JMP® Introductory Lab Activities

Activity 7: Sampling Variation

Data Set: Pollen.jmp

Summary

Cornelius Growbetter has tried a new fertilizer for his corn. He has 3,848 stalks of corn in his field. He is investigating if this product will produce better kernels of corn. He would like to measure the number of kernels on the nub, edge, and crack of the corncob for a random sample of ears of corn. You have access to all the corn stalks through this data set.

You will compare the distributions of the sample mean from samples of size 50 and samples of size 10, and will investigate how the distribution is related to the entire population. You will summarize your findings in a report (required output and discussion is in italics).

The Pollen Data

Open the data table **Pollen.jmp** from the **Sample Data**. The data table has five columns of data and 3,848 rows of data. The key variables of interest are **edge** and **nub**, which are measures of the number of kernels on stalks of corn.

Part 1: Using the entire data set (the population)

Use **Analyze > Distribution** to graph and summarize the data. Select **edge** and **nub** for **Y, columns**, and click **OK**. This will produce graphs, quantiles, and summary statistics.

From the top red triangle next to **Distribution**, select **Stack** to display the output in a horizontal layout, and select **Uniform Scaling** to apply the same x-axis scaling to both histograms.

1. *Cut and paste the histograms and box plots into your report. Also include the Quantiles and Summary Statistics tables.*

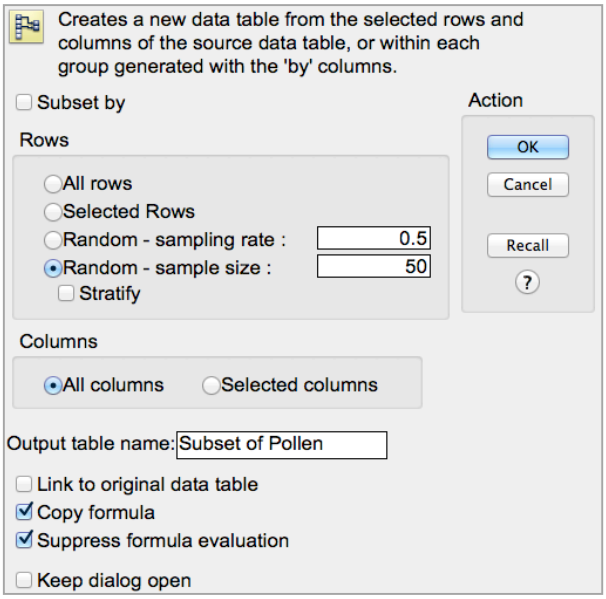
   *Describe and contrast the distributions (**edge** and **nub**) using the histogram and box plot. Looking at this information, how would you describe the distributions of the number of kernels found on the edge of the corn as compared to the nub?*

Part 2: Using a subset of the data set

Go to **Tables > Subset**.

Select **Random – sample size**, enter **50**, and click **OK**.

This will pick a random sample size of 50 from the 3,848 and store the resulting data in a new data table.



Using **Analyze > Distribution**, select **edge** and **nub** for **Y, columns** and click **OK**.

2. *Cut and paste histograms and box plots into your report. Include the Quantiles and Summary Statistics tables.*

   Describe and contrast the distributions (nub and edge) using the histogram and box plot. Looking at this information, how would you describe the distribution of the number of kernels found on the edge of the corn as compared to the nub?

3. *Compare and contrast the distributions for the sample of size 50 with the population distributions in problem 1 using the histograms, box plots and summary statistics. How close are the sample means for edge and nub to the population means?*

   Look at the graphs and tables obtained by two other students. Describe the shape of their distributions and how they compare to yours.

4. *Now take a sample size of 10 and compare the shape of your distribution to the distribution from two other students.*

   How are the distributions for the sample of size 10 different from the population distributions?

   How are the distributions for the sample size of 10 different from the distributions for the sample size of 50?

5. *How is variability related to the sampling size?*

Write Your Report

Your report should include:

- *A histogram of the population distributions for edge and nub.*
- *A histogram of the values from your sample of size 50 and 10 distribution for both edge and nub.*
- *A description of both sample distributions using numeric and graphical information.*
- *A comparison of both sample distributions with the corresponding population distributions.*

Be sure to write in the context of the situation.