Lecture 7: Composite Hypotheses

STA261 – Probability & Statistics II

Ofir Harari

Department of Statistical Sciences

University of Toronto

Outline

Composite Hypotheses

- When the alternative hypothesis is of the form $\mathcal{H}_1 : \theta \in \Theta_1$, where Θ_1 consists of more than a single possible value, we call $\mathcal{H}_1$ a *composite* alternative.
- Consider for example $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ (with σ^2 known for now). One may choose to test –
  1. $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu > \mu_0$ (a *right-tailed test*),
  2. $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu < \mu_0$ (a *left-tailed test*), or
  3. $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu \neq \mu_0$ (a *two-tailed test*),
- Recall that when we tested $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta = \theta_1$, the power of the test was defined to be

rejecting when we should reject

$$\pi = 1 - \beta = \mathbb{P}\left(\begin{array}{c} \text{reject} \\ \mathcal{H}_0 \end{array} \middle| \theta = \theta_1 \right).$$

- Now there is no θ_1 anymore, but rather a set of alternatives $\theta \in \Theta_1$.

cannot talk about power anymore, dunno which alternative we are talking about..

Power Functions

Definition the graph of a function

The *power function* of a statistical test is

$$\pi(\theta^*) = \mathbb{P}\left(\begin{array}{c} \text{reject} \\ \mathcal{H}_0 \end{array} \middle| \theta = \theta^* \right).$$

- Consider for example the LRT test we developed for the simple hypotheses $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu = \mu_1$ ($\mu_1 > \mu_0$) for Normal data with known variance, based on the rejection region

$$\mathcal{C} = \left\{ \underline{x} \in \mathbb{R}^n \, : \, \overline{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}.$$

- Note that $\mathcal{C}$ does not depend on μ_1: the test is the same for any $\mu_1 > \mu_0$.
- We will use it then for the right-tailed test – $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu > \mu_0$.

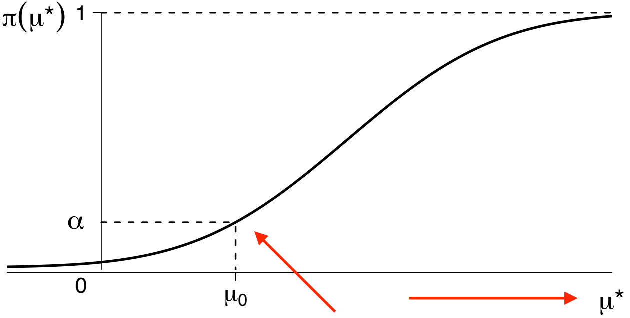
Power Functions (cont.)

power: reject null when alternative is true

$$\pi(\mu^*) = \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| \mu = \mu^*\right) = \mathbb{P}\left(\overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \middle| \mu = \mu^*\right)$$

$$= 1 - \mathbb{P}\left(\frac{\overline{X} - \mu^*}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu^*}{\sigma/\sqrt{n}} + z_{1-\alpha} \middle| \mu = \mu^*\right) = 1 - \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu^*)}{\sigma} + z_{1-\alpha}\right).$$

standardize w.r.t. mu^*, not mu_0 ….



when mu = mu_0; then \pi is probability of rejecting when null is true, equivalent to type 1 error

Power calculations

$$\pi(\mu^*) = 1 - \Phi\left(\frac{-\sqrt{n}(\mu^* - \mu_0)}{\sigma} + z_{1-\alpha}\right)$$

- A quick look at $\pi(\mu^*)$ reveals that –
  1. The larger n is – the more powerful the test is (explain)
  2. The further apart the null hypothesis and the alternative are – the greater the power is (explain) expect to reject null easier
  3. The larger σ is – the less powerful the test is (explain)
  4. The smaller α is – the smaller the power is (has been discussed already)
- With that said, for any given α and we can make β arbitrarily small by increasing n – so long as we know the kind of differences $\mu^* - \mu$ that we wish to uncover.

since H_1 and alpha are set before the test. and sigma is based on the data.
only thing that can change is the size of samples.

UNIVERSITY OF TORONTO

note that we only need to ensure that power at mu_1 is greater than 1 - beta, because for mu > mu_1, power is larger further apart null and alternative hypothesis are

Power calculations (cont.)

- Suppose that we wish to test $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu > \mu_0$ with σ^2 assumed to be known, as before, at significance level α.
- In addition, differences smaller than $\delta = \mu_1 - \mu_0$ are deemed non-consequential, but beyond that difference we wish to keep the probability of a type II error at less than β.

note power function is higher for higher mu_1

$$1 - \beta \leq \pi(\mu_1) = 1 - \Phi\left(-\frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma} + z_{1-\alpha}\right)$$

calculate n required to have higher output to power function than specified beta

$$\Longrightarrow \Phi\left(-\frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma} + z_{1-\alpha}\right) \leq \beta$$

$$\Longrightarrow -\frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma} + z_{1-\alpha} \leq z_\beta = -z_{1-\beta}$$

$$\Longrightarrow \frac{\sqrt{n}(\mu_1 - \mu_0)}{\sigma} \geq z_{1-\alpha} + z_{1-\beta} \Longrightarrow n \geq \left\{\frac{\sigma(z_{1-\alpha} + z_{1-\beta})}{\mu_1 - \mu_0}\right\}^2.$$

calculate minimum sample size to achieve the specified alpha, beta beyond the \mu_1

Power calculations (cont.)

- For example, Suppose we know that the serum cholesterol levels for all 20-24 year-old males in Canada is normally distributed with $\mu = 180$ mg/100ml and the $\sigma = 46$ mg/100ml.

- We would expect that the mean cholesterol level of a special diet group in this population to be higher than 180 mg/100ml. (Assuming the same $\sigma = 46$ mg/100ml.)

mu_1

beyong what mu_1 do we care, i.e. ensure power is at least 90% beyond 180 + 20 = 200

- If the mean cholesterol level increases by at least 20 mg/100ml, we would like to be able to show that the new diet is effective (i.e. reject $\mathcal{H}_0 : \mu = 180$ in favor of $\mathcal{H}_1 : \mu > 180$) at a significance level of 5% and with a power of 90%.

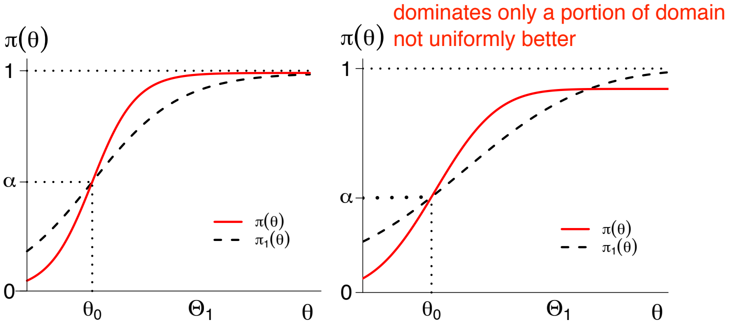
- To do that we will need a sample of size

$$n \geq \left\{ \frac{\sigma(z_{1-\alpha} + z_{1-\beta})}{\mu_1 - \mu_0} \right\}^2 = \left\{ \frac{46(\overbrace{1.645}^{z_{0.95}} + \overbrace{1.282}^{z_{0.9}})}{200 - 180} \right\}^2 = 45.3 \Longrightarrow n \geq 46.$$

Uniformly Most Powerful Tests

Definition

Consider testing $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta \in \Theta_1$ (a composite alternative). We say that a test at level α with power function $\pi(\theta)$ is a *uniformly most powerful* (UMP) test, if for any other test at level α with power function $\pi_1(\theta)$, we have $\pi_1(\theta) \le \pi(\theta)$ for all $\theta \in \Theta_1$.



UMP (left panel) and not UMP (right panel).

An example of a UMP test

- For $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ (with a known σ^2), we have proposed the rejection region
$$\mathcal{C} = \left\{ \underline{x} \in \mathbb{R}^n \ : \ \overline{x} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$
to test $\mathcal{H}_0 : \mu = \mu_0$ vs. the right-tailed alternative $\mathcal{H}_1 : \mu > \mu_0$.
- Here $\Theta_1 = (\mu_0, \infty]$ (the set of alternatives)
- This has been shown to be the most powerful (MP) test at level α for testing $\mathcal{H}_0 : \mu = \mu_0$ vs. $\mathcal{H}_1 : \mu = \mu_1$ $(\mu_1 > \mu_0)$ for simple hypotheses
- It does not depend on the value of μ_1, hence it is the most powerful test for any $\mu \in \Theta_1$ test is MP for all mu_1 > mu_0 ==> test is UMP
- But this is the definition of a UMP test!

The p-value

- We have just shown that

$$\mathcal{C}_{\text{right}} = \left\{ \overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

forms a rejection for a UMP test at level α for testing $\mathcal{H}_0 : \mu = \mu_0$ vs. the right-tailed alternative $\mathcal{H}_1 : \mu > \mu_0$.

- It is equally simple to show that

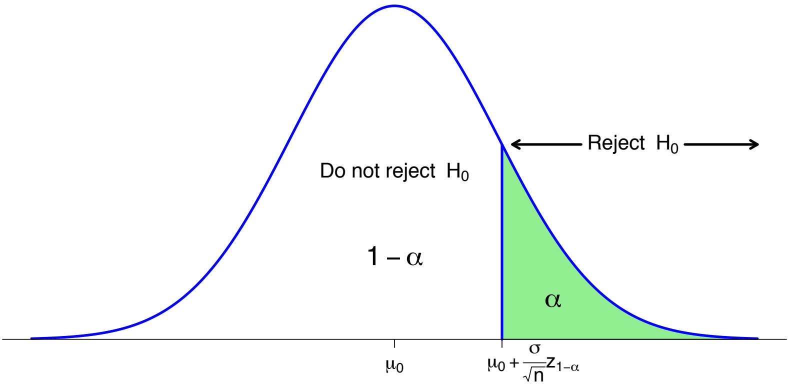
$$\mathcal{C}_{\text{left}} = \left\{ \overline{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}$$

forms a rejection region for a UMP test at level α for testing $\mathcal{H}_0 : \mu = \mu_0$ vs. the left-tailed alternative $\mathcal{H}_1 : \mu < \mu_0$.

- If $\mathcal{H}_0 : \mu = \mu_0$ is true, $\overline{X} \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$

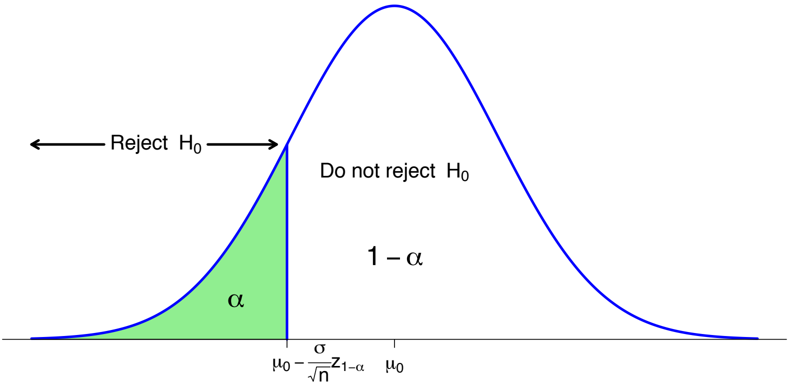
The p-value (cont.)

The distribution of $\overline{X}$ under $\mathcal{H}_0$ and the rejection region for the right-tailed test:



The p-value (cont.)

The distribution of $\overline{X}$ under $\mathcal{H}_0$ and the rejection region for the left-tailed test:



Example

- Suppose that we have a sample of size 25 from a $\mathcal{N}(\mu, 5^2)$ distribution, and we wish to test $\mathcal{H}_0 : \mu = 175$ vs. $\mathcal{H}_1 : \mu > 175$ at 5% level. The observed sample mean was 177.

- The rejection region is

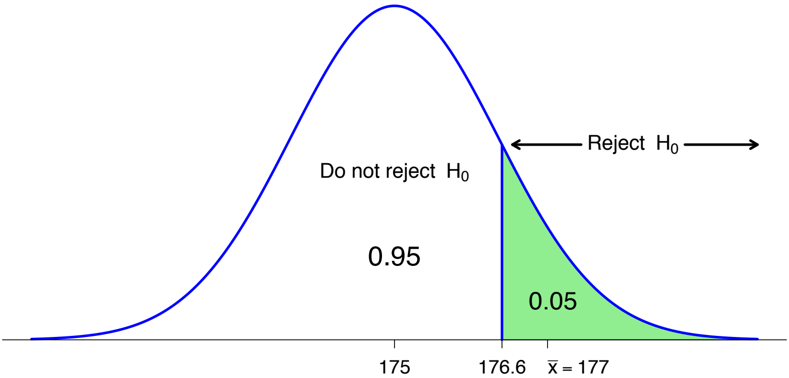
$$\mathcal{C} = \left\{ \overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\} = \left\{ \overline{X} \geq 175 + \frac{5}{5} \underbrace{z_{0.95}}_{1.645} \right\} = \{ \overline{X} \geq 176.6 \}.$$

- In this example we are told that $\overline{x} = 177$ was observed – inside the rejection region.

- at 5% significance level then, we reject $\mathcal{H}_0$.

Example (cont.)

The distribution of $\overline{X}$ under $\mathcal{H}_0$ and the rejection region:



if alpha is set lower, i..e decreases, then its harder to reject,

UNIVERSITY OF TORONTO

Example (cont.)

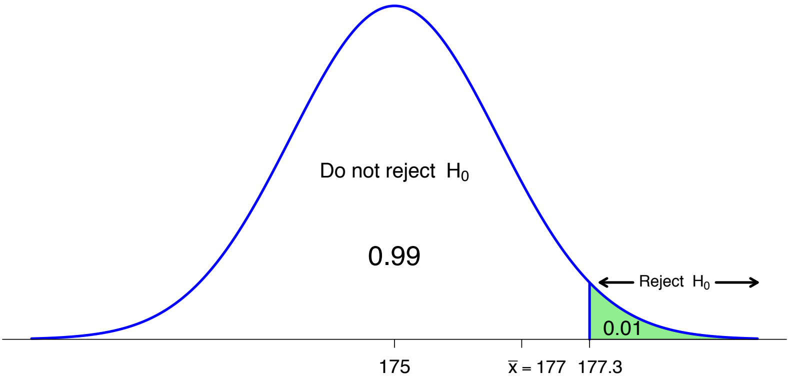
- What if changed our mind and would now like to test the hypotheses at 1% level?
- The new rejection region is

$$\mathcal{C} = \left\{ \overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\} = \left\{ \overline{X} \geq 175 + \frac{5}{5} \underbrace{z_{0.99}}_{2.326} \right\} = \{\overline{X} \geq 177.3\}.$$

- Now the observed sample mean $\overline{x} = 177$ is outside of the rejection region.
- At a 1% significance level then, we do not reject $\mathcal{H}_0$.

Example (cont.)

The distribution of $\overline{X}$ under $\mathcal{H}_0$ and the rejection region:



decrease alpha, line move from left to right, area of rejection gets smaller

The p-value

- We rejected $\mathcal{H}_0$ at 5% level – when $\overline{X}$ was outside of the rejection region
- We did not reject $\mathcal{H}_0$ at 1% level – when $\overline{X}$ was inside the rejection region
- What if $\overline{X}$ was right at the boundary of the rejection region?
  - i.e. the rejection region was $\mathcal{C} = \{\overline{X} \geq \overline{x} = 177\}$

gives smallest alpha that we will still reject H_0

- That would give us the minimum significance level for which $\mathcal{H}_0$ would be rejected. We call that number the *p-value*.
- Let us calculate it:

$$\text{p-value} = \mathbb{P}\left(\begin{array}{c}\text{Type I}\\ \text{error}\end{array}\right) = \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| \mu = 175\right) = \mathbb{P}\left(\overline{X} \geq 177 \middle| \mu = 175\right)$$

$$= \mathbb{P}\left(\frac{\overline{X} - 175}{5/\sqrt{25}} \geq \frac{177 - 175}{5/\sqrt{25}} \middle| \mu = 175\right) = 1 - \Phi(2) = 0.023.$$

Understanding p-values

- We calculated that the p-value for the last example was 0.023

smaller alpha => harder to reject

- It is the minimum α for which $\mathcal{H}_0$ would be rejected – hence we rejected $\mathcal{H}_0$ at 5% but did not reject it at 1% level.

- We no longer need to calculate a new rejection region for every α like we did before, because the rule is

$$\boxed{\text{Reject } \mathcal{H}_0 \text{ at level } \alpha \iff \text{p-value} \leq \alpha}$$

  - At 3% level we would reject $\mathcal{H}_0$
  - At 2% level we would not reject $\mathcal{H}_0$ because 0.02 < 0.023
  - And so on...

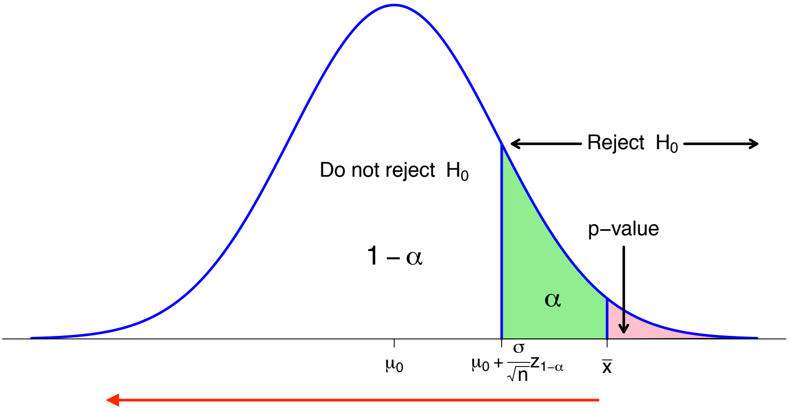
no longer need to calculate different rejection region for differernt alpha

- Recall that we calculated

$$\text{p-value} = \mathbb{P}\left(\overline{X} \geq \overline{x} \middle| \mu = 175\right)$$

Understanding p-values (cont.)

The distribution of $\overline{X}$ under $\mathcal{H}_0$:



p-value = minimum alpha to reject

$$\text{p-value} = \mathbb{P}\left(\overline{X} \geq \overline{x} \middle| \mu = \mu_0\right)$$

Understanding p-values (cont.)

$$\text{p-value} = \mathbb{P}\left(\overline{X} \geq \overline{x} \middle| \mu = \mu_0\right)$$

- The p-value is therefore the probability of observing an effect "at least as extreme" as the one in your data, assuming the truth of $\mathcal{H}_0$.
- Remember that we reject $\mathcal{H}_0 \iff$ p-value $\leq \alpha$ (for a predetermined α)
- The philosophy of hypothesis testing is then, in a nutshell –

> i.e. P(Type I error) is less than pre-determined alpha
>
> If the probability of observing an affect at least as extreme as the one in your data, assuming the truth of $\mathcal{H}_0$, is very low, then $\mathcal{H}_0$ is most likely false. so we reject null

Notes on composite null hypotheses

- Clearly, if we reject $\mathcal{H}_0 : \mu = 175$ in favor of $\mathcal{H}_1 : \mu > 175$, we would also reject $\mathcal{H}_0 : \mu = 170$, $\mathcal{H}_0 : \mu = 150$ or in general $\mathcal{H}_0 : \mu = \mu_0$ for any $\mu_0 \leq 175$.

- This allows us to write the hypotheses in a more "complete" way:

  writing null and alternative hypotheses complementing each other

$$\begin{cases} \mathcal{H}_0 : \underline{\mu \leq \mu_0} \\ \mathcal{H}_1 : \mu > \mu_0 \end{cases} \text{(right-tailed)}, \quad \text{or} \quad \begin{cases} \mathcal{H}_0 : \underline{\mu \geq \mu_0} \\ \mathcal{H}_1 : \mu < \mu_0 \end{cases} \text{(left-tailed)}$$

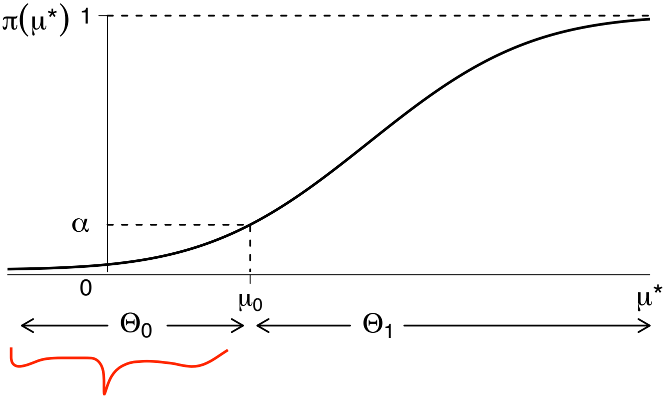
- One could also write down the hypotheses as appearing above, and define the significance level of the test to be

$$\boxed{\alpha = \sup_{\theta \in \Theta_0} \pi(\theta),}$$

  where $\pi(\theta)$ is the power function of the test, as discussed at the beginning of this lecture.

Composite null-hypotheses (cont.)

The power function of the right-tailed test and the significance level:



Two-tailed test for the Normal mean (known σ^2)

- Consider now the problem of testing $\mathcal{H}_0 : \mu = \mu_0$ vs. the two-tailed alternative $\mathcal{H}_1 : \mu \neq \mu_0$.
- When testing $\mathcal{H}_0$ vs. the right-tailed alternative, very positive values of $\overline{X} - \mu_0$ support the alternative, hence the rejection region is of the form
$$\mathcal{C}_{\text{right}} = \left\{ \overline{X} - \mu_0 \geq \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}.$$
- Likewise, when testing $\mathcal{H}_0$ vs. the left-tailed alternative, very negative values of $\overline{X} - \mu_0$ are in support of the alternative, hence the rejection region is of the form
$$\mathcal{C}_{\text{left}} = \left\{ \overline{X} - \mu_0 \leq -\frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right\}.$$
- However, when testing $\mathcal{H}_0$ vs. the two-tailed alternative, a strong evidence against $\mathcal{H}_0$ would be large values of $\left|\overline{X} - \mu_0\right|$.

want absolute value to be larger -> evidence against null

Two-tailed test for μ with known σ^2 (cont.)

- Trivially, we look for a rejection region of the form $\mathcal{C} = \left\{ |\overline{X} - \mu_0| \boxed{\geq c} \right\}$.
- For a test at significance level α, c must satisfy

$$\alpha = \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| \mu = \mu_0\right) = \mathbb{P}\left(|\overline{X} - \mu_0| \geq c \middle| \mu = \mu_0\right)$$

$$= \mathbb{P}\left(\overline{X} - \mu_0 \geq c \middle| \mu = \mu_0\right) + \mathbb{P}\left(\overline{X} - \mu_0 \leq -c \middle| \underline{\mu = \mu_0}\right)$$

$$= \mathbb{P}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{c}{\sigma/\sqrt{n}} \middle| \mu = \mu_0\right) + \mathbb{P}\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \leq -\frac{c}{\sigma/\sqrt{n}} \middle| \mu = \mu_0\right)$$

$$= 1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right) + \Phi\left(-\frac{c\sqrt{n}}{\sigma}\right) = 2\left\{1 - \Phi\left(\frac{c\sqrt{n}}{\sigma}\right)\right\}$$

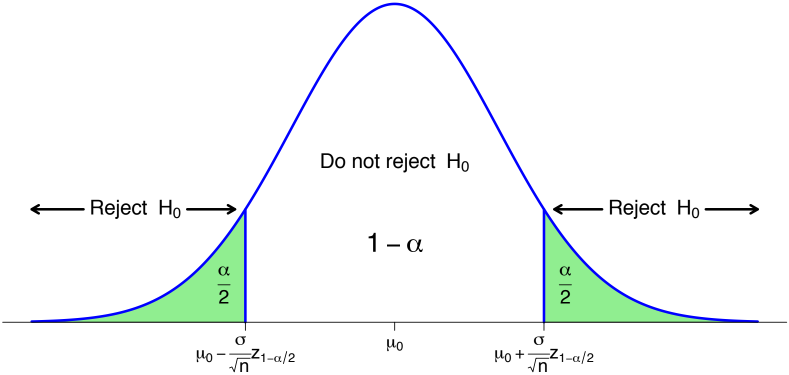
$$\boxed{\Longrightarrow \Phi\left(\frac{c\sqrt{n}}{\sigma}\right) = 1 - \alpha/2 \Longrightarrow \frac{c\sqrt{n}}{\sigma} = z_{1-\alpha/2} \Longrightarrow c = \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}.}$$

note its 1 - alpha/2 for two tailed test; 1 - alpha for one tailed test

Two-tailed test for μ with known σ^2 (cont.)

$$\mathcal{C} = \left\{ \left|\overline{X} - \mu_0\right| \geq \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right\} = \left\{ \overline{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right\} \bigcup \left\{ \overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right\}$$

- The distribution of $\overline{X}$ under $\mathcal{H}_0$ and the rejection region:



Two-tailed test for μ with known σ^2 (cont.)

- Let us now test $\mathcal{H}_0 : \mu = 175$ vs. $\mathcal{H}_1 : \mu \neq 175$ at 4% level, using the same data as before: $n = 25$, $\overline{x} = 177$, $\sigma = 5$.
- The same data yielded p-value = 0.023 for the right-tailed test, hence we would reject $\mathcal{H}_0$ at any level $\alpha \geq 0.023$ – in particular we would reject $\mathcal{H}_0$ at level 4%.
- The rejection region is

$$\mathcal{C} = \left\{ \overline{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right\} \bigcup \left\{ \overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right\}$$

$$= \left\{ \overline{X} \leq 175 - \frac{5}{\sqrt{25}} \underbrace{z_{0.98}}_{2.05} \right\} \bigcup \left\{ \overline{X} \geq 175 + \frac{5}{\sqrt{25}} \underbrace{z_{0.98}}_{2.05} \right\}$$

$$= \left\{ \overline{X} \leq 172.95 \right\} \bigcup \left\{ \overline{X} \geq 177.05 \right\} \implies \text{we do not reject } \mathcal{H}_0!$$

because x = 177

p-value for the two-tailed test

one tail -> hints to where alternative hypothesis lies… so needs more evidence

two tail -> first time, no prior knowledge of where H_1 lies

- Well, that's confusing – the same data that lead us to conclude at 4% level that $\mu > 175$, does not provide enough evidence to conclude that $\mu \neq 175$...?

- The supposed paradox is resolved by re-examining our initial standpoint: testing $\mathcal{H}_0$ vs. a one-tailed alternative suggests that we have an idea as to where to expect the parameter to be with respect to $\mathcal{H}_0$, ergo, we need less convincing.

easier to reject null for one tail alternative than two tail alternative

- Let us reaffirm the result by calculating the p-value.

- Here, observing an effect that is "at least as extreme" as that in your data (assuming the truth of $\mathcal{H}_0$) means observing $\left|\overline{X} - \mu_0\right| \geq |\overline{x} - \mu_0|$, hence

$$\text{p-value} = \mathbb{P}\left(\underline{\left|\overline{X} - \mu_0\right| \geq |\overline{x} - \mu_0|}\,\middle|\,\mu = \mu_0\right).$$

p-value for the two-tailed test (cont.)

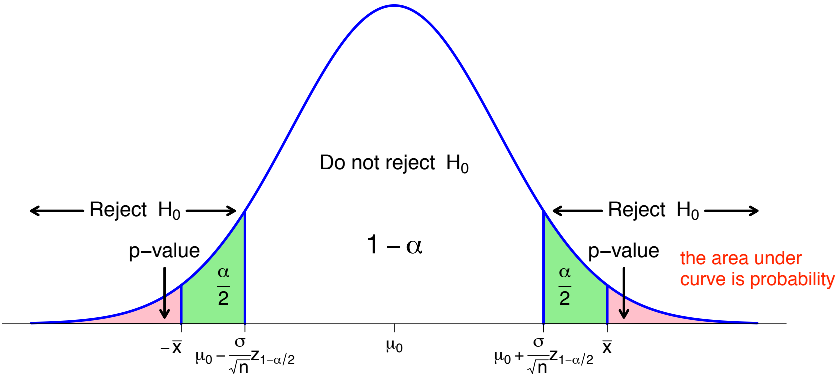
$$
\begin{aligned}
\text{p-value} &= \mathbb{P}\left(\left|\overline{X}-\mu_0\right| \geq |\overline{x}-\mu_0| \,\middle|\, \mu=\mu_0\right) \\
&= \mathbb{P}\left(\frac{\left|\overline{X}-\mu_0\right|}{\sigma/\sqrt{n}} \geq \frac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}} \,\middle|\, \mu=\mu_0\right) \\
&= 2\mathbb{P}\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \geq \frac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}} \,\middle|\, \mu=\mu_0\right) = 2\left\{1-\Phi\left(\frac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}}\right)\right\} \\
&= 2\left\{1-\Phi\left(\frac{|177-175|}{5/\sqrt{25}}\right)\right\} = 2\left\{1-\Phi(2)\right\} = 0.046.
\end{aligned}
$$

- Note that this is exactly double the p-value of the right-tailed test.
- We did not reject $\mathcal{H}_0$ at 4% level, but we would have rejected it at 5% level.

reject as long as p value is smaller than alpha

p-value for the two-tailed test (cont.)

The distribution of $\overline{X}$ under $\mathcal{H}_0$ and the rejection region:



rejecting two tail tests is harder because p-value is doubled in this case

Power function for the two-tailed test

We may now proceed to calculate the power function for the two-tailed test –

$$\pi(\mu^*) = \mathbb{P}\left(\underline{X} \in \mathcal{C} \middle| \mu = \mu^*\right)$$

$$= \mathbb{P}\left(\overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \middle| \mu = \mu^*\right) + \mathbb{P}\left(\overline{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \middle| \mu = \mu^*\right)$$

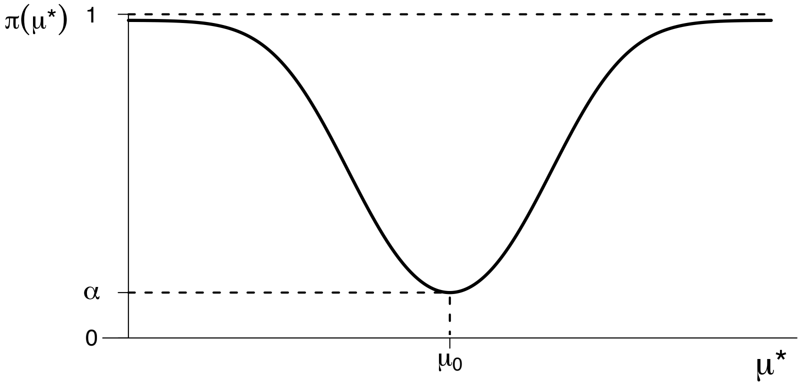
$$= \mathbb{P}\left(\frac{\overline{X} - \mu^*}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu^* + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}}{\sigma/\sqrt{n}} \middle| \mu = \mu^*\right)$$ standardize with respect to mu^*

$$+ \mathbb{P}\left(\frac{\overline{X} - \mu^*}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu^* - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}}{\sigma/\sqrt{n}} \middle| \mu = \mu^*\right)$$

$$= 1 - \Phi\left(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} + z_{1-\alpha/2}\right) + \Phi\left(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} - z_{1-\alpha/2}\right).$$

Power function for the two-tailed test (cont.)

$$\pi(\mu^*) = 1 - \Phi\left(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} + z_{1-\alpha/2}\right) + \Phi\left(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} - z_{1-\alpha/2}\right)$$



Comments on the two-tailed tests

two tail test not more powerful than one tail test on their respective domain

- Is the two-tailed test for the Normal mean (known variance) a UMP test? No!
  - The right-tailed test is more powerful for any alternative $\mu^* > \mu_0$
  - The left-tailed test is more powerful for any $\mu^* < \mu_0$
- For large n and small α we have $\Phi\left(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} - z_{1-\alpha/2}\right) \approx 0$, hence

$$\pi(\mu^*) \approx 1 - \Phi\left(-\frac{\sqrt{n}(\mu^* - \mu_0)}{\sigma} + z_{1-\alpha/2}\right),$$

allowing us to calculate the required n for any prespecified α, β and $\delta = \mu_1 - \mu_0$ (see practice problems).

Two-tailed tests and confidence intervals

equivalent

- Recall the two-tailed test for the normal mean with known variance at level α, based on the rejection region

$$\mathcal{C} = \left\{ \overline{X} \leq \mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right\} \bigcup \left\{ \overline{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \right\}.$$

- Note that

We do not reject $\mathcal{H}_0 : \mu = \mu_0$ $\iff \overline{X} \notin \mathcal{C}$

1 - \alpha Confidence interval for normal mean

$$\iff \boxed{\mu_0 - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \overline{X} \leq \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}}$$

$$\iff \overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \mu_0 \leq \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha/2}$$

$\iff$ μ_0 is contained in the $(1-\alpha)100\%$ confidence interval for μ.

Two-tailed tests and confidence intervals (cont.)

$$\begin{array}{c}\text{We do not reject} \\ \mathcal{H}_0 : \mu = \mu_0 \text{ at level } \alpha\end{array} \iff \begin{array}{c}\mu_0 \text{ is contained in the } (1-\alpha)100\% \\ \text{confidence interval for } \mu.\end{array}$$

- This is hardly surprising if we think of a confidence interval as a set of "plausible" values for μ: we reject $\mathcal{H}_0 : \mu = \mu_0 \iff \mu_0$ is not a plausible value for μ.

isnt theta_0 a single number not a set even for composite hypothesis

- In general, the set of all θ_0 for which $\mathcal{H}_0 : \theta = \theta_0$ would not get rejected in a two-tailed set at level α, forms a $(1-\alpha)100\%$ *confidence set* for θ.
- Conversely, the set of all $\underline{X}$'s for which the $(1-\alpha)100\%$ confidence interval for θ based on $\underline{X}$ would not contain θ_0, forms a rejection region of size α for $\mathcal{H}_0 : \theta = \theta_0$ vs. $\mathcal{H}_1 : \theta \neq \theta_0$.
- In other words: every confidence set has a corresponding two-tailed test and vice versa.