Lecture 15: Hypothesis Testing (II)

Ye Tian

Department of Statistics, Columbia University
Calculus-based Introduction to Statistics (S1201)

August 3, 2022

Recap: Hypothesis testing

Statistical hypothesis: A claim or assertion either about the value of a parameter of population.
E.g.: the lifetime of the light bulb > 500 hrs, population mean height of men is equal to 70 inches etc..
Two hypotheses: Usually people decide between two claims. The **null hypothesis**, denoted by H_0, is the claim that is initially assumed to be true. The **alternative hypothesis**, denoted by H_a or H_1, is the assertion that is contradictory to H_0.

- Usually if we want to claim some new findings, we put them as H_a
- H_0 and H_a are NOT symmetric --- we tend to "**protect**" H_0, and require **very strong evidence** to claim H_a is true

A **test (of hypotheses)** is a method for using **sample data** to decide whether the null hypothesis should be rejected. It is a **function of sample data**, whose value indicates the result of hypothesis testing.

Recap: Type-I and Type-II errors

There are two competing hypotheses: the null and the alternative. In a hypothesis test, we make a decision about which might be true, but our choice might be incorrect.

		Decision	
		accept Ho	reject Ho
Truth	Ho true	✓	Type 1 error
	Ha true	Type 2 error	✓

- A Type 1 Error is rejecting the null hypothesis when H_0 is true.
- A Type 2 Error is failing to reject the null hypothesis when H_a is true.
- We never know if H_0 or H_a is true, but we need to consider all possibilities

Recap: Type-I and Type-II error **rates**

We call probability $\mathbb{P}(\text{Type-I error}|H_0 \text{ true}) = \mathbb{P}(\text{reject } H_0|H_0 \text{ true})$ as **Type-I error rate** (of the test), and call $\mathbb{P}(\text{Type-II error}|H_1 \text{ true}) = \mathbb{P}(\text{accept } H_0|H_1 \text{ true})$ as **Type-II error rate** (of the test). And $1-$ **Type-II error** rate is called the **power** (of the test).

- This definition can be problematic. We will discuss more on this.
- Recall that Type-I error is more serious

The ideal scenario: Both Type-I and Type-II errors are small.

But it's IMPOSSIBLE! (We will see this later)

Neyman-Pearson Criterion: We would like to construct a test that:

(1) controls Type-I error rates under some small level α

(2) under (1), minimizes the Type-II error

In this course, we will focus on the goal (1).

Recap: General Procedure

(1) Construct the null and alternative hypotheses

(2) Given the level α, develop a test that satisfies Type-1 error rate $\mathbb{P}(\text{reject } H_0 | H_0 \text{ true}) \le \alpha$

How to do step (2):

- Construct a **test statistic** $T(X_1, \ldots, X_n)$, which is a function of samples $X_1, \ldots, X_n$ and whose distribution under H_0 is **known**
- Find a random region B such that $\mathbb{P}(\text{reject } H_0 | H_0 \text{ true}) = \mathbb{P}(T(X_1, \ldots, X_n) \in B) = \alpha$ (why "$=$" instead of $\le$?)
- Then the rejection region is B, and the corresponding test is

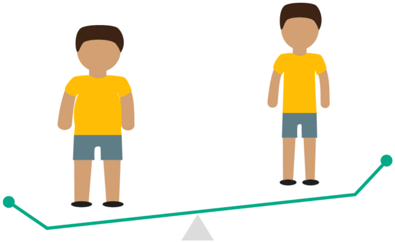
$$\begin{cases} \text{reject } H_0, & \text{if } T(X_1, \ldots, X_n) \in B, \\ \text{accept } H_0, & \text{if } T(X_1, \ldots, X_n) \notin B. \end{cases}$$

Example: revisited

"A common belief among the lay public is that body weight increases after entry into college, and the phrase 'freshman 15' has been coined to describe the 15 pounds that students presumably gain over their freshman year." Suppose everyone's weight is **normally distributed**.

Let μ denote the true average weight gain of students over the course of their first year in college.

$$H_0 : \mu = 15 \quad \text{v.s.} \quad H_a : \mu \neq 15$$



Example: revisited

$$H_0 : \mu = 15 \quad \text{v.s.} \quad H_a : \mu \neq 15$$

- Under H_0, $T = \frac{\sqrt{n}(\bar{X}-15)}{S} \sim t_{n-1}$-distribution would be the test statistic.
- We want to find c such that Type-I error rate $= \mathbb{P}(|T| > c) = 5\%$

It suffices to let $c = t_{n-1,0.025}$.
Therefore, we reject H_0 if $T > t_{n-1,0.025}$ or $T < -t_{n-1,0.025}$. We call $\big(-\infty, -t_{n-1,0.025}\big) \cup \big(t_{n-1,0.025}, +\infty\big)$ the **rejection region**.

- Its Type-I error rate $= 5\%$ (by construction of the test!).
- Given $\mu = \mu_0$, Type-II error rate $= \mathbb{P}(\text{accept } H_0 | H_1 \text{ true}) = \mathbb{P}(|T| \leq t_{n-1,0.025} | H_1 \text{ true}) = \mathbb{P}(|T| \leq t_{n-1,0.025} | \mu = \mu_0)$.
- If the data collected satisfies $\bar{x} = 20$, $s = 3$, $n = 21$, then the realization of T i.e. $t = \frac{\sqrt{n}(\bar{x}-15)}{s} = 3.06 > t_{20,0.025} = 2.086 \Rightarrow$ reject H_0

Example: revisited

$$H_0 : \mu = 15 \quad \text{v.s.} \quad H_a : \mu \neq 15$$

The rejection region: $\big(-\infty, -t_{n-1,0.025} \big) \cup \big(t_{n-1,0.025}, +\infty \big)$.

- If the data collected satisfies $\bar{x} = 20$, $s = 3$, $n = 21$, then the realization of T i.e. $t = \frac{\sqrt{n}(\bar{x}-15)}{s} = 3.06 > t_{20,0.025} = 2.086 \Rightarrow$ reject H_0
- **p-value** is the probability of obtaining a value of the test statistic at least as contradictory to H_0 as the value calculated from the available sample data, when assuming that the null hypothesis is true.
- $T_0 = \frac{\sqrt{n}(\bar{x}-15)}{3} = \frac{\sqrt{21}\times(17-15)}{3} = 3.06$. Then p-value $=$ $\mathbb{P}(|T| > |T_0||\mu = 15) = \mathbb{P}(\underbrace{T}_{\sim t_{20}} > 3.06) + \mathbb{P}(\underbrace{T}_{\sim t_{20}} < -3.06)$

  $= 2 \times 0.003 = 0.006$

Today's goal

- Understand two-sided and one-sided hypothesis testing procedure when the samples are **normally** distributed
  - Test the mean ($H_0 : \mu = \mu_0$ for some constant μ_0)
    - σ known $\Rightarrow$ one-sample z-test
    - σ unknown $\Rightarrow$ one-sample t-test
  - Test the variance/SD ($H_0 : \sigma = \sigma_0$ for some constant σ_0) (You will derive this in HW5)
- Know how to calculate Type-I, Type-II and p-value under the aforementioned scenarios

One-sample z-test

Example: fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the distribution of activation temperature is **normal with standard deviation** 1.5 °F, does the data contradict the manufacturer's claim at significance level $\alpha = 0.05$?

H_0 : average activation temperature $\mu = 130\,^\circ$F   v.s. $H_a : \mu \neq 130\,^\circ$F

Two-sided one-sample z-test: derivation

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with σ^2 **known**.
Our hypothesis testing problem is

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu \neq \mu_0,$$

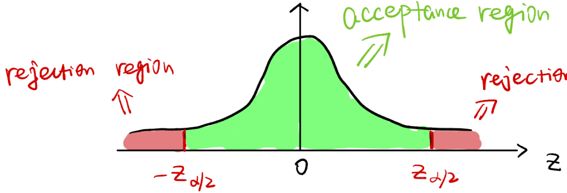
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

Step 1: Under H_0, $\bar{X} \sim N(\mu_0, \sigma^2/n) \Rightarrow Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim N(0,1)$

Step 2: The **larger** $|Z|$ is, more likely H_a is true. Therefore the rejection region would be in the form $|Z| > c$. Note that

$$\mathbb{P}(|Z| > z_{\alpha/2}) = \alpha.$$

Step 3: The rejection region is $|Z| > z_{\alpha/2}$. p-value $= \mathbb{P}(|Z| > |z|)$
$= \mathbb{P}\Big(|Z| > \big|\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}\big|\Big) = 1 - 2\Phi\Big(-\big|\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}\big|\Big)$



Two-sided one-sample z-test

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with σ^2 **known**.
Our hypothesis testing problem is

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu \neq \mu_0,$$

for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

- Test statistic $Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim N(0,1)$
- The test: reject H_0 if $|Z| > z_{\alpha/2}$, otherwise accept H_0. This is called the **two-sided one-sample z-test** with level α.
- p-value $= \mathbb{P}(|Z| > |z|) = 2\Phi\Big(-\Big|\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}\Big|\Big)$
- When $\mu = \mu_1$, $\bar{X} \sim N(\mu_1, \sigma^2/n)$.Type-II error rate

$$= \mathbb{P}(-z_{\alpha/2} \leq Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \leq z_{\alpha/2})$$

$$= \mathbb{P}\big(\frac{\sqrt{n}(\mu_0-\mu_1)}{\sigma} - z_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X}-\mu_1)}{\sigma} \leq \frac{\sqrt{n}(\mu_0-\mu_1)}{\sigma} + z_{\alpha/2}\big)$$

$$= \Phi\Big(\frac{\sqrt{n}(\mu_0-\mu_1)}{\sigma} + z_{\alpha/2}\Big) - \Phi\Big(\frac{\sqrt{n}(\mu_0-\mu_1)}{\sigma} - z_{\alpha/2}\Big)$$

Example: fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the distribution of activation temperature is **normal with standard deviation** 1.5 °F, does the data contradict the manufacturer's claim at significance level $\alpha = 0.05$?

$$H_0 : \text{average activation temperature } \mu = 130 \quad \text{v.s.} \quad H_a : \mu \neq 130$$

- The test: reject H_0 if $|Z| > z_{\alpha/2} = z_{0.025} = 1.96$, otherwise accept H_0
- Sample z-statistic we observed $z = \frac{\sqrt{9}(131.08-130)}{1.5} = 2.16 > 1.96 \Rightarrow$ reject H_0, i.e. under 5% significance level, the data contradicts the manufacturer's claim
- Test statistic $Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim N(0, 1)$
- p-value $= \mathbb{P}(|Z| > |z|) = \mathbb{P}(|Z| > 2.16) = 2\Phi(-2.16) = 0.0308$ $< 0.05 \Rightarrow$ reject H_0

Example: fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the distribution of activation times is **normal with standard deviation** 1.5 °F, does the data contradict the manufacturer's claim at significance level $\alpha = 0.05$?

$$H_0 : \text{average activation temperature } \mu = 130 \quad \text{v.s.} \quad H_a : \mu \neq 130$$

- The test: reject H_0 if $|Z| > z_{\alpha/2} = z_{0.025} = 1.96$, otherwise accept H_0
- When $\mu = 131$ (H_a true), we'd make Type-II error if accept H_0 ($-1.96 \leq Z \leq 1.96$).

  Type-II error rate $= \mathbb{P}(-1.96 \leq Z = \frac{\sqrt{9}(\bar{X}-130)}{1.5} \leq 1.96)$

  $$= \mathbb{P}\Big(-1.96 - \tfrac{\sqrt{9}}{1.5} \leq \underbrace{\frac{\sqrt{9}(\bar{X} - 131)}{1.5}}_{\sim N(0,1)} \leq 1.96 - \tfrac{\sqrt{9}}{1.5} \Big)$$

  $$= \Phi(-0.04) - \Phi(-3.96) \approx 0.4840$$

Example: fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the distribution of activation temperature is **normal with standard deviation** 1.5 °F, does the data contradict the manufacturer's claim at significance level $\alpha = 0.05$?

$$H_0 : \text{average activation temperature } \mu = 130 \quad \text{v.s.} \quad H_a : \mu \neq 130$$

- The test: reject H_0 if $|Z| > z_{\alpha/2} = z_{0.025} = 1.96$, otherwise accept H_0
- When $\mu = 131$ (H_a true), Type-II error rate of the test $= 0.4840$.
- But based on current data, we DO NOT make a Type-II error! (because we reject H_0)
- Type-II error **rate** is the long-run probability that we make a Type-II error if we do the test with independent data many times

Example (revisited): fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature μ is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the distribution of activation temperature is **normal with standard deviation** 1.5 °F, does the data suggest that $\mu > 130$ at significance level $\alpha = 0.05$?

H_0 : average activation temperature $\mu = 130\,^\circ$F v.s. $H_a : \mu > 130\,^\circ$F

One-sided one-sample z-test: derivation

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with σ^2 **known**.
Our hypothesis testing problem is

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu > \mu_0,$$

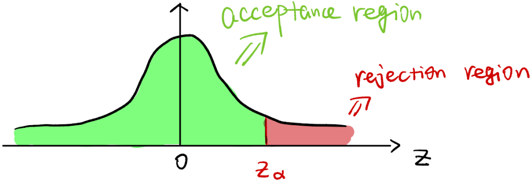
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

Step 1: Under H_0, $\bar{X} \sim N(\mu_0, \sigma^2/n) \Rightarrow Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0,1)$

Step 2: The **larger** $|Z|$ is, more likely H_a is true.Therefore the rejection region would bein the form $Z > c$. Note that

$$\mathbb{P}(Z > z_\alpha) = \alpha.$$

Step 3: The rejection region is $Z > z_\alpha$. p-value $= \mathbb{P}(Z > z)$
$= \mathbb{P}\left(Z > \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}\right)$



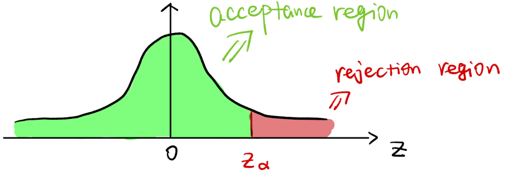
One-sided one-sample z-test

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with σ^2 **known**.
Our hypothesis testing problem is

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu > \mu_0,$$

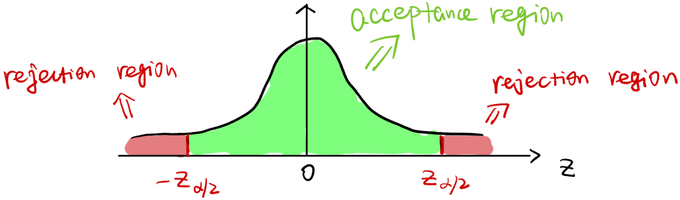
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

- Test statistic $Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1)$
- The test: reject H_0 if $Z > z_\alpha$, otherwise accept H_0. This is called the **one-sided one-sample z-test** with level α.
- p-value $= \mathbb{P}(Z > z) = 1 - \Phi\left(\frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma}\right)$
- When $\mu = \mu_1$, $\bar{X} \sim N(\mu_1, \sigma^2/n)$. Type-II error rate
  $= \mathbb{P}(Z = \frac{\sqrt{n}(\bar{X} - \mu_1)}{\sigma} \leq z_\alpha) = \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu_1)}{\sigma} \leq \frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + z_\alpha\right)$
  $= \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + z_\alpha\right)$

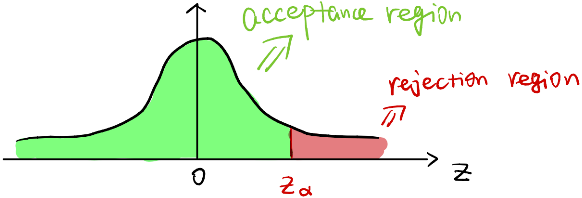


Compare two-sided and one-sided z-tests

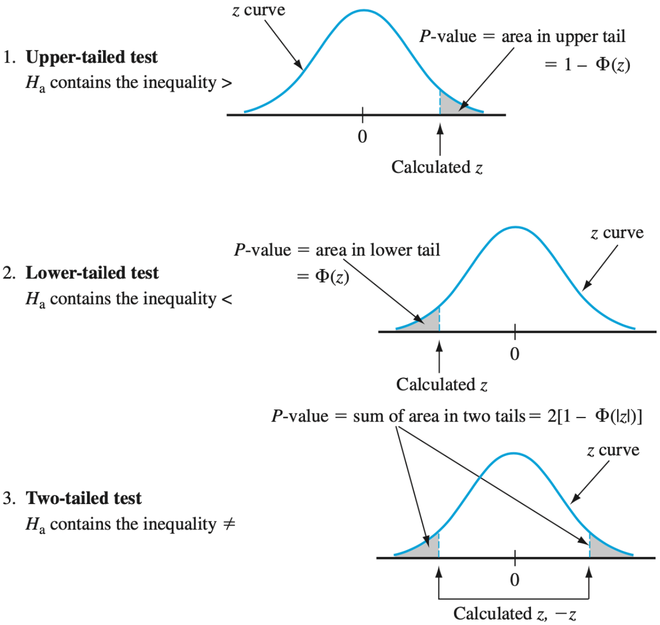
Two-sided z-test: H_a includes both directions ($\mu \neq \mu_0$)



One-sided z-test: H_a includes only one direction ($\mu > \mu_0$ or $\mu < \mu_0$)



p-value in two-sided and one-sided z-tests



Example (revisited): fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the distribution of activation temperature is **normal with standard deviation** 1.5 °F, does the data suggest that $\mu > 130$ at significance level $\alpha = 0.05$?

$$H_0 : \text{average activation temperature } \mu = 130 \quad \text{v.s.} \quad H_a : \mu > 130$$

- The test: reject H_0 if $Z > z_{0.05} = 1.645$, otherwise accept H_0
- Sample z-statistic we observed $z = 2.16 > 1.645 \Rightarrow$ reject H_0, i.e. under 5% significance level, the data contradicts the manufacturer's claim
- Test statistic $Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0, 1)$
- p-value $= \mathbb{P}(Z > z) = \mathbb{P}(Z > 2.16) = 1 - \Phi(2.16) = 0.0154 < 0.05 \Rightarrow$ reject H_0

Example (revisited): fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the distribution of activation temperature is **normal with standard deviation** 1.5 °F, does the data suggest that $\mu > 130$ at significance level $\alpha = 0.05$?

$$H_0 : \text{average activation temperature } \mu = 130 \quad \text{v.s.} \quad H_a : \mu > 130$$

- The test: reject H_0 if $Z > z_{0.05} = 1.645$, otherwise accept H_0
- When $\mu = 131$ (H_a true), we'd make Type-II error if accept H_0 ($Z \le 1.645$).

  Type-II error rate $= \mathbb{P}\Big(Z = \frac{\sqrt{9}(\bar{X}-130)}{1.5} \le 1.645\Big)$

  $= \mathbb{P}\Big(\underbrace{\frac{\sqrt{9}(\bar{X}-131)}{1.5}}_{\sim N(0,1)} \le 1.645 - \frac{\sqrt{9}}{1.5}\Big)$

  $= \Phi(-0.36) \approx 0.3594$

One-sample t-test

Example (again): fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F and **sample standard deviation** 1.21 °F. If the distribution of activation temperature is **normal**, does the data contradict the manufacturer's claim at significance level $\alpha = 0.05$?

H_0 : average activation temperature $\mu = 130\,^{\circ}$F v.s. $H_a : \mu \neq 130\,^{\circ}$F

Two-sided one-sample t-test: derivation

Suppose we have data $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with σ^2 **unknown**.
Our hypothesis testing problem is

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu \neq \mu_0,$$

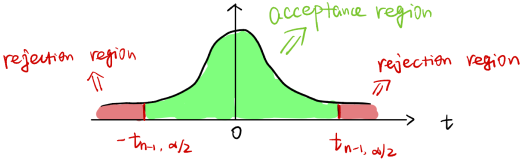
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

Step 1: Under H_0, $T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \sim t_{n-1}$

Step 2: The **larger** $|T|$ is, more likely H_a is true. Therefore the rejection region would be in the form $|T| > c$. Note that

$$\mathbb{P}(|S| > t_{n-1,\alpha/2}) = \alpha.$$

Step 3: The rejection region is $|T| > t_{n-1,\alpha/2}$, i.e. $\bar{X} < \mu_0 - \frac{S}{\sqrt{n}} t_{n-1,\alpha/2}$ or $\bar{X} > \mu_0 + \frac{S}{\sqrt{n}} t_{n-1,\alpha/2}$. p-value $= \mathbb{P}(|T| > |t|) = \mathbb{P}\Big(|T| > \Big|\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\Big|\Big)$ $= 2F\Big(-\Big|\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\Big|\Big)$, where F is the cdf of t_{n-1}-distribution



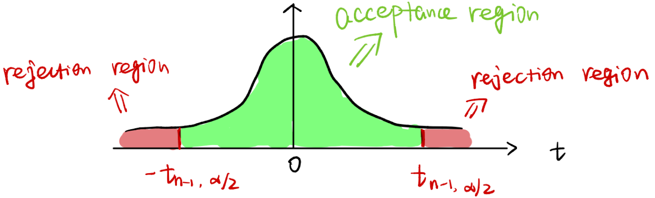
Two-sided one-sample t-test

Suppose we have data $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with σ^2 **unknown**.
Our hypothesis testing problem is

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu \neq \mu_0,$$

for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

- Test statistic $T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \sim t_{n-1}$
- The test: reject H_0 if $T < -t_{n-1,\alpha/2}$ or $T > t_{n-1,\alpha/2}$, otherwise accept H_0. This is called the **two-sided one-sample z-test** with level α.
- p-value $= \mathbb{P}(|T| > |t|) = 2F\left(-\left|\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\right|\right)$, F : cdf of t_{n-1}-distribution
- Type-II error rate is a bit harder to calculate in this case so we won't discuss it



Example: fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F and **sample standard deviation** 1.21 °F. If the distribution of activation temperature is **normal**, does the data contradict the manufacturer's claim at significance level $\alpha = 0.05$?

$$H_0 : \text{average activation temperature } \mu = 130 \quad \text{v.s.} \quad H_a : \mu \neq 130$$

- The test: reject H_0 if $|T| > t_{8,0.025} = 2.306$, otherwise accept H_0
- Sample test statistic we observed $t = \frac{\sqrt{n}(\bar{x}-130)}{s} = \frac{3\times(131.08-130)}{1.21}$ $= 2.678 > 2.306 \Rightarrow$ reject H_0, i.e. under 5% significance level, the data contradicts the manufacturer's claim
- Test statistic $T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \sim t_{n-1} = t_8$
- p-value $= \mathbb{P}(|T| > |t|) = \mathbb{P}(|T| > 2.678) = 2F(-2.678) = 0.028$ $< 0.05 \Rightarrow$ reject H_0

Example (again): fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature μ is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F and **sample standard deviation** 1.21 °F. If the distribution of activation temperature is **normal**, does the data suggest that $\mu > 130$ at significance level $\alpha = 0.05$?

H_0 : average activation temperature $\mu = 130\,^{\circ}$F v.s. $H_a : \mu > 130\,^{\circ}$F

One-sided one-sample t-test: derivation

Suppose we have data $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with σ^2 **unknown**.
Our hypothesis testing problem is

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu > \mu_0,$$

for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

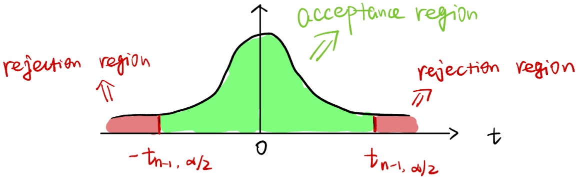
Step 1: Under H_0, $T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \sim t_{n-1}$

Step 2: The **larger** T is, more likely H_a is true. Therefore the rejection region would be in the form $T > c$. Note that

$$\mathbb{P}(|S| > t_{n-1,\alpha/2}) = \alpha.$$

Step 3: The rejection region is $T > t_{n-1,\alpha}$, i.e. $\bar{X} > \mu_0 + \frac{S}{\sqrt{n}} t_{n-1,\alpha}$.

p-value $= \mathbb{P}(T > t) = \mathbb{P}\left(T > \frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\right) = 1 - F\left(\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\right)$, where F is the cdf of t_{n-1}-distribution



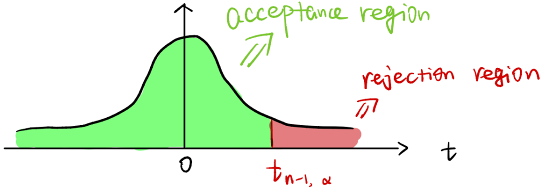
One-sided one-sample t-test

Suppose we have data $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with σ^2 **unknown**.
Our hypothesis testing problem is

$$H_0 : \mu = \mu_0 \quad \text{v.s.} \quad H_a : \mu > \mu_0,$$

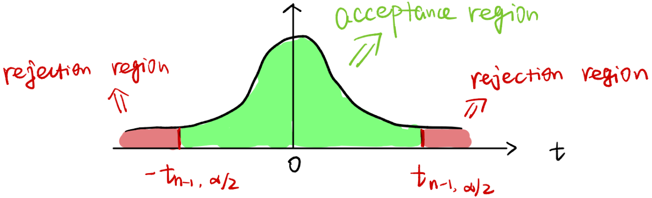
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

- Test statistic $T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim t_{n-1}$
- The test: reject H_0 if $T > t_{n-1,\alpha}$, otherwise accept H_0. This is called the **one-sided one-sample z-test** with level α.
- p-value $= \mathbb{P}(T > t) = 1 - F\left(\frac{\sqrt{n}(\bar{x} - \mu_0)}{s}\right)$, F : cdf of t_{n-1}-distribution
- Type-II error rate is a bit harder to calculate in this case so we won't discuss it

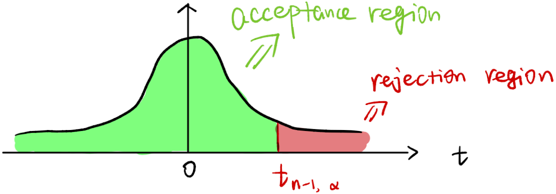


Compare two-sided and one-sided t-tests

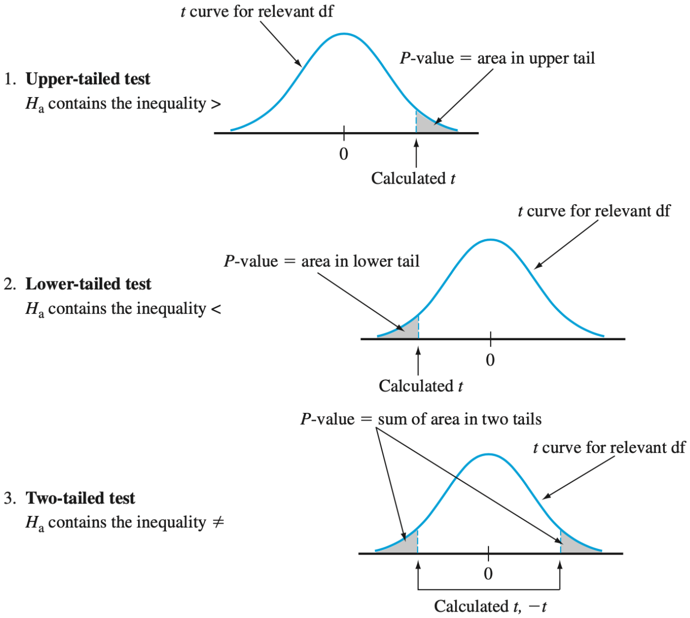
Two-sided t-test: H_a includes both directions ($\mu \neq \mu_0$)



One-sided t-test: H_a includes only one direction ($\mu > \mu_0$ or $\mu < \mu_0$)



p-value in two-sided and one-sided t-tests



Example (finally!): fire protection system

A manufacturer of sprinkler systems used for fire protection in office buildings claims that the true average system-activation temperature is 130 °F. A sample of $n = 9$ systems, when tested, yields a sample average activation temperature of 131.08 °F and **sample standard deviation** 1.21 °F. If the distribution of activation temperature is **normal**, does the data suggest that $\mu > 130$ at significance level $\alpha = 0.05$?

H_0 : average activation temperature $\mu = 130$ v.s. $H_a : \mu > 130$

- The test: reject H_0 if $T > t_{8,0.05} = 1.860$, otherwise accept H_0
- Sample test statistic we observed $t = \frac{\sqrt{n}(\bar{x}-130)}{s} = \frac{3\times(131.08-130)}{1.21}$ $= 2.678 > 1.860 \Rightarrow$ reject H_0, i.e. under 5% significance level, the data contradicts the manufacturer's claim
- Test statistic $T = \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \sim t_{n-1} = t_8$
- p-value $= \mathbb{P}(T > t) = \mathbb{P}(T > 2.678) = 1 - F(2.678) = 0.014 < 0.05 \Rightarrow$ reject H_0

More about p-values

- p-value is a **probability**
- We reject the null when p-value $<$ significance level α, and accept the null when p-value $\geq$ significance level α. This is **equivalent** to the test we came up with before by calculating the rejection region.
- For one-sided test like $H_0 : \mu = \mu_0$ v.s. $H_a : \mu > \mu_0$, the p-value usually looks like $\mathbb{P}(T > t)$, where T is the test statistic and T_0 is the current value of T calculated from the available sample
- For one-sided test like $H_0 : \mu = \mu_0$ v.s. $H_a : \mu < \mu_0$, the p-value usually looks like $\mathbb{P}(T < t)$
- For two-sided test like $H_0 : \mu = \mu_0$ v.s. $H_a : \mu \neq 15$, the p-value usually looks like $\mathbb{P}(|T| > |t|)$
- The smaller p-value we have, more evidence we have against H_0
- p-value is **NOT** the probability that H_0 is true!!! (H_0 can only be true or false, so it's deterministic, although we don't know.)

Many thanks to

- Chengliang Tang
- Joyce Robbins
- Yang Feng
- Owen Ward
- Wenda Zhou
- And all my teachers in the past 25 years