Statistical Investigations and Analyses with FREE GeoGebra Software

Stephen J. Miller
Winchester Thurston School
Pittsburgh, PA

17 April 2015

Today IS NOT about

Today is **not** going to be about learning statistical techniques, although we might be able to discuss some techniques and address specific questions you might have.

NCTM 2015
Boston
Session 408
Stephen J. Miller
About Today
Plan for Today
About GeoGebra
Tour of GeoGebra
Data Entry
Univariate Data
Bivariate Data
Investigation
Probability
Inference
Contact Information

Today IS about

Our goal for today is to learn how to **use** GeoGebra to carry out some statistical calculations and techniques, and to learn how to do some statistical investigations.

NCTM 2015
Boston
Session 408
Stephen J. Miller
About Today
Plan for Today
About GeoGebra
Tour of GeoGebra
Data Entry
Univariate Data
Bivariate Data
Investigation
Probability
Inference
Contact Information

Outline

The **Plan for Today:**

- About GeoGebra
- GeoGebra Interface overview
- Data Entry
- Univariate Data (graphs and summaries)
- Bivariate Data (graphs and summaries)
- Investigation: Regression and Correlation
- Probability Calculations
- Statistical Inference

NCTM 2015
Boston
Session 408
Stephen J. Miller
About Today
Plan for Today
About GeoGebra
Tour of GeoGebra
Data Entry
Univariate Data
Bivariate Data
Investigation
Probability
Inference
Contact Information

About GeoGebra

- Any GeoGebra users here?
- Multi-platform (Windows, Mac, Linux)
- Tablet versions
- Portable version
- CAS
- Geometry
- Algebra/Precalculus/Calculus
- Statistics
- FREE!

NCTM 2015
Boston
Session 408
Stephen J. Miller
About Today
Plan for Today
About GeoGebra
Tour of GeoGebra
Data Entry
Univariate Data
Bivariate Data
Investigation
Probability
Inference
Contact Information

Tour of GeoGebra

Quick overview of the GeoGebra user interface

NCTM 2015
Boston
Session 408
Stephen J. Miller
About Today
Plan for Today
About GeoGebra
Tour of GeoGebra
Data Entry
Univariate Data
Bivariate Data
Investigation
Probability
Inference
Contact Information

Data Entry

Data entry can be achieved through "copy and paste" as well as direct entry.

Enter these data values, the heights (in inches) of students in my AP Statistics classes:

62	62	63	63	63	63	64	65	65
65	65	66	66	66	66	66	67	67
67	67	69	69	69	69	70	70	70
71	71	71	71	71	72	72	72	72
73	73	73	73	73	73	73	74	77

Univariate Graphical Displays

- Using the data previously entered, we will create:
  - histogram
  - dotplot
  - boxplot
  - stemplot
  - normal quantile plot (normal probability plot)

Univariate Numerial Analysis

- Using the data previously entered, calculate numerical summaries
  - mean
  - median
  - standard deviation
  - quartiles
- Including/excluding values

Bivariate Graphical Displays

Here are data on the hours of study per week for eight college students and their GPA.

Hours	14	25	15	5	10	12	5	21
GPA	3.2	3.6	3.4	3.0	3.1	3.3	2.7	3.8

- Enter the data
- Create a scatterplot of the data

Least-Squares Regression/Line of Best Fit

Using the hours of study and GPA data:

- Find the equation of the LSRL
- Plot LSRL on scatterplot
- Create a residual plot
- Compute correlation (r) and coefficient of determination (R^2)
- Use LSRL to predict GPA given the number of hours of study per week

Regression and Correlation Investigation

- Close out of the **Spreadsheet View** and go to **Algebra View**
- Be sure you have the **Input Bar** visible
- Using the **Point Tool**, create a “cloud” of about 10 points on the graph
- Using the **Select Tool** (the arrow), highlight and select all your points
- Select the **Best Fit Line** tool from the fourth button (all about lines)
- The line will show, as will the LSRL equation
- In the **Input Bar**, type **r = correlation[A,B,C,...]** until you have all your points listed

Regression and Correlation Investigation

What happens to the LSRL and correlation in the following scenarios?

- Select and drag one of your points far from your "cloud" of points.
- Select and drag your points until they all follow a linear pattern with a **positive slope**. Then drag a central point straight up.
- Select and drag your points until they all follow a linear pattern with a **negative slope**. Then drag a central point straight up.
- Select and drag your points until they form a **quadratic pattern**.

Regression and Correlation Investigation

- Arrange your points such that the slope of your LSRL is negative, and the correlation is approximately $r = -1$
- Arrange your points such that the slope of your LSRL is positive, and the correlation is approximately $r = 0.5$
- Arrange your points such that there is a strong nonlinear relationship between the x and y variables. What is the correlation?

Probability Calculations

Probability calculations can be carried out from the **Distribution** tab of the **Probability Calculator**.
We will do two examples (one continuous distribution and one discrete distribution):

- Normal Probability Calculations
- Binomial Probability Calculations

Normal Probability Calculations

For 17-year-olds in the United States, blood cholesterol levels in milligrams per deciliter have an approximately normal distribution with mean 176 mg/dL and standard deviation 30 mg/dL.

- What proportion of 17-year-olds have blood cholesterol levels between 165 and 185 mg/dL?
- What proportion of 17-year-olds have blood cholesterol levels above 200 mg/dL?
- What proportion of 17-year-olds have blood cholesterol levels below 160 mg/dL?
- What cholesterol level is at the 90th percentile of blood cholesterol levels for 17-year-olds?

Binomial Probability Calculations

The proportion of adults age 25 and older in the United States with at least a bachelor's degree is 0.27. Suppose you pick seven adults at random.

- What is the probability that exactly three will have a bachelor's degree or higher?
- What is the probability that three or four will have a bachelor's degree or higher?
- What is the probability that at least three will have a bachelor's degree or higher?
- What is the probability that less than three will have a bachelor's degree or higher?

Statistical Inference

Inferential methods (one-and two-sample hypothesis tests and confidence intervals, as well chi-square analyses and ANOVA) can be carried out from the **statistics tab** of the **Probability Calculator**

Examples (as time allows)

- Inference for Proportions
  - Testing a claim about a population proportion
  - Estimating a population proportion
  - Chi-Square Test
- Inference for Means
  - Testing a claim about a population mean
  - Estimating a population mean

Inference for Proportions

The General Social Survey (GSS) asked randomly selected subjects if they would be willing to pay much higher prices in order to protect the environment. Of the $n = 1154$ respondents, 518 indicated a willingness to do so.

Find a 95% confidence interval for the population proportion of adult Americans willing to do so at the time of that survey.

Inference for Proportions

An article summarized a study in which 2,205 adolescents ages 12 to 19 took a cardiovascular treadmill test.

The researchers wanted to test the claim that more than 30% of adolescents have a poor level of cardiovascular fitness. As such, they will test $H_0 : p = 0.30$ *versus* $H_a : p > 0.30$.

Of the 2,205 adolescents tested, 750 had a poor level of cardiovascular fitness. Does this sample provide support for the claim that more than thirty percent of adolescents have a poor level of cardiovascular fitness?

Inference for Proportions

A 2011 survey asked 806 randomly selected adult Facebook users about their Facebook privacy settings. One of the questions on the survey was “Do you know how to adjust your Facebook privacy settings to control what people can and cannot see?” Their responses are cross-tabulated based on gender:

	Male	Female
Yes	288	378
No	61	62
Not sure	10	7

Carry out a **chi-square test of association** to see if there is evidence of an association between gender and knowing how to adjust Facebook privacy settings to control what people can and cannot see.

Inference for Means

Data in the table below are the weights (in ounces) of a random sample of 15 bags of potato chips. The potato chip manufacturer claims that their bags of potato chips, on average, 10 ounces.

9.8	9.9	10.2	9.9	9.9	9.9	10.0	9.7
9.6	9.7	10.1	10.1	9.9	9.8	9.3	

- Estimate, using a 95% confidence interval, the true mean weight of bags of this type of potato chip.
- A consumer group believes the manufacturer’s claim is overstated. They believe the true mean weight of this type of bag of potato chip is actually less than 10 ounces. The consumer group will test the claim $H_0 : \mu = 10$ *versus* $H_a : \mu < 10$.

Contact Information

Feel free to contact me:

Stephen Miller
Winchester Thurston School
Pittsburgh, PA
sm1016@gmail.com