**The final is due Thursday, Dec. 14th by 12:00 PM (PST) on Canvas.**

Instructions:

* This exam includes 7 questions and 2 extra credit questions and is worth 120 (out of 100) points.
* For essay and short answer questions, answers should be typed. You are encouraged to use bullet points to organize your ideas as you answer questions (it also makes it easier to grade!).
* All answers should be reported below the question. **To make it easier for the TAs to grade, please type all answers in blue.**
* For questions requiring computer work, answers should either be typed or be copy/pasted from the **edited, relevant output** (i.e., include only what is needed and highlight your answer; please do not include unedited or irrelevant output). You may include your code at **the very end** of the final if you would like, but it is not required.
* You may include a photo of any work done by hand, but please try to make your answers legible. Please make any answers/solutions easy for the TA’s to find (if we have to guess what your answer is from a sheet of hand-written work, we may not find the answer you intended to submit).
* This exam gives you the opportunity to work independently. You may consult with the TA’s or Emilio concerning the final, but not with anyone else. You can also consult your book, class notes, and previous labs and homework.
* You will not be able to submit corrections for the final.

**1. (30 points) Imagine that you are interested in examining the relationship between mood and weather. You ask 4 people to fill out a questionnaire about their mood for 70 consecutive days and also record the maximum temperature for each day. The data for weather and mood from 4 individuals are in the “tempmood.csv” data set.**

(a) Plot the data in a way that you find meaningful to illustrate the relationship between mood and weather. Comment on your plot. You may consider changing your plot after you complete the rest of the questions, to best illustrate your research findings. [Note: the plot deemed “the best” (e.g., most informative, most interesting, most understandable, etc.) by your teaching team will be awarded bonus points! Have fun! [5 pts]

(b) Compute means and standard deviations for each of the variables in the data set. [ 4 pts]

|  |  |  |
| --- | --- | --- |
|  | Mean | SD |
| Temp | 80.0 | 3.0 |
| Part1 | 3.0 | 0.8 |
| Part2 | 4.3 | 1.2 |
| Part3 | 5.1 | 1.8 |
| Part4 | 5.9 | 2.3 |

(c) Compute the sum of cross-products and covariance between weather and mood across all participants and report them below. Then explain what each of these indices indicate about the relation between weather and mood? [4 pts]

(d) What is the correlation between weatherand mood across all participants? What does the correlation indicate about the relationship between weather and mood? [2 pts]

*r* = -0.244. That means that there is a weak negative relationship between mood and weather.

(e) Estimate the regression equation of the line that best represents the relationship between weather and mood for all individuals as a single group. [2 pts]

(f) Estimate and report the regression equations for the lines best representing the relationship between weather and mood for each individual participant. [4 pts]

(g) Calculate and report the standardized beta-weight for the association between weather and mood (rounded to two decimal places) for each participant. Identify which participant had the strongest association between weather and mood, and which participant had the weakest association between weather and mood. [4 pts]

Participant 2 had the weakest association between mood and weather, while participant 4 had the strongest association between these two variables.

(h) What can you say about differences in the relationship between weather and mood across individual participants? [2 pts]

For participant 1, weather and mood have a positive relationship, meaning that as the temperature increases or decreases, their mood moves in the same direction. For Participant 2, the association was not statistically different from zero, implying that for this participant there is no association between the two variables. For Participants 3 and 4, the association is negative. As the temperature increases the mood decreases and vice-versa. Participant 4 also seems to be the most affected by temperature, since they have the strongest regression coefficient among all subjects.

(h) You submit the result of all these analyses for publication but the editor rejects the manuscript on the basis of: (i) a lack of power to examine your research questions, and (ii) the fact that there are only 4 individuals in your data set and, thus – they claim – you cannot generalize to the population. Nevertheless, you are convinced – or just have a hunch – that there might be something valuable here and write back arguing that the data and analyses are worth disseminating. What would you say to support your argument? [3 pts]

**2. (10 points) The following matrices are the covariance and correlation matrix, respectively, of variables *X*1, *X*2, *X*3, *X*4, and *X*5.Using the information provided in the matrices, fill in the gray boxes with the appropriate values.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **X1** | **X2** | **X3** | **X4** | **X5** |
|  | **X1** |  | - | - | - | - |
|  | **X2** | -0.5 | 0.25 | - | - | - |
| **Covariance** | **X3** | 1.8 |  |  | - | - |
|  | **X4** | -1.08 | -0.135 | -2.43 | 7.29 | - |
|  | **X5** | 6.48 |  | 6.48 | 4.374 |  |
|  |  |  |  |  |  |  |
|  |  | **X1** | **X2** | **X3** | **X4** | **X5** |
|  | **X1** |  | - | - | - | - |
| **Correlation** | **X2** | -0.5 |  | - | - | - |
|  | **X3** |  | 0.25 |  | - | - |
|  | **X4** |  |  | -0.3 |  | - |
|  | **X5** |  | 0.5 |  | 0.3 |  |

**3. (20 points) Researchers were interested in the role of extracurricular activities (sports: 0 = other extracurricular activities; 1 = participation in sports) and biological sex (female: 0 = male; 1 = female) on standard normal score of adolescent perceptions of social acceptance (PSA).**

**The data can be found in the “socialacceptance.csv” file. Determine whether factors of extracurricular activity type and biological sex are associated with adolescent PSA.**

a) State the type of design of the study. [2 pts]

b) Test whether there are group differences in adolescent PSA based on extracurricular activity, biological sex, and their interaction (use Type II SS). If there is a significant interaction effect, be sure to conduct appropriate follow-up analyses and report their outcomes. Write a report (no longer than a page) in which you report your findings as you would in a journal article (i.e., include text, table, and figure). [18 pts]

**4. (10 points) Explain why *r* must be between -1 and +1. Please, do not use more than 1 or 2 paragraphs. You can append calculations, if you need them.**

**5. (20 points) Say you run a simple regression with predictor variable and outcome variable .** **You fit the following model:**

a) What is the interpretation of ? What is the interpretation of ? **[4 points]**

b) When will be equal to the correlation between and ? Why? **[4 points]**

**After running the analysis, you remember a covariate that you believe is related to but is not substantively of interest.**

c) What are the benefits of including the covariate in the model? Include two benefits and explain in them in detail. **[4 points]**

**Finally, you include the covariate within the analysis and fit the following model:**

d) What is the interpretation of each of the coefficients in this model? **[4 points]**

e) When will be equal to the correlation between and ? When will be equal to the correlation between and ? Why? **[4 points]**

**6. (10 points) Suppose you are hired to serve as a statistical consultant. In each of the following cases, what advice would you give to your client concerning the procedures and/or conclusions he or she has drawn, or about the kind of statistical techniques most suitable? Be sure to briefly explain the reasoning underlying your advice.**

(a) A researcher studies the effects of education (HS or less, Some College, 4 Year College Degree, Graduate/Professional Degree) on income by randomly calling 5,000 participants in the United States. At a presentation of his results, several colleagues suggest that effects of education on income may not be robust when considering other predictors such as work experience, time with their current employer, age, personal investments, etc. What sort of analysis did the researcher most likely conduct, and how can the researcher address these criticisms of his research? **[2]**

(b) A researcher is interested in predicting the mental health (mentally stable versus mentally unstable) of college students based on their reported level of stress. What kind of sample should she collect, and what statistical technique would be best to address her research question? **[2 ]**

(c) A researcher collected data from undergraduate and graduate students at universities across the country in a study of the relation between age (range: 18 – 46 years; *M* = 23.5) and openness to experience. The researcher found a significant, negative relation between age and openness to experience (*r*(2,998) = -0.13, *p* < .05). She used this finding to argue that as people age, their openness to new experiences decreases, and that this explains why elderly individuals (aged 60 years and above) have difficulty in learning about novel technology and ideological shifts. Is this a reasonable conclusion? Why or why not? **[2 points]**

(d) A researcher studied a group of 100 students by having them complete a survey once a quarter, every quarter, for two years via an online survey form. The survey consisted of several items meant to measure anxiety, self-competence, and academic performance. What method(s) of analysis would be applicable to this type of data? Justify your recommendations. **[2 points]**

(e) A researcher received a small grant to conduct a study and is debating on how to spend the money. Her options are to (1) give a test to 300 individuals on one occasion; (2) give a test to one individual on 300 occasions; (3) give a test to 30 individuals on 10 occasions; (4) give a test to 10 individuals on 30 occasions; or (5) any combination of the above. What factors would you need to consider in consider in your recommendation of which data collection methods should she should use and why? **[2 points]**

**7. (20 points) Imagine you were hired to help a school board analyze data. The school board created different promotional videos to get middle schoolers interested in learning to play an instrument. One of the videos promoted the school’s marching band, and a second video promoted the school’s orchestra. The school board planned to randomly assign children to view one of these two videos, but after consulting with a researcher, the school board agreed to include a control group in the study. Hence, children were randomly assigned to view the marching band video, the orchestra video, or the generic educational video. After children viewed the videos, they took a brief survey to assess their interest in playing a musical instrument. The means and SD are given in the table below.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Educational Video | Marching Band Video | Orchestra Video | Total |
| Mean | 5.2 | 22.17 | 30.8 | 19.39 |
| SD | 7.2 | 7.2 | 7.2 | 12.11 |
| n | 6 | 6 | 6 | 18 |
| *ssB = 2035.6* |  |  |  |  |

(a) If you were to use orthogonal contrasts to test for differences between the means, what is the best way that you could assign contrast weights to each group to assess whether the promotional videos were generally effective in increasing children’s interest in playing an instrument, as well as whether one video was more effective than the other? Fill in the table below with these weights, and explain why these contrast weights are more appropriate than other weights. **[6 points]**

(b) Using the means in the table above, perform these contrasts. Report and interpret the results of the orthogonal contrasts, including the t-value associated with each contrast (use a two-tail alpha criterion of .05). Then, summarize your conclusions as if you were writing to the school board, and make a recommendation about which video(s) to continue showing students. **[7 points]**

(c) Explain why contrasts weights must be chosen with care to test specific hypotheses. To illustrate your point, repeat the orthogonal contrasts using a different set of orthogonal contrast weights, and explain how your conclusions would change if you had reported the results of these contrasts to the school board. Based on these changes, make an argument for why erroneous conclusions might be drawn if inappropriate contrast weights are use. **[7 points]**

**Extra Credit**

**8. (10 points) Explain what it means to say that a correlation is a covariance expressed in *z*-scores? Derive numerically the formula for a correlation based on the formula from a covariance (and describe the steps in your own words).**

**9. (10 points) Imagine that you are hired by the superintendent of a local school district to serve as the statistical consultant on a project examining children’s acquisition of mathematics skills over the course of a school year. Students were measured at 5 times throughout the school year and approximately 25% of the data are missing due to students being absent on testing days. The superintendent lets you know that children tend to start at much different levels of math ability at the start of the school year, and that some students progress more rapidly than others in increasing their ability, and some even decline. The superintendent wants you to use repeated measures ANOVA to evaluate whether the trajectories of students with mothers who completed college differ from students whose mothers did not earn a college degree. What advice would you give the superintendent (e.g., appropriateness of RM ANOVA to answer research questions, assumptions of RM ANOVA, etc.)? Explain your answer in a paragraph or two.**