Applied Statistics

Hypothesis Testing

Cesar O. Aguilar
SUNY Geneseo

Portions of these notes were created from *Learning statistics with R* by Danielle Navarro, *Learning statistics with jamovi* by David Foxcroft, and *Introduction to Statistical Thinking* by Benjamin Yakir.

Hypothesis Testing

- Parameter estimation, covered in the last chapter, is one of the two "big ideas" in inferential statistics
- The other big idea is **hypothesis testing**
- In hypothesis testing, you as the researcher have some theory about the world and you want to determine whether the data supports the theory
- A hypothesis test is a statistical technique used to evaluate competing claims using data
- Specifically, a hypothesis test is a procedure that allows us to determine whether the data provide sufficient evidence to reject a **null hypothesis** in favor of an **alternative hypothesis**
- Frequently, the null hypothesis takes a stance of no difference or no effect
- If the null hypothesis and the data notably disagree, then we will reject the null hypothesis in favor of the alternative hypothesis

Hypothesis Testing

- Suppose that we want to determine whether a certain coin is fair or not

- We have no reason to believe that the coin is biased towards heads or tails
- We perform an experiment by flipping the coin $N = 100$ times and obtain $X = 62$ heads
- Do we have enough evidence to conclude that the coin is not fair?
- Let's denote the probability of getting a head as θ

Hypothesis Testing

- Below are different statistical hypotheses that we could test:
  - If the coin is fair, we should expect the coin to land heads half the time and so the statistical hypothesis is that $\theta = 0.5$
  - If the coin is not fair and biased towards landing heads, then the statistical hypothesis is that $\theta > 0.5$
  - If the coin is not fair but biased towards landing tails, then the statistical hypothesis is that $\theta < 0.5$
  - If the coin is not fair then we may not know whether it is biased towards heads or tails, so the statistical hypothesis is that $\theta \neq 0.5$
- Because we have no reason to believe that the coin is biased towards heads or tails, we will test the hypothesis that $\theta = 0.5$
- This is called the **null hypothesis** and is denoted by H_0

The Null Hypothesis

- The **null hypothesis** in a statistical test is the hypothesis that the data were generated by chance; it represents a skeptical perspective and is often a claim of no change or no difference
- In our example, the null hypothesis is that the coin is fair and any deviation from the expected number of heads is due to chance
- The null hypothesis is deemed to be true unless you, the researcher, can prove beyond a reasonable doubt that it is false
- You can think of a statistical test as a trial where the null hypothesis is the defendant and you are the prosecutor
- You are free to design your experiment however you like (within reason, obviously!) and your goal is to maximise the chance that the data will yield a conviction for the crime of being **false**
- The statistical test sets the rules of the trial to protect the null hypothesis, specifically to ensure that if the null hypothesis is true the chances of a false conviction are guaranteed to be low

The Alternative Hypothesis

- The **alternative hypothesis** is the hypothesis that the data were not generated by chance
- It is the claim researchers hope to prove or find evidence for, and it often asserts that there has been a change or an effect
- The alternative hypothesis is denoted by H_1 or H_A
- Because the alternative hypothesis is the hypothesis that we are trying to prove, it should be formulated before the experiment is conducted
- In our case, the alternative hypothesis is that the coin is not fair but we don't know whether it is biased towards heads or tails
- Thus, the alternative hypothesis is that $\theta \neq 0.5$

Hypothesis Test

- Our hypothesis test is as follows:

$$H_0\text{: The coin is fair with probability of heads } \theta = 0.5$$
$$H_A\text{: The coin is not fair with probability of heads } \theta \neq 0.5$$

- Depending on the outcome of the test, we will either **reject the null hypothesis** in favor of the alternative hypothesis or **fail to reject the null hypothesis** or **retain the null hypothesis**

- We will never "accept the null hypothesis" or "accept the alternative hypothesis"

Hypothesis Test Errors

- After running a hypothesis test and making a decision, there are four possible outcomes
- If the null hypothesis is **TRUE**, we can either make the correct decision or make an error
- If the null hypothesis is **FALSE**, we can either make the correct decision or make an error
- Thus, there are two types of errors that can occur in a hypothesis test:

	Retain H_0	**Reject H_0**
H_0 is true	correct decision	error (Type I)
H_0 is false	error (Type II)	correct decision

- The single most important design principle of the test is to control the probability of a Type I error, to keep it below some fixed probability

Hypothesis Test Errors

- The probability of making a Type I error is called the **significance level** of the test and is denoted by α
- A hypothesis test is said to have significance level α if the Type I error rate is no larger than α
- The probability of making a Type II error is denoted by β
- The **power** of the test is $1 - \beta$ and is the probability of rejecting the null hypothesis when it is false

	Retain H_0	**Reject H_0**
H_0 is true	$1 - \alpha$ (correct decision)	α (Type I error)
H_0 is false	β (Type II error)	$1 - \beta$ (power of test)

- By convention, scientists make use of three different levels: $\alpha = 0.05$, $\alpha = 0.01$, and $\alpha = 0.001$

Test Statistics and Sampling Distributions

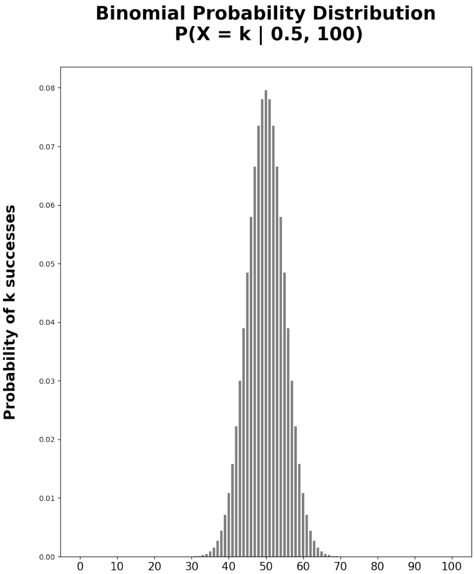
- To make a decision about whether to reject the null hypothesis, we need to calculate a **test statistic**
- In the case of our coin flip experiment, the test statistic is the number of heads X in N flips
- In this case, the underlying distribution of the test statistic is the **binomial distribution**:

$$X \sim \text{Binomial}(N, \theta)$$

- The null hypothesis is that $\theta = 0.5$ and if it were true then the most likely value of X would be $\mu = N \cdot \theta = 50$
- Moreover, if the null hypothesis were true then we're almost certain to see somewhere between 40 and 60 heads because

$$P(40 \leq X \leq 60) = 0.9648$$

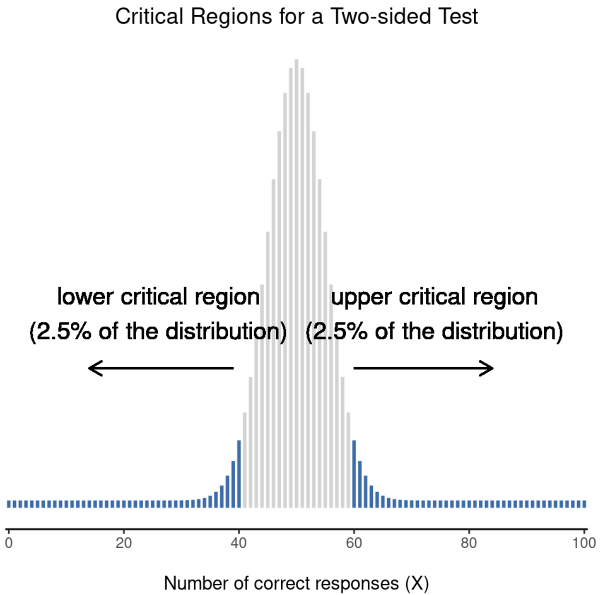
Test Statistics and Sampling Distributions



Test Statistics and Sampling Distributions

- In general, the sampling distribution in question describes the probability that we would obtain a particular value of X if the null hypothesis were actually true
- To decide to either reject or retain the null hypothesis, we introduce the concept of a **critical region** or **rejection region** for the test statistics X
- The critical region of the test corresponds to those values of X that would lead us to reject null hypothesis
- The critical region is defined by the significance level α
- For example, if $\alpha = 0.05$ then the critical region must cover 5% of the probability of the sampling distribution
- Hence, if the null hypothesis is true then the probability of incorrectly rejecting the null hypothesis is α

Test Statistics and Sampling Distributions



Test Statistics and Sampling Distributions

- At this point, our hypothesis test is essentially complete:
  1. We choose an α level (e.g., $\alpha = .05$)
  2. Come up with some test statistic (e.g., X) that does a good job (in some meaningful sense) of comparing H_0 to H_1
  3. Figure out the sampling distribution of the test statistic on the assumption that the null hypothesis is true (in this case, binomial)
  4. Then calculate the critical region that produces an appropriate α level (0-40 and 60-100).
- In our running example, $X = 62$ is in the critical region and so we reject the null hypothesis
- We then say that the result is **statistically significant** at the $\alpha = 0.05$ level
- All that "statistically significant" means is that the data allowed us to reject a null hypothesis

One-Sided and Two-Sided Tests

- The hypothesis test we conducted took the form

$$H_0: \quad \theta = 0.5$$
$$H_1: \quad \theta \neq 0.5$$

- Therefore, the alternative hypothesis H_1 covers **both** the possibility that $\theta < 0.5$ and the possibility that $\theta > 0.5$
- This is an example of a **two-sided test**
- It's called this because the alternative hypothesis covers the area on both "sides" of the null hypothesis
- As a consequence, the critical region of the test covers both tails of the sampling distribution (2.5% on either side if $\alpha = 0.05$)

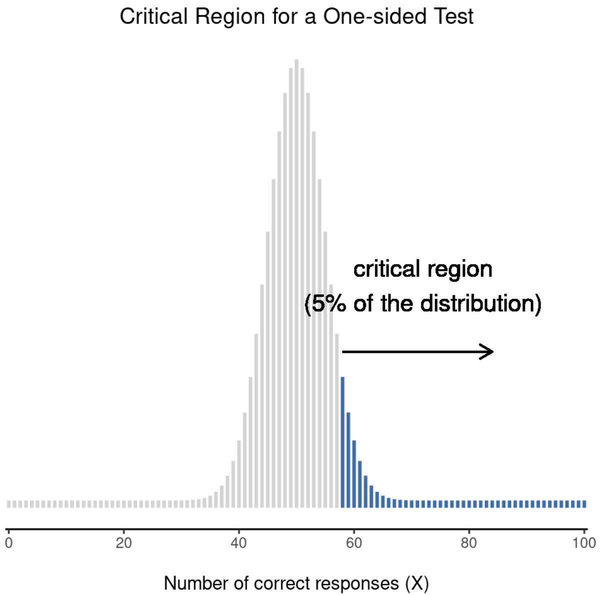
One-Sided and Two-Sided Tests

- We might instead want to perform the following hypothesis test:

$$\begin{aligned} H_0 &: \quad \theta \leq 0.5 \\ H_1 &: \quad \theta > 0.5 \end{aligned}$$

- In the coin flip example, this would correspond to the null hypothesis that the coin is fair or biased towards tails
- This is a **one-sided test**
- In this case, the critical region only covers one tail of the sampling distribution

One-Sided and Two-Sided Tests



General Hypothesis Testing

- In general, we want to test a null hypothesis H_0 regarding some parameter θ against an alternative hypothesis H_1
- The null hypothesis is $H_0 : \theta = \theta_0$
- We need an estimator $\hat{\theta}$ that has a known sampling distribution under the null hypothesis
- We fix the significance level α, if not specified use $\alpha = 0.05$
- To test the null hypothesis, we use the Z-statistic

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

- The rejection region depends on the alternative hypothesis H_1

General Hypothesis Testing

- If the alternative hypothesis is that $H_1 : \theta \neq \theta_0$ then the test is two-sided and the rejection region is

$$\left\{ Z < -z_{\alpha/2} \text{ or } Z > z_{\alpha/2} \right\}$$

- If the alternative hypothesis is that $H_1 : \theta > \theta_0$ then the test is one-sided and the rejection region is

$$\left\{ Z > z_{\alpha} \right\}$$

- If the alternative hypothesis is that $H_1 : \theta < \theta_0$ then the test is one-sided and the rejection region is

$$\left\{ Z < -z_{\alpha} \right\}$$

Summary: General Hypothesis Testing

$H_0 : \theta = \theta_0$

$$H_1 : \begin{cases} \theta > \theta_0 & \text{(upper-tail alternative)} \\ \theta < \theta_0 & \text{(lower-tail alternative)} \\ \theta \neq \theta_0 & \text{(two-tail alternative)} \end{cases}$$

Test statistic: $Z = \dfrac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$

$$\text{Rejection region: } \begin{cases} Z > z_\alpha & \text{(upper-tail test)} \\ Z < -z_\alpha & \text{(lower-tail test)} \\ Z < -z_{\alpha/2} \text{ or } Z > z_{\alpha/2} & \text{(two-tail test)} \end{cases}$$

Example: Hypothesis Test for a Population Mean

Example. A call center manager believes that employees are averaging more than 15 sales per week. To test the claim, the manager selects a random week and 36 random employees. The sample mean and sample standard deviation of sales for the 36 employees were recorded as 17 and 3, respectively. Does the collected data support the manager's belief?

- Let μ denote the average number of sales per week per employee
- The null hypothesis is $H_0 : \mu = 15$
- The alternative hypothesis is $H_1 : \mu > 15$
- The test statistic is the sample mean $\bar{X}$, which for $n \geq 30$, has an approximately normal sampling distribution with mean μ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$
- The Z-statistic is $Z = \dfrac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

Example: Hypothesis Test for a Population Mean

- The rejection region is $\{Z > z_\alpha\}$ where the critical value z_α is determined by the significance level $\alpha = 0.05$
- Using SALT we get that $z_\alpha = 1.645$ because $P(Z > 1.645) = \alpha$
- Although σ is not known, we can use the sample standard deviation $s = 3$ as an estimate because $n \geq 30$ is sufficiently large
- The observed value of the test statistic is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{17 - 15}{3/\sqrt{36}} = 4$$

- Thus, the observed Z-score is inside the rejection region, so we reject the null hypothesis and conclude that the data supports the manager's belief

Example: Hypothesis Test for a Population Proportion

Example. A grocery store receives a large order of eggs. The store will not accept the order if more than 10% of the eggs are cracked. The store manager randomly selects 100 eggs from the order and finds that 15 of them are cracked. Should the store accept the order? Use a signficance level of 0.01.

- Let p denote the proportion of eggs that are cracked in the order
- The null hypothesis is $H_0 : p = 0.1$
- The alternative hypothesis is $H_1 : p > 0.1$
- The test statistic is the sample proportion $\hat{p}$, which for $n \geq 30$, has an approximately normal sampling distribution with mean p and standard deviation $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$
- The Z-statistic is

$$Z = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$

Example: Hypothesis Test for a Population Proportion

- The rejection region is $\{Z > z_\alpha\}$ where the critical value z_α is determined by the significance level $\alpha = 0.01$
- Using SALT we get that $z_\alpha = 2.326$ because $P(Z > 2.326) = \alpha$
- The observed sample proportion is $\hat{p} = 15/100 = 0.15$
- The observed value of the test statistic is

$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} = \frac{0.15 - 0.1}{\sqrt{0.1(1-0.1)/100}} = 1.667$$

- Thus, the observed Z-score is not inside the rejection region, so we fail to reject the null hypothesis and conclude that the data does not support the store manager's belief

Mean and Variances of Commonly Used Test Statistics

θ	Sample Size	Estimator	Mean	Std. Deviation
μ	n	$\bar{X}$	μ	$\sigma/\sqrt{n}$
p	n	$\hat{p} = X/n$	p	$\sqrt{\frac{p(1-p)}{n}}$
$\mu_1 - \mu_2$	n_1 and n_2	$\bar{X}_1 - \bar{X}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	n_1 and n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

- For two samples, the samples are assumed to be independent

Example: Hypothesis Test for a Difference in Population Means

Example. A study was performed to measure the reaction times for two groups, Group A and Group B. Independent samples of 50 were gathered for each group. The average reaction time for the Group A sample was 3.6 seconds, with variance 0.18. The average reaction time for the Group B sample was 3.8 seconds, with variance 0.14. Does the collected data support the claim that there exists a difference in the average reaction times for the groups?

- Let μ_A and μ_B denote the average reaction times for Group A and Group B, respectively
- The null hypothesis is $H_0 : \mu_A - \mu_B = 0$
- The alternative hypothesis is $H_1 : \mu_A - \mu_B \neq 0$
- The Z-statistics is $Z = \dfrac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$

Example: Hypothesis Test for a Difference in Population Means

- The rejection region is $\left\{Z < -z_{\alpha/2} \text{ or } Z > z_{\alpha/2}\right\}$ where the critical value $z_{\alpha/2}$ is determined by the significance level $\alpha = 0.05$
- Using SALT we get that $z_{\alpha/2} = 1.96$ because $P(|Z| > 1.96) = \alpha$
- Although we don't have the population variances, we can use the sample variances as estimates
- The observed value of the test statistic is

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \frac{(3.6 - 3.8) - 0}{\sqrt{\frac{0.18}{50} + \frac{0.14}{50}}} = -2.5$$

- Thus, the observed Z-score is inside the rejection region, so we reject the null hypothesis and conclude that the data supports the claim that there exists a difference in the average reaction times for the groups

Example: Hypothesis Test for a Difference in Population Proportions

- The cause of racial wage gap
- Skew the Script

https://www.youtube.com/watch?v=Fbhf9L8SS4M

The *p*-value of a Test

- If in the running coin flip example we observed $X = 97$ then we would have also rejected the null hypothesis
- The value of $X = 97$ is obviously more extreme than $X = 62$ and loosely speaking more "significant"
- However, the procedure that we've already described makes no distinction between these two observed values of X
- That's where the *p*-value comes in
- The *p*-value is **the smallest value of α for which the null hypothesis can be rejected**
- In other words, the *p*-value of a test is **the probability of obtaining a value of the test statistic that is at least as extreme as the one we observed, given that the null hypothesis is true**

The p-value of a Test

- The smaller the p-value, the more compelling is the evidence that the null hypothesis should be rejected
- Reporting the p-value allows the reader of published research to evaluate the extent to which the observed data disagree with the null hypothesis
- Reporting p-value permits each reader to use their own choice for α in deciding whether the observed data should lead to rejection of the null hypothesis

The p-value of a Test

- If the observed Z-score of the test statistics is z_0, and we are performing a two-sided test, then the p-value is

$$p = P(Z < -|z_0| \text{ or } Z > |z_0|)$$

- If it is an upper-tail test then

$$p = P(Z > z_0)$$

- And if it is a lower-tail test then

$$p = P(Z < z_0)$$

Example: Hypothesis Test

Example. A class survey in a large class for first-year college students asked, "About how many hours do you study during a typical week?" The mean response of the 463 students was 13.7 hours. Previous studies have shown that study time follows a normal distribution with standard deviation of 7.4 hours in the population of all first-year students at this university. Does the current study support students' claim to study more than 13 hours per week on average?

- Let μ denote the average number of hours studied per week by first-year students at this university
- The null hypothesis is $H_0 : \mu = 13$
- The alternative hypothesis is $H_1 : \mu > 13$
- The Z-statistic is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Example: Hypothesis Test

- The rejection region is $\{Z > z_\alpha\}$ where the critical value z_α is determined by the significance level $\alpha = 0.05$
- Using SALT we get that $z_\alpha = 1.645$ because $P(Z > 1.645) = \alpha$
- The observed value of the test statistic is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{13.7 - 13}{7.4/\sqrt{463}} = 2.035$$

- Thus, the observed Z-score is inside the rejection region, so we reject the null hypothesis and conclude that the data supports the claim that students study more than 13 hours per week on average
- Because this is a two-sided test, and the observed value was $Z = 2.035$, the p-value is

$$p = P(Z < -2.035 \text{ or } Z > 2.035) = 0.042$$

- If we had chosen $\alpha = 0.01$ then $p > \alpha$ and we would not have rejected

Example: Difference of Two Proportions

Example. Advertising on a news website resulted in 412 clicks out of 5000 visitors and advertising on a real estate website resulted in 312 clicks out of 3000 visitors. Is advertising on a real estate website more effective? Use a significane level of 0.01

- Let p_1 denote the proportion of visitors to the news website who click on an ad and p_2 denote the proportion of visitors to the real estate website who click on an ad
- The null hypothesis is $H_0 : p_1 - p_2 = 0$
- The alternative hypothesis is $H_1 : p_1 - p_2 > 0$
- The Z-statistic is

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}_c(1 - \hat{p}_c)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where $\hat{p}_c = \frac{x_1 + x_2}{n_1 + n_2}$ is the **combined** proportion of clicks

Example: Difference of Two Proportions

- We are given the following data: $n_1 = 5000$, $x_1 = 412$, $n_2 = 3000$, $x_2 = 312$
- We compute that $\hat{p}_1 = x_1/n_1 = 0.0824$, $\hat{p}_2 = x_2/n_2 = 0.078$, and $\hat{p}_c = (x_1 + x_2)/(n_1 + n_2) = 0.0804$
- The z-value is

$$z = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}_c(1 - \hat{p}_c)(\frac{1}{n_1} + \frac{1}{n_2})}} = 0.7626221$$

- Using the standard normal distribution, the p-value is

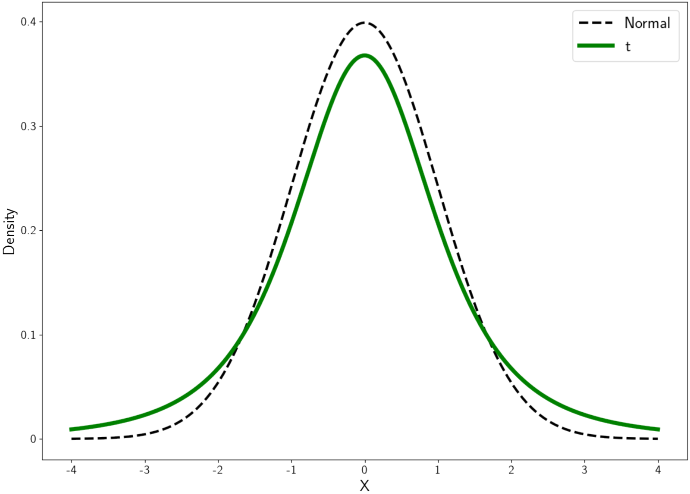
$$p = P(Z > 0.7626221) = 0.223$$

- Since $p > \alpha = 0.05$, we fail to reject the null hypothesis and conclude that there is not enough statistical evidence to support the claim that advertising on a real estate website is more effective

t-tests

- When running a hypothesis test for the mean, we used the sample standard deviation s to estimate the population standard deviation σ
- When the sample size is large, this is a good approximation but in practice we need to make some adjustment for the fact that we have some uncertainty about what the true population standard deviation actually is, especially for **small samples**
- The t-test is a modification of the Z-test that makes this adjustment
- The main difference is that instead of using the normal distribution to approximate the sampling distribution of the test statistic, we use the **Student's t-distribution**
- The t-distribution tends to arise in situations where you think that the data actually follow a normal distribution, but you don't know the mean or standard deviation
- The **t-distribution** is a continuous distribution that looks very similar to a normal distribution

t-distribution



t-distribution

- When running a hypothesis test for the mean using the t-test, the test statistic is
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

- In other words, we use the same form as the Z-test, but we use the sample standard deviation s instead of the population standard deviation σ

- The t-distribution has one parameter called the **degrees of freedom**, which is equal to $df = n - 1$

- The t-distribution is symmetric about zero, and the mean and median are both zero

- As n increases, the t-distribution approaches the normal distribution

- For large sample size n, the t-test behaves exactly the same way as a z-test

Example: One Sample t-test

Example. Historically, the mean grade in a course taught by Professor Lopez is 67.5. In the current version of the course consisting of 20 students, the average grade is 72.3 with sample standard deviation 9.52. Does the data support the claim that the current average in the course is statistically different than the historical average?

- Let μ be the average grade in the course
- The null hypothesis is $H_0 : \mu = 67.5$
- The alternative hypothesis is $H_1 : \mu \neq 67.5$
- The t-statistic is
$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$
where $s = 9.52$ and $n = 20$ and $\bar{X} = 72.3$
- The t-statistic is $t = 2.25$
- Using SALT we get a p-value of $p = 0.036$

Example: One Sample t-test

- Since $p < 0.05$, the observed t-score is inside the rejection region when $\alpha = 0.05$
- Thus, we reject the null hypothesis and conclude that the data supports the claim that the average grade in the course is statistically different than the historical average
- To run a one-sample t-test in `jamovi`, we can use the **T-Tests** menu
- `zeppo.csv` contains the data from the example above

Comparing Two Means: Independent Samples t-test

- We now consider the problem of comparing two means from independent samples with unknown population standard deviations
- The null hypothesis is $H_0 : \mu_1 - \mu_2 = 0$
- The t-statistic is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

  where s_p is the **pooled standard deviation**

- The pooled standard deviation is

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

  where s_1 and s_2 are the sample standard deviations of the two samples

- We use a t-distribution with $df = n_1 + n_2 - 2$ degrees of freedom

Example: Independent Samples t-test

Example. A course contains 33 students and 2 tutors. There are 15 students in Tutor's A tutorials and 18 students in Tutor's B tutorials. The research question that we are interested in is whether there is a difference in students' performance based on the tutor. Below is a table summarzing the data:

Tutor	mean	std dev	N
Tutor A	74.53	9.00	15
Tutor B	69.06	5.77	18

- Let μ_1 be the average grade in Tutor A's tutorials and μ_2 be the average grade in Tutor B's tutorials
- The null hypothesis is $H_0 : \mu_1 - \mu_2 = 0$
- The alternative hypothesis is $H_1 : \mu_1 - \mu_2 \neq 0$

Example: Independent Samples t-test

- The t-statistic is

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- We are given that $X_1 = 74.53$, $X_2 = 69.06$, $s_1 = 9.00$, $s_2 = 5.77$, $n_1 = 15$, and $n_2 = 18$

- The pooled standard deviation is

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 7.4052$$

- The t-statistic value is $t = 2.12$ and using SALT we get a p-value of $p = 0.043$

- Since $p < 0.05$, the observed t-score is inside the rejection region when $\alpha = 0.05$, suggesting that Tutor A's effectiveness is statistically different than Tutor B's

Example: Independent Samples t-test

- To run an independent samples t-test in `jamovi`, we can use the **T-Tests** menu
- The data for this example is in `harpo.csv`

Paired Samples t-test

- One of the main assumptions in the Independent Samples t-test is that the two samples are independent
- There are many research scenarios where this assumption clearly does not hold
- As an example, a professor may want to know if the students in their course had any improvement from one test to the next
- In this case, we would collect one sample of test 1 scores and another sample of test 2 scores
- The problem is that the same students are in both samples, so the samples are not independent
- In this case, we can use a **paired samples** t-test
- This test is just a one sample t-test where the sample mean is the difference between the two sample means
- And the data is the difference between the two samples

Paired Samples t-test

- Suppose that we have two samples of size n:

$$x_1, x_2, \ldots, x_n$$
$$y_1, y_2, \ldots, y_n$$

- The null hypothesis is $H_0 : \mu_X - \mu_Y = 0$
- And the alternative hypothesis is $H_1 : \mu_X - \mu_Y \neq 0$
- We create a new variable $Z = X - Y$ and run a one sample t-test on Z
- The t-statistic is

$$t = \frac{\bar{Z}}{\frac{s_Z}{\sqrt{n}}}$$

where s_Z is the sample standard deviation of Z

Example: Paired Samples t-test

Example. In Professor Chico's class, students take two major tests, one early in the semester and one later in the semester. Professor Chico runs a very hard class, one that most students find very challenging. But she argues that by setting hard assessments students are encouraged to work harder. Her theory is that the first test is a bit of a "wake up call" for students. When they realize how hard her class really is, they'll work harder for the second test and get a better grade. Is she right? The data is contained in the file `chico.csv`.

- Perform a one-sample t-test on the difference variable
- Or use `jamovi`'s `Paired Samples t-test` from the `T-Tests` menu

Example: Paired Samples t-test

Example. Data was collected of the number of work hours lost to accidents before and after a safety program was implemented in 10 factories. Use the 0.05 level of significance to test the claim that the safety program was effective.

Before	After
45	36
57	51
73	60
83	77
46	44
34	29
124	119
26	24
33	35
17	11

Example: Paired Samples t-test

- We need to use a paired samples t-test because the two samples are not independent
- The variable of interest is then the difference between the two samples
- Let μ be the mean difference between the two samples, specifically, $\mu = \mu_{\text{after}} - \mu_{\text{before}}$, where
- The null hypothesis is $H_0 : \mu = 0$
- The alternative hypothesis is $H_1 : \mu < 0$
- The t-statistic is

$$t = \frac{\bar{x} - \mu}{(s/\sqrt{n})}$$

where $\bar{x}$ is the sample mean of the difference variable, s is the sample standard deviation of the difference variable, and n is the sample size

- The mean difference is computed to be $\bar{x} = -5.20$ and the sample standard deviation is $s = 4.08$

Example: Paired Samples t-test

- The t-statistic is $t = -4.03$ and the p-value is $p = 0.001$
- The rejection region is $t < -1.8331$
- Since $t < -1.8331$ (or equivalently because $p < \alpha$), we reject H_0 and conclude that there is statistically significant evidence that the safety program was effective