Test Statistics for TCS/BIT (Module 6 -201800421) ,

Friday the 31st of January 2020, 8.45-11.00 h. Lecturer Dick Meijer, module-coordinator Randy Klaassen

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

1. A new interface for a smart (programmable) heating thermostat was designed by students: the aim was that users could intuitively program the weekly heating schedule without consulting the user`s guide.
30 potential users were asked to program a given schedule for the thermostat.
In the table below you find the ordered task completion times (TCT), in minutes.

2.28	2.29	2.29	2.41	2.44	2.45	2.56	2.62	3.05	3.21
3.22	3.26	3.32	3.37	3.44	3.46	3.88	4.28	4.35	4.42
4.54	4.55	4.57	5.05	5.09	5.13	5.29	5.83	5.97	7.34

a. Determine the 90th percentile of these measurements.

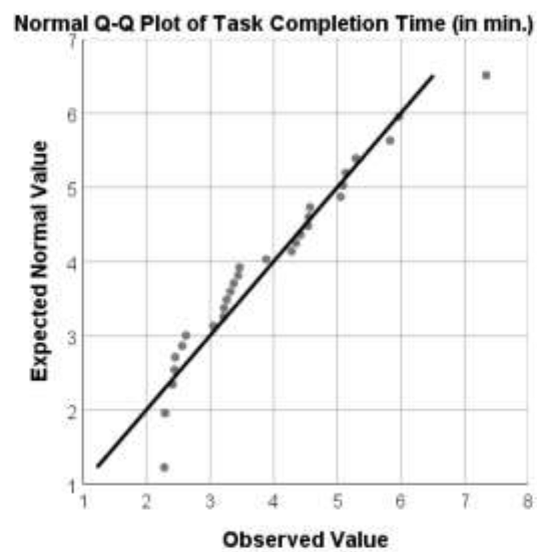
b. Determine the 5-number-summary of the observations and determine outliers, using the $1.5 \times IQR$-rule.

SPSS provided the following numerical summary (note that SPSS reports "**Kurtosis – 3**"), Shapiro Wilk`s test statistic and the normal Q-Q plot

Descriptive Statistics

	N	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Task Completion Time	30	3.8653	1.29561	1.679	0.715	0.427	0.121	0.833

Test on normality: Shapiro Wilk's $W = 0.929$



c. Can we assume a normal distribution for the observed task completion times?
Comment on:
1. The numerical summary,
2. The normal Q-Q plot and
3. The observed value of Shapiro Wilk's W

and draw your overall conclusion.

d. Assuming normality, give a 95%-confidence interval for the expected TCT and give the proper interpretation of the numerical interval.

2. An item of the Dutch 8 o`clock news on the 6th of August 2019 concerned the wearing of safety belts in cars. The presenter stated: "Traffic data revealed that last year (2018) 18 of the 58 deaths in car accidents did not wear a safety belt, whereas the year before (2017) only 11 out of 60 deaths did not wear a safety belt, although wearing a belt is compulsory." The interviewed official called the increase "substantial".

a. Is this increase also statistically significant at a 5% level? Conduct an appropriate test in 8 steps.

b. If we would test on the equality of the proportions against the inequality, a Chi-squared test is an alternative for the test, conducted in a.: give for this test (only) the test statistic and its observed value.

3. After some complaints about slow service in a fast food restaurant the management decided to compare the service times at this restaurant and another restaurant (of the same company) in the same town.
The observed 31 service times in the first restaurant were on average 80 seconds and the standard deviation was 8 seconds. In the second restaurant the mean of the 26 observed times was 73 seconds with a standard deviation of 6 seconds.

a. First test, with a 5% level of significance, whether assuming equal variances is allowed.
Only report: 1. The hypotheses
2. The test statistic and its observed value.
3. The rejection region
4. Your conclusion (in words).

b. Test whether there is a difference in mean service times between the two restaurants.
Use the appropriate parametric test (in 8 steps), with $\alpha = 5\%$.

c. Which non-parametric alternative would you apply for the test in b. ?
Give only the formula of the test statistic.

4. Is the (expensive) training for sales persons in a large insurance company effective?
Lately 10 of the sales persons were trained. For each of them the sales in a month time after the training were compared to the sales of the month before the training. The numbers of sold insurances were as follows:

Person	1	2	3	4	5	6	7	8	9	10
After	24	18	17	16	19	13	16	24	25	20
Before	22	19	16	15	16	12	15	24	23	18

The data analyst of the company advised **not** to use a parametric method (assuming normal distributions for the observations) to answer the question whether the training was effective.

a. Explain why you can support the choice of a non-parametric test in this case and which non-parametric test is appropriate.

b. Conduct the test in a. in 8 steps at a 5% level of significance.

c. Determine the power of the test in b. if in reality 80% of the sales persons improve their sales numbers.

5. A random number generator produces a random real number X with a unknown mean μ.
Assume that X has a uniform distribution on the interval $[0, 2\mu]$:then $E(X) = \mu$ and $var(X) = \frac{\mu^2}{3}$.
$X_1, \ldots, X_n$ is a random sample of these numbers.

a. Show that $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator of μ.

b. Consider the family of estimators of μ given by $T = a\overline{X}$, where a is a positive real number.
For which value of a is T the best estimator of μ within this family?

--- END ---

Grade $= 1 + \frac{\#\text{ points}}{46} \times 9$,
rounded at 1 decimal

1				2		3			4			5		Tot
a	b	c	d	a	b	a	b	c	a	b	c	a	b	
1	4	4	4	6	2	4	6	1	2	6	3	1	2	46

Solutions

Exercise 1

a. The 90th percentile is $\frac{x_{(27)}+x_{(28)}}{2} = \frac{5.29+5.83}{2} = 5.56$ since 90% of 30 is 27.

b. The minimum is 2.28 and the maximum 7.34. The median is $\frac{x_{(15)}+x_{(16)}}{2} = \frac{3.44+3.46}{2} = 3.45$

$Q_1 = x_{(8)} = 2.62$, since 25% of 30 is 7.5, similarly $Q_3 = x_{(23)} = 4.57$.

5-number-summary: 2.28, 2.62, 3.45, 4.57, 7.34

$IQR = Q_3 - Q_1 = 1.95$

$(Q_1 - 1.5\times IQR, Q_3 + 1.5\times IQR) = (-0.30, 7.50)$: this interval contains all measurements, no outliers

c. 1. Both the skewness coefficient and the kurtosis are close to 0 (reference values of the normal distribution), since the deviation is less than 2 standard errors, which makes the normality assumption reasonable.

2. The normal Q-Q plot shows some deviations from the line y=x, but the deviations do not seem to be large or systematic: normal distribution could be plausible.

3. Rejection Region: $W \leq 0.927$, from the Shapiro-Wilk table of critical values with $n = 30$ and $\alpha = 5\%$. $W = 0.929$ does not lie in the RR: at a 5% significance level we could not prove that the TCT's are **not** normally distributed.

All in all the normal distribution seems to be a good model for the TCT's.

d. Requested is a confidence interval for μ, the expected TCT. Using the formula sheet:

95%-CI(μ) $= (\overline{X} - c\cdot\frac{S}{\sqrt{n}}, \overline{X} + c\cdot\frac{S}{\sqrt{n}})$, where $n = 30$, $\overline{x} = 3.8653$, $s = 1.29561$

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

95%-CI(μ) $= \left(3.8653 - 2.045\cdot\frac{1.29561}{\sqrt{30}}, 3.8653 + 0.4837\right) \approx (3.38, 4.35)$

Interpr.: "We are 95% confident that the <u>expected</u> task completion time is between 3.38 and 4.35 minutes".

Exercise 2

a. We have a situation with two independent binomial counts:

1. X = "the number of times that the person did not wear a belt among 58 deaths in 2018" and Y = "the corresponding number among 60 deaths in 2017" are independent and $B(58, p_1)$ resp. $B(60, p_2)$.
2. Test $H_0: p_1 = p_2$ against $H_1: p_1 > p_2$ with $\alpha = 5\%$.
3. $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $\hat{p} = \frac{X+Y}{n_1+n_2}$
4. Under H_0, Z is $N(0,1)$.
5. Observed is $X = 18$ and $Y = 11$, so $\hat{p} = \frac{X+Y}{n_1+n_2} = \frac{29}{118}$ and $z = \frac{\frac{18}{58}-\frac{11}{60}}{\sqrt{\frac{29}{118}\cdot\frac{89}{118}\cdot\left(\frac{1}{58}+\frac{1}{60}\right)}} \approx 1.602$
6. Right-sided test: Reject H_0 if $Z \geq c$, where $\Phi(c) = 1 - \alpha = 0.95$: $c = 1.645$.
7. $z = 1.60 < 1.645 = c \implies$ do not reject H_0 .
8. At a 5% level of significance we cannot claim that the proportion of deaths without belt has increased.

b. The alternative is the **Chi-squared test on homogeneity** for the following 2×2 table: (***in a. the test is one sided: then no χ^2-test***!)

		Year		
		2018	2017	Total
Belt	No	18	11	29
	Yes	40	49	89
	Total	58	60	118

The expected values in case of homogeneity (N_{ij}`s as usual):

$\hat{E}_0(N_{11}) = \frac{29\times 58}{118} \approx 14.25$, $\hat{E}_0(N_{12}) \approx 29 - 14.25 = 14.75$,

$\hat{E}_0(N_{21}) = 58 - 14.25 = 43.75$ and $\hat{E}_0(N_{22}) \approx 45.25$

$$\chi^2 = \sum_{j=1}^{2}\sum_{i=1}^{2}\frac{\left(N_{ij}-\hat{E}_0 N_{ij}\right)^2}{\hat{E}_0 N_{ij}}$$

$$= \frac{(18-14.25)^2}{14.25} + \frac{(40-43.75)^2}{43.75} + \frac{(11-14.75)^2}{14.75} + \frac{(49-45.25)^2}{45.25} \approx 2.572$$

(*Note* 1. *the RR for $\alpha = 5\%$ is [3.84, ∞): H_0 is not rejected and* 2. (*in* b.) $\chi^2 \approx 2.572 \approx 2.566 = 1.602^2 = z^2$ (*in* a.))

Exercise 3

a. The F-test:

1. Test $H_0: \sigma_1^2 = \sigma_2^2$ against $H_1: \sigma_1^2 \neq \sigma_2^2$ with $\alpha = 5\%$
2. $F = \frac{s_X^2}{s_Y^2} = \frac{8^2}{6^2} = 1.78$
3. Two sided test: reject H_0 if $F \leq c_1$ or $F \geq c_2$, where $c_2 = 2.18$ such that $P(F_{25}^{30} \geq c_2) = \frac{\alpha}{2} = 0.025$

   and $c_1 = \frac{1}{2.11} \approx 0.47$, since $P(F_{30}^{25} \geq 2.11) = \frac{\alpha}{2} = 0.025$
4. 1.78 is not in the RR: at a 5% level of significance we could not prove that the variances are different: we may assume that they are the same.

b. 1. Model assumptions ("statistical assumptions"):

We have **two independent random samples** of service times here, $X_1, \dots, X_{31}$ drawn from a $N(\mu_1, \sigma^2)$-distribution for restaurant 1 and $Y_1, \dots, Y_{26}$ from a $N(\mu_2, \sigma^2)$-distribution for restaurant 2 (equal σ's!)

2. We will test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 \neq \mu_2$ with α = 5%
3. Test statistic $T = \frac{\overline{X} - \overline{Y}}{\sqrt{S^2\left(\frac{1}{31}+\frac{1}{26}\right)}}$ with $S^2 = \frac{30S_X^2 + 25S_Y^2}{31+26-2}$
4. T is under H_0 t-distributed with $df = n_1 + n_2 - 2 = 55$
5. Observed: $s^2 = \frac{30\times 8^2 + 25\times 6^2}{55} \approx 51.27$, so $t = \frac{80 - 73}{\sqrt{51.27\left(\frac{1}{31}+\frac{1}{26}\right)}} \approx 3.68$
6. This test is two-tailed: **reject H_0 if $T \leq -c$ or $T \geq c$.**

   where $c = 2.005$, taken from the t_{55}-table (average of the values in the t_{50}- and the t_{60}-tables).
7. $t = 3.68$ ($> c$) lies in the Rejection Region, so reject H_0.
8. The mean service times of the two restaurants are significantly different, at a 5% level of significance.

c. We could apply Wilcoxon's Rank Sam test as an alternative: test statistic $W = \sum_{i=1}^{31} R(X_i)$

Exercise 4

a. Clearly the observations are paired (dependent), since the sales number of each sales person are observed twice. A normal model for the differences is not correct since the differences are integers close to 0 and they attain only a few integer values. Hence a non-parametric approach seems better in this case.

Note: some students argue that, since we have only $n = 10$ differences, we will have to use the sign test instead of the t-test, but if applicable the t-test can also be used for small samples and is more powerful than the sign test!

b. Sign test on the differences of the sales numbers After – Before, +2, -1, +1, +1, +3, +1, +1, 0, +2, +2 :
8 positive and 1 negative differences, the 0-difference is cancelled: we are left with $n = 9$ differences.

1. The non-zero differences After – Before $X_1, \dots, X_9$ are independent and have the same distribution.
   X is the number of positive differences and p is the probability that after the training the sales numbers of a sales person increase.
2. Test $H_0: p = \frac{1}{2}$ against $H_1: p > \frac{1}{2}$ with $\alpha = 5\%$.
3. Test statistic: X (the number of positive differences)
4. Under H_0 X has a $B\left(9, \frac{1}{2}\right)$-ditribution
5. $X = 8$
6. Reject H_0 if the p-value $\leq \alpha$.

   The p-value of this test is $P(X \geq 8|H_0) = 1 - P(X \leq 7|p = 0.5) \overset{B(9,\ 0.5)-\text{table}}{\approx} 1 - 0.980 = 2.0\%$

   (Or use the binomial formula $P(X \geq 8|H_0) = P\left(X = 8|p = \frac{1}{2}\right) + P\left(X = 9\middle|p = \frac{1}{2}\right)$

   $= \binom{9}{8}\left(\frac{1}{2}\right)^9 + \left(\frac{1}{2}\right)^9 \approx 1.95\%$)
7. The p-value is less than $\alpha = 5\%$, so reject H_0.
8. At a 5% level of significance the observed sales number show that the training pays.

c. We first need to establish the RR "$X \geq c$": since $P(X \geq 8|H_0) = 2.0\% < \alpha = 5\%$ (b.) and $P(X \geq 7|H_0) = 1 - P(X \leq 6|p = 0.5) \overset{B(9,\ 0.5)-\text{table}}{\approx} 1 - 0.910 = 9.0\% > \alpha$, the RR is $X \geq 8$.
(*this rejection region can also be found by choosing the smallest integer such that* $P(X \geq c|H_0) \leq 5\%$)
The power of the test for $p = 0.8$ is:

$$P(X \geq 8|p = 0.8) = P(X = 8|p = 0.8) + P(X = 9|p = 0.8) = \binom{9}{8}0.8^8 0.2 + 0.8^9 \approx 43.6\%.$$

(*Or use the* $B(9, 0.2)$*-table for* $Y \sim B(9,0.2)$: $P(X \geq 8|p = 0.8) = P(Y \leq 1|p = 0.2) = 43.6\%$.)

Exercise 5

a. $E(\overline{X}) = \mu$, so $\overline{X}$ is an unbiased estimator.
(*It is not necessary to show that* $E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n\mu = \mu$).

b. Since $var(\overline{X}) = \frac{\sigma^2}{n}$, where $\sigma^2 = var(X) = \frac{(b-a)^2}{12}\ (formula) = \frac{(2\mu)^2}{12} = \frac{\mu^2}{3}$, we have Should be:

$$MSE(T) = \left(E(a\overline{X}) - \mu\right)^2 + var(a\overline{X}) = (a\mu - \mu)^2 + a^2 \cdot var(\overline{X}) = \mu^2\left[(a-1)^2 + \frac{a^2}{3}\right].$$

$g(a) = (a-1)^2 + \frac{a^2}{3}$ attains its extrema if $g`(a) = 2(a-1) + \frac{2}{3n}a = 0$, or $\frac{3n+1}{3n}a = 1$ or $a = \frac{3n}{3n+1}$.
This value of a minimizes $MSE(T)$, since $g``(a) = 2 + \frac{2}{3n} > 0$.