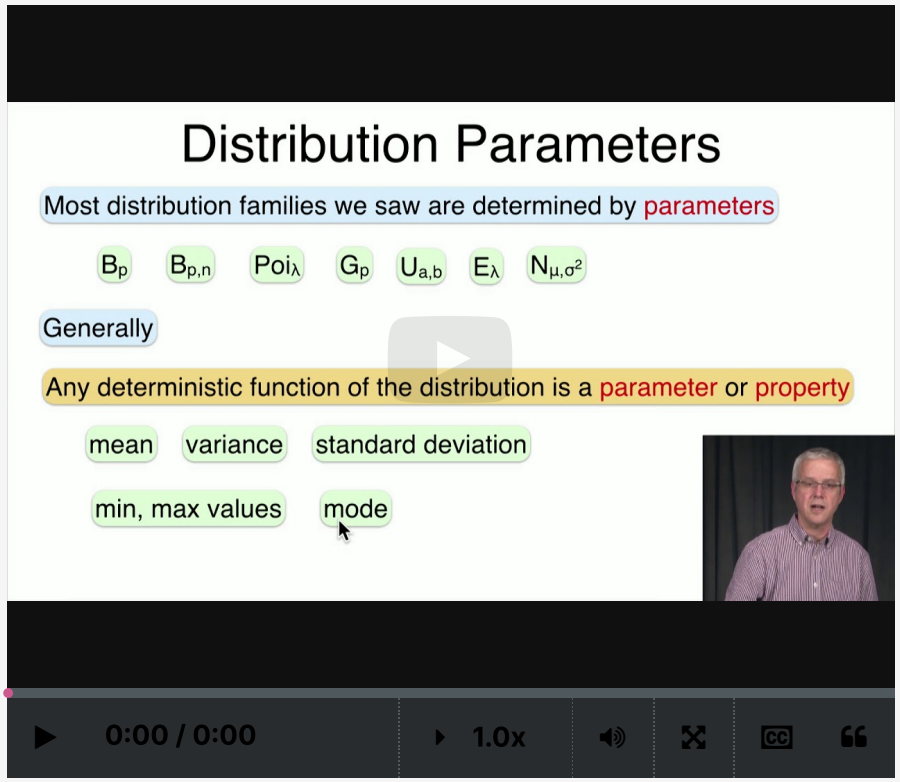
Problem Sets due Jul 26, 2022 04:34 +03

Video



Start of transcript. Skip to the end.

- Hello, and welcome back.

So far, we talked about probability

where everything was designed by us.

Like, we said, okay, this is a uniform distribution,

it behaves exactly in that way.

11.1_Statistics

POLL

Recall a statistic is a single value calculated from the sample. Which of the following is a statistic?

RESULTS

○	**sample max**	**3%**
◉	**sample mean**	**3%**
○	**sample median**	**3%**

○	**all of the above**	**91%**

Submit

Results gathered from 32 respondents.

FEEDBACK

All of them are statistics.

---

1 (Graded)

3/3 points (graded)

225 iPhones go on sale on black friday, and 100 customers are in line to buy them. If the random number of iPhones that each customer wishes to buy is distributed Poisson with mean 2, approximate the probability that all 100 customers get their desired number of iPhones?

0.96

✔ **Answer:** 0.9615

0.96

Explanation

The total iPhone demand may be expressed as a sum $S = X_1 + \ldots + X_{100}$, where each X_i is distributed Poission(2), denoting the number of iPhones demanded by the ith custorer. By the central limit theorem, $S = X_1 + \ldots + X_{100}$ is distributed approximately $\mathcal{N}(200, 200)$. Therefore we may approximate the probability as

$P(S \leq 225) = P\left(\frac{S-200}{\sqrt{200}} \leq \frac{25}{\sqrt{200}}\right) \approx \Phi\left(\frac{25}{\sqrt{200}}\right) = 0.9615.$

? **Hint (1 of 1):** Central Limit Theorom will be helpful. Next Hint

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

2 (Graded)

3/3 points (graded)

The number of years a Bulldog lives is a random variable with mean 9 and standard deviation 3, while for Chihuahuas, the mean is 15 and the standard deviation is 4. Approximate the probability the that in a kennel of 100 Bulldogs and 100 Chihuahuas, the average Chihuahua lives at least 7 years longer than the average Bulldog.

The checker accepts answers with tolerance 0.001

0.022

✔ **Answer:** 0.0228

0.022

Explanation

Let B_i, C_i, $i \in \{1, \ldots, 100\}$ denote the number of years the ith Bulldog, Chihuahua lives respectively, hence $B_i \sim \mathcal{N}(9, 9), C_i \sim \mathcal{N}(15, 16)$. Then, by the central limit theorem, the difference in average lifetime, $D = \sum_{i=1}^{100} \frac{C_i - B_i}{100}$ is distributed $\mathcal{N}(6, 25/100)$.

Therefore

$$P(D \geq 7) = P\left(\frac{D-6}{\sqrt{25/100}} \geq \frac{1}{\sqrt{25/100}}\right) \approx 1 - \Phi\left(\sqrt{100/25}\right) = 1 - \Phi(2) = 0.0228.$$

? **Hint (1 of 1):** Central Limit Theorom will be helpful. Next Hint

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Topic 11 / Statistics

Add a Post

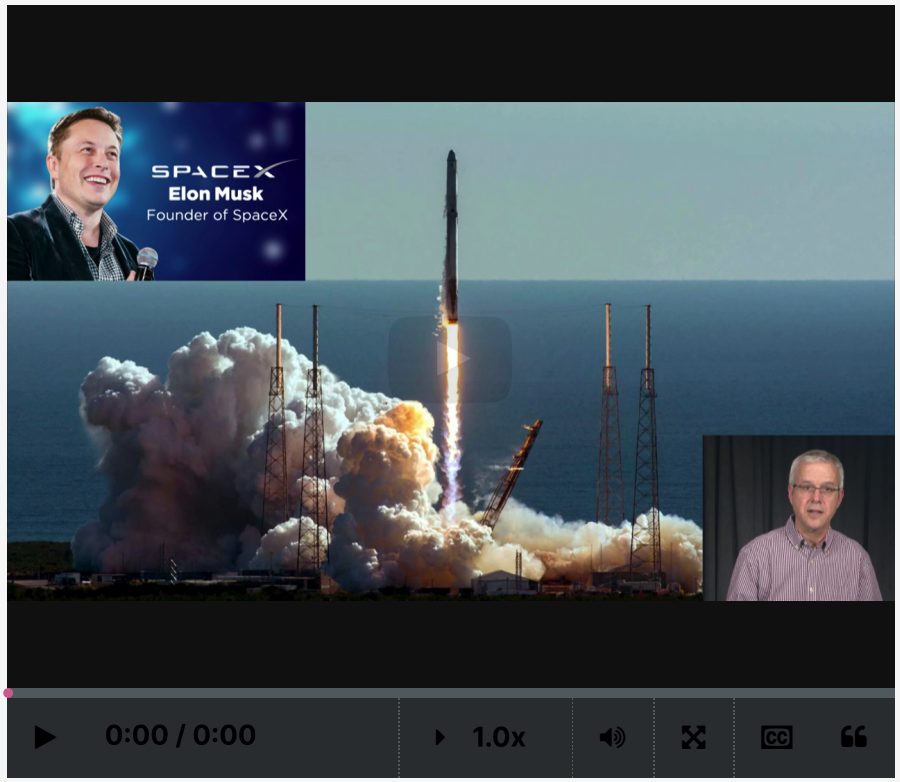
Show all posts by recent activity

There are no posts in this topic yet.

✖

Problem Sets due Jul 26, 2022 04:34 +03

Video



Start of transcript. Skip to the end.

- Hello and welcome back.

Now that we know what we're going to do,

we're going to start by looking

at estimating the first property,

and so then talk about parameter estimation.

We'll also have a mean example

POLL

A distribution has mean 5 and variance 10. If we collect a sample by making 20 independent observations, what is the variance of the sample mean?

RESULTS

○	2	14%
○	1/2	68%
◉	1/4	7%
○	1/40	11%

Submit

Results gathered from 28 respondents.

FEEDBACK

The answer is 10/20 = 1/2.

11.2_Parameter_Estimation

1

0 points possible (ungraded)

If an estimator is unbiased, then

- [ ] its value is always the value of the parameter,
- [x] its expected value is always the value of the parameter, ✔
- [x] it variance is the same as the variance of the parameter.

✖

Explanation

An estimator $\hat{A}$ for a parameter A is unbiased if $E(\hat{A}) = A$.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

2 (Graded)

1/1 point (graded)

If $\{X_1, \ldots, X_n\}$ are the observed values of n sample items, which of the following are unbiased estimators for distribution mean?

- [x] X_1
- [x] $\frac{1}{n}\sum_{i=1}^{n} X_i$
- [ ] $\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}$

✔

Explanation

Denote the distribution mean as μ

- True. $E(X_1) = \mu$.

- True. $E\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n}n\mu = \mu$.

- False. $E(\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}) \neq \mu$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

3

0 points possible (ungraded)

As the sample size n grows, the sample mean estimates the distribution mean better. Because

- [ ] its bias decreases,
- [ ] its variance decreases,
- [ ] none of the above.

Submit You have used 0 of 2 attempts

4 (Graded)

1/1 point (graded)

A sample of size n has sample mean 20.20. After adding a new observed value 21, the sample mean increases to 20.25. What is n?

15

✔ **Answer:** 15

15

Explanation

Let S_n denote the sum of the n samples. Thus $S_n/n = 20.20$ or $S_n = 20.20n$. We're also told $20.20n + 21 = S_{n+1} = 20.25 \cdot (n+1)$. Hence $0.75 = 0.05n$ or $n = 15$.

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

5

0 points possible (ungraded)

To estimate the average alcohol consumption of UCSD students, we take three random samples of 40, 45 and 50 students respectively, and their sample means turn out to be 3.15, 3.20 and 2.76 pints per week respectively. What is the sample mean of the collection of all three samples?

Submit You have used 0 of 4 attempts

6 (Graded)

3/3 points (graded)

Let $X_1, X_2, \ldots, X_n$ be independent samples from a distribution with pdf $f_X(x) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}}$ $(X \geq 0)$. Which of the following is an unbiased estimator for θ?

- ○ $\bar{X}$
- ◉ $\frac{\bar{X}}{2}$
- ○ $\frac{\bar{X}}{3}$
- ○ $\frac{\bar{X}}{6}$

✔

Explanation

By linearity of expectation, $E(\bar{X}) = E(X_1) = \int_0^\infty x \frac{1}{\theta^2} x e^{-\frac{x}{\theta}} dx = 2\theta$. Thus $E(\frac{\bar{X}}{2}) = \theta$ and is therefore an unbiased estimator for θ.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

7

0 points possible (ungraded)

For $i \in \{1, \cdots, n\}$, let $X_i \sim \mathrm{U}(0, W)$ independently of each other, and let $M_n = \max_{i \in \{1, \cdots, n\}} X_i$. For what value of c is $c \cdot M_n$ an unbiased mean estimator?

- ○ $\frac{n+1}{2n}$

- $\frac{n}{2(n-1)}$
- $\frac{2n+1}{4n}$
- $\frac{2n}{4n-1}$

Submit You have used 0 of 2 attempts

8

0 points possible (ungraded)

Let X be distributed $\text{Poisson}(\lambda)$. Which of the following is an unbiased estimator for λ^2.

- X^2
- $X^2 - X$
- $2X^2 - X$
- $3X^2 - 2X$

Submit You have used 0 of 2 attempts

Discussion

Hide Discussion

Topic: Topic 11 / Mean and Variance

Add a Post

Show all posts by recent activity

There are no posts in this topic yet.

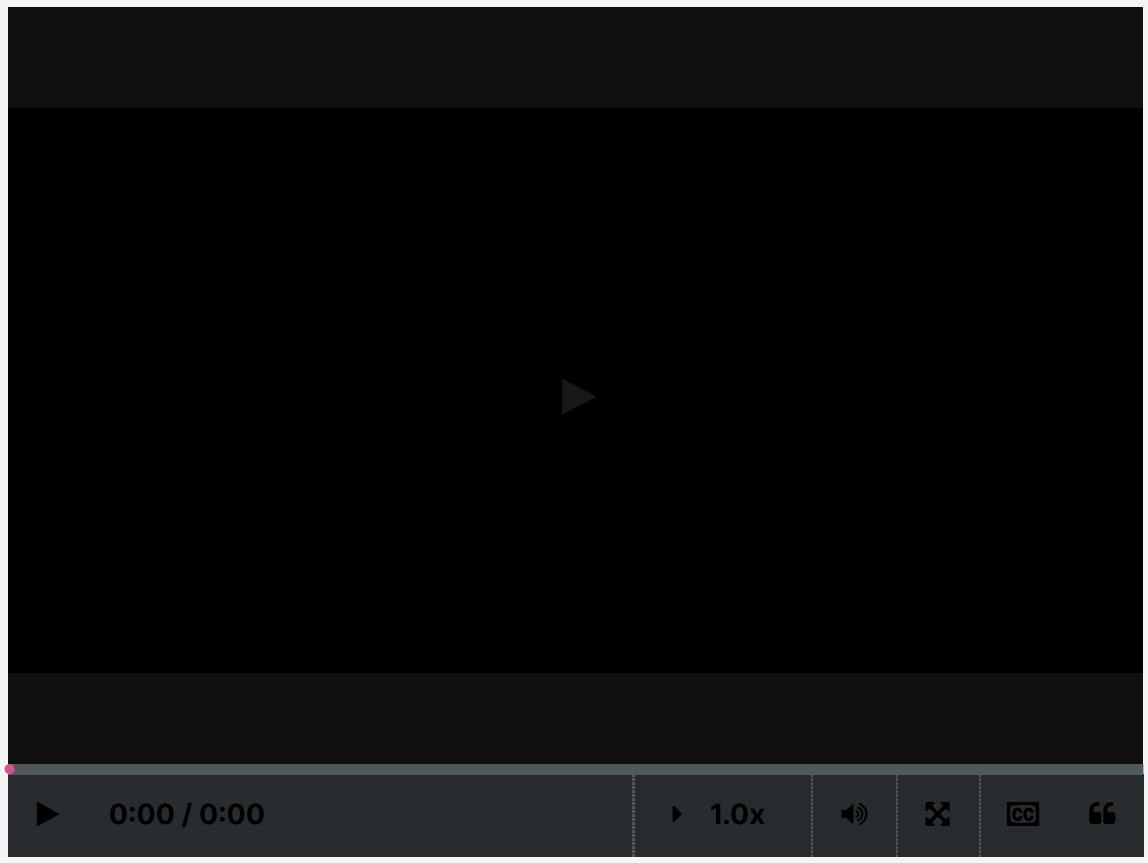
✖

Variance Estimation

Bookmark this page

Problem Sets due Jul 26, 2022 04:34 +03

Video



Start of transcript. Skip to the end.

- Hello and welcome back.

Last time we saw how to estimate the mean of a distribution.

And now we're going to talk about the next

most natural parameter, which is the variance

and see how to estimate that.

And we're going follow one of the early data scientists

POLL

As an estimator for distribution variance, the "raw" sample variance is

RESULTS

biased	**90%**
unbiased	**10%**

Submit

Results gathered from 29 respondents.

FEEDBACK

It's biased.

11.3_Variance_Estimation

1 (Graded)

2/2 points (graded)

Let $\overline{X}_n$ and S_n^2 be the sample mean and the sample variance of $\{X_1, \ldots, X_n\}$. Let $\overline{X}_{n+1}$ and S_{n+1}^2 be the sample mean and the sample variance of $\{X_1, \ldots, X_n, \overline{X}_n\}$. Which of the following hold

- for sample means,

- [ ] $\overline{X}_n > \overline{X}_{n+1}$
- [ ] $\overline{X}_n < \overline{X}_{n+1}$
- [x] $\overline{X}_n = \overline{X}_{n+1}$

✔

Explanation

$\overline{X}_{n+1} = \frac{\sum_{i=1}^{n} X_i + \overline{X}_n}{n+1} = \frac{n \cdot \overline{X}_n + \overline{X}_n}{n+1} = \overline{X}_n.$

- for sample variances?

- [x] $S_n^2 > S_{n+1}^2$
- [ ] $S_n^2 < S_{n+1}^2$
- [ ] $S_n^2 = S_{n+1}^2$

✔

Explanation

$S_{n+1}^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X}_{n+1})^2 + (\overline{X}_n - \overline{X}_{n+1})^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2 + (\overline{X}_n - \overline{X}_n)^2}{n} < \frac{\sum_{i=1}^{n} (X_i - \overline{X}_n)^2}{n-1} = S_n^2.$

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

2 (Graded)

2/2 points (graded)

Consider the following array of $m \times n$ random variables $X_{11}, X_{12}, \cdots, X_{1n}, \cdots, X_{i1}, X_{i2}, \cdots, X_{in}, \cdots, X_{m1}, X_{m2}, \cdots, X_{mn}$. For $i = 1, \cdots, m$, let $\overline{X}_i$ be the sample mean of $\{X_{i1}, X_{i2}, \cdots, X_{in}\}$, and $\overline{S}^2$ be the "raw" sample variance of $\{\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_m\}$. If $\forall i, j, V(X_{ij}) = \sigma^2$, what is $E(\overline{S}^2)$?

- [ ] $\frac{n-1}{n}\sigma^2$
- [ ] $\frac{m-1}{m}\sigma^2$
- [ ] $\frac{1}{n}\sigma^2$
- [ ] $\frac{1}{m}\sigma^2$
- [ ] $\frac{n-1}{mn}\sigma^2$
- [x] $\frac{m-1}{mn}\sigma^2$

✔

Explanation

As $V(X_{ij}) = \sigma^2$, we have $V(\overline{X}_i) = \frac{\sigma^2}{n}$.

$E(\overline{S}^2) = \frac{m-1}{m} V(\overline{X}_i) = \frac{m-1}{mn}\sigma^2.$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

3

0 points possible (ungraded)

If all the observations in a sample increase by 5

- [ ] the sample mean increases by 5,
- [ ] the sample mean stays the same,
- [ ] the sample variance increases by 5,
- [ ] the sample variance stays the same.

Submit You have used 0 of 2 attempts

Discussion

Hide Discussion

Topic: Topic 11 / Variance Estimation

Add a Post

Show all posts by recent activity

There are no posts in this topic yet.

✖

< Previous Next >

© All Rights Reserved

Connect

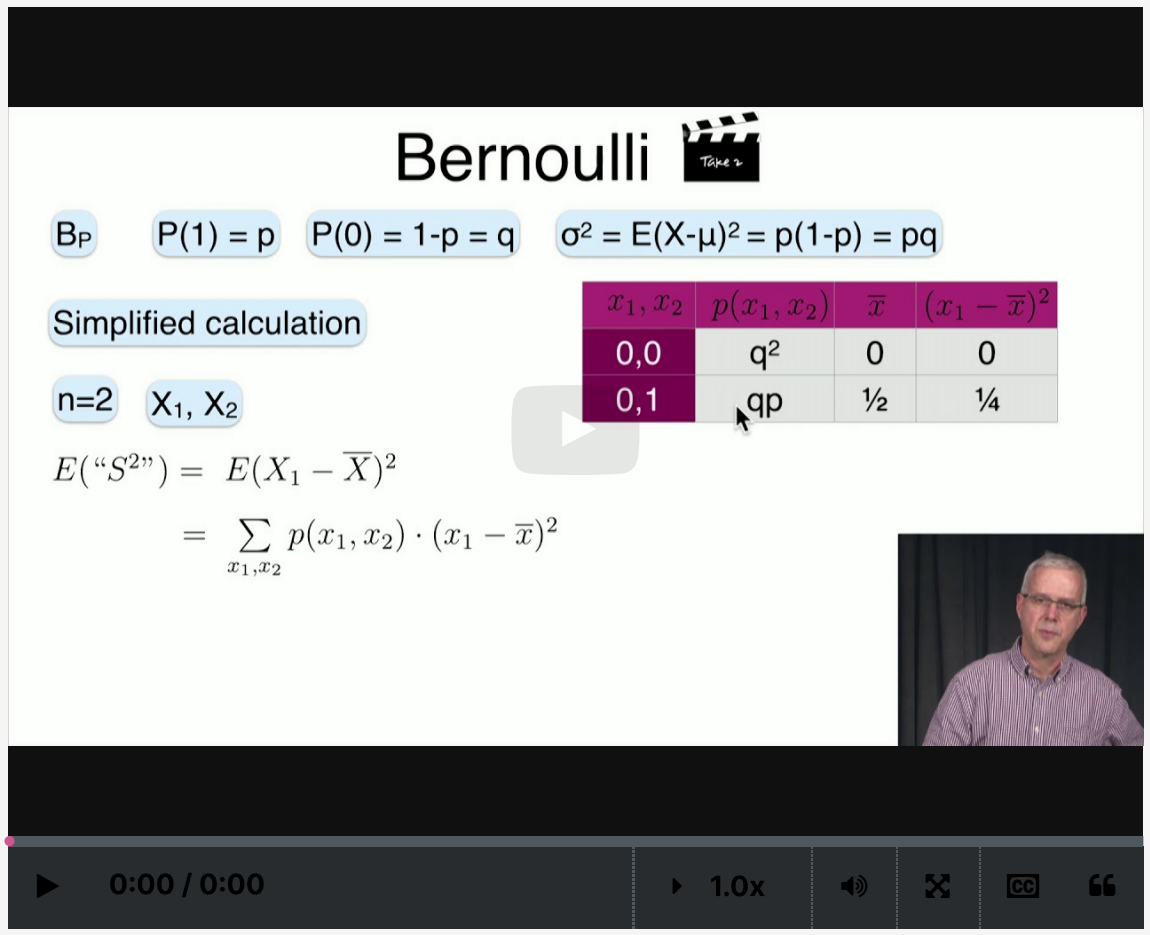
© 2022 edX LLC. All rights reserved.
深圳市恒宇博科技有限公司 粤ICP备17044299号-2

Unbiased Estimation

Bookmark this page

Problem Sets due Jul 26, 2022 04:34 +03

Video



Start of transcript. Skip to the end.

- Hello and welcome back.

Last time we discussed estimating the variance.

Today we're going to talk about unbiased estimation

of the variance.

And we're going to resolve the mystery of the missing man.

So what we're going to do is evaluate the bias

11.4_Unbiased_Variance_Estimation

POLL

Compared to the distribution variance, the expectation of the biased "raw" sample variance is

RESULTS

always larger	4%
always smaller	81%
always equal	4%
could be any of the above	11%

Submit

Results gathered from 27 respondents.

FEEDBACK

It's always smaller.

1 (Graded)

1/1 point (graded)

As the sample size n grows, the effect of the Bessel correction

- becomes larger,
- becomes smaller,
- stays the same.

Explanation
Sample variance $S_n^2 = \frac{n-1}{n}\sigma^2$. When n grows, $\frac{n-1}{n} \to 1$ and the effect becomes smaller.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

2 (Graded)

2/2 points (graded)
According to the U.S. Department of Agriculture, ten to twenty earthworms per cubic foot is a sign of healthy soil. The soil of a garden is checked by digging 8 holes, each of one-cubic-foot, and counting the earthworms, and the following counts are found: 5, 25, 15, 10, 7, 12, 16, 20. Use the unbiased estimators discussed in the video to estimate

- the true mean,

13.75 **Answer:** 13.75

13.75

Explanation
If X_i, $1 \leq i \leq 8$ denotes the samples, the unbiased estimator for the true mean is given by $\overline{X} = \sum_{i=1}^{8} X_i/n = 110/8 = 13.75.$

- the true variance.

44.5 **Answer:** 44.5

44.5

Explanation
The unbiased estimator for the true variance is given by $S = \sum_{i=1}^{8} \left(X_i - \overline{X}\right)^2 / (n-1) = 311.5/7 = 44.5.$

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Topic 11 / Unbiased Estimation

Add a Post

Show all posts by recent activity

There are no posts in this topic yet.

✖

< Previous Next >

© All Rights Reserved

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

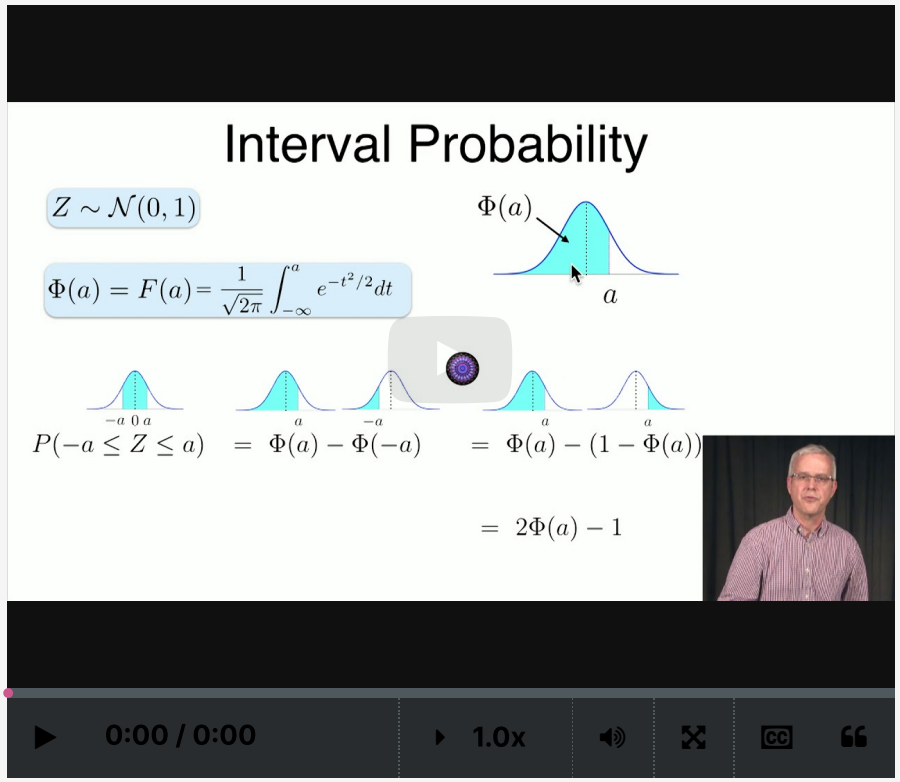
Blog
Contact Us
Help Center
Media Kit

GET IT ON Google play
Download on the App Store

© 2022 edX LLC. All rights reserved.
深圳市恒宇博科技有限公司 粤ICP备17044299号-2

Problem Sets due Jul 26, 2022 04:34 +03

Part 1



Start of transcript. Skip to the end.

- Hello and welcome back.

In the previous lectures,

we talked about parameter estimation,

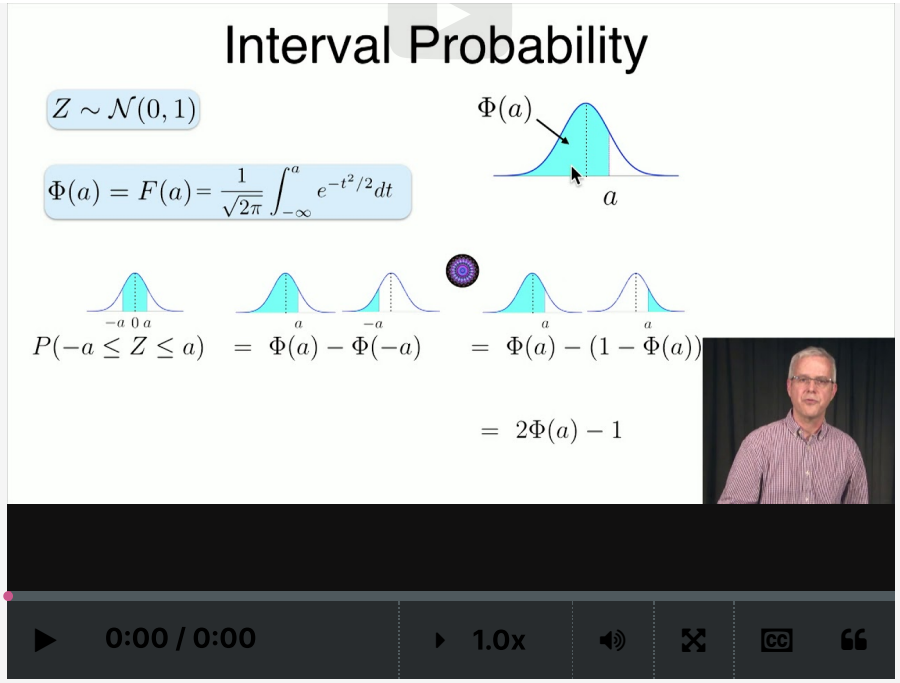
and we tried to do our best to find the value of parameter.

Correction: When p = 98%, Zp = 2.326, not 2.056.

Part 2

Start of transcript. Skip to the end.

All right, so first,



let's see what is the distribution of the sample mean,

because we're trying to estimate the mean.

Let's look and we're trying

11.6_Confidence_Intervals

POLL

The margin of error of confidence interval with 100% confidence level will be

RESULTS

○	**Zero**	**54%**
◉	**One standard deviation**	**4%**
○	**Infinity**	**43%**
○	**None of the above**	**0%**

Submit

Results gathered from 28 respondents.

FEEDBACK

Infinity

1 (Graded)

1/1 point (graded)

Which of the following will **increase** the length of the confidence interval?

- [x] Increase confidence level
- [ ] Decrease confidence level
- [ ] Increase sample size
- [x] Decrease sample size

✔

Explanation

- As $P\left(-a \leq Z \leq a\right) = \textbf{confidence level}$, when confidence level increase, the the length of the interval $\left(-a, a\right)$ will increase.
- When sample size decreases, the variance increases, and probability under the original interval decreases. Since the confidence level is not changed, the length of the confidence interval will increase.

Submit

You have used 1 of 2 attempts

Answers are displayed within the problem

2 (Graded)

2/2 points (graded)

A psychologist estimates the standard deviation of a driver's reaction time to be 0.05 seconds. How large a sample of measurements must be taken to derive a confidence interval for the mean with margin of error at most 0.01 second, and confidence level 95%?

97

✔ **Answer:** 97

97

Explanation

For the margin of error of 0.01, to be within the 95% confidence interval, we require $1.96 * (0.05/\sqrt{n}) \leq 0.01$ or $n \geq 1.96^2 * 25 = 96.04$. A choice of $n = 97$ thus gives the answer.

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

3

0 points possible (ungraded)

A sample of size $n = 25$ with the population standard deviation $\sigma = 3$, compute the margin of error of a 90% confidence interval for the mean μ.

Submit You have used 0 of 4 attempts

Discussion

Hide Discussion

Topic: Topic 11 / Confidence Interval

Add a Post

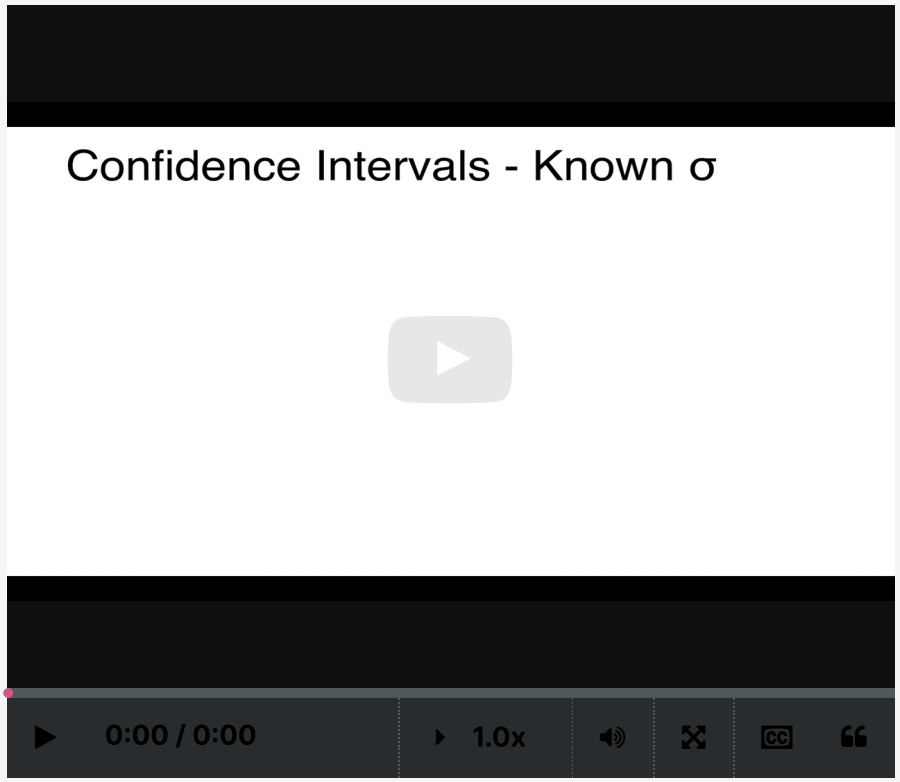
Show all posts by recent activity

There are no posts in this topic yet.

✖

Problem Sets due Jul 26, 2022 04:34 +03

Part 1



Start of transcript. Skip to the end.

- [Instructor] Hello, and welcome back.

In the last video, we talked about confidence intervals,

but we assumed that sigma was known, the standard deviation.

That, of course, is almost never the case,

so, in this lecture, we'll see how we can come up

Part 2: Example

Start of transcript. Skip to the end.

- [Narrator] So, let's say that you wanted

to find the average length

of an elephant's trunk,

in this case mature African elephant trunks.

So you want to do that, then what you would do,

you would go out to the field and take some measurements

POLL

A confidence interval of mean has confidence level 95%. It means

RESULTS

○	**The confidence interval includes distribution mean with probability 95%**	**54%**
◉	**95% of the observations in the sample fall into this interval**	**21%**
○	**If we take a new sample point, it falls into this interval 95% of the time**	**18%**
○	**Non of the above**	**7%**

Submit

Results gathered from 28 respondents.

FEEDBACK

The confidence interval includes distribution mean with probability 95%.

11.7_Confidence_Intervals_-_unknown_sigma

1 (Graded)

1/1 point (graded)

Student's t-distribution can be used to form confidence intervals only when the samples are normal distributed.

- ○ True
- ◉ False

✔

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

2 (Graded)

2/2 points (graded)

To find the average SAT verbal score in a class, six students are sampled and their scores are 560, 610, 500, 470, 660, and 640. Assuming that students' SAT verbal scores follow normal distribution, what is the upper limit for the confidence interval of the distribution mean with confidence level 90%?

636.2

✔ **Answer:** 636.5

636.2

Explanation

Let $X_i,\ 1 \le i \le 6$ denote the samples. Here the sample mean $\overline{X} = \sum_{i=1}^{6} X_i/6 = 3440/6 = 573.333$. The sample variance $S^2 = \sum_{i=1}^{6} \left(X_i - \overline{X}\right)^2 / (6-1) = 29533.33333/5 = 5906.667$.

Since the data is assumed to be generated from a normal distribution, the 90% confidence interval for the mean μ is given by $[\overline{X} - \frac{t_{0.9,5}S}{\sqrt{6}}, \overline{X} + \frac{t_{0.9,5}S}{\sqrt{6}}]$. Using the fact that $t_{0.9,5} = 2.015$, $\overline{X} + \frac{t_{0.9,5}S}{\sqrt{6}} = 573.333 + 2.015\frac{\sqrt{5906.667}}{\sqrt{6}} = 636.542$ (A tolerance of

± 0.5 about 636.5 is an acceptable answer).

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

3

0 points possible (ungraded)

What is the critical t for a 92% confidence interval with a sample size 10?

What is the critical t for a 92% confidence interval with a sample size 1000?

What is the critical z for a 92% confidence interval?

Submit You have used 0 of 4 attempts

Discussion

Hide Discussion

Topic: Topic 11 / Sigma Unknown

Add a Post

Show all posts by recent activity

There are no posts in this topic yet.

✖