Hypothesis Testing

553.211 Intro to Prob. & Stats.

Johns Hopkins University

EN.553.211

Hypothesis testing

- Often, when performing an experiment, we have to choose between two "hypotheses" or assumptions
  E.g. Is a vaccine effective or not? Is a coin that we are tossing fair or unfair?
- To answer such questions, we look at the data sample $x_1, \ldots, x_n$ and decide between the two hypotheses. More precisely, we verify which one of the two is more compatible with the data sample.

Hypothesis testing continued

The set-up for a statistical hypothesis test can be thought of as being the prosecutor in a criminal court case:

1. A claim is made about a parameter of a population ($\mu = 5$, for example). Analogous to the statement "*the defendant is not guilty*".

Hypothesis testing continued

The set-up for a statistical hypothesis test can be thought of as being the prosecutor in a criminal court case:

1. A claim is made about a parameter of a population ($\mu = 5$, for example). Analogous to the statement "*the defendant is not guilty*".
2. We would like to present evidence that the claim on the parameter is false. E.g. we the prosecution want to demonstrate the defendant is guilty.

Hypothesis testing continued

The set-up for a statistical hypothesis test can be thought of as being the prosecutor in a criminal court case:

1. A claim is made about a parameter of a population ($\mu = 5$, for example). Analogous to the statement "*the defendant is not guilty*".
2. We would like to present evidence that the claim on the parameter is false. E.g. we the prosecution want to demonstrate the defendant is guilty.
3. Importantly, we can NEVER know what the actual value of the parameter is; so we collect evidence to hopefully demonstrate that the claimed value of the parameter is "unreasonable".

Hypothesis testing continued

4 Similarly, the jury must find the defendant guilty *beyond all reasonable doubt*. What "reasonable doubt" means in a real life criminal case can often be tricky.

Hypothesis testing continued

4 Similarly, the jury must find the defendant guilty *beyond all reasonable doubt*. What "reasonable doubt" means in a real life criminal case can often be tricky.

5 But in certain experiments we can give a numerical measurement of this called the significance level.

Large sample tests on population mean

Let's start with a claim being made on a normal population mean μ with known standard deviation σ:

$$\mu = \mu_0.$$

Large sample tests on population mean

Let's start with a claim being made on a normal population mean μ with known standard deviation σ:

$$\mu = \mu_0.$$

We can collect an independent random sample from this population $X_1, X_2, \ldots, X_n$. Important: We consider these as random variables until the moment we observe them.

Large sample tests on population mean

Let's start with a claim being made on a normal population mean μ with known standard deviation σ:

$$\mu = \mu_0.$$

We can collect an independent random sample from this population $X_1, X_2, \ldots, X_n$. Important: We consider these as random variables until the moment we observe them.
We know from Ch. 6 that

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Large sample tests on population mean continued

Therefore, if the claim were correct, then $\mu = \mu_0$. Now, if we take a random sample average $\bar{X}$, what is the probability that $\bar{X} = \mu_0$?

Large sample tests on population mean continued

Therefore, if the claim were correct, then $\mu = \mu_0$. Now, if we take a random sample average $\bar{X}$, what is the probability that $\bar{X} = \mu_0$? It's zero! So, we could make the rule that we'll decide $\mu = \mu_0$ is false no matter what. But this should hopefully sound silly to you. The question becomes

Question

How far should an observed sample mean be from μ_0 before I find the claim $\mu = \mu_0$ to be unreasonable?
In the criminal case analogy: How likely is it that I find a not guilty person guilty?

Large sample tests on population mean continued

To answer this, let's suppose for a moment that $\mu = \mu_0$ were true. Then, after observing a sample average of $\bar{x}$, ask ourselves: "What's the probability of seeing a sample mean at least as far from μ_0 as what we did?"

Large sample tests on population mean continued

To answer this, let's suppose for a moment that $\mu = \mu_0$ were true. Then, after observing a sample average of $\bar{x}$, ask ourselves: "What's the probability of seeing a sample mean at least as far from μ_0 as what we did?"
This probability is given by

$$P\left(Z \geq \left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right|\right)$$

Large sample tests on population mean continued

To answer this, let's suppose for a moment that $\mu = \mu_0$ were true. Then, after observing a sample average of $\bar{x}$, ask ourselves: "What's the probability of seeing a sample mean at least as far from μ_0 as what we did?"
This probability is given by

$$P\left(Z \geq \left|\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right|\right)$$

Therefore, if we tolerate a probability of no more than $0 < \alpha < 1$ of making the mistake of rejecting the claim $\mu = \mu_0$, then the rule for this decision should be to reject if the above probability is less than α.

Large sample tests on population mean continued

It should be noted that sampling from a population with known standard deviation but unknown mean is pretty unrealistic. Luckily, we may take it for granted that for $n \geq 30$ we have

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \overset{\text{approx}}{\sim} \mathcal{N}(0, 1),$$

which we will exploit throughout this chapter (and learn what to do when we can't in the next).

Large sample tests on population mean continued

It should be noted that sampling from a population with known standard deviation but unknown mean is pretty unrealistic. Luckily, we may take it for granted that for $n \geq 30$ we have

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \overset{\text{approx}}{\sim} \mathcal{N}(0, 1),$$

which we will exploit throughout this chapter (and learn what to do when we can't in the next).
The next slide names the key terms; see if you can match the terms with the theoretical concepts we just went over.

Hypothesis testing; key terms

Definition (Hypothesis testing)

1. The **null hypothesis** H_0, that represents the "status quo"
2. The **alternative hypothesis** H_a, that represents the research question (what we are testing for)
3. The **test statistic**, which is a numerical value that we compute from the given sample (e.g $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$) and its corresponding **p-value** to be compared to the test's **significance level** α
4. The **conclusion** on which one of the two statement is more compatible with the data sample.
   - If p-value$\leq \alpha \Rightarrow$we reject H_0
   - If p-value$> \alpha \Rightarrow$we fail to reject H_0
5. The **rejection region** R_0 (for the null hypothesis H_0). The above test is equivalent to
   - If test statistic$\in R_0 \Rightarrow$we reject H_0
   - If test statistic$\notin R_0 \Rightarrow$we fail to reject H_0

Hypothesis testing

Remark

We have two possible conclusions:

1. *we reject H_0, and we have evidence to support H_a.*
2. *Or, we fail to reject H_0, and we have no evidence to support H_a.*

Hypothesis testing

Example

In 2001, Hopkins' students' average GPA was $\mu_0 = 3.37$. Suppose today's population mean is μ. I want to test whether there has been grade inflation in the past few years. To do so, I get a random sample from the current population. Suppose the size of our specific sample is $n = 50$ and that the sample yields the following statistics: the sample mean $\overline{x} = 3.47$ and a sample std. dev. $s = 0.31$. We use a significance level $\alpha = 0.05$ to test this hypothesis.

Null and alternative hypothesis

We first establish null and alternative hypothesis

- $H_0 : \mu = \mu_0$
- $H_a : \mu > \mu_0$ "there has been grade inflation in the past few years" (the research question)

where $\mu_0 = 3.37$ is called the **null value**
We want to establish which one of the two hypotheses is more compatible with the sample

Null and alternative hypothesis

We first establish null and alternative hypothesis

- $H_0 : \mu = \mu_0$
- $H_a : \mu > \mu_0$ "there has been grade inflation in the past few years" (the research question)

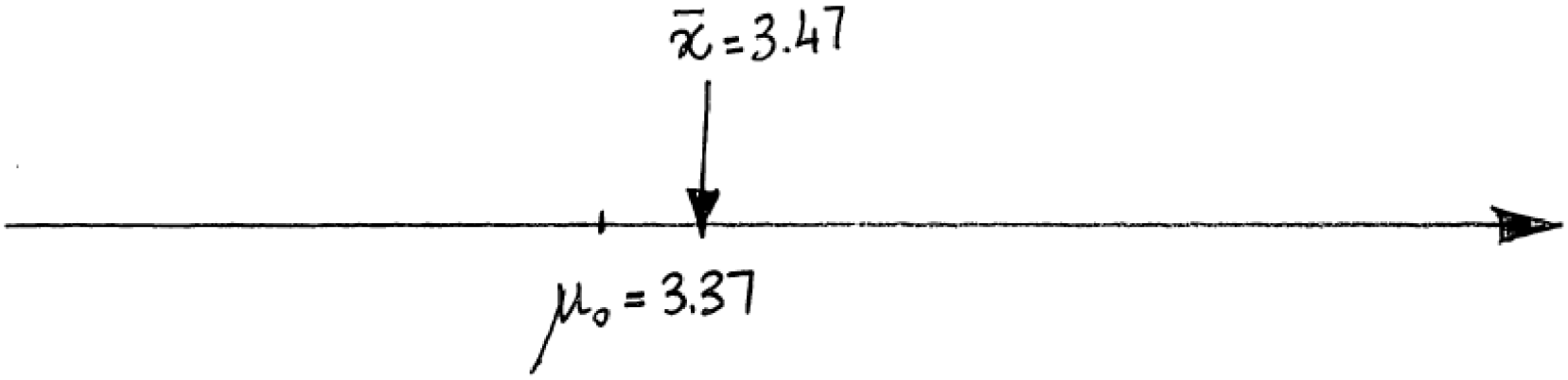
where $\mu_0 = 3.37$ is called the **null value**
We want to establish which one of the two hypotheses is more compatible with the sample

Remark

An alternative hypothesis of the type $\mu > \mu_0$ is called **one-sided**. *An alternative hypothesis of the type $\mu < \mu_0$ is also called one-sided. On the other hand, an alternative hypothesis $\mu \neq \mu_0$ (which is equivalent to $\mu > \mu_0$ or $\mu < \mu_0$) is called* **two-sided**.

Test statistic

- $\overline{x}$ is an estimate of μ. So, since $\overline{x} = 3.47 > 3.37 = \mu_0$, it seems to suggest that $H_a : \mu > \mu_0$.



- But is a difference of just $\overline{x} - \mu_0 = 3.47 - 3.37 = 0.1$ conclusive? Couldn't it be just caused by the randomness of the sampling process, and still be compatible with the null hypothesis $H_0 : \mu = \mu_0$?
- So the question that we are going to try to answer is:
  If the null hypothesis is true, how likely are we to get a sample with $\overline{x} = 3.47$?

Test statistic

- To answer this question, we have to look at the sampling distribution of the sample mean $\overline{X}$ as a RV. For this example, because $n = 50 \geq 30$ we know (we don't know σ here and have to use s instead of σ)

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim N(0, 1)$$

- Now, under the null hypothesis H_0 we have $\mu = \mu_0$. So if the null hypothesis is true, we have

$$\frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim N(0, 1)$$

  Thus $N(0, 1)$ is called the **null distribution of the test statistic** $\frac{\overline{X}-\mu_0}{s/\sqrt{n}}$**.**

- The test statistic $\frac{\overline{X}-\mu_0}{s/\sqrt{n}}$ is a RV because it depends on the sample, which is random itself.

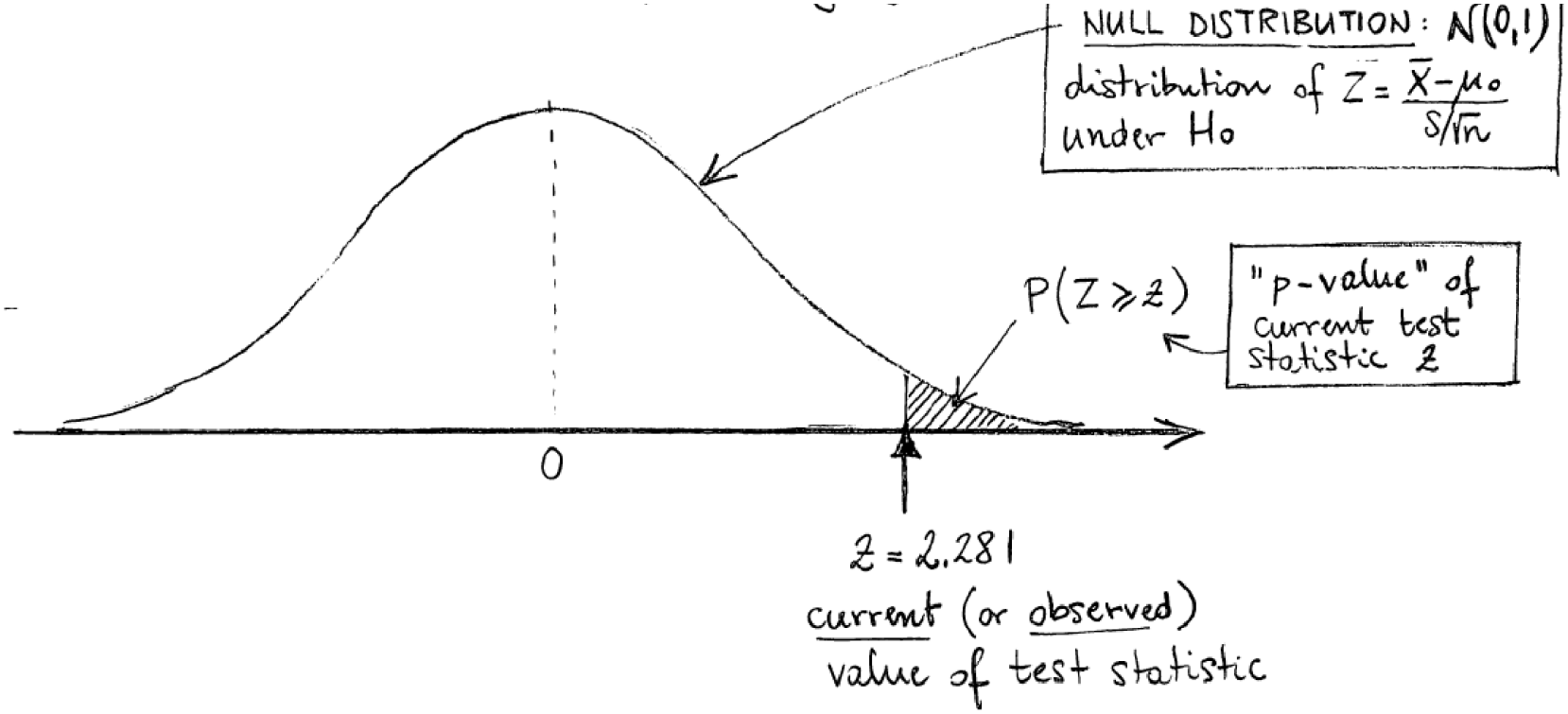
Test statistic

- For our current (or observed) sample, the specific value of the test statistic is given by the z-score of the sample mean $\overline{x}$ (which respect to the normal distribution when $\mu = \mu_0$). i.e.

$$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{3.47 - 3.37}{0.31/\sqrt{50}} \approx 2.281.$$

- This value seems a bit high, considering it is a sample from a standard normal RV (*).
- How can we quantify this? Let's compare the specific value z computed from out specific sample with the distribution (*), i.e. let's compare it, in a sense, with all the possible samples from our population.

p-value


NULL DISTRIBUTION: $N(0,1)$
distribution of $Z = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ under H_0
$P(Z \geq z)$
"p-value" of current test statistic z
0
$z = 2.281$
current (or observed) value of test statistic

p-value

- To quantify how "extreme" the observed test statistic z is (with respect to the null distribution and the direction of the alternative hypothesis) we compute

$$P(Z \geq z) = P(Z \geq 2.281) = 1 - \Phi(2.281) = 0.0113 = 1.13\%$$

- The number $P(Z \geq z) = 0.0113$, called the **p-value of the test statistic** $z = 2.281$ (we'll give a more general and precise definition of p-value later) is very small and gives a measure of how "extreme" the statistic $z = 2.281$ is with respect to the null distribution $N(0, 1)$.
- This says that, under the null hypothesis H_0, the current sample is very unlikely. In other words, the small p-value seems to suggest that the null hypothesis is statistical incompatible with the current (observed) data, and therefore we should reject the null hypothesis H_0.
- So: if p-value of the test statistic is small, then we reject H_0

p-value

Definition (p-value)

It is the probability, under the assumption that the null hypothesis H_0 is true, that a generic test statistic (a RV) is at least as favorable as to the alternative hypothesis H_a as the current (observed) statistic (a constant).

E.g. for the hypothesis testing about population mean μ

- For $H_a : \mu > \mu_0 \Rightarrow$p-value$= P(Z \geq z)$
- For $H_a : \mu < \mu_0 \Rightarrow$p-value$= P(Z \leq z)$
- For $H_a : \mu \neq \mu_0 \Rightarrow$ p-value $= P(Z \leq -|z|) + P(Z \geq |z|) = 2P(Z \geq |z|)$

Significance level

- Recall if the p-value is small enough, then we reject H_0.
- When do we decide to reject the null hypothesis? We need a threshold for p-value. This is given by the so-called **significance level** α, which is normally assigned by the problem (Typically, $\alpha = 0.05$). So
  - if p-value$\leq \alpha \Rightarrow$we reject H_0
  - if p-value$> \alpha \Rightarrow$we fail to reject H_0 (we retain H_0)
- For this example, the significance level $\alpha = 0.05$. We obtained p-value$=0.0113 < 0.05$, then we reject H_0 at significance level $\alpha = 0.05$, and we have strong evidence to conclude that there has been grade inflation in the past few years.

Significance level

The p-value quantifies the evidence for/against a hypothesis
Rough guideline:

- if p-value$\leq 0.01 \Rightarrow$ very strong evidence against H_0. H_0 is rejected. The conclusion for H_a is **highly significant**.
- if $0.01 <$p-value$\leq 0.05 \Rightarrow$ strong evidence against H_0. H_0 is rejected at $\alpha = 0.05$. The conclusion for H_a is **statistically significant**.
- if $0.05 <$p-value$\leq 0.1 \Rightarrow$ weak evidence against H_0. H_0 is usually not rejected. The conclusion for H_a is only **tending toward statistical significance**.
- if $0.1 <$p-value$\Rightarrow$ little or no evidence against H_0. H_0 is not rejected. The conclusion for H_a is **not statistically significant**.

We always set a significance level α in the question. Directly use it to decide whether the p-value is small enough

Significance level & critical values

Definition

A **Type I error** for a statistical test is the error of rejecting H_0 even though it's actually true. The **significance level**, denoted by α, is the probability

$$\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}).$$

The **critical value** is z_α for one-tailed test and $z_{\alpha/2}$ for two tailed test.

Rejection region

Definition (Rejection region)

Rejection region indicated by R_0, is the set of values of the current test statistic for which we reject the null hypothesis H_0.

E.g. for the hypothesis testing about population mean μ, when we know the test statistic is normal

- For $H_a : \mu > \mu_0 \Rightarrow R_0 = [z_\alpha, \infty)$
- For $H_a : \mu < \mu_0 \Rightarrow R_0 = (-\infty, -z_\alpha]$
- For $H_a : \mu \neq \mu_0 \Rightarrow R_0 = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$

Rejection region

- For this example, $\alpha = 0.05$ and $H_a : \mu > \mu_0$. Thus

$$R_0 = [z_\alpha, \infty) = [z_{0.05}, \infty) = [1.64, \infty)$$

- The test statistic $z = 2.281 \in [1.64, \infty)$. Then we reject H_0 at significance level $\alpha = 0.05$, and we have strong evidence to conclude that there has been grade inflation in the past few years.

Hypothesis testing steps

1. Establish null and alternative hypothesis

   E.g. for population mean μ

   Null Hypothesis
   - $H_0 : \mu = \mu_0$

   Choice of Alternative Hypothesis:
   - $H_a : \mu \neq \mu_0$ (two-sided)
   - $H_a : \mu < \mu_0$ (one-sided)
   - $H_a : \mu > \mu_0$ (one-sided)
2. Check the conditions and specify the null distribution of the test statistic
3. Compute the test statistic

   E.g. for population mean μ, when σ is unknown

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Hypothesis testing steps

4. Find p-value with respect to the direction of H_a
   E.g. for the hypothesis testing about population mean μ
   - For $H_a : \mu > \mu_0 \Rightarrow$p-value$= P(Z \geq z)$
   - For $H_a : \mu < \mu_0 \Rightarrow$p-value$= P(Z \leq z)$
   - For $H_a : \mu \neq \mu_0 \Rightarrow$ p-value $= P(Z \leq -|z|) + P(Z \geq |z|) = 2P(Z \geq |z|)$

5. Decide whether the p-value is small enough for the given α
   - if p-value$\leq \alpha \Rightarrow$we reject H_0
   - if p-value$> \alpha \Rightarrow$we fail to reject H_0

Hypothesis testing steps

Or

4. Find rejection region R_0 for the given α with respect to the direction of H_a
   E.g. for the hypothesis testing about population mean μ, when we know the test statistic is normal
   - For $H_a : \mu > \mu_0 \Rightarrow R_0 = [z_\alpha, \infty)$
   - For $H_a : \mu < \mu_0 \Rightarrow R_0 = (-\infty, -z_\alpha]$
   - For $H_a : \mu \neq \mu_0 \Rightarrow R_0 = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty)$
5. Decide whether the test statistic belongs to the rejection region
   - if the test statistic $\in R_0 \Rightarrow$we reject H_0
   - if the test statistic $\notin R_0 \Rightarrow$we fail to reject H_0

Hypothesis testing steps

6. Make conclusion

   We finally state our conclusion in the context of the problem. E.g.

   - We reject H_0 at significance level $\alpha = \ldots$, and we have evidence to conclude that ...
   - We fail to reject H_0 at significance level $\alpha = \ldots$, and we have no evidence to conclude that ...

Hypothesis testing for population mean

Hypothesis testing for μ

Null Hypothesis

- $H_0 : \mu = \mu_0$

Choice of Alternative Hypothesis:

- $H_a : \mu \neq \mu_0$ (two-sided)
- $H_a : \mu < \mu_0$ (one-sided)
- $H_a : \mu > \mu_0$ (one-sided)

Hypothesis testing for population mean

Hypothesis testing for μ

Suppose the population std. dev. σ is known.

- If the population distribution is normal
- Or if the sample size is large enough ($n \geq 30$), then regardless of the population distribution

Then the null distribution for the test statistic

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

We compute the test statistic for the given sample as

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Hypothesis testing for population mean

Hypothesis testing for μ

Suppose the population std. dev. σ is unknown.

- If the sample size is large enough ($n \geq 30$), then regardless of the population distribution

Then the null distribution for the test statistic

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim N(0, 1)$$

We compute the test statistic for the given sample as

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Hypothesis testing for population mean

Example (Carpentry)

You wish to show that the average hourly wage of carpenters in a particular city is different from \$19, the city average. You survey a sample of 100 carpenters and find the sample average of \$20 with a sample std. dev. of \$6. Use a significance level of $\alpha = 0.05$ to test your hypothesis.

Hypothesis testing for population mean

1. Establish null and alternative hypothesis
   - $H_0 : \mu = 19$
   - $H_a : \mu \neq 19$; the average carpenter wage is different than the city average.

The null value $\mu_0 = 19$

Hypothesis testing for population mean

2. Check the conditions and specify the null distribution of the test statistic
σ is unknown and $n = 100 \geq 30$. Thus

$$\frac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim N(0, 1)$$

3. Compute the test statistic

$$z = \frac{20 - 19}{6/\sqrt{100}} = 1.67$$

Hypothesis testing for population mean

4. Find p-value with respect to the direction of H_a

$$\begin{aligned}\text{p-value} &= P(Z < -|z|) + P(Z > |z|) = \Phi(-1.67) + (1 - \Phi(1.67)) \\ &= 2 \times (1 - \Phi(1.67)) \approx 2(1 - 0.9525) = 0.095\end{aligned}$$

5. Decide whether the p-value is small enough with respect to α

$$\text{p-value} \approx 0.095 > 0.05 = \alpha$$

Thus we fail to reject H_0

Hypothesis testing for population mean

Or

4. Find rejection region R_0 with respect to the direction of H_a

$$\begin{aligned} R_0 &= (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, \infty) = (-\infty, -z_{0.025}] \cup [z_{0.025}, \infty) \\ &= (-\infty, -1.96] \cup [1.96, \infty) \end{aligned}$$

5. Decide whether the test statistic belongs to the rejection region

$$z = 1.67 \notin R_0$$

Thus we fail to reject H_0

Hypothesis testing for population mean

6. Make conclusion

We fail to reject H_0 at significance level $\alpha = 0.05$, and we have no evidence to conclude that the average carpenter wage is different than the city average.

Power of a Stat. Test

- Recall the significance level of a hypothesis test, α, is the probability of type I error (reject H_0 even though it's true).
- There's another way we could mistakenly come to the wrong conclusion. We could retain H_0 even though it's false
- The power of a statistical test can be thought of as a measurement of the probability that this $\uparrow$ *doesn't* happen.

Power of a Stat. Test

Definition

A **Type II** error is to retain (fail to reject) H_0 even though it's false.
E.g. for the hypothesis testing about population mean μ
When H_0 is false, the true mean μ is something other than μ_0.
Thus, the probability of a Type II error depends on the significance level *and* what the true population parameter is:

$$\beta(\mu, \alpha) = P(\text{Retain } H_0 \text{ at sig. lv. } \alpha \mid \text{True mean is } \mu).$$

The **power** of a statistical test is

$$1 - \beta(\mu, \alpha) = P(\text{Reject } H_0 \text{ at sig. lv. } \alpha \mid \text{True mean is } \mu)$$

Which we will usually abbreviate as just $1 - \beta$.

Power of a Stat. Test

In the above example (carpentry), we retain H_0.
What is the power of the test if it were true that the average wage of carpenters actually was \$20/hr. not \$19/hr.?

Power of a Stat. Test

In the above example (carpentry), we retain H_0.
What is the power of the test if it were true that the average wage of carpenters actually was \$20/hr. not \$19/hr.?
In this case... (recall when n is large enough, $n \geq 30$, we have $\overline{X} \sim \mathcal{N}(\mu, s/\sqrt{n})$)

$$\begin{aligned}
\beta(20, 0.05) &= P(\text{Retain } H_0 \text{ sig. lv. } \alpha = 0.05 \mid \text{True mean } \mu = 20) \\
&= P\left(-1.96 \leq \underbrace{\frac{\overline{X} - 19}{6/\sqrt{100}}}_{\text{test stat. } Z} \leq 1.96 \mid \underbrace{\overline{X} \sim \mathcal{N}(20, 0.6)}_{\substack{\text{true distribution} \\ \text{of sample average } \overline{X}}} \right) \\
&= P\Big(17.824 \leq \overline{X} \leq 20.176 \,\Big|\, \overline{X} \sim \mathcal{N}(20, 0.6)\Big) \\
&\approx 0.6152.
\end{aligned}$$

Power of a Stat. Test

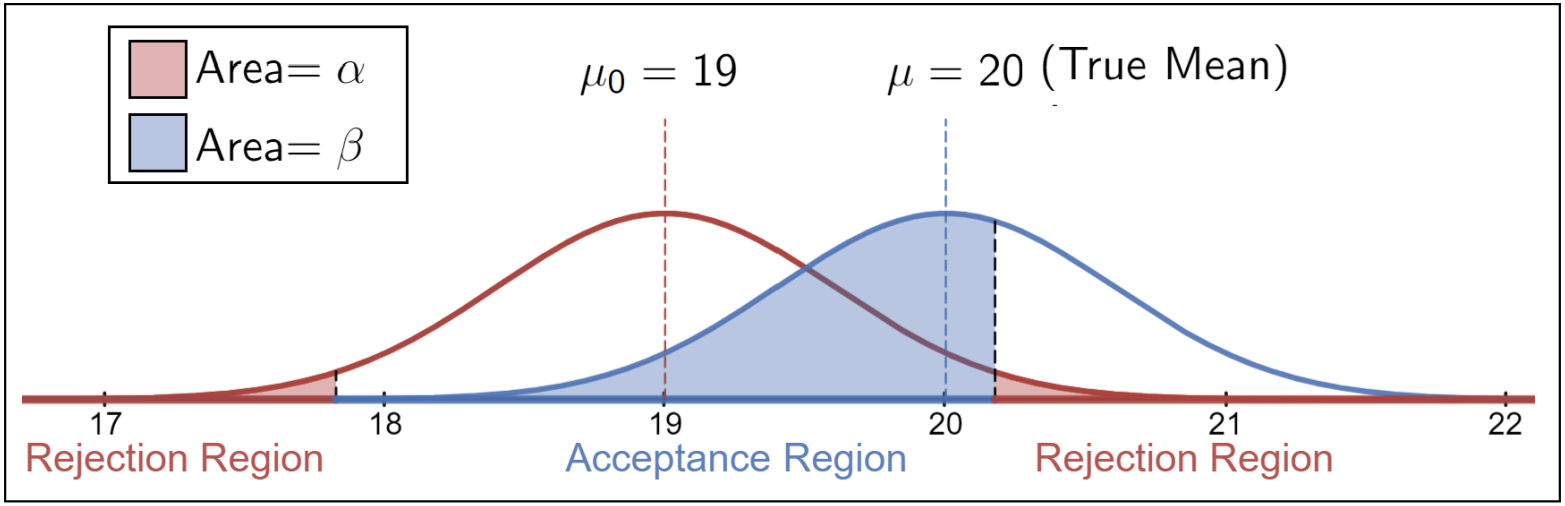
In the above example (carpentry), we retain H_0.
What is the power of the test if it were true that the average wage of carpenters actually was \$20/hr. not \$19/hr.?
In this case... (recall when n is large enough, $n \geq 30$, we have $\overline{X} \sim \mathcal{N}(\mu, s/\sqrt{n})$)

$$\begin{aligned}
\beta(20, 0.05) &= P(\text{Retain } H_0 \text{ sig. lv. } \alpha = 0.05 \mid \text{True mean } \mu = 20) \\
&= P\left(-1.96 \leq \underbrace{\frac{\overline{X} - 19}{6/\sqrt{100}}}_{\text{test stat. } Z} \leq 1.96 \mid \underbrace{\overline{X} \sim \mathcal{N}(20, 0.6)}_{\substack{\text{true distribution} \\ \text{of sample average } \overline{X}}} \right) \\
&= P\Big(17.824 \leq \overline{X} \leq 20.176 \,\Big|\, \overline{X} \sim \mathcal{N}(20, 0.6)\Big) \\
&\approx 0.6152.
\end{aligned}$$

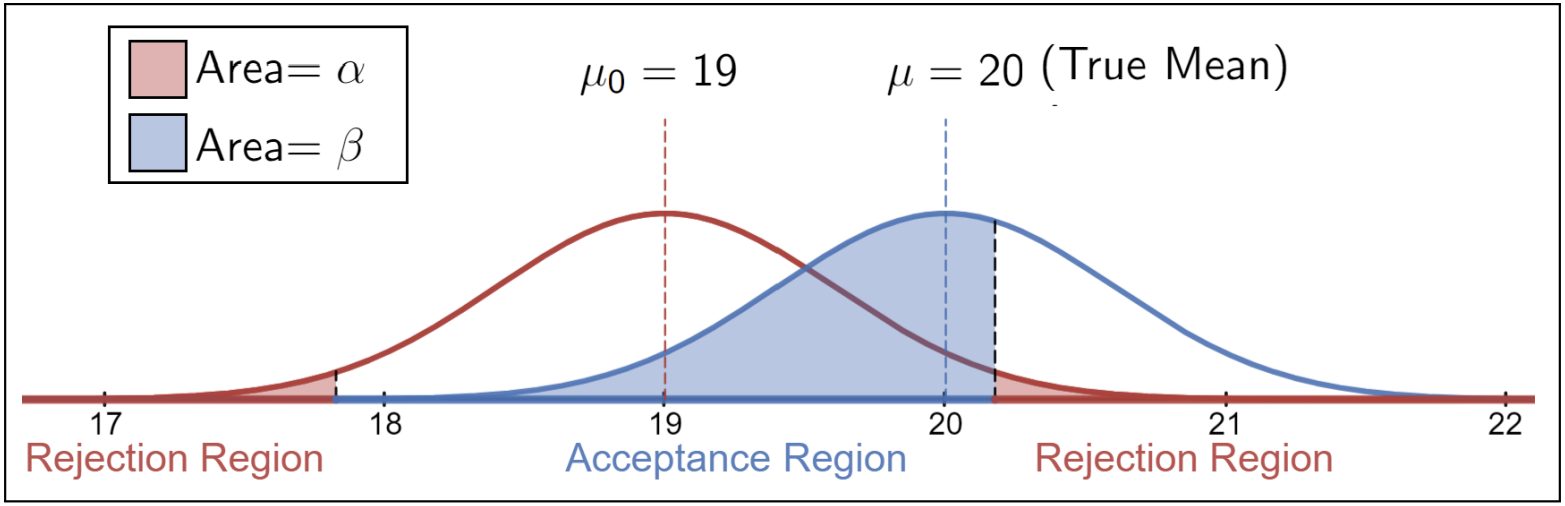
Thus, the power of the test is

$$1 - \beta = 1 - 0.6152 = 0.3848.$$

Power of a Stat. Test



Power of a Stat. Test



Notice that

- for fixed μ : Increasing α decreases β. I.e. the larger the rejection region, the *less* likely you are to mistakenly retain.
- for fixed α : The closer μ is to μ_0, the larger β becomes. I.e. When the true mean is only slightly different from μ_0, you are *more* likely to think they are the same.

Hypothesis testing for difference between population means

In many situations, the statistical question to be answered involves a comparison of two population means.

Hypothesis testing for $\mu_1 - \mu_2$

Null Hypothesis

- $H_0 : \mu_1 - \mu_2 = \Delta_0$

Choice of Alternative Hypothesis:

- $H_a : \mu_1 - \mu_2 \neq \Delta_0$ (two-sided)
- $H_a : \mu_1 - \mu_2 < \Delta_0$ (one-sided)
- $H_a : \mu_1 - \mu_2 > \Delta_0$ (one-sided)

Hypothesis testing for difference between population means

Hypothesis testing for $\mu_1 - \mu_2$

Suppose the population std. dev. σ_1, σ_2 are known.

- If the population distributions are normal
- Or if the sample sizes are large enough $n_1 \geq 30, n_2 \geq 30$, then regardless of the population distributions

Then the null distribution for the test statistic

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

We compute the test statistic for the given sample as

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Hypothesis testing for difference between population means

Hypothesis testing for $\mu_1 - \mu_2$

Suppose the population std. dev.s σ_1, σ_2 are unknown.

- If the sample sizes are large enough $n_1 \geq 30, n_2 \geq 30$, then regardless of the population distributions

Then the null distribution for the test statistic

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$$

We compute the test statistic for the given sample as

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Hypothesis testing for difference between population means

Example (GPA)

Independent random samples of 36 students and 45 students were draw from two large high schools A and B. The sample data summary is shown here:

	School A	**School** B
Sample Size	36	45
Avg. GPA	2.24	2.31
Sample Var.	0.0560	0.0540

Does the data present sufficient evidence to indicate that the mean for School A is less than the mean for School B? Use significance level $\alpha = 0.05$.

Hypothesis testing for difference between population means

1. Establish null and alternative hypothesis
   - $H_0 : \mu_1 - \mu_2 = 0$.
   - $H_a : \mu_1 - \mu_2 < 0$. The average GPA of School A is *lower* than that of School B.

The null value $\Delta_0 = 0$

Hypothesis testing for difference between population means

2. Check the conditions and specify the null distribution of the test statistic
$n_1 = 36 \geq 30, n_2 = 45 \geq 30$ thus

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$$

3. Compute the test statistic

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - \Delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(2.24 - 2.31) - 0}{\sqrt{\frac{0.056}{36} + \frac{0.054}{45}}} \approx -1.3335$$

Hypothesis testing for difference between population means

4. Find p-value with respect to the direction of H_a

$$\text{p-value} = P(Z < z) \approx \Phi(-1.33) \approx 0.0918.$$

5. Decide whether the p-value is small enough with respect to α

$$\text{p-value} \approx 0.0918 > 0.05 = \alpha$$

Thus we fail to reject H_0

Hypothesis testing for difference between population means

Or

4. Find rejection region R_0 with respect to the direction of H_a

$$R_0 = (-\infty, -z_\alpha] = (-\infty, -z_{0.05}] = (-\infty, -1.64]$$

5. Decide whether the test statistic belongs to the rejection region

$$z = -1.33 \notin R_0$$

Thus we fail to reject H_0

Hypothesis testing for difference between population means

6. Make conclusion

We fail to reject H_0 at significance level $\alpha = 0.05$, and we have insufficient evidence to conclude that the average GPA of School A is *lower* than that of School B.

Subtle differences

Suppose you want to test the influence that cellphone screen time has on academic performance. Consider two experiments below:

1. A classroom is selected and the grades of the students are recorded for the first 6 weeks with no limit to cellphone screen time. For the next 6 weeks, screen time is limited to 1 hr per day for each student, and again grades are recorded.
2. Two classrooms (of the same grade level) are randomly selected and the grades are recorded for 6 weeks. The students in classroom A have no restriction on cellphone screen time. The students in classroom B are limited to 1 hr of screen time per day.

What's different between these two experiments?

Subtle Differences; Continued

In the second experiment, the experimental and control group represent samples from two different populations. So we would have

$$\overline{x}_1 = \text{avg grade of class } A \quad \& \quad \overline{x}_2 = \text{avg grade of class } B.$$

Our statistic of interest would be $\overline{x}_1 - \overline{x}_2$, and our hypothesises would be

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{and} \quad H_a : \mu_1 - \mu_2 \neq 0.$$

Paired-difference

In the first experiment, the students are *their own* control group.

paired-difference

If there is a one-to-one relationship between the two samples. This is called a **paired-difference**.

For the paired-difference, we first compute the difference for each pair and then deal with it as only one sample from one population.

The observation for each student i is

$$x_i = (\text{1st 6 weeks grade}) - (\text{2nd 6 weeks grade})$$

and the sample avg used in the test stat. is $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
So we would do a single sample test with hypothesises

$$H_0 : \mu = 0 \quad \text{and} \quad H_a : \mu \neq 0.$$

Hypothesis testing for population proportion

Hypothesis testing for p

Null Hypothesis

- $H_0 : p = p_0$

Choice of Alternative Hypothesis:

- $H_a : p \neq p_0$ (two-sided)
- $H_a : p < p_0$ (one-sided)
- $H_a : p > p_0$ (one-sided)

Hypothesis testing for population proportion

Hypothesis testing for p

Suppose the population proportion p is unknown.

- If $np_0 > 5, nq_0 > 5$

Then the null distribution for the test statistic

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0, 1)$$

We compute the test statistic for the given sample as

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

Notice we use p_0 and q_0 because we assume null hypothesis is true.

Hypothesis testing for population proportion

Example (Spare Parts)

A computer chip manufacturer currently averages 0.3% of their products to be defective. However, a new processing method has been developed which claims to reduce the proportion of defective products, p.

Suppose a sample of 10000 computer chips produced under the new method yielded only 15 defective. Is this enough evidence to suggest that the new processing method reduces p? Suppose you want to be 99% sure in your conclusion, if you ultimately decide that the new method does make a difference.

Hypothesis testing for population proportion

1. Establish null and alternative hypothesis
   - $H_0 : p = 0.003$.
   - $H_a : p < 0.003$. The new method does reduce the proportion of defects

The null value $p_0 = 0.003$

Hypothesis testing for population proportion

2. Check the conditions and specify the null distribution of the test statistic
$np_0 = 10000 \times 0.003 = 30 > 5, nq_0 = 10000 \times (1 - 0.003) = 9970 > 5$ thus

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0, 1)$$

3. Compute the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.0015 - 0.003}{\sqrt{\frac{0.003(1-0.003)}{10000}}} = -2.73$$

Hypothesis testing for population proportion

4. Find p-value with respect to the direction of H_a

$$\text{p-value} = P(Z < z) = \Phi(-2.73) \approx 0.0032$$

5. Decide whether the p-value is small enough with respect to α

$$\text{p-value} \approx 0.0032 \leq 0.01 = \alpha$$

Thus we reject H_0

Hypothesis testing for population proportion

Or

4. Find rejection region R_0 with respect to the direction of H_a

$$R_0 = (-\infty, -z_\alpha] = (-\infty, -z_{0.01}] = (-\infty, -2.33]$$

5. Decide whether the test statistic belongs to the rejection region

$$z = -2.73 \in R_0$$

Thus we reject H_0

Hypothesis testing for population proportion

6. Make conclusion

We reject H_0 at significance level $\alpha = 0.01$, and we have evidence to conclude that the new method does reduce the proportion of defects.

Hypothesis testing for population proportion

Example

A large survey concluded that 20% of all US adults participate in fitness activities at least one a week. However, you suspect that the proportion is lower for adults over 40. To test this, you conduct a survey of $n = 2500$ US adults over 40 and found that 375 of them indicated that they participate in fitness activities at least once a week. Use $\alpha = 0.05$ to test your hypothesis, and find the power of the test if the actual proportion of adults over 40 who exercise at least once a week is $p = 0.179$.

Hypothesis testing for population proportion

1. Establish null and alternative hypothesis
   - $H_0 : p = 0.2$.
   - $H_a : p < 0.2$. The proportion of adults over 40 who exercise weekly is less than all adults.

The null value $p_0 = 0.2$

Hypothesis testing for population proportion

2. Check the conditions and specify the null distribution of the test statistic

$np_0 = 2500 \times 0.2 = 500 > 5, nq_0 = 2500 \times (1 - 0.2) = 2000 > 5$

thus

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \sim N(0, 1)$$

3. Compute the test statistic

$\hat{p} = 375/2500 = 0.15$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.15 - 0.2}{\sqrt{\frac{0.2(1-0.2)}{2500}}} = -6.25$$

Hypothesis testing for population proportion

4. Find p-value with respect to the direction of H_a

$$\text{p-value} = P(Z < z) = \Phi(-6.25) \approx 0$$

5. Decide whether the p-value is small enough with respect to α

$$\text{p-value} \approx 0 \leq 0.05 = \alpha$$

Thus we reject H_0

Hypothesis testing for population proportion

Or

4. Find rejection region R_0 with respect to the direction of H_a

$$R_0 = (-\infty, -z_\alpha] = (-\infty, -z_{0.05}] = (-\infty, -1.64]$$

5. Decide whether the test statistic belongs to the rejection region

$$z = -6.25 \in R_0$$

Thus we reject H_0

Hypothesis testing for population proportion

6. Make conclusion

We reject H_0 at significance level $\alpha = 0.05$, and we have evidence to conclude that the proportion of adults over 40 who exercise weekly is less than all adults.

Power Curves

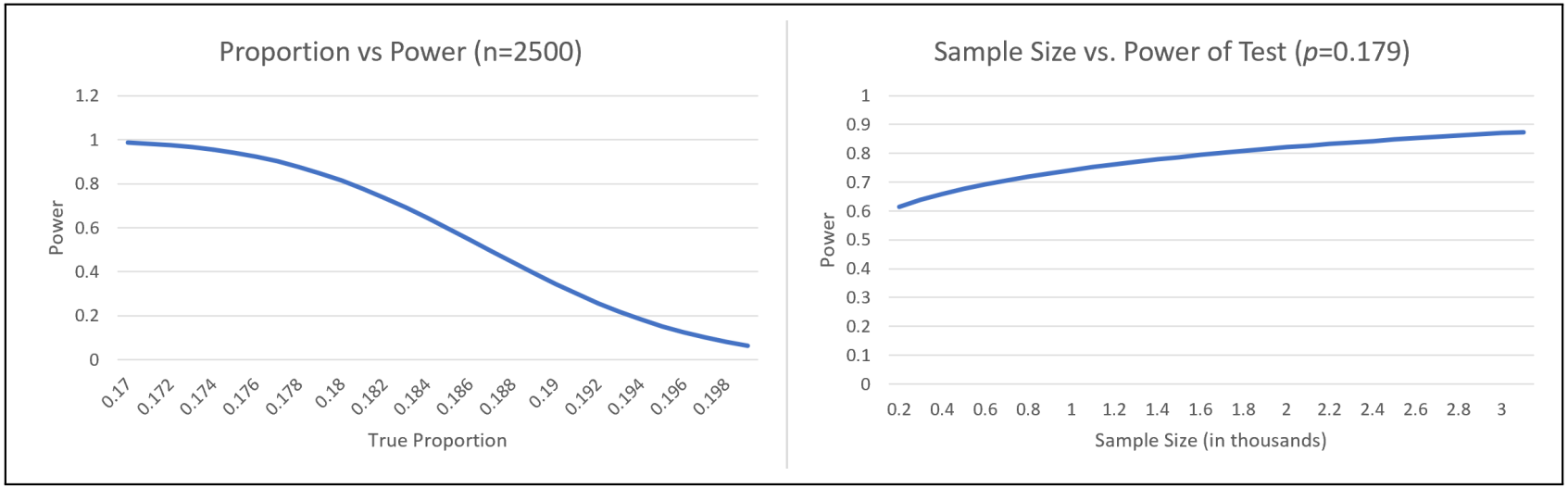
If the actual proportion is $p = 0.179$, then the sample proportion $\hat{p}$ approximately follows a $\mathcal{N}(p, \sqrt{p(1-p)/n}) = \mathcal{N}(0.179, \sqrt{(0.179)(0.821)/2500}) = \mathcal{N}(0.179, 0.0077)$ distribution
The condition for *not* rejecting H_0 is

$$z = \frac{\hat{p} - 0.2}{\sqrt{\frac{0.2(1-0.2)}{2500}}} \geq -1.64 \iff \hat{p} \geq 0.18688.$$

Thus, the power of the test is

$$\begin{aligned} 1 - \beta(p = 0.179, \alpha = 0.05) &= 1 - P(\hat{p} \geq 0.18688 \,\big|\, p = 0.179) \\ &\approx 1 - \left(1 - \Phi\left(\frac{0.18688 - 0.179}{\sqrt{\frac{0.179(1-0.179)}{2500}}}\right)\right) \\ &= \Phi(1.03) \\ &= 0.8485 \end{aligned}$$

Power Curves



- Left: A graph of $1 - \beta$ for fixed $\alpha = 0.05$ and sample size $n = 2500$. The power shrinks the closer the true proportion moves towards 0.20.
- Right: A graph of $1 - \beta$ for fixed $\alpha = 0.05$ and true proportion $p = 0.179$. The larger the sample size, the more power the test has.

Power of a Stat. Test

Remark

- *A good test is one with a high power* $1 - \beta$*, and a low significance level* α*.*
- *For fixed* α *and parameter (*μ *for mean, or p for proportion), the only way to increase power is by decreasing the standard error SE.*
- *The only way to decrease the SE is to increase the sample size. (So if you want more power, you have to pay for it!)*

Hypothesis testing for difference of population proportions

Hypothesis testing for $p_1 - p_2$

Null Hypothesis

- $H_0 : p_1 - p_2 = 0$

Choice of Alternative Hypothesis:

- $H_a : p_1 - p_2 \neq 0$ (two-sided)
- $H_a : p_1 - p_2 < 0$ (one-sided)
- $H_a : p_1 - p_2 > 0$ (one-sided)

Note: For the difference of population proportions p_1, p_2, we always study whether $p_1 = p_2, p_1 < p_2, p_1 > p_2$ or not, so the null value is always 0

Hypothesis testing for difference of population proportions

Hypothesis testing for $p_1 - p_2$

Suppose the population proportions p_1, p_2 are unknown.

- If $n_1\hat{p}_1 > 5, n_1\hat{q}_1 > 5$ and $n_2\hat{p}_2 > 5, n_2\hat{q}_2 > 5$

Then the null distribution for the test statistic

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

We compute the test statistic for the given sample as

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

Here $\hat{p}$ is the **pooled sample proportion**. Given null hypothesis is true, we have $p_1 = p_2 = p$. Then we estimate p as $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

Hypothesis testing for difference of population proportions

Example (Heart Disease)

The records of a hospital shows that 52 men in a sample of 1000 men versus 23 women in a sample of 1000 women were admitted because of heart disease. Does this data present sufficient evidence to indicate a higher rate of heart disease among men admitted to the hospital? Use $\alpha = 0.05$.

Hypothesis testing for difference of population proportions

Let p_1 and p_2 be the proportions of patients admitted for heart disease in men and women, respectively.

1. Establish null and alternative hypothesis
   - $H_0 : p_1 - p_2 = 0$.
   - $H_a : p_1 - p_2 > 0$. The proportion of heart disease is greater among men than women.

(The null value $\Delta_0 = 0$)

Hypothesis testing for difference of population proportions

2. Check the conditions and specify the null distribution of the test statistic

$n_1\hat{p}_1 = 52 > 5, n_1\hat{q}_1 = 948 > 5$ and
$n_2\hat{p}_2 = 23 > 5, n_2\hat{q}_2 = 977 > 5$ thus

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

3. Compute the test statistic

The **pooled sample proportion:**

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{52 + 23}{1000 + 1000} = 0.0375.$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1} + \frac{1}{n_2})}} = \frac{0.052 - 0.023}{\sqrt{0.0375(1 - 0.0375)(\frac{1}{1000} + \frac{1}{1000})}} = 3.4118$$

Hypothesis testing for difference of population proportions

4. Find p-value with respect to the direction of H_a

$$\text{p-value} = P(Z > z) = 1 - \Phi(3.4118) \approx 0.0003$$

5. Decide whether the p-value is small enough with respect to α

$$\text{p-value} \approx 0.0003 \leq 0.05 = \alpha$$

Thus we reject H_0

Hypothesis testing for difference of population proportions

Or

4. Find rejection region R_0 with respect to the direction of H_a

$$R_0 = [z_\alpha, \infty) = [z_{0.05}, \infty) = [1.64, \infty)$$

5. Decide whether the test statistic belongs to the rejection region

$$z = 3.4118 \in R_0$$

Thus we reject H_0

Hypothesis testing for difference of population proportions

6. Make conclusion

We reject H_0 at significance level $\alpha = 0.05$, and we have evidence to conclude that the proportion of heart disease is greater among men than women.