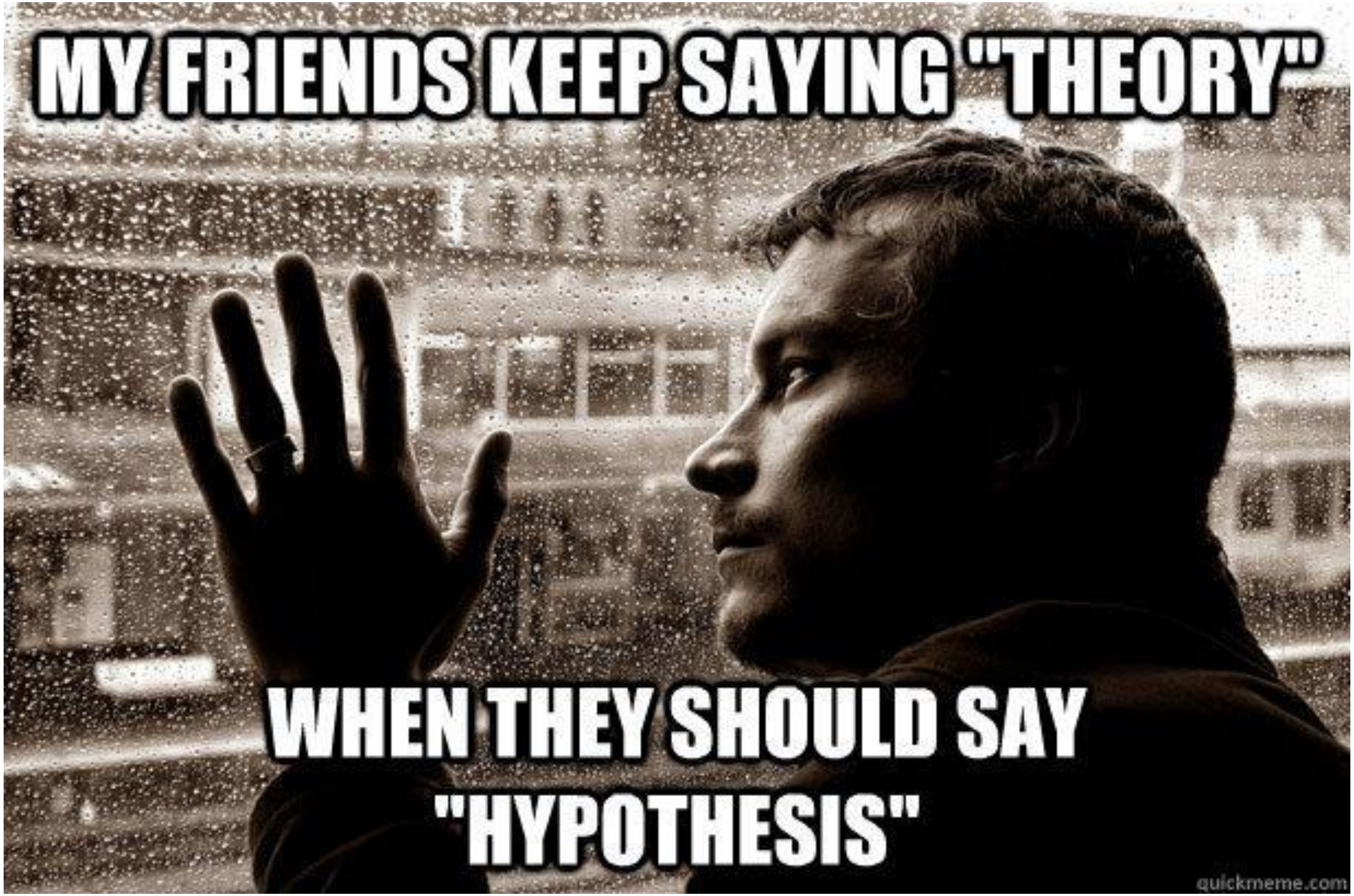
CS3121 - Introduction to Data Science

Hypothesis Testing

Dr. Nisansa de Silva,
Department of Computer Science & Engineering
http://nisansads.staff.uom.lk/


MY FRIENDS KEEP SAYING "THEORY"
WHEN THEY SHOULD SAY "HYPOTHESIS"
quickmeme.com

Hypothesis Testing

Hypothesis testing is the key to our scientific inquiry.

Through out experience and survey of previous researches related to our problem, we are able to uncover pertinent information and learn about the methods used by other researchers in investigating similar topics.

Because of this, we will be able to define our hypothesis.

Terms

- **Population** ≡ all possible values
- **Sample** ≡ a portion of the population
- **Statistical inference** ≡ generalizing from a sample to a population with calculated degree of certainty
- Two forms of statistical inference
  – Hypothesis testing
  – Estimation
- **Parameter** ≡ a characteristic of population
  e.g., population mean (μ)
- **Statistic** ≡ calculated from data in the sample
  e.g., sample mean ($\bar{X}$)

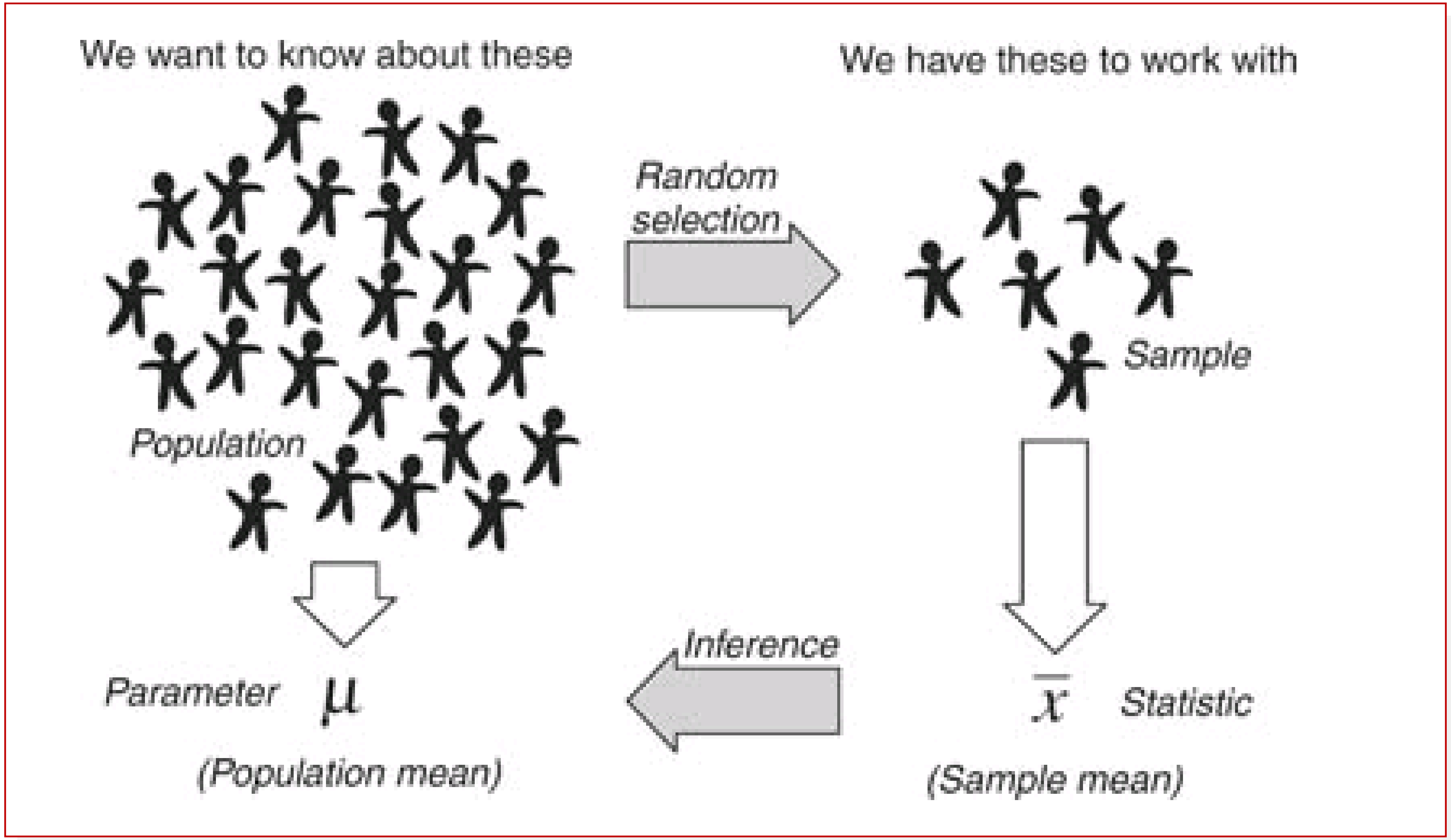
"The Tortoise And The Hare" is actually a fable about small sample sizes.

What is a Hypothesis?

A **hypothesis** is a prediction based on the body of knowledge, scientific theory, or observations.

- A statement of circumstances in the **population** that the statistical process will examine and decide whether it is true or false.
- In hypothesis testing, we test our prediction about one or more of the population parameters (or characteristics) that will either be accepted or rejected on the basis of the information obtained from the sample.
- Sample data provided us with estimates of population parameters or characteristics.
  - Statistical hypotheses are discussed in terms of the **population**, not the sample, yet tested on **samples**
- These estimates are used in arriving at a decision to either accept or reject a hypothesis.
- Based on the mathematical concept of probability
- The technique is introduced by considering a one-sample z test

Quick Overview of Statistics



What are the Kinds or Types of Hypothesis?

Null hypothesis

Alternative hypothesis

Activity: Arrange the Following Steps as you Think Fit

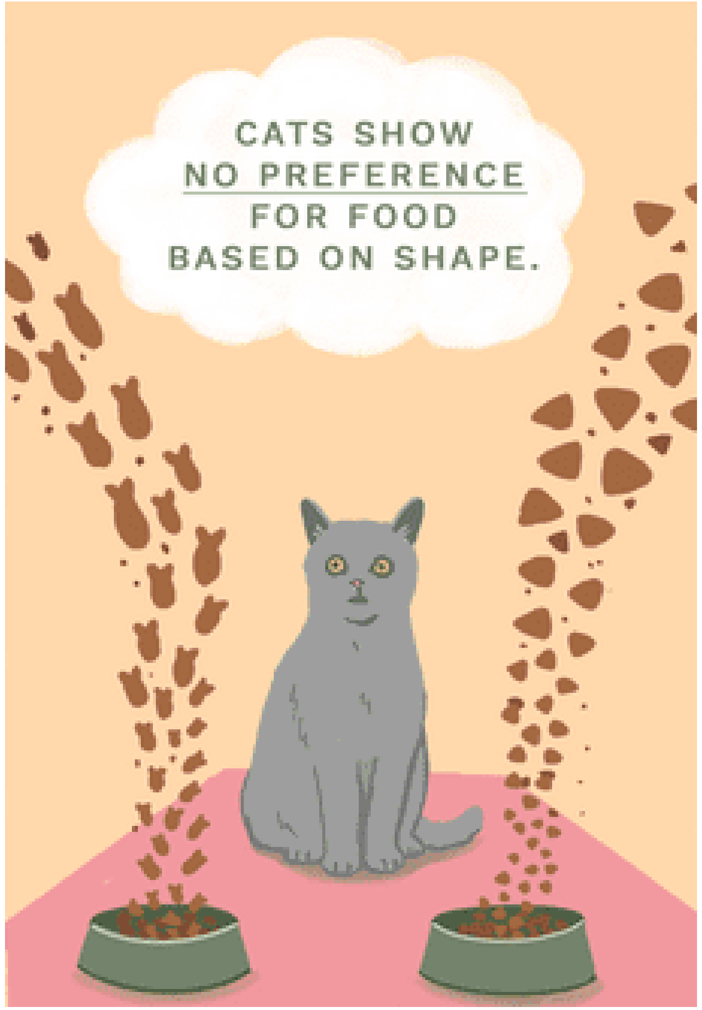
- Choose the sampling distribution and specify the critical region
- State the Null and Alternative hypothesis
- State the level of significance and sample size
- Decide on whether to accept or reject the null hypothesis
- Choose the appropriate test statistic to be used

Steps in Hypothesis Testing

1. State the Null and Alternative hypothesis
2. Choose the appropriate test statistic to be used
3. State the level of significance and sample size
4. Choose the sampling distribution and specify the critical region
5. Decide on whether to accept or reject the null hypothesis

1. State the Null and Alternative Hypothesis

- Null Hypothesis ($\boldsymbol{H_0}$)
  - Indicates the value of the population parameter to be tested
  - This is the hypothesis of "**No Difference**" (The case when the two groups are equal; population means are the same)
  - Is usually formulated for the sole purpose of being **REJECTED**.



1. State the Null and Alternative Hypothesis

- Null Hypothesis (H_0)
  - Indicates the value of the population parameter to be tested
  - This is the hypothesis of "**No Difference**" (The case when the two groups are equal; population means are the same)
  - Is usually formulated for the sole purpose of being **REJECTED**.
  - We assume this statement is true until proven otherwise.
  - This is what is called a "**Test of Significance**"

1. State the Null and Alternative Hypothesis

- Null Hypothesis ($\boldsymbol{H_0}$)
  - Indicates the value of the population parameter to be tested
  - This is the hypothesis of "**No Difference**" (The case when the two groups are equal; population means are the same)
  - Is usually formulated for the sole purpose of being **REJECTED**.
  - We assume this statement is true until proven otherwise.
  - This is what is called a "**Test of Significance**"
    - Which is a determination of the truth of a prediction.

1. State the Null and Alternative Hypothesis

- Alternative Hypothesis ($\boldsymbol{H_A}$ or $\boldsymbol{H_1}$)
  - The case when the two groups are not equal; when there is some treatment difference; when other possibilities exist
  - Is the operational statement of the experimenter's research hypothesis.
  - The Null Hypothesis is the one being tested statistically. Once the null hypothesis is rejected, the Alternative Hypothesis is **assumed** to be **TRUE** or **ACCEPTED**.
- Statistical Hypotheses
  - The $\boldsymbol{H_0}$ and $\boldsymbol{H_A}$ must be mutually exclusive
  - The $\boldsymbol{H_0}$ and $\boldsymbol{H_A}$ must be exhaustive; that is, no other possibilities can exist
  - The $\boldsymbol{H_A}$ contains our **Research Hypothesis which** is the prediction derived from the theory being tested.

1. State the Null and Alternative Hypothesis

Example

There is no significant difference between the booster from Is-is fruit extract and the commercial one in terms of improving the weight of chickens.

Why do you think so? What type of hypothesis?

1. State the Null and Alternative Hypothesis

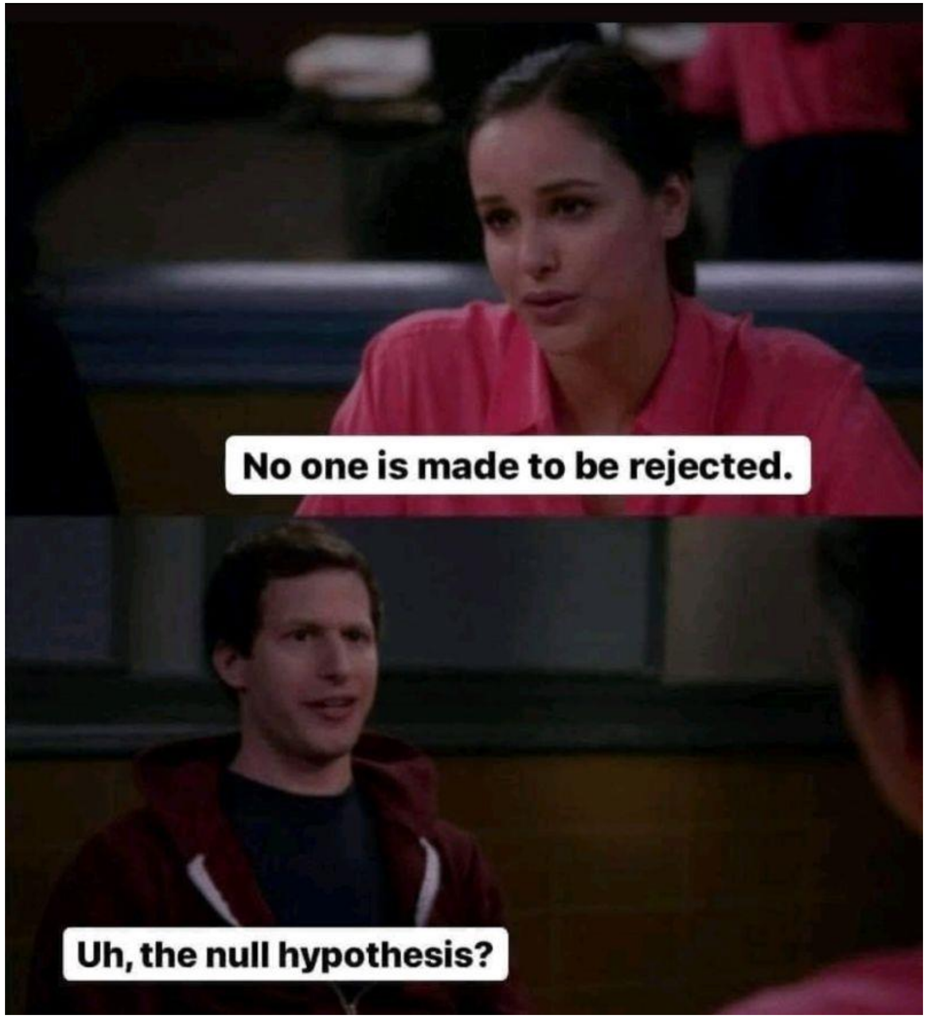
Example

Is-is plant have phytochemical properties that enables it to be a potential source of alternative booster for chickens.

Why do you think so? What type of hypothesis?

1. State the Null and Alternative Hypothesis

Test of Significance



- It is therefore a problem of deciding between the null and the alternative hypotheses on the basis of the information contained in the sample.
- In order to gain support for our research hypothesis, we must **REJECT** the null hypothesis.
- Thereby concluding that the alternative hypothesis (**likely**) reflects what is going on in the population.
- You can never “prove” the Alternative Hypothesis!

1. State the Null and Alternative Hypothesis

Test of Significance: Example

- Suppose that the standard medication for influenza is effective in 70% of all cases. A drug company believes that its new drug, “MEDIFLU”, is more effective than the old treatment.
- Formulate the null and alternative hypothesis to test if there is statistical evidence to support the new drug.
- Let μ denote the cure rate of the new drug. Since the drug company believes that MEDIFLU is better than the standard medication, the research hypothesis should be, $H_A: \mu > 70\%$
- And the null hypothesis is $H_0: \mu \leq 70\%$
- If H_0 is rejected in this case, we have to accept that H_A is true. Otherwise the reverse must be accepted.

2. Choose the Appropriate Test Statistic to be Used

Statistical Test or Test Statistic

- Is a calculated number that is used to decide whether to reject or accept the null hypothesis.
- The formula to be used for the test statistics depends on the variable we are testing.
- The level of measurement of the variable is the basis for choosing the appropriate statistics to be used in testing the hypothesis.
- Qualitative (Nominal and Ordinal Scale)
  - Cannot make use of **PARAMETRIC** statistical tests.
  - Only uses **NON-PARAMETRIC** statistical tests.
- Quantitative (Interval and Ratio Scale)
  - Can make use of both **PARAMETRIC** and **NON-PARAMETRIC** statistical tests.

2. Choose the Appropriate Test Statistic to be Used

Appropriate Statistics for the Different Scales of Measurement

Scale of Measurement	Relations being Defined	Appropriate Statistical Test to be Used	Examples of Statistics that Can be Used
Nominal	Equivalence	Non Parametric Test	Mode, Frequency, Chi Square test
Ordinal	Equivalence, Grater than, Less than	Non Parametric Test	Median, Spearman rank, Friedmann's test, Kendall's tau Percentile
Interval	Equivalence, Grater than, Less than, Known Ratio of any two intervals	Non Parametric and Parametric Test	Mean, Standard Deviation, z-test, t-test, ANOVA, Pearson's r.
Ratio	Equivalence, Grater than, Less than, Known Ratio of any two ratios	Non Parametric and Parametric Test	Mean, Standard Deviation, Coefficient of variation, Pearson's r, z-test, t-test

Note: Interval and ratio measurements can be reduced to nominal and ordinal measurements; while nominal and ordinal measured cannot be upgraded to interval or ratio measures.

2. Choose the Appropriate Test Statistic to be Used

Parametric Tests

- One sample hypothesis test of means
  - z-test
  - t-test (When the sample size is small)
- Two sample hypothesis test of means
  - t-Test (Two independent sample means)
  - t-Test (Two dependent samples/Matched-pair design experiments)
- Three or more sample hypothesis test of means
  - ANOVA (**AN**alysis **O**f **VA**riance)

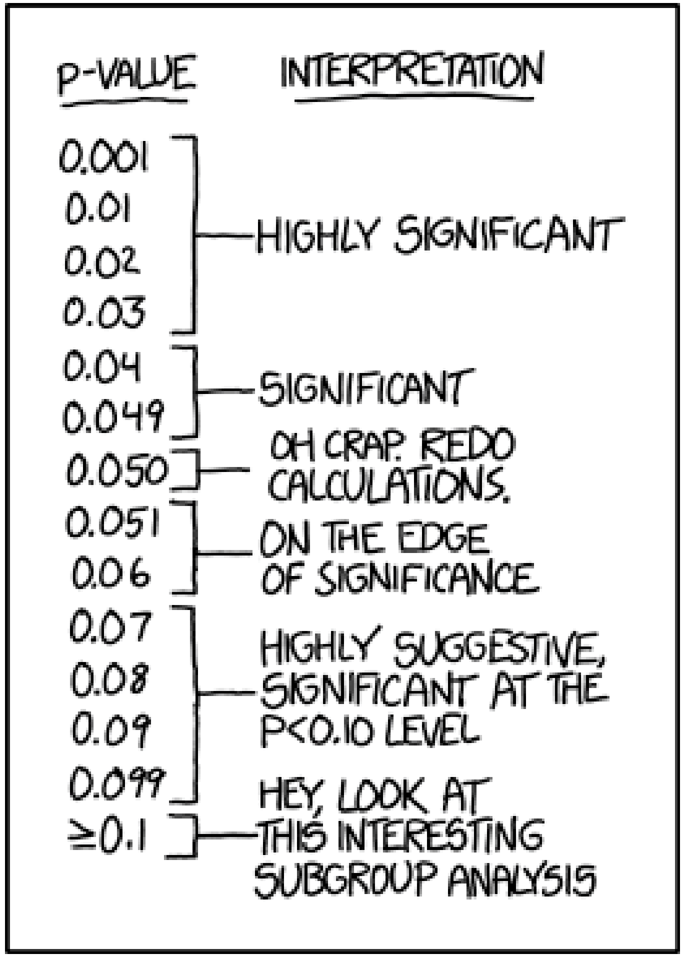
2. Choose the Appropriate Test Statistic to be Used

Non-Parametric Tests

- Chi-Square test (One way/Two way)
- The sign test
- Wilcoxon Sign Test
- Kruskal-wallis H test
- Friedman test

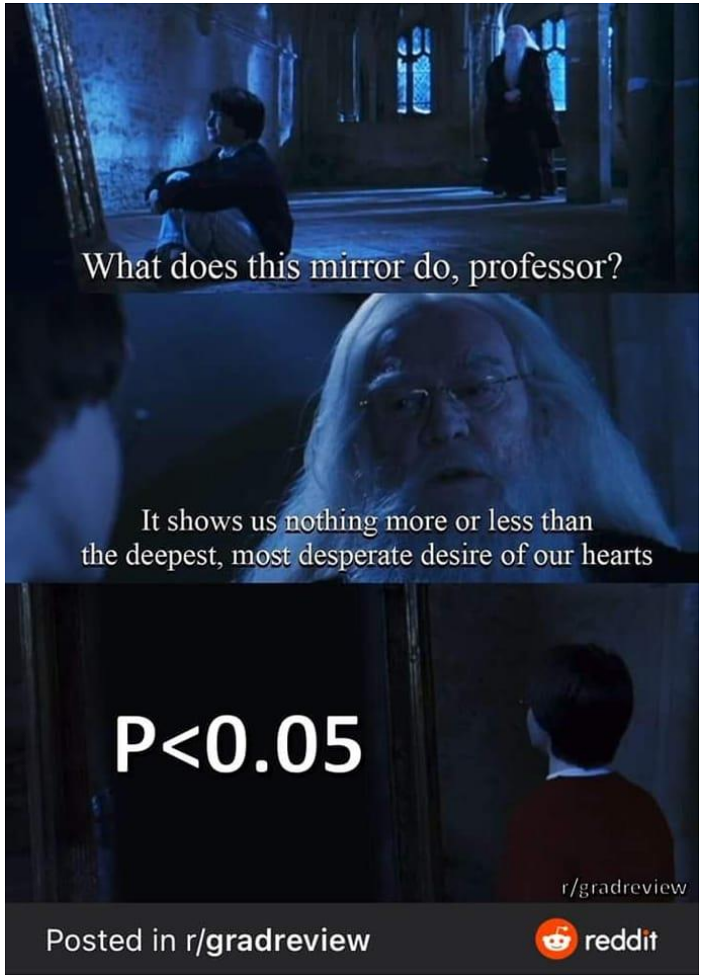
3. State the Level of Significance and Sample Size

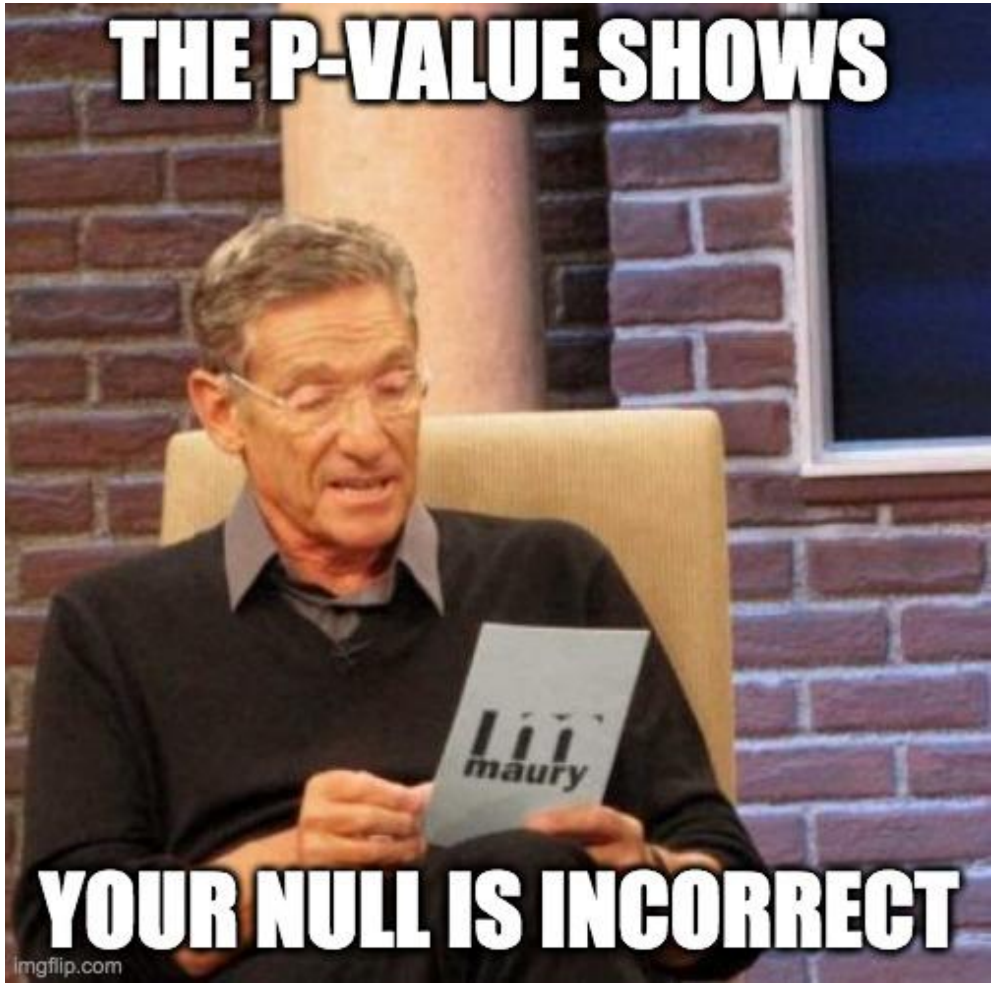
- In stating the level of significance (α), the researcher sets up the rejection and acceptance region for the null hypothesis.
- Is the probability that the test statistic will reject the null hypothesis when the null hypothesis is true
- Significance is a property of the distribution of a test statistic, not of any particular draw of the statistic
- **REJECTION REGION** – is called the critical region.
  - For results with a 90% level of confidence, the value of α is $\mathbf{1 - 0.90 = 0.10}$.
  - For results with a 95% level of confidence, the value of α is $\mathbf{1 - 0.95 = 0.05}$.
  - Typically set at 5% (.05) or 1% (.01)
- **ACCEPTANCE REGION** – is the remaining region.



3. State the Level of Significance and Sample Size

p-value



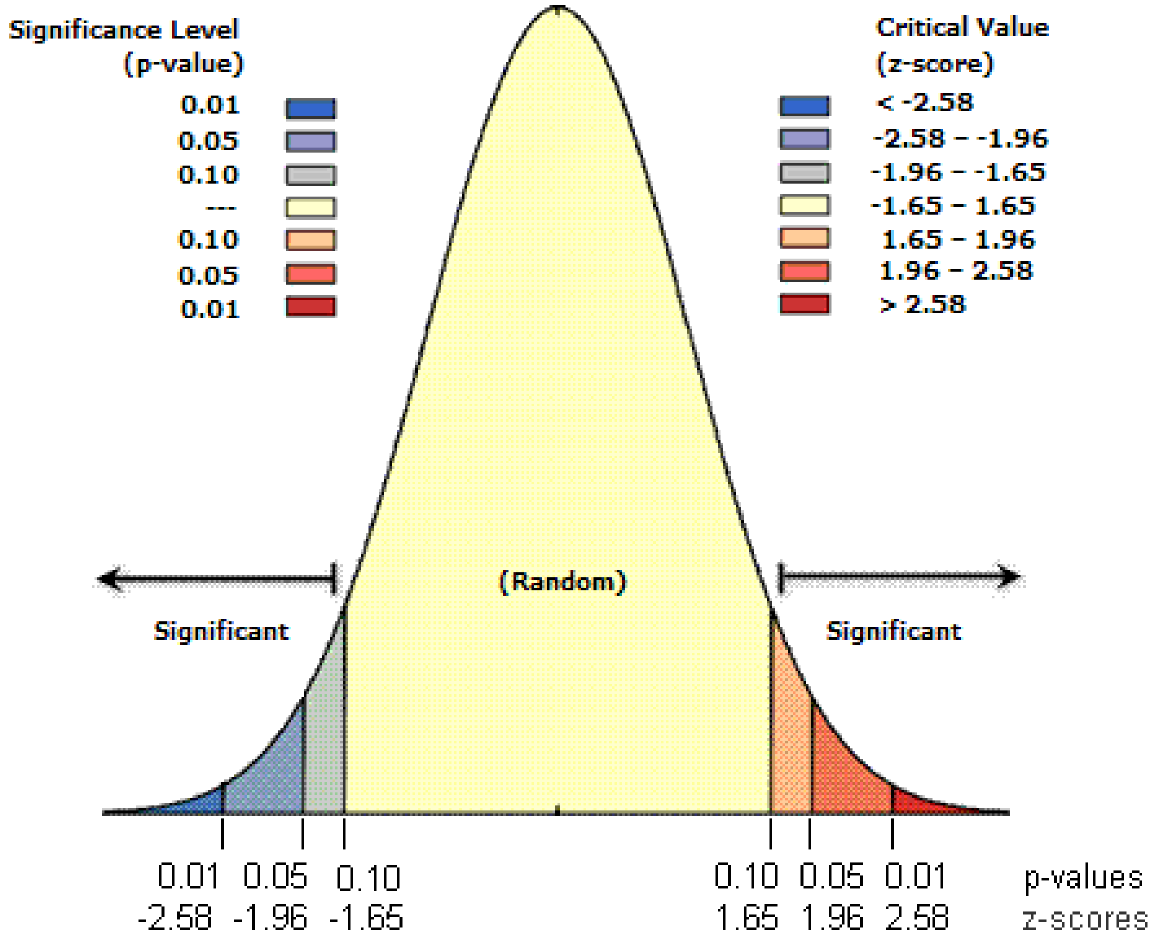


3. State the Level of Significance and Sample Size

p-value

- The ***p*-value**, or calculated probability, is the estimated probability of **rejecting** the null hypothesis (H_0) of a study question when that hypothesis is **true**.
- Probability that the observed statistic occurred by chance alone.

Values of p	Inference
$p > 0.10$	No evidence against the null hypothesis.
$0.05 < p < 0.10$	Weak evidence against the null hypothesis
$0.01 < p < 0.05$	Moderate evidence against the null hypothesis
$0.05 < p < 0.001$	Good evidence against null hypothesis.
$0.001 < p < 0.01$	Strong evidence against the null hypothesis
$p < 0.001$	Very strong evidence against the null hypothesis



3. State the Level of Significance and Sample Size

Types of Tests

Type of symbol used in the alternative hypothesis	$<$ (Less than)	$\neq$ (Not equal)	$>$ (Less than)
Example of $\boldsymbol{H_A}$	$\boldsymbol{H_A\colon \mu < \mu_0}$	$\boldsymbol{H_A\colon \mu \neq \mu_0}$	$\boldsymbol{H_A\colon \mu > \mu_0}$
Location of the rejection region to be used	One region on the left side of the distribution curve	Two regions on each side of the distribution curve	One region on the right side of the distribution curve
Type of Test	One-tailed test	Two-tailed test	One-tailed test

The **rejection regions** contain the critical value.

3. State the Level of Significance and Sample Size

Illustrative Example: “Body Weight”

- **The problem:** In the 1970s, 20–29 year old men in the U.S. had a mean μ body weight of 170 pounds. Standard deviation σ was 40 pounds. We test whether mean body weight in the population now differs.
- **Null hypothesis** H_0: $\mu = 170$ (“no difference”)
- The **alternative hypothesis** can be either
  - H_A: $\mu > 170$ (**one-sided test**) or
  - H_A: $\mu \neq 170$ (**two-sided test**)

3. State the Level of Significance and Sample Size

Level of Significance

- It is the probability that the test statistic falls within the rejection region.
- When a null hypothesis rejected or accepted, there is a risk of making an error. There are two types of errors that can be made at this point.
  - **TYPE I ERROR**: When we **reject** a **true** null hypothesis that we **should not have rejected**.
    - Same as a “false positive”
    - The level of significance (α value) gives us the probability of a Type I error.
    - Which is most commonly either 0.05 and 0.01 level.
    - For example, α = .05 = 5% chance of rejecting a true null hypothesis
    - These are the risks taken in rejecting a **TRUE HYPOTHESIS**.
  - **TYPE II ERROR**: When we **do not reject** a **false** null hypothesis that we **should have rejected**.
    - Same as a “false negative”
    - The probability of a Type II error is given by the Greek letter beta (β). This number is related to the power or sensitivity of the hypothesis test, denoted by $1 - \beta$
- The probability of making a correct decision is $1 - \alpha$

3. State the Level of Significance and Sample Size

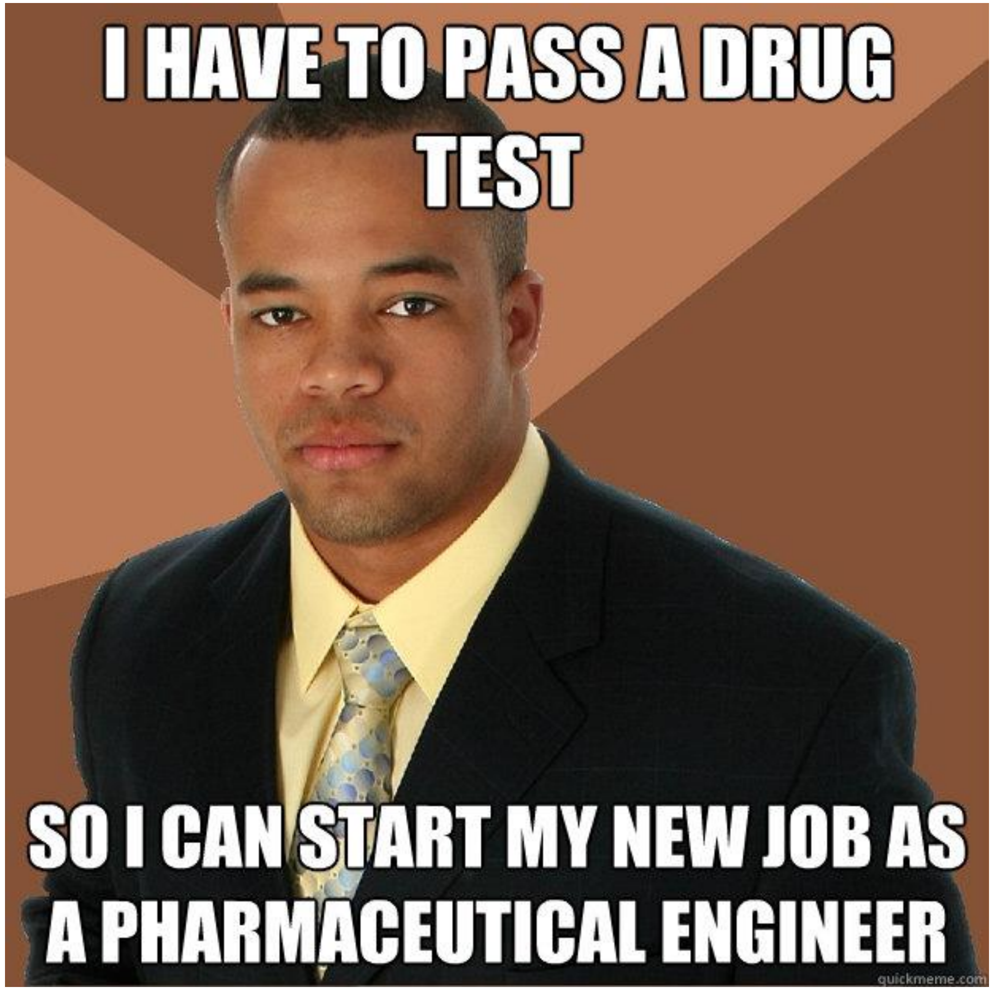
Example: Hypothesis Test as a Trial

- If we think of a hypothesis test as a criminal trial: Declaring the defendant innocent when they are actually guilty
  - H_0 : Defendant is innocent
  - H_A : Defendant is guilty
- **Type-I error**: Declaring the defendant innocent when they are actually guilty
- **Type-II error**: Declaring the defendant guilty when they are actually innocent
- Which error do you think is the worse error to make?

> "Better that ten guilty persons escape than that one innocent suffer"
> - William Blackstone

3. State the Level of Significance and Sample Size

Level of Significance: Type I Error



- Example:
  - H_0 = drug has no effect
  - H_A = drug has an effect
  - Reject H_0 and instead claim H_A is correct, so claim that the drug has an effect when indeed it does not.
  - Therefore the drug is falsely claimed to have an effect.

3. State the Level of Significance and Sample Size

Level of Significance: Type I Error

- Controlling Type I Error
  - α is the maximum probability of having a Type I error.
    - E.g., 95% Confidence Interval, chance of having a Type I is 5%
    - Therefore, a 5% chance of rejecting H_0 when H_0 is true
    - That is, 1 out of 20 hypotheses tested will result in Type I error
  - We can control Type I error by setting a different α level.
  - Particularly important to change α level to be more conservative if calculating several statistical tests and comparisons.
  - We have a 5% chance of getting a significant result just by chance. So, if running 10 comparisons, should set a more conservative α level to control for Type I error
    - Bonferroni correction: .05/10 = .005 α

3. State the Level of Significance and Sample Size

Level of Significance: Type II Error

- **Power**: The power of a test sometimes, less formally, refers to the probability of rejecting the null when it is not correct.
- $Power = P\left(\frac{reject\ H_0}{H_1\ is\ true}\right) = P\left(\frac{accept\ H_1}{H_1\ is\ true}\right)$
- As the power increases, the chances of a Type II error (false negative; β) decreases.
- Power = ***1-β***
- Power increases with sample size.

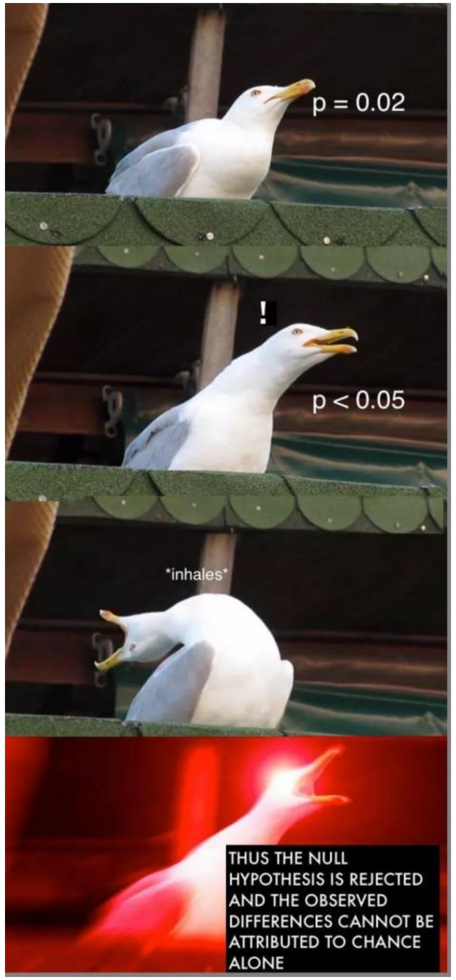
4. Choose the Sampling Distribution and Specify the Critical Region

- It is the theoretical distribution associated with the test statistic applied.
- Example
  - When we are testing a hypothesis concerning means, we may either use the standard normal distribution table (z-table) or the table for t-values or t distribution table.

5. Decide on Whether to Accept or Reject the Null Hypothesis

- We say the value of the computed statistic is significant when it falls within the rejection region. Otherwise, we have to accept the null hypothesis.

- This is done by comparing the values of $\boldsymbol{\alpha}$ and the ***p*-value**. There are two possibilities that emerge:
  - The <u>*p*-value is less than or equal to alpha</u> (e.g., ***p* ≤ .05**). In this case we reject the null hypothesis. When this happens we say that the result is statistically significant. In other words, we are reasonably sure that there is something besides chance alone that gave us an observed sample.
  - The <u>*p*-value is greater than alpha</u> (e.g., **p > .05**). In this case we fail to reject the $\boldsymbol{H_0}$. Therefore, not statistically significant. Observed data are likely due to chance alone.



5. Decide on Whether to Accept or Reject the Null Hypothesis

- In an ideal world we would always reject the null hypothesis when it is false, and we would not reject the null hypothesis when it is indeed true. But there are two other scenarios that are possible, each of which will result in an error
- There are 4 possible decisions a researcher can make in a test of hypothesis.

Null Hypothesis (H_0)	**Decision**	
	Reject H_0	**Accept H_0**
True	Type I Error (α)	Correct Decision ($1-\alpha$)
False	Correct Decision ($1-\beta$)	Type II Error (β)

- Therefore: In any decision that we make, there is always a risk of making an error.

Sample Size Requirement

Sample size for one-sample *z* test:

$$n = \frac{\boldsymbol{\sigma}^2\left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{\Delta^2}$$

where

$1 - \beta$ ≡ desired power

α ≡ desired significance level (two-sided)

σ ≡ population standard deviation

$\Delta = \mu_0 - \mu_a$ ≡ the **difference worth detecting**

Sample Size Requirement: Example

How large a sample is needed for a one-sample z test with 90% power and α = 0.05 (two-tailed) when σ = 40?

Let H_0: μ = 170 and H_a: μ = 190 (thus, $\Delta = \mu_0 - \mu_a$ = 170 – 190 = −20)

Use the z-table

$$n = \frac{\sigma^2\left(z_{1-\beta} + z_{1-\frac{\alpha}{2}}\right)^2}{\Delta^2}$$

$$= \frac{\sigma^2\left(z_{0.900} + z_{0.975}\right)^2}{\Delta^2}$$

$$= \frac{40^2(1.28 + 1.96)^2}{-20^2}$$

Round up to 42 to ensure adequate power.

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
$\boxed{1.2}$	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	$\boxed{.89973}$	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
$\boxed{1.9}$	.97128	.97193	.97257	.97320	.97381	.97441	$\boxed{.97500}$	.97558	.97615	.97670

Example 1: Independent t-test

- Research Question: Is there a difference between the reading abilities of boys and girls?
- Hypotheses:
  - H_0: There is not a difference between the reading abilities of boys and girls.
  - H_1: There is a difference between the reading abilities of boys and girls.
- **Dataset:** Reading test scores (out of 100)
- α = .05, two-tailed test
- $df = n_1 + n_2 - 2 = 10 + 10 - 2 = 18$
- Use t-Table to determine $t_{crit} = \pm 2.101$

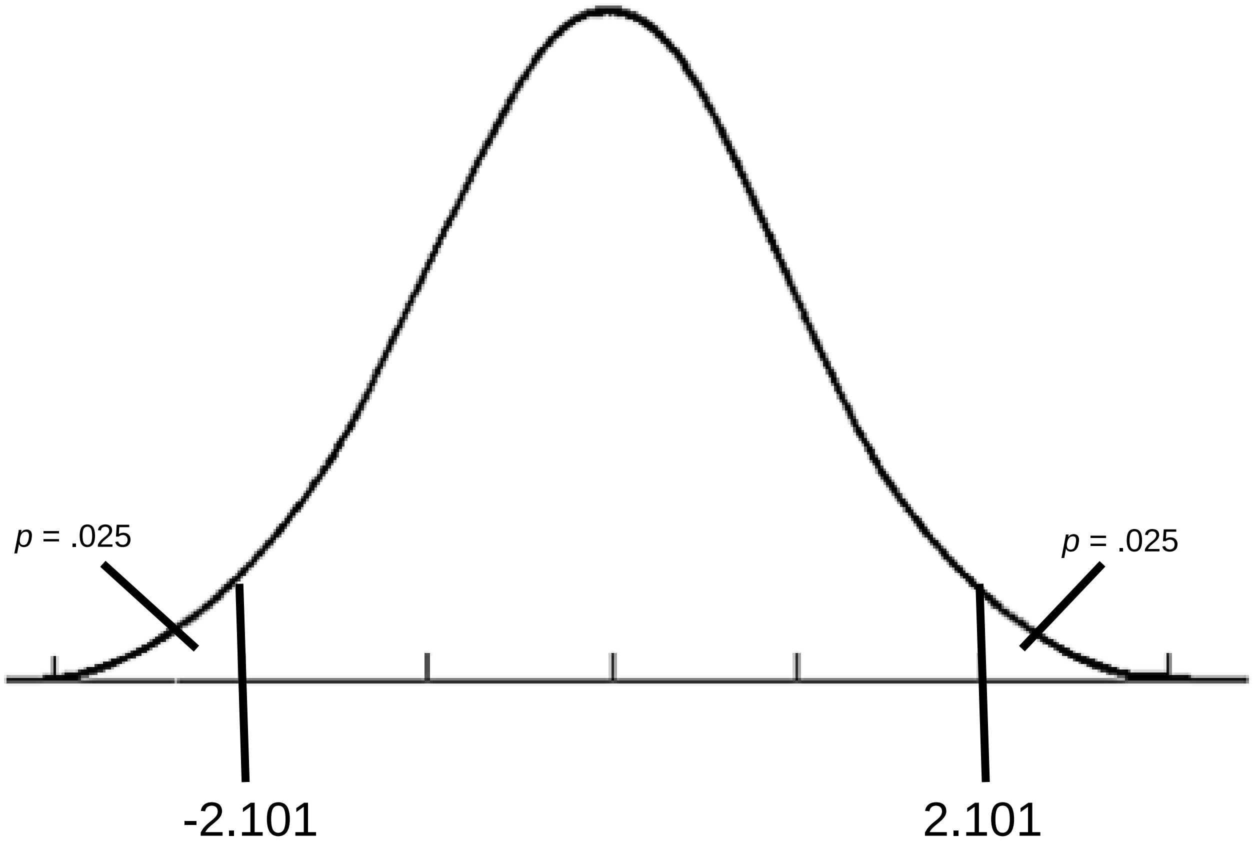
t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01
df									
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845

Example 1: Independent t-test

Decision Rules

- If $t_{calc} > t_{crit}$, then $p_{calc} < p_{crit}$
  – Reject $\boldsymbol{H_0}$
- If $t_{calc} \leq t_{crit}$, then $p_{calc} \geq p_{crit}$
  – Fail to reject $\boldsymbol{H_0}$



Example 1: Independent t-test

Computations

$$\bar{X} = \frac{\sum x_i}{n}$$

$$s^2 = \frac{\sum(x - \bar{X})^2}{n - 1}$$

$$s = \sqrt{\frac{\sum(x - \bar{X})^2}{n - 1}}$$

Boys	Girls
88	88
82	90
70	95
92	81
80	93
71	86
73	79
80	93
85	89
86	87

	Boys	Girls
Frequency (*n*)	10	10
Sum (Σ)	807	881
Mean ($\bar{X}$)	80.70	88.10
Variance (S^2)	55.34	26.54
Standard Deviation (*S*)	7.44	5.15

- Pooled variance: $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{(10-1)*55.34+(10-1)*26.54}{10+10-2} = \frac{736.92}{18} = 40.94$
- Standard error: $SE_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{40.94}{10} + \frac{40.94}{10}} = \sqrt{8.188} = 2.862$
- Compute t_{calc}: $t = \frac{\bar{X}_1 - \bar{X}_2}{SE_{\bar{X}_1-\bar{X}_2}} = \frac{80.70 - 88.10}{2.862} = -2.586$
- **Decision**: Reject $\boldsymbol{H_0}$. Girls scored statistically significantly higher on the reading test than boys did.

Example 1: Independent t-test

Confidence Intervals

- Sample means provide a **point estimate** of our population means.
- Due to sampling error, our sample estimates may not perfectly represent our populations of interest.
- It would be useful to have an **interval estimate** of our population means so we know a plausible range of values that our population means may fall within.
- 95% confidence intervals do this.
- Can help reinforce the results of the significance test.

$$CI_{95} = \bar{x} \pm t_{crit}(SE) = (80.70 - 88.10) \pm 2.101(2.862)$$

$$= -7.4 \pm 2.101(2.862) = [-13.412, -1.387]$$

Statistical Significance vs. Importance of Effect

- Does finding that $p < .05$ mean the finding is relevant to the real world?
  - Not necessarily… (More info in the "Dance of the p Values" video)
- Effect size provides a measure of the magnitude of an effect
  - Practical significance
- Cohen's **d**, **η²**, and ***R*²** are all types of effect sizes

Example 1: Independent t-test

Cohen's d (Standardized Mean Difference)

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_p} = \frac{80.70 - 88.10}{\sqrt{40.944}} = -\frac{7.4}{6.3987} = -1.16$$

- Guidelines:
  - d = .2 = small
  - d = .5 = moderate
  - d = .8 = large

- Not only is our effect statistically significant, but the effect size is large.

Example 2

The drying time of a type of paint under specified test conditions is known to be normally distributed with mean value 75 minutes and standard deviation 9 minutes. Chemists have proposed a new additive designed to **decrease** average drying time μ. It is believed that drying times with this additive will remain normally distributed with $\sigma = 9$.

- Develop a test that controls Type-I error under 5%.
  - $\boldsymbol{H_0}: \mu = \mathbf{75}$ v.s. $\boldsymbol{H_1}: \mu < \mathbf{75}$
  - This is a **one-sided/one-tailed** test. Why? We are only looking at **decrease**.
  - Test statistic $\mathrm{T} = \frac{\sqrt{n}(\bar{X}-75)}{9} \sim N(0,1)$ when $\mu = 75$ (H_0 true)
  - **Intuition:** When $T <$ some positive constant c, we would reject H_0.
  - By Type-I error rate control: Let, $\mathbb{P}\left(\frac{T<c}{\mu=75}\right) = 5\%$ i.e. $c = z_{0.95}$
  - Implying a rejection region $\bar{X} < 75 - \frac{9}{\sqrt{n}} z_{0.05} = 72.039$

Example 2

The drying time of a type of paint under specified test conditions is known to be normally distributed with mean value 75 minutes and standard deviation 9 minutes. Chemists have proposed a new additive designed to **decrease** average drying time μ. It is believed that drying times with this additive will remain normally distributed with $\sigma = 9$.

- We have collected drying time $X_1, \ldots, X_{25}$ with $\bar{X} = 69.5$ min. What's the conclusion?
  - We calculated a rejection region of $\bar{X} < 75 - \frac{9}{\sqrt{n}} z_{0.05} = 72.039$, which contains 69.5.
  - Therefore, we have sufficient evidence to reject H_0 under significance level (or Type-I error rate control level) 5%,
  - i.e. we believe that the new additive can indeed decrease average drying time μ under significance level 5%.
  - Suppose $T_0 = \frac{\sqrt{n}(\bar{X}-75)}{9} = \frac{\sqrt{25}*(69.5-75)}{9} = -3.06$ then, $\mathbb{P}\left(\frac{T<T_0}{\mu=75}\right) = \mathbb{P}(N(0,1) < T_0) = \Phi(-3.06) = 0.0011$
  - This was, obviously, the **p-value** of the test. $0.0011 < 0.05$

Example 3 with Margin of Error instead of p-value

- A drug company claims that their new drug relieves migraine 70% of the time. A news paper investigates this claim by getting migraine sufferers to try the new drug. They get 100 results that say it relives migraine 62% of the time.
- **CLAIM (H_0)**: The new drug relieves migraine 70% of the time.
- **SAMPLE:** A newspaper investigates this claim by getting migraine sufferers to try the new drug. They get 100 results that say it relieves migraine 62% of the time.
- **MARGIN OF ERROR:** $\frac{1}{\sqrt{100}} = \pm 10\%$
- **TEST :**
  Using Sample Proportion

52% **62%** **P** **70%** **72%**

Fail to Reject

Example 4 with Margin of Error instead of p-value

- A lecturer claims that 30% of university students in Sri Lanka are 172cm or taller. You collected 200 results from your batch as the results of the simple random sample of all university students and found that 34 are 172cm or taller.
- **CLAIM (H_0)**: 30% of university students in Sri Lanka are 172cm or taller.
- **SAMPLE:** Of the 200 students in the CSE sample, 34 are 172cm or taller. Is this sufficient evidence to reject the lecturer's claim, at the 5% level of significance?
- **MARGIN OF ERROR:** $\frac{1}{\sqrt{200}} = \pm 7\%$

$\frac{34}{200} = 17\%\ are\ 172cm\ or\ taller$

30%

Reject

- **TEST :**
  Using Sample Proportion

10% ← **P** → **24%**

17%

References and Links

- Hypothesis Testing: Scientific Research Manual (2008) by Ethel Grace Pedroche
- Hypothesis Testing: Minding your Ps and Qs and Error Rates by Bonnie Halpern-Felsher
- Lecture 14: Hypothesis Testing (I) by Ye Tian
- What is a z-score? What is a p-value? by ArcGIS Pro 2.9
- Hypothesis Testing by Project Maths
- Stats Primer by Jayme Palka, Peter Boedeker, Marcus Fagan, and Trey Dejong
- P-Values by xkcd
- P Value, Statistical Significance and Clinical Significance by Padam Singh
- Basics of Hypothesis Testing by B. Burt Gerstman
- Hypothesis Testing by Statistics How To
- Two-Sample T-Test: When to Use it
- Z Test: Definition & Two Proportion Z-Test