CSCI 3022-002 Intro to Data Science
Hypotheses and P-Values

H_0: status quo, present/current port

H_a: new port better

Small!

errors: T1: reject H_0 but it was true.
Result: use new port, wasn't better

Example:

Suppose a company is considering putting a new type of coating on bearings that it produces. Let denote the true mean life for the new coating. The company would not want to make any (costly) changes unless evidence **strongly** suggested that exceeds 1000 hours.

μ

new value

present!
borderline

T2: did not reject H_0, but we should have
Result: used old port, new was better

Opening sol: Test for Population Mean (known variance)

Example: Suppose a company is considering putting a new type of coating on bearings that it produces. Let denote the true mean <u>life</u> for the new coating. The company would not want to make any (costly) changes unless evidence **strongly** suggested that exceeds 1000 hours.

State the null and alternative hypotheses:

H_0: new isn't so good

H_A: new process better

option $\mu \le 1000$

$\mu = 1000$

$\mu > 1000$

Reality

	H_0	H_a
H_0	$1-\alpha$	Error Prob $\| \| \beta$
Decision H_a	Error Prob α	$1-\beta$

Now, suppose the company tests $n = 25$ randomly selected bearings with the new coating, and finds that the sample mean for life of the new coating is 1090 hours. Suppose we know that the population standard deviation is 130 hours. Conduct a test for the mean.

$\bar{X} = 1090$ $\sigma = 130$

Opening sol: Test for Population Mean (known variance)

Example: Suppose a company is considering putting a new type of coating on bearings that it produces. Let denote the true mean life for the new coating. The company would not want to make any (costly) changes unless evidence **strongly** suggested that exceeds 1000 hours.

State the null and alternative hypotheses:

Null hypothesis: $H_0 : \mu = \mu_0$

Now, suppose the company tests $n = 25$ randomly selected bearings with the new coating, and finds that the sample mean for life of the new coating is 1090 hours. Suppose we know that the population standard deviation is 130 hours. Conduct a test for the mean.

Opening sol: Test for Population Mean (known variance)

Example: Suppose a company is considering putting a new type of coating on bearings that it produces. Let denote the true mean life for the new coating. The company would not want to make any (costly) changes unless evidence **strongly** suggested that exceeds 1000 hours.

State the null and alternative hypotheses:

Null hypothesis: $H_0 : \mu = \mu_0$
Alt hypothesis: $H_a : \mu > \mu_0$

Now, suppose the company tests $n = 25$ randomly selected bearings with the new coating, and finds that the sample mean for life of the new coating is 1090 hours. Suppose we know that the population standard deviation is 130 hours. Conduct a test for the mean.

Opening sol: Test for Population Mean (known variance)

Example: Suppose a company is considering putting a new type of coating on bearings that it produces. Let denote the true mean life for the new coating. The company would not want to make any (costly) changes unless evidence **strongly** suggested that exceeds 1000 hours.

99% ?

State the null and alternative hypotheses:

Null hypothesis: $H_0 : \mu = \mu_0$
Alt hypothesis: $H_a : \mu > \mu_0$

Now, suppose the company tests $n = 25$ randomly selected bearings with the new coating, and finds that the sample mean for life of the new coating is 1090 hours. Suppose we know that the population standard deviation is 130 hours. Conduct a test for the mean. Compare

$\overset{\bar{X} - \mu_0}{}$

$Z = \frac{1090 - 1000}{130/\sqrt{25}} = \boxed{3.45}$ to $z_{.01} = 2.32$ and **Reject**

$\underset{\sigma/\sqrt{n}}{}$

Announcements and Reminders

- Short HW for next week.
- NB day Friday.

Where we at?

A summary of our process:

1. State hypothesis: H_0 : the baseline or "nothing is interesting result." For the coin: a fair coin, with $p = .5$.
   H_a : what we *want* to test or demonstrate. For the coin: an unfair coin, with $p \neq .5$

2. Collect some data

3. Compute a *test statistic* from our data. Maybe a sample proportion of heads $\hat{p}$?

4. Decide whether the *test statistic* $\hat{p}$ is **too far** from it's assumed baseline value in H_0, and make a decision accordingly. E.g. was $\hat{p}$ *far enough* from $p = .5$ to actually assert that they're different?

5. α is the value that describes the probability of rejecting a null hypotheses *given* that the hypothesis was true.

Rejection Regions or Probabilities?

How would we know when the test statistic is "sufficiently rare" under the null hypothesis such that we might regard the null as false? We could define a rejection region: a range of values that leads a researcher to reject the null hypothesis.

Today we discuss the second option. We can *either:*

1. Define a range of x-values - in the *units* of the data - that correspond to z-values (on the standard normal) that represent "extremely far" from the hypothesized mean. Reject if they're far enough, where far enough is beyond the z_{crit} value that depends on α

2. Compute a *probability*: if the null hypothesis is true, exactly how "extreme" is our data, as a probability? If it's in the α proportion of most extreme or outlying outcomes when the null is true, maybe we should conclude the null *wasn't* true.

Test for Population Proportion

Example (continued):
Null hypothesis:

Test statistic value:

Alternative Hypothesis   Rejection Region for α level test:

Test for Population Proportion

Example (continued):
Null hypothesis: $H_0 : p = p_0$

Test statistic value:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

<u>Alternative Hypothesis</u>   <u>Rejection Region</u> for α level test:

Test for Population Proportion

Example (continued):
Null hypothesis: $H_0 : p = p_0$

Test statistic value:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Alternative Hypothesis   Rejection Region for α level test:

$H_a : p > p_0$
$H_a : p < p_0$
$H_a : p \neq p_0$

Test for Population Proportion

Example (continued):
Null hypothesis: $H_0 : p = p_0$

Test statistic value:

$$Z = \frac{(\hat{p} - p_0)}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

diff b/t Sample prop & (true H_0) prop.

→ Var of the Mean of n Bernoullis.

Alternative Hypothesis	Rejection Region for α level test:
$H_a : p > p_0$	$z > z_{\alpha}$
$H_a : p < p_0$	$z < -z_{\alpha}$
$H_a : p \neq p_0$	$z < -z_{\alpha/2}$ **or** $z > z_{\alpha/2}$

Test for Population Mean (known variance)

Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

<u>Alternative Hypothesis</u>   <u>Rejection Region</u> for α level test:

Test for Population Mean (known variance)

Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Alternative Hypothesis   Rejection Region for α level test:

Test for Population Mean (known variance)

Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Alternative Hypothesis   Rejection Region for α level test:

$H_a : \mu > \mu_0$

$H_a : \mu < \mu_0$

$H_a : \mu \neq \mu_0$

Test for Population Mean (known variance)

Continuos

Null hypothesis: $H_0 : \mu = \mu_0$

data $- H_0$

Test statistic value:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$: Scaled by $\frac{\sigma}{\sqrt{n}}$: $\sqrt{V_{ar}[\bar{X}]}$

Alternative Hypothesis	Rejection Region for α level test:
$H_a : \mu > \mu_0$	$z > z_\alpha$
$H_a : \mu < \mu_0$	$z < -z_\alpha$
$H_a : \mu \neq \mu_0$	$z < -z_{\alpha/2}$ **or** $z > z_{\alpha/2}$

Errors in Hypothesis Testing

Definitions: *Type One and Type Two Errors*

A *type I error* is when the null hypothesis is rejected, but it is true. (False Positive).

A *type II error* is not rejecting H_0 when H_0 is false. (False Negative)

Errors in Hypothesis Testing

Definitions: *Type One and Type Two Errors*

A *type I error* is when the null hypothesis is rejected, but it is true. (False Positive).

We commit type I errors at a rate of α.

A *type II error* is not rejecting H_0 when H_0 is false. (False Negative)

We commit type II errors at a rate of β.

harder to compute... $H_a: \mu > 50$
$P(\text{reject})$ depends on how much bigger than 50 μ is.

Errors in Hypothesis Testing

Suppose we are evaluating a medical procedure. It has the following outcomes:

1. Procedure works: subject is cured!

2. Procedure doesn't work: subject dies.

Do we want our test - which compared this procedure as H_a to a null hypothesis of "don't get the procedure" - to focus on decreasing α or β?

Errors in Hypothesis Testing

Suppose we are evaluating a medical procedure. It has the following outcomes:

1. Procedure works: subject is cured!

2. Procedure doesn't work: subject dies.

Do we want our test - which compared this procedure as H_a to a null hypothesis of "don't get the procedure" - to focus on decreasing α or β?

$\boxed{\alpha}$: probability of endorsing the test when it doesn't actually work. Result: subjects die.

$\boxed{\beta}$: probability of *not* endorsing the test when it really works! Result: subjects don't get treated.

Errors in Hypothesis Testing

Suppose we are evaluating a medical procedure. It has the following outcomes:

1. Procedure works: subject is cured!

2. Procedure doesn't work: subject dies.

Do we want our test - which compared this procedure as H_a to a null hypothesis of "don't get the procedure" - to focus on decreasing α or β?

$\boxed{\alpha}$: probability of endorsing the test when it doesn't actually work. Result: subjects die.

Where is your risk tolerance?

Large Sample HT Example

Example An inventor has developed a new, energy-efficient lawn mower engine. He claims that the engine will run continuously for more than 5 hours (300 minutes) on a single gallon of regular gasoline. (The leading brand lawnmower engine runs for 300 minutes on 1 gallon of gasoline.)

$$H_0: \mu = 300$$
$$H_a: \mu > 300$$

$$n = 50$$

From his stock of engines, the inventor selects a simple random sample of 50 engines for testing. The engines run for an average of 305 minutes. The *sample* standard deviation s is found at 30 minutes.

$$s = 30$$

$$\bar{x} = 305$$

Test the hypothesis that the mean run time is more than 300 minutes. Use a 0.01 level of significance.

Large Sample Testing

When the sample size is large, the z tests are easily modified to yield valid test procedures without requiring either a normal population distribution or known standard deviation.

Earlier, we used the key result to justify large-sample confidence intervals:
A large $n\,(>30)$ implies that the standardized variable

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

has approximately a standard normal distribution.

Large Sample Testing

When the sample size is large, the z tests are easily modified to yield valid test procedures without requiring either a normal population distribution or known standard deviation.

Earlier, we used the key result to justify large-sample confidence intervals:
A large $n \, (> 30)$ implies that the standardized variable

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

has approximately a standard normal distribution.

Test for Population Mean (Large Sample)

Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

<u>Alternative Hypothesis</u>   <u>Rejection Region</u> for α level test:

Test for Population Mean (Large Sample)

Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

$$Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Alternative Hypothesis   Rejection Region for α level test:

Test for Population Mean (Large Sample)

Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

$$Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Alternative Hypothesis   Rejection Region for α level test:

$H_a : \mu > \mu_0$

$H_a : \mu < \mu_0$

$H_a : \mu \neq \mu_0$

Test for Population Mean (Large Sample)

Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

$$Z_{stat} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

(handwritten annotations: "Data", "Hyp true value")

Alternative Hypothesis	Rejection Region for α level test:
$H_a : \mu > \mu_0$	1) $z > z_\alpha$
$H_a : \mu < \mu_0$	$z < -z_\alpha$
$H_a : \mu \neq \mu_0$	$z < -z_{\alpha/2}$ **or** $z > z_{\alpha/2}$

(handwritten:)

or $P(Z < z_{stat}) < \alpha$

OR reject if

$P(|Z| > z_{stat}) < \frac{\alpha}{2}$

1) $P(Z > z_{stat}) < \alpha$

Example, cont'd: An inventor has developed a new, energy-efficient lawn mower engine. He claims that the engine will run continuously for more than 5 hours (300 minutes) on a single gallon of regular gasoline. (The leading brand lawnmower engine runs for 300 minutes on 1 gallon of gasoline.)
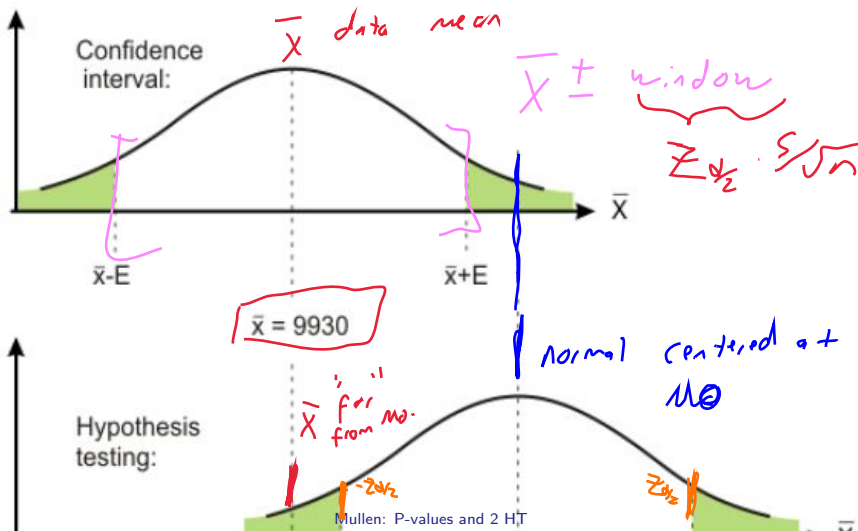
$n = 50; \bar{X} = 305, s = 30$. Is $\mu > 300$? at $\alpha = .01$?

Test the hypothesis that the mean run time is more than 300 minutes. Use a 0.01 level of significance.

Errors

Example, cont'd: An inventor has developed a new, energy-efficient lawn mower engine. He claims that the engine will run continuously for more than 5 hours (300 minutes) on a single gallon of regular gasoline. (The leading brand lawnmower engine runs for 300 minutes on 1 gallon of gasoline.)

$n = 50$; $\bar{X} = 305, s = 30$. Is $\mu > 300$? at $\alpha = .01$?

Test the hypothesis that the mean run time is more than 300 minutes. Use a 0.01 level of significance.

Compare $Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{305 - 300}{30/\sqrt{50}} = 1.17$ to $z_{.01} = 2.32$ and **Fail to Reject**.

Rejection Regions and CIs

Rejection regions have a lot in common with confidence intervals.

Confidence interval:

$\bar{x}$ data   mean

$\bar{x} \pm$ window

$Z_{\alpha/2} \cdot s/\sqrt{n}$

$\bar{x}-E$ $\bar{x}+E$

$\bar{x} = 9930$

Hypothesis testing:

$\bar{x}$ "far from no."

normal centered at μ_0

$-Z_{\alpha/2}$ $Z_{\alpha/2}$

p-values

The p-value measures the "extremeness" of the test statistic.

Definition: *p-value:*

A *p-value* is the probability, under the null hypothesis, that we would get a test statistic at least as "extreme" as the one we calculated.

given H_0

$\hookrightarrow H_A:$ $>$ $<$

both: $|z_{stat}|$ far from Z.

Idea: So, the smaller the p-value, the more evidence there is in the sample data against the null hypothesis (so the story goes...).

$\hookrightarrow$ *foul!*

So what constitutes "sufficiently small" and "extreme enough" to make a decision about the null hypothesis?

p-values: a Hypothesis Testing decision

1. Select a significance level (as before, the desired type I error probability), which would define the rejection region.

$$\alpha = 05, \quad .0)$$

p-values: a Hypothesis Testing decision

1. Select a significance level (as before, the desired type I error probability), which would define the (rejection region)

2. Instead, calculate your test statistic (e.g. a z_{stat} or Z-score),

p-values: a Hypothesis Testing decision

1. Select a significance level (as before, the desired type I error probability), which would define the rejection region.

2. Instead, calculate your test statistic (e.g. a z_{stat} or Z-score),

3. ...and from it calculate the associated probability e.g. $p = P(Z > z_{stat})$ - or the less than, two sided variant - and use that.

a normal Z under H_0

vs. data value

p-values: a Hypothesis Testing decision

1. Select a significance level (as before, the desired type I error probability), which would define the rejection region.

2. Instead, calculate your test statistic (e.g. a z_{stat} or Z-score),

3. ...and from it calculate the associated probability e.g. $p = P(Z > z_{stat})$ - or the less than, two sided variant - and use that.

4. decision rule is:

$$\textbf{Reject if:} \quad p \leq \alpha$$

p-values: a Hypothesis Testing decision

1. Select a significance level (as before, the desired type I error probability), which would define the rejection region.

2. Instead, calculate your test statistic (e.g. a z_{stat} or Z-score),

4. decision rule is:

Reject if: $p \leq \alpha$

Thus if the p-value exceeds the chosen significance level, the null hypothesis cannot be rejected at that level.

Note, the p-value can be thought of as the smallest significance level at which H_0 can be rejected.

$P = \frac{1}{1024}$ reject $\cancel{\theta\theta}$ / all $\alpha > \frac{1}{1024}$

p-values

The p-value measures the "extremeness" of the test statistic.

Note:
  1. This probability is calculated assuming that the null hypothesis is true.
  2. Beware: The p-value is not the probability that H_0 is true, nor is it an error probability!
  3. The p-value is between 0 and 1. $\rightarrow$ baseline: $\bar{X} = 0$ if $\mu = 0$.

0: data nearly impossible if H_0 true.

p-values

The calculation of the p-value depends on whether the test is upper-, lower-, or two-tailed.

$\neq$

$> \quad <$

Each of these is the probability of getting a value at least as extreme as what was obtained (assuming H_0 true).
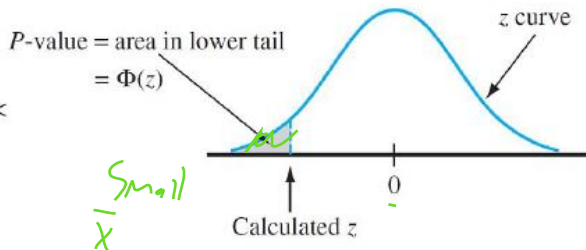
p-values for Z tests

1. **Upper-tailed test**
   H_a contains the inequality >

   z curve

   P-value = area in upper tail
   = $1 - \Phi(z)$

   $P(\bar{X} \geq$ observed $\bar{X}$ | $H_0)$
   any

   0

   Calculated *z* from data given H_0
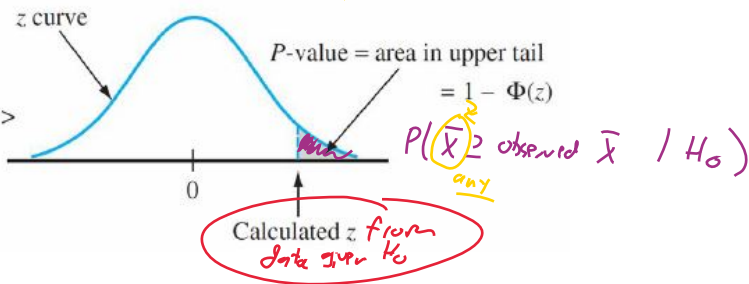
2. **Lower-tailed test**
   H_a contains the inequality <

   P-value = area in lower tail
   = $\Phi(z)$

   z curve

   Small
   $\overline{X}$

   Calculated *z*

   0

p-values for Z tests

$P\text{-value} = \text{sum of area in two tails} = 2[1 - \Phi(|z|)]$
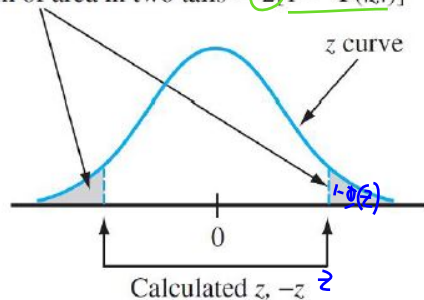
z curve

3. **Two-tailed test**

H_a contains the inequality $\neq$

1-$\Phi(z)$

0

Calculated z, $-z$ $\geq$

unfair T

|| +/

p-values for Z tests

Back to the lawnmower engine example: There, we had

$H_0 : \mu = 300$ vs $H_a : \mu > 300$

and $Z = 1.17$.

What is the p-value for this result? What decision would you make?

p-values for Z tests

Back to the lawnmower engine example: There, we had

$H_0 : \mu = 300$ vs $H_a : \mu > 300$

and $Z = 1.17$.

What is the p-value for this result? What decision would you make?

Hopefully the same one! In particular,
$P(Z > 1.17) = $ `1-stats.norm.cdf(1.17)` $= .12 = p > \alpha$
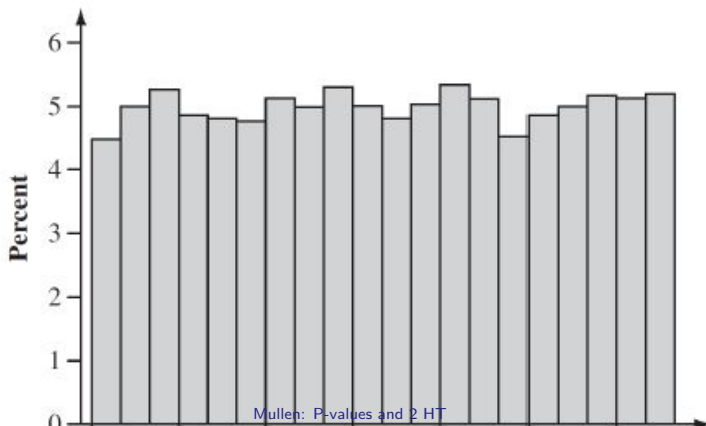
Hyp. Test:

1) RR: reject if
   $1.17 > 2$

2) p-value
   compare
   $p(data) = .12$ to $\alpha = .05$

Distribution of p-values

Figure below shows a histogram of the 10,000 P-values from a simulation experiment under a null $\mu = 20$ (with n = 4 and = 2).
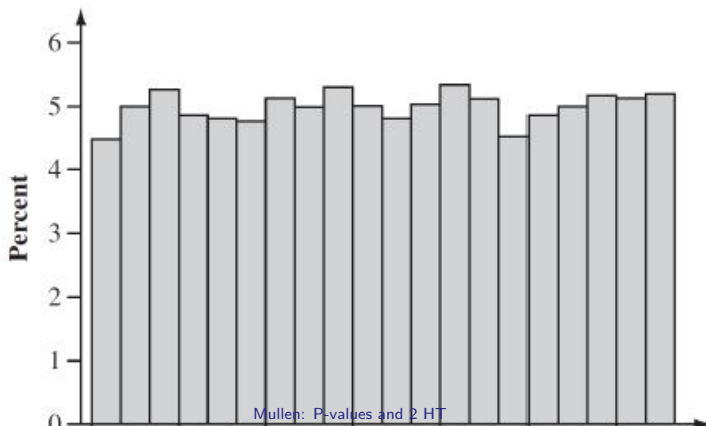
When H0 is true, the probability distribution of the P-value is a uniform distribution on the interval from 0 to 1.

Distribution of p-values

Figure below shows a histogram of the 10,000 P-values from a simulation experiment under a null $\mu = 20$ (with n = 4 and = 2).

When H0 is true, the probability distribution of the P-value is a uniform distribution on the interval from 0 to 1.
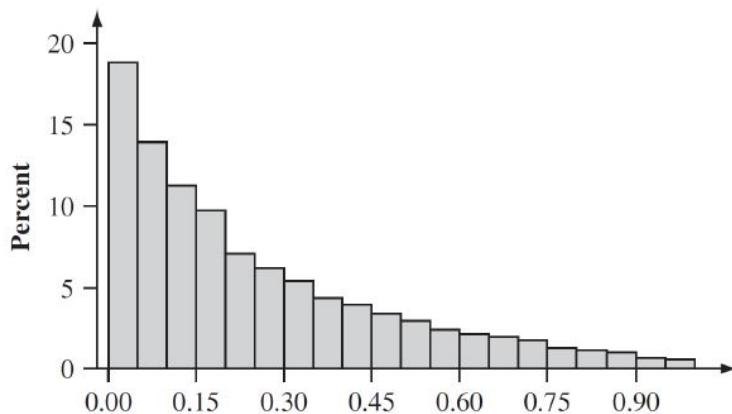
Distribution of p-values

About 4.5% of these p-values are in the interval from 0 to .05.

Thus when using a significance level of .05, the null hypothesis is rejected in roughly 4.5% of these 10,000 tests.

If we continued to generate samples and carry out the test for each sample at significance level .05, in the long run, 5% of the p-values would be in the first class interval.

Distribution of p-values

A histogram of the p-values when we simulate under an alternative hypothesis. There is a much greater tendency for the p-value to be small (closer to 0) when $\mu = 21$ than when $\mu = 20$
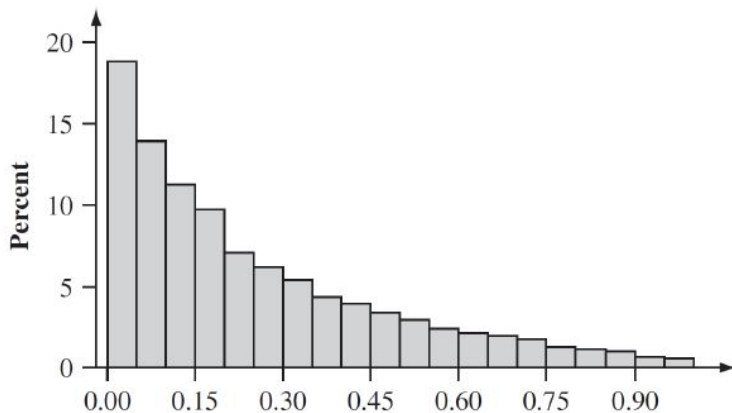
Distribution of p-values

A histogram of the p-values when we simulate under an alternative hypothesis. There is a much greater tendency for the p-value to be small (closer to 0) when $\mu = 21$ than when $\mu = 20$
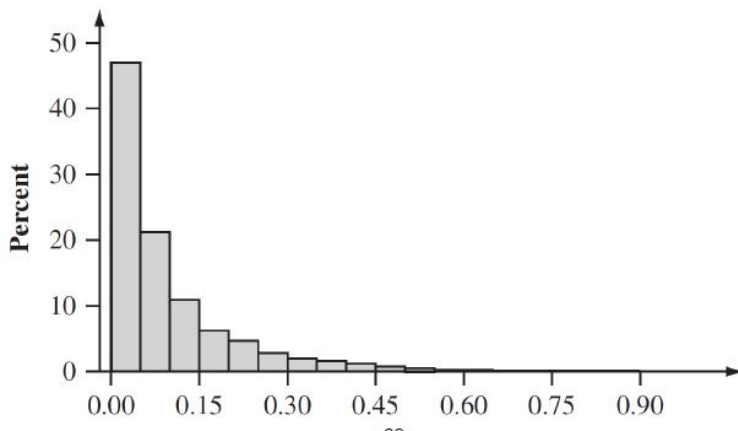
Distribution of p-values

Again, H_0 is rejected at significance level .05 whenever the p-value is at most .05 (in the first bin).

Unfortunately, this is the case for only about 19% of the p-values. So only about 19% of the 10,000 tests correctly reject the null hypothesis; for the other 81%, a type II error is committed.

The difficulty is that the sample size is quite small and 21 is not very different from the value asserted by the null hypothesis (the "effect size" is small).
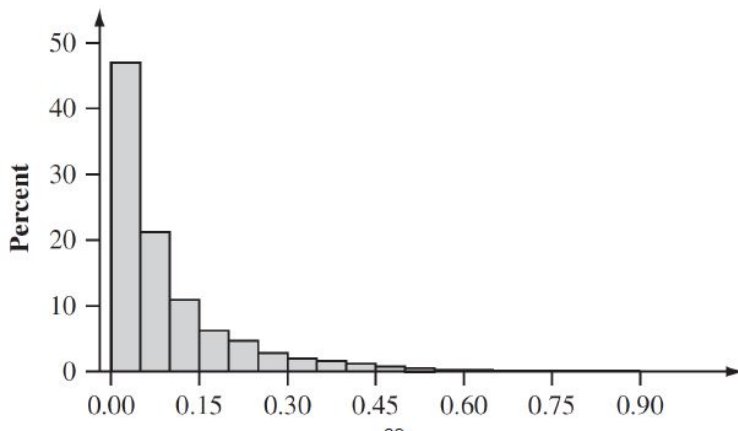
Distribution of p-values

The figure below illustrates what happens to the pvalue when H_0 is false because $\mu = 22$.

Distribution of p-values

The figure below illustrates what happens to the pvalue when H_0 is false because $\mu = 22$.

Again, these data come from H_a, but we reach the correct conclusion more often. The

Distribution of p-values

The histogram is even more concentrated toward values close to 0 than was the case when $\mu = 21$.

In general, as moves further to the right of the null value 20, the distribution of the p-value will become more and more concentrated on values close to 0.

Even here a bit fewer than 50% of the p-values are smaller than .05. So it is still slightly more likely than not that the null hypothesis is incorrectly not rejected. Only for values of much larger than 20 (e.g., at least 24 or 25) is it highly likely that the p-value will be smaller than .05 and thus give the correct conclusion.

Test for Population Mean (Small Sample)

When the sample size is small and the population is normal, we can use a t-test.
Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

Alternative Hypothesis   Rejection Region for α level test:

Test for Population Mean (Small Sample)

When the sample size is small and the population is normal, we can use a t-test.
Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Alternative Hypothesis   Rejection Region for α level test:

Test for Population Mean (Small Sample)

When the sample size is small and the population is normal, we can use a t-test.
Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Alternative Hypothesis   Rejection Region for α level test:

$H_a : \mu > \mu_0$
$H_a : \mu < \mu_0$
$H_a : \mu \neq \mu_0$

Test for Population Mean (Small Sample) *Variance unknown*

When the sample size is small and the population is normal, we can use a t-test.

Null hypothesis: $H_0 : \mu = \mu_0$

Test statistic value:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

Alternative Hypothesis   Rejection Region for α level test:

$\nu = n - 1$

$H_a : \mu > \mu_0$ $t > t_{\alpha,}$

$H_a : \mu < \mu_0$ $t < -t_{\alpha,}$

$H_a : \mu \neq \mu_0$ $t < -t_{\alpha/2,}$ **or** $t > t_{\alpha/2,}$

t example.

The Brinell scale is a measure of how hard a material is. An engineer hypothesizes that the mean Brinell score of all subcritically annealed ductile iron pieces is not equal to 170. It is known that these scores follow a normal distribution.

$$\mu = 170$$
$$\mu_0$$

$$H_A: \quad \mu \neq 170$$

The engineer measured the Brinell score of 25 pieces of this type of iron and calculated the sample mean to be 174.52 and the sample standard deviation to be 10.31.

$$\bar{x} \qquad n = 25$$

$$s = 10.31$$

Perform a hypothesis test that the true average Brinell score is not equal to 170, as well as the corresponding confidence interval. Set alpha = 0.01.

t sol:

We have: $\bar{X} = 174.52$; $n = 25$; $s = 10.31$, and underlying normal. That means that

t sol:

We have: $\bar{X} = 174.52$; $n = 25$; $s = 10.31$, and underlying normal. That means that

$$Z = \frac{\bar{X} - 170}{\sigma/\sqrt{n}}$$

is normally distributed... but we need to use s! It's t-time! That would give us a test statistic of

$$t = \frac{17\bar{4}.52 - 170}{10.31/\sqrt{25}} = 2.19$$

t sol:

We have: $\bar{X} = 174.52$; $n = 25$; $s = 10.31$, and underlying normal. That means that

$$Z = \frac{\bar{X} - 170}{\sigma/\sqrt{n}}$$

is normally distributed... but we need to use s! It's t-time! That would give us a test statistic of

$$t = \frac{17\bar{4}.52 - 170}{10.31/\sqrt{25}} = 2.19$$

From 'is not equal to,' we know we're testing the two sided $H_a : \mu \neq 170$, so we compared this to $t_{\alpha/2} = $ stats.t.ppf(.995, 24) $= 2.79$

$$.995$$
$$= 1 - \frac{.01}{2}$$

t sol:

We have: $\bar{X} = 174.52$; $n = 25$; $s = 10.31$, and underlying normal. That means that

$$Z = \frac{\bar{X} - 170}{\sigma/\sqrt{n}}$$

is normally distributed... but we need to use s! It's t-time! That would give us a test statistic of

$$t = \frac{17\bar{4}.52 - 170}{10.31/\sqrt{25}} = 2.19$$

From 'is not equal to,' we know we're testing the two sided $H_a : \mu \neq 170$, so we compared this to $t_{\alpha/2} = \texttt{stats.t.ppf(.995, 24)} = 2.79$

Our test statistic t is less than $t_{\alpha/2}$, so we **do not reject**, or there is not sufficient evidence to discount the null hypothesis.

Comparing 2 Means

$$\frac{\left(\overline{x} - \overline{Y}\right) - \left(\mu_x - \mu_Y\right)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_Y^2}{m}}}$$

We now have the framework to finish asking most of the questions you'll most often see for binary decisions. We need to flesh out what to do with 2 samples for both Hypothesis testing and for t statistics.

Spoilers: The same way we could test a null of $H_0 : \mu = \mu_0$ against a two-sided alternative by looking at the CI for μ of our sample, we can look at a CI for $\mu_1 - \mu_2$ to get similar inference to make a decision like $\mu_1 \neq \mu_2$.

Comparing 2 Means

Basic Assumptions:

Note: We haven't made any distributional assumptions, for now.

Comparing 2 Means

Basic Assumptions:

1. $X_1, X_2, \ldots X_n$ are a random sample from distribution 1 with mean μ_1 (or μ_X) and SD σ_1.

2. $Y_1, Y_2, \ldots Y_m$ are a random sample from distribution 2 with mean μ_2 and SD σ_2.

3. The X and Y sample are independent of one another.

keep this in mind

Note: We haven't made any distributional assumptions, for now.

Comparing 2 Means: Review

The natural estimator of $\mu_1 - \mu_2$ is _____.

Inferential procedures are based on standardizing estimators, so as before we need the mean and standard deviation of _____.

Mean of _____ :

Variance/Standard Deviation of _____ :

Comparing 2 Means: Review

The natural estimator of $\mu_1 - \mu_2$ is $\underline{\bar{X} - \bar{Y}}$.

Inferential procedures are based on standardizing estimators, so as before we need the mean and standard deviation of $\underline{\bar{X} - \bar{Y}}$.

Mean of $\bar{X} - \bar{Y}$:

Variance/Standard Deviation of $\bar{X} - \bar{Y}$:

Comparing 2 Means: Review

The natural estimator of $\mu_1 - \mu_2$ is $\underline{\bar{X} - \bar{Y}}$.

Inferential procedures are based on standardizing estimators, so as before we need the mean and standard deviation of $\underline{\bar{X} - \bar{Y}}$.

Mean of $\bar{X} - \bar{Y}$:

$$E[\underbrace{\bar{X} - \bar{Y}}] = E\left[\frac{\sum_i X_i}{n} - \frac{\sum_j Y_j}{m}\right] = \cdots = \underline{\mu_1 - \mu_2}$$

Variance/Standard Deviation of $\bar{X} - \bar{Y}$:

$$Var\left[\bar{X} - \bar{Y}\right] = Var\left[\frac{\sum_i X_i}{n} - \frac{\sum_j Y_j}{m}\right] = Var[\bar{X}] + Var[\bar{Y}] = \ldots$$

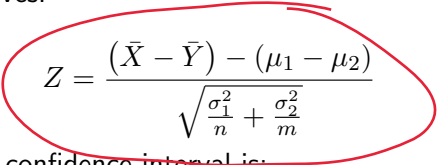$$= \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$$

Comparing 2 Means

Normal Populations with known variances:

If both populations are normal and independent, $\bar{X} - \bar{Y}$ is normally distributed with expected value $\mu_1 - \mu_2$ and standard deviation: $\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$. So:

$$\left(\bar{X} - \bar{Y}\right) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

Standardizing our estimator gives:

$$Z = \frac{\left(\bar{X} - \bar{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

Therefore, the $(1 - \alpha) \cdot 100\%$ confidence interval is:

$$\left(\bar{X} - \bar{Y}\right) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

Test Procedures for Normal Populations with Known Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

Alt Hypothesis   Rejection Region   p-value:

Test Procedures for Normal Populations with Known Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

Sample diff   *theoretical diff H_0*

$$z_{stat} = \frac{(\bar{X} - \bar{Y}) - (\Delta_0)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

Std. error

Alt Hypothesis	Rejection Region	p-value:

Test Procedures for Normal Populations with Known Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$z_{stat} = \frac{(\bar{X} - \bar{Y}) - (\Delta_0)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$		
$H_a : \mu < \mu_0$		
$H_a : \mu \neq \mu_0$		

Test Procedures for Normal Populations with Known Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$z_{stat} = \frac{\left(\bar{X} - \bar{Y}\right) - (\Delta_0)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$	$z_{stat} > z_\alpha$	
$H_a : \mu < \mu_0$	$z_{stat} < -z_\alpha$	
$H_a : \mu \neq \mu_0$	$|z_{stat}| > z_{\alpha/2}$	

Test Procedures for Normal Populations with Known Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$z_{stat} = \frac{(\bar{X} - \bar{Y}) - (\Delta_0)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$	$z_{stat} > z_\alpha$ ✓	$P(Z > z_{stat})$ ✓
$H_a : \mu < \mu_0$	$z_{stat} < -z_\alpha$	$P(Z < z_{stat})$
$H_a : \mu \neq \mu_0$	$|z_{stat}| > z_{\alpha/2}$	$P(|Z| > |z_{stat}|)$

$\mu > \mu_0 :$

$\Delta > \Delta_0$

$\mu_1 - \mu_2 > \Delta_0$

$\mu_1 > \mu_2 + \Delta_2$

Large Sample Tests

The assumptions of normal population distributions and known standard deviations are fortunately unnecessary when both sample sizes are sufficiently large. WHY?

Furthermore, using _____ and _____ in place of ___ and ___ gives a variable whose distribution is approximately standard normal:

These tests are usually appropriate if both $m > 30$ and $n > 30$.

Large Sample Tests

The assumptions of normal population distributions and known standard deviations are fortunately unnecessary when both sample sizes are sufficiently large. WHY?

Furthermore, using $\underline{s_x^2}$ and $\underline{s_y^2}$ in place of $\underline{\sigma_x^2}$ and $\underline{\sigma_y^2}$ gives a variable whose distribution is approximately standard normal:

These tests are usually appropriate if both $m > 30$ and $n > 30$.

Large Sample Tests

The assumptions of normal population distributions and known standard deviations are fortunately unnecessary when both sample sizes are sufficiently large. WHY?

Furthermore, using $\underline{s_x^2}$ and $\underline{s_y^2}$ in place of $\underline{\sigma_x^2}$ and $\underline{\sigma_y^2}$ gives a variable whose distribution is approximately standard normal:

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

These tests are usually appropriate if both $m > 30$ and $n > 30$.

Large Sample Tests

Example:

Data on daily calorie intake both for a sample of teens who said they did not typically eat fast food and another sample of teens who said they did usually eat fast food.

Eat Fast Food	Sample Size	Sample Mean	Sample SD
No	663	2258	1519
Yes	413	2637	1138

Does this data provide strong evidence for concluding that true average calorie intake for teens who typically eat fast food exceeds more than 200 calories per day the true average intake for those who don't typically eat fast food? Let's investigate by carrying out a test of hypotheses at a significance level of 0.05.

Large Sample Tests

Example:
Data on daily calorie intake both for a sample of teens who said they did not typically eat fast food and another sample of teens who said they did usually eat fast food.

Eat Fast Food	Sample Size	Sample Mean	Sample SD
No	663	2258	1519
Yes	413	2637	1138

Does this data provide strong evidence for concluding that true average calorie intake for teens who typically eat fast food exceeds more than 200 calories per day the true average intake for those who don't typically eat fast food? Let's investigate by carrying out a test of hypotheses at a significance level of 0.05.

$$H_0 : \mu_2 - \mu_1 = 200; \qquad H_a : \mu_2 - \mu_1 > 200$$

Plan: Reject if $p < .05$.

Large Sample Tests

Example:

Data on daily calorie intake both for a sample of teens who said they did not typically eat fast food and another sample of teens who said they did usually eat fast food.

Eat Fast Food	Sample Size	Sample Mean	Sample SD
No	663	2258	1519
Yes	413	2637	1138

Does this data provide strong evidence for concluding that true average calorie intake for teens who typically eat fast food exceeds more than 200 calories per day the true average intake for those who don't typically eat fast food? Let's investigate by carrying out a test of hypotheses at a significance level of 0.05.

$$z_{stat} = \frac{(2637 - 2258) - (200)}{\sqrt{\frac{1519^2}{663} + \frac{1138^2}{413}}} = 2.20 \qquad P(Z < 2.20) = .987$$

Comparing 2 Means: Small Sample

For large samples, the CLT allows us to use these methods we have discussed even when the two populations of interest are not normal.

In practice, it can happen that at least one sample size is small and the population variances have unknown values.

Without the CLT at our disposal, we proceed by making specific assumptions about the underlying population distributions.

Comparing 2 Means: Small Sample

When the population distributions are both normal, the standardized variable

has approximately a t distribution with df ν estimated from the data by:

Comparing 2 Means: Small Sample

When the population distributions are both normal, the standardized variable

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

has approximately a t distribution with df ν estimated from the data by:

Comparing 2 Means: Small Sample

When the population distributions are both normal, the standardized variable

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

has approximately a t distribution with df ν estimated from the data by:

$$\nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{(s_1^2/m)^2}{m-1} + \frac{(s_2^2/n)^2}{n-1}}$$

Comparing 2 Means: Small Sample

The two-sample t confidence interval for $\mu_1 - \mu_2$ with confidence level $(1 - \alpha) \cdot 100\%$ is then:

$$\left(\bar{X} - \bar{Y} \right) \pm t_{\alpha/2, \nu} \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$$

Test Procedures for Normal Populations with Unknown Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

Alt Hypothesis   Rejection Region   p-value:

Test Procedures for Normal Populations with Unknown Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$t_{stat} = \frac{\left(\bar{X} - \bar{Y} \right) - (\Delta_0)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

Alt Hypothesis   Rejection Region   p-value:

Test Procedures for Normal Populations with Unknown Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$t_{stat} = \frac{(\bar{X} - \bar{Y}) - (\Delta_0)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$		
$H_a : \mu < \mu_0$		
$H_a : \mu \neq \mu_0$		

Test Procedures for Normal Populations with Unknown Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$t_{stat} = \frac{(\bar{X} - \bar{Y}) - (\Delta_0)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$	$t_{stat} > t_{\alpha,\nu}$	
$H_a : \mu < \mu_0$	$t_{stat} < -t_{\alpha,\nu}$	
$H_a : \mu \neq \mu_0$	$|t_{stat}| > t_{\alpha/2,\nu}$	

Test Procedures for Normal Populations with Unknown Variances

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$t_{stat} = \frac{(\bar{X} - \bar{Y}) - (\Delta_0)}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$	$t_{stat} > t_{\alpha,\nu}$	$P(T > t_{stat})$
$H_a : \mu < \mu_0$	$t_{stat} < -t_{\alpha,\nu}$	$P(T < t_{stat})$
$H_a : \mu \neq \mu_0$	$|t_{stat}| > t_{\alpha/2,\nu}$	$P(|T| > |t_{stat}|)$

Test for Equivalence of Proportions

Theoretically, we know that:

has approximately a standard normal distribution.

When $H_0 : p_1 - p_2 = 0$ is true, we have $p_1 = p_2$, which simplifies this:

However, this Z cannot serve as a test statistic because the value of p is unknown; H_0 asserts only that there is a common value of p, but does not say what that value is.

Test for Equivalence of Proportions

Theoretically, we know that:

$$Z = \frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

has approximately a standard normal distribution.

When $H_0 : p_1 - p_2 = 0$ is true, we have $p_1 = p_2$, which simplifies this:

However, this Z cannot serve as a test statistic because the value of p is unknown; H_0 asserts only that there is a common value of p, but does not say what that value is.

Test for Equivalence of Proportions

Theoretically, we know that:

$$Z = \frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

has approximately a standard normal distribution.

When $H_0 : p_1 - p_2 = 0$ is true, we have $p_1 = p_2$, which simplifies this:

$$Z = \frac{(\hat{p_1} - \hat{p_2})}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}}$$

However, this Z cannot serve as a test statistic because the value of p is unknown; H_0 asserts only that there is a common value of p, but does not say what that value is.

Test for Equivalence of Proportions

Under the null hypothesis, we assume that $p_1 = p_2 = p$, instead of separate samples of size m and n from two different populations (two different binomial distributions).

So, we really have a single sample of size $m + n$ from one population with proportion p.

The total number of individuals in this combined sample having the characteristic of interest is $X + Y$.

The estimator of p is then:

Test for Equivalence of Proportions

Under the null hypothesis, we assume that $p_1 = p_2 = p$, instead of separate samples of size m and n from two different populations (two different binomial distributions).

So, we really have a single sample of size $m + n$ from one population with proportion p.

The total number of individuals in this combined sample having the characteristic of interest is $X + Y$.

The estimator of p is then: $\hat{p} = \frac{X+Y}{n+m}$

Test for Equivalence of Proportions

Using and $\hat{p}; 1 - \hat{p}$ in place of p and $1 - p$ in our old equation gives a test statistic having approximately a standard normal distribution when H_0 is true.

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

Alt Hypothesis   Rejection Region   p-value:

Test for Equivalence of Proportions

Using and $\hat{p}; 1 - \hat{p}$ in place of p and $1 - p$ in our old equation gives a test statistic having approximately a standard normal distribution when H_0 is true.

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$z_{stat} = \frac{(\hat{p_1} - \hat{p_2})}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

Alt Hypothesis   Rejection Region   p-value:

Test for Equivalence of Proportions

Using and $\hat{p}; 1 - \hat{p}$ in place of p and $1 - p$ in our old equation gives a test statistic having approximately a standard normal distribution when H_0 is true.

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$z_{stat} = \frac{(\hat{p_1} - \hat{p_2})}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$		
$H_a : \mu < \mu_0$		
$H_a : \mu \neq \mu_0$		

Test for Equivalence of Proportions

Using and $\hat{p}; 1 - \hat{p}$ in place of p and $1 - p$ in our old equation gives a test statistic having approximately a standard normal distribution when H_0 is true.

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$z_{stat} = \frac{(\hat{p_1} - \hat{p_2})}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$	$z_{stat} > z_\alpha$	
$H_a : \mu < \mu_0$	$z_{stat} < -z_\alpha$	
$H_a : \mu \neq \mu_0$	$|z_{stat}| > z_{\alpha/2}$	

Test for Equivalence of Proportions

Using and $\hat{p}; 1 - \hat{p}$ in place of p and $1 - p$ in our old equation gives a test statistic having approximately a standard normal distribution when H_0 is true.

Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$ (often 0)

Test statistic value:

$$z_{stat} = \frac{(\hat{p_1} - \hat{p_2})}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}}$$

Alt Hypothesis	Rejection Region	p-value:
$H_a : \mu > \mu_0$	$z_{stat} > z_\alpha$	$P(Z > z_{stat})$
$H_a : \mu < \mu_0$	$z_{stat} < -z_\alpha$	$P(Z < z_{stat})$
$H_a : \mu \neq \mu_0$	$|z_{stat}| > z_{\alpha/2}$	$P(|Z| > |z_{stat}|)$

Now what?

We're almost done talking about CI's and Hypothesis tests. Where are our gaps?

1. We can compare samples or do inference on a single sample when one of the following conditions is met:

    1.1 The sample or samples are $n > 30$ (or success/fail>10) (use Z!)

    1.2 The sample or samples are small and underlying normal (use t!)

2. What are we missing?

    2.1 The samples are small and *not normal*

    2.2 We aren't trying to do inference on means at all, but something else!

    2.3 We will cover 2 cases of this Wednesday: variances and bootstrapping.

Daily Recap

Today we learned

1. Comparing multiple large or normal samples for equivalence of the mean!

Moving forward:

- nb day Friday

Next time in lecture:

- Bootstrapping!