§5.1
HYPOTHESIS TESTS USING NORMAL DISTRIBUTIONS

Tom Lewis

Spring Semester
2024

Outline

Three things

Three things

1. In many cases, a bootstrap or randomization dotplot will follow a normal distribution. The mean and standard deviation of a normal curve are related to the center and spread of the curve.
2. Counting the proportion of dots in the tail of a dotplot can be carried out by calculating the area in the tail of a normal distribution.
3. We can compute the p-value of an observed statistic using a standardized test statistic and a normal distribution.

Examples

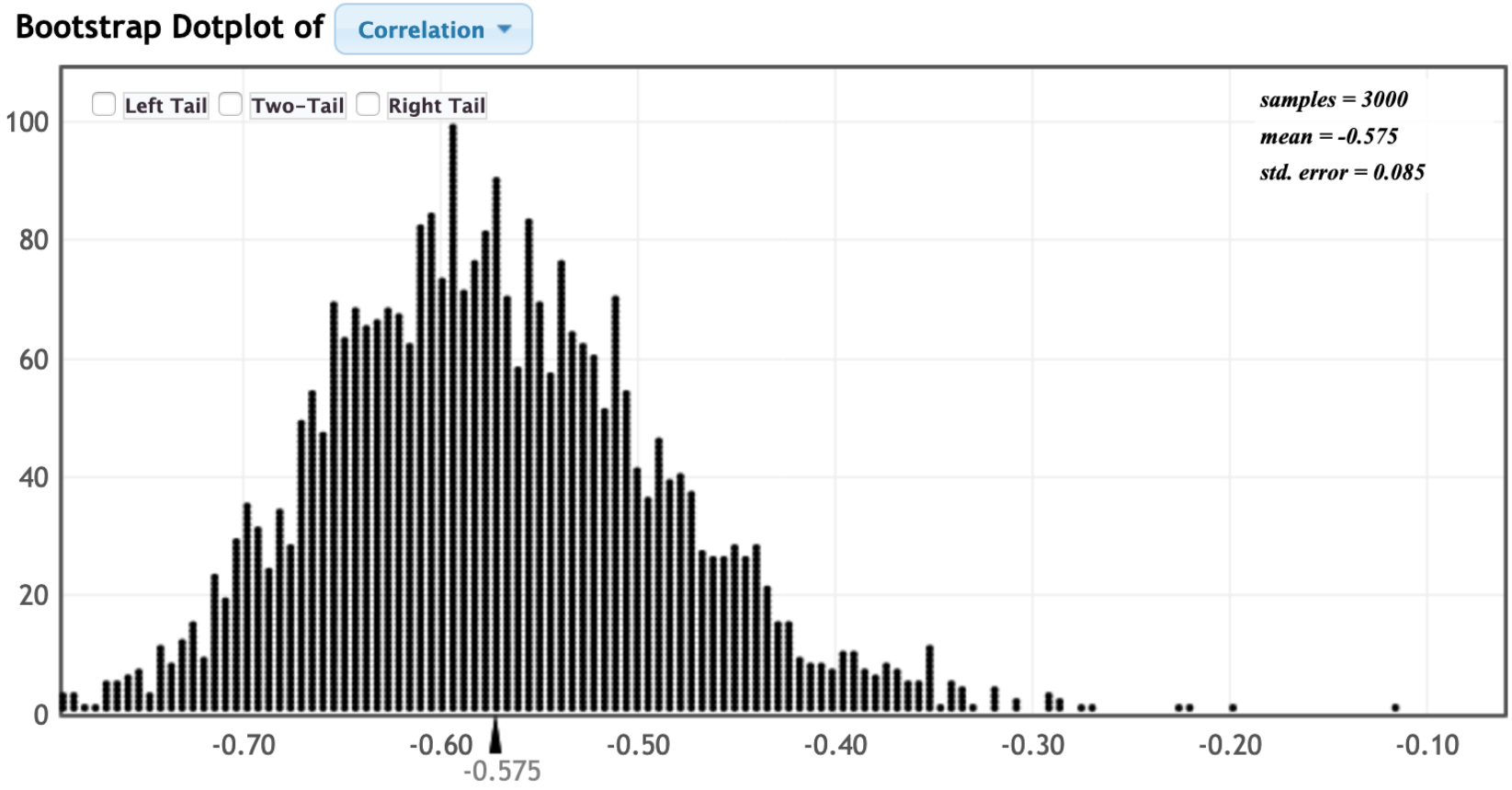


Figure: This is the bootstrap dotplot for the correlation between the pH and mercury levels in the lakes of Florida.

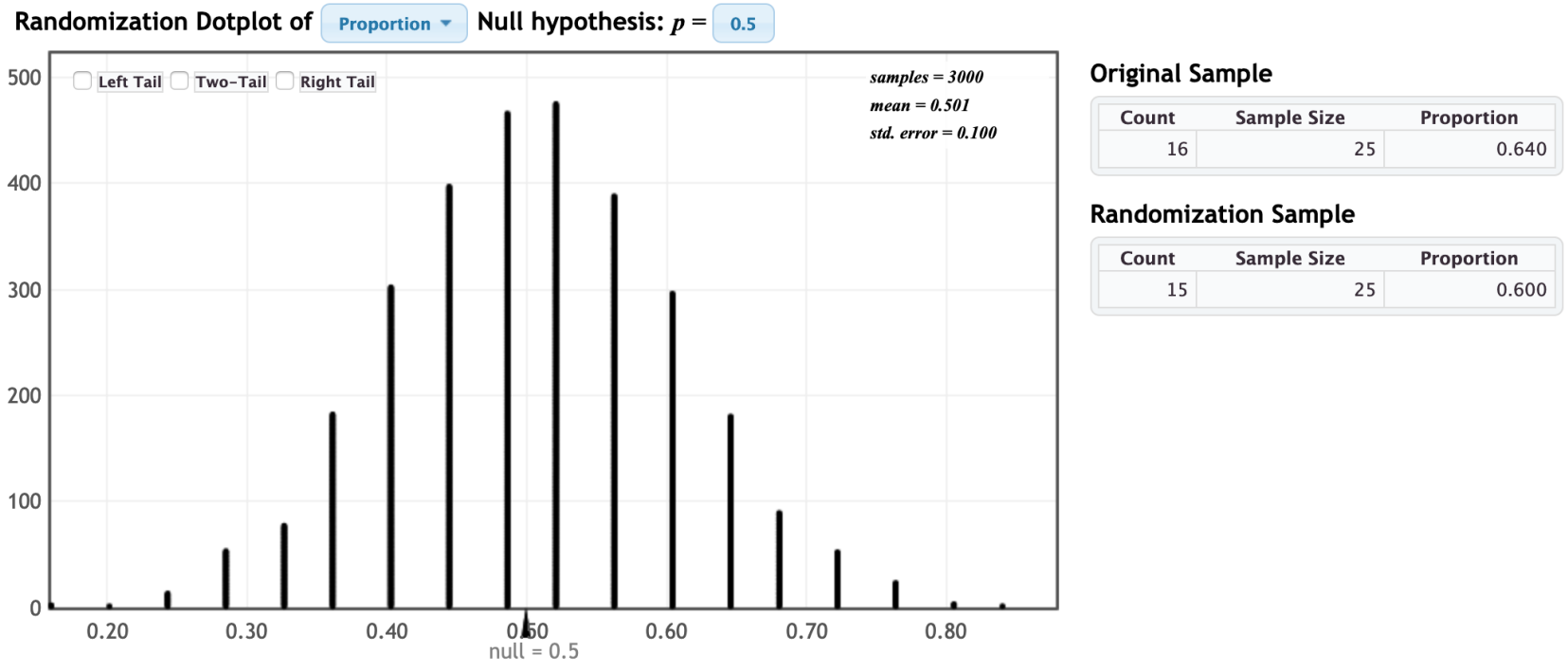


Figure: This is the randomization dotplot for the proportion of correct matches of dogs and their owners.

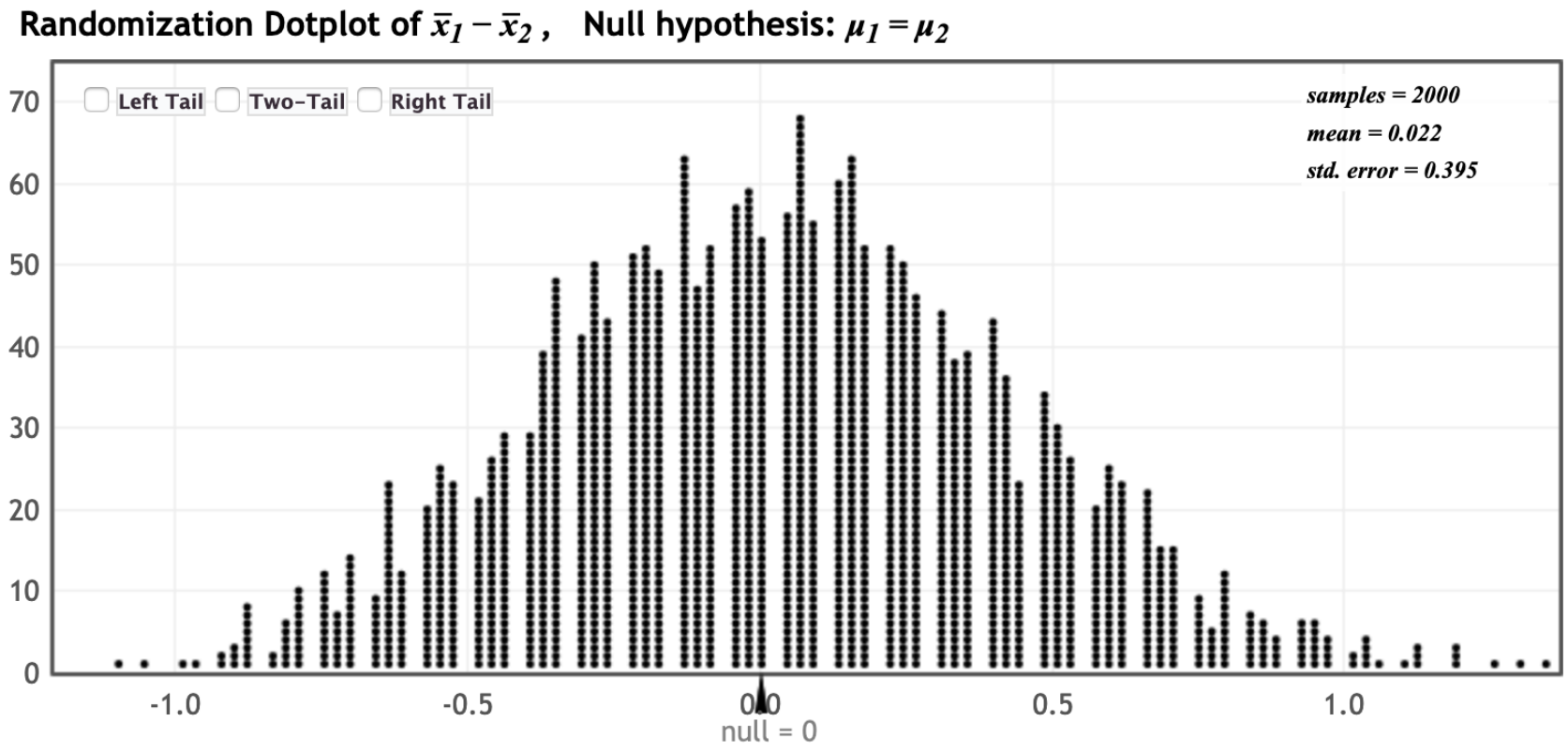


Figure: This is the randomization dotplot for the difference in means for leniency scores between smiling and neutral facial expressions.

An empirical law

Under fairly general circumstances, the distribution of many common statistics will follow the same bell-shaped pattern. The formal name for this shape is a normal distribution.

Normal distributions

Definition (Normal Distribution)

The normal distributions are a family of distribution curves. Each member of the family is specified by two parameters:

1. the mean, denoted by μ;
2. the standard deviation, denoted by σ.

A normal distribution follows a bell-shaped curve.

For shorthand we often use the notation $N(\mu, \sigma)$ to specify a normal distribution with parameters μ and σ.

Features of a normal distribution curve

A normal distribution curve with parameters μ and σ will have the following features:

1. it will be symmetric about the line $x = \mu$;
2. The curve is "effectively positive" on the interval from $\mu - 4\sigma$ to $\mu + 4\sigma$.
3. There is a unit of area trapped between a normal curve and the x-axis.

Problem

Go to StatKey and select ***Normal*** *from the* ***Theoretical Distributions*** *panel. Examine the graphs of various normal distributions.*

Problem (Calculating areas)

Use StatKey to calculate the following:

1. *Find the area trapped between the* $N(5, 2)$ *distribution curve and the* x*-axis for* $x \geqslant 7.6$*.*
2. *Find the area trapped between the* $N(150, 15)$ *distribution curve and the* x*-axis for* $x < 130$ *and* $x > 170$*.*

The standard normal distribution

Definition

The standard normal distribution is a normal distribution with mean zero and standard deviation equal to one, that is, $N(0, 1)$.

Standardizing

Let x be a quantity distributed according to a $N(\mu, \sigma)$ scale. Then the standard score for x, that is,

$$z = \frac{x - \mu}{\sigma}$$

will be distributed on a $N(0, 1)$ scale.

The effect of standardizing

A standardized test has a mean of 70 and a standard deviation of 10. Here are the scores on 500 randomly sampled tests.

$$t = \{35, 37, 40, 43, 44, \ldots, 96, 97, 97, 97, 100\}.$$

Here are the standardized scores of the test data:

$$s = \{-3.5, -3.3, -3.0, -2.7, -2.6, \ldots, 2.6, 2.7, 2.7, 2.7, 3.0\}$$

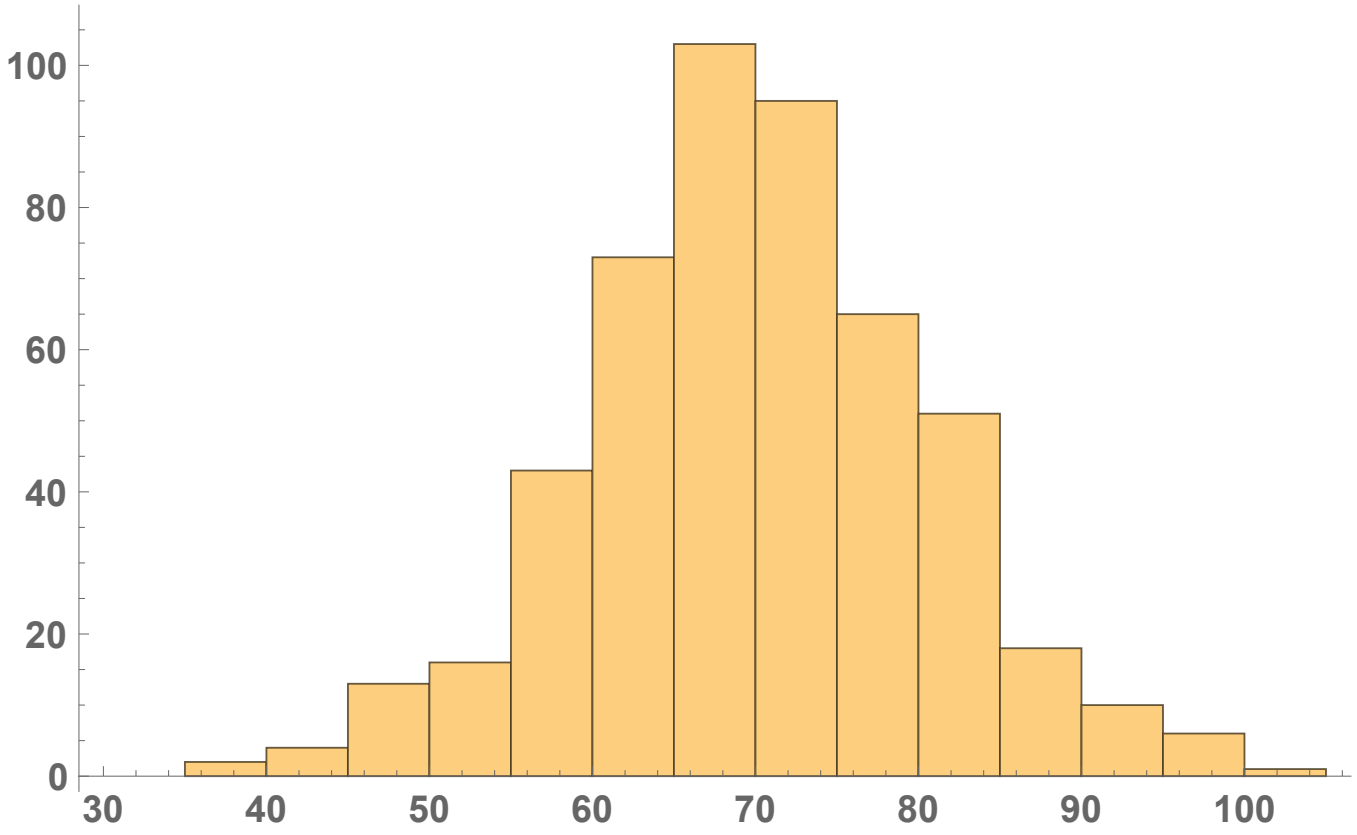


Figure: The histogram of the test data, t

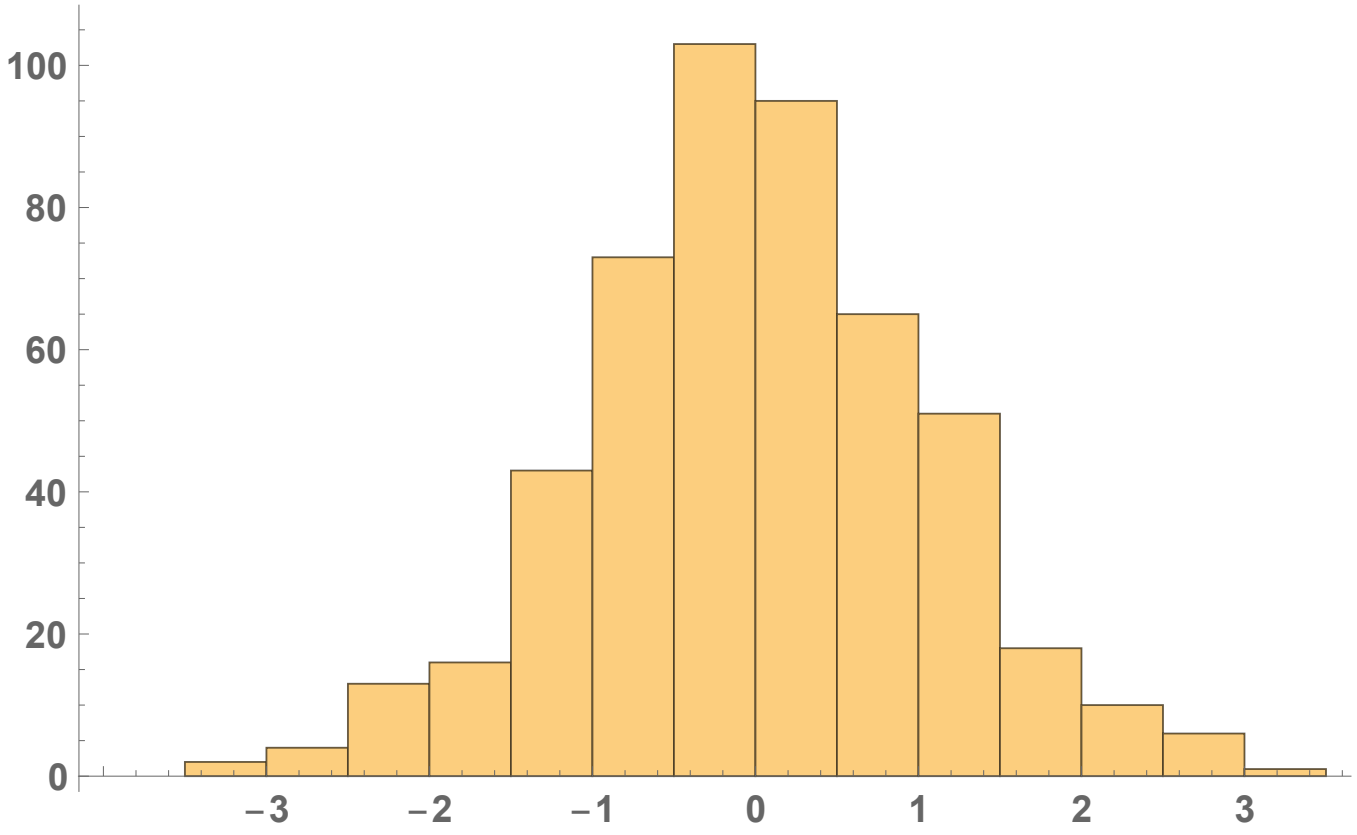


Figure: The histogram of the standardized test data, s

Problem

Find the area to the right of 95 in a normal distribution with mean 80 and standard deviation 10. Do this in two different ways:

1. *Calculate the area directly through StatKey.*
2. *Find the z-score for the endpoint of 95 and find the area to the right of that point under a standard normal distribution. Use StatKey.*

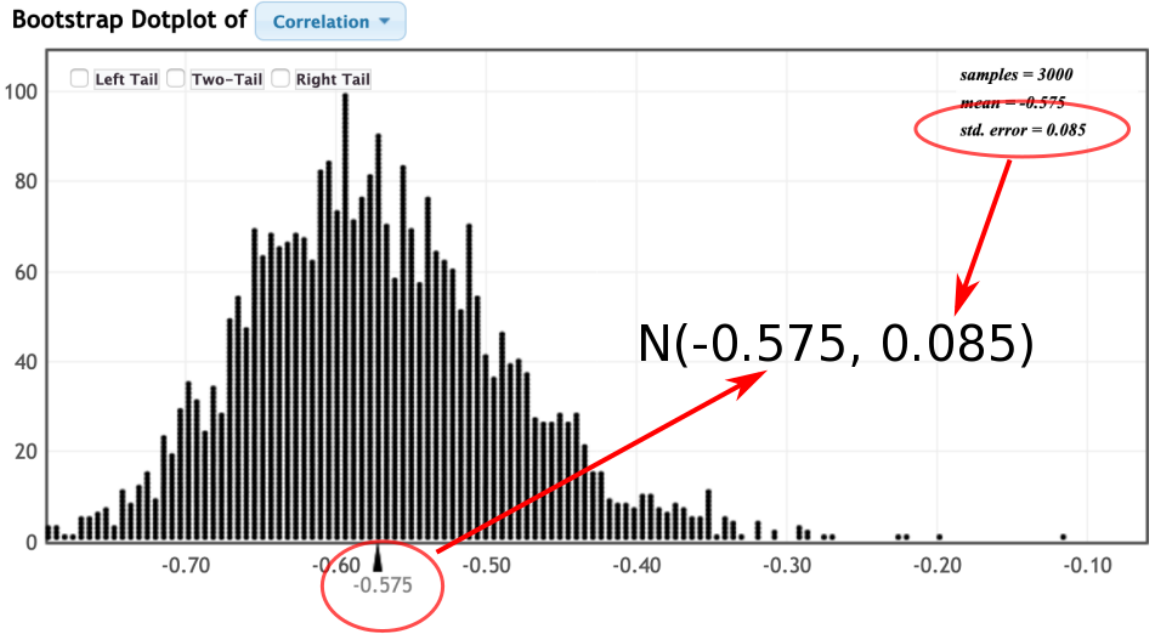
Dotplots

A dotplot and its corresponding normal curve

The normal curve $N(\mu, \sigma)$ corresponding to a dotplot can be found by setting

$\mu =$ mean of the dotplot

$\sigma =$ standard error (SE) of the dotplot



A p-value calculation

The proportion of the dots that are greater than or equal to r can be calculated in three different ways:

- Method 1 Count the number of dots that are greater than or equal to r and divide by the total number of dots in the plot.
- Method 2 Find the area trapped under the $N(\mu, \sigma)$ curve to the right of r.
- Method 3 Find the area trapped under the $N(0, 1)$ curve to the right of the z-score of r.

The left tail and two tail analogs hold as well.

Problem (Calculating p-values through normal curves)

Go to StatKey. Go to Randomization Hypothesis Tests $\rightarrow$ Test for Difference in Means and load the file Leniency and Smiles. Generate several thousand dots.

1. *What are the mean and standard error of the dotplot?*
2. *Calculate the p-value of the observed statistic $\overline{x}_1 - \overline{x}_2 = .79$ using the dotplot.*
3. *Calculate the p-value using a standard normal curve.*

Hypothesis testing, II

Problem (5.22, Part I)

In a random sample of 1500 adults of all ages in the US, 990 said television was one of their main sources of news. Does this provide evidence that more than 65% of all adults in the US used television as one of their main sources for news in 2010?

A randomization distribution for this test shows SE = 0.013. Find a standardized test statistic and compare it to the standard normal to find the p-value. Show all details of the test.

Problem (5.22, Part II)

1. *State H_0 and H_a*
2. *Find the z-score of the observed statistic.*
3. *Calculate the p-value of the observed statistic using the standard normal curve.*

Problem (5.24, Part I)

A study of 138 penalty shots in World Cup Finals games between 1982 and 1994 found that the goalkeeper correctly guessed the direction of the kick only 41% of the time. The article notes that this is "slightly worse than random chance."

We will treat these data as a "sample" of all World Cup penalty shots ever. Test at a 5% significance level to see whether there is evidence that the percent guessed correctly is less than 50%. The sample size is large enough to use the normal distribution. The standard error from a randomization distribution under the null hypothesis is SE = 0.043.

Problem (5.24, Part II)

1. *State H_0 and H_a*
2. *Find the z-score of the observed statistic.*
3. *Calculate the p-value of the observed statistic using the standard normal curve.*