

HH/PSYC 2022 A – Statistical Methods II

Final Assignment

Due Monday, December 21, 2020 at 11:59:59 PM

Overview

Submit a single word-processed document using appropriate word-processing software (e.g., Word, LibreOffice, L^AT_EX, R Markdown). Embed all images in the document. Please submit the document in .doc, .docx, .odt, or .pdf file format (.pdf is preferred). Submit the document using the *Final Assignment* submission portal on the course eClass page. All data required to complete the assignment are available in text format throughout this document. To aid students who opt to use statistical software, all the text-formatted data in this document have been duplicated into a series of .csv files also located in the Final Assignment folder. Each .csv file is intuitively labeled according to the Long Answer question for which it is intended (e.g., file “Q1.csv” contains the data for Long Answer question 1).

Long Answer (70 marks)

For these questions, you are given brief summaries of research scenarios, specific hypotheses, and accompanying data. You are required to conduct appropriate statistical analyses of the data to draw conclusions for each research hypothesis. Write a brief report of the results in which you

- Highlight all inferential statistical effects (e.g., There was a strong main effect of Factor A)
- Discuss the qualitative nature of any such effects (e.g., Group A scored higher on <dependent variable> than group B)
- Report the results of your statistical tests using the appropriate APA formatting for each respective statistical test (e.g., $t[12] = 2.53$, $p = .026$, $d = 0.74$, 95% CI = [1.34, 4.78])
- Refer back to the research scenario to make a broad declarative statement about how well the research hypothesis is supported

For all analyses, assume two-tailed hypothesis tests with $\alpha = .05$ unless otherwise specified. All parametric analyses must include effect sizes. Any t -tests must include 95% confidence intervals. Any analysis that includes an F -test must also include an accompanying ANOVA table. Any analysis that includes a correlation must also include the variances of the variables and their cross product (SP ; covariability). Any analysis that includes linear regression must also include all relevant descriptive statistics (i.e., $SS_{Residual}$, standard error of the residual, regression equation). Any non-parametric analysis must include the ranked scores and the sum of ranks (ΣR).

You are free to use statistical software (e.g., SPSS, SAS, R, sciPy, MATLAB) to conduct statistical analyses and construct data plots. However, you must provide screenshots of your output as part of your submission. Please embed these images in the word-processed document, rather than submitting them separately. If you would prefer to compute

statistical tests and construct data plots by hand on paper, please take photos of your work and embed them into your submission in much the same manner. Alternatively, you could type out your hand calculations, but this would likely be rather tedious and we therefore advise against it.

If you opt to use statistical software, please report exact p values for each statistical test (e.g., $p = .026$). Otherwise, if you perform hand calculations, report p values in relation to (e.g., less than) the appropriate α -level (e.g., $\alpha = .05$) and include the critical values from the relevant statistical tables provided in the textbook for each statistical test (e.g., $p < .05$; $t_{\text{crit}} = 2.18$).

Question 1 **(24 marks)**

A human-computer interaction (HCI) researcher was interested in examining whether humans are better able to use a joystick or a mouse to point a computer cursor. She therefore constructed an experiment in which participants used either a joystick (*joystick* condition) or a mouse (*mouse* condition) to point a cursor to a target displayed on a computer monitor. She measured the time (seconds) that it took to place the cursor over the target (dependent variable referred to as *time*, where a higher *time* score indicates poorer performance). To determine whether potential benefits of using the joystick or mouse generalized to difficult HCI scenarios, the target was either stationary (*static* condition) or moved slowly across the computer screen at a constant velocity (*motion* condition). Participants were randomly and uniquely assigned to 1 of 4 conditions in the interface (levels: *joystick*, *mouse*) \times target type (levels: *static*, *motion*) experimental design. Given the data collected by the researcher (see Table 1 below), what can she conclude about how easily humans interact these HCI interfaces and how are these HCI interfaces influenced by target type? Include a line graph of the means.

Table 1. Target localization times in the interface \times target type conditions.

Subject	Interface	Target	<i>Time</i>
1	Joystick	Static	2.53
2	Joystick	Static	2.24
3	Joystick	Static	3.16
4	Joystick	Static	1.50
5	Joystick	Static	3.06
6	Joystick	Static	2.76
7	Joystick	Static	2.57
8	Joystick	Static	4.03
9	Joystick	Static	1.70
10	Joystick	Static	0.70
11	Joystick	Motion	2.00
12	Joystick	Motion	2.54
13	Joystick	Motion	2.23
14	Joystick	Motion	0.10
15	Joystick	Motion	0.75
16	Joystick	Motion	1.30
17	Joystick	Motion	2.11
18	Joystick	Motion	2.60
19	Joystick	Motion	1.05
20	Joystick	Motion	2.05
21	Mouse	Static	1.39
22	Mouse	Static	0.67
23	Mouse	Static	1.62
24	Mouse	Static	1.69
25	Mouse	Static	0.29
26	Mouse	Static	2.29
27	Mouse	Static	0.93
28	Mouse	Static	1.32
29	Mouse	Static	0.33
30	Mouse	Static	1.67
31	Mouse	Motion	1.35
32	Mouse	Motion	1.29
33	Mouse	Motion	1.31
34	Mouse	Motion	2.48
35	Mouse	Motion	1.83
36	Mouse	Motion	1.03
37	Mouse	Motion	2.18
38	Mouse	Motion	2.01
39	Mouse	Motion	1.79
40	Mouse	Motion	1.87

Question 2
(11 marks)

A marketing team wants to determine which of two prospective product lines consumers might prefer. They randomly select subjects to participate in a quantitative focus group in which half of the participants were given product *A*, while the other half of participants were given product *B*. Participants completed a questionnaire that probed their affinity for the product they inspected during the focus group. Questionnaires items were collapsed into a continuous-valued composite index of product affinity. In a preliminary analysis, the marketing team ensured that the assumption of homogeneity of variance was met:

$$F_{max} = \frac{s_{largest}^2}{s_{smallest}^2} = \frac{5.59}{4.33} = 1.29, F_{max.crit} = 4.04.$$

Given the product affinity data they collected (see Table 2 below), which product are consumers more likely to prefer? Show all relevant descriptive statistics.

Table 2. Product affinity scores for Product *A* and product *B*.

Subject	Product	Scores
1	<i>A</i>	3.27
2	<i>A</i>	2.52
3	<i>A</i>	4.83
4	<i>A</i>	0.68
5	<i>A</i>	4.59
6	<i>A</i>	3.84
7	<i>A</i>	3.35
8	<i>A</i>	7.00
9	<i>A</i>	1.19
10	<i>A</i>	-1.32
11	<i>B</i>	2.69
12	<i>B</i>	4.04
13	<i>B</i>	3.26
14	<i>B</i>	-2.08
15	<i>B</i>	-0.44
16	<i>B</i>	0.93
17	<i>B</i>	2.97
18	<i>B</i>	4.17
19	<i>B</i>	0.30
20	<i>B</i>	2.80

Question 3
(18 marks)

Sociologists investigated whether there is an association between salary and life enjoyment. They administered questionnaires to randomly selected participants who reported their salary in thousands of dollars (*salary*) and a battery of questionnaire items that probe life enjoyment. The life enjoyment items were collapsed into a continuous-valued composite index of life enjoyment (*LE*). Given their data (see Table 3 below), is there an association between salary and life enjoyment? If so, how does a change in salary quantitatively relate to a change in life enjoyment? Include a scatterplot of the data and the line of best fit.

Table 3. Salary in thousands of dollars (*salary*) and life enjoyment composite index (*LE*).

Subject	<i>salary</i>	<i>LE</i>
1	29	24
2	25	13
3	37	30
4	16	21
5	35	13
6	32	36
7	29	18
8	48	32
9	18	5
10	6	16
11	28	14
12	35	16
13	31	14
14	5	16
15	13	12
16	20	5
17	30	25
18	36	26
19	17	13
20	29	20

Question 4
(10 marks)

Cognitive psychologists are examining the effect of visuospatial cueing on perceptual processing speed. They designed an experiment in which a square randomly appeared on either the left or right side of a computer monitor and participants were required to push a button as soon as they detected the square. Each participant repeated this action hundreds of times. On half of the trials, a quick flash of light preceded the appearance of the square and it always appeared on the same side of the computer monitor as the square (*Cued* condition). On the other half of the trials, no such flash occurred (*Control* condition). The researchers measured the average time it took participants to detect the square on trials in both the *Cued* and *Control* conditions (see Table 4 below). What can the researchers conclude about the effect of visuospatial cueing on perceptual processing speed? Show all relevant descriptive statistics.

Table 4. Mean subject reaction times (milliseconds) in the *Cued* and *Control* conditions.

Subject	<i>Condition</i>	
	<i>Cued</i>	<i>Control</i>
1	204	218
2	200	225
3	212	221
4	191	195
5	210	203
6	207	210
7	204	220
8	223	226
9	193	207
10	181	219

Question 5 (7 marks)

A pharmaceutical company is interested in examining the efficacy of a new experimental drug to reduce allergic reactions and therefore recruited subjects to participate in a randomized clinical trial. The researchers exposed participants to a benign allergen to elicit allergic reactions. Half of the participants were assigned to an experimental drug treatment condition in which they were administered the new experimental drug (*Drug* condition), while the remaining half of participants were placed in a placebo condition in which they received a sham pharmacological treatment (*Placebo* condition). They measured a continuous-valued, composite index of allergic reaction symptomology to examine whether those who received the experimental drug treatment showed a reduction in allergic reaction symptomology (see Table 5 below). Upon preliminary analysis, the researchers discovered that the assumption of homogeneity of variance was violated:

$$F_{max} = \frac{s_{largest}^2}{s_{smallest}^2} = \frac{6.79}{1.67} = 4.07, F_{max.crit} = 4.04.$$

Furthermore, they realized that the dependent variable was not normally distributed, as scores were heavily positively skewed. Given these observations, using a parametric analysis of the mean difference between conditions is not appropriate. What can the researchers conclude about the efficacy of the drug using a non-parametric analysis? Show all relevant descriptive statistics.

Table 5. Composite allergic reaction symptomology scores in the Drug and Control conditions.

Subject	Condition	Scores
1	Drug	1.05
2	Drug	0.70
3	Drug	2.10
4	Drug	0.20
5	Drug	1.91
6	Drug	1.38
7	Drug	1.09
8	Drug	4.41
9	Drug	0.29
10	Drug	0.02
11	Placebo	5.73
12	Placebo	8.21
13	Placebo	6.70
14	Placebo	1.00
15	Placebo	2.02
16	Placebo	3.33
17	Placebo	6.19
18	Placebo	8.47
19	Placebo	2.67
20	Placebo	5.92

Short Answer (30 marks)

For these questions, you are given brief summaries of research scenarios and specific hypotheses. You need to choose which statistical test is appropriate to analyze the hypothesis given the context of the research scenario. Each scenario is worth 5 marks. 1 mark is for correctly identifying the appropriate statistical test. An additional 4 marks are awarded for justifying your answer with respect to

1. The type of data collected by researchers (i.e., continuous, ordinal, or nominal)
2. Whether researchers are investigating an association/relationship between at least two variables or are instead investigating differences between conditions using the same variable
3. If applicable, the number of conditions and/or factors
4. If applicable, the independence/dependence of observations between conditions

You may format your answers as either short paragraphs or bullet lists.

Question 1 (5 marks)

A vision scientist was examining whether human depth perception relies on contextual visual information (i.e., depth cues). He constructed an experiment in which participants viewed an object in virtual reality and had to estimate how far away the object was. Participants repeated this many times and the researcher computed how accurately participants were able to infer the depth of the object. Accuracy was operationalized as the difference between participant estimates of object distance and the true distance of the object. On half of the trials, the object appeared in a long hallway with realistic décor, thus providing contextual depth cues. On the other half of trials, the object appeared in a white void of empty space. Participants experienced both trial types in randomized order and their accuracy was compared between the hallway and void conditions. If depth is inferred from visual cues, accuracy should be lower in the void condition than in the hallway condition. What analysis should the vision scientist perform to examine whether subjects were less able to infer depth when deprived of depth cues?

Question 2 (5 marks)

A zoologist was interested in investigating whether non-human primates (NHPs) experience attentional “pop-out”; that is, when you are visually searching for something in a cluttered environment, the object you are looking for can either be difficult to find or is immediately obvious—it “pops out”. Therefore, they constructed an experiment in which NHPs were trained to locate an object amongst clutter and he measured how long it took the NHPs to locate the object. In the baseline condition, the object was presented alone. The researcher also included 2 additional conditions in which the object was presented amongst low-levels of clutter or high-levels of clutter. The NHPs completed each condition in randomized order. If NHPs experience “pop-out”, locating the object should take equally as long in the baseline condition as it does in both cluttered conditions. Which analysis should the zoologist perform?

Question 3
(5 marks)

A sociologist wanted to investigate how well a person's salary can be predicted by the combination of their intelligence quotient (IQ) and the number of years they have been formally educated. She therefore published a survey in which participants self-reported their salary in dollars, completed an IQ battery, and self-reported the number of years they have spent in post-secondary education. The IQ questionnaire was collapsed into a continuous-valued composite IQ index. What analysis should the researcher perform to assess the combined influence of IQ and educational attainment on salary?

Question 4
(5 marks)

An educational psychologist was curious whether students who consistently sit in the front row of lectures achieved higher grades than students who consistently sit in the back row. He published an online survey available to university students in which participants self-reported which row they sat in during lectures throughout the semester and their final grade point average at the end of the semester. He was interested in comparing the grade point averages of two groups of students: those who sat in front of the class and those who sat in the back. What analysis should he perform?

Question 5
(5 marks)

A self-proclaimed health and wellness company contracted statisticians to examine the efficacy of their newly developed dieting program. They provided the statisticians with data from a publicized weight loss competition that they hosted to advertise the release of their new product. For the competition, participants completed the dieting program and were ranked in order of the amount of weight they lost pre- and post-completion of the program. What statistical analysis could the statisticians perform to assess the efficacy of the dieting program?

Question 6
(5 marks)

A personality psychologist was interested in examining whether the prevalence of narcissism differed between people based on the type of pet they own: cats, dogs, or exotic pets. She administered a questionnaire that probed narcissistic personality traits to randomly selected participants. She collapsed the questionnaire items into a continuous-valued composite index of narcissism prevalence. She included biological sex as a factor in her analysis to examine whether biological sex interacts with pet type to influence the prevalence of narcissism. This gave her a 2×3 experimental design with sex (men, women) and pet type (cat, dog, exotic) as factors.

Total: /100 marks