

Mid-term exam, 28 October 2021

All aids are allowed, except a computer and personal assistance. Restricted use of some computer-like devices (including tablets and smartphones) is permitted under the rules described at the VHM 801 course homepage. The exam consists of one question with five parts, or subquestions, (labeled by letters **a**) – **e**), with weights as indicated) that should all be answered to achieve the maximal 15 points. The subquestions can be answered independently of each other. The mid-term exam accounts for 15% of the course mark; however, every student may choose to waive the result of the mid-term exam. The duration of the mid-term exam is 1 hour.

Generally, **statistical models and methods should be specified**. If you realize that you need more information than is provided to carry out the analysis, specify what information you need, how you would obtain it using Minitab or Stata, and how you would use it.

Question 1 (15 points)

Subliminal stimuli are any sensory stimuli below an individual's absolute threshold for conscious perception. It is sometimes claimed that subliminal messages can affect people's behaviour and thinking. A study carried out at a university investigated the effect of subliminal messages on math skills. The subliminal messages were flashed on a screen too rapidly to be consciously read. The study subjects were 18 students who participated in a summer course designed to raise their math skills after they had failed a certain mathematics test. The students were randomly assigned to receive daily either a positive subliminal message ("Each day I am getting better in math.") or a neutral subliminal message ("People are walking in the street.") The students took a test at the beginning of the summer course and took another (similar) test at the end of the course.

The test scores have been entered into Minitab in a worksheet containing the following variables recorded for all 18 subjects: **message** (positive/negative), **before**, **after** (test scores before and after the course), and **diff** (computed as **after-before**).

The attached Minitab output contains a full listing of the data, as well as some descriptive statistics and statistical analyses for the data. You may use the information in any way you want to answer the questions. Note that you are **not** required to duplicate the calculation of statistics already shown in the Minitab output.

- a) (2 points) Characterize the study type (e.g., experimental or another type) and design (e.g., one-sample or another design). Describe also the type of each of the variables in the data set (e.g., quantitative, or another type).
- b) (2 points) Use the information provided to carry out a **brief** descriptive analysis of the distribution of test scores **before** the summer course. Focus on the features of relevance for assessing whether it can reasonably be described by a normal distribution.

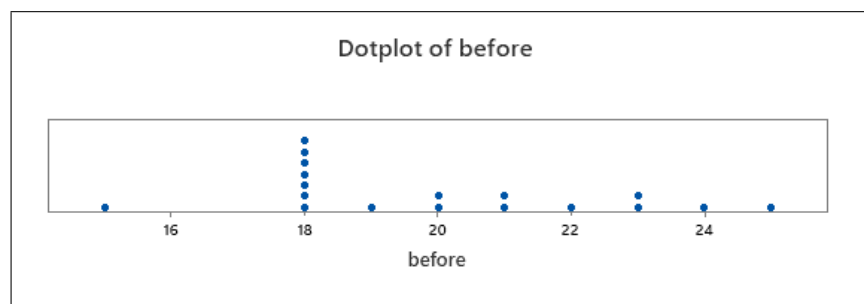
- c) (5 points) The research question of primary interest was whether the positive subliminal message brought about a greater improvement in math skills than the neutral message. Estimate the difference in improvements between the two groups with a suitable confidence interval, and use also a statistical test to determine the statistical significance of the data for the research question. Draw conclusions of your analysis.
- d) (3 points) Another question of interest to the researchers was whether the study gave evidence of a beneficial effect of the instruction that took place during the summer course. Use the data to quantify the improvement of math skills by an estimate and a measure of its precision (of your choice). Interpret the values you report carefully, in the context of the question of interest.
- e) (3 points) For this final subquestion, you need to answer only **one of the two parts i) and ii) below**. It is allowed (but not recommended) to answer both parts, in which case your score for e) will be for the best among the two answers.
- i) Upon reviewing the test results of the study, the course instructor reacted with great enthusiasm to the fact that all students in the group receiving the neutral message improved their score on the final test compared to the initial test, saying that “this was highly unlikely to happen by chance alone”. Carry out the relevant computation to substantiate this claim, and draw conclusions.
- ii) The random assignment of students to the two groups was intended to ensure that the two groups were comparable at the beginning of the course. Without carrying out any calculations, explain how you would give a statistical assessment of whether the random assignment was successful in balancing the math skills in the two groups. Make sure to fully specify the statistical procedure you would use and to explain the rationale for your chosen approach.

MIDTERM.MWX

Data Display

Data

Row	message	before	after	diff
1	positive	18	24	6
2	positive	18	25	7
3	positive	21	33	12
4	positive	18	29	11
5	positive	18	33	15
6	positive	20	36	16
7	positive	23	34	11
8	positive	23	36	13
9	positive	21	34	13
10	positive	18	27	9
11	neutral	18	29	11
12	neutral	24	29	5
13	neutral	20	24	4
14	neutral	18	26	8
15	neutral	25	38	13
16	neutral	22	27	5
17	neutral	15	22	7
18	neutral	19	31	12



Descriptive Statistics: before, after, diff

Statistics

Variable	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Skewness
before	18	19.944	2.645	15.000	18.000	19.500	22.250	25.000	0.32
after	18	29.83	4.76	22.00	25.75	29.00	34.00	38.00	0.08
diff	18	9.889	3.628	4.000	6.750	11.000	13.000	16.000	-0.11

Descriptive Statistics: before, after, diff

Statistics

Variable	message	N	Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Skewness
before	positive	10	19.800	2.098	18.000	18.000	19.000	21.500	23.000	0.60
	neutral	8	20.13	3.36	15.00	18.00	19.50	23.50	25.00	0.13
after	positive	10	31.10	4.48	24.00	26.50	33.00	34.50	36.00	-0.56
	neutral	8	28.25	4.89	22.00	24.50	28.00	30.50	38.00	1.00
diff	positive	10	11.30	3.23	6.00	8.50	11.50	13.50	16.00	-0.31
	neutral	8	8.13	3.48	4.00	5.00	7.50	11.75	13.00	0.29

Two-Sample T-Test and CI: after, message

Method

μ_1 : population mean of after when message = positive

μ_2 : population mean of after when message = neutral

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: after

message	N	Mean	StDev	SE Mean
positive	10	31.10	4.48	1.4
neutral	8	28.25	4.89	1.7

Estimation for Difference

95% CI for	
Difference	Difference
2.85	(-1.95, 7.65)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.27	14	0.223

Two-Sample T-Test and CI: diff, message

Method

μ_1 : population mean of diff when message = positive

μ_2 : population mean of diff when message = neutral

Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Descriptive Statistics: diff

message	N	Mean	StDev	SE Mean
positive	10	11.30	3.23	1.0
neutral	8	8.13	3.48	1.2

Estimation for Difference

95% CI for	
Difference	Difference
3.18	(-0.26, 6.61)

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
1.98	14	0.067