Technologies for Data Analysis for Experimental Biologists

Melanie I Stefan, PhD

HMS

8 May 2014

MI Stefan (HMS)

Data Analysis with JMP

3 8 May 2014 1 / 21

-

If you do not have JMP installed yet

- Go to https://rc.hms.harvard.edu
- Navigate to Downloads and then to the JMP download page
- Download and install JMP or JMP Pro

Get datasets for today's class

- Go to https://nanosandothercourses.hms.harvard.edu/ and log in
- Download the datasets London2012.xls and Chicago2013.xls

► 4 Ξ ► 4

Come to Quant Bio Club!



Quant Bio Club, Mondays 3-5, Goldenson 229

MI Stefan (HMS)

Data Analysis with JMP

8 May 2014 3 / 21

4 ∃ > < ∃ >

First steps with JMP 1

- Normality

< □ > < □ > < □ > < □ > < □ >

Beginners' Tutorial

Beginners' Tutorial

- Open up JMP
- Navigate to Help \rightarrow Tutorials \rightarrow Beginners Tutorial 0

MI Stefan (HMS)

Data Analysis with JMP

э 8 May 2014 5 / 21

イロト イポト イヨト イヨト

First steps with JMP

2 Descriptive statistics

Normality



MI Stefan (HMS)

Data Analysis with JMP

3 8 May 2014 6 / 21

< □ > < □ > < □ > < □ > < □ >

Concepts covered

- Opening files
- Tabulate
- Saving/Exporting results
- Graph Builder •
- Distribution 0

イロト イポト イヨト イヨト

Descriptive statistics exercise

Analyse results from the Chicago marathon

- Open dataset Chicago2013.xls This is a random sample of 85 participants in the 2013 Chicago marathon and their finishing times.
- Use Analyze \rightarrow Tabulate to compute mean, standard deviation, median and interquartile range
- Save the result of your "tabulate" command in a format of your choosing (and verify that this worked)
- Use Graph \rightarrow Graph Builder to plot finishing time as a function of rank.
- Use Analyze \rightarrow Distribution to plot a histogram and box plot. Look at the Summary Statistics in the Distribution window and check that they are the same as in the file you saved earlier.

Sac

イロト イポト イヨト イヨト

First steps with JMP





< □ > < □ > < □ > < □ > < □ >

Concepts covered

- Is my data normal?
- Fitting to a normal curve
- Can we test for normality?

- 4 ⊒ →

Image: A match a ma

Is my data normal?

What does normal data look like?

- Open file bunnies.xls
- Draw a histogram of the entire dataset. Is it normal?
- Use Table \rightarrow Subset \rightarrow Random sample size to draw a sample of a particular size.
- Plot your sample. Does it look normal?
- What indicators would you look for?

Fitting to a normal curve

Fitting a normal curve your data

- Go back to your Chicago Marathon dataset
- Do you think this is a normal distribution? Examine indicators of normality (mean and median, skew, where most data points are, ...)
- Are there other reasons to believe that the data should/should not be normal?
- Fit the data to a normal curve (Little red arrow above your histogram \rightarrow Continuous Fit \rightarrow Normal). What do you think?

Can we test for normality

Can we test for normality?

- How might a test for normality work?
- Is there a way to do this in JMP? •
- When is this useful, and when is it not?

Normality

Let's take a break



MI Stefan (HMS)

Data Analysis with JMP

First steps with JMP

- Normality



イロト イポト イヨト イヨト

Concepts covered

- Computing a Z score
- Creating custom columns in JMP
- Interpreting a Z score
- Computing number of samples within one standard deviation from the mean

イロト イポト イヨト イヨト

Z scores

Z score tutorial

- Go to http://stat.fsu.edu/tutorials/zscore/mac.html (Mac) or http://stat.fsu.edu/tutorials/zscore/pc.html (PC)
- Work through the tutorial
- Compute z scores for the Chicago 2013 dataset
- My friend ran the marathon in 3.08 hours. How many standard deviations from the mean is that?
- How many runners lie within one standard deviation from the mean? What percentage is that?

A (10) < (10) < (10) </p>

- First steps with JMP



イロト イポト イヨト イヨト

Sac

Concepts covered

- Taking samples from a population
- The subset function
- Distribution of sample means
- Confidence intervals
- Small vs large samples

-

Sampling from a larger dataset

Take samples of size 3

- Go back to your Chicago 2013 dataset
- Take a sample size of size 3 (how?)
- For your sample, report the mean and the upper and lower 95 % confidence interval

Sampling from a larger dataset

Take samples of size 3

- Go back to your Chicago 2013 dataset
- Take a sample size of size 3 (how?)
- For your sample, report the mean and the upper and lower 95 % confidence interval

Comparing your samples

- What is the true population mean?
- How often does the 95 % confidence interval contain the true mean?
- How are the sample means distributed?

Sampling from a larger dataset

Take samples of size 3

- Go back to your Chicago 2013 dataset
- Take a sample size of size 3 (how?)
- For your sample, report the mean and the upper and lower 95 % confidence interval

Comparing your samples

- What is the true population mean?
- How often does the 95 % confidence interval contain the true mean?
- How are the sample means distributed?

Take samples of size 50

- Repeat the same exercise for samples of size 50
- What differences do you notice?

Sneak Preview

Homework (posted later today):

- Reinforcing concepts learned today
- Exploring new functions of JMP

Next week:

- Hypothesis testing
- Parametric and nonparametric tests
- Correlation

- ∃ >