Chapter 9. Hypothesis Testing (Vocabulary and Procedure)

- A **"Statistical Hypothesis"** is a statement about a distribution, usually about a parameter
- H_0 **"Null Hypothesis"** is that nothing interesting or unusual is happening
- H_A **"Alternative Hypothesis"** is the opposite of the Null Hypothesis (includes your expected value)

Example. If $H_0: p = \frac{1}{2}$ then $H_A: p \neq \frac{1}{2}$.
If $H_0: \mu \leq 8$ then $H_A: p > 8$.

Goal of hypothesis testing is to show that data is incompatible with null hypothesis.
"Reject the null hypothesis"

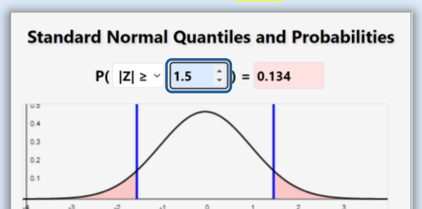
1. Compute **statistic** estimating parameter
2. Compute ***p*-value** (probability of statistic if H_0 is true)
3. Compare p-value to chosen **significance level** α
4. If $p < \alpha$, then *"Reject the Null Hypothesis"*
   If $p \geq \alpha$, then *"Fail to Reject the Null Hypothesis"*

In practice, p-value is either a **one-tailed** or **two-tailed** probability of a z-score or t-score

Two-Tailed Test

$\boldsymbol{H_0: \mu = 8}$ **and** $\boldsymbol{H_A: \mu \neq 8}$

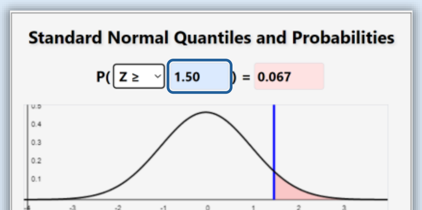
- Statistic is $\boxed{z = \frac{\overline{x} - 8}{\sigma/\sqrt{n}}}$
- p-value is $P(|Z| \geq \boxed{z})$



One-Tailed Test (Right)

$\boldsymbol{H_0: \mu \leq 8}$ **and** $\boldsymbol{H_A: \mu > 8}$

- Statistic is $\boxed{z = \frac{\overline{x} - 8}{\sigma/\sqrt{n}}}$
- p-value is $P(Z \geq \boxed{z})$



Usually tests are **two-tailed.**

Note: p-values for two-tailed tests are $2\times$ as big

- *Use one-tailed test to show something is **bigger** than something else.*

Note: α is usually 1% or 5% and **cannot** be changed.

To get $p < \alpha$, we need **small** p-value.

- **Best way:** Use lots of data. Big n yields small p.
- **Alternate:** Decrease σ? Partition data to subsets?

Note: If statistic $\overline{x}$ is far from null hypothesis μ_0, then p-value will be small even without much data!

*The closer statistic is to null hypothesis, the more data will be needed to get **significance**!*

Testing mean μ:

- ***z*-test** if σ **known** and either data is Normal or else number of samples $n > 30$
- ***t*-test** if σ **unknown** and data is Normal

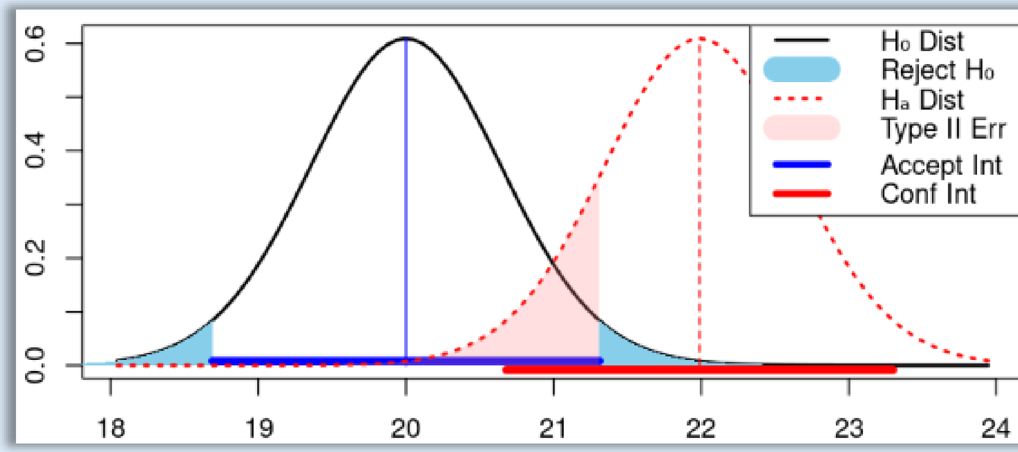
Testing population proportions:

- **Proportion test**

Chapter 9. Hypothesis Testing (Errors and Power)

	Reject H_0	Fail to reject H_0
H_0 is True	Type I Error	no error
H_0 is False	no error	Type II Error

- **Type I Error** is when null hypothesis is **true**... but your data leads you to incorrectly **reject** it.
- **Type II Error** is when null hypothesis is **false**... but your data isn't significant enough to reject.



The **significance** cutoff α used for rejecting is equal to **Probability of Type I Error** *(blue area in plot above)*

The complementary probability $(1-\alpha)$ is called the **"Confidence Level"**.
Confidence level is probability of correct conclusion if the null hypothesis is **true**.

The **Probability of Type II Error** is written β.
(red shaded area in plot above)

The complementary probability $(1-\beta)$ is called the **"Power"**.
Power of test is the probability of correct conclusion if the null hypothesis is **false**.

Goal: Design experiment to have **high confidence** and **high power** $\Longleftrightarrow$ error probabilities $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ **small**

Problem: Directly changing α to be smaller will make β bigger, and vice versa.

- Type I error is **WORSE** than Type II error
- Set α at a *small* value like 5% or 1%
- To get β small, use more data (bigger n)
  (If statistic is far from null hypothesis, then β might be small even without much data!)

Note: "**statistically significant**" does not mean "**important**".
If you have enough data, then even tiny, unimportant differences can be statistically significant.