STAT 583 Lecture 2

Linjun (Leon) Zhang

Department of Statistics
Rutgers University

Jan. 28th

Misc

- Average score: 6.4
- keep basic math concepts; remove some hard math concepts, and add a few applied examples
- Office Hours:
  - Instructor: 9:30-11:30am, 9:30pm-10pm every Tuesday,
  - TA: TBD

Last time

- The relationship between statistics and probability theory
  - Data and random variables
  - Sample and population
  - Estimates and parameters
- Sampling distributions
  - Central Limit Theorem
  - chi-squared distribution and t-distribution

Recap

Random variables / Data:

By data we mean the observed value of a random variable once some experiment has been performed.

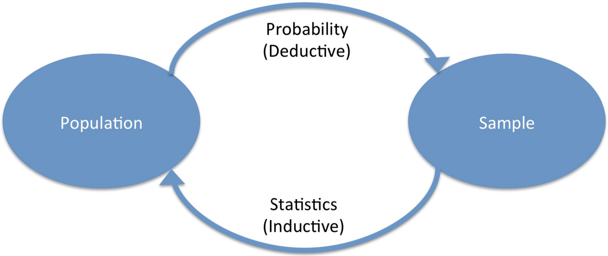
- Random variables: before the experiment
- Data: after the experiment

We typically write X_1 , . . . , X_n for hypothetical random sample, and the lower case $x_1, ..., x_n$ for their observed values.

Recap

Population / Sample

- Population: the complete collection of units about which info is sought
- Sample: a subset of a population that is actually observed



Recap

Parameter / Estimate

- Parameter: a numerical characteristic of a population
- Estimate: a numerical function of the sample data, used to make inference about the unknown parameter

Example:

- "true" kidney cancer death rate in a county (e.g., if we were to keep observing for an infinitely long time)

Recap

Parameter / Estimate

- Parameter: a numerical characteristic of a population
- Estimate: a numerical function of the sample data, used to make inference about the unknown parameter

Example:

- "true" kidney cancer death rate in a county (e.g., if we were to keep observing for an infinitely long time)
- observed kidney cancer death rate in a county in any given year

Recap

Parameter / Estimate

- Parameter: a numerical characteristic of a population
- Estimate: a numerical function of the sample data, used to make inference about the unknown parameter

Example:

- "true" kidney cancer death rate in a county (e.g., if we were to keep observing for an infinitely long time)
- observed kidney cancer death rate in a county in any given year
- A company wishes to estimate the mean service time for customers.

Recap

Parameter / Estimate

- Parameter: a numerical characteristic of a population
- Estimate: a numerical function of the sample data, used to make inference about the unknown parameter

Example:

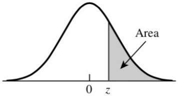
- "true" kidney cancer death rate in a county (e.g., if we were to keep observing for an infinitely long time)
- observed kidney cancer death rate in a county in any given year
- A company wishes to estimate the mean service time for customers.
- A manufacturer wishes to estimate the standard deviation of the diameters of a part produced in a factory.

Warm-up

If $X \sim N(3, 2^2)$

- Find the probability $\mathbb{P}(X \leq 4)$.

Normal Curve Areas
Standard normal probability in right-hand tail (for negative values of z, areas are found by symmetry)



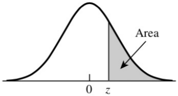
	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681

Warm-up

If $X \sim N(3, 2^2)$

- Find the probability $\mathbb{P}(X \leq 4)$.
- $\mathbb{P}(X \leq 4) = \mathbb{P}(\frac{X-3}{2} \leq \frac{4-3}{2}) = \mathbb{P}(Z \leq 0.5) = 1 - 0.3085 = 0.6915.$

Normal Curve Areas
Standard normal probability in right-hand tail (for negative values of z, areas are found by symmetry)



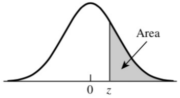
	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681

Warm-up

If $X \sim N(3, 2^2)$

- Find the probability $\mathbb{P}(1 \leq X \leq 4)$.

Normal Curve Areas
Standard normal probability in right-hand tail (for negative values of z, areas are found by symmetry)



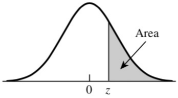
	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985

Warm-up

If $X \sim N(3, 2^2)$

- Find the probability $\mathbb{P}(1 \le X \le 4)$.
- $\mathbb{P}(1 \le X \le 4) = \mathbb{P}(\frac{1-3}{2} \le \frac{X-3}{2} \le \frac{4-3}{2}) = \mathbb{P}(-1 \le Z \le 0.5) = 1 - \mathbb{P}(Z > 1) - \mathbb{P}(Z > 0.5) = 1 - 0.1587 - 0.3085 = 0.5328.$

Normal Curve Areas
Standard normal probability in right-hand tail (for negative values of *z*, areas are found by symmetry)



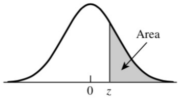
	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985

Warm-up

If $X \sim N(3, 2^2)$

- Find x, such that $\mathbb{P}(X \leq x) = 0.9$.

Normal Curve Areas
Standard normal probability in right-hand tail (for negative values of z, areas are found by symmetry)



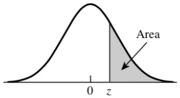
	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Warm-up

If $X \sim N(3, 2^2)$

- Find x, such that $\mathbb{P}(X \leq x) = 0.9$.
- $0.9 = \mathbb{P}(X \leq x) = \mathbb{P}(\frac{X-3}{2} \leq \frac{x-3}{2}) = \mathbb{P}(Z \leq \frac{x-3}{2})$. Since $\mathbb{P}(Z \leq 1.2816) = 0.9$, we have $\frac{x-3}{2} = 1.2816$, and this implies $x = 5.5632$.

Normal Curve Areas
Standard normal probability in right-hand tail (for negative values of *z*, areas are found by symmetry)



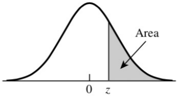
	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Warm-up

If $X \sim N(3, 2^2)$

- Find x, such that $\mathbb{P}(X \leq x) = 0.9$.
- $0.9 = \mathbb{P}(X \leq x) = \mathbb{P}(\frac{X-3}{2} \leq \frac{x-3}{2}) = \mathbb{P}(Z \leq \frac{x-3}{2})$. Since $\mathbb{P}(Z \leq 1.2816) = 0.9$, we have $\frac{x-3}{2} = 1.2816$, and this implies $x = 5.5632$.
- In R: qnorm and pnorm

Normal Curve Areas
Standard normal probability in right-hand tail (for negative values of *z*, areas are found by symmetry)

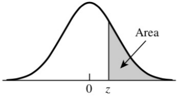


z	Second decimal place of z .00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Example

Roll a die 100 times. Find the chance that the sample mean is less than 3.7 and more than 3.3.

Normal Curve Areas
Standard normal probability in right-hand tail (for negative values of *z*, areas are found by symmetry)



	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455

Example

Answer

- Define X as the number to turn up. We then have $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(X = 2) = ... = \mathbb{P}(X = 6) = \frac{1}{6}$.

Example

Answer

- Define X as the number to turn up. We then have $\mathbb{P}(X=0)=\mathbb{P}(X=1)=\mathbb{P}(X=2)=\ldots=\mathbb{P}(X=6)=\frac{1}{6}$.
- We have $\mathbb{E}(X)=3.5$, $Var(X)=\frac{1}{6}\cdot(1^2+\ldots+6^2)-3.5^2=2.917$.

Example

Answer

- Define X as the number to turn up. We then have $\mathbb{P}(X=0)=\mathbb{P}(X=1)=\mathbb{P}(X=2)=\ldots=\mathbb{P}(X=6)=\frac{1}{6}$.
- We have $\mathbb{E}(X)=3.5$, $Var(X)=\frac{1}{6}\cdot(1^2+\ldots+6^2)-3.5^2=2.917$.
- By CLT, $\bar{X}=\frac{X_1+\ldots+X_{100}}{100}\sim N(3.5,\frac{2.917}{100})$.

Example

Answer

- Define X as the number to turn up. We then have $\mathbb{P}(X=0)=\mathbb{P}(X=1)=\mathbb{P}(X=2)=\ldots=\mathbb{P}(X=6)=\frac{1}{6}$.
- We have $\mathbb{E}(X)=3.5$, $Var(X)=\frac{1}{6}\cdot(1^2+\ldots+6^2)-3.5^2=2.917$.
- By CLT, $\bar{X}=\frac{X_1+\ldots+X_{100}}{100}\sim N(3.5,\frac{2.917}{100})$.
- To find $\mathbb{P}(3.3<\bar{X}<3.7)$.

Example

Answer

- Define X as the number to turn up. We then have $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(X = 2) = ... = \mathbb{P}(X = 6) = \frac{1}{6}$.
- We have $\mathbb{E}(X) = 3.5$, $Var(X) = \frac{1}{6} \cdot (1^2 + ... + 6^2) - 3.5^2 = 2.917$.
- By CLT, $\bar{X} = \frac{X_1 + ... + X_{100}}{100} \sim N(3.5, \frac{2.917}{100})$.
- To find $\mathbb{P}(3.3 < \bar{X} < 3.7)$.
- $\mathbb{P}(3.3 < \bar{X} < 3.7) = \mathbb{P}(\frac{3.3-3.5}{\sqrt{\frac{2.917}{100}}} < Z < \frac{3.3-3.5}{\sqrt{\frac{2.917}{100}}}) = \mathbb{P}(-1.17 < Z < 1.17) \approx \mathbb{P}(|Z| \leq 1.2) = 0.7699$.

Example

Suppose a sample is drawn *i.i.d* from a population whose mean is μ and whose standard deviation is σ. What is the approximate chance that the sample mean is within $2 \cdot \frac{\sigma}{\sqrt{n}}$ of the population mean μ?

Example

Suppose a sample is drawn $i.i.d$ from a population whose mean is μ and whose standard deviation is σ. What is the approximate chance that the sample mean is within $2 \cdot \frac{\sigma}{\sqrt{n}}$ of the population mean μ?

Answer

Answer: Since the sample mean $\bar{X}$ is approximately normal $N(\mu, \frac{\sigma^2}{n})$:

$$\mathbb{P}(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95.$$

This example is very important for the following lectures.

Overview

- Parameter Estimation
- Confidence Interval

An Overview of Statistical Inference Problems

- Making probabilistic statements about an unknown population parameter based on a random sample from the population
- Estimation
  - Point estimation: estimate the value of the unknown parameter
  - Confidence interval (CI): estimate an interval in which the parameter lies (how accurate is the estimate)
- Hypothesis Testing
  - Make decision (yes/no) on a hypothetical statement about the parameter

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

- An estimator $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

- An estimator $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$
  - A statistic computed from the sample data

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

- An estimator $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$
  - A statistic computed from the sample data
  - Examples: $\theta = \mu$, $\hat{\theta} = \bar{X}; \theta = \sigma^2, \hat{\theta} = S^2$.

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

- An estimator $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$
  - A statistic computed from the sample data
  - Examples: $\theta = \mu$, $\hat{\theta} = \bar{X}; \theta = \sigma^2, \hat{\theta} = S^2$.
  - Convention: When we are estimating a parameter (eg. μ, p, σ^2), we denote its estimate by putting a hat on this parameter (eg. $\hat{\mu}$, $\hat{p}$, $\widehat{\sigma^2}$).

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

- An estimator $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$
  - A statistic computed from the sample data
  - Examples: $\theta = \mu$, $\hat{\theta} = \bar{X}; \theta = \sigma^2, \hat{\theta} = S^2$.
  - Convention: When we are estimating a parameter (eg. μ, p, σ^2), we denote its estimate by putting a hat on this parameter (eg. $\hat{\mu}$, $\hat{p}$, $\widehat{\sigma^2}$).
- Estimator vs. Estimate

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

- An estimator $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$
  - A statistic computed from the sample data
  - Examples: $\theta = \mu$, $\hat{\theta} = \bar{X}; \theta = \sigma^2, \hat{\theta} = S^2$.
  - Convention: When we are estimating a parameter (eg. μ, p, σ^2), we denote its estimate by putting a hat on this parameter (eg. $\hat{\mu}$, $\hat{p}$, $\widehat{\sigma^2}$).
- Estimator vs. Estimate
  - Estimator: a random variable, a rule to compute the desired value from data

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

- An estimator $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$
  - A statistic computed from the sample data
  - Examples: $\theta = \mu$, $\hat{\theta} = \bar{X}; \theta = \sigma^2, \hat{\theta} = S^2$.
  - Convention: When we are estimating a parameter (eg. μ, p, σ^2), we denote its estimate by putting a hat on this parameter (eg. $\hat{\mu}$, $\hat{p}$, $\widehat{\sigma^2}$).
- Estimator vs. Estimate
  - Estimator: a random variable, a rule to compute the desired value from data
  - Estimate: the specific value obtained using the rule on the observed data

Point estimation

- Basic setup: a random sample from a population with unknown parameter θ:

$$X_1, ..., X_n \overset{i.i.d.}{\sim} F_\theta$$

- An estimator $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$
  - A statistic computed from the sample data
  - Examples: $\theta = \mu$, $\hat{\theta} = \bar{X}; \theta = \sigma^2, \hat{\theta} = S^2$.
  - Convention: When we are estimating a parameter (eg. μ, p, σ^2), we denote its estimate by putting a hat on this parameter (eg. $\hat{\mu}$, $\hat{p}$, $\widehat{\sigma^2}$).
- Estimator vs. Estimate
  - Estimator: a random variable, a rule to compute the desired value from data
  - Estimate: the specific value obtained using the rule on the observed data
  - Example: $\bar{X}$ vs. $\bar{x}$

Quality of a Point Estimator

- Bias

Quality of a Point Estimator

- Bias
  - Definition: $Bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$, measuring average performance

Quality of a Point Estimator

- Bias
  - Definition: $Bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$, measuring average performance
  - An estimator is unbiased if $Bias(\hat{\theta}) = 0$, i.e., $E(\hat{\theta}) = \theta$

Quality of a Point Estimator

- Bias
  - Definition: $Bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$, measuring average performance
  - An estimator is unbiased if $Bias(\hat{\theta}) = 0$, i.e., $E(\hat{\theta}) = \theta$
- Variance

Quality of a Point Estimator

- Bias
  - Definition: $Bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$, measuring average performance
  - An estimator is unbiased if $Bias(\hat{\theta}) = 0$, i.e., $E(\hat{\theta}) = \theta$
- Variance
  - Definition: $Var(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$

Quality of a Point Estimator

- Bias
  - Definition: $\mathit{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$, measuring average performance
  - An estimator is unbiased if $\mathit{Bias}(\hat{\theta}) = 0$, i.e., $E(\hat{\theta}) = \theta$
- Variance
  - Definition: $\mathit{Var}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$
  - Precision is the reciprocal of variance

Quality of a Point Estimator

- Bias
  - Definition: $Bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$, measuring average performance
  - An estimator is unbiased if $Bias(\hat{\theta}) = 0$, i.e., $E(\hat{\theta}) = \theta$
- Variance
  - Definition: $Var(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$
  - Precision is the reciprocal of variance
- Mean squared error (MSE)

Quality of a Point Estimator

- Bias
  - Definition: $Bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$, measuring average performance
  - An estimator is unbiased if $Bias(\hat{\theta}) = 0$, i.e., $E(\hat{\theta}) = \theta$
- Variance
  - Definition: $Var(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$
  - Precision is the reciprocal of variance
- Mean squared error (MSE)
  - Definition: $MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$, i.e., the expected squared error loss

Quality of a Point Estimator

- Bias
  - Definition: $Bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$, measuring average performance
  - An estimator is unbiased if $Bias(\hat{\theta}) = 0$, i.e., $E(\hat{\theta}) = \theta$
- Variance
  - Definition: $Var(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2]$
  - Precision is the reciprocal of variance
- Mean squared error (MSE)
  - Definition: $MSE(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$, i.e., the expected squared error loss
  - Basic identity:

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + [Bias(\hat{\theta})]^2$$

Quality of a Point Estimator

Example: Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.

Estimate μ by $\bar{X}$

- Bias: $Bias(\bar{X}) = \mathbb{E}[\bar{X}] - \mu = 0$, i.e. $\bar{X}$ is unbiased

Quality of a Point Estimator

Example: Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.

Estimate μ by $\bar{X}$

- Bias: $Bias(\bar{X}) = \mathbb{E}[\bar{X}] - \mu = 0$, i.e. $\bar{X}$ is unbiased
- Variance: $Var(\bar{X}) = \frac{\sigma^2}{n}$

Quality of a Point Estimator

Example: Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.

Estimate μ by $\bar{X}$

- Bias: $Bias(\bar{X}) = \mathbb{E}[\bar{X}] - \mu = 0$, i.e. $\bar{X}$ is unbiased
- Variance: $Var(\bar{X}) = \frac{\sigma^2}{n}$
- MSE: $MSE(\bar{X}) = Var((\bar{X}) + [Bias(\bar{X})]^2 = Var(\bar{X}) = \frac{\sigma^2}{n}$

Example

Example

Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p} = \frac{Y}{n}$. Compute $MSE(\hat{p})$

Example

Example

Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p} = \frac{Y}{n}$. Compute $MSE(\hat{p})$

Answer

- Bias: $Bias(\hat{p}) = \mathbb{E}[\hat{p}] - p = 0$, i.e. $\hat{p}$ is unbiased

Example

Example

Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p} = \frac{Y}{n}$. Compute $MSE(\hat{p})$

Answer

- Bias: $Bias(\hat{p}) = \mathbb{E}[\hat{p}] - p = 0$, i.e. $\hat{p}$ is unbiased
- Variance: $Var(\hat{p}) = \frac{p(1-p)}{n}$

Example

Example

Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p} = \frac{Y}{n}$. Compute $MSE(\hat{p})$

Answer

- Bias: $Bias(\hat{p}) = \mathbb{E}[\hat{p}] - p = 0$, i.e. $\hat{p}$ is unbiased
- Variance: $Var(\hat{p}) = \frac{p(1-p)}{n}$
- MSE: $MSE(\hat{p}) = Var(\hat{p}) + [Bias(\hat{p})]^2 = Var(\hat{p}) = \frac{p(1-p)}{n}$

Example

Example

Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p}^* = \frac{Y+1}{n+2}$. Compute $MSE(\hat{p}^*)$.

Example

Example

Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p}^* = \frac{Y+1}{n+2}$. Compute $MSE(\hat{p}^*)$.

Answer

- We have $\mathbb{E}[Y] = np$, $Var(Y) = np(1-p)$

Example

Example

Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p}^* = \frac{Y+1}{n+2}$. Compute $MSE(\hat{p}^*)$.

Answer

- We have $\mathbb{E}[Y] = np$, $Var(Y) = np(1-p)$
- Bias: $Bias(\hat{p}^*) = \mathbb{E}[\frac{Y+1}{n+2}] - p = \frac{np+1}{n+2} - p = \frac{1-2p}{n+2}$

Example

Example

Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p}^* = \frac{Y+1}{n+2}$. Compute $MSE(\hat{p}^*)$.

Answer

- We have $\mathbb{E}[Y] = np$, $Var(Y) = np(1-p)$
- Bias: $Bias(\hat{p}^*) = \mathbb{E}[\frac{Y+1}{n+2}] - p = \frac{np+1}{n+2} - p = \frac{1-2p}{n+2}$
- Variance: $Var(\hat{p}^*) = Var(\frac{Y}{n+2}) = \frac{np(1-p)}{(n+2)^2}$

Example

Example

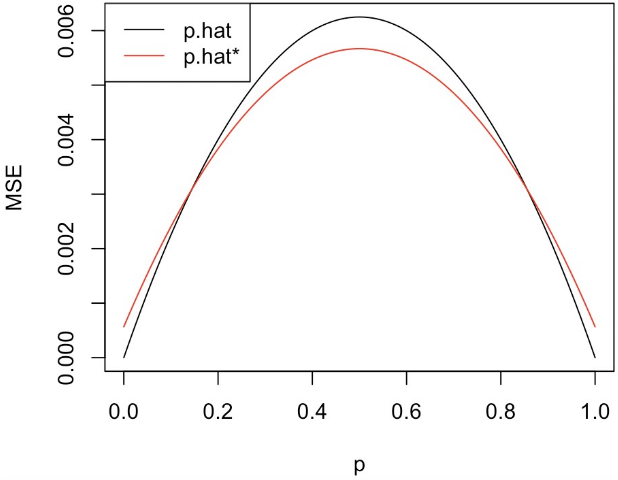
Let Y be a single observation from a binomial distribution with known n and unknown p (success probability). We wish to estimate p. Let $\hat{p}^* = \frac{Y+1}{n+2}$. Compute $MSE(\hat{p}^*)$.

Answer

- We have $\mathbb{E}[Y] = np$, $Var(Y) = np(1-p)$
- Bias: $Bias(\hat{p}^*) = \mathbb{E}[\frac{Y+1}{n+2}] - p = \frac{np+1}{n+2} - p = \frac{1-2p}{n+2}$
- Variance: $Var(\hat{p}^*) = Var(\frac{Y}{n+2}) = \frac{np(1-p)}{(n+2)^2}$
- MSE: $MSE(\hat{p}^*) = Var(\hat{p}^*) + [Bias(\hat{p}^*)]^2 = \frac{np(1-p)}{(n+2)^2} + \frac{(1-2p)^2}{(n+2)^2} = \frac{np-np^2+1+4p^2-4p}{(n+2)^2}$

Comparison

For a given $n = 40$, plot $MSE(\hat{p})$ and $MSE(\hat{p}^*)$ against p.



Some common unbiased estimators

- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.

Some common unbiased estimators

- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.
  - μ: $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + ... + X_n)$

Some common unbiased estimators

- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.
  - μ: $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + ... + X_n)$
  - σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Some common unbiased estimators

- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.
  - μ: $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + ... + X_n)$
  - σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - Note: $S = \sqrt{S^2}$ is not an unbiased estimator for σ.

Some common unbiased estimators

- Suppose we have $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.
  - μ: $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$
  - σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - Note: $S = \sqrt{S^2}$ is not an unbiased estimator for σ.
- Suppose we have $X_1, \ldots, X_n \overset{i.i.d.}{\sim} Ber(p)$.

Some common unbiased estimators

- Suppose we have $X_1, \ldots, X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.
  - μ: $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$
  - σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - Note: $S = \sqrt{S^2}$ is not an unbiased estimator for σ.
- Suppose we have $X_1, \ldots, X_n \overset{i.i.d.}{\sim} Ber(p)$.
  - p: $\hat{p} = \frac{1}{n}(X_1 + \ldots + X_n)$

Some common unbiased estimators

- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.
  - μ: $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + ... + X_n)$
  - σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - Note: $S = \sqrt{S^2}$ is not an unbiased estimator for σ.
- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} Ber(p)$.
  - p: $\hat{p} = \frac{1}{n}(X_1 + ... + X_n)$
- Two independent samples: $X_1, ..., X_{n_1} \overset{i.i.d.}{\sim} P_1$ with mean μ_1 and σ^2, $Y_1, ..., Y_{n_2} \overset{i.i.d.}{\sim} P_2$ with mean μ_2 and σ^2.

Some common unbiased estimators

- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.
  - μ: $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + ... + X_n)$
  - σ^2: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - Note: $S = \sqrt{S^2}$ is not an unbiased estimator for σ.
- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} Ber(p)$.
  - p: $\hat{p} = \frac{1}{n}(X_1 + ... + X_n)$
- Two independent samples: $X_1, ..., X_{n_1} \overset{i.i.d.}{\sim} P_1$ with mean μ_1 and σ^2, $Y_1, ..., Y_{n_2} \overset{i.i.d.}{\sim} P_2$ with mean μ_2 and σ^2.
  - $\mu_1 - \mu_2$: $\hat{\mu}_1 - \hat{\mu}_2 = \bar{X} - \bar{Y}$

Some common unbiased estimators

- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} P$ with mean μ and σ^2.
  - μ: $\hat{\mu} = \bar{X} = \frac{1}{n}(X_1 + ... + X_n)$
  - σ^2: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
  - Note: $S = \sqrt{S^2}$ is not an unbiased estimator for σ.
- Suppose we have $X_1, ..., X_n \overset{i.i.d.}{\sim} Ber(p)$.
  - p: $\hat{p} = \frac{1}{n}(X_1 + ... + X_n)$
- Two independent samples: $X_1, ..., X_{n_1} \overset{i.i.d.}{\sim} P_1$ with mean μ_1 and σ^2, $Y_1, ..., Y_{n_2} \overset{i.i.d.}{\sim} P_2$ with mean μ_2 and σ^2.
  - $\mu_1 - \mu_2$: $\hat{\mu}_1 - \hat{\mu}_2 = \bar{X} - \bar{Y}$
  - σ^2: $S_{pool}^2 = \frac{1}{n_1 + n_2 - 2} (\sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2)$

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ
- If $\hat{\theta}$ has an approximately normal sampling distribution (eg. by CLT) with mean θ and variance $\sigma^2_{\hat{\theta}}$.

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ
- If $\hat{\theta}$ has an approximately normal sampling distribution (eg. by CLT) with mean θ and variance $\sigma^2_{\hat{\theta}}$.
- Then $\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \approx N(0, 1)$

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ
- If $\hat{\theta}$ has an approximately normal sampling distribution (eg. by CLT) with mean θ and variance $\sigma^2_{\hat{\theta}}$.
- Then $\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \approx N(0,1)$
- As a result, $\mathbb{P}(\epsilon \leq c) = \mathbb{P}(\frac{|\hat{\theta}-\theta|}{\sigma_{\hat{\theta}}} \leq \frac{c}{\sigma_{\hat{\theta}}}) = \mathbb{P}(|Z| \leq \frac{c}{\sigma_{\hat{\theta}}})$

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ
- If $\hat{\theta}$ has an approximately normal sampling distribution (eg. by CLT) with mean θ and variance $\sigma^2_{\hat{\theta}}$.
- Then $\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \approx N(0,1)$
- As a result, $\mathbb{P}(\epsilon \leq c) = \mathbb{P}(\frac{|\hat{\theta}-\theta|}{\sigma_{\hat{\theta}}} \leq \frac{c}{\sigma_{\hat{\theta}}}) = \mathbb{P}(|Z| \leq \frac{c}{\sigma_{\hat{\theta}}})$
- By basic normal probabilities, $\mathbb{P}(|Z| \leq 2) = 0.95$

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ
- If $\hat{\theta}$ has an approximately normal sampling distribution (eg. by CLT) with mean θ and variance $\sigma^2_{\hat{\theta}}$.
- Then $\frac{\hat{\theta}-\theta}{\sigma_{\hat{\theta}}} \approx N(0,1)$
- As a result, $\mathbb{P}(\epsilon \leq c) = \mathbb{P}(\frac{|\hat{\theta}-\theta|}{\sigma_{\hat{\theta}}} \leq \frac{c}{\sigma_{\hat{\theta}}}) = \mathbb{P}(|Z| \leq \frac{c}{\sigma_{\hat{\theta}}})$
- By basic normal probabilities, $\mathbb{P}(|Z| \leq 2) = 0.95$
- As a result, $\epsilon \in (0, 2\sigma_{\hat{\theta}})$ forms an approximate 95% bound on ϵ: $\mathbb{P}(\epsilon \in (0, 2\sigma_{\hat{\theta}})) = 0.95$

Example

Example

Suppose there is an unknown population with mean μ and variance 5^2, To estimate the mean, we take a random sample of size 49, and we calculate $\bar{Y}$. What is an approximate 95% bound on $\epsilon = |\bar{Y} - \mu|$?

Example

Example

Suppose there is an unknown population with mean μ and variance 5^2, To estimate the mean, we take a random sample of size 49, and we calculate $\bar{Y}$. What is an approximate 95% bound on $\epsilon = |\bar{Y} - \mu|$?

Answer

- By CLT, $\bar{Y}$ is approximately normal
- As a result, we use the formula $(0, 2\sigma_{\hat{\theta}})$

Example

Example

Suppose there is an unknown population with mean μ and variance 5^2, To estimate the mean, we take a random sample of size 49, and we calculate $\bar{Y}$. What is an approximate 95% bound on $\epsilon = |\bar{Y} - \mu|$?

Answer

- By CLT, $\bar{Y}$ is approximately normal
- As a result, we use the formula $(0, 2\sigma_{\hat{\theta}})$
- $\sigma_{\hat{\theta}} = \frac{5}{7}$

Example

Example

Suppose there is an unknown population with mean μ and variance 5^2, To estimate the mean, we take a random sample of size 49, and we calculate $\bar{Y}$. What is an approximate 95% bound on $\epsilon = |\bar{Y} - \mu|$?

Answer

- By CLT, $\bar{Y}$ is approximately normal
- As a result, we use the formula $(0, 2\sigma_{\hat{\theta}})$
- $\sigma_{\hat{\theta}} = \frac{5}{7}$
- The 95% bound is $(0, \frac{10}{7})$.

Example

Example

A comparison of the durability of two types of automobile tires was obtained by road testing samples of $n_1 = n_2 = 100$ tires of each type. The number of miles until wear-out was recorded, where wear-out was defined as the number of miles until the amount of remaining tread reached a prespecified small value. The measurements for the two types of tires were obtained independently, and the following means and variances were computed:

$$\bar{y}_1 = 26,400 \text{ miles}, \bar{y}_2 = 25,100 \text{ miles},$$

$$s_1^2 = 1,440,000, s_2^2 = 1,960,000.$$

Estimate the difference in mean miles to wear-out and place a 2-standard-error bound on the error of estimation.

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ
- If $\hat{\theta}$ does not have an approximately normal sampling distribution, we can still get a conservative bound on ϵ

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ
- If $\hat{\theta}$ does not have an approximately normal sampling distribution, we can still get a conservative bound on ϵ
- Chebyshev's inequality, if $\mathbb{E}[\hat{\theta}] = \theta$

$$\mathbb{P}(|\frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}| > k) \leq \frac{1}{k^2}$$

Estimation error bounds

- Let $\epsilon = |\hat{\theta} - \theta|$ be the error in estimation for an estimator $\hat{\theta}$ of θ
- If $\hat{\theta}$ does not have an approximately normal sampling distribution, we can still get a conservative bound on ϵ
- Chebyshev's inequality, if $\mathbb{E}[\hat{\theta}] = \theta$

$$\mathbb{P}(|\frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}| > k) \leq \frac{1}{k^2}$$

- Let $\frac{1}{k^2} = 0.05$ (i.e. $k = \sqrt{20} = 4.47$), a 95% bound is then $\epsilon \in (0, 4.47\sigma_{\hat{\theta}})$.

Estimation of the population mean

- When $X_1, ..., X_n$ is an *i.i.d.* sample from a population with mean μ and variance σ^2

Estimation of the population mean

- When $X_1, ..., X_n$ is an *i.i.d.* sample from a population with mean μ and variance σ^2
- By the Law of Large Numbers, we can estimate μ by the sample mean $\bar{X} = \frac{X_1 + ... + X_n}{n}$. Denote it by $\hat{\mu} = \bar{X}$.

Estimation of the population mean

- When $X_1, ..., X_n$ is an *i.i.d.* sample from a population with mean μ and variance σ^2
- By the Law of Large Numbers, we can estimate μ by the sample mean $\bar{X} = \frac{X_1 + ... + X_n}{n}$. Denote it by $\hat{\mu} = \bar{X}$.
- You could give a single number as an estimate, but you would almost certainly be wrong.

Estimation of the population mean

- When $X_1, ..., X_n$ is an $i.i.d.$ sample from a population with mean μ and variance σ^2
- By the Law of Large Numbers, we can estimate μ by the sample mean $\bar{X} = \frac{X_1 + ... + X_n}{n}$. Denote it by $\hat{\mu} = \bar{X}$.
- You could give a single number as an estimate, but you would almost certainly be wrong.
- You could give a range of numbers, which is more realistic.

Estimation of the population mean

- When $X_1, ..., X_n$ is an $i.i.d.$ sample from a population with mean μ and variance σ^2
- By the Law of Large Numbers, we can estimate μ by the sample mean $\bar{X} = \frac{X_1 + ... + X_n}{n}$. Denote it by $\hat{\mu} = \bar{X}$.
- You could give a single number as an estimate, but you would almost certainly be wrong.
- You could give a range of numbers, which is more realistic.
- Or, you could give a range of numbers with a statement which conveys your confidence in the interval itself.

From Point to Interval

- Point estimation: from the sample $X_1, ..., X_n$ to a random variable $\hat{\theta}$

From Point to Interval

- Point estimation: from the sample $X_1, ..., X_n$ to a random variable $\hat{\theta}$
- Confidence interval: from the sample $X_1, ..., X_n$ to a random interval $[L, U]$

From Point to Interval

- Point estimation: from the sample $X_1, ..., X_n$ to a random variable $\hat{\theta}$
- Confidence interval: from the sample $X_1, ..., X_n$ to a random interval $[L, U]$
- Both L and U are random, i.e., $L = L(X_1, ..., X_n)$, $U = U(X_1, ..., X_n)$ They are called the confidence limits.

From Point to Interval

- Point estimation: from the sample $X_1, ..., X_n$ to a random variable $\hat{\theta}$
- Confidence interval: from the sample $X_1, ..., X_n$ to a random interval $[L, U]$
- Both L and U are random, i.e., $L = L(X_1, ..., X_n)$, $U = U(X_1, ..., X_n)$ They are called the confidence limits.
- A confidence interval $[L, U]$ is a $100(1-\alpha)\%$ CI if

$$\mathbb{P}_\theta(L \leq \theta \leq U) \geq 1 - \alpha$$

From Point to Interval

- Point estimation: from the sample $X_1, ..., X_n$ to a random variable $\hat{\theta}$
- Confidence interval: from the sample $X_1, ..., X_n$ to a random interval $[L, U]$
- Both L and U are random, i.e., $L = L(X_1, ..., X_n)$, $U = U(X_1, ..., X_n)$ They are called the confidence limits.
- A confidence interval $[L, U]$ is a $100(1-\alpha)\%$ CI if

$$\mathbb{P}_\theta(L \le \theta \le U) \ge 1 - \alpha$$

- After a sample is collected, we observe $x_1, ... x_n$. The confidence limits $= L(x_1, ..., x_n)$, $u = U(x_1, ..., x_n)$ can be computed as numbers.

From Point to Interval

- Point estimation: from the sample $X_1, ..., X_n$ to a random variable $\hat{\theta}$
- Confidence interval: from the sample $X_1, ..., X_n$ to a random interval $[L, U]$
- Both L and U are random, i.e., $L = L(X_1, ..., X_n)$, $U = U(X_1, ..., X_n)$ They are called the confidence limits.
- A confidence interval $[L, U]$ is a $100(1-\alpha)\%$ CI if

$$\mathbb{P}_\theta(L \leq \theta \leq U) \geq 1 - \alpha$$

- After a sample is collected, we observe $x_1, ...x_n$. The confidence limits $= L(x_1, ..., x_n)$, $u = U(x_1, ..., x_n)$ can be computed as numbers.
- It is also possible to form a one-sided confidence interval such that:

From Point to Interval

- Point estimation: from the sample $X_1, ..., X_n$ to a random variable $\hat{\theta}$
- Confidence interval: from the sample $X_1, ..., X_n$ to a random interval $[L, U]$
- Both L and U are random, i.e., $L = L(X_1, ..., X_n)$, $U = U(X_1, ..., X_n)$ They are called the confidence limits.
- A confidence interval $[L, U]$ is a $100(1-\alpha)\%$ CI if
$$\mathbb{P}_\theta(L \leq \theta \leq U) \geq 1 - \alpha$$
- After a sample is collected, we observe $x_1, ...x_n$. The confidence limits $= L(x_1, ..., x_n)$, $u = U(x_1, ..., x_n)$ can be computed as numbers.
- It is also possible to form a one-sided confidence interval such that:
    - upper one-sided: $\mathbb{P}_\theta(\theta \leq U) \geq 1 - \alpha$

From Point to Interval

- Point estimation: from the sample $X_1, ..., X_n$ to a random variable $\hat{\theta}$
- Confidence interval: from the sample $X_1, ..., X_n$ to a random interval $[L, U]$
- Both L and U are random, i.e., $L = L(X_1, ..., X_n)$, $U = U(X_1, ..., X_n)$ They are called the confidence limits.
- A confidence interval $[L, U]$ is a $100(1-\alpha)\%$ CI if
$$\mathbb{P}_\theta(L \leq \theta \leq U) \geq 1 - \alpha$$
- After a sample is collected, we observe $x_1, ...x_n$. The confidence limits $= L(x_1, ..., x_n)$, $u = U(x_1, ..., x_n)$ can be computed as numbers.
- It is also possible to form a one-sided confidence interval such that:
  - upper one-sided: $\mathbb{P}_\theta(\theta \leq U) \geq 1 - \alpha$
  - lower one-sided: $\mathbb{P}_\theta(\theta \geq L) \geq 1 - \alpha$

From Point to Interval

- Point estimation: from the sample $X_1, \ldots, X_n$ to a random variable $\hat{\theta}$
- Confidence interval: from the sample $X_1, \ldots, X_n$ to a random interval $[L, U]$
- Both L and U are random, i.e., $L = L(X_1, \ldots, X_n)$, $U = U(X_1, \ldots, X_n)$ They are called the confidence limits.
- A confidence interval $[L, U]$ is a $100(1-\alpha)\%$ CI if
$$\mathbb{P}_\theta(L \leq \theta \leq U) \geq 1 - \alpha$$
- After a sample is collected, we observe $x_1, \ldots x_n$. The confidence limits $= L(x_1, \ldots, x_n)$, $u = U(x_1, \ldots, x_n)$ can be computed as numbers.
- It is also possible to form a one-sided confidence interval such that:
  - upper one-sided: $\mathbb{P}_\theta(\theta \leq U) \geq 1 - \alpha$
  - lower one-sided: $\mathbb{P}_\theta(\theta \geq L) \geq 1 - \alpha$
- How to construct these intervals?

Estimation of the binomial parameter p

Let's consider a special example of population mean estimation. If we have an $i.i.d.$ sample $X_1, ..., X_n$ with mean μ and variance σ^2, how to estimate μ?

- Recall that $\bar{X} = \sum_{i=1}^{n} X_i$, and by CLT, $\bar{X}$ is approximately $N(\mu, \sigma^2/n)$

Estimation of the binomial parameter p

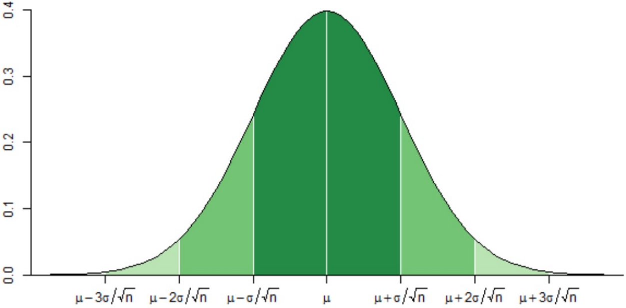
Let's consider a special example of population mean estimation. If we have an $i.i.d.$ sample $X_1, ..., X_n$ with mean μ and variance σ^2, how to estimate μ?

- Recall that $\bar{X} = \sum_{i=1}^{n} X_i$, and by CLT, $\bar{X}$ is approximately $N(\mu, \sigma^2/n)$
- How to construct an interval that contains μ with large probability?

The sampling distribution of $\bar{X}$

Since $\bar{X} \approx N(\mu, \frac{\sigma^2}{n})$, CLT enables us to make the following statements

1. $\mathbb{P}(\mu - \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + \frac{\sigma}{\sqrt{n}}) \approx 0.68$
2. $\mathbb{P}(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$
3. $\mathbb{P}(\mu - 3\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 3\frac{\sigma}{\sqrt{n}}) \approx 0.997$



The key idea

- Assuming that n is large enough. By the CLT and the Empirical Rule, we believe there is a 95% chance that a new $\bar{X}$ will be within $2\sigma/\sqrt{n}$ from μ:

$$\mathbb{P}(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$$

The key idea

- Assuming that n is large enough. By the CLT and the Empirical Rule, we believe there is a 95% chance that a new $\bar{X}$ will be within $2\sigma/\sqrt{n}$ from μ:

$$\mathbb{P}(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$$

- We don't know what μ is, but we can invert the previous statement to say that there is a 95% chance that μ is within $2\sigma/\sqrt{n}$ from $\bar{X}$:

$$\mathbb{P}(\bar{X} - 2\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$$

The key idea

- Assuming that n is large enough. By the CLT and the Empirical Rule, we believe there is a 95% chance that a new $\bar{X}$ will be within $2\sigma/\sqrt{n}$ from μ:

$$\mathbb{P}(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$$

- We don't know what μ is, but we can invert the previous statement to say that there is a 95% chance that μ is within $2\sigma/\sqrt{n}$ from $\bar{X}$:

$$\mathbb{P}(\bar{X} - 2\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$$

- Putting this statement into a formula: $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$ provides a 95% confidence interval for μ.

The key idea

- Assuming that n is large enough. By the CLT and the Empirical Rule, we believe there is a 95% chance that a new $\bar{X}$ will be within $2\sigma/\sqrt{n}$ from μ:

$$\mathbb{P}(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$$

- We don't know what μ is, but we can invert the previous statement to say that there is a 95% chance that μ is within $2\sigma/\sqrt{n}$ from $\bar{X}$:

$$\mathbb{P}(\bar{X} - 2\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 2\frac{\sigma}{\sqrt{n}}) \approx 0.95$$

- Putting this statement into a formula: $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$ provides a 95% confidence interval for μ.
- Confidence Intervals do not have to be 95%. 95% is just a convention. If you want something else (say 90%) then switch the 2 to 1.645.

Confidence interval

- When $X_1, ..., X_n$ is an *i.i.d.* sample from a population, and n is large,

$$[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$$

is called an (approximate) 95% confidence interval (CI) for the population mean.

Confidence interval

- When $X_1, ..., X_n$ is an $i.i.d.$ sample from a population, and n is large,

$$[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$$

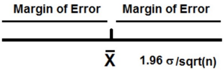
  is called an (approximate) 95% confidence interval (CI) for the population mean.
- Another definition: $2\frac{\sigma}{\sqrt{n}}$ is called Margin of Error (MoE).

Confidence interval

- When $X_1, ..., X_n$ is an $i.i.d.$ sample from a population, and n is large,

$$[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$$

  is called an (approximate) 95% confidence interval (CI) for the population mean.
- Another definition: $2\frac{\sigma}{\sqrt{n}}$ is called Margin of Error (MoE).
- The MoE is the distance from the center to the edge of the interval.

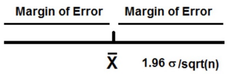


Confidence interval

- When $X_1, ..., X_n$ is an *i.i.d.* sample from a population, and n is large,

$$[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$$

  is called an (approximate) 95% confidence interval (CI) for the population mean.
- Another definition: $2\frac{\sigma}{\sqrt{n}}$ is called Margin of Error (MoE).
- The MoE is the distance from the center to the edge of the interval.



- Remark: Keep this in mind, the MoE is the SD of the estimate multiplied by some constant, where $sd(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

Interpreting a confidence interval

- What does the 95% confidence in the interval really mean? You must **not** say: There is a 95% probability that μ lies in the interval.
- You can't say this because μ is not a random variable, just some fixed but unknown number. Therefore there are no probability statements to be made about it.

Interpreting a confidence interval

- What does the 95% confidence in the interval really mean? You must **not** say: There is a 95% probability that μ lies in the interval.
- You can't say this because μ is not a random variable, just some fixed but unknown number. Therefore there are no probability statements to be made about it.
- The 95% is a property of the procedure, not a specific interval. You can say *In approximately 95% of all samples the confidence interval created according to this procedure will contain μ.*

Interpreting a confidence interval

- What does the 95% confidence in the interval really mean? You must **not** say: There is a 95% probability that μ lies in the interval.
- You can't say this because μ is not a random variable, just some fixed but unknown number. Therefore there are no probability statements to be made about it.
- The 95% is a property of the procedure, not a specific interval. You can say *In approximately 95% of all samples the confidence interval created according to this procedure will contain μ.*
- In real life, we only get one sample. For this sample, the confidence interval either covers the true parameter or it does not cover. After the sample is taken, there is no chance only *confidence*. The word *confidence* used in this context is a term of art.

Interpreting a confidence interval

- What does the 95% confidence in the interval really mean? You must **not** say: There is a 95% probability that μ lies in the interval.
- You can't say this because μ is not a random variable, just some fixed but unknown number. Therefore there are no probability statements to be made about it.
- The 95% is a property of the procedure, not a specific interval. You can say *In approximately 95% of all samples the confidence interval created according to this procedure will contain μ.*
- In real life, we only get one sample. For this sample, the confidence interval either covers the true parameter or it does not cover. After the sample is taken, there is no chance only *confidence*. The word *confidence* used in this context is a term of art.
- An illustration applet: `http://www.rossmanchance.com/applets/ConfSim.html`

Example

The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. What is a 95% CI for the true mean?

Example

The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. What is a 95% CI for the true mean?

Answer

- $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$

Example

The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. What is a 95% CI for the true mean?

Answer

- $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$
- $12.8 \pm 2 \cdot 2/\sqrt{100} = [12.4, 13.2]$.

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.
- You can be 95% confident that the true mean is somewhere between 12.4 and 13.2.

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.
- You can be 95% confident that the true mean is somewhere between 12.4 and 13.2.
- Correct interpretations:

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.
- You can be 95% confident that the true mean is somewhere between 12.4 and 13.2.
- Correct interpretations:
  1. in 95% of samples, the true mean will be covered by the confidence interval $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$: i.e. μ is in the interval.

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.
- You can be 95% confident that the true mean is somewhere between 12.4 and 13.2.
- Correct interpretations:
  1. in 95% of samples, the true mean will be covered by the confidence interval $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$: i.e. μ is in the interval.
  2. every sample has a 95% chance of producing a confidence interval that covers the truth.

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.
- You can be 95% confident that the true mean is somewhere between 12.4 and 13.2.
- Correct interpretations:
  1. in 95% of samples, the true mean will be covered by the confidence interval $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$: i.e. μ is in the interval.
  2. every sample has a 95% chance of producing a confidence interval that covers the truth.
- Incorrect but tempting interpretations:

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.
- You can be 95% confident that the true mean is somewhere between 12.4 and 13.2.
- Correct interpretations:
  1. in 95% of samples, the true mean will be covered by the confidence interval $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$: i.e. μ is in the interval.
  2. every sample has a 95% chance of producing a confidence interval that covers the truth.
- Incorrect but tempting interpretations:
  1. There is a 95% chance that the true mean is in the interval $[12.4, 13.2]$.

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.
- You can be 95% confident that the true mean is somewhere between 12.4 and 13.2.
- Correct interpretations:
  1. in 95% of samples, the true mean will be covered by the confidence interval $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$: i.e. μ is in the interval.
  2. every sample has a 95% chance of producing a confidence interval that covers the truth.
- Incorrect but tempting interpretations:
  1. There is a 95% chance that the true mean is in the interval $[12.4, 13.2]$.
  2. In 95% of samples of size 100, the true mean will be covered by the interval [12.4, 13.2].

Example: Interpreting a confidence interval

- The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. A 95% CI for the true mean is $[12.4, 13.2]$.
- You can be 95% confident that the true mean is somewhere between 12.4 and 13.2.
- Correct interpretations:
  1. in 95% of samples, the true mean will be covered by the confidence interval $[\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}]$: i.e. μ is in the interval.
  2. every sample has a 95% chance of producing a confidence interval that covers the truth.
- Incorrect but tempting interpretations:
  1. There is a 95% chance that the true mean is in the interval $[12.4, 13.2]$.
  2. In 95% of samples of size 100, the true mean will be covered by the interval [12.4, 13.2].
  3. 95% of samples of size 100 will have sample means between 12.4 and 13.2.

Example: Interpreting a confidence interval

- KEY FACT: since the sample mean is random, the confidence interval is random.
- If two people draw two independent samples, the resulting confidence intervals will not be the same.

More General Confidence Intervals

- The length of the confidence interval is determined by how accurate you wish to be. A 2 standard deviation (in each direction) CI is approximately 95% accurate.

More General Confidence Intervals

- The length of the confidence interval is determined by how accurate you wish to be. A 2 standard deviation (in each direction) CI is approximately 95% accurate.
- In general, you can go $z_{\alpha/2}$ standard deviation away from $\bar{X}$ in each direction, where $z_{\alpha/2}$ is the value such that $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$.
- $z_{\alpha/2}$ is the critical value with area $(1 - \alpha)$ between $-z_{\alpha/2}$ and $z_{\alpha/2}$ on the Normal curve: $\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$.

More General Confidence Intervals

- The length of the confidence interval is determined by how accurate you wish to be. A 2 standard deviation (in each direction) CI is approximately 95% accurate.
- In general, you can go $z_{\alpha/2}$ standard deviation away from $\bar{X}$ in each direction, where $z_{\alpha/2}$ is the value such that $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$.
- $z_{\alpha/2}$ is the critical value with area $(1-\alpha)$ between $-z_{\alpha/2}$ and $z_{\alpha/2}$ on the Normal curve: $\mathbb{P}(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1-\alpha$.
- Definition : A $(1-\alpha)$ level confidence interval for μ is $[\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n}, \ \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}]$.

More General Confidence Intervals

- The length of the confidence interval is determined by how accurate you wish to be. A 2 standard deviation (in each direction) CI is approximately 95% accurate.
- In general, you can go $z_{\alpha/2}$ standard deviation away from $\bar{X}$ in each direction, where $z_{\alpha/2}$ is the value such that $\mathbb{P}(Z > z_{\alpha/2}) = \alpha/2$.
- $z_{\alpha/2}$ is the critical value with area $(1-\alpha)$ between $-z_{\alpha/2}$ and $z_{\alpha/2}$ on the Normal curve: $\mathbb{P}(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$.
- Definition : A $(1-\alpha)$ level confidence interval for μ is $[\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n},\ \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}]$.
- If you want a CI with higher confidence $1-\alpha$, you should increase the length of CI

More General Confidence Intervals

A $(1-\alpha)$ level confidence interval for μ is $[\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n},\ \ \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}]$.

Examples:

- A more accurate 95% confidence interval for μ is $[\bar{X} - 1.96 \cdot \sigma/\sqrt{n},\ \ \bar{X} + 1.96 \cdot \sigma/\sqrt{n}]$, since $\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95$.

More General Confidence Intervals

A $(1-\alpha)$ level confidence interval for μ is $[\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n},\ \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}]$.

Examples:

- A more accurate 95% confidence interval for μ is $[\bar{X} - 1.96 \cdot \sigma/\sqrt{n},\ \bar{X} + 1.96 \cdot \sigma/\sqrt{n}]$, since $\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95$.
- If we are going to construct a 99% confidence interval for μ, then we need to look for $z_{0.005}$, such that $\mathbb{P}(-z_{0.005} \leq Z \leq z_{0.005}) = 0.99$.
- Since we have $\mathbb{P}(-2.576 < Z < +2.576) = 0.99$, $z_{0.005} = 2.576$

More General Confidence Intervals

A $(1-\alpha)$ level confidence interval for μ is $[\bar{X} - z_{\alpha/2} \cdot \sigma/\sqrt{n},\ \bar{X} + z_{\alpha/2} \cdot \sigma/\sqrt{n}]$.

Examples:

- A more accurate 95% confidence interval for μ is $[\bar{X} - 1.96 \cdot \sigma/\sqrt{n},\ \bar{X} + 1.96 \cdot \sigma/\sqrt{n}]$, since $\mathbb{P}(-1.96 \leq Z \leq 1.96) = 0.95$.
- If we are going to construct a 99% confidence interval for μ, then we need to look for $z_{0.005}$, such that $\mathbb{P}(-z_{0.005} \leq Z \leq z_{0.005}) = 0.99$.
- Since we have $\mathbb{P}(-2.576 < Z < +2.576) = 0.99$, $z_{0.005} = 2.576$
- A 99% level confidence interval for μ is $[\bar{X} - 2.576 \cdot \sigma/\sqrt{n},\ \bar{X} + 2.576 \cdot \sigma/\sqrt{n}]$.

Example

The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. What is a 99% CI for the true mean?

Example

The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. What is a 99% CI for the true mean?

Answer

- $[\bar{X} - 2.576\frac{\sigma}{\sqrt{n}}, \bar{X} + 2.576\frac{\sigma}{\sqrt{n}}]$

Example

The SD of a normal population is 2 and the mean is unknown. A sample of 100 observations is drawn and the sample mean is 12.8. What is a 99% CI for the true mean?

Answer

- $[\bar{X} - 2.576\frac{\sigma}{\sqrt{n}}, \bar{X} + 2.576\frac{\sigma}{\sqrt{n}}]$
- $12.8 \pm 2.576 \cdot 2/\sqrt{100} = [12.3, 13.3]$.

Confidence Intervals with unknown σ

- Because $SD(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ is typically unknown in practice, it is usually estimated by $s_{\bar{X}} = s/\sqrt{n}$, where $s = \sqrt{\frac{(X_1 - \hat{\mu})^2 + \ldots + (X_n - \hat{\mu})^2}{n}}$.

Confidence Intervals with unknown σ

- Because $SD(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ is typically unknown in practice, it is usually estimated by $s_{\bar{X}} = s/\sqrt{n}$, where $s = \sqrt{\frac{(X_1 - \hat{\mu})^2 + \ldots + (X_n - \hat{\mu})^2}{n}}$.
- $s_{\bar{X}}$ is also called the standard error. Note the difference between $\sigma_{\bar{X}}$ and $s_{\bar{X}}$: the former uses population parameters while the latter estimates the $\sigma_{\bar{X}}$ by substituting sample statistics for population parameters.

Confidence Intervals with unknown σ

- Because $SD(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ is typically unknown in practice, it is usually estimated by $s_{\bar{X}} = s/\sqrt{n}$, where $s = \sqrt{\frac{(X_1-\hat{\mu})^2+\ldots+(X_n-\hat{\mu})^2}{n}}$.
- $s_{\bar{X}}$ is also called the standard error. Note the difference between $\sigma_{\bar{X}}$ and $s_{\bar{X}}$: the former uses population parameters while the latter estimates the $\sigma_{\bar{X}}$ by substituting sample statistics for population parameters.
- For large n, you can safely substitute s for σ to obtain a $(1-\alpha)$ confidence interval

$$[\bar{X} - z_{\alpha/2} \cdot s/\sqrt{n}, \bar{X} + z_{\alpha/2} \cdot s/\sqrt{n}]$$

Confidence Intervals with unknown σ

- Because $SD(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ is typically unknown in practice, it is usually estimated by $s_{\bar{X}} = s/\sqrt{n}$, where $s = \sqrt{\frac{(X_1-\hat{\mu})^2+\ldots+(X_n-\hat{\mu})^2}{n}}$.
- $s_{\bar{X}}$ is also called the standard error. Note the difference between $\sigma_{\bar{X}}$ and $s_{\bar{X}}$: the former uses population parameters while the latter estimates the $\sigma_{\bar{X}}$ by substituting sample statistics for population parameters.
- For large n, you can safely substitute s for σ to obtain a $(1-\alpha)$ confidence interval

$$[\bar{X} - z_{\alpha/2} \cdot s/\sqrt{n}, \bar{X} + z_{\alpha/2} \cdot s/\sqrt{n}]$$

- We will discuss the case where the sample size n is small next week.

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:
- We randomly sample an observation from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:
- We randomly sample an observation from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.
- If $Y \sim Exp(\theta)$, then $f_Y(y) = \frac{1}{\theta} e^{-y/\theta}$ for $y \geq 0$.

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:
- We randomly sample an observation from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.
- If $Y \sim Exp(\theta)$, then $f_Y(y) = \frac{1}{\theta} e^{-y/\theta}$ for $y \geq 0$.
- Let $U = \frac{Y}{\theta}$, we have $f_U(u) = f_Y(u\theta) \cdot \frac{dy}{du} = \frac{1}{\theta} e^{-u} \cdot \theta = e^{-u}$ for $u > 0$, that is, $U \sim Exp(1)$

Other confidence intervals

- A useful method for deriving confidence intervals is to use a pivotal quantity
- A pivotal quantity
  - is a function of the sample data, the unknown target parameter, and no other quantities
  - has a distribution that does not depend on the target parameter
- Example:
- We randomly sample an observation from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.
- If $Y \sim Exp(\theta)$, then $f_Y(y) = \frac{1}{\theta} e^{-y/\theta}$ for $y \geq 0$.
- Let $U = \frac{Y}{\theta}$, we have $f_U(u) = f_Y(u\theta) \cdot \frac{dy}{du} = \frac{1}{\theta} e^{-u} \cdot \theta = e^{-u}$ for $u > 0$, that is, $U \sim Exp(1)$
- Thus, we can use $U = \frac{Y}{\theta}$ as a pivotal quantity

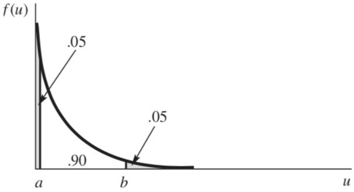
Example

- We need to find two numbers a and b such that

$$\mathbb{P}(a \leq U \leq b) = 90\%$$

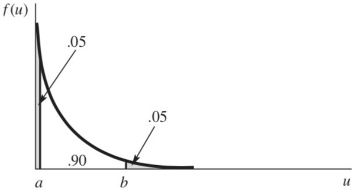
Example

- We need to find two numbers a and b such that
$$\mathbb{P}(a \leq U \leq b) = 90\%$$
- One way to do this is to choose a and b to satisfy $\mathbb{P}(U < a) = \mathbb{P}(U > b) = 5\%$



Example

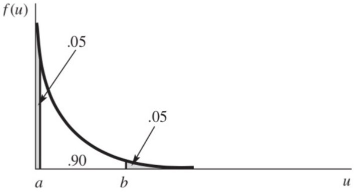
- We need to find two numbers a and b such that
$$\mathbb{P}(a \le U \le b) = 90\%$$
- One way to do this is to choose a and b to satisfy $\mathbb{P}(U < a) = \mathbb{P}(U > b) = 5\%$



- $1 - e^{-a} = .05$ and $e^{-b} = .05$, equivalently, $a = 0.051$, $b = 2.996$

Example

- We need to find two numbers a and b such that
$$\mathbb{P}(a \le U \le b) = 90\%$$
- One way to do this is to choose a and b to satisfy $\mathbb{P}(U < a) = \mathbb{P}(U > b) = 5\%$



- $1 - e^{-a} = .05$ and $e^{-b} = .05$, equivalently, $a = 0.051$, $b = 2.996$
- $0.9 = \mathbb{P}(0.051 \le \frac{Y}{\theta} \le 2.996) = \mathbb{P}(\frac{Y}{2.996} \le \theta \le \frac{Y}{0.051})$

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, ..., X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+...+\frac{1}{\theta_n}})$.

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, ..., X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+...+\frac{1}{\theta_n}})$.
- $\frac{X_1}{\theta}, ... \frac{X_{10}}{\theta} \overset{i.i.d.}{\sim} Exp(1)$

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, ..., X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+...+\frac{1}{\theta_n}})$.
- $\frac{X_1}{\theta}, ... \frac{X_{10}}{\theta} \overset{i.i.d.}{\sim} Exp(1)$
- $U = \min\{\frac{X_1}{\theta}, ... \frac{X_{10}}{\theta}\} \sim Exp(\frac{1}{10})$

Example

Example

We randomly sample 10 observations $X_1, ..., X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, ..., X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+\ldots+\frac{1}{\theta_n}})$.
- $\frac{X_1}{\theta}, \ldots \frac{X_{10}}{\theta} \overset{i.i.d.}{\sim} Exp(1)$
- $U = \min\{\frac{X_1}{\theta}, \ldots \frac{X_{10}}{\theta}\} \sim Exp(\frac{1}{10})$
- Find a, b such that $\mathbb{P}(U < a) = 1 - e^{-10a} = .05$ and $\mathbb{P}(U > b) = e^{-10b} = .05$, equivalently, $a = 0.005$, $b = 0.300$

Example

Example

We randomly sample 10 observations $X_1, \ldots, X_{10}$ from an exponential distribution with unknown mean θ. Find a formula for a 90% CI for θ.

Answer

- Hint: If $X_i \sim Exp(\theta_i)$ then $\min\{X_1, \ldots, X_n\} \sim Exp(\frac{1}{\frac{1}{\theta_1}+\ldots+\frac{1}{\theta_n}})$.
- $\frac{X_1}{\theta}, \ldots \frac{X_{10}}{\theta} \overset{i.i.d.}{\sim} Exp(1)$
- $U = \min\{\frac{X_1}{\theta}, \ldots \frac{X_{10}}{\theta}\} \sim Exp(\frac{1}{10})$
- Find a, b such that $\mathbb{P}(U < a) = 1 - e^{-10a} = .05$ and $\mathbb{P}(U > b) = e^{-10b} = .05$, equivalently, $a = 0.005,\ b = 0.300$
- $0.9 = \mathbb{P}(0.005 < U < 0.300) = \mathbb{P}(0.005 < \frac{\min\{X_1,\ldots,X_{10}\}}{\theta} < 0.300) = \mathbb{P}(\frac{\min\{X_1,\ldots,X_{10}\}}{0.300} < \theta < \frac{\min\{X_1,\ldots,X_{10}\}}{0.005})$

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Answer

- Consider $U = \frac{Y}{\theta}$

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Answer

- Consider $U = \frac{Y}{\theta}$
- $U \sim U(0, 1)$

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Answer

- Consider $U = \frac{Y}{\theta}$
- $U \sim U(0, 1)$
- $\mathbb{P}(U < 0.95) = 0.95$

Example

Example

Suppose that we take a sample Y from a uniform distribution defined on the interval $[0, \theta]$, where θ is unknown. Find a 95% lower confidence bound for θ, that is, find $L(Y)$, such that $\mathbb{P}(\theta \geq L(Y)) = 0.95$.

Answer

- Consider $U = \frac{Y}{\theta}$
- $U \sim U(0, 1)$
- $\mathbb{P}(U < 0.95) = 0.95$
- $0.95 = \mathbb{P}(U < 0.95) = \mathbb{P}(\frac{Y}{\theta} < 0.95) = \mathbb{P}(\frac{Y}{0.95} \leq \theta)$

Manipulating confidence intervals

- The key idea: you are allowed to transform the ends of the confidence interval to obtain a new confidence interval for the transformed parameter (the transformation must be monotone though).

Manipulating confidence intervals

- The key idea: you are allowed to transform the ends of the confidence interval to obtain a new confidence interval for the transformed parameter (the transformation must be monotone though).
- Example: Let X be the temperature at 9am tomorrow in degree Celsius. You have a 95% CI for $\mu = \mathbb{E}(X)$: (12°C, 14°C).

Manipulating confidence intervals

- The key idea: you are allowed to transform the ends of the confidence interval to obtain a new confidence interval for the transformed parameter (the transformation must be monotone though).
- Example: Let X be the temperature at 9am tomorrow in degree Celsius. You have a 95% CI for $\mu = \mathbb{E}(X)$: (12°C, 14°C).
- Recall the formula that converts Celsius to Fahrenheit, $F = 32 + 9C$.

Manipulating confidence intervals

- The key idea: you are allowed to transform the ends of the confidence interval to obtain a new confidence interval for the transformed parameter (the transformation must be monotone though).
- Example: Let X be the temperature at 9am tomorrow in degree Celsius. You have a 95% CI for $\mu = \mathbb{E}(X)$: (12°C, 14°C).
- Recall the formula that converts Celsius to Fahrenheit, $F = 32 + 9C$.
- Find a 95% CI for μ when it is measured in degree Fahrenheit.

Manipulating confidence intervals

- The key idea: you are allowed to transform the ends of the confidence interval to obtain a new confidence interval for the transformed parameter (the transformation must be monotone though).
- Example: Let X be the temperature at 9am tomorrow in degree Celsius. You have a 95% CI for $\mu = \mathbb{E}(X)$: (12°C, 14°C).
- Recall the formula that converts Celsius to Fahrenheit, $F = 32 + 9C$.
- Find a 95% CI for μ when it is measured in degree Fahrenheit.
- The interval on the transformed scale is: $(12 \times 9 + 32, 14 \times 9 + 32)°\text{F} = (140°\text{F}, 158°\text{F})$.

Manipulating confidence intervals

- In many medical settings clinicians like to report results in terms of odds ratios rather than probabilities.
- If the probability is given by p then the odds ratio is $p/(1-p)$. A study has found the 95% confidence interval for p to be (0.35, 0.38).
- Find the 95% CI for the odds ratio.

Manipulating confidence intervals

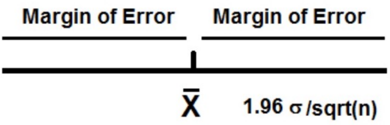
- In many medical settings clinicians like to report results in terms of odds ratios rather than probabilities.
- If the probability is given by p then the odds ratio is $p/(1-p)$. A study has found the 95% confidence interval for p to be (0.35, 0.38).
- Find the 95% CI for the odds ratio.

Answer

$(0.35/0.65,\ 0.38/0.62) = (0.538,\ 0.613)$.

Sample Size Choice

- How large a sample size do I need?
- Think about three things:
  1. What confidence level you want (say 95%).
  2. What Margin of Error (MoE) you want. $(2\sigma/\sqrt{n})$
  3. If σ is unknown, then you need an estimate for σ.
- Recall that the MoE is the distance from the center to the edge of the interval.



The sample size formula for a population mean, .

- As the $MoE = 1.96\sigma/\sqrt{n}$ in the 95% confidence interval for μ, it follows that:

$$n = (\frac{1.96\sigma}{MoE})^2 \approx (\frac{2\sigma}{MoE})^2 \approx (\frac{2s}{MoE})^2.$$

- There are different formulas for different statistical questions (not just the mean), but sample size is always a legitimate question.

Example

- An insurance company is being sued because it has paid bills late and then failed to pay interest on the late payments.
- Lawyers need to estimate the average amount of unpaid interest on the late bills.
- The population of bills is 300,000, far too many to review individually.
- How large a sample size do I need to make a 95% confidence interval for the mean amount of unpaid interest on bills paid late?
- We decide on a MoE of $\pm\$2.00$ and from a previous study we estimate σ by $s = \$25.25$.

Example

- An insurance company is being sued because it has paid bills late and then failed to pay interest on the late payments.
- Lawyers need to estimate the average amount of unpaid interest on the late bills.
- The population of bills is 300,000, far too many to review individually.
- How large a sample size do I need to make a 95% confidence interval for the mean amount of unpaid interest on bills paid late?
- We decide on a MoE of $\pm\$2.00$ and from a previous study we estimate σ by $s = \$25.25$.
- Answer:

$$n = \left(\frac{2 \times 25.25}{2}\right)^2 = 638.$$

Misc.

- Ideally, we would like a narrow 99.9999% confidence interval. Such interval conveys that we have a precise estimate of the unknown population parameter. However, this is not possible, because $z_{\alpha/2}$ is too large in this case. The statisticians often settle for a 95% confidence interval.
- What happens when you increase sample size n? Standard deviation becomes smaller and therefore CI becomes more narrow.

Summary

- Point Estimation
- Confidence intervals.
- Confidence interval for a population mean (μ).

Summary

- Point Estimation
- Confidence intervals.
- Confidence interval for a population mean (μ).
- More confidence intervals based on the pivotal quantity

Homework

- Suppose $Y \sim Bin(n, p)$, calculate $MSE(\frac{Y+2}{n+3})$ and plot the MSE against p
- Prove that S^2 is an unbiased estimator of σ^2
- 7.52, 8.6(a), 8.10(a)
- Sample size calculations and the Margin of Error.