Test "Statistics" for IEM in module 3 (201300108) – April 16, 2020, 13.45-16.00 V6

Lecturer: Dick Meijer – module coordinator: Adina Aldea.

This online test consists of 5 exercises. You can use **the reader**, including the formula sheet and the probability tables. **Your own notes** are allowed as well. Other resources are not allowed. Use a common (non-programmable) calculator, not a graphical one (GR). The test should be hand written. Number the pages before copying and uploading them.

Above the resources for this test are mentioned. Read the following carefully:
By testing you remotely in this fashion, we express our trust that you will adhere to the ethical standard of behaviour expected of you. This means that we trust you to answer the questions and perform the assignments in this test to the best of your own ability, without seeking or accepting the help of any source that is not explicitly allowed by the conditions of this test.
Copy the following statement to your first answering sheet and **sign it with your name and student number**: "*I will make this test to the best of my own ability, without seeking or accepting the help of any source not explicitly allowed by the conditions of the test.*"

1. Because of health issues simple radon detectors are sold to customers, but how accurate are these detectors? To answer this question a researcher placed 12 of these detectors in a room that was exposed to a concentration of 105 picocurie per litre (pCi/ltr) radon. The measurements are mentioned below.

91.9	97.8	111.4	122.3	105.4	95.0
103.8	99.6	96.6	119.3	104.8	101.7

a. Determine: 1. Sample mean,
2. The sample standard deviation,
3. The median and
4. the interquartile range

b. Determine a 95%-confidence interval for the standard deviation of radon concentration measurements

c. Is there convincing evidence that measurements with this radon detector deviates structurally from the real value 105? Conduct a test in 8 steps at a 5% level of significance.

Consider for the exercises 2, 3 and 4 the following problems (indicated with A, B and C):

A. Do lecturer grade the same test significantly different?
Two math teachers, A and B, were asked to both grade the test of 10 students, using a given detailed grading, to assess whether systematic differences occurred. Below you find the results:

Student	1	2	3	4	5	6	7	8	9	10	**Mean**	**Standard deviation**
Grade A	3.5	4.5	8	6	7	6	5	7	5.5	6.5	5.9	1.33
Grade B	4	5	9	6.5	6.5	6.5	5.5	8	6.5	7	6.45	1.42
Difference	-0.5	-0.5	-1	-0.5	0.5	-0.5	-0.5	-1	-1	-0.5	-0.55	0.44

B. A cocaine addiction is hard to overcome. Addicted people need cocaine to feel comfortable. Hence researchers assumed that medicines against depressions might be suitable in a recovery therapy.
In a 3 years investigation of 72 chronic cocaine users the effect of using the antidepressant desipramine was compared to lithium and a placebo. Lithium is a standard drug used in treatments against cocaine addiction. A placebo is a fake drug to see what the effect is of participation in the programme without real medicine. The 72 users were randomly assigned to the 3 equally large treatment groups. Here are the results:

		Regression to cocaine addiction	
		Yes	No
Treatment	Desipramine	10	14
	Lithium	18	6
	Placebo	20	4

C. A marketing company supplies manufacturers with information about sales of their products, on the basis of samples amongst small shop owners. Marketing managers are inclined to simply look at estimates and neglect the possible errors in sampling. Assume we have a random sample of 75 of these small shops: the mean sales is reported to be 52 pieces in a month with a sample standard deviation of 12.8 pieces. The manager compared this to the result of a random sample of 53 shops last year when the average sales of 49 pieces was reported, with a samples standard deviation of 11.3 pieces. The increase from 49 to 52 (6%) made the manager euphoric (but is euphoria appropriate?)

2. First we will apply **parametric tests** to the problems A, B and C: these tests include binomial and t-tests for one and two samples, the F-test on the equality of variances and Chi-square tests on the variance and for one or two categorical variables (amongst which the test on independence and the test on homogeneity). Determine for each of the questions under a., b. and c. below (only!):

1. The type of test. Motivate your choice in one sentence.
2. The hypotheses
3. The test statistic and its observed value.
4. The rejection region at a 5% level of significance
5. The conclusion (reject H_0 or not)

a. Problem A: Is there a systematic difference in grading between the two teachers?
b. Problem B: Is the effect of the three treatments on the probability of regression different?
c. Problem C: Is the increase of the sales this year statistically significantly higher than last year?.

3. Referring to the test in exercise 2a. with respect to problem A:
a. Determine the p-value of this test and show that we can draw the same conclusion as in 2a.
b. What is the non-parametric alternative for this test? Include in your answer:

1. Define the appropriate test statistic for this alternative test and give its observed value.
2. Give the hypotheses.
3. Determine the p-value.
4. Draw your conclusion in words at a 5% level of significance.

4. Referring to the test in exercise 2c. (problem C)
a. Give all necessary statistical assumptions for the test in 2c. (the probability model).
b. Apply the test on the equality of variances to problem C. Give (only):

1. The hypotheses
2. The test statistic and its distribution under H_0.
3. Determine the rejection region for $\alpha = 5\%$ (you can round to the nearest table values).
4. Use the observed value to draw a conclusion on the assumption of equal variances.

5. Elections are coming up and a poll expert wants to determine the support of party A, that scored 20% in last elections. Let p be the current population proportion of party A voters and X the number of party A voters in a random sample of n voters.
a. Determine the sample size n such that the 90%-confidence interval for p has an error margin of 0.02 (assuming that the support is still around 20%).

The poll expert chooses a sample size of $n = 1000$ and states in advance that if the number of party A voters is at least 220 (22% in the sample, 2% more than the 20% of last elections), the sample shows convincingly that the support of party has increased.
b. Determine the probability of a type I error (or level of significance) of the described test.
c. Determine the probability of a type II error of the same test if in reality the current support is 23%.

-- END TEST --

Grade = $1 + \frac{\#\text{ points}}{50} \times 9$,
rounded on 1 decimal.

1			2			3		4		5			Tot
a	b	c	a	b	c	a	b	a	b	a	b	c	
4	3	6	5	5	5	3	4	2	4	3	3	3	50

Solutions
Exercise 1

a. Using the calculator: 1. Sample mean $\bar{x} = 104.133$
2. Sample standard deviation $s = 9.397$

Using the order statistics (below): 3. Median = mean of the 6th and 7th observation = 102.75
4. $IQR = Q_3 - Q_1 = 108.45 - 97.2 = 11.25$,

where $Q_1 = \frac{x_{(3)}+x_{(4)}}{2} = \frac{96.6+97.8}{2} = 97.2$ (25% of 12 is 3) and $Q_3 = \frac{x_{(9)}+x_{(10)}}{2} = \frac{105.5+111.4}{2} = 108.45$

Rank	1	2	3	4	5	6	7	8	9	10	11	12
Observation	91.9	95.0	96.6	97.8	99.6	101.7	103.8	104.8	105.5	111.4	119.3	122.3

b. We will use formula for σ^2: $\left(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}\right)$, where $P(\chi^2_{n-1} \leq c_1) = P(\chi^2_{n-1} \geq c_2) = \frac{\alpha}{2} = 0.025$

In this case $n = 12$ and $s = 9.397$ are given and in the χ^2_{12-1}-table we find $c_1 = 3.82$ and $c_2 = 21.92$

So a 95%-CI(σ) = $\left(\sqrt{\frac{12s^2}{c_2}}, \sqrt{\frac{12s^2}{c_1}}\right) = \left(\sqrt{\frac{12 \cdot 9.397^2}{21.92}}, \sqrt{\frac{12 \cdot 9.397^2}{3.82}}\right) \approx (6.95,\ 16.66)$

c. One sample t-test on μ = the expected radon concentration.

1. Model: the observed radon concentrations $X_1, \ldots, X_{12}$ are independent and all $N(\mu, \sigma^2)$-distributed with unknown μ and σ^2.
2. Test $H_0: \mu = 105$ against $H_1: \mu \neq 105$ with $\alpha = 5\%$.
3. Test statistic $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X} - 105}{S/\sqrt{12}}$
4. Under $H_0: T \sim t_{n-1}$
5. Observed $t = \frac{104.133 - 105}{9.397/\sqrt{12}} \approx -0.320$
6. **Reject H_0, if $T \geq c$ or $T \leq -c$**, where $c = 2.201$ such that $P(T_{12-1} \geq c) = \frac{1}{2}\alpha = 2.5\%$.
   (Or determine the p-value = $2 \times P(T \geq +0.320) > 2 \times 25\%$)
7. $t \approx -0.320$ is not in the RR (or p-value > α): we fail to reject H_0.
8. At a 5% level of significance the sample did not show that the observed Radon concentations deviate systematically from the real concentration.

Exercise 2

a. 1. We have paired samples here since one test (of one student) is graded twice: we will apply the one sample t-test on the expected difference in grading, using the row of differences in the table.
2. Test $H_0: \mu = 0$ against $H_1: \mu \neq 0$
3. Test statistic $T = \frac{\bar{X}}{S/\sqrt{10}}$, with observed value $t = \frac{-0.55}{0.44/\sqrt{10}} \approx -3.95$
4. Reject H_0 if $T \leq -c$ or $T \geq c = 2.262$ from the t_{10-1}-table at tail probability $\frac{1}{2}\alpha = 0.025$
5. $t \approx -3.95$ lies in the rejection region, so reject H_0 (*the grading by teachers A and B do deviate structurally).*

b. 1. In this case we compare the success rates of the treatments, using a random sample with fixed size (24) for each treatment: we will apply the test on homogeneity (of the binomial regression variables).
2. Test $H_0: p_{11} = p_{21} = p_{31}$ (the success rates are the same for the 3 treatments) against
H_1: the inequality of at least 2 of the 3 success rates p_{11}, p_{21} and p_{31}.
3. Test statistic $\chi^2 = \sum_{j=1}^{c}\sum_{i=1}^{r} \frac{(N_{ij} - \hat{E}_0 N_{ij})^2}{\hat{E}_0 N_{ij}}$, with $\hat{E}_0 N_{ij} = \frac{\text{row total} \times \text{column total}}{n}$

The expectations are given in the table below, e.g. $\hat{E}_0 N_{11} = \frac{24 \times 48}{72} = 16$

$\chi^2 = \frac{(10-16)^2}{16} + \cdots + \frac{(4-8)^2}{8} \approx 10.5$

4. Reject H_0 if $\chi^2 \geq c = 5.99$, using the χ^2-table with $df = (r-1)(c-1) = 2$ and $\alpha = 5\%$.
5. $\chi^2 \approx 10.5$ lies in the rejection region: H_0 is rejected (the effect of the treatments is different)

	Regression		
	Yes	No	Total

Treatment	Desipramine	Observed: 10 Expected: 16	Observed: 14 Expected: 8	24
	Lithium	Observed: 18 Expected: 16	Observed: 6 Expected: 8	24
	Placebo	Observed: 20 Expected: 16	Observed: 4 Expected: 8	24
	Total	48	24	72

c. 1. Since the observed means of both years were based on random samples, it is reasonable to assume we have independent random samples (the total number of shops is likely to be much larger than the sample sizes). Hence we can apply the 2 independent random samples t-test (assuming normality and equal variances) (*or alternatively the large samples z-test for two independent samples, since both sample sizes are larger than 40*)

2. Test $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$. (index 1 is for this year , 2 for last year)

3. $T = \frac{\overline{X}-\overline{Y}}{\sqrt{S^2\left(\frac{1}{75}+\frac{1}{53}\right)}} = \frac{52-49}{\sqrt{148.92\left(\frac{1}{75}+\frac{1}{53}\right)}} \approx 1.370,$

where $S^2 = \frac{n_1-1}{n_1+n_2-2}S_X^2 + \frac{n_2-1}{n_1+n_2-2}S_Y^2 = \frac{74}{126}\cdot 12.8^2 + \frac{52}{126}\cdot 11.3^2 \approx 148.92 \quad (s \approx 12.20)$

4. Reject H_0 if $T \geq c = 1.658$ from the t-table with $df = 120$, close to $df = n_1 + n_2 - 2 = 126$.

5. $t \approx 1.370$ is not in the RR: we failed to reject H_0 (*the sales increase is not significant at a 5% level*) (If you chose for the large sample approach the observed value $z = 1.40$ and the critical value $c = 1.645$ only differ slightly, not affecting the conclusion.)

Exercise 3

a. The p-value of this two-sided test, given the observed value -3.95 of T is: $2\times P(T \geq |-3.95|\,|H_0)$ is between 1% and 0.2%, Since $P(T_9 \geq 3.95)$ is between 0.005 and 0.001. It follows that the p-value is less than 5% = α: reject H_0, as in 3a. with the RR.

b. We will use the sign test as an alternative of the t-test on the differences:

1. Test statistic is $X =$"# of positive reductions", observed $X = 1$
2. Test $H_0: p = \frac{1}{2}$ against $H_1: p \neq \frac{1}{2}$
3. The p-value of this two-sided test is $2 \times P(X \leq 1|p = \frac{1}{2}) = 2 \times 0.011 = 2.2\%$
4. Since the p-value is less than α = 5%, we can conclude that at a 5% level; of significance the data contain sufficient evidence to state that teachers do grade differently.

Exercise 4

a. The sales of the shops this year $X_1, \dots, X_{75}$ and the sales of the shops last year $Y_1, \dots, Y_{53}$ are two independent random samples and $X_i \sim N(\mu_1, \sigma^2)$ and $Y_j \sim N(\mu_2, \sigma^2)$ (equal variances) (If you chose for the large samples Z-test, the normality assumptions nor the equal variances are necessary)

b. 1. Test $H_0: \sigma_X^2 = \sigma_Y^2$ against $H_1: \sigma_X^2 = \sigma_Y^2$

2. The test statistic $F = \frac{S_X^2}{S_Y^2} \sim F_{53-1}^{75-1}$ under H_0

3. Reject H_0 if $F \leq c_1$ or $F \geq c_2$, where $P(F_{52}^{74} \geq c_2) \approx P(F_{60}^{60} \geq c_2) = \frac{1}{2}\alpha = 0.025 \Rightarrow c_2 = 1.67$

and $P(F_{52}^{74} \leq c_1) = P\left(F_{74}^{52} \leq \frac{1}{c_1}\right) \approx P\left(F_{60}^{60} \leq \frac{1}{c_1}\right) = 0.025 \Rightarrow c_1 = \frac{1}{1.67} \approx 0.60$

4. $F = \frac{12.8^2}{11.3^2} \approx 1.28$ does not lie in the rejection region. Hence we did not prove that the variances of the sales this year and last year are are different, at a 5% level of significance.

Exercise 5

a. In the formula $\hat{p}0.8 \pm c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ the error margin $c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ should be less than 0.02.

Since for a 90% level of confidence $c = 1.645$, such that $\Phi(c) = 1 - \frac{1}{2}\alpha = 0.95$ and $\hat{p} \approx 0.2$, we have:

$$c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645\sqrt{\frac{0.16}{n}} < 0.02 \Rightarrow \frac{1.645}{0.02} \times 0.4 < \sqrt{n} \Longrightarrow n > 32.9^2 = 1082.41$$

The sample size n has to be at least 1083.

b. The probability of a type I error is

$$P(X \geq 220|H_0) \overset{\text{c.c.}}{=} P(X \geq 219.5|H_0) \overset{\text{CLT}}{\approx} P\left(Z \geq \frac{219.5-200}{\sqrt{1000 \cdot 0.2 \cdot 0.8}}\right) \approx 1 - \Phi(1.50) = 6.68\%.$$

c. The power of this test for $p = 0.23$:

$$P(X \geq 220|p = 0.23) \overset{\text{c.c.}}{=} P(X \geq 219.5|p = 0.23) \overset{\text{CLT}}{\approx} P\left(Z \geq \frac{219.5-230}{\sqrt{177.1}}\right) \approx \Phi(0.79) = 78.52\%.$$

		1			2			3		4		5					
IEM M3	Online test Statistics	a	b	c	a	b	c	a	b	a	b	a	b	c		Total	Result
	April 16, 2020	4	3	6	5	5	5	3	4	2	4	3	3	3		50	10,0
2371286	Oost, Lisa van	4	3	6	3	4	4	3	2	2	4	2	2	1		40	8,2