**This take-home final is worth 20 points. make sure you show all your assumptions and calculations to get full (or partial) credit. Insert text, tables and figures in the spaces provided, between the yellow highlight. NOTE: Make sure you add your name to the file name, or you will be fined 1 point.**

**Datasets: You will use two datasets for this exam:**

* **Dataset #1: http://www.pelagicos.net/BIOL4090\_6090/final/Dataset1.xlsx**
* **Dataset #2: http://www.pelagicos.net/ BIOL4090\_6090/final/Dataset2.xlsx**

**Turning in Exam: Due Date / Time: December 6th at 9 am**

* **Email a single word file to** [**khyrenba@gmail.com**](mailto:khyrenba@gmail.com) **using “Biol6090\_4090\_final” as the title.**

**Late Work:**

**Late submissions will be penalized 2 points per day (or fraction) starting December 6th at 9 am.**

**Feed-back:**

**I will answer all email questions I receive every morning through December 5th, by sending my answer to the entire class.**

**DISCLAIMER:**

**This exam recreates homework questions using tests you have seen and have practiced in class. There are no tricks – the computer will run these tests and you will get results.**

**You will complete this take-home final working ALONE, but are allowed any class notes, books and reference materials. By turning in this assignment, you certify that you have worked on this problem set without help from anyone else, whether enrolled in this class or not. Thanks.**

**Please sign that this statement above is true: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**(print name or add scanned signature in space provided above)**

**NOTE: I will not accept take-home exams without the proper signature.**

1a) Testing for normality.

You have the following distribution (use file Dataset1) of 100 fish lengths, which you want to test for normality using two criteria: (i) the kurtosis /skewness and (ii) the SW test. NOTE: for the skewness / kurtosis, use the rules of thumb discussed in class (mean +/- 2 SE) to assess whether they are significant. For the SW test, use the p-value.

Summarize the “length” data, using a histogram (+0.125). Do these data look normal?

Copy and paste a Q-Q plot. Do these data look normal (+0.125)? Do these data look normal?

Report and interpret kurtosis (+0.125) skew (+0.125) of this distribution (Are they significant)?

What are the null and the alternate hypotheses being tested by the SW test (+0.125)?

Paste the SW test result (+0.125) and explain the outcome of the test. Explicitly state the meaning of the p value and state whether this distribution is normal? (+0.125)

1b) Perform a data transformation of the original length data using Rcmdr. Explain why you selected the specific transformation (+0.125), and if you added any constants to the data? (+0.125)

Re-evaluate the data, after the transformation. Explicitly report and interpret kurtosis (+0.125) skew (+0.125) of the transformed data (Are they significant)?

Paste the SW test result (+0.125) and explain the outcome of the test. Explicitly state the meaning of the p value and state whether this distribution is normal? (+0.125)

1c) Perform a second data transformation of the original length data using Rcmdr. Explain why you selected this specific transformation (+0.125), and if you added any constants to the data? (+0.125)

Report and interpret kurtosis (+0.125) skew (+0.125) of the transformed data (Are they significant)?

Paste the SW test result (+0.125) and explain the outcome of the test. Explicitly state the meaning of the p value and state whether this distribution is normal? (+0.125)

1d) Briefly summarize (in your own words) how the data transformations changed the datasets. Refer to the figures / results provided above:

* First transformation (+0.25):
* Second transformation (+0.25):

2a) You have data on the body lengths and the tail lengths of 12 endangered monkeys (use file Dataset2.xlsx) and want to determine whether these two variables vary together (e.g., larger monkeys have larger tails). Briefly explain what statistical test you would use to analyze these data (+1):

What is the null / alternate hypothesis for this statistical test? (+0.125):

Paste a scatter-plot of these data (+0.250). Does it look like there is a relationship? Can you describe the pattern you observe? (+0.125):

Perform two tests (one assuming normal distributions and one not assuming normal distributions) of these data and compare your results: Hint – use the nonparametric version that is identical to the parametric version, but uses ranks rather than the raw data.

* Parametric test name : \_\_\_\_\_\_\_\_\_\_ (+0.125) - Degrees of freedom (+0.125):\_\_\_\_
* Non-parametric test name : \_\_\_\_\_\_\_\_\_\_(+0.125) - Degrees of freedom (+0.125):\_\_\_\_

Copy and paste the results of parametric test (+0.125):

Copy and paste the results of non-parametric test (+0.125):

What are the results of these tests? Did the results agree? (+0.125)

Based on the similarity / lack of similarity of these two test results, would you guess that these distributions are normally distributed (Why / why not)? (+0.25)

2b) You want to compare the lengths of the tails and the bodies of these monkeys, but are unsure about what statistical test to use. Which type of test should you use to account for the fact that larger monkeys have longer tails (+0.125) – Hint: are these data independent?

Briefly explain why you selected this specific test (+1):

Perform two tests (one assuming normal distributions and one not assuming normal distributions) of these data and compare your results:

* Parametric test name : \_\_\_\_\_\_\_\_\_\_(+0.125) - Degrees of freedom (+0.125):\_\_\_\_
* Non-parametric test name : \_\_\_\_\_\_\_\_\_\_(+0.125) - Degrees of freedom (+0.125):\_\_\_\_

Copy and paste the results of parametric test (+0.125):

Copy and paste the results of non-parametric test (+0.125):

What was the outcome of these tests? Did the results agree? (+0.125):

2c) Customs agents obtained two monkey tails from a wildlife trader and want to use your data to figure out the body length of the animals these tails were taken from. The tails are 11.0 cm and 12.0 cm long.

What test can you use to predict the body length of a monkey, if you know its tail length? Explain how this process works (+0.125)– what is a critical inherent assumption of this approach (+0.125):

Briefly explain why you selected this specific test (+1 point):

What are the null / alternate hypotheses of this test? (+0.125):

Perform this test and paste the results (+0.125):

What are the degrees of freedom of this test? (+0.125):

What is the result of this test (significant / not significant) (+0.125) – Explain)?

Create a scatterplot of these data, including a trend line (+0.125):

How much variance in tail length does the body length model account for (from 0% to 100%)? (+0.125):

Interpret estimates of the intercept (+/- SE) (+0.125) and the slope (+/- SE) of the linear model (+0.125) – Hint: use the rule of thumb (mean +/- 2 SE of the estimates to approximate their 90% Confidence limits):

Using these estimates, calculate the body length of a monkey that had a 11.0 cm long tail (+0.125):

Please show your work (calculations) for full credit (+0.125). Does this value make sense? Explain (+0.125)

Using these estimates, calculate the body length of a monkey that had a 12.0 m long tail (+0.125):

Please show your work (calculations) for full credit (+0.125). Does this value make sense? Explain (+0.125)

Extract the residuals from this test and determine if they are , in fact, normally distributed (+0.125):

Calculate the Cook’s distances for the residuals. Create a box plot of the Cook’s values and identify any outliers (+0.125).

Based on the box plot you just created, do any data points have undue influence on the slope? Explain why / why not using the rule of thumb discussed in class? (+0.125):

If you had monkey body lengths, what equation would you use to estimate their tail lengths? Explain (+0.125):

Paste the scatter plot including this new linear regression line (with the equation), showing the data (+0.125):

3) You want to determine the effect of four different diets (feed types) on pig weight weights, but are unsure if you have normally-distributed data, so you will perform two tests: one parametric and one non-parametric using the following data: 20 pigs: 5 pigs assigned to four different diet treatments (see below) . Enter data into Excel and import the file into R (+1 POINT FOR SETTING UP DATA)

|  |  |  |  |
| --- | --- | --- | --- |
| Feed 1 | Feed2 | Feed3 | Feed4 |
| 60.8 | 68.7 | 102.6 | 87.9 |
| 57 | 67.7 | 102.1 | 84.2 |
| 65 | 74.0 | 100.2 | 83.1 |
| 58.6 | 66.3 | 96.5 | 85.7 |
| 61.7 | 69.8 | 97.2 | 90.3 |

3a) Use the descriptive statistics to summarize these four datasets. Report the means / medians / SDs / ranges of the data below (+0.125 for each dataset \* 4 datasets):

Paste a bar graph (with error bars) illustrating the means of these four groups (+0.250). Do these four group data distributions look different? (Why / Why not) (+0.125):

What is the grand mean? (+0.125):

Perform two tests (one assuming normal distributions and one not assuming normal distributions) of these data and compare your results:

* Parametric test name : \_\_\_\_\_\_\_\_\_\_(+0.125) - Degrees of freedom (+0.125): \_\_\_\_\_\_
* Non-parametric test name : \_\_\_\_\_\_\_\_\_\_(+0.125) - Degrees of freedom (+0.125): \_\_\_\_\_\_

Briefly explain why you selected these specific tests (+1 point):

Copy and paste tabular results of nonparametric test (+0.125) – State the null / alternate hypothesis of this test (+0.125):

Copy and paste tabular results of parametric test (+0.125) – State the null / alternate hypothesis of this test (+0.125):

* For ANOVA, what were the total degrees of freedom (+0.125): \_\_\_\_\_\_\_\_\_
* For ANOVA, what were the Model degrees of freedom (+0.125):\_\_\_\_\_\_\_\_
* For ANOVA, what were the Error degrees of freedom (+0.125): \_\_\_\_\_\_\_\_\_

What was the outcome of these tests? Did the results agree? (+0.125):

Were the variances of the groups equal? (Why / why not? (Explain – refer to the R results)): (+0.125)

Based on this result, should you use the result of the parametric test or the non-parametric test: (+0.125):

3b) After getting significance in the ANOVA test, are we done with the analysis? Why / Why not? (+0.250):

Perform the parametric test again, this time asking for post-hoc tests to compare all the pair-wise comparisons of these four groups. Paste the tabular results (+0.125):

Based on this result, which group(s) were responsible for the significant ANOVA result? Explain (+0.250):

How can you implement this next step in the non-parametric example? Explain (+0.125):

Perform the required non-parametric tests to determine which group(s) were responsible for the significant non-parametric ANOVA result. Explain the outcome of these tests (+0.250):

4a) You want to test the effects of two factors (sex: male / female) and medical treatment (placebo /drugs) on hormone levels. Your sample size is 20 values from 20 individuals. Enter data into Excel and import the file into R. Perform the required parametric ANOVA test. (+1 POINT FOR SETTING UP DATA)

|  |  |  |  |
| --- | --- | --- | --- |
| **No Drug** | **No Drug** | **Drug** | **Drug** |
| **Female** | **male** | **Female** | **Male** |
| 16.5 | 14.5 | 39.1 | 32 |
| 18.4 | 11 | 26.2 | 23.8 |
| 12.7 | 10.8 | 21.3 | 28.8 |
| 14 | 14.3 | 35.8 | 25 |
| 12.8 | 10.0 | 40.2 | 29.3 |

What type of ANOVA would you perform? (highlight one with bold text):

dependent OR independent (+0.125)

one-way OR two-way (+0.125)

Briefly explain why you selected this specific test (+1 point):

Before you start, answer these questions:

* Total degrees of freedom (+0.125): \_\_\_\_\_\_\_\_ (Explain Why?) (+0.125):
* Model degrees of freedom (+0.125): \_\_\_\_\_\_\_\_ (Explain Why?) (+0.125):
* Interaction (AXB) degrees of freedom (+0.125): \_\_\_\_\_\_\_\_ (Explain Why?) (+0.125):
* Error degrees of freedom (+0.125): \_\_\_\_\_\_\_\_ (Explain Why?) (+0.125):

4b) Use the descriptive statistics to summarize these data: Report the means / medians / SDs / ranges of each dataset data below (+0.125 for each one \* 4 datasets):

Paste a bar graph (with error bars) illustrating the means of these four groups (+0.250).

What is the grand mean (+0.125)?:

What is the “male” mean (+0.125)?:

What is the “female” mean (+0.125)?:

What is the “no drug” mean (+0.125)?

What is the “drug” mean (+0.125)?

After looking at these data summaries, which factors will be significant in this analysis. Explain (+0.125):

Perform a parametric test of these data. Copy and paste tabular results of parametric test (+0.125):

What was the effect of “sex”: Explain the outcome using the results of the table above (+0.125). State the null / alternate hypothesis of this test (+0.125), and explain what is the meaning of the test result (+0.125):

What was the effect of “drug”: Explain the outcome using the results of the table above (+0.125). State the null / alternate hypothesis of this test (+0.125), and explain what is the meaning of the test result (+0.125):

What was the effect of “sex \* drug” interaction: Explain the outcome using the results of the table above (+0.125). State the null / alternate hypothesis of this test (+0.125), and explain what is the meaning of the test result (+0.125):

Paste a figure illustrating the interaction of “sex” and “drug” (+0.250). Please interpret the figure (does it suggest there was a significant interaction – Explain (+0.250)).