Final Exam Instructions

* Turn your **cell phones** completely **off**.
* You will have the full final exam period to complete this exam, although this exam is only about 1.5 times the length of a midterm. Relax.
* Point values appear in [brackets].
* If you have multiple solutions to a problem, I will give credit for the most incorrect solution unless you indicate which solution you want me to grade. Please identify the solution you want graded.
* For any question requiring a justification or explanation, note that the most of the points are awarded for the proper justification, not the proper answer to the question. *In most cases, a single sentence justification will be sufficient.*
* You must **show your work** for full (or partial credit).
* Cell phones, iPods, tablets, laptops, etc. are not calculators. These devices cannot be used during the exam. Sharing calculators is not permitted. Passing a calculator during the exam will be considered as cheating. See me if you do not have a calculator.
* There are 83 points possible on the exam. This exam comprises 15% of your course grade, just a few percent more than a regular midterm.
* **Sign the integrity agreement below. You will get a score of zero if the agreement is not signed.**

## Integrity Agreement

I understand that direct or indirect sharing of information (giving or receiving) about this exam with other students who are taking this exam *or will take this exam* will result in an F grade. In addition a report will be submitted to the Office of Students Rights and Responsibilities.

Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Printed Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

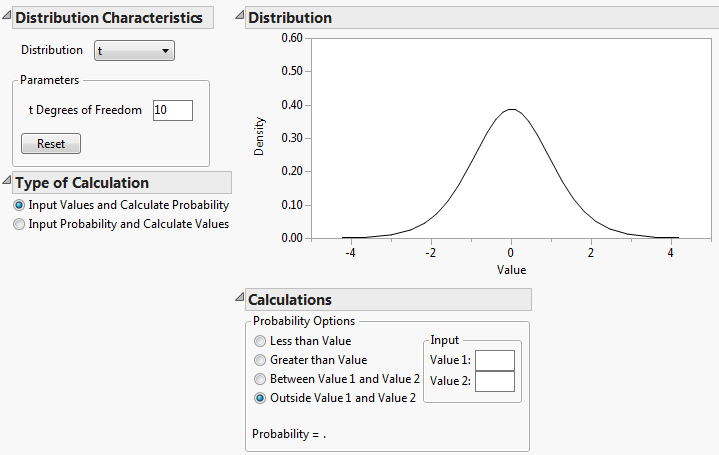
Exam Questions

# Question A

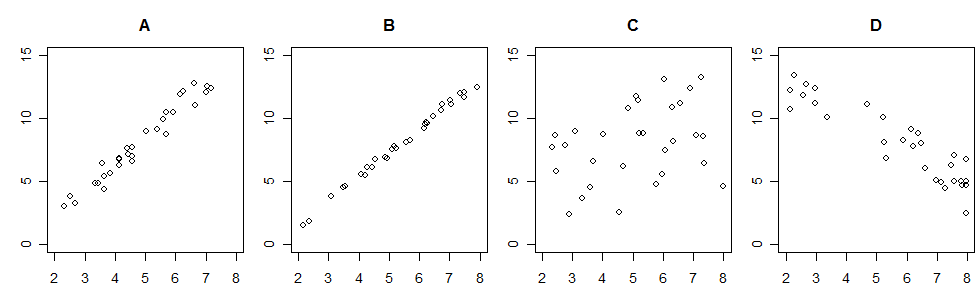
Is there a relationship between wing length (mm) and wingbeat frequency (Hz) among hummingbirds? An article in the *Journal of Experimental Biology* entitled “Muscle performance in hovering hummingbirds” presented data from 12 hummingbirds to examine this relationship. The questions below are based on these data which are plotted to the right and summarized in the table below.

| **Variable** | **N** | **Mean** | **Std Dev** |
| --- | --- | --- | --- |
| Wing Length (mm) | 12 | 39.4034 | 11.0090 |
| Wingbeat Frequency (Hz) | 12 | 41.6858 | 5.4176 |
| r = -0.9061 | | | |

1. [1] The summaries listed above are for all the data in the plot (including the points in the circle and triangle). If the circled point was removed from this dataset, would the value of the sample correlation **increase in magnitude**, **decrease in magnitude**, or **stay about the same**? *Circle one.*
2. [1] The residual SD for all these data (including the points in the circle and triangle) is Hz. If the point in the triangle were removed, the residual SD would **increase**, **decrease**, or **stay about the same**? *Circle one.*
3. [3] What percentage of the variation in Wingbeat Frequency is explained by the relationship between Wing Length and Wingbeat Frequency?
4. [1] Formally speaking, an important adjective is missing from #3 which describes the nature of the relationship being described (maybe you didn’t notice). What is the missing adjective?
5. [1] Consider the research question: Do longer wings tend to beat more slowly (i.e., lower frequency)? In plain English, what are the null and alternative hypotheses to be tested?
6. [3] Compute the value of used to test the hypotheses stated in part (5). *No conclusions are needed. Only compute the value.*
7. [3] Below is the dialog box in JMP to find the P-value for this test. Fill in the three missing numbers to find the *non-directional* P-value associated with your t-statistic in part 6.



1. [3] Consider the research question: Do longer wings tend to beat more slowly? If JMP reported the *non-directional* P-value as 0.004, what would you report as the *directional* P-value and what conclusions would you draw?
2. [4] Using the summaries listed at the start of the main question (which include all the data), find the equation of the fitted regression line.
3. [2] When excluding the circled gray point, the fitted regression line is . Using the fitted model from this reduced dataset, suppose we compared two hummingbirds: A and B. If hummingbird A’s wing was 1mm longer than hummingbird B’s, then we would expect that hummingbird A’s wings would beat \_\_\_\_\_\_\_\_ Hz (**faster**/**slower**) hummingbird B’s. *Fill in the box with a number and circle the right adjective (faster or slower).*
4. [2] When excluding the circled gray point, the fitted regression line is . Using the fitted model from this reduced data set, predict the mean wingbeat frequency for hummingbirds with 30mm wings.
5. [1] Is your prediction in 11 an **interpolation** or **extrapolation**? *Circle one.*
6. [1] In most all practical problems, the regression line is less steep (i.e., more flat) than the SD line. **True** or **False**? *Circle one.*
7. [1] Arrange the following plots in order from weakest to strongest correlation. Mark the plots with the digits 1 through 4. Use “1” for the weakest correlation and “4” for the strongest.



1. [1] Consider testing the null hypothesis of no linear relationship between *X* and *Y* for the data shown in each of the plots in #14 above using a *t*-test. Which dataset (**A**, **B**, **C**, or **D**) would yield the largest P-value? (Note all samples are the same size). *Circle one: A, B, C, or D.*

# Question B

The *Journal of the Science of Food and Agriculture* contained anarticle entitled “Influence of hydroponic and soil cultivation on quality and shelf life of ready-to-eat lamb’s lettuce,” researchers studied the effects of different hydroponic growing methods on the nitrate content of lettuce. Researchers randomly assigned 34 lettuce seedlings to one of three growing conditions: soil, hydroponic A, or hydroponic B. At the end of the growing period (60 days), nitrate measurements of the lettuces were taken. The data is summarized in the table below.

Table 1: Leaf nitrate concentration (mg/kg)

|  |  |  |  |
| --- | --- | --- | --- |
| Growing Method | Soil | Hydroponic A | Hydroponic B |
| Mean ± SE | 3806 ± 124 | 4695 ± 107 | 3878 ± 103 |
| n | 9 | 12 | 13 |

1. [2] One of the researcher’s main questions is to determine whether the growing method affects nitrate concentration in lettuce. Considering how this study was executed (i.e., an observational study or an experiment), can one address this question with the data provided above? *Briefly justify your answer.*
2. [3] Averaging the three group means we obtain the following: However, the mean nitrate concentration for all the leaves was reported as 4147.29. Why do these two numbers not agree?

Below is an incomplete ANOVA table summarizing the data. You may use this information for the subsequent problems.

Table 2: ANOVA summary of Lettuce Nitrate Concentration

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Ratio** | **Prob > F** |
| --- | --- | --- | --- | --- | --- |
| Growing Method | XX | 5590867.1 | XX | 20.1695 | <.0001\* |
| Error | XX | XX | 138597 |  |  |
| C. Total | XX | XX |  |  |  |

1. [2] In the context of the research and in plain English, what are the null and alternative hypotheses investigated using the ANOVA summary above (Table 2)?
2. [1] What is the (symbolic) null hypotheses investigated using the ANOVA summary above (Table 2)?
3. [1] **True** or **False**? *Circle one.* The alternative hypothesis investigated in the ANOVA summary above is . *No explanation needed.*
4. [1] What is the value of df(within)?
5. [1] What is the value of df(between)?
6. [1] **True** or **False**? *Circle one.* The value of the *F*-statistic listed above would clearly be larger if the nitrate standard deviations were smaller for each group.
7. [1] **True** or **False**? *Circle one.* The value of the *F*-statistic listed above would clearly be larger if the nitrate means were smaller for each group.
8. [1] **True** or **False**? *Circle one.* The value of the *F*-statistic listed above would clearly be larger if the nitrate means were more different across the groups. *No explanation needed.*
9. [3] Citing values from the ANOVA table above (Table 2) to support your answer, what conclusions could be drawn regarding the hypotheses stated in (4) and (5)?



1. To the right is a plot of the residuals versus fitted values for these data.
   1. [1] Which ANOVA condition is this plot used to verify?
   2. [2] Explain how this plot suggests the data meet this condition.
2. [3] Compute using the information provided in the ANOVA table (Table 2). *Show your work for credit.*
3. Suppose the value of mg/kg (this is not the correct answer, but use it for computations below).
   1. [3] Compute an approximate 95% Fisher LSD interval comparing the mean nitrate concentration of the “Hydro A” and “Soil” growing methods.
   2. [3] Interpret the interval you just computed in the context of this problem.
4. [3] Below is JMP output for all three Fisher 95% LSD comparisons. The sample mean nitrate level was lowest for lettuce grown in soil. If we wish to minimize nitrates in the lettuce, is growing the lettuce plants in soil clearly the best choice? *Briefly explain your reasoning citing one or more parts of the output below.*

| **Level** |  |  |  |  |  |  | **Mean** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Hydro A | A |  |  |  |  |  | 4695.0000 |
| Hydro B |  | B |  |  |  |  | 3878.0000 |
| Soil |  | B |  |  |  |  | 3806.0000 |

Levels not connected by same letter are significantly different.

| **Level** | **- Level** | **Difference** | **Std Err Dif** | **Lower CL** | **Upper CL** |
| --- | --- | --- | --- | --- | --- |
| Hydro A | Soil | 889.0000 | 164.1626 | 554.188 | 1223.812 |
| Hydro A | Hydro B | 817.0000 | 149.0336 | 513.044 | 1120.956 |
| Hydro B | Soil | 72.0000 | 161.4340 | -257.247 | 401.247 |

1. [1] The confidence intervals in the output in (14) above are Fisher 95% confidence intervals. If Bonferroni intervals were computed instead, would the intervals be **wider** or **narrower**? *Circle one.*

# Question C

Suppose that the adult weight of wild female mountain lions is normally distributed with a mean of 100 lbs and a standard deviation of 20 lbs.

1. [3] Draw a picture of this normal curve and label the axis so that the value of the mean and standard deviation are clear.
2. [3] What percentage of adult wild female mountain lions weigh between 80 and 120 lbs? (Note: you should be able to solve this without a normal table.)
3. Two researchers are proposing studies of mountain lions. The first, Dr. Cinco, will capture 5 female mountain lions. The second, Dr. Diez, will capture 10 female mountain lions. Each researcher will compute the sample mean weight of their captured lions.
   1. [3] If Dr. Cinco were to go out and repeat her study many times, she’d get different means for each sample. By approximately what amount would you expect these sample means to vary from the true population mean?
   2. [1] What is the statistical name of the number you computed in part (a)?
   3. [1] Is the following statement **true**, **false**, or **not enough info**? *Circle one.* The sample mean of Dr. Cinco is more likely to lie in the interval (98, 102) than the sample mean of Dr. Diez.

# Question D – Multiple Choice

For each of the following, circle the single *best* answer.

1. [1] In the context of correlation, if and …
   1. There is significant evidence that the variables and are linearly related.
   2. There is a lack of evidence that the variables and are linearly related.
   3. There is significant evidence that the variables and Y are not linearly related.
   4. No claim of significance (or lack of significance) can be made with this information alone.
2. [1] In the context of regression, if and …
   1. The scatter of the data closely follows a straight line.
   2. The scatter of the data is nearly random.
   3. The scatter of the data closely follows a 45° diagonal line.
   4. Nothing can be said about the scatter of the points without seeing the plot.
3. [1] A researcher compared PCB contamination levels for three watersheds. Her report lists the results of an ANOVA F-test: “.” Which of the following claims is valid?
   1. There is significant evidence that the three watersheds have the same amount of contamination, on average.
   2. There is significant evidence that at least one of the three watersheds has a different amount of contamination, on average.
   3. There is no substantial evidence that the mean contamination levels are different among the three watersheds.
   4. None of the above.
4. [1] A researcher compared phosphate contamination levels for three watersheds and failed to find a statistically significant difference in the phosphate levels. Which of the following statements is true?
   1. If the researcher conducts a larger study, she would be unlikely to find any difference in phosphate levels.
   2. If the researcher conducts a larger study, she would be likely to find a difference in phosphate levels.
   3. If the researcher conducts a larger study, she could potentially find a difference in phosphate levels despite of her current findings.
   4. None of the above.
5. [1] The histogram to the right is of the number of La Graciosa thistle seedlings emerging in 40 1m×1m plots. Regarding this data, which of the following statements is true?
   1. The data is skewed to the left and the mean will be larger than the median.
   2. The data is skewed to the left and the median will be larger than the mean.
   3. The data is skewed to the right and the median will be larger than the mean.
   4. The data is skewed to the right and the mean   
      will be larger than the median.
6. [1] Which of the boxplots to the right (a, b, c, or d) corresponds the histogram appearing in part 5 above?
7. [1] A researcher has just collected data on the pH of the acids in 17 cow’s abomasums (4th stomach). The mean pH was 2.6 with a standard deviation of 0.06. If the researcher is trying to describe the variability of pH levels among cows, which of the following values will be of *most* interest to her?
   1. Sample standard deviation
   2. Standard error of the mean
   3. Sample mean
   4. Confidence Interval for the population mean
8. [1] The histogram to the right is a histogram of the total carbohydrate content (g) of 38 red delicious apples. Which of the following most closely represents a 95% confidence interval for the population mean carbohydrate content of red delicious apples?
   1. (11.2, 17.7) g
   2. (12.7, 16.2) g
   3. (13.1,15.9) g
   4. (13.9, 15.2) g
9. [1] The following are two histograms of musk thiol concentrations (the stinky stuff in percent by volume) based on data drawn from the same population of striped skunks. One is a histogram of thiol concentrations for a single sample of 70 skunks. The other is a histogram of sample mean concentrations for 70 samples of 10 skunks per sample. Which is the histogram of the thiol concentration of the single sample of 70 skunks? A or B?



# Question E

Choosing from the following list, for each of the scenarios below indicate what analysis procedure would *most* likely be appropriate. Please write the letter of the procedure in the box to the right.

1. One-sample t
2. Two-sample t
3. Paired-t
4. Sign
5. Wilcoxon-Mann-Whitney
6. Chi-squared test of independence
7. Chi-squared goodness of fit
8. ANOVA
9. Regression

|  |  |  |
| --- | --- | --- |
|  | [1] A geneticist self-pollinated pink-flowered snapdragon plants and produced 97 progeny with the following colors: 22 red plants, 52 pink plants, and 23 white plants. The purpose of this experiment was to investigate a genetic model which states that the probabilities of red, pink, and white are 0.25, 0.50, and 0.25. |  |
|  | [1] A sample of 15 patients was randomly split into two groups as part of a double-blind experiment to compare two pain relievers. The seven patients in the first group were given Demerol and reported the following numbers of hours of pain relief: 2, 6, 4, 13, 5, 21, 4. The eight patients in the second group were given an experimental drug and reported the following number of hours of pain relief: 0, 12, 1, 10, 2, 2, 1, 3. |  |
|  | [1] The Diablo nuclear power plant discharges its cooling water into Diablo Cove north of Avila Beach. Since many marine organisms are sensitive to water temperature, a study was conducted to investigate if there was a relationship between water temperature and intertidal red algae. Temperature and algae density data were obtained for 40 intertidal locations in and around Diablo Cove. |  |
|  | [1] Heat shock proteins (HSP’s) are a type of protein produced by some organisms as protection against damage from exposure to high temperature. In the fruit fly Drosophila melanogaster, the genes that encode HSP’s are found on chromosomes that uncoil and appear to puff out when viewed under a microscope. To determine the effect, if any, of heat shock on HSP’s a biologist counted the number of puffs per chromosomal arm from the salivary glands of 40 Drosphila larvae that had been heat shocked at 37°C for 30 minutes, 40 larvae that had been heat shocked for 60 minutes, and 40 control larvae. |  |
|  | [1] Physical exercise has become an important part of standard therapy for patients who have had acute myocardial infarction and cardiac surgery; however, its role after heart transplantation has not been well defined. Doctors from UCLA randomly assigned 68 heart transplant patients to either a six-month structured cardiac-rehabilitation exercise program (36 patients) or were given unstructured therapy at home (32 patients). For each patient, doctors measured cardiopulmonary stress at the beginning of the study, and again 6-months later. The difference in stress scores was then used to measure improvement (or decline) for each patient in each group. |  |
|  | [1] Random samples of oranges grown under two different irrigation conditions (A and B) were classified into two categories (1 = ship for sale, 2 = ship for juice). Growers were interested in whether the irrigation affected the overall quality of the crop. |  |