Test Statistical Techniques for BIT-TCS (Module 6 -201800421) , Thursday the 20th of December 2018, 8.45-11.00 h. Lecturer Dick Meijer, module-coordinator Dennis Reidsma

This test consists of 5 exercises. A formula sheet and the probability tables are provided.
An ordinary scientific calculator is allowed, not a programmable one (GR).

1. The effectiveness of a helpdesk is assessed by researchers, by measuring the service times of customers who asked for support. The researchers intend to estimate the expected service time, using a confidence interval, and want to compare the mean to the bench mark for helpdesks.
In the table below you will find the observed service times (the order statistics) of 30 customers.

0.20	0.62	1.02	1.08	1.23	1.24	1.45	1.80	1.85	1.91
1.93	2.10	2.11	2.21	2.24	2.26	2.37	2.41	2.49	2.57
2.94	3.10	3.34	3.69	3.81	4.52	4.67	5.22	5.76	6.44

The numerical summary of the observations is given below (SPSS-output): note that the reported value of the kurtosis is, as usual in SPSS, "Kurtosis – 3".

Descriptive Statistics

	N	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Service Time (sec.)	30	2,6193	1,50601	2,268	0,912	0,427	0,486	0,833
Valid N (listwise)	30							

a. Compare the mean and the median: what does the difference tell you about the distribution of the service times?
b. Determine the 5-numbers-summary of the data set and use the quartiles to check whether there are outliers.
c. Assess the assumption of normality of the service times, using the numerical summary.
d. In the SPSS-output the value of Shapiro-Wilk's test statistic was reported: $W = 0.930$.
1. Give the null and the alternative hypothesis for this test.
2. Determine the coefficients $a_1$and a_{29} in the formula of Shapiro Wilk's W.
3. Determine the rejection region of the test.
4. What is your conclusion (in words) at a 5% significance level?

e. Determine a 95%-confidence interval for the expected service time (assuming normality) and give the proper interpretation of the numerical interval.

2. Do students at the UT in majority support the opinion that English should be the sole official language at their university. In a random sample of 200 UT-students 111 supported the opinion.

a. Conduct an appropriate test to verify whether the sample proves that a majority of all UT-students supports the opinion. Apply the testing procedure with $\alpha = 5\%$, using the **p-value** of the test.
b. Determine, additionally, the rejection region of the test in a.
c. Determine the power of the test in a. (*If you could not find the rejection region, use* $\{112, 113, \dots\}$)

3. A computer scientist is investigating the usefulness of two different design languages in improving programming tasks. Sixteen expert programmers, familiar with both languages, are asked to code a standard function in both languages, and the time (in minutes) is recorded. The data follow:

Expert	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	$\overline{x}$	s
Language 1	17	16	21	14	27	19	18	15	18	24	16	14	21	24	13	20	18.56	4.05
Language 2	18	14	19	11	18	18	13	12	23	21	10	14	19	23	12	17	16.38	4.18
Difference	**−1**	+2	+2	+3	+9	+1	+5	+3	**−5**	+3	+6	0	+2	+1	+1	+3	2.19	3.08

a. Assume for this part that normal distribution can be assumed for the observed times.
Conduct an appropriate (parametric) test in 8 steps at a 1% significance level, to check whether there is a structural difference in time needed to complete the task in language 1 and in language 2.

b. If the computer scientist considers it unreasonable to assume the normal distribution for the observed values, which test would you advise him as an alternative for the test in a.
Additionally, (only) define the test statistic of this test and give its observed value, the hypotheses and the distribution of the test statistic if H_0 is true.

4. Suppose we have two independent random samples, both with sample size 12.
We want to test whether there is a difference in the population means, but to apply the two independent samples t-test we need to check the assumption of normality for both samples and the assumption of equal variances.

a. Which test should we use to check the equality of variances?
Give for this test the rejection region if $\alpha = 5\%$ (use the given sample sizes 12).

b. For the test in **a.** a p-value of 1.2% is reported: what conclusion can you draw from that result?

c. Considering the result of the test on normality, it is obvious that the normal distribution does not apply for one of the samples. Which alternative non-parametric test can we apply here? Give the formula for the appropriate test statistic and determine the distribution under H_0 (if both sample sizes are 12).

5. The yearly return (in %) of state bonds is modelled as a normally distributed variable with unknown expected return μ and standard deviation $\sigma = \mu$. Based on a random sample of four returns X_1, X_2, X_3 and X_4 we consider two estimators of μ: $T_1 = \overline{X} = \frac{1}{4}\sum_{i=1}^{4} X_i$ and $T_2 = \frac{1}{5}\sum_{i=1}^{4} X_i$.

a. Are T_1 and T_2 unbiased estimators of μ? Motivate both answers.

b. Determine the mean squared errors of T_1 and T_2

c. Is T_2 a better estimator of μ than T_1?

Grade $= 1 + \frac{\#\text{ points}}{45} \times 9$,

rounded at 1 decimal

1					2			3		4			5			Tot
a	b	c	d	e	a	b	c	a	b	a	b	c	a	b	c	
2	3	1	4	4	6	2	2	6	4	2	2	2	2	2	1	45

Solutions

Exercise 1

a. The median $M = \frac{X_{(15)}+X_{(16)}}{2} = \frac{2.24+2.26}{2} = 2.25$ and $\overline{x} = 2.6193$

$\overline{x} > M$ indicates skewness to the right (which is confirmed numerically: the skewness coefficient > 0)

b. $Q_1 = x_{(8)} = 1.80$, since 25% of 30 is 7.5, similarly $Q_3 = x_{(23)} = 3.34$. So $IQR = Q_3 - Q_1 = 1.54$.

The 5-number-summary is $X_{(1)}, Q_1, M, Q_3, x_{(30)}$: 0.20, 1.80, 2.25, 3.34, 6.44

$(Q_1 - 1.5\times IQR, Q_3 - 1.5\times IQR) = (-0.51, 5.65)$, so both 5.76 and 6.44 are outliers.

c. Since the skewness 0.912 indicates a slight skewness to the right and both the skewness and the kurtosis deviate from the normal reference value 0, we cannot assume a normal distribution.

d. 1. Test H_0: $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ ("the service times are normally distributed") against

H_1: $F(x) \neq \Phi\left(\frac{x-\mu}{\sigma}\right)$ "the service times are not normally distributed".

2. $a_1 = -0.4254$ and $a_{29} = 0.2944$

3. Rejection Region: $W \leq 0.927$

4. $W = 0.930$ does not lie in the RR, so we cannot reject H_0:

"At a 1% significance level we could not prove that the service times are not normally distributed".

e. Requested is a conf. int for μ, the expected service time. Using the formula sheet:

95%-CI(μ) $= (\overline{X} - c \cdot \frac{S}{\sqrt{n}}, \overline{X} + c \cdot \frac{S}{\sqrt{n}})$, where $n = 30, \overline{x} = 2.6193$, $s = 1.50601$ and

$c = 2.045$ such that $P(T_{30-1} \geq c) = \frac{1}{2}\alpha = 0.025$.

95%-CI(μ) $= \left(2.6193 - 2.045 \cdot \frac{1.50601}{\sqrt{30}}, 2.6193 + 0.5623\right) \approx (2.06, 3.18)$

Interpretation: "we are 95% confident that the expected service time is between 2.06 and 3.18 minutes".

Exercise 2

a. Evidently we have to apply a one-sample-binomial test:

1. X is the number of supporters of the opinion that English should be the only official language. $X \sim B(200, p)$ where p is the unknown population proportion.
2. Test H_0: $p = \frac{1}{2}$ against H_1: $p > \frac{1}{2}$ with α = 5%.
3. Test statistic: X
4. If H_0 is true, we have: $X \sim B\left(200, \frac{1}{2}\right)$,

   so approximately according to the CLT: $X \sim N(np_0, np_0(1-p_0)) = N(100, 50)$
5. Observed value of the test statistic: $x = 111$.
6. Reject H_0 if the p-value $\leq \alpha = 5\%$, where the p-value =

   $P(X \geq 111|H_0) \overset{\text{c.c.}}{=} P(X \geq 110.5|H_0) \overset{\text{CLT}}{\approx} P\left(Z \geq \frac{110.5-100}{\sqrt{50}}\right) \approx 1 - \Phi(1.48) = 6.94\%$
7. Since the p-value > 5% = α, we cannot reject H_0.
8. At a 5% level of significance the sample did not prove convincingly that the majority supports the opinion about English as official language at the UT.

b. We reject H_0 if $X \geq c$, where $P(X \geq c|H_0) \leq \alpha_0 = 0.05$.

$P(X \geq c|H_0) \overset{\text{c.c.}}{=} P(X \geq c - 0.5|H_0) \overset{\text{CLT}}{\approx} P(Z \geq \frac{c-0.5-100}{\sqrt{50}}) \leq 0.05$ and from $\Phi\left(\frac{c-100.5}{\sqrt{50}}\right) \geq 0.95$ it follows: $\frac{c-100.5}{\sqrt{50}} \geq 1.645$, so $c \geq 100.5 + 1.645\sqrt{50} \approx 112.1$: $\boldsymbol{c = 113}$

c. The power is $P(X \geq 113|H_1)$ and the power of the test is computed for any $p_1 > 0.5$ in H_1:

$\beta(p_1) = P(X \geq 113|p = p_1) \overset{\text{c.c.+CLT}}{=} P\left(Z \geq \frac{112.5-200p_1}{\sqrt{200p_1(1-p_1)}}\right) = 1 - \Phi\left(\frac{112.5-200p_1}{\sqrt{200p_1(1-p_1)}}\right)$

Instead, it is sufficient if you showed how to calculate the power for a specific value of p_1, such as $p = 0.6$.

$$\beta(0.60) = P(X \geq 113 | p = 0.60) \overset{\text{c.c.+CLT}}{=} P\left(Z \geq \frac{112.5-120}{\sqrt{200\times0.6\times0.4}}\right) \approx P(Z \geq -1.08) = \Phi(1.08) = 85.99\%$$

Exercise 3

a. We have **paired samples** in this case, since each expert produces two times (we will continue with the last line of the table).

1. The differences $(L1 - L2)$ $X_1, \ldots, X_{16}$ are independent and $N(\mu, \sigma^2)$-distributed.
2. Test $H_0: \mu = 0$ against $H_1: \mu \neq 0$ with $\alpha = 0.01$ (two-sided test).
3. Test statistic: $T = \frac{\overline{X}}{S/\sqrt{16}}$
4. Under H_0: $T \sim t_{15}$.
5. The observed value: $\overline{x} = 2.38$ and $s = 3.08$, so: $t = \frac{2.19}{3.08/\sqrt{16}} \approx 2.84$
6. Reject H_0 if $T \geq c$ or $T \leq -c$, here $c = 2.947$ such that $P(T_{15} \geq c) = \frac{1}{2}\alpha = 0.005$.
7. The observed $t = 2.84 < 2.947$ does not lie in the Rejection Region: we cannot reject H_0.
8. We could not prove at a 5% level of significance that there is a difference in expected time needed to program in language 1 and 2.

b. Assuming non-normality for the differences the sign test is an alternative testing procedure: there are 13 positive, 2 negative and 1 "undecided" differences (*the "zero" observation is cancelled!*)

1. $X =$ "the number of positive differences among the 15 non-zero differences $L1 - L2$".
2. Observed value $X = 13$.
3. Test $H_0: p = \frac{1}{2}$ (or with the median: $m = 0$) (*no structural difference*) against
   $H_1: p \neq \frac{1}{2}$ (or $m \neq 0$)) (*difference between the languages*) with $\alpha = 5\%$.
4. If H_0 is true, X is $B\left(15, \frac{1}{2}\right)$.

Exercise 4

a. The F-test: Two sided test: reject H_0 if $F \leq c_1$ or $F \geq c_2$, where
$c_2 = 3.47$ such that $P(F_{11}^{11} \geq c_2) = \frac{\alpha}{2} = 0.025$ and $c_1 = \frac{1}{3.47} \approx 0.29$.

b. The p-value 1.2% < $\alpha = 5\%$, so reject H_0.
At a 5% level of significance the data showed that the variances are different.

c. Wilcoxon's rank sum test: $W = \sum_{i=1}^{12} R(X_i)$,
W is approximately **normally distributed** (since $n_1 > 5$ and $n_2 > 5$) with:
$\mu_W = \frac{1}{2}n_1(N+1) = 150$ and $\sigma_W^2 = \frac{1}{12}n_1 n_2(N+1) = 300$ $(\sigma_W \approx 17.32)$

Exercise 5

a. $E(T_1) = E(\overline{X}) = \mu$: unbiased
$E(T_2) = \frac{1}{5}\sum_{i=1}^{4} E(X_i) = \frac{1}{5} \cdot 4 \cdot \mu = \frac{4}{5}\mu$: not unbiased

b. $T_1 = \overline{X}$ is unbiased, so the $MSE(\overline{X}) = var(\overline{X}) = \frac{\sigma^2}{n} = \frac{1}{4}\mu^2$

Since $var(T_2) = var\left(\frac{1}{5}\sum_{i=1}^{4} X_i\right) = \left(\frac{1}{5}\right)^2 \cdot \sum_{i=1}^{4} var(X_i) = \frac{1}{25} \cdot 4 \cdot \mu^2 = \frac{4}{25}\mu^2$

$MSE(T_2) = (ET_2 - \mu)^2 + var(T_2) = \left(\frac{4}{5}\mu - \mu\right)^2 + \frac{4}{25}\mu^2 = \frac{5}{25}\mu^2 = \frac{1}{5}\mu^2$

c. $MSE(T_2) = \frac{1}{5}\mu^2 < \frac{1}{4}\mu^2 = MSE(T_1) \implies$ T_2 is better than T_1.