

1 **MA 354: Data Analysis I – Fall 2021**
2 **Midterm Exam:**

3 **Instructions:**

- 4 • You have 2.5 hours to complete this exam.
- 5 • If you can't complete something you want to do in R describe it so I can understand your thought
6 process and so that you can leave yourself a to do note for later.
- 7 • Take a deep breath. You're going to do well and the worst case is that it will be productive.

8 **1 In-exam Portion:**

9 **Part I (20 points)**

10 In Part I, I'm simply evaluating your engagement with the material. If you've worked through the material,
11 there should be clear distinctions to make. I have provided as much room as I think is necessary to answer
12 these questions. Take a minute to think or do some scratch work – your answer should fit in the space
13 provided, only keep the important distinctions. I do not expect you to recite the formulas but to explain the
14 procedures, their hypotheses, conclusions and/or their differences.

15
16 **Guidance:** For these questions, do not look up these topics. If you spend time doing that, you will not
17 have enough time to conduct the data analysis below. Aim for 20-30 minutes here.

18
19 **Part II (50 points)**

20 In Part II, you're completing a data analysis. In this analysis you should provide numerical and graphical
21 summaries that provide information for the researcher related to their research question.

22
23 **Guidance:** For this part, sketch out what needs to be done first and then start executing your plan. Here,
24 I want to see that you know what needs to be done. This will not be perfect given the time constraints.

25
26 **Submit your exam through Moodle**

- 27 • The LAST-FIRST.RNW
- 28 • LAST-FIRST.pdf of your entire exam

29 **Guidance:** Ensure to compile your RNW file often. If you compile each paragraph then, when it doesn't
30 compile you'll know exactly where the issue is. If you don't compile often, you might find yourself trying to
31 find an issue in pages of content.

2 Out-of-exam Portion:

Part III (10 points)

Shortly after the exam, you will receive an email to anonymously review two exams. These will be graded on a mix of completeness and helpfulness. You should review their data analysis for completeness, correctness, and communication. You will type up **constructive** notes to make the response better.

The idea is to provide guidance for what's needed for the full data analysis to be effectively communicated to where you can understand the logic and the conclusions made about the data analysis. The format is discussed below.

- Write a paragraph about the general pros and cons of the paper you're reviewing. There is something good about every paper – find it and discuss that part. Also provide, in broad strokes a **constructive** critique of the response.
- Provide a list of major issues and suggestions.
- Provide a list of minor issues and suggestions.
- Provide a list of typographical errors you've found while reading.
- Ensure to provide specific line item comments where applicable; e.g.,
 - On page 1, line 2, you appear to interpret the statistic incorrectly.
 - On page 2, line 4, you're missing a period at the end of the sentence.

Part VI (20 points)

After you receive comments about your work, revisit your analysis from the exam. Write a final draft of your analysis and provide responses to reviewer comments.

- Write a revision of your original solution which incorporates comments made in the reviews you've received.
- For each review, provide a 1-5 score and a sentence or two evaluating how helpful and or thorough the review was.
- Provide a list of responses to specific line item comments; e.g.,
 - On page 1, line 2, you appear to interpret the statistics incorrectly.
Response: This was actually done correctly because I was treating the predictor as categorical and not continuous. I've added a sentence to make this distinction clear when fitting the model.
 - On page 2, line 4, you're missing a period at the end of the sentence.
Response: Thank you for pointing this out; I've added the missing period to the end of the sentence.

Part I – Use only the space provided to answer the following. If you go over, the text of your answer will be unreadable as it will overlap the next question. Only provide what is essential to answering the question.

1. Succinctly describe the intuition behind method of moments estimation.

Solution:

2. Succinctly describe the intuition behind maximum likelihood estimation.

Solution:

3. Succinctly describe the difference between a PMF and a CDF.

Solution:

4. Succinctly describe the results provided by The Central Limit Theorem.

Solution:

End Page

77
78

5. Succinctly describe what a p -value tells us. Consider why it may make sense to change the default threshold from $\alpha = 0.05$ to $\alpha = 0.005$.

Solution:

79

6. Suppose that X_1, X_2, X_3 , and X_4 are independent Poisson random variables with corresponding means $\lambda_1, \lambda_2, \lambda_3 = \lambda_1 + \lambda_2$, and $\lambda_4 = \lambda_1 \lambda_2$. In R, we can simulate random observations from the Poisson distribution using `rpois()`. How can I use this information to estimate the expected value of

$$U = \frac{X_1 + X_2}{1 + X_3 + X_4},$$

80

in the case that $\lambda_1 = \lambda_2 = 1$?

Solution:

81
82
83

End Page

84

Optional Bonus: Consider Question 6. Create a graphical display that provides compelling evidence for your answer.

85

Part II

Researchers have documented high rates of mental health issues among undergraduate students for years. Xiao et al. (2017) report data on 340 university and college counseling centers over five years. Their results suggest significant increasing trends for self-reported distress in generalized anxiety, depression, social anxiety, family distress, and academic distress, with the largest effect sizes observed for generalized anxiety, depression, and social anxiety. Further O'Neill et al. (2016) used World Health Organization World Mental Health Surveys to examine this issue across 21 countries to conclude that mental health issues are common among college students more broadly and that problems are typically untreated.

On top of this underlying "crisis," COVID-19 proved to be a severe disruption for colleges and universities. Administrations decided to move students off-campus and out of residence halls in the middle of Spring 2020, and some schools didn't fully return until Fall 2021. Now that many institutions are returning to on-campus activities, there are still downstream effects of the pandemic. Students and faculty who live with persons at high risk are still scared to become a disease vector. Students stressed about school are faced with additional academic and social expectations.

Son et al. (2020) interviewed 195 students at a large public university and found that 71% indicated increased stress and anxiety due to the COVID-19 outbreak, and 91% reported negative impacts of the pandemic overall. Wang et al. (2020) came to a similar conclusion among 2031 participants at another large public university. Students with usual and new stressors need to find ways to cope and seek support. This leads students to adopt either negative or positive coping mechanisms. Colleges and Universities are motivated to support students and guide them toward positive coping mechanisms.

Binfet et al. (2021) recruited $n = 284$ student participants from a mid-sized western Canadian university through their psychology classes. The goal was to evaluate the effects of canine contact on wellbeing outcomes. The researchers randomly assigned each participant to one of three treatments: (a) direct touch contact ($n_a = 95$); (b) indirect no-touch contact ($n_b = 95$); (c) control where students meet a dog handler without a dog ($n_c = 94$). Each participant completed pre- and post-test surveys to collect data described below.

- **Demographics** (pre-test only)
- **Flourishing Scale** (Diener et al., 2010): How participants feel about their life – higher scores indicate greater psychological wellbeing.
- **Positive Affect Scale** (Thompson, 2007): How participants rate their alignment with positive characteristics – higher scores indicate more positive affectivity.
- **Social Connectedness Scale** (Lee et al., 2001): How participants feel about their social surroundings and relationships – higher scores indicate a stronger sense of social connectedness.
- **Happiness Scale** (Lyubomirsky and Lepper, 1999): How participants feel about their happiness – higher scores indicate better happier participants.
- **Integration into Campus community**: How participants perceive their integration into campus life – higher scores indicate better integration.
- **Stress Scale**: How participants perceive their stress levels – higher scores indicate more stress.
- **Homesickness Scale**: a self-reported level of homesickness – higher scores indicate higher levels of homesickness.
- **Loneliness Scale** (Russell, 1996): a 20-item scale designed to measure participant loneliness – higher scores indicate more loneliness.
- **Negative Affect Scale** (Thompson, 2007): How participants rate their alignment with negative characteristics – higher scores indicate more negative affectivity.

The researchers want to evaluate whether there is evidence in the data for three hypotheses.

Hypothesis 1: All treatments groups will have significantly higher measures of wellbeing and lower measures of illbeing.

Hypothesis 2: The treatments groups that interact with dogs will have significantly higher measures of wellbeing and lower measures of illbeing.

Hypothesis 3: Direct contact with a therapy dog would yield greater benefits than the indirect contact treatment.

You can load the data into R using the following R code. Note that I kindly combined the scales described above for the pre-test (1), the post-test (2), and the follow-up (3). Note that the pre- and post-tests are of primary interest. You may consider exploring whether the effects last through to the follow-up after the experiment had completed.

Notes on coding:

- **Gender:** Male (1), Female (2), Non-Binary(3).
- **Year_of_Study:** First (1), Second (2), Third (3), Fourth (4), Fifth or more (5).
- **Live_Pets:** Pets back home (1), no pets back home (2).
- **Consumer_BARK:** Low (1), Medium (2), or High (3) consumer of BARK program.

Load the Data:

```
library(tidyverse,quietly = T)
dat.canine <- read_csv("canine_contact.csv",
                      na=c("x","X",""))
```

154 **Create Wellbeing Features:**

```
#####
# Well-being measurements
#####
dat.canine <- dat.canine %>%
  mutate() %>%
  mutate(posaff1 = rowMeans(select(dat.canine,
                                   num_range("PN1_", range = c(3,5,7,8,10))),
          na.rm = TRUE),
         posaff2 = rowMeans(select(dat.canine,
                                   num_range("PN2_", range = c(3,5,7,8,10))),
          na.rm = TRUE),
         posaff3 = rowMeans(select(dat.canine,
                                   num_range("PN3_", range = c(3,5,7,8,10))),
          na.rm = TRUE)) %>%
  mutate(happy1 = rowMeans(select(dat.canine, starts_with("HA1_")), na.rm = TRUE),
         happy2 = rowMeans(select(dat.canine, starts_with("HA2_")), na.rm = TRUE),
         happy3 = rowMeans(select(dat.canine, starts_with("HA3_")), na.rm = TRUE)) %>%
  mutate(social1 = rowMeans(select(dat.canine, starts_with("SC1_")), na.rm = TRUE),
         social2 = rowMeans(select(dat.canine, starts_with("SC2_")), na.rm = TRUE),
         social3 = rowMeans(select(dat.canine, starts_with("SC3_")), na.rm = TRUE)) %>%
  mutate(integration1 = H01_2,
         integration2 = H02_2,
         integration3 = H03_2)%>%
  mutate(flourish1 = rowMeans(select(dat.canine, starts_with("F1_")), na.rm = TRUE),
         flourish2 = rowMeans(select(dat.canine, starts_with("F2_")), na.rm = TRUE),
         flourish3 = rowMeans(select(dat.canine, starts_with("F3_")), na.rm = TRUE))
```

155 **Create Illbeing Features:**

```
#####
# Ill-being measurements
#####
dat.canine <- dat.canine %>%
  mutate(stress1 = S1_1,
         stress2 = S2_1,
         stress3 = S3_1) %>%
  mutate(homesick1 = H01_1,
         homesick2 = H02_1,
         homesick3 = H03_1) %>%
  mutate(lonely1 = rowMeans(select(dat.canine, starts_with("L1_")), na.rm = TRUE),
         lonely2 = rowMeans(select(dat.canine, starts_with("L2_")), na.rm = TRUE),
         lonely3 = rowMeans(select(dat.canine, starts_with("L3_")), na.rm = TRUE)) %>%
  mutate(negaff1 = rowMeans(select(dat.canine,
                                   num_range("PN1_", range = c(1,2,4,6,9))),
          na.rm = TRUE),
         negaff2 = rowMeans(select(dat.canine,
                                   num_range("PN2_", range = c(1,2,4,6,9))),
          na.rm = TRUE),
         negaff3 = rowMeans(select(dat.canine,
                                   num_range("PN3_", range = c(1,2,4,6,9))),
          na.rm = TRUE))
```


156 Use what you've learned and practiced so far this semester to help the researchers evaluate their research
157 questions. This obviously can not be done fully or completely in two hours. Get started exploring the data
158 and show that you know what needs to be done to help the researchers answer their questions.

References

- Binfet, J.-T., Green, F. L., and Draper, Z. A. (2021). The importance of client–canine contact in canine-assisted interventions: A randomized controlled trial. *Anthrozoös*, pages 1–22.
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-w., Oishi, S., and Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social indicators research*, 97(2):143–156.
- Lee, R. M., Draper, M., and Lee, S. (2001). Social connectedness, dysfunctional interpersonal behaviors, and psychological distress: Testing a mediator model. *Journal of counseling psychology*, 48(3):310.
- Lyubomirsky, S. and Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social indicators research*, 46(2):137–155.
- O’Neill, S., Auerbach, R., Alonso, J., Axinn, W., Cuijpers, P., Ebert, D., Green, J., Hwang, I., Kessler, R., Liu, H., et al. (2016). Mental disorders among college students in the who world mental health surveys. *Psychological Medicine*, 46(14):2955–2970.
- Russell, D. W. (1996). UCLA loneliness scale (version 3): Reliability, validity, and factor structure. *Journal of personality assessment*, 66(1):20–40.
- Son, C., Hegde, S., Smith, A., Wang, X., and Sasangohar, F. (2020). Effects of covid-19 on college students’ mental health in the united states: Interview survey study. *Journal of medical internet research*, 22(9):e21279.
- Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (panas). *Journal of cross-cultural psychology*, 38(2):227–242.
- Wang, X., Hegde, S., Son, C., Keller, B., Smith, A., and Sasangohar, F. (2020). Investigating mental health of us college students during the covid-19 pandemic: cross-sectional survey study. *Journal of medical Internet research*, 22(9):e22817.
- Xiao, H., Carney, D. M., Youn, S. J., Janis, R. A., Castonguay, L. G., Hayes, J. A., and Locke, B. D. (2017). Are we in crisis? national mental health and treatment trends in college counseling centers. *Psychological services*, 14(4):407.