# Sampling Distributions and Null Hypothesis Statistical Testing: Go Fish

Ben Babcock
University of Minnesota

# Sampling Distributions: The Basic Idea

Imagine that we wanted to get a better idea about the mean height of all Americans ($\mu$ of height). We would draw *a bunch* of samples of 9 people and measure the height of each person. We would then take the means height of each sample. We would finally made a histogram of MEANS. What would happen?

A. We would get bored.
B. We would spend a lot of money on participants.
C. A new type of distribution would emerge.
D. All of the above.

# Sampling Distributions

The new type of distribution is called a sampling distribution. A sampling distribution is like a *density (histogram-ish) of a statistic* from a bunch of different samples. The distribution has variability because of sampling error. Standard error is the standard deviation of a sampling distribution. H, 148.

Take the height example.
How could we have sampling error?

What would your sampling distribution look like if your samples had *no sampling error* (perfect sampling from the pop.)?

# Sampling Distribution of the Mean

The sampling distribution of the mean is the most used sampling distribution in statistics. *Conceptually* you obtain the distribution from taking a bunch of samples from the population, finding the mean for each sample, and plotting all the means. (H, 149)

The central limit theorem (H, 281) helps define the behavior of sampling distributions of the mean. This theorem states that with most distributions, the sampling distribution of the mean will become ___________________ as the sample size increases.

# The Central Limit Theorem

You need to know what the Central Limit Theorem (CLT) says.

The CLT states that the following is true for most distributions:

If a population distribution has a mean $\mu$ and a standard deviation of $\sigma$, the sampling distribution of the mean using sample size $N$ will have a mean of $\mu$ with a standard deviation of $\sigma / \sqrt{N}$ . As $N$ becomes large, the sampling distribution of the mean will become normal. The quantity $\sigma / \sqrt{N}$ is called the ______________________.

H, 281

# The Central Limit Theorem in Symbols

1. Start with most any population distribution.

2. Sample size = $N$

3. $\mu$ (mean of population) = $\mu$ (mean of sampling dist. of mean)

4. Standard deviation (sd of population) = $\sigma$

Standard error of the mean (sd of sampling dist. of mean)

$$= \frac{\sigma}{\sqrt{(N)}}$$

5. As $N$ goes up, the sampling dist of the mean looks more normal. To the dry-erase board!

# Example

Suppose you gave the WAIS to 25 University of Minnesota students. You sample has a mean of 105.3 with a standard deviation of 15. Let's estimate the mean and standard error of the sampling distribution of the mean for this population.

$\mu$ estimate: $\sigma$ estimate:

standard error estimate:

*What means* encompass the middle 95% (aprox.)?

*What % of means* will be greater than 111.3?

# Hypothesis Testing: Vocabulary

Null hypotheses. This is the hypothesis that you are attempting to ______. IT DOES NOT HAVE TO BE A ZERO EFFECT HYPOTHESIS!!!!!!!!!!! The null hypothesis includes that the *parameter(s)* of interest is ________ to some other value.
Notation: $H_0$, $H_N$

Alternative hypothesis. This hypothesis generally deals with a parameter of interest not being equal to some value.
Notation: $H_1$, $H_A$

$H_0$: $\mu_1 = \mu_2$
$H_1$: $\mu_1 \neq \mu_2$

$H_0$: $\mu_1 \leq \mu_2$
$H_1$: $\mu_1 > \mu_2$

H, 150-152.

# Sampling Distributions: Testing Hypotheses

Among other reasons, sampling distributions are important in the context of hypothesis testing. Hypothesis testing is a decision-making process about __________. Abbreviated NHST. (H, 147)

Examples of questions followed by an alternative hypothesis:
Is the mean score on the WAIS higher for males or females?
I hypothesize that the mean female score will be higher than the male score.

Is the mean height of men who enlist in the U. S. Army different than the mean height of U. S. men?
I hypothesize that the Army male mean height is different than the overall male mean height.

# Hypothesis Testing: Why is it Hard?

Let's take the question from the previous page: Is the mean height of men who enlist in the U. S. Army different than the mean height of U. S. men?

This should be an easy question to answer, right? Just sample the nearest men in uniform, find their mean height, and see if the mean is greater than the mean height for all men.

Why isn't it that easy?

# Hypothesis Testing: The Logic

We state our null and our alternative hypotheses. When taken together, the two hypotheses are:

1. Mutually exclusive.

2. Mutually exhaustive.

We test the <u>null</u> hypothesis. We assume that the null is true. <u>If we have a low probability that our data or more extreme data occurred *given* the null is true, we reject the null hypothesis</u>.

If $P(\text{data}|\text{null})$ is small, we reject the null.

YOU NEVER <u>PROVE</u> THAT THE ALTERNATIVE IS TRUE!
YOU NEVER <u>PROVE</u> THAT THE NULL IS TRUE!

# Hypothesis Testing: More Vocabulary

Howell covers some important vocabulary on pages 155-156.

*p*-value: Many tests will yield a so-called "*p*-value". This is the probability that our result or a result _________________ occurred given that the null hypothesis is true.

*p*-value = $P(\text{data}|\text{null})$.

I repeat, *p*-value = $P(\text{data}|\text{null})$.

If your *p*-value is smaller than your rejection level, you reject the null hypothesis. The decision to set a rejection level (or critical value) at a given number is an ____________ decision. To the dry-erase board for some NHST steps!

# Hypothesis Testing: The Formalities

Example question: Is the mean height of men who enlist in the U. S. Army different than the average height of U. S. men?

$H_0$: $\mu_{\text{Army}} = \mu_{\text{Overall}}$ or $\mu_{\text{Army}} - \mu_{\text{Overall}} = 0$

$H_1$: $\mu_{\text{Army}} \neq \mu_{\text{Overall}}$ or $\mu_{\text{Army}} - \mu_{\text{Overall}} \neq 0$

These are <u>non-directional hypotheses</u>.

Example question: Do females score more than five points higher than males on the WAIS?

$H_0$: $\mu_{\text{Females}} \leq \mu_{\text{Males}} + 5$ or $\mu_{\text{Females}} - \mu_{\text{Males}} \leq 5$

$H_1$: $\mu_{\text{Females}} > \mu_{\text{Males}} + 5$ or $\mu_{\text{Females}} - \mu_{\text{Males}} > 5$

These are <u>directional hypotheses</u>.

If we reject the null, we are confident that are data are _________.
If we do not reject, we really do not learn much at all...

# A Simple Hypothesis Test: A *z*-score Area

Suppose we knew that the WAIS $\mu = 100$ and that $\sigma = 15$. We want to know if 2801 students are different than the population.

$H_0$: $H_1$:

critical *p*-value:

Suppose that we sampled 25 "2801" students and found a mean WAIS of 109.

$\mu$: $\sigma$:

standard error:

*What % of means* will be __________________ than 109?

# Hypothesis Testing: The Error Matrix

In deciding to reject or not reject a null hypothesis, there is always a chance that you will get it wrong.

Type I Error: You reject the null when the null is true in reality.
Type II Error: You do not reject the null when the null is false in reality.

$\alpha$ and $\beta$ (H, 158-161)
Other term for $\alpha$:

Fact: Jack Bauer has never made a Type I error.

How do we know if we have made an error?

# Hypothesis Testing: The Error "Matrix"

http://video.google.com/videoplay?docid=-3170787386244810664

# Hypothesis Testing Errors and Sampling Distributions

# NHST: Some Factors that Affect Rejection

1. The size of the __________ (mean difference, etc.)

2. Sampling variation (increases the ______________________)

3. Sample size

4. The alpha level you choose

5. Conducting a one or two-tailed test (directional vs non-d.)

Which are the easiest for you to change, and which should you?
H, 293

# Hypothesis Testing: The Saga Continues

There are a LOT of people who hate hypothesis testing. Others think that hypothesis testing is useful, but most researchers are currently misusing NHST. Here are some issues with NHST:

NHST does not address “how large” an effect is.

Most people always use the nill (no effect, difference is 0) as the null hypothesis. This isn't all that interesting.

People misinterpret the _______. It isn't what we want to know!

News flash: every “equal to” null hypothesis is ________!