Test Mathematical Statistics
Module 5 Statistics and Analysis

Bachelor 2 Applied Mathematics

Module code:	201400218
Date:	Friday the 27th of October 2017
Time:	8:45 - 11:45 hrs.
Module coordinator:	Dr. P.K. Mandal
Lecturer	Dr. K. Poortema

Allowed: a simple scientific calculator, **not** a graphical one (GR)

Annexes:
Formula sheet Mathematical Statistics (2 pages),
Separately:
Standard normal table
t-table,
Chi squared table,
Tables for F distribution (4 pages),
Tables for Shapiro-Wilk (3 pages),
Tables for binomial distribution (3 pages),
Tables for Poisson distribution (2 pages).

Grading:

1	2a	2b	2c	3	4	5a	5b	5c	6a	6b	6c	7
4	1	2	4	4	3	2	2	3	2	3	1	5

Total: 36 points. Test grade = (# of points + 4)/4, rounded at one decimal.

Exercise 1

A double blind experiment was carried out to investigate the effect of the stimulant caffeine on a performance on a simple physical task. Twenty male college students were trained in finger tapping. They were then divided at random into two groups of 10 and the groups received different doses of caffeine (0 and 100 mg). Two hours after treatment each man was required to do finger tapping and the number of taps was recorded. Does caffeine affect performance on this task?

Group 0 (0 mg caffeine):

242	245	244	248	247	248	242	244	246	242

Group1 (100 mg caffeine)

248	246	245	247	248	250	247	246	243	244

Investigate first whether the data are normally distributed. Use the following information with respect to both groups. Use **all six parts** of the output, without consulting a probability table.

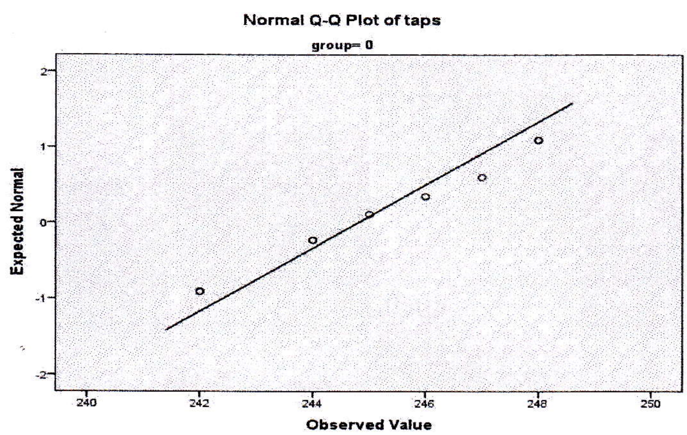
Descriptives[a]

		Statistic	Std. Error
taps	Mean	244,80	,757
	Std. Deviation	2,394	
	Skewness	,112	,687
	Kurtosis	-1,548	1,334

a. group = 0

Note that in SPPS the target value of the kurtosis is 0, because the kurtosis has been replaced by the kurtosis minus 3.

	Shapiro-Wilk (group 0)		
	Statistic	df	Sig.
taps	,889	10	,166

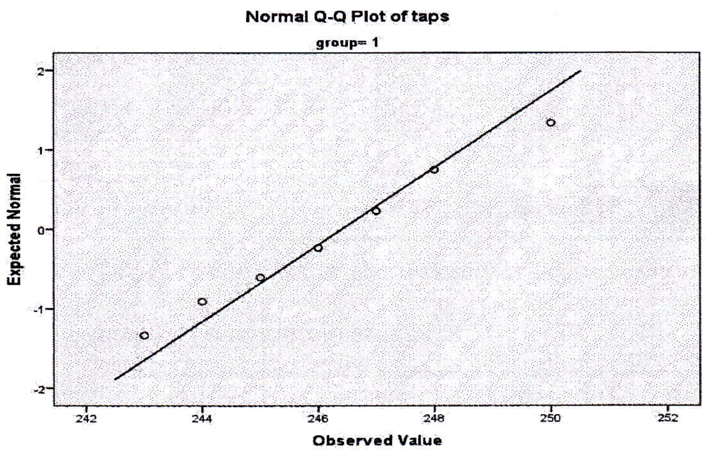


Descriptives[a]

		Statistic	Std. Error
taps	Mean	246,40	,653
	Std. Deviation	2,066	
	Skewness	-,011	,687
	Kurtosis	-,119	1,334

a. group = 1

	Shapiro-Wilk (group 1)		
	Statistic	df	Sig.
taps	,981	10	,971



Exercise 2

We again study the data of exercise 1.

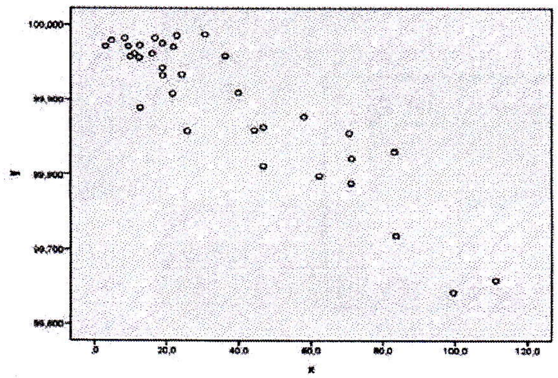
a. Assume normal distributions. Let S_X^2 (group 0) and S_Y^2 (group 1) denote the two sample variances, and let σ_X^2 and σ_Y^2 denote the true variances. Show that the ratio $\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$ has the F distribution with 9 and 9 degrees of freedom.

b. Construct a 95% confidence interval for σ_X^2/σ_Y^2 .

c. Test whether the two groups differ systematically with respect to the average number of finger taps in order to answer the question 'Does caffeine affect performance on this task?'

Exercise 3

The data give, for 34 batches of peanuts, the average of aflatoxin (parts per billion) in a mini-lot sample of 120 pounds of peanuts (x) and the percentage of noncontaminated peanuts in the batch (Y). These data are plotted in the next scatter plot.



Regression output obtained by SPSS is as follows:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	,239	1	,239	154,619	,000[b]
	Residual	,049	32	,002		
	Total	,289	33			

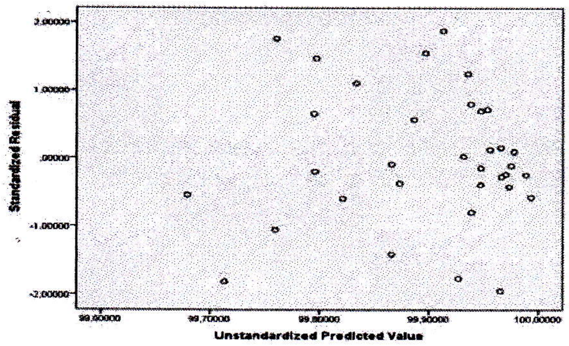
a. Dependent Variable: y

b. Predictors: (Constant), x

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	100,002	,011		9184,910	,000
	x	-,003	,000	-,910	-12,435	,000

a. Dependent Variable: y



Investigate whether Y really depends on x by using a statistical test. Choose $\alpha = 1\%$, give all eight steps of the testing procedure.

Exercise 4

Heart disease	Snore				Total
	Non-snorers	Occasional snorers	Snore nearly every night	Snore every night	
Yes	24	35	21	30	110
No	1355	603	192	224	2374
Total	1379	638	213	254	2484

The data come from a report of a survey which investigated whether snoring was related to various diseases. One sample of $n = 2484$ was realized. Those surveyed were classified according to the amount they snored, on the basis of reports from their partners. These particular data relate to the presence or absence of heart disease. The question is whether the presence or absence of heart disease depends on the amount of snoring. Conduct a statistical test in order to answer the question. Use level of significance 5% and give the eight steps of a test.

Exercise 5

We observe independent random variables $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ which are distributed as follows:

$P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$

for some unknown probability p, with $0 < p < 1$. We want to estimate the parameter $p(1-p)$, which is the variance of the observations X_i. We define $X = X_1 + X_2 + \cdots + X_n$.

a. Show that X is a sufficient statistics.

b. Show that X is a complete statistic.

c. Does the unique Minimum Variance Unbiased (MVU) estimator of $p(1-p)$ exist and is it given by the estimator $\frac{X}{n-1} - \frac{X^2}{n(n-1)}$? Motivate/explain your answer.

Exercise 6

We observe independent random variables $X_1, X_2, \ldots, X_n$ distributed according to density

$f(x \mid \lambda) = \frac{1}{2}\lambda \exp(-\lambda\,|x|)$ for $x \in \mathbb{R}$ (for all real numbers x),

where $\lambda > 0$.

a. Show that $1/\left(\frac{1}{n}\sum_i |X_i|\right)$ is the maximum likelihood estimator of λ.

b. Derive the most powerful (MP) test for testing H_0: $\lambda = 1$ against H_1: $\lambda = \frac{1}{2}$, using level of significance 5%. Use a normal approximation for the calculation of the critical value c.

c. Consider now testing of H_0: $\lambda = 1$ against H_1: $\lambda < 1$. Does there exist an uniformly most powerful (UMP) test? Explain why, or why not.

Exercise 7

We observe independent random variables $X_1, X_2, \dots, X_n$ $(n \geq 2)$ which all are distributed according to a normal distribution with unknown expectation μ and unknown variance σ^2 . Prove that $(n-1)S^2/\sigma^2$ is distributed according to a chi square distribution with $n-1$ degrees of freedom, with S^2 being the sample variance, $S^2 = \sum_i (X_i - \bar{X})^2/(n-1)$.

Hints:

Use properties of the multivariate normal distribution.

Use an orthonormal basis $u_1, u_2, \dots, u_n$ for n-dimensional vectors and define the first vector in an appropriate way.