CSC380: Principles of Data Science

Statistics 2

Chicheng Zhang

Outline

- Interval estimation
- Hypothesis testing

Interval estimation

Motivation

- Point estimation makes statements of the form:
  - "Given the data, I estimate the bias of the coin to be 0.73"
  - "Given the data, I estimate the mean height of UA students to be 172cm"
- In many applications, we'd like to make statements with uncertainty quantifications
  - "Given the data, I estimate the bias of the coin to be 0.73 ± 0.05"
  - "Given the data, I estimate the mean height of UA students to be 172 ± 2cm"
- This is called *interval estimation*

Interval Estimation: basic setup

$$\theta \;\to X_1, \dots, X_n \;\to I_n = [\hat{\theta}_n \pm b_n]$$

data generation process   Confidence Interval (CI) for θ

Examples

Coin toss: $\theta = p, \;\; X_1, \dots, X_n \sim \text{Bernoulli}(p)$

Student height: $\theta = \mu, X_1, \dots, X_n \sim N(\mu, 8^2)$

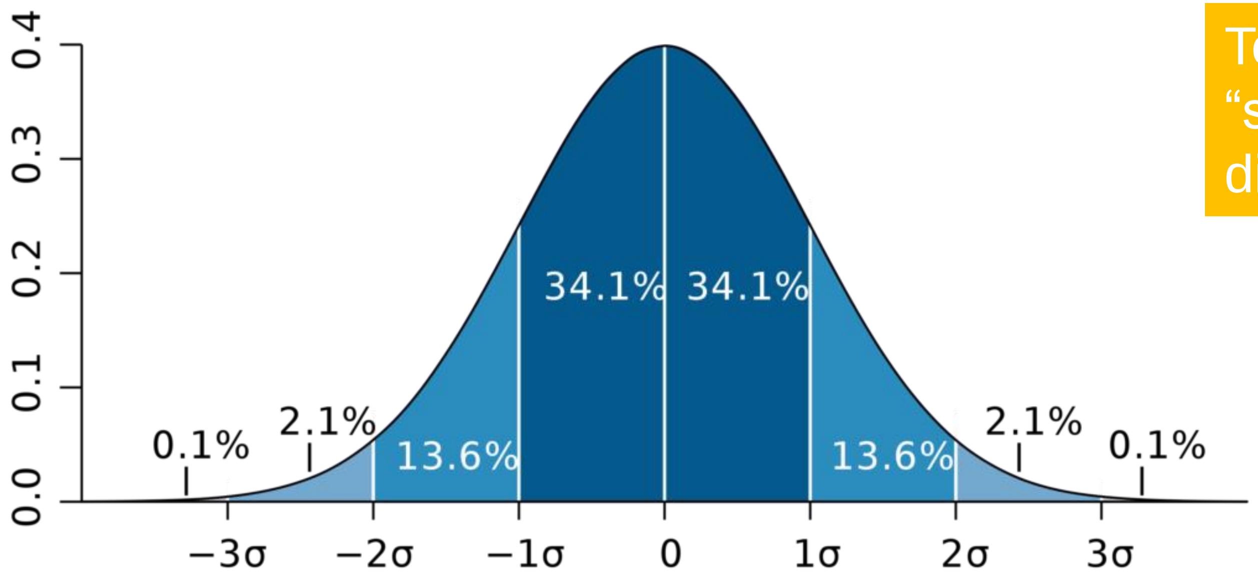
Goal: construct I_n using data, such that with 95% confidence (say),

$$\theta \in I_n$$

We will mostly focus on estimating θ = population mean, and will take $\hat{\theta}_n$ = sample mean.

How to choose b_n? **uncertainty of our estimate**

Recall: Normal distribution



Terminology: "standard" normal distribution := $N(0,1)$

Fact If $X \sim N(\mu, \sigma^2)$, then

$$P(-1.96\sigma \leq X - \mu \leq 1.96\sigma) = 0.95$$

In words:

with 95% confidence, X falls within 1.96 standard deviation of μ

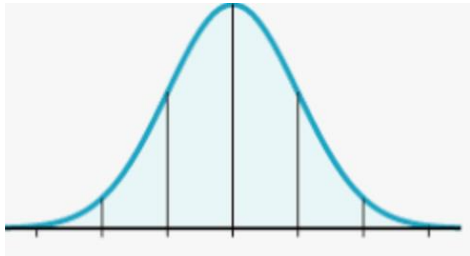
i.e, with 95% confidence, μ falls within 1.96 standard deviation of X

$[X - 1.96\sigma, X + 1.96\sigma]$ **is a 95% confidence interval for μ**

Recall: Central Limit Theorem

- Let $X_1, \ldots, X_n$ be iid with mean μ and variance σ^2. Then for large n, the sample mean $\bar{X}_n$ roughly follow a normal distribution:

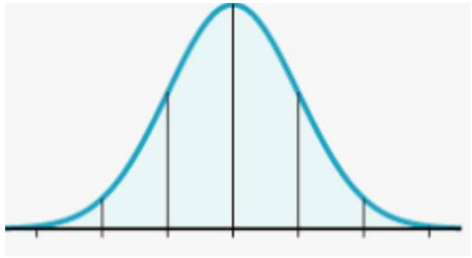
$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$



Constructing confidence interval

- Let $X_1, \dots, X_n$ be iid with mean μ and variance σ^2. Then for large n, the sample mean $\bar{X}_n$ roughly follow a normal distribution:

$$\bar{X}_n \approx N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$$



Corollary with 95% confidence, μ lies within $1.96\frac{\sigma}{\sqrt{n}}$ of $\bar{X}_n$

Our confidence interval for μ: $I_n = [\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}]$

Example: UA student height

Example Assume that UA students' heights (in centimeters) follow $N(\mu, 8^2)$, and we observe 4 students' heights:

$$163, 171, 179, 167$$

Find a 95% confidence interval for μ

Solution

Sample mean　　population stddev　　sample size

our CI for μ: $I_n = [\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}]$

$= 170$　　$\sigma = 8$　　n=4

Plugging in all values, $I_n = [170 \pm 7.84] = [162.1, 177.8]$

Confidence intervals: extensions

- What if we'd like to find 99% confidence interval? 99.9%?

Fact If $X \sim N(\mu, \sigma^2)$, then

standard normal CDF

$$P(-k\,\sigma \le X - \mu \le k\,\sigma) = 2\Phi(k) - 1$$

Setting

$$2\Phi(k) - 1 = p \Rightarrow\ k = \Phi^{-1}\left(\frac{p+1}{2}\right)$$

Our p confidence interval for μ:

$$I_n = [\bar{X}_n \pm \Phi^{-1}\left(\frac{p+1}{2}\right)\frac{\sigma}{\sqrt{n}}]$$

p	$\Phi^{-1}\left(\frac{p+1}{2}\right)$
0.95	1.96
0.99	2.58
0.999	3.29

Example: UA student height

Example Assume that UA students' heights (in centimeters) follow $N(\mu, 8^2)$, and we observe 4 students' heights:

163, 171, 179, 167

Find 99%, 99.9% confidence intervals for μ

Solution

our p-CI for μ: $I_n = [\bar{X}_n \pm \Phi^{-1}\left(\frac{p+1}{2}\right)\frac{\sigma}{\sqrt{n}}]$

$$p = 0.99 \Rightarrow [159.7, 180.3]$$

$$p = 0.999 \Rightarrow [156.9, 183.1]$$

p	$\Phi^{-1}\left(\frac{p+1}{2}\right)$
0.95	1.96
0.99	2.58
0.999	3.29

Confidence interval: observations

$$p\text{-CI for } \mu\text{: } I_n = [\bar{X}_n \pm \Phi^{-1}\left(\frac{p+1}{2}\right)\frac{\sigma}{\sqrt{n}}]$$

$p = 0.95 \Rightarrow [162.1, 177.8]$
$p = 0.99 \Rightarrow [159.7, 180.3]$
$p = 0.999 \Rightarrow [156.9, 183.1]$

The center is always at $\bar{X}_n$

The width of the interval depends on:

- Sample size n: width smaller when n larger
- Confidence level p: width larger when p closer to 1
- Population stddev σ: width larger when σ large (more noise)

What if σ is unknown?

- We will address this soon..

Confidence interval: interpretation

Example Assume that UA students' heights (in centimeters) follow $N(\mu, 8^2)$, and we observe 4 students' heights:

$$163, 171, 179, 167$$

we found that a 95% CI for μ is $[162.1, 177.8]$

Can we say "with probability 95%, the population mean height μ lies in interval $[162.1, 177.8]$"?

No! This is a common misinterpretation

- μ is deterministic, and $[162.1, 177.8]$ is deterministic,
- Proposition $\mu \in [162.1, 177.8]$ is either true or false!

Then, what does "95% probability" mean?

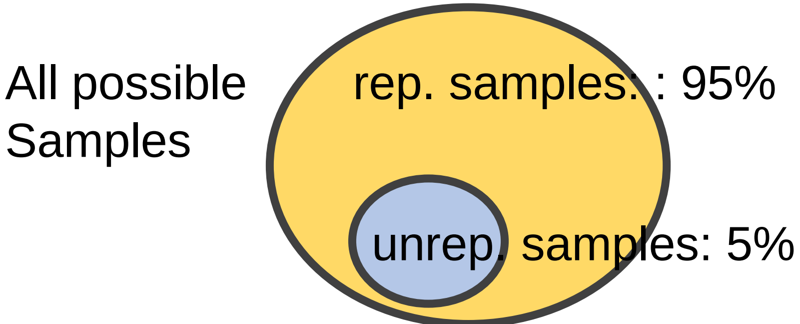
Confidence interval: interpretation

Example Assume that UA students' heights (in centimeters) follow $N(\mu, 8^2)$, and we observe 4 students' heights:

$$163, 171, 179, 167$$

True: With probability 0.95 *over the draw of a sample*, $[\bar{X}_n \pm 7.84]$ contains μ

As long as we are not extremely unlucky / our sample is mildly representative, my CI contains μ



Confidence interval: interpretation

Example Assume that UA students' heights (in centimeters) follow $N(\mu, 8^2)$, and we observe 4 students' heights:

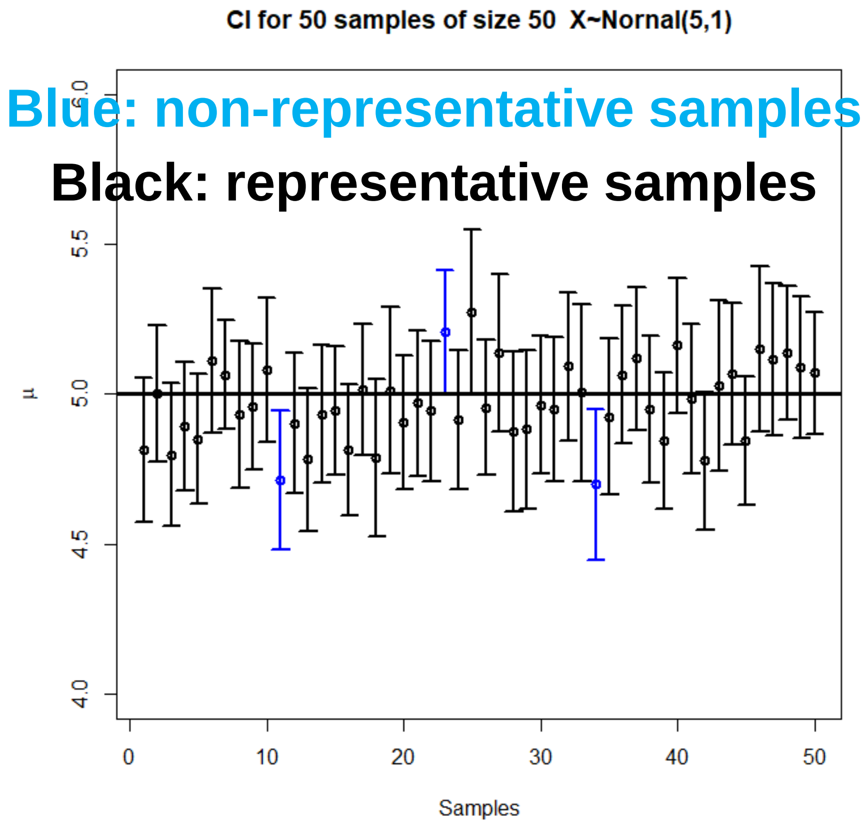
$$163, 171, 179, 167$$

True: With probability 0.95 *over the draw of a sample*, $[\bar{X}_n \pm 1.96 \frac{\sigma}{\sqrt{n}}]$ contains μ

50 draws of samples

$\Rightarrow$ 50 CIs

$\Rightarrow$ expect $50\times 95\% = 47.5$ CI's to contain μ



Confidence interval for mean with unknown variance

Example Assume that UA students' weights (in kgs) follow $N(\mu, \sigma^2)$, and we observe 4 students' weights:

60, 65, 70, 75

Find a 95% confidence interval for μ

Note The CI construction before $[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}]$ no longer works, since σ is *unknown*

How to fix this?

Confidence interval for mean with unknown variance

- $[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}]$ no longer works, since σ is unknown

- We can try replace σ with its estimate

$$\hat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

and return $[\bar{X}_n - 1.96\frac{\hat{\sigma}_n}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\hat{\sigma}_n}{\sqrt{n}}]$

- This is OK, but it is too aggressive to be valid
  - may contain μ with probability much lower than 95%..

The student-t distribution

Fact $X_1, \ldots, X_n$ is an iid sample with unknown μ & σ^2. Let $\hat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$. Then, approximately:

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\hat{\sigma}_n} \sim \text{student-t}(n-1)$$

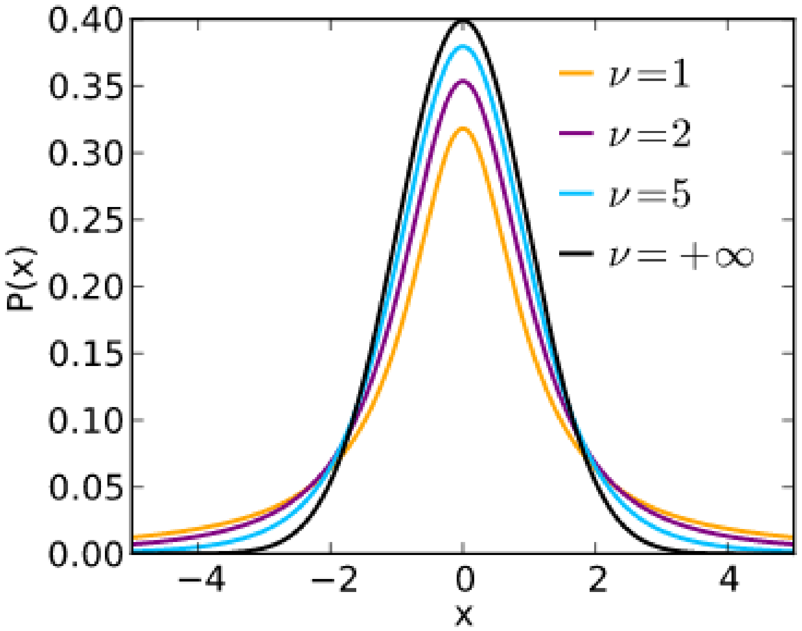
Abbreviated as $t(n-1)$

degree of freedom

student-t(ν) is a family of distributions

William S Gosset (aka *Student*)

While working at Guinness



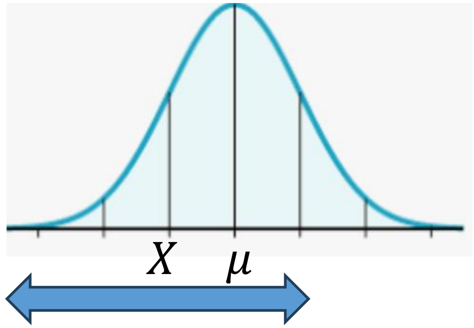
Announcements 4/30

- Quiz 12 next Monday

- We have a list of sample questions to help prepare for the final

Recap 4/30

- Confidence interval: estimate unknown quantity (e.g. population mean μ) with *uncertainty quantification*



- Baby confidence interval (sample of size 1)
  if $X \sim N(\mu, 1)$,
  then $[X - 1.96, X + 1.96]$ is a 95% confidence interval for μ

- General confidence interval
  if $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$,
  $[\bar{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X}_n + 1.96\frac{\sigma}{\sqrt{n}}]$ is a 95% confidence interval for μ

What if σ is unknown?

The student-t distribution

Fact $X_1, \ldots, X_n$ is an iid sample with unknown μ & σ^2. Let $\hat{\sigma}_n =$ sample stddev. Then, approximately:

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\hat{\sigma}_n} \sim \text{student-t}(n-1)$$

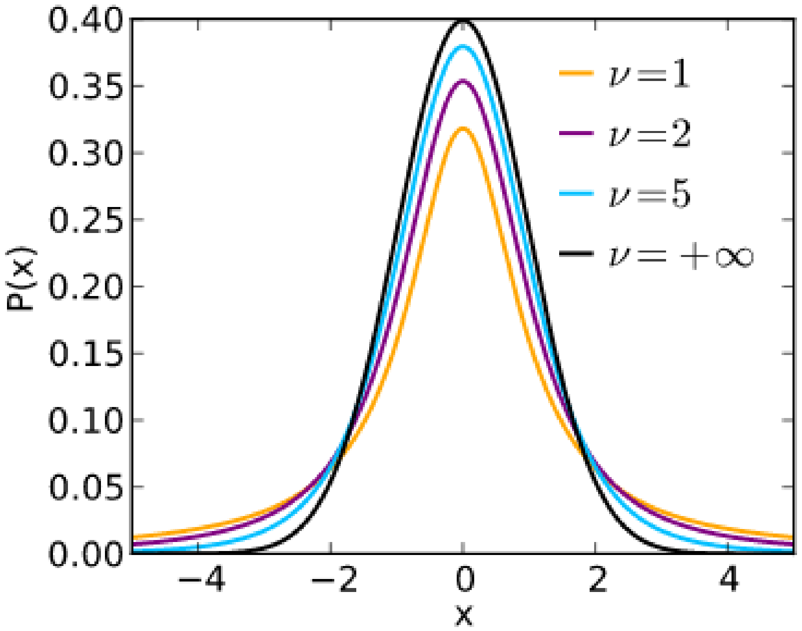
degree of freedom

Abbreviated as $t(n-1)$

student-t(ν) is a family of distributions

William S Gosset (aka *Student*)

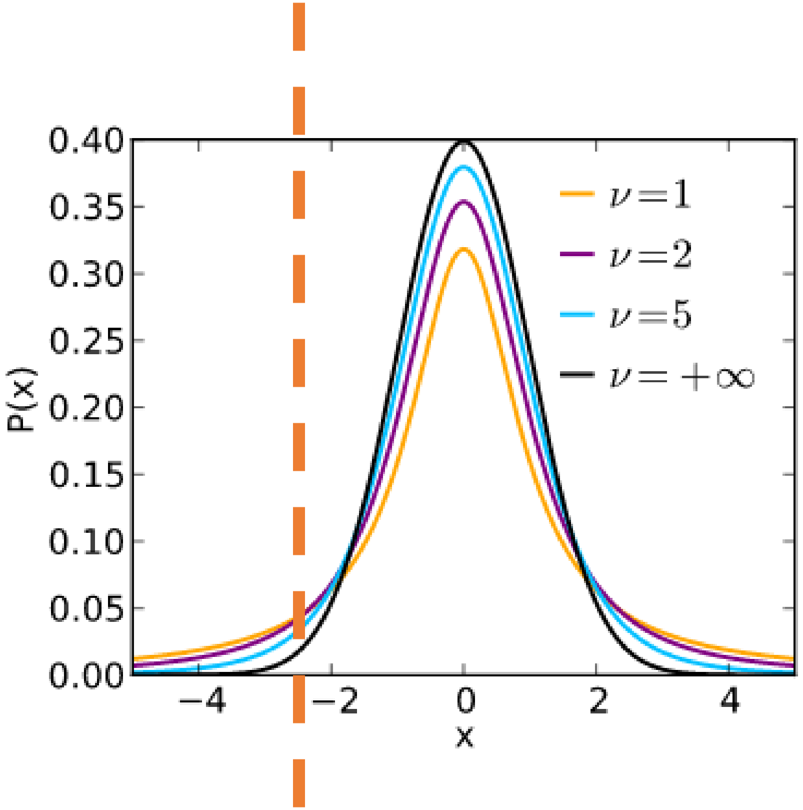
While working at Guinness



The student-t distribution

student-t(ν) distribution family

- goes to Gaussian when ν is large
- generally has heavier tail than Gaussian




```
import scipy.stats as st

st.t.ppf(0.975,df=2)
=> 4.302652729911275

st.t.ppf(0.975,df=5)
=> 2.5705818366147395

st.t.ppf(0.975,df=10)
=> 2.2281388519649385

st.t.ppf(0.975,df=100)
⇒ 1.9839715184496334

Recall:
st.norm.ppf(0.975) gives 1.96
```

Confidence interval for mean with unknown variance

Example Assume that UA students' weights (in kgs) follow $N(\mu, \sigma^2)$, and we observe 4 students' weights:

60, 65, 70, 75

Find a 95% confidence interval for μ

Solution

sample mean

$$\sqrt{4}\frac{\bar{X}_4 - \mu}{\hat{\sigma}_4} \sim t(3)$$

sample stddev

st.t.ppf(0.975,df=3)
=> 3.18

Therefore, $P\left(\left|\sqrt{4}\frac{\bar{X}_4 - \mu}{\hat{\sigma}_4}\right| \leq 3.18\right) \geq 0.95$

Example Assume that UA students' weights (in kgs) follow $N(\mu, \sigma^2)$, and we observe 4 students' weights:

$$60, 65, 70, 75$$

Find a 95% confidence interval for μ

Solution With 95% confidence,

$$\left|\sqrt{4}\frac{\bar{X}_4 - \mu}{\hat{\sigma}_4}\right| \le 3.18$$

$$\Rightarrow \mu \in \left[\underbrace{\bar{X}_4}_{= 67.5} - 3.18\overbrace{\frac{\hat{\sigma}_4}{\sqrt{4}}}^{= 6.45}, \bar{X}_4 + 3.18\frac{\hat{\sigma}_4}{\sqrt{4}}\right]$$

Our confidence interval

Plugging data,

our CI is $[67.5 - 10.3, 67.5 + 10.3] = [57.2, 77.8]$

Confidence interval for mean with unknown variance

General result given a sample $X_1, \dots, X_n$ drawn from a distribution with mean μ, a p-confidence interval (e.g. p=95%) is

$$\left[\bar{X}_n - w\frac{\hat{\sigma}_n}{\sqrt{n}}, \bar{X}_n + w\frac{\hat{\sigma}_n}{\sqrt{n}}\right],$$

where w is the $\left(\frac{1+p}{2}\right)$-quantile of the $t(n-1)$ distribution

Example p=0.95, n=4 $\Rightarrow w = 3.18$

```
st.t.ppf(0.975,df=3)
=>  3.18
```

p=0.99, n=4 $\Rightarrow$ $w = 5.84$

p=0.99, n=9 $\Rightarrow w = 3.35$

Confidence interval: closing remarks

- How to construct confidence intervals for μ?

- We consider the probability distributions of the following statistics:

z-statistic	$\sqrt{n}\dfrac{\bar{X}_n-\mu}{\sigma}$	$\sim\ N(0,1)$	when σ is known
t-statistic	$\sqrt{n}\dfrac{\bar{X}_n-\mu}{\hat{\sigma}_n}$	$\sim t(n-1)$	when σ is unknown

and make use of that they cannot take values that are too extreme

- general CI construction follows a similar principle
  - Define [pivotal quantity] and use its quantile to guide CI construction

Hypothesis testing

Hypothesis

- Statements about parameter / property θ of a distribution / population

Examples

- Average GPA < 2.8
- Probability of head of a coin > 0.6
- People eat more on weekends than weekdays

Simple vs. composite hypotheses

$\theta = 3.2$ (simple) , $\theta \in \{3.2, 4\}$ (composite), $\theta \in [3.2, 4]$ (composite)

One-sided vs. two-sided

$\theta > 3.2$ (one-sided) , $\theta < 1.5 \text{ or } \theta > 3.2$ (two-sided), $\theta \neq 2$ (two-sided)

Hypothesis testing

Hypothesis testing: choosing from two hypotheses:

- Null hypothesis H_0
  - Status quo, assumption believed to be true
  - Coin in my pocket, probability of head $p = 0.5$
- Alternative hypothesis H_1: Complement of H_0
  - Novel finding after research
  - Coin has probability of head $p \neq 0.5$

Hypothesis testing

- How to test?
- Design experiment, collect data, check:

If data shows strong evidence against H_0:

Reject H_0 (in favor of H_1)

Else

Do not reject H_0 **Note: does not necessarily mean "accept H_0 "**

- Analogy with the legal principle:
  - Presume innocent (H_0) until proven guilty (H_1) with strong evidence against innocence

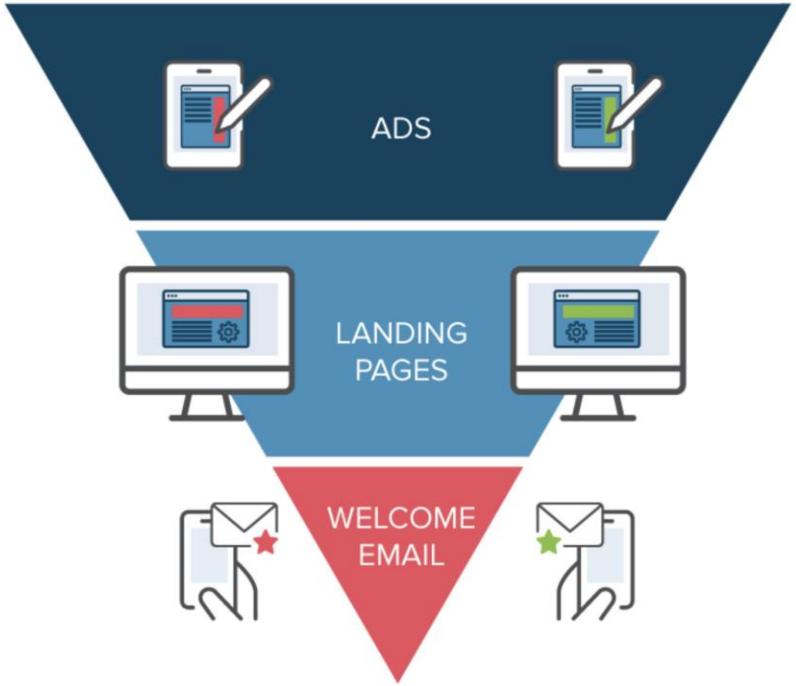
Application: A/B Testing

These days, Internet companies run A/B testing extensively

Try out an alternative of user interface (UI) on **<u>randomly chosen subset of users</u>** to collect their feedback (e.g. rating)

- E.g, choosing b/w **list view** vs grid view

How do we know if the new UI is better than older one? (i.e., statistically significant)



(from optimizely.com)

Application: A/B Testing

Evaluator:	**1**	**2**	**3**	**4**	**5**	**6**
Old UI	5	2	2	5	4	2
New UI	4	4	1	3	3	5

Compute the score differences:

Evaluator:	**1**	**2**	**3**	**4**	**5**	**6**
Score difference X	-1	+2	-1	-2	-1	+3

Can view X's as drawn from some distribution with unknown mean μ

"Does new UI improve over old UI?" is now a hypothesis testing problem:

$$H_0\colon\ \mu \leq 0, \qquad\qquad H_1\colon\ \mu > 0$$

we can perform e.g. t-test based on data (we will see)

Hypothesis testing

- More specifically:
  - Design experiment
  - Design test statistic W (related to hypothesis)
  - Find distribution of W under H_0
  - Compute w, value of W applied on the data $X_1, \dots, X_n$
  - Reject H_0 if $w \in R$, for “reasonable” rejection region R

Example Assume that UA students' heights (in centimeters) follow $N(\mu, 8^2)$, test the hypothesis

$$H_0\colon \mu = 168, \qquad\qquad H_1\colon \mu \neq 168$$

Suppose we observe 4 students' heights: 173, 181, 189, 177

Informally: reject H_0 if $\bar{X}_n$ is far from 168

Formally: define $Z = \frac{\sqrt{n}(\bar{X}_n - 168)}{8}$. Reject H_0 if $|Z| \geq c$

We will see where the $\frac{\sqrt{n}}{8}$ factor comes from..

How to choose c ?

Hypothesis testing

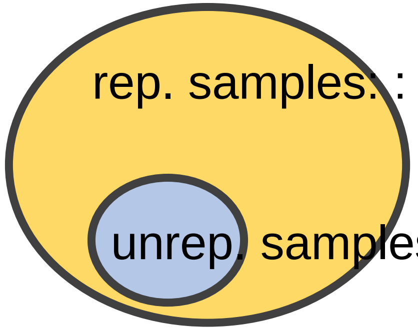
- How to choose c?

- Significance level α:

$$Z = \frac{\sqrt{n}(\bar{X}_n - 168)}{8}$$

$$P_{H_0}(|Z| \geq c) \leq \alpha$$

Type-I error: we reject H_0 (due to $\bar{X}_n$ far from 168), but H_0 is true



- Usually α is small, e.g. 0.05
- I.e., stay with the null hypothesis as long as our sample is 95%-representative

Smaller α => more inclined to stay with H_0 => Need stronger evidence to reject H_0

Hypothesis testing

Choose c such that

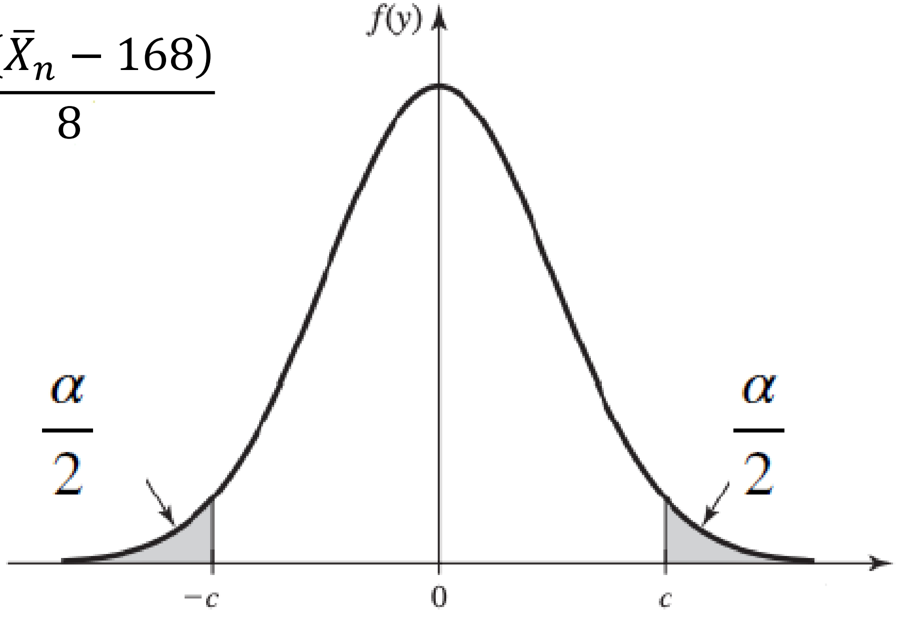
$$P_{H_0}(|Z| \geq c) = \alpha = 5\%$$

Reject H_0 if $|Z| \geq c$, i.e.

Z falls in the shaded region

Let's find the value of c..

$$Z = \frac{\sqrt{n}(\bar{X}_n - 168)}{8}$$



PDF of Z

Hypothesis testing

z-statistic: a statistic that is supposed to follow $N(0,1)$

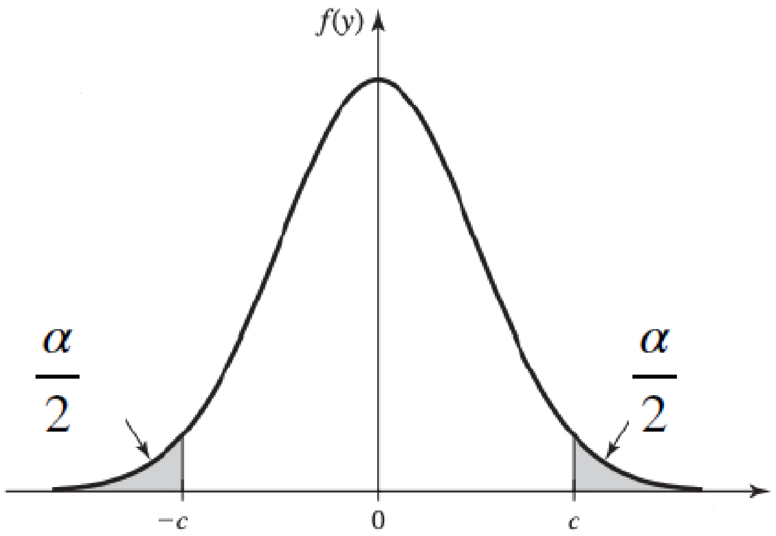
- under H_0, by central limit theorem:

$$Z = \frac{\sqrt{n}(\bar{X}_n - 168)}{8} \sim N(0,1)$$

Z **is a valid z-statistic**

c is such that $P_{Z\sim N(0,1)}(|Z| \geq c) = \alpha$

$$\Rightarrow c = \Phi^{-1}(1 - \alpha/2)$$



E.g. $\alpha = 0.05 \Rightarrow c = \Phi^{-1}(0.975) = \text{st.norm.ppf}(0.975) = 1.96$

z-test

Example Assume that UA students' heights (in centimeters) follow $N(\mu, 8^2)$, test the hypothesis

$$H_0: \mu = 168, \qquad H_1: \mu \neq 168$$

Suppose we observe 4 students' heights: 173, 181, 189, 177

We reject if $Z = \frac{\sqrt{n}}{8}|\bar{X}_n - 168| \geq \Phi^{-1}(0.975)$

1.96

z-statistic: a statistic that is supposed to follow $N(0,1)$

This is called a z-test

From data, $Z = 3$, so we reject H_0.

z-test: general case

General fact Assume that we have a set of samples $X_1, \ldots, X_n$ that follow $N(\mu, \sigma^2)$, test the hypothesis

$$H_0: \mu = \mu_0, \qquad H_1: \mu \neq \mu_0$$

with significance level α

We can use the z-test:

$$\text{Reject if } |Z| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \textit{ where } Z = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu_0)$$

Larger *n* ⇒ more reject

rejection threshold r

Larger α ⇒ more reject

Larger σ ⇒ less reject

Hypothesis testing

General fact Assume that we have a set of samples $X_1, \ldots, X_n$ that follow $N(\mu, \sigma^2)$, test the hypothesis

$$H_0\colon\ \mu = \mu_0, \qquad\qquad H_1\colon\ \mu \neq \mu_0$$

with significance level α

z-test: Reject if $|Z| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

rejection threshold r

Example $\sigma = 8$, $n = 4$, $\bar{X}_n = 180$, use z-test to test if $\mu = 168$

$$Z = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu_0) = 3$$

$\alpha = 0.05 \Rightarrow r = 1.96$ reject H_0

$\alpha = 0.01 \Rightarrow r = 2.58$ reject H_0

$\alpha = 0.001 \Rightarrow r = 3.29$ do not reject H_0

p-value

Example $\sigma = 8$, $n = 4$, $\bar{X}_n = 180$, test if $\mu = 168$

$\alpha = 0.05 \Rightarrow r = 1.96$ reject H_0

$\alpha = 0.01 \Rightarrow r = 2.58$ reject H_0

$\alpha = 0.001 \Rightarrow r = 3.29$ do not reject H_0

The p-value of z-test given this data, is between 0.01 and 0.001 (we will see its exact value soon..)

p-value: given the data & hypothesis test,

p = the smallest α so that we can still reject H_0

E.g. $p \leq 0.01$ => even with a strong bias in keeping H_0 ($\alpha = 0.01$), we still reject H_0

Smaller p-value implies stronger evidence in rejecting H_0

p-value

Example $\sigma = 8$, $n = 4$, $\bar{X}_n = 180$, use z-test to test if $\mu = 168$. Find the test's p-value.

- Reject with significance level α if

$$3 = |Z| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

- p-value = smallest α s.t. $3 \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

  = solution of eqn "$3 = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$"

  = 0.0027

```
from scipy.stats import norm
2*(1-norm.cdf(3))
```

```
np.float64(0.002699796063260207)
```

- => We reject $\mu = 168$ with strong evidence

Other z-tests

- Other tests can be found using the same reasoning

Reject H_0 if:

- $H_0\colon\ \mu = \mu_0$, vs $H_1\colon\ \mu \neq \mu_0$

$|Z| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$ $\quad Z = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu_0)$

one-sided hypothesis testing problem

- $H_0\colon\ \mu \leq \mu_0$, vs $H_1\colon\ \mu > \mu_0$

$Z \geq \Phi^{-1}(1 - \alpha)$

- $H_0\colon\ \mu \geq \mu_0$, vs $H_1\colon\ \mu < \mu_0$

$Z \leq \Phi^{-1}(\alpha)$

All these are z-tests, since it uses the z-statistic

$Z = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu_0)$

Other z-tests

- Other tests can be derived using the same reasoning

p-value of the test

$H_0\colon \mu = \mu_0$, vs $H_1\colon \mu \neq \mu_0$	$2\big(1 - \Phi(\lvert Z\rvert)\big)$
$H_0\colon \mu \leq \mu_0$, vs $H_1\colon \mu > \mu_0$	$1 - \Phi(Z)$
$H_0\colon \mu \geq \mu_0$, vs $H_1\colon \mu < \mu_0$	$\Phi(Z)$

One-sided z-test: example

Example Assume that UA students' heights follow $N(\mu, 8^2)$, test the hypothesis

$$H_0\colon \mu \leq 168, \qquad H_1\colon \mu > 168$$

we have collected a sample of size $n = 4$, with mean $\bar{X}_n = 180$. Find the p-value of the z-test.

Solution $Z = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} = 3$

$$p = 1 - \Phi(Z) = 1 - \Phi(3) = 0.0013$$

=> we reject H_0 with strong evidence

Announcements 5/5

- Final exam logistics
  - You are welcome to bring a calculator
  - You are welcome to bring a letter-sized “cheatsheet”
  - LMK if you’d like me to discuss specific practice final questions!
- 7Student course survey will close on 5/7
  - The bonus will only increase your scores
- We will close collection of participation self-report this week
  - We need time in calculating the final grades

Quiz 12

You manage a snack room at work with a vending machine that fills soda cans with 12 oz of soda on average. The machine's standard deviation is known to be 1 oz, based on manufacturer specs.

A curious employee thinks the machine is cheating people—maybe overfilling or underfilling. You decide to investigate.

You take a random sample of 36 cans and find the average is 12.3 oz.

1. Compute the z-statistic (hint: $Z = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma}$)
2. If our z-test has significance level $\alpha = 0.05$, should we reach the conclusion that the machine is cheating people?

   (hint: norm.ppf(0.95) = 1.64, norm.ppf(0.975) = 1.96)

Quiz 12

- Soda volume X: mean μ, stddev $\sigma = 1$
- The hypothesis testing problem:

$$H_0: \mu = 12, \qquad H_1: \mu \neq 12$$

no cheating cheating

- Data: $n = 36$, $\bar{X}_n = 12.3$

- $Z = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma} = \frac{6 \times (12.3 - 12)}{1} = 1.8$

Quiz 12

If our z-test has significance level α =0.05, should we reach the conclusion that the machine is cheating people?

(hint: norm.ppf(0.95) = 1.64, norm.ppf(0.975) = 1.96)

our z-test:

$$|Z| \geq \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$
$$= \Phi^{-1}(0.975) = 1.96$$

$|1.8| < 1.96 \Rightarrow$ do not reject $H_0 \Rightarrow$ should not reach this conclusion

t-test

- Drawback of z-test: needs to know population stddev σ

Example Suppose the #of medical inpatient days in nursing homes follow a distribution with mean μ and variance σ^2. We'd like to perform hypothesis test between:

$$H_0\colon \mu = 200, \qquad H_1\colon \mu \neq 200$$

and we observe $n = 18$ samples with $\bar{X}_n = 182.17$ and $\hat{\sigma}_n = 17.72$

Should I reject H_0?

t-test

Example Suppose the #of medical inpatient days in nursing homes follow a distribution with mean μ and variance σ^2. We'd like to perform hypothesis test between:

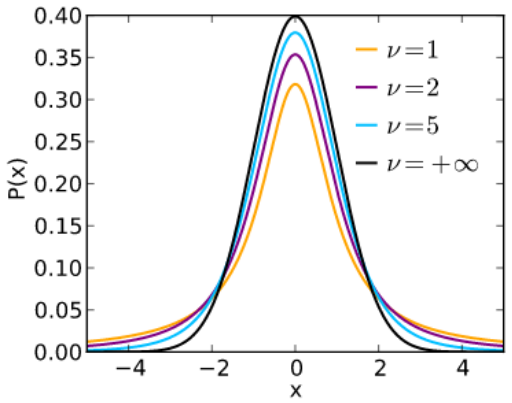
$$H_0\colon \mu = 200, \qquad\qquad H_1\colon \mu \neq 200$$

and we observe $n = 18$ samples with $\bar{X}_n = 182.17$ and $\hat{\sigma}_n = 17.72$

Approach When H_0 happens,

this is called a t-statistic, i.e, a statistic that follows t-distribution

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_n} \sim t(n-1)$$

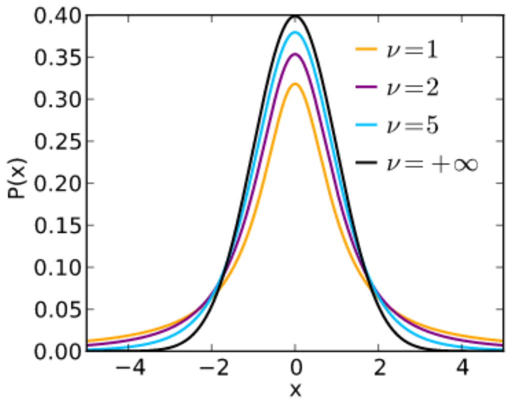


observed value $\frac{\sqrt{18}(182.17-200)}{17.72} = -1.018$

t-test

Approach We've seen that under H_0,

$$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_n} \sim t(n-1)$$



Our test with significance α:

$$\text{reject when } |T| > F^{-1}\left(1 - \frac{\alpha}{2}\right)$$

F is now the CDF of the $t(n-1)$ distribution

(Note how similar this is to the z-test)

$|T| = 1.018$

$F^{-1}\left(1 - \frac{\alpha}{2}\right) = 2.11$

```
1 st.t.ppf(1-0.05/2, 17)
```

```
np.float64(2.1098155778331806)
```

thus, we do not reject H_0: $\mu = 200$

Other t-tests

- Other tests can be found using the same reasoning

Reject H_0 if:

- $H_0\colon \mu = \mu_0$, vs $H_1\colon \mu \neq \mu_0$ $\quad |T| \geq F^{-1}\left(1 - \frac{\alpha}{2}\right)$ $\quad T = \frac{\sqrt{n}}{\hat{\sigma}_n}(\bar{X}_n - \mu_0)$

F: CDF of $t(n-1)$

- $H_0\colon \mu \leq \mu_0$, vs $H_1\colon \mu > \mu_0$ $\quad T \geq F^{-1}(1-\alpha)$

- $H_0\colon \mu \geq \mu_0$, vs $H_1\colon \mu < \mu_0$ $\quad T \leq F^{-1}(\alpha)$

All these are called t-test, since it relies on computing T, a t-statistic

Other t-tests

- Other tests can be found using the same reasoning

	p-value of the test
$H_0\colon \mu = \mu_0$, vs $H_1\colon \mu \neq \mu_0$	$2\big(1 - F(\lvert T\rvert)\big)$
$H_0\colon \mu \leq \mu_0$, vs $H_1\colon \mu > \mu_0$	$1 - F(T)$
$H_0\colon \mu \geq \mu_0$, vs $H_1\colon \mu < \mu_0$	$F(T)$

t-test: a practice final question

Example Metal fibers produced, length in millimeters; use t-test to test

$$H_0\colon\ \mu \le 5.2, \qquad\qquad H_1\colon\ \mu > 5.2$$

n=15 fibers measured, $\bar{X}_n = 5.4$, $\hat{\sigma}_n = 0.4226$.

Shall we reject H_0 at significance 0.05?

Solution The t-test is "reject if $T \ge F^{-1}(1-\alpha)$"

t-statistic $T = \frac{\sqrt{n}}{\hat{\sigma}_n}(\bar{X}_n - \mu_0) = 1.83$

rejection threshold $F^{-1}(1-\alpha) =$ t.ppf(0.95, 14) = 1.76 we should reject

Backup