JMP Instructional data set 024

Penguins: Male/Female Comparison

JMP STATISTICAL DISCOVERY

JMP platforms and statistical methods

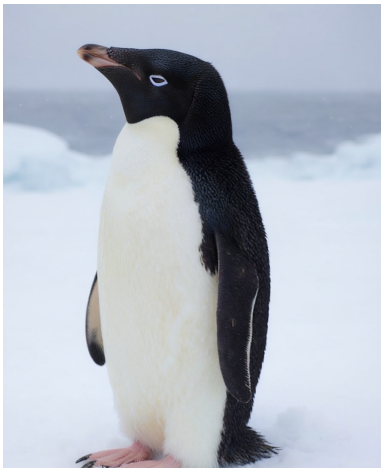
Graph Builder and Graphs:	Comparative dot plots; 3D scatterplot
Fit Y by X:	One factor analysis of variance (ANOVA)
Multivariate:	Scatterplot matrix
Distribution:	Confidence interval for a population mean; tolerance interval; prediction Interval

Problem statement

To better understand similarities and differences in morphological features between males and females of different penguin species, a researcher from Simon Fraser University, in collaboration with The Palmer Station Antarctica Long-Term Ecological Research Network, obtained measurements on a sample of penguins from three different species [1].

The dataset consists of measurements across 4 physical features of 34 male and 34 female Chinstrap penguins, 71 male and 72 female Adélie penguins, and 61 male and 58 female Gentoo penguins.

Chinstrap



Adélie

Gentoo

Focus of the analysis will be to compare the physical features between the different species and sexes of the penguins. interval estimation methods (confidence intervals, tolerance intervals, and prediction intervals) will also be used to estimate the range of sizes of a feature expected for the mean, a population, and an individual penguin.

Data

IDS_024.jmp

Species	Species of penguin (Chinstrap, Adélie, Gentoo)
Sex	Sex (Male , Female)
Culmen Length	Length (mm) of culmen*
Culmen Depth	Depth (mm) of culmen*
Flipper Length	Length (mm) of flipper
Body Mass	Body mass (grams)

** Culmen is the upper ridge of the beak.*

Exercises

Before you begin the analyses, it would be best to create row states using a unique color for each species and symbol for each sex. To do, select Rows > Color or Mark by Column. Select 'Species' and click OK. To use a different symbol for each sex, open the column header by clicking the icon in the upper left of the data table. 

In the 'Sex' column, select the Male group. Rows > Markers, and choose the larger solid circle (4th one down from the top left). Then select the Female group and choose the open circle (3rd one down from the top left). Now all graphs created will utilize these row states.

1. Create comparative dot plots across species and sex for each of the four features. Describe the similarities and differences observed in these features.

> ***Instructions:***
> *Graph > Graph Builder. Place 'Culmen Length' on the Y axis, 'Sex' on the X axis, and 'Species' on the X axis below 'Sex'*
> *Instead of creating 4 different comparative dot plots, it would be best to utilize the Column Switcher tool so all the variables can be viewed from the one graph by selecting the variable to display on the Y Axis. To do so, select Redo > Column Switcher under the red triangle. Select the current variable used for the Y as the Column to Switch ('Culmen Length'), and choose that along with all the others to use as the Replacement Columns. Now any of the variables can be selected to display on the graph.*

2. Perform a two-sample t-Test for each feature comparing the means between males vs. females for each of the species separately. Identify all those comparisons where there is statistically significant evidence of a difference in means. For all those where there is a significant difference found, provide an interpretation of the confidence intervals.

Instructions:
Analyze > Fit Y by X. Place 'Culmen Length' in the Y, Response, 'Sex' in the X, Factor, and 'Species' in the By role. Click OK.
Choose Layout > Arrange in Rows under the top red triangle. Enter "3" and click OK.

Instead of creating separate reports for each variable, it can be advantageous to use the Column Switcher tool. Select Redo > Column Switcher under the top red triangle. Select the current variable used for the Y as the Column to Switch ('Culmen Length'), and choose that along with the other three variables to use as the Replacement Columns. Now any of the variables can be selected and used for the analysis.

Select t Test under the two red triangle. Examine the p-values and confidence intervals for all 3 x 4 = 12 comparisons.

3. Perform a one-way ANOVA for each feature comparing the means across the three species for each of the sexes separately. Identify all those comparisons where there is statistically significant evidence of a difference in means. For all those where there is a significant difference found, conduct a student t's multiple comparisons analysis. Identify those where is a significant difference. For all those where there is a significant difference found in the paired comparisons, provide an interpretation of the confidence intervals.

Instructions:
Analyze > Fit Y by X. Place 'Culmen Length' in the Y, Response, 'Species' in the X, Factor, and 'Sex' in the By role. Click OK.
Choose Layout > Arrange in Rows under the top red triangle. Enter "2" and click OK.

Select Redo > Column Switcher under the top red triangle. Select the current variable used for the Y as the Column to Switch ('Culmen Length'), and choose that along with the other three variables to use as the Replacement Columns. Now any of the variables can be selected and used for the analysis.

Select Means/ANOVA under the top red triangle. Examine the p-value for all 2 x 4 = 8 comparisons.

Under the top red triangle, choose Compare Means > Each Pair, Student's t. Examine the p-values and confidence intervals 3 x 2 x 4 comparisons.

4. Perform a one factor ANOVA for each feature comparing males vs females for each of the species separately. the X variable. Identify any statistically significant differences observed in the means of each feature between the species and the sexes. Why are these analyses not the best ways to test/quantify statistical differences between the 2 sexes and between the 3 features?

Instructions:
In the Fit Y by X report, select both pooled t test and the t Test.
If there were any variables where you believed the variances were not equal, it's best to view the t Test as that assumes unequal variability between the group. Otherwise, examine the Pooled t test as that's a more powerful analysis if the assumption of equal variance holds.
Use significance level of 0.05 for the analyses.

5. For each species/sex combination, produce 95% confidence intervals for the population mean for each feature. Describe how one would interpret these intervals by demonstrating it for one of the morphological variables across the six species/sex groups.

Instructions:
Analyze > Distribution. Place all four morphological feature variables in the Y, Columns. Place both 'Species' and 'Sex' in the By role. Click OK.
The report will contain 24 histograms, summary statistics, and confidence intervals for the mean (4 features for each of 3x2=6 species/sex combination).

6. For the female gentoo penguins, evaluate if the normal distribution is a good model to describe each of the four morphological variables. Are there any concerns that the normal distribution is not a good model?

Instructions:
First set up a data filter so the analyses can be based only on that group.
Select the Filter View tab on the top left side of the data table. Click the "+" button.
Name the Filter "Species & Sex".
In the Table Filter that has opened on the right of the data table, select both 'Species' and 'Sex' variables and click "+". In this Table Filter, select the Female Gentoo.
The data table has now been filtered to include only those penguins. Any analyses produces while this filter is active will only use these data. Selecting "X Exit View" will return to the full data table.

For the analysis: Analyze > Distribution. Place all four morphological variables in the Y, Columns. Click OK. In the distribution report, select Continuous Fit > Fit Normal under the red triangle next to each of the variable names.
In the Fitted Normal Distribution tables added to the report window, select Goodness of Fit under the red triangles. Using the Anderson-Darling test for normality, identify if there are any variables that have p-values < 0.05.

7. For the female gentoo penguins, produce a 99% tolerance interval (at 95% confidence level) for each feature. Interpret these intervals.

Instructions:
In the Distribution report, select Tolerance Interval under the red triangle for each variable. Specify confidence level of 0.95 and 0.99 for proportion to cover. Confirm the Normal Distribution is the distribution being used. Click OK.

8. For the female gentoo penguins, produce a 95% prediction Interval for each feature. Interpret these intervals.

Instructions:
In the Distribution report, select Prediction Interval under the red triangle for each variable. Specify confidence level of 0.95 and 1 for number of samples. Click OK.

9. Calculate the correlations between all pairs of variables and create a scatterplot matrix. Identify the pairs of variables that have the strongest and weakest correlation. Examine the scatterplot matrix. Do these plots provide any additional insights into differences between the species and sexes that the analyses comparing the variables individually did not identify?

> ***Instructions:***
> *Note: You'll need to first exit from the Table Filter to return to the full data table.*
> *Analyze > Multivariate Methods > Multivariate. Place all the feature variables in the Y, Columns role. Click OK.*

Supporting materials

1. Gorman KB, Williams TD, Fraser WR (2014) "Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus *Pygoscelis*)". *PLoS ONE*, 9(3): e90081, 2014.
https://doi.org/10.1371/journal.pone.0090081
https://pallter.marine.rutgers.edu

Resources to learn more about these penguin species.

https://en.wikipedia.org/wiki/Pygoscelis

https://www.penguinworld.com/types/pygoscelis.html

** Please adhere to any citation requirements for distribution and use of these data.*