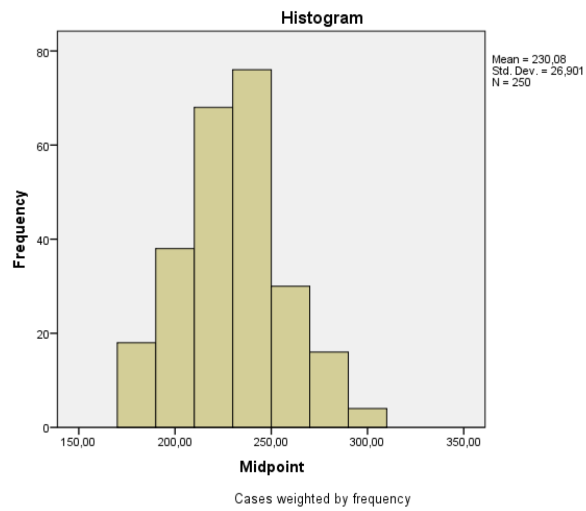
Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	487.333	29.307		16.629	0.000
	food amount	8.000	1.491	0.937	5.367	0.006
a. Dependent Variable: chicken weight						

4 ANSWERS TO EXERCISES

4.1 DESCRIPTIVE STATISTICS

1)

a) Histogram:



b) By using the formulas given in sections 1.2.3 and 1.2.4, the following can be obtained:
$\bar{x} = 230.08; s^2 = 723.69; s = 26.9.$

2) By using the formulas given in section 1.1.2, the following can be obtained:

a) 125.7; from 2010 to 2013 there is 25.7 percent increase.

b) $V_{2008} = 104;\ V_{2009} = 104;\ V_{2010} = 108.8;\ V_{2011} = 111.4;\ V_{2012} = 105.6;\ V_{2013} = 106.9; I = 106.7;$ this means that from 2009 to 2010 there is an 8.8 percent increase.

3) By using the formulas given in section 1.1.2, the following can be obtained:

Year	2011	2012	2013	2014	2015
Fixed base index numbers (2014)	95.3	111.7	126.9	100	146.6

4) By using the formula given in section 1.2.3, it follows that:
$h = 106.4.$

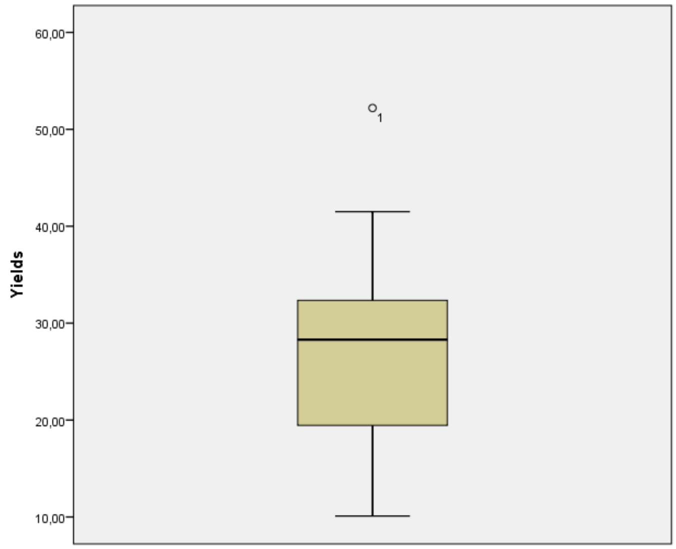
5) By using the formula given in section 1.2.3, it follows that:
$\bar{x} = 110.$

6) Using the formulas given in sections 1.2.2, 1.2.3 and 1.2.4, we get:

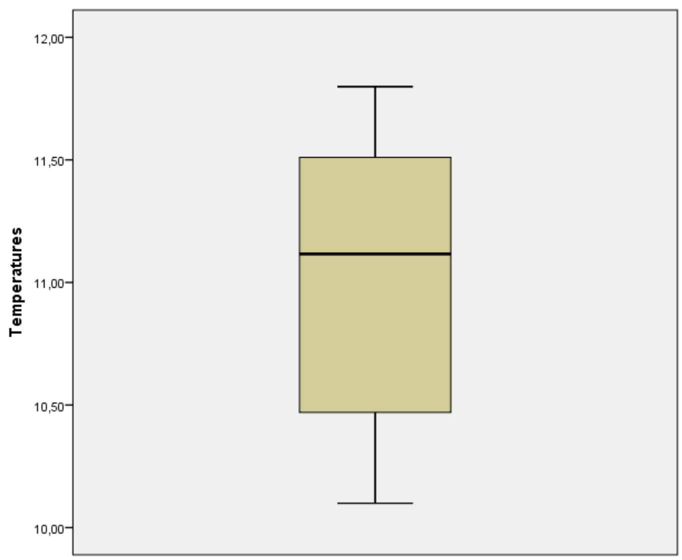
a) $\bar{x} = 27.685; s = 9.826.$

b) $me = 28.3; q_1 = 19.45;\ q_3 = 32.35; q_r = 12.90.$

c) Box-plot:



7) Using the formulas given in sections 1.2.2, 1.2.3 and 1.2.4, we obtain:
   a) $\bar{x} = 11.0285; s = 0.5801; me = 11.115.$
   b) Box-plot:



   c) 10.78.

4.2 NORMAL DISTRIBUTION

8) If we define a new continuous random variable $\boldsymbol{Z}$, in terms of $\boldsymbol{X}$ as $\boldsymbol{Z} = \frac{\boldsymbol{X-M}}{\boldsymbol{\sigma}}$ and used that $\boldsymbol{P(X_1 \leq X \leq X_2) = P(Z_1 \leq Z \leq Z_2)}$, then, by considering the standard normal distribution table 2.1 and symmetry of the normal distribution, we obtain the following areas under the normal curve:
   a) z =1.5; $P(X \leq 13) = P(Z \leq 1.5) = 0.9332.$
   b) $z_1 = -6;\ z_2 = -1;\ P(-2 \leq X \leq 8) = P(-6 \leq Z \leq -1) = 0.1587.$
   c) $z = 0.75;\ P(X \geq 11.5) = P(Z \geq 0.75) = 0.2266.$

   You can easily verify the above results by using the normal distribution calculator.

9) You can solve a) and b) by following all hints given in exercise 8):

a) $z_1 = -1.125;\ z_2 = 1.375;\ 78.5$ %.

b) $z = -3;\ P(X \leq 150) = 0.0013$.

c) First convert the area of the standard normal curve to 0.01, find the closest value to 0.01 which gives you z-score of 2.33 and then insert it into the equation $X = Z\sigma + M$ which equals $x = 192.64$.

d) You can get the solution by following hints given in case c) and by considering the symmetry of the normal distribution: $x_1 = 158.32;\ x_2 = 189.68$.

10) When we sample from a normal population, then sampling distribution of $\bar{X}$ is a normal distribution and the variable $Z = \frac{\bar{X}-M}{\frac{\sigma}{\sqrt{n}}}$ follows the standard normal curve which, by considering the standard normal distribution table 2.1 and symmetry, gives :

a) $z = -1.5; 0.0668$.

b) $z_1 = -1;\ z_2 = 0.4; 0.4967$.

c) $z = 2;\ 0.0228$.

4.3 CONFIDENCE INTERVAL

11) By using the formulas given in section 1.3.1, it follows that:
$68.42 \leq \mathrm{M} \leq 71.58; 67.91 \leq \mathrm{M} \leq 72.09$.
You can verify the above results by using the confidence interval calculator.

12) First calculate the sample mean and sample standard deviation, then, by using the formulas given in section 1.3.1, you get:
$8.9 \leq \mathrm{M} \leq 11$.

4.4 ONE SAMPLE T-TEST

13) The solution is obtained by calculating the sample mean and sample standard deviation and then using the formulas given in section 1.3.1:
$H_0: M = 62.0;\ H_1: M \neq 62.0; \alpha = 0.05;\ t_{comp} = -1.663;\ df = 11;\ t_{crit} = 2.201$; the null hypothesis is retained, thus at the 5% level of significance, the population mean egg weight is not significantly different from 62.0. You can verify the obtained results by using the one sample t-test calculator.

14) By using the formulas given in the section 1.3.1 it follows that:
$H_0: M = 150.0;\ H_1: M \neq 150.0; \alpha = 0.05;\ t_{comp} = 3.33;\ df = 24;\ t_{crit} = 2.064$; the null hypothesis is rejected, thus at the 5% level of significance, we can claim that the new variety has the higher grain yield.

15)

a) $H_0: M = 57.62$.
b) $t = -0.612$.
c) $\alpha = 0.05$.
d) $p = 0.548$.
e) At a 5% level of significance the population mean does not differ significantly from the 57.62 mean.

4.5 INDEPENDENT SAMPLES T-TEST

16) The solution is obtained by calculating the sample mean and sample standard deviation for both varieties and then using the formulas given in the section 1.3.2:
$H_0: M_1 = M_2$; $H_1: M_1 \neq M_2$; $\alpha = 0.05$; $t_{comp} = -0.294$; $\mathrm{df} = 8$; $t_{crit} = 2.306$; we cannot reject the null hypothesis, thus we conclude that the varieties do no differ significantly with respect to grain yield. You can check the obtained results by using the t-test calculator (choose a unpaired t-test).

17) Apply the formulas given in section 1.3.2 and you get the following:
$H_0: M_1 = M_2$; $H_1: M_1 \neq M_2$; $\alpha = 0.05$; $t_{comp} = 1.457$; $\mathrm{df} = 89$; $t_{crit} = 1.99$; the null hypothesis is retained, we cannot say that the chemical fertilizer produces a statistically significant higher yield.

18)

a) $H_0: M_1 = M_2$.
b) $t = -2.364$.
c) $\alpha = 0.05$.
d) $p = 0.030$.
e) The yields differ significantly.

4.6 PAIRED SAMPLES T-TEST

19) First calculate the differences and then use the formulas from 1.3.2:
$H_0: M_1 - M_2 = 0$; $H_1: M_1 - M_2 \neq 0$; $\alpha = 0.05$; $t_{comp} = -0.357$; $\mathrm{df} = 7$; $t_{crit} = 2.365$; we cannot conclude that a significant difference exists. You can verify the obtained results by using the t-test calculator (choose a paired t-test).

20)

a) $H_0: M_1 - M_2 = 0$
b) $t = -2.63$.
c) $\alpha = 0.05$.
d) $p = 0.034$.
e) There is a statistical difference between standing and supine blood pressures

4.7 ANALYSIS OF VARIANCE

21) Apply the formulas given in section 1.3.3, and you get the following:
$H_0: M_1 = M_2 = M_3$; $H_1: M_i \neq M_j$ for some $i, j \in \{1,2,3\}$; $\alpha = 0.05$; $F_{comp} = 0.496$; $df1 = 2$; $df2 = 12$; $F_{crit} = 3.885$; we cannot reject the null hypothesis, thus we conclude that the varieties do not differ significantly with respect to grain yield. You can verify the obtained results by using the one-way ANOVA calculator.

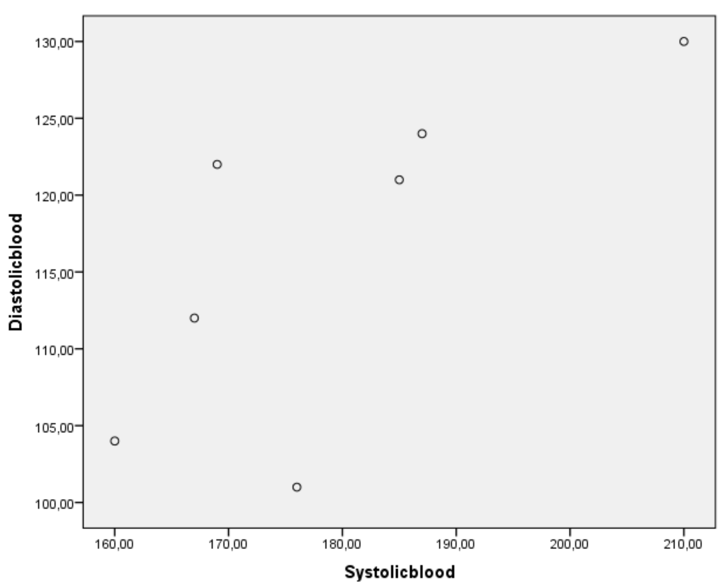
22)

a) $H_0: M_1 = M_2 = M_3 = M_4$.
b) $F = 14.218$.
c) $\alpha = 0.05$.
d) $p = 0.000$.
e) We have evidence at $\alpha = 0.05$ that at least two means for weight loss differ significantly.

4.8 PEARSON' S CORRELATION COEFFICIENT

23) Apply the formula given in section 1.3.4, and you get the following:
$H_0: \rho = 0$; $H_1: \rho \neq 0$; $\alpha = 0.05$; $r_{xy} = 0.726$; $r_{crit} = 0.754$; we cannot reject the null hypothesis, at a 5% level of significance, the evidence is not strong enough to indicate any correlation between the systolic and diastolic blood pressures. The above can be checked by using the Pearson correlation coefficient calculator.

Scatter diagram:



24) $H_0: \rho = 0$; $H_1: \rho \neq 0$; $\alpha = 0.01$; $r_{xy} = 0.5$; $r_{crit} = 0.479$; at 1% level of significance we can reject the null hypothesis and conclude that a statistically significant positive linear correlation exists in the population.

25)

a) $H_0: \rho = 0$.
b) $r_{xy} = 0.896$.
c) $\alpha = 0.05$.
d) $p = 0.000$.
e) At $\alpha = 0.05$ (and also $\alpha = 0.001$) we can claim that a significant positive linear correlation between the variables exists.

4.9 SPEARMAN' S CORRELATION COEFFICIENT

26) Convert the raw data in each variable into ranks and apply the formula given in section 1.3.4 and you get the following:

$H_0: \rho_S = 0$; $H_1: \rho_S \neq 0; \alpha = 0.05; r_s = 0.937$ $r_{crit} = 0.587$; at 5 % level of significance we can reject the null hypothesis and conclude that a statistically significant positive correlation exists between the maximum monthly averaged daily temperature and monthly averaged sunshine hours. The above can be checked by using the Spearman rank-order correlation coefficient calculator.

27) By applying the formula given in section 1.3.4 , you get the following:

$H_0: \rho_S = 0$; $H_1: \rho_S \neq 0; \alpha = 0.05; r_s = 0.903$; $r_{crit} = 0.648$; at 5 % level of significance we can reject the null hypothesis and conclude that there is convincing evidence of agreement.

28)

a) $H_0: \rho_S = 0$.
b) $r_S = 0.672$.
c) $\alpha = 0.01$.
d) $p = 0.000$.
e) At $\alpha = 0.01$ (also $\alpha = 0.001$) we can claim that there is a significant positive correlation between the variables. It turns out that an increase of agricultural land in use affects the increase in gross value added.

4.10 CHI-SQUARED TEST

29) By applying the formulas from 1.3.4, you obtain:

H_0: In this population, the two categorical variables are independent; H_1: In this population, the two categorical variables are dependent;$\alpha = 0.05$; $\chi^2_{comp} = 125.024$; $df = 6$; $\chi^2_{crit} = 12.592$; a sale of different varieties of apples and different properties are dependent at level 0.05. You can use chi-squared test calculator to check the answer.

30) By applying the formulas from section 1.3.4, you obtain:

H_0: In the population, the two categorical variables are independent; H_1: In this population, the two categorical variables are dependent; $\alpha = 0.05$; $\chi^2_{comp} = 5.774$; $df = 2$; $\chi^2_{crit} = 5.991$; at a 0.05 level of significance we cannot claim that insects are attacking one of the species more than the others.

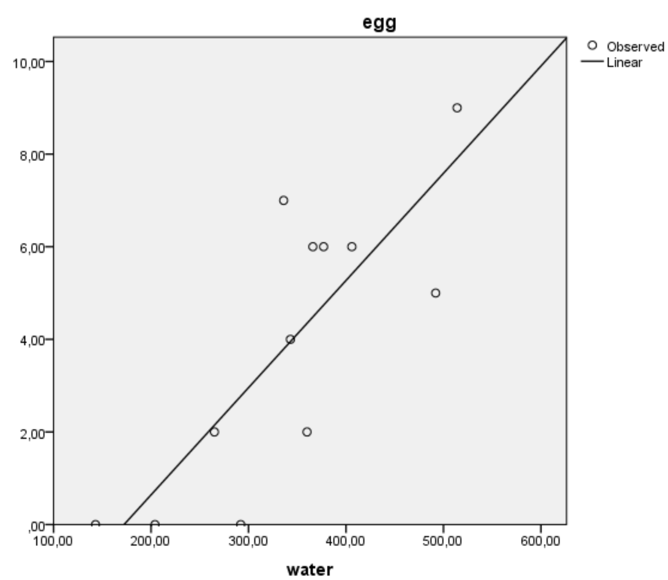
31) By applying the formula from section 1.3.4, you obtain:

H_0: In the population, the two categorical variables are independent; H_1: In this population, the two categorical variables are dependent; $\alpha = 0.05$; $\chi^2_{comp} = 2.645$; $df = 1$; $\chi^2_{crit} = 3.841$; at a 0.05 level of significance we cannot claim that variables are dependent. You can verify the results by applying a 2x2 contingency table calculator.

4.11 LINEAR REGRESSION

32)

a) Scatter diagram:



By applying the formulas from section 1.3.5, the following can be obtained:

b) $\hat{Y} = 0.023 \cdot X - 3.990$.

c) $x = 450$; $\hat{y} = 6.36$.

d) $r^2_{xy} = 0.653$.

e) 65.3 %.

f) 34.7 %.

g) $H_1: \rho \neq 0$; $\alpha = 0.01$; $F_{comp} = 18.831$; $df1 = 1$; $df2 = 10$; $F_{crit} = 10.04$; we can accept the hypothesis that $\rho \neq 0$, thus, water intake linearly increases the eggs production.

h) $H_1: \beta \neq 0$; $\alpha = 0.01$; $t_{comp} = 4.34$; $df = 10$; $t_{crit} = 3.169$; we can accept the hypothesis that $\beta \neq 0$, thus, water intake linearly increases the eggs production.

33)

a) $\hat{Y} = 8.00 \cdot X + 487.33$.

b) $x = 21$; $\hat{y} = 655.33$.

c) $H_1: \beta \neq 0$; $\alpha = 0.01$; $t_{comp} = 5.367$; we can accept the hypothesis that $\beta \neq 0$, we can thus conclude that the amount of addition linearly increases the chickens' weight.

d) $87.8\ \%$.

5 REFERENCES

Brase, C.H., Brase, C.P. Understandable statistics (8th ed.). Boston, Houghton Mifflin, 2006.

Graphpad.com. (2018). Analyze a 2x2 contingency table. [online] Available at: https://www.graphpad.com/quickcalcs/contingency1.cfm [Accessed 8 May 2018].

Graphpad.com. (2018). One sample t test. [online] Available at: https://www.graphpad.com/quickcalcs/oneSampleT1/ [Accessed 8 May 2018].

Graphpad.com. (2018). t test calculator. [online] Available at: https://www.graphpad.com/quickcalcs/ttest1.cfm [Accessed 8 May 2018].

Mathportal.org. (2018). Normal distribution calculator. [online] Available at: https://www.mathportal.org/calculators/statistics-calculator/normal-distribution-calculator.php [Accessed 8 May 2018].

Mathsisfun.com. (2018). Confidence interval calculator. [online] Available at: https://www.mathsisfun.com/data/confidence-interval-calculator.html [Accessed 8 May 2018].

Metoffice. gov.uk. (2018) [online] Available at: https://www.metoffice.gov.uk/public/weather/climate/gcpn7mp10 [Accessed 8 May 2018].

Palaniswamy, U.R., Palaniswamy, K.M. Handbook of statistics for teaching and research in plant and crop science. Binghamton, NY: The Haworth Press, Inc, 2006.

Shayib, M.A. Apllied Statistics 1 st edition. (2013). [online] Available at: http://thuvienso.bvu.edu.vn/bitstream/TVDHBRVT/15780/1/Applied-Statistics.pdf [Accessed 8 May 2018].

Sheskin, D.J. Handbook of parametric and nonparametric statistical procedures (2nd ed.). Boca Raton, Chapman & Hall /CRC, 2000.

Slovenian Environment Agency – ARSO. (2018) [online] Available at: http://www.arso.gov.si/en/ [Accessed 8 May 2018].

Socscistatistics.com. (2018). Chi-square test calculator. [online] Available at: http://www.socscistatistics.com/tests/chisquare2/Default2.aspx [Accessed 8 May 2018].

Socscistatistics.com. (2018). One-way ANOVA calculator. [online] Available at:

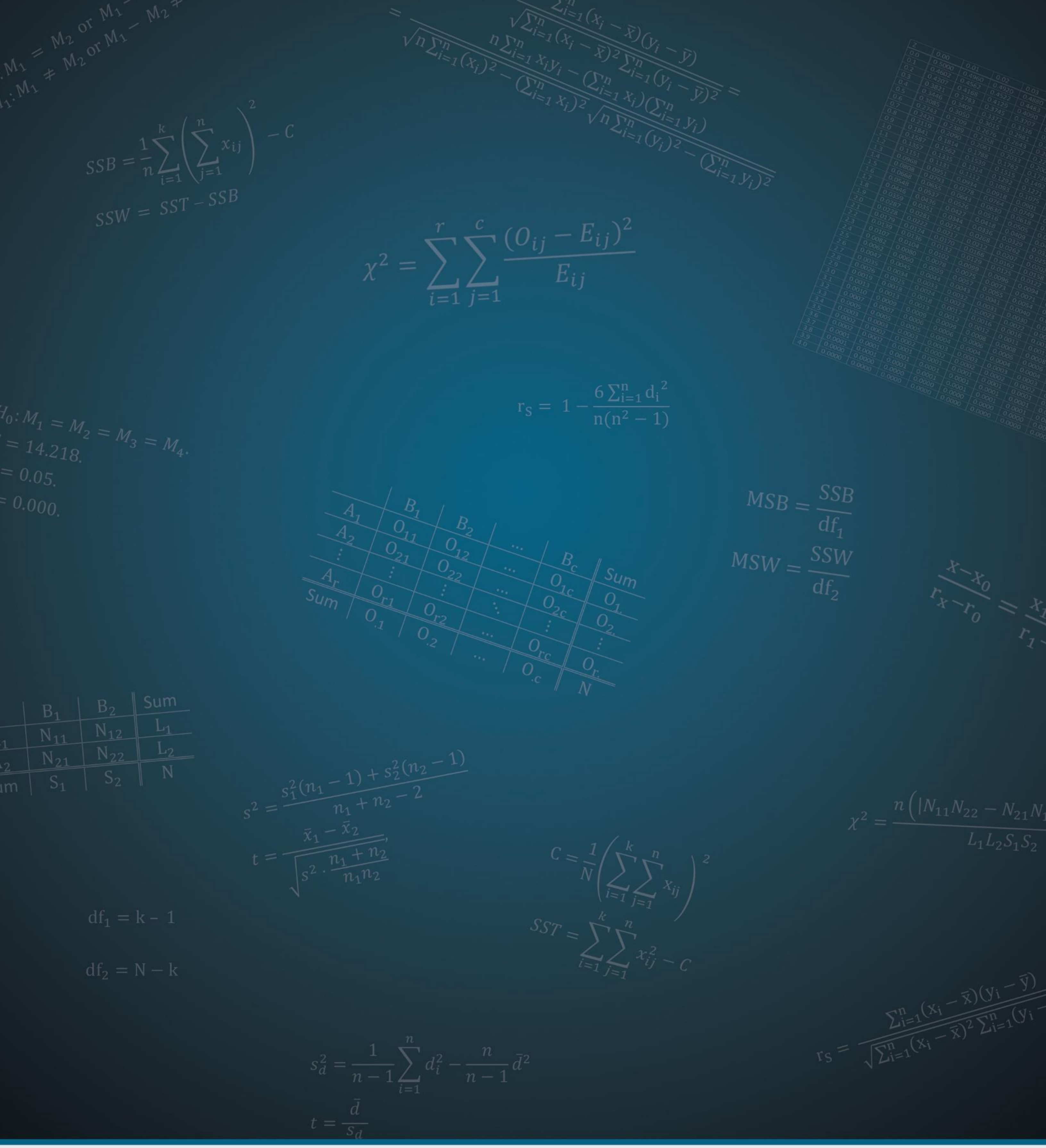
http://www.socscistatistics.com/tests/anova/default2.Aspx [Accessed 8 May 2018].

Socscistatistics.com. (2018). Pearson correlation coefficient calculator. [online] Available at: http://www.socscistatistics.com/tests/pearson/Default2.aspx [Accessed 8 May 2018].

Trpin Šikart, D. Analysis of the results of agricultural holdings in Slovenia processed according to FADN methodology. Master thesis, Faculty of agriculture and life sciences, 2016.

Vassarstats.net. (2018). Spearman rank-order correlation coefficient. [online] Available at: http://vassarstats.net/corr_rank.html [Accessed 8 May 2018].

6 INDEX



University of Maribor

Faculty of Agriculture
and Life Sciences