Lecture 16: Hypothesis Testing (III)

Ye Tian

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

August 2, 2022

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

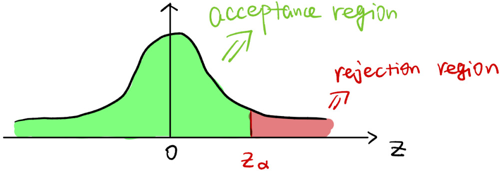
One-sided one-sample z-test

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

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

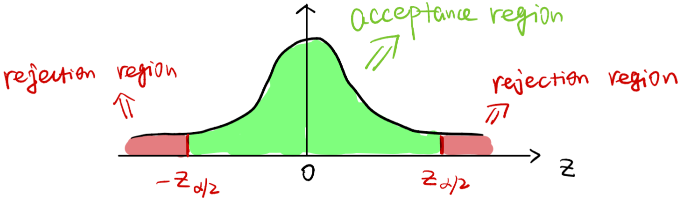
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

- 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 $= \mathbb{P}(Z = \frac{\sqrt{n}(\bar{X} - \mu_1)}{\sigma} \leq z_\alpha) = \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu_1)}{\sigma} \leq \frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + z_\alpha\right)$
$= \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + z_\alpha\right)$

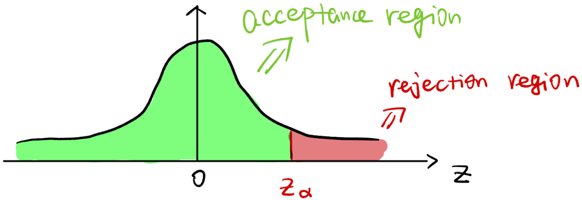


Compare two-sided and one-sided z-tests

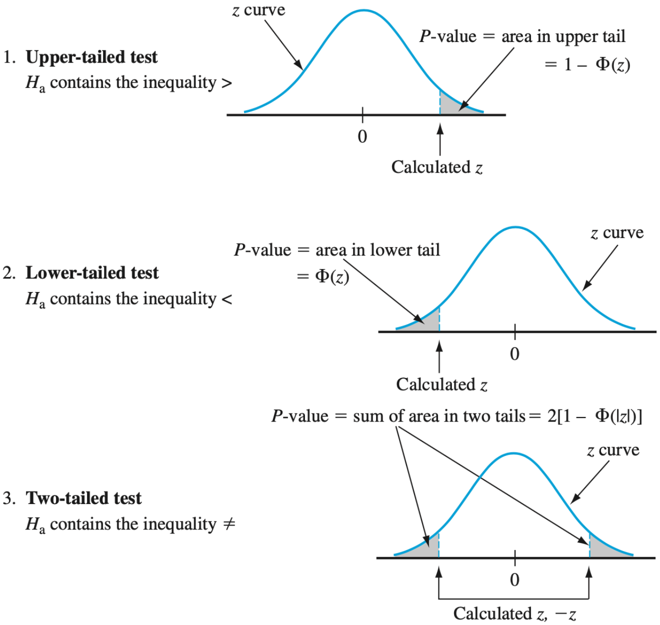
Two-sided z-test: H_a includes both directions ($\mu \neq \mu_0$)



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



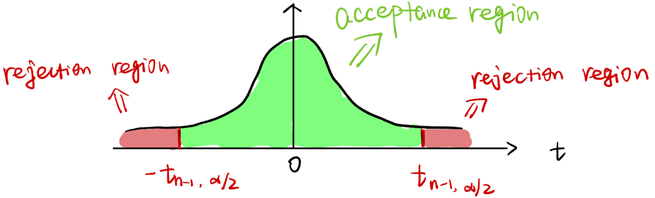
Two-sided one-sample t-test

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ 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



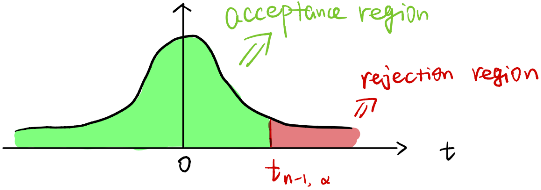
One-sided one-sample t-test

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

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

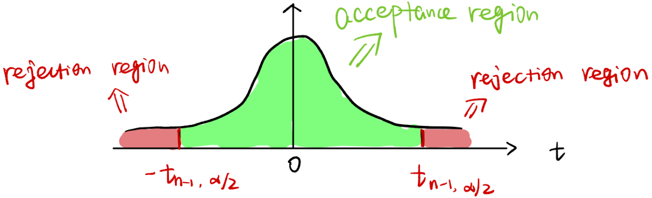
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

- 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

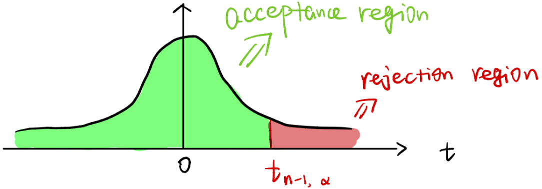


Compare two-sided and one-sided t-tests

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

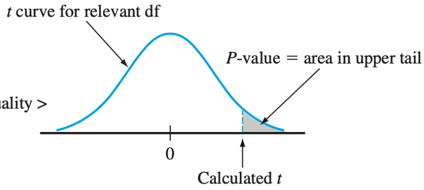


One-sided t-test: H_a includes only one direction ($\mu > \mu_0$ or $\mu < \mu_0$)

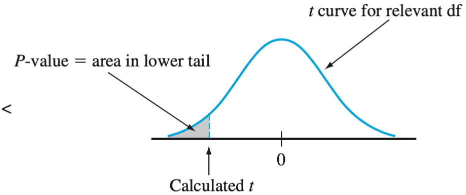


p-value in two-sided and one-sided t-tests

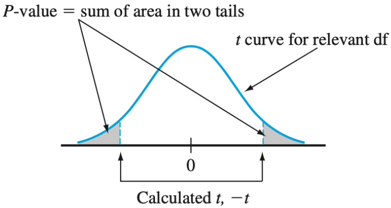
1. **Upper-tailed test**
   H_a contains the inequality $>$



2. **Lower-tailed test**
   H_a contains the inequality $<$



3. **Two-tailed test**
   H_a contains the inequality $\neq$



Summary of testing mean of normal distribution

- σ known, $H_0 : \mu = \mu_0$
  - $H_a : \mu \neq \mu_0 \Rightarrow$ two-sided z-test
  - $H_a : \mu > \mu_0$ (or $\mu < \mu_0$) $\Rightarrow$ one-sided z-test
- σ unknown, $H_0 : \mu = \mu_0$
  - $H_a : \mu \neq \mu_0 \Rightarrow$ two-sided t-test
  - $H_a : \mu > \mu_0$ (or $\mu < \mu_0$) $\Rightarrow$ one-sided t-test

Today's goal

- Know how to use z-test to test the mean of general distributions (valid only when n is very large, hence called "large-sample z-test")
- Understand two-sided and one-sided hypothesis testing procedure for the population proportion p (i.e. the parameter p in Bernoulli(p) or Bin(n, p), valid only when n is very large, also called "large-sample z-test")
- Know how to calculate Type-I, Type-II and p-value under the aforementioned scenarios

Large-sample z-test for Mean of General Distributions

Example (the last time): 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 $n = 90$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the sample standard deviation of activation times is 2.25 °F **[with NO normal distribution!]**, 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

Recall the general procedure to construct a test

(1) Construct the null and alternative hypotheses. ✓

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

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** **✗ We don't know the distribution! --- But we may use CLT!**
- Find a random region B such that Type-1 error rate $\mathbb{P}(\text{reject } H_0 | H_0 \text{ true}) = \mathbb{P}(T(X_1, \ldots, X_n) \in B) = \alpha$
- Then the rejection region is B, and the corresponding test is

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

Two-sided large-scale z-test: derivation

Suppose we have data $X_1, \dots, X_n \overset{i.i.d.}{\sim}$ **some unknown distribution** with both mean μ and standard deviation 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, by CLT and Law of Large Numbers,
$\bar{X} \overset{d}{\approx} N(\mu_0, \sigma^2/n)$, $S \approx \sigma^2 \Rightarrow Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} \approx \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \overset{d}{\approx} N(0,1)$
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

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

Step 3: The rejection region is $|Z| > z_{\alpha/2}$, i.e. $\bar{X} < \mu_0 - \frac{S}{\sqrt{n}} z_{\alpha/2}$ or $\bar{X} > \mu_0 + \frac{S}{\sqrt{n}} z_{\alpha/2}$. p-value $= \mathbb{P}(|Z| > |z|) = \mathbb{P}\Big(|Z| > \big|\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\big|\Big)$
$\approx 2\Phi\Big(-\big|\frac{\sqrt{n}(\bar{x}-\mu_0)}{s}\big|\Big)$

Two-sided large-scale z-test

Suppose we have data $X_1, \dots, X_n \overset{i.i.d.}{\sim}$ **some unknown distribution** with both mean μ and standard deviation 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 $Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \overset{d}{\approx} N(0, 1)$
- The test: reject H_0 if $|Z| > z_{\alpha/2}$, otherwise accept H_0. This is called the **two-sided large-sample (one-sample) z-test** with level α.
- p-value $= \mathbb{P}(|Z| > |z|) \approx 2\Phi\Big(- \Big|\frac{\sqrt{n}(\bar{x} - \mu_0)}{s}\Big|\Big)$
- When $\mu = \mu_1$, $\bar{X} \overset{d}{\approx} N(\mu_1, \sigma^2/n) \Rightarrow Z' = \frac{\sqrt{n}(\bar{X} - \mu_1)}{S} \overset{d}{\approx} N(0, 1)$. Type-II error rate $= \mathbb{P}(-z_{\alpha/2} \leq Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \leq z_{\alpha/2})$
  $\approx \mathbb{P}\big(\frac{\sqrt{n}(\mu_0 - \mu_1)}{S} - z_{\alpha/2} \leq \frac{\sqrt{n}(\bar{X} - \mu_1)}{S} \leq \frac{\sqrt{n}(\mu_0 - \mu_1)}{S} + z_{\alpha/2}\big)$
  $\approx \Phi\Big(\frac{\sqrt{n}(\mu_0 - \mu_1)}{S} + z_{\alpha/2}\Big) - \Phi\Big(\frac{\sqrt{n}(\mu_0 - \mu_1)}{S} - z_{\alpha/2}\Big)$

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 $n = 90$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the sample standard deviation of activation times is 2.25 °F **[with NO normal distribution!]**, 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 $|Z| = \left|\frac{\sqrt{n}(\bar{X}-130)}{S}\right| > z_{0.025} = 1.96$, otherwise accept H_0
- Sample z-statistic we observed $z = \frac{\sqrt{9}(131.08-130)}{2.25} = 1.44 < 1.96 \Rightarrow$ accept H_0, i.e. under 5% significance level, the data does not contradict the manufacturer's claim
- Test statistic $Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \sim N(0,1)$
- p-value $= \mathbb{P}(|Z| > |z|) = \mathbb{P}(|Z| > 1.44) = 2\mathbb{P}(Z > 1.44) = 0.1498$ $> 0.05 \Rightarrow$ accept 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 $n = 90$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the sample standard deviation of activation times is 2.25 °F **[with NO normal distribution!]**, does the data contradict the manufacturer's claim at significance level $\alpha = 0.05$?

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

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

Type-II error rate

$$= \mathbb{P}\Big(-1.96 \leq \frac{\sqrt{9}(\bar{X} - 130)}{2.25} \leq 1.96 \Big)$$

$$= \mathbb{P}\Big(-1.96 - \frac{\sqrt{9}}{2.25} \leq \underbrace{\frac{\sqrt{9}(\bar{X} - 131)}{2.25}}_{\sim N(0,1)} \leq 1.96 - \frac{\sqrt{9}}{2.25} \Big)$$

$$= \Phi(0.63) - \Phi(-3.29) \approx 0.7352$$

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 $n = 90$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the sample standard deviation of activation times is 2.25 °F **[with NO normal distribution!]**, 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 large-scale z-test: derivation

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ **some unknown distribution** with both mean μ and standard deviation σ^2 **unknown**.

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

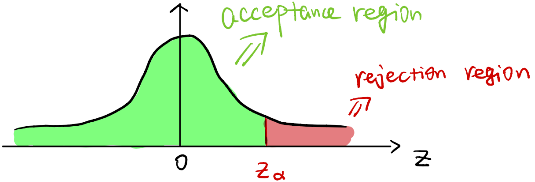
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

Step 1: Under H_0, by CLT and Law of Large Numbers, $Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \sim N(0, 1)$

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

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

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



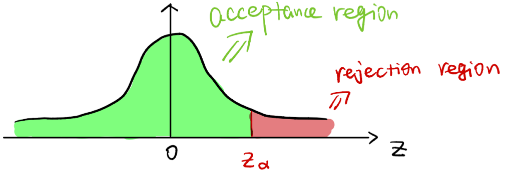
One-sided one-sample z-test

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

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

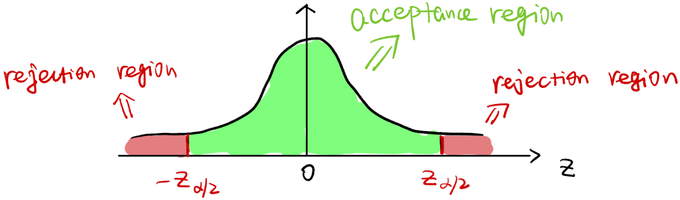
for some constant μ_0. We want to control Type-I error rate $\leq \alpha$.

- 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 large-sample (one-sample) z-test** with level α.
- p-value $= \mathbb{P}(Z > z) = 1 - \Phi\left(\frac{\sqrt{n}(\bar{x} - \mu_0)}{s}\right)$
- When $\mu = \mu_1$, $\bar{X} \sim N(\mu_1, \sigma^2/n)$. Type-II error rate
  $= \mathbb{P}(Z = \frac{\sqrt{n}(\bar{X} - \mu_1)}{S} \leq z_\alpha) = \mathbb{P}\left(\frac{\sqrt{n}(\bar{X} - \mu_1)}{S} \leq \frac{\sqrt{n}(\mu_0 - \mu_1)}{S} + z_\alpha\right)$
  $= \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{S} + z_\alpha\right)$

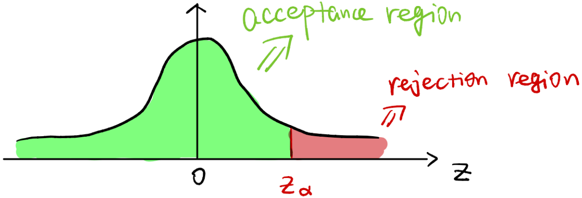


Compare two-sided and one-sided z-tests

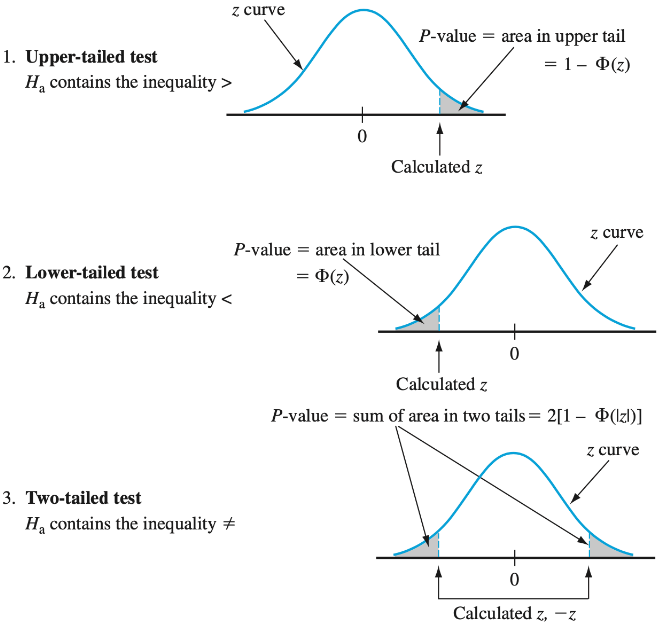
Two-sided z-test: H_a includes both directions ($\mu \neq \mu_0$)



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 $n = 90$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the sample standard deviation of activation times is 2.25 °F **[with NO normal distribution!]**, 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 $Z > z_{0.05} = 1.645$, otherwise accept H_0
- Sample z-statistic we observed $z = \frac{\sqrt{9}(131.08-130)}{2.25} = 1.44 < 1.645 \Rightarrow$ 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 $= \mathbb{P}(Z > z) = \mathbb{P}(Z > 1.44) = 0.0749 > 0.05 \Rightarrow$ accept 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 $n = 90$ systems, when tested, yields a sample average activation temperature of 131.08 °F. If the sample standard deviation of activation times is 2.25 °F **[with NO normal distribution!]**, does the data suggest that $\mu > 130$ at significance level $\alpha = 0.05$?

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

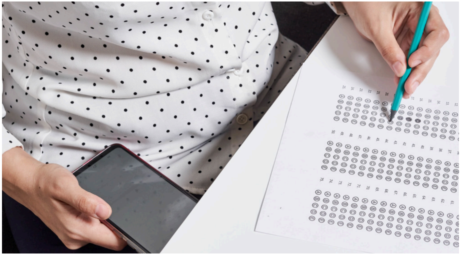
- The test: reject H_0 if $Z > z_{0.05} = 1.645$, otherwise accept H_0
- When $\mu = 131$ (H_a true), we'd make Type-II error if accept H_0 ($\bar{X} \leq 130.98$).

$$\begin{aligned}
\text{Type-II error rate} &= \mathbb{P}\left(Z = \frac{\sqrt{9}(\bar{X} - 130)}{2.25} \leq 1.645\right) \\
&= \mathbb{P}\Big(\underbrace{\frac{\sqrt{9}(\bar{X} - 131)}{2.25}}_{\sim N(0,1)} \leq 1.645 - \frac{\sqrt{9}}{2.25}\Big) \\
&= \Phi(0.31) \approx 0.3783
\end{aligned}$$

Large-sample z-test for Population Proportion

Example (new!): proportion of cheating students

27 of the 267 students in a sample admitted to texting during an exam. Can it be concluded at significance level 0.01 that the proportion of students that texted during the exam is significantly different from 5%?



H_0 : the proportion of cheating students $p = 0.05$ v.s. $H_a : p \neq 0.05$

Two-sided large-sample z-test for proportion: derivation

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ Bernoulli(p).

$$H_0 : p = p_0 \quad \text{v.s.} \quad H_a : p \neq p_0,$$

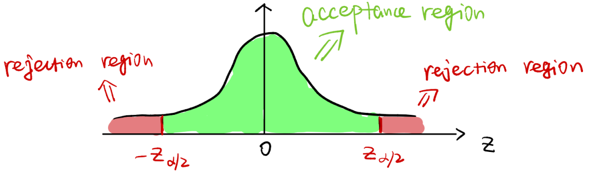
for some constant p_0. We want to control Type-I error rate $\leq \alpha$.

Step 1: Under H_0, by CLT, $\bar{p} = \bar{X} \overset{d}{\approx} N(p_0, p_0(1-p_0)/n) \Rightarrow$
$Z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \overset{d}{\approx} N(0,1)$

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

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

Step 3: The rejection region is $|Z| > z_{\alpha/2}$. p-value $= \mathbb{P}(|Z| > |z|)$
$= \mathbb{P}\Big(|Z| > \big|\frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}\big|\Big) = 2\Phi\Big(-\big|\frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}\big|\Big)$



Two-sided large-sample z-test for proportion

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ Bernoulli(p).

$$H_0 : p = p_0 \quad \text{v.s.} \quad H_a : p \neq p_0,$$

for some constant p_0. We want to control Type-I error rate $\leq \alpha$.

- Test statistic $Z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \overset{d}{\approx} N(0,1)$
- The test: reject H_0 if $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$, otherwise accept H_0. This is called the **two-sided large-sample z-test for proportion** with level α.
- p-value $= \mathbb{P}(|Z| > |z|) = 2\Phi\Big(-\Big|\frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}\Big|\Big)$
- When $p = p_1$, $\hat{p} \overset{d}{\approx} N(p_1, p_1(1-p_1)/n) \Rightarrow \frac{\sqrt{n}(\bar{X}-\mu_1)}{S} \overset{d}{\approx} N(0,1)$.
  Type-II error rate $= \mathbb{P}(-z_{\alpha/2} \leq Z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \leq z_{\alpha/2}) \approx$
  $\mathbb{P}\left(\frac{\sqrt{n}(p_0-p_1)}{\sqrt{p_1(1-p_1)}} - \frac{\sqrt{p_0(1-p_0)}}{\sqrt{p_1(1-p_1)}} z_{\alpha/2} \leq \frac{\sqrt{n}(\hat{p}-p_1)}{\sqrt{p_1(1-p_1)}} \leq \frac{\sqrt{n}(p_0-p_1)}{\sqrt{p_1(1-p_1)}} + \frac{\sqrt{p_0(1-p_0)}}{\sqrt{p_1(1-p_1)}} z_{\alpha/2}\right)$
  $\approx \Phi\Big(\frac{\sqrt{n}(p_0-p_1)}{\sqrt{p_1(1-p_1)}} + \frac{\sqrt{p_0(1-p_0)}}{\sqrt{p_1(1-p_1)}} z_{\alpha/2}\Big) - \Phi\Big(\frac{\sqrt{n}(p_0-p_1)}{\sqrt{p_1(1-p_1)}} - \frac{\sqrt{p_0(1-p_0)}}{\sqrt{p_1(1-p_1)}} z_{\alpha/2}\Big)$

Example: proportion of cheating students

27 of the 267 students in a sample admitted to texting during an exam. Can it be concluded at significance level 0.01 that the proportion of students that texted during the exam is significantly different from 5%?

H_0 : the proportion of cheating students $p = 0.05$ v.s. $H_a : p \neq 0.05$

- The test: reject H_0 if $|Z| > z_{0.005} = 2.58$, otherwise accept H_0
- Sample proportion $\hat{p} = \frac{27}{267} = 0.10$, sample test statistic we observed $z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} = \frac{\sqrt{267}\times(0.10-0.05)}{\sqrt{0.05\times0.95}} = 3.75 > 2.58 \Rightarrow$ reject H_0, i.e. under 5% significance level, there's sufficient evidence that the proportion significantly deviates from 5%
- Under H_0, test statistic $Z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \overset{d}{\approx} N(0,1)$
- p-value $= \mathbb{P}(|Z| > |z|) = \mathbb{P}(|Z| > 3.75) = 2\Phi(-3.75) \approx 0 < 0.01 \Rightarrow$ reject H_0

Example: proportion of cheating students

27 of the 267 students in a sample admitted to texting during an exam. Can it be concluded at significance level 0.01 that the proportion of students that texted during the exam is significantly different from 5%?

H_0 : the proportion of cheating students $p = 0.05$ v.s. $H_a : p \neq 0.05$

- The test: reject H_0 if $|Z| > 2.58$, otherwise accept H_0
- When $p = 0.03$ (H_a true), we'd make Type-II error if accept H_0 ($|Z| \leq 2.58$).

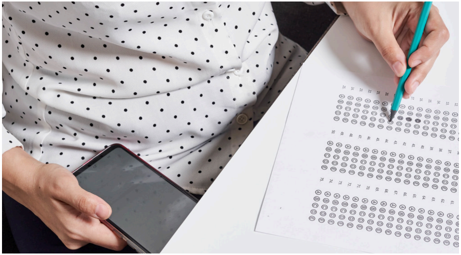
  Type-II error rate $= \mathbb{P}\Big(-2.58 \leq Z = \frac{\sqrt{267}(\hat{p}-0.05)}{\sqrt{0.05\times 0.95}} \leq 2.58 \Big)$

  $$= \mathbb{P}\Big(-2.58\frac{\sqrt{0.05\times0.95}}{\sqrt{0.03\times0.97}} + \frac{\sqrt{267}\times0.02}{\sqrt{0.03\times0.97}} \leq \underbrace{\frac{\sqrt{267}(\hat{p}-0.03)}{\sqrt{0.03\times0.97}}}_{\overset{d}{\approx}N(0,1)} \leq 2.58\frac{\sqrt{0.05\times0.95}}{\sqrt{0.03\times0.97}} + \frac{\sqrt{267}\times0.02}{\sqrt{0.03\times0.97}} \Big)$$

  $\approx \Phi(5.21) - \Phi(-1.38) = 0.9172$

Example (revisited): proportion of cheating students

27 of the 267 students in a sample admitted to texting during an exam. Can it be concluded at significance level 0.01 that more than 5% of all students in the population sampled had texted during an exam?



H_0 : the proportion of cheating students $p = 0.05$ v.s. $H_a : p > 0.05$

One-sided large-sample z-test for proportion: derivation

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ Bernoulli(p).

$$H_0 : p = p_0 \quad \text{v.s.} \quad H_a : p > p_0,$$

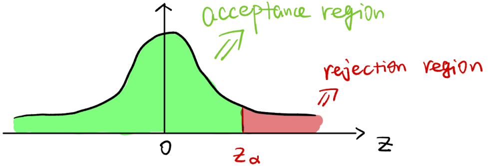
for some constant p_0. We want to control Type-I error rate $\leq \alpha$.

Step 1: Under H_0, by CLT, $\bar{p} = \bar{X} \overset{d}{\approx} N(p_0, p_0(1-p_0)/n) \Rightarrow$
$Z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \overset{d}{\approx} N(0,1)$

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

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

Step 3: The rejection region is $Z > z_\alpha$. p-value $= \mathbb{P}(Z > z)$
$= \mathbb{P}\left(Z > \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}\right) \approx 1 - \Phi\left(\frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}\right)$



One-sided large-sample z-test for proportion

Suppose we have data $X_1, \ldots, X_n \overset{i.i.d.}{\sim}$ Bernoulli(p).

$$H_0 : p = p_0 \quad \text{v.s.} \quad H_a : p > p_0,$$

for some constant p_0. We want to control Type-I error rate $\leq \alpha$.

- Test statistic $Z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \overset{d}{\approx} N(0,1)$
- The test: reject H_0 if $Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$, otherwise accept H_0. This is called the **two-sided large-sample z-test for proportion** with level α.
- p-value $= \mathbb{P}(Z > z) = 1 - \Phi\Big(-\frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}\Big)$
- When $p = p_1$, $\hat{p} \overset{d}{\approx} N(p_1, p_1(1-p_1)/n) \Rightarrow \frac{\sqrt{n}(\bar{X}-\mu_1)}{S} \overset{d}{\approx} N(0,1)$.
  Type-II error rate $= \mathbb{P}\Big(Z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \leq z_{\alpha/2}\Big)$
  $\approx \mathbb{P}\left(\frac{\sqrt{n}(\hat{p}-p_1)}{\sqrt{p_1(1-p_1)}} \leq \frac{\sqrt{n}(p_0-p_1)}{\sqrt{p_1(1-p_1)}} + \frac{\sqrt{p_0(1-p_0)}}{\sqrt{p_1(1-p_1)}} z_{\alpha/2}\right)$
  $\approx \Phi\Big(\frac{\sqrt{n}(p_0-p_1)}{\sqrt{p_1(1-p_1)}} + \frac{\sqrt{p_0(1-p_0)}}{\sqrt{p_1(1-p_1)}} z_{\alpha/2}\Big)$

Example: proportion of cheating students

27 of the 267 students in a sample admitted to texting during an exam. Can it be concluded at significance level 0.01 that more than 5% of all students in the population sampled had texted during an exam?

H_0 : the proportion of cheating students $p = 0.05$ v.s. $H_a : p > 0.05$

- The test: reject H_0 if $Z > z_{0.01} = 2.33$, otherwise accept H_0
- Sample proportion $\hat{p} = \frac{27}{267} = 0.10$, sample test statistic we observed $z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} = \frac{\sqrt{267}\times(0.10-0.05)}{\sqrt{0.05\times 0.95}} = 3.75 > 2.33 \Rightarrow$ reject H_0, i.e. under 5% significance level, there's sufficient evidence that the proportion significantly deviates from 5%
- Under H_0, test statistic $Z = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}} \overset{d}{\approx} N(0,1)$
- p-value $= \mathbb{P}(Z > z) = \mathbb{P}(Z > 3.75) = \Phi(-3.75) \approx 0 < 0.01 \Rightarrow$ reject H_0

Example: proportion of cheating students

27 of the 267 students in a sample admitted to texting during an exam. Can it be concluded at significance level 0.01 that more than 5% of all students in the population sampled had texted during an exam?

H_0 : the proportion of cheating students $p = 0.05$ v.s. $H_a : p > 0.05$

- The test: reject H_0 if $Z > 2.33$, otherwise accept H_0
- When $p = 0.08$ (H_a true), we'd make Type-II error if accept H_0 ($Z \leq 2.33$).

  Type-II error rate $= \mathbb{P}\Big(Z = \frac{\sqrt{267}(\hat{p}-0.05)}{\sqrt{0.05\times 0.95}} \leq 2.33\Big)$

  $$= \mathbb{P}\Big(\underbrace{\frac{\sqrt{267}(\hat{p}-0.08)}{\sqrt{0.08\times 0.92}}}_{\overset{d}{\approx} N(0,1)} \leq 2.33\frac{\sqrt{0.05\times 0.95}}{\sqrt{0.08\times 0.92}} - \frac{\sqrt{267}\times 0.03}{\sqrt{0.08\times 0.92}}\Big)$$

  $\approx \Phi(0.06) = 0.5239$

Many thanks to

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