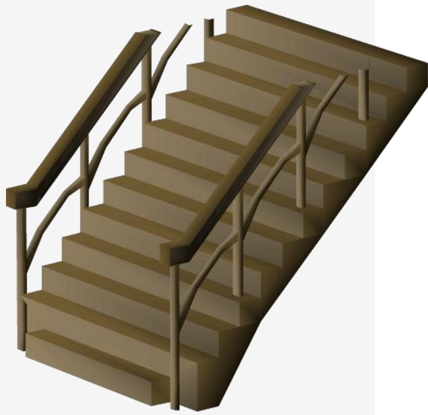
Class Business

- Office hours are today, Wed 3:30 – 4:30 pm, or Thurs 8 – 9:30 am, in case you have last-minute questions about the exam corrections assignment.
  - Assignment due on Moodle 12:30 pm tomorrow

OVERVIEW of STEPS of NHST

1. State 2 hypotheses about unknown population (null & alternative)
2. **Set your significance level (threshold for rejecting null)**
3. Calculate the test statistic & p-value
4. Make the decision to retain or reject null



- Before collecting data, need to decide what counts as good evidence for discrediting or rejecting the null hypothesis.
- **Think critically: Why should we do this BEFORE collecting data?**
  - *It allows for a fairer, more objective, less biased test of our hypothesis.*

NHST will involve asking the question . . .

- **Assuming the <u>null hypothesis were true (credible)</u>**, what's the likelihood that we'd obtain <u>the observed data (i.e., these sample statistics) or data more extreme</u>?

H_0: $\mu_{b\text{-}s} = 80$

- **Assuming <u>the mean test score in the b-s population were 80</u>**, what's the likelihood that we'd obtain <u>a sample with a mean test score of</u> ______ or something more extreme?

Ex: Observed data

 $\bar{x}$ = 79.2

Ans: High likelihood

(*null seems very credible*)

Ex: Observed data

 $\bar{x}$ = 80.1

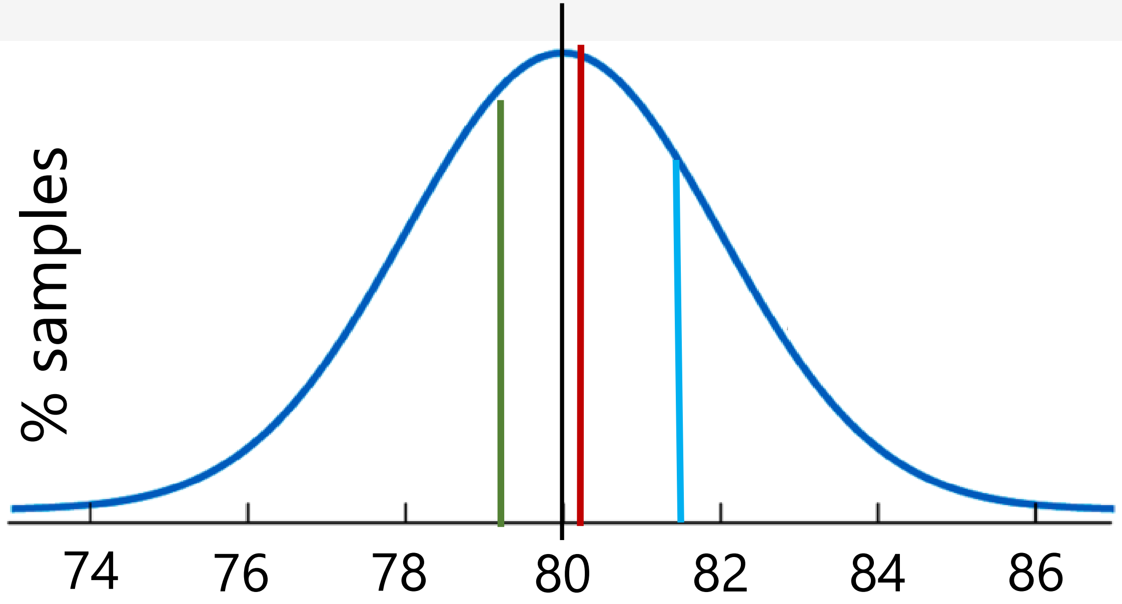
Ans: Very high likelihood (*null seems extremely credible*)

Ex: Observed data

 $\bar{x}$= 81.5

Ans: High-ish likelihood

(*null seems very credible*)



Step 2: Set your Significance Level (aka Alpha)

- Alpha (α): the probability value below which the null hypothesis will be rejected

Researchers in psych typically use a non-directional ("two tailed") test with either:

α = .05 or α = .01

Rejection region is colored in red.
Area of red region = ***α***

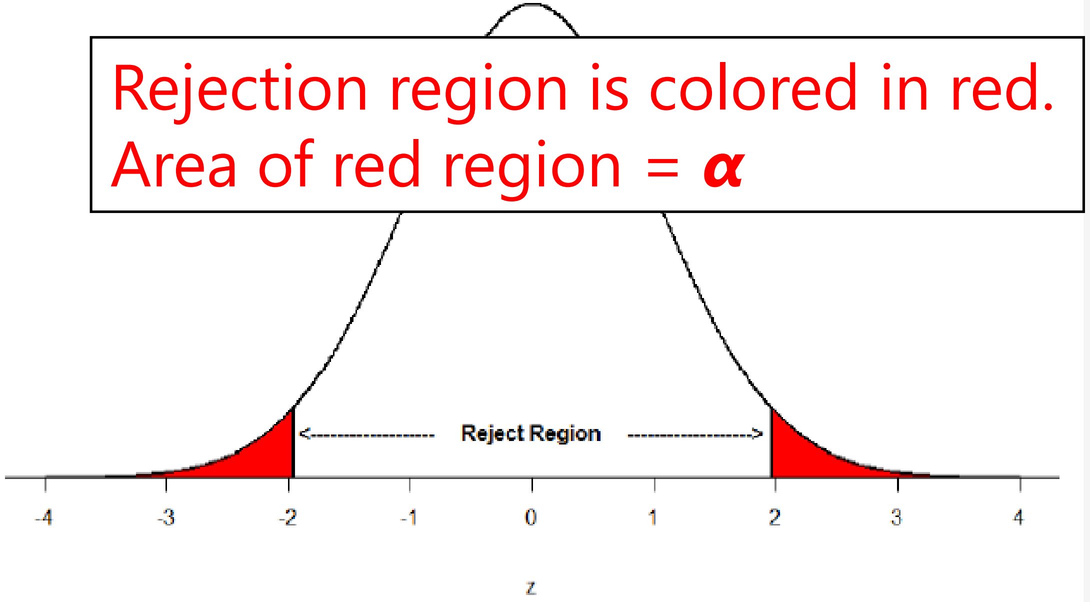
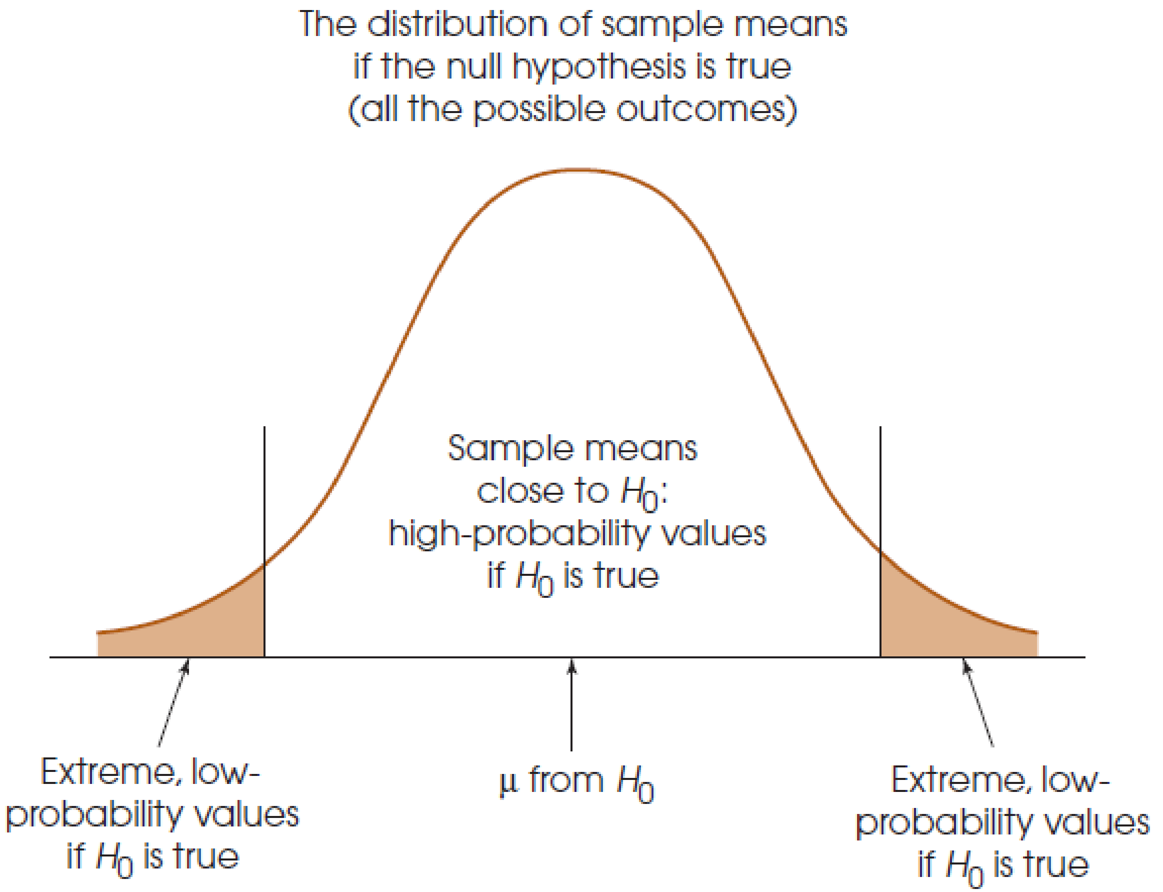


Figure 2: Two-tailed rejection region



Step 2: Set your Significance Level (aka Alpha)

- Alpha (α): the probability value below which the null hypothesis will be rejected
- Researchers in psych typically use a non-directional ("two tailed") test with: **α = .05** or **α = .01**

Recall from earlier: Assuming the null hypothesis were true (credible), what's the **likelihood** that we'd obtain the observed data or data more extreme?

- *Likelihood* is also known as *probability*

If in step 2, prior to data collection, we set α = .05 …
…we're deciding up front that we'll reject the null hypothesis if the observed sample data has less than a 5% chance of occurring when the null is true.

REMINDER --- H_0: $\mu_{b\text{-}s} = 80$
H_1: $\mu_{b\text{-}s} \neq 80$

Practice your understanding by indicating "true" or "false," and correcting false statements.

1. Alpha (α) is also know as the "significance level".
2. Researchers typically set an alpha of 95% or 99%.
3. The significance level is another way of saying the *threshold* for deciding whether you have evidence that's strong enough to discredit the null.
4. Setting the alpha at .01 indicates that if the data we obtain has a less than 1% probability of occurring if the null were true, we will decide to reject the null.
5. If you set alpha at .01, the area in the rejection region is equal to 99%.

1. True
2. FALSE – 5% or .05, not 95%
3. True
4. True
5. FALSE – if alpha is .01, the area in the rejection region is equal to 1% (.01 = 1%)

OVERVIEW OF STEPS in NHST

collect data from a sample here

1. State 2 hypotheses about unknown population (null & alternative)
2. Set your significance level (threshold for rejecting null, alpha)
3. **Calculate the test statistic & p-value *using our sample data***
4. Make the decision to retain or reject null hypothesis

Step 3: Calculate test statistic & p-value *using our sample data*

- REMINDER --- Assuming the null hypothesis were true, what's the likelihood that we'd obtain the observed (sample) data or data more extreme?

- Calculate a test statistic using the observed data & determine the p-value

- *Conceptually*, a **test statistic** = $\dfrac{\textbf{Size of the effect}\text{/difference/change/relationship}}{\textbf{Error}\text{ in measurement of that effect/difference/change/relationship}}$

There are many different test statistics, and we'll spend the rest of the semester learning about them.

Step 3: Calculate test statistic & p-value *using our sample data*

- REMINDER --- Ask the Q: Assuming the null hypothesis is true, what is the **likelihood** that we'd obtain the observed (sample) data or data more extreme?
- Calculate a test statistic using the observed data

$$\frac{\text{effect}}{\text{error}}$$

- What can the test statistic give us (indirectly)?
  - the likelihood of obtaining the observed data or data that are more extreme, assuming the null hypothesis is true

probability

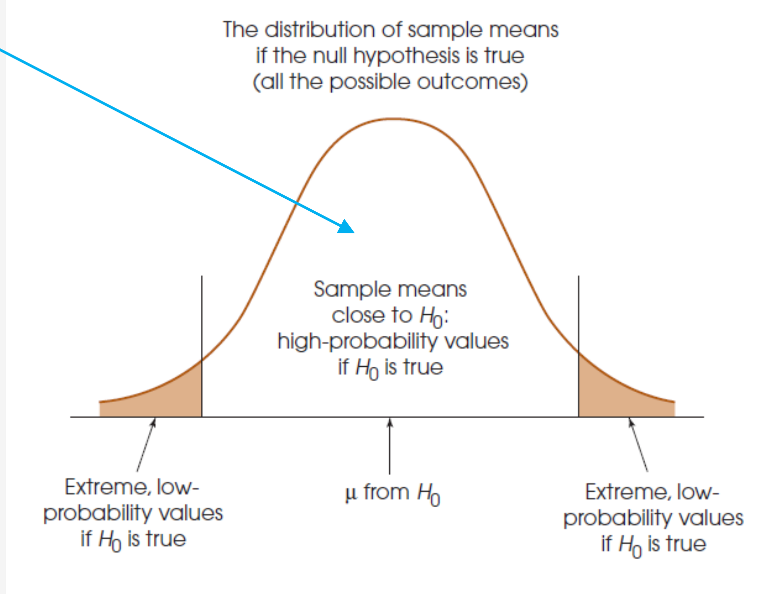
...otherwise known as the "*p*-value"

What do low vs. high p-values tell us?

Smaller test statistics are associated with **higher p-values**, which tell us . . .

- that there is a **high probability** that we'd have obtained the observed data or data more extreme, if the null hypothesis were true

. . . which lends credibility to the null hypothesis.



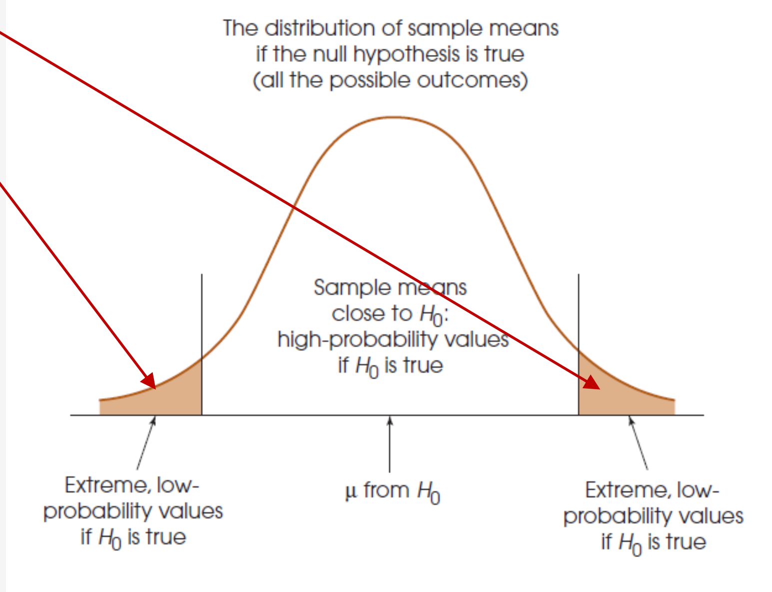
REMINDER: the p-value tells us the <u>probability of obtaining the observed data (or data that are more extreme) assuming the null hypothesis is true</u>

What do low vs. high p-values tell us?

Larger test statistics are associated with **lower p-values**, which tell us . . .

- that there is a **low probability** that we'd have obtained the observed data or data that are more extreme, if the null hypothesis were true

. . . which creates doubt that the null hypothesis is true (i.e., the data discredit the null).



Q: How "low" does the p-value need to be to create enough doubt that the null hypothesis is true (i.e., *to reject the null*)?

It needs to be lower than your alpha!

REMINDER: the p-value tells us the probability of obtaining the observed data (or data that are more extreme) assuming the null hypothesis is true

Practice your understanding by indicating “true” or “false,” and correcting false statements.

1. A p-value is conceptually defined as the ratio of effect to error.
2. Large test statistics are associated with high p-values.
3. A p-value can be defined as the probability that your sample came from the population described by the null hypothesis.
4. Low p-values suggest that the null hypothesis is *not* credible.
5. Low p-values suggest that our data has discredited the null hyp.

1 – False. This is the definition of a test statistic, not a p-value. See #3
2 – false. Large test statistics are associated with *lower* p-values, and vice versa.
3 – true.
4 – true
5 – true.

OVERVIEW OF STEPS in NHST

1. State 2 hypotheses about unknown population (null & alternative)
2. Set your significance level (threshold for rejecting null)
3. Calculate the test statistic & p-value *using our sample data*
4. **Make the decision to retain or reject the null hypothesis**

Step 4: Make the decision to retain or reject null hypothesis

HOW?

Compare p-value to your significance level (alpha).

If $p < \alpha$ (e.g., $p < .05$): **Reject** the null hypothesis

- There is less than a 5% chance that we'd have obtained the observed data or data more extreme if the null hypothesis were true. This suggests that the null is *not credible*.
- There is a *significant* effect/relationship/difference, etc.

If $p >= \alpha$ (e.g., $p >= .05$): **Retain** the null hypothesis

- There is a 5% or greater chance that we'd have obtained the observed data or data more extreme if the null hypothesis were true. This suggests that the null *is* credible.
- There is *no significant* effect/relationship/difference, etc.

Considering our blueberry supplement example . . .

- *If p* $< \alpha$ (e.g., if our p-value is $< .05$)
  - *There is a significant effect of taking b-s on cognitive functioning.*
  - Or, *taking b-s causes a significant change in cognitive functioning.*
- *p* $>= \alpha$ (e.g., if our p-value is $>= .05$)
  - *There is no significant effect of taking b-s on cognitive functioning.*
  - Or, *taking b-s causes no significant change in cognitive functioning.*

Interpretations of p-values

Example of all four steps

Johnson, Tuomisto, & Patching (2011) wanted to know if stress created and measured in the lab was higher than stress in a natural setting. They plan to compare a measure of physiological stress (heart rate) in people who have their stress created and measured in the lab with people whose stress is created and measured in a natural setting.

1. State hypotheses H_0: $\mu_{lab} = \mu_{natural\ setting}$ / H_1: $\mu_{lab} \neq \mu_{natural\ setting}$

2. Set significance/alpha level. $\alpha = .05$ *(After this point, researchers collect data.)*

3. Use statistical program to calculate test statistic from the data, and determine corresponding p-value. Let's assume test statistic = 3.05, and $p = .04$

4. Make decision to retain or reject null. Since $p = .04$, which is less than .05 (α), we reject the null and conclude: *Stress created in the lab* ***is significantly different*** *from stress created in a natural setting.*

Example of all four steps

Johnson, Tuomisto, & Patching (2011) wanted to know if stress created and measured in the lab was higher than stress in a natural setting. They plan to compare a measure of physiological stress (heart rate) in people who have their stress created and measured in the lab with people whose stress is created and measured in a natural setting.

1. State hypotheses H_0: $\mu_{lab} = \mu_{natural\ setting}$ / H_1: $\mu_{lab} \neq \mu_{natural\ setting}$

2. Set significance/alpha level. $\alpha = .05$ *(After this point, collect data.)*

3. Use statistical program to calculate test statistic from your data, and determine corresponding p-value. Let's assume test statistic = 0.15, and $p = .99$

4. Make decision to retain or reject null. Since $p = .99$, which is greater than .05 (α), we **do not reject** (i.e., we ***retain***) the null and conclude: *Stress created in the lab is **not significantly different** from stress created in a natural setting.*

Outline for Ch. 7

1. Underlying logic of null hypothesis significance testing (NHST)
2. Steps of NHST
3. **Common misunderstandings and criticisms related to NHST**
4. Effect sizes
5. Using confidence intervals for hypothesis testing (Ch. 8 pp. 154-158)
6. Types of Errors when engaging in NHST

What does a ***significant*** *p-value* mean, conceptually?

- ***It's unlikely that our sample data would look the way it does if the null hypothesis were true.***

A "significant" p-value **does not mean** the null hypothesis *is false*

A "not significant" p-value **does not mean** the null hypothesis *is true*

- "statistically significant" **does not mean** "important" (or "practically significant")
  - Is improvement from a test score of 80 to a test score of 80.8 going to make a real difference in older adults' lives?
    - NHST cannot answer this question.

Criticisms of NHST

1. It tells us nothing about practical significance (i.e., whether the effect/difference/change matters in the real world)

2. It encourages "all-or-nothing" thinking.