

Test Statistics for BIT and TCS (Module 6 -201700269)

December 22, 2017, 13.45-16.00 hr in SC. Lecturer Dick Meijer, module-coordinator Dennis Reitsma.

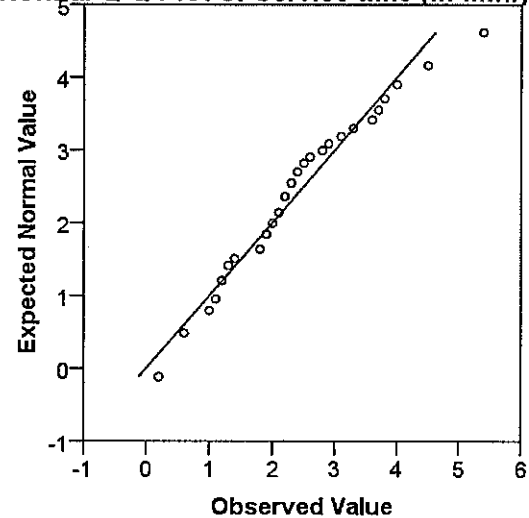
This test consists of 4 exercises and a formula sheet. Probability tables are provided in a separate document. An ordinary scientific calculator is allowed, not a programmable one (GR)..

1. In a survey on the effectivity of a helpdesk the length of the service times of customers, who call the helpdesk with problems, was assessed. Below you will find the observed service times (in tenths of minutes) of a sample of 38 customers. A numerical summary and the normal Q-Q plot are given as well:

0.2	0.6	0.6	1.0	1.1	1.2	1.2	1.2	1.3	1.4
1.8	1.8	1.9	1.9	1.9	2.0	2.1	2.1	2.1	2.2
2.2	2.2	2.3	2.3	2.4	2.4	2.5	2.6	2.8	2.9
3.1	3.3	3.6	3.7	3.8	4.0	4.5	5.4		

Sample -	
size	38
mean	2.25
standard deviation	1.11
skewness coefficient	0.69
kurtosis	3.74

Normal Q-Q Plot of Service time (in min.)



- Determine the 40th and the 80th percentiles of these observations.
 - Determine the median and compare it to the sample mean: what does the difference tell you about (the distribution of) the data?
 - Is the assumption of a normal distribution for the service times reasonable? Comment on this assumption based on each of the following aspects:
 - The numerical summary.
 - The Q-Q-plot.
 - The observed value of Shapiro Wilk's $W = 0.960$
 - Determine a 95%-confidence interval for the expected service time, assuming a normal model for the service times.
 - Determine a 95%-confidence interval for the variance of the service times.
2. Marketing experts investigated the claim that cola-drinkers in majority prefer the taste of Coca Cola (CC) over the taste of Pepsi. A random sample of 1000 cola-drinkers was asked to taste both cola's (blindly) and choose which of the two tastes best. The experts determined in advance that if at least 527 out of 1000 cola-drinkers prefer CC, this would be sufficient proof in favour of the claim.
- Define the test statistic X and give the probability model and the hypotheses of the described test.
 - Show that the given rejection region of the test has been determined using a 5% level of significance.
 - Compute the power of the test if in the population 55% of the cola drinkers prefer Coca Cola.
- For parts d. and e. of the exercise you can use the sample result of 550 preferences for CC (out of 1000).

- d. Determine the p-value of the test based on this result.
- e. Determine the width of the 95%-confidence interval for the proportion of CC-preferences and then compute the sample size, necessary to reduce the width of the interval to (at most) 0.02
- f. Show that $\hat{p} = \frac{x}{1000}$ is an unbiased estimator of p and express its Mean Squared Error in p .
3. Researchers were interested in comparing the long-term psychological effects of being on a high-carbohydrate, low-fat (LF) diet versus a high-fat, low carbohydrate (LC) diet. A group of 47 overweight and obese participants in the survey were randomly assigned to one of these energy-restricted diets. In week 52 the mood of the 21 LC dieters and the 26 LF dieters was assessed using a “total mood disturbance score” (TMDS). The results are shown in the table below.

Group	n	\bar{x}	s
LC	21	37.3	28.3
LF	26	19.3	20.8

- a. Test with $\alpha = 5\%$ whether the variances of the scores are different. Give (only) 1. The appropriate test statistic and its observed value.
2. The Rejection Region
3. Your conclusion.
- b. Is there a difference in expected TMDS-score in week 52? Give all steps of the testing procedure with $\alpha = 5\%$.
- c. Which alternative test would you choose in b. if the assumption of a normal distribution does not hold? Give the appropriate test and determine its distribution under H_0 (for these data)..
4. In a survey on customer satisfaction the energy provider Power4People distinguished two groups of new customers: customers contracted via the website “Independer” and other (new) customers. After 6 months random samples of both groups were asked if they seriously consider extending their energy contract after 1 year. 160 of the 250 Independer customers seriously considered extension and 110 of the 200 other customers considered extension as well.
- a. Use a binomial test to show whether the proportions of intended extension of the contract are different for the two groups, at a 5% significance level.
- b. The test in a. can be conducted, using a Chi square test: determine for this test:
1. the table with observed and expected values,
2. compute the value of the test statistic and
3. give the appropriate Rejection Region ($\alpha = 5\%$).

Grading: test result = $1 + \frac{\# \text{ points}}{48} \times 9$,
rounded at 1 decimal.

1					2						3			4		Tot
a	b	c	d	e	a	b	c	d	e	f	a	b	c	a	b	
2	2	4	3	3	2	3	2	2	3	2	3	6	2	6	3	48