Sample Size (7.4), Confidence Intervals for Variance and Standard Deviation

Stat 400 - April 1, 2021

Today's topics

Review:

- Chi Squared Distribution, t distribution, CI for mean

New:

- Confidence Interval for variance (or sd)
- Required Sample Size

Chi-squared distribution (review) $\boxed{\chi^2}$

$Z^2 \sim \chi^2_1$

The Chi-Squared distribution is a special case of the Gamma distribution where $\theta = 2$.

Also, sum of squares of normal distributions

https://homepage.divms.uiowa.edu/~mbognar/applets/chisq.html notes

Let $X \sim \text{Gamma}(\alpha = 5, \theta = 4)$

$$\int_{16}^{\infty} f(x)\,dx$$

$\frac{1}{2}X \sim \text{Gamma}(\alpha = 5, \theta = 2)$

$\frac{1}{2}X \sim \chi^2_{10}$

$$P[X > 16] = P\left[\tfrac{1}{2}X > 8\right]$$

$$P[\chi^2_{10} > 8] = ?$$

$\underline{.6288}$

notes

SE

t distribution: $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

$\sqrt{s^2}$ n-1

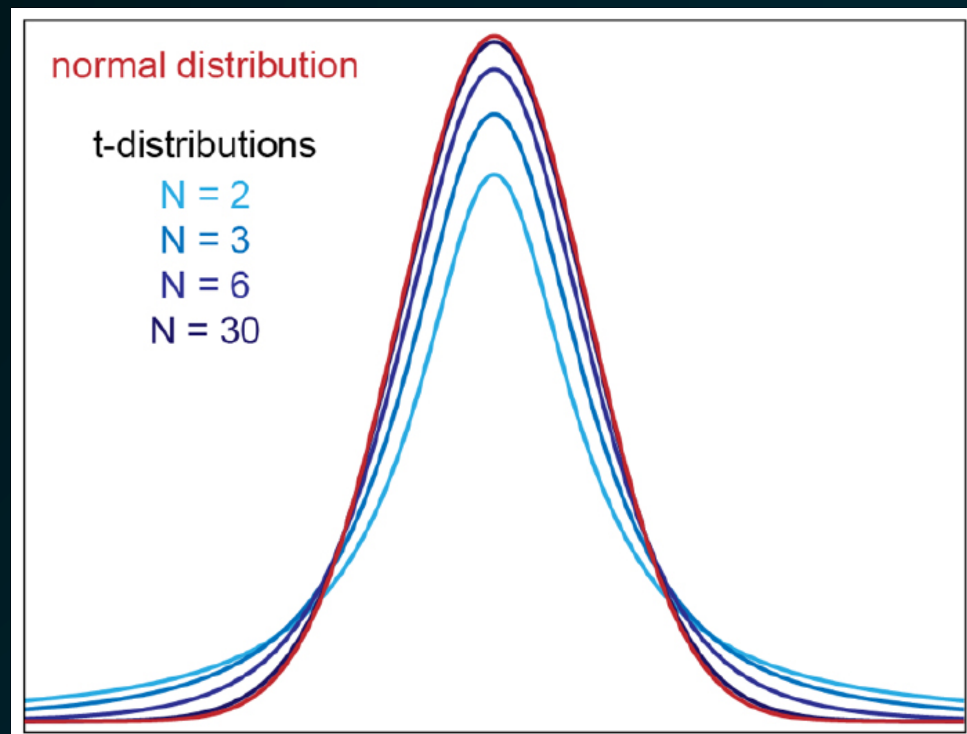
s

If σ is unknown:

Use s instead of σ

$$T = \boxed{\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}}$$

if I "standardize" the sample mean using 's' instead of 'σ'.



t distribution

Theorem 5.5-3 **(Student's *t* distribution)** Let

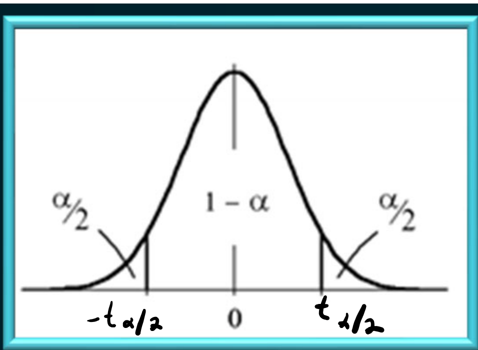
$$T = \frac{Z}{\sqrt{U/r}},$$

where Z is a random variable that is $N(0,1)$, U is a random variable that is $\chi^2(r)$, and Z and U are independent. Then T has a t distribution with pdf

$$f(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\,\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}, \qquad -\infty < t < \infty.$$

If interested, please refer to textbook for proof. (You are not expected to know how to do it).

Review: CI for mean



Given the following sample: {16, 12, 18, 13, 21, 15, 8, 17} $n=8$

Construct a 92% confidence interval for the true mean.

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

$\bar{x} = 15$, $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{112}{7} = 16$, s=4

$\alpha = 0.08, \alpha/2 = 0.04$

df = n -1 = 7, $t_{7,\,0.04} = 2.046$

x	$x-\bar{x}$	$(x-\bar{x})^2$
16	1	1
12	−3	9
18	3	9
13	−2	4
21	6	36
15	0	0
8	−7	49
17	2	4
	0	112

CI: $\mathbf{15 \pm 2.046 \cdot \frac{4}{\sqrt{8}}}$ = (12.107 , 17.893)

notes

Estimate ± margin of error

General Form of Confidence Interval

0.05 1.645

90% CI

$\alpha = 0.1$

$\alpha/2 = 0.05$

Estimate ± (Critical Value * SE of estimate)

e.g. if σ is known: z

$$\bar{x} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

CI for mean μ

if σ is unknown:

t_{n-1}

$$\bar{x} \pm t_{n-1,\alpha/2} * \frac{s}{\sqrt{n}}$$

7.4

Required Sample Size

margin of error

$\bar{X} \pm \varepsilon$

- $\bar{x} \pm z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$

- $\varepsilon = z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$

m.o.e.

$$\boxed{n = \left[\frac{z_{\alpha/2} * \sigma}{\varepsilon}\right]^2}$$

$\sigma^2 = 6.25 \rightarrow \underbrace{\sigma = 2.5}$ $\varepsilon = 0.5$

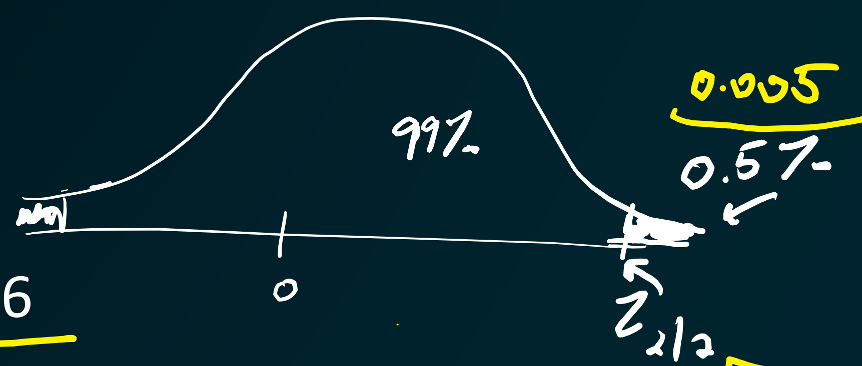
Required Sample Size

How many test runs of an automobile are required for determining its average miles-per-gallon rating on the highway to within 0.5 miles per gallon with 99% confidence? Suppose that the variance of the population in mpg² is 6.25.

- $$\boxed{n = \left[\frac{z_{\alpha/2} * \sigma}{\varepsilon}\right]^2}$$

$\alpha/2 = 0.005$

- $\alpha = 0.01$ $z_{\alpha/2} = 2.576$



$$n = \left[\frac{2.576 \cdot 2.5}{0.5}\right]^2 = 165.89$$

$\boxed{n = 166}$

Round up!

last example, found $s^2 = 16$ ← point estimate

Confidence Interval for σ^2 ← σ

Now we want to make a confidence interval for σ^2 based on s^2.

Given the following sample: {16, 12, 18, 13, 21, 15, 8, 17}

- The distribution of s^2 is not Normal.

- It can be shown that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$ ←

will not need to prove this

χ

Proof of $\frac{(n-1)\boxed{S^2}}{\sigma^2} \sim \chi^2_{n-1}$

$S^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$

Assume

1. $\bar{X}$ (the sample mean) and S^2 are independent.
2. If $Z \sim N(0,1)$ then $Z^2 \sim \chi^2(1)$.
3. If $X_i \sim \chi^2(1)$ and the X_i are independent then $\sum_{i=1}^n X_i \sim \chi^2(n)$.
4. A $\chi^2(n)$ random variable has the moment generating function $(1-2t)^{-n/2}$.

χ^2_1

$E[e^{tX}]$

With some algebra, you can show, by adding $-\bar{X} + \bar{X}$ inside the parentheses and grouping appropriately, that $\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$. Then, dividing through by σ^2 yields

$$\chi^2_n \quad \underbrace{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2}_{U} = \underbrace{\boxed{\sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2}}_{V.} + \underbrace{\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2}_{W}. \quad \chi^2_1$$

Denote these expressions by U, V, and W, respectively, so that the formula reads $U = V + W$. By facts (2) and (3) above, $U \sim \chi^2(n)$ and $W \sim \chi^2(1)$. Also, $V = \frac{(n-1)S^2}{\sigma^2}$.

Since $\bar{X}$ and S^2 are independent, so are V and W. Thus $M_U(t) = M_V(t)M_W(t)$, where $M_X(t)$ denotes the moment generating function of the random variable X. By fact (4) above, this says that

$$\frac{1}{(1-2t)^{n/2}} = M_V(t)\frac{1}{(1-2t)^{1/2}}.$$

$M_V(t) = \frac{1}{(1-2t)^{(n-1)/2}}$, so $V \sim \chi^2_{(n-1)}$

$\boxed{S^2} = \frac{1}{\cancel{n-1}}\sum_{i=1}^n (X_i - \bar{X})^2$

notes

$\frac{1}{3} < \frac{1}{2} < \frac{1}{1}$

$3 > 2 > 1$

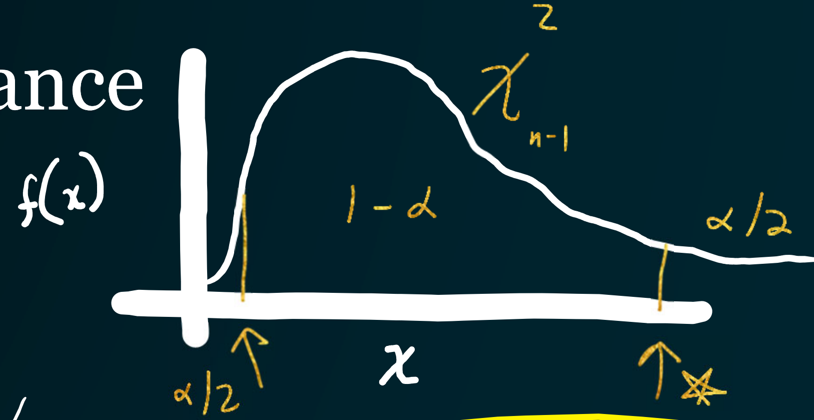
$qchisq(1-\alpha/2, df)$

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ e.g. $\chi^2_{(10)}$

Derivation of CI for Variance

$1 < 2 < 3$

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}}$$



$$P\left[\chi^2_{n-1,\,1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1,\,\alpha/2}\right] = 1-\alpha$$

on the next slide

$$P\left[\frac{\chi^2_{n-1,\,1-\alpha/2}}{(n-1)S^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{n-1,\,\alpha/2}}{(n-1)S^2}\right] = 1-\alpha \Rightarrow P\left[\frac{(n-1)S^2}{\chi^2_{n-1,\,\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1,\,1-\alpha/2}}\right] = 1-\alpha$$

Confidence Interval for σ^2

Confidence Interval for σ^2 : $\left(\dfrac{(n-1)\cdot s^2}{\chi^2_{\alpha/2}}, \dfrac{(n-1)\cdot s^2}{\chi^2_{1-\alpha/2}} \right)$

CI for Variance

Confidence Interval for σ^2 : $\left(\dfrac{(n-1)\cdot s^2}{\underbrace{\chi^2_{\alpha/2}}_{\text{big}}}, \dfrac{(n-1)\cdot s^2}{\underbrace{\chi^2_{1-\alpha/2}}_{\text{small}}}\right)$

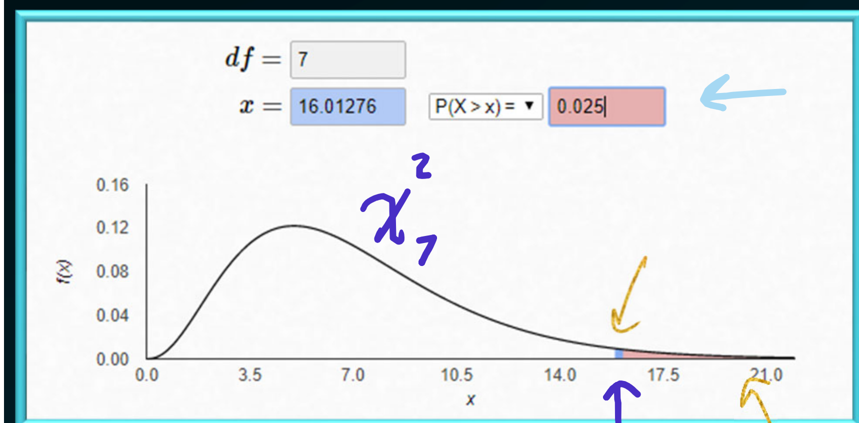
Given the following sample: {16, 12, 18, 13, 21, 15, 8, 17} ← $n=8$

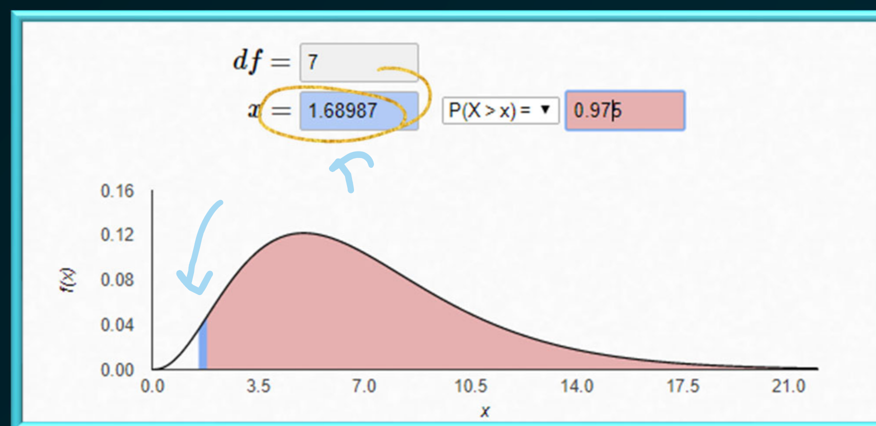
Construct a 95% CI for the true variance.

$s^2=16$, $\alpha = 0.05$, $\alpha/2 = 0.025$, $1-\alpha/2 = 0.975$, df = 7

$\chi^2_{(7,\ 0.025)} = 16.013$ $\chi^2_{(7,\ 0.975)} = 1.690$

$$\left(\frac{(8-1)16}{16.013}, \frac{(8-1)16}{1.68987}\right) = \boxed{(6.996, 66.272)}$$





Frequentist

$\rightarrow (6.9, 66)$

σ^2

Confidence Interval

- For a 100(1-alpha)% CI, if we continue to construct an infinite number of intervals, we expect that 100(1-alpha)% of these intervals will contain the true parameter.

Say 100 CI

Expect 95 to contain true σ^2

Bayesian

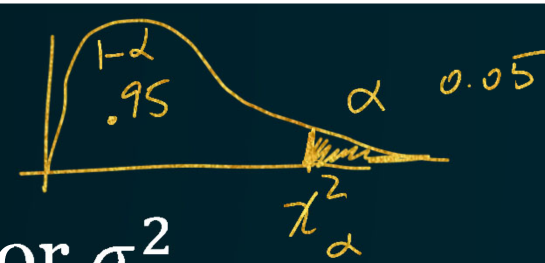
20%

$(0, \infty)$

$(6.9, 66)$

LB UB

1-sided Confidence Intervals for σ^2

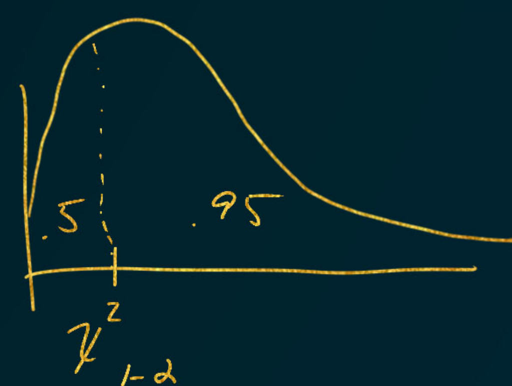


95% confidence lower bound:

$$\boxed{\left(\frac{(n-1)s^2}{\chi^2_{(df,\alpha)}}, \infty\right)}$$

95% confidence upper bound:

$$\boxed{\left(\cancel{-\infty}\ 0, \frac{(n-1)s^2}{\chi^2_{(df,1-\alpha)}}\right)}$$



Blank page

Given the following sample: {16, 12, 18, 13, 21, 15, 8, 17}
Construct a 95% CI for the true variance.

$s^2 = 16$

$n-1 = 7$

$\alpha = 0.05$

95% confidence lower bound:

$$\left(\frac{(n-1)s^2}{\chi^2_{(df,\alpha)}}, \infty\right)$$

$\chi^2_{7,0.05}$

$$\left(\frac{7 \cdot 16}{14.07}, \infty\right)$$

$$\boxed{(7.96, \infty)}$$

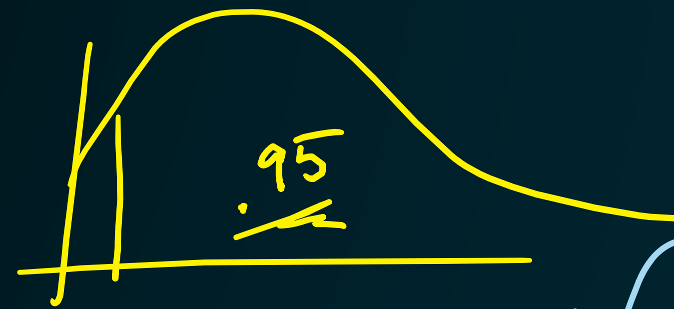
notes

Given the following sample: {16, 12, 18, 13, 21, 15, 8, 17}
Construct a 95% CI for the true variance.

95% confidence upper bound:

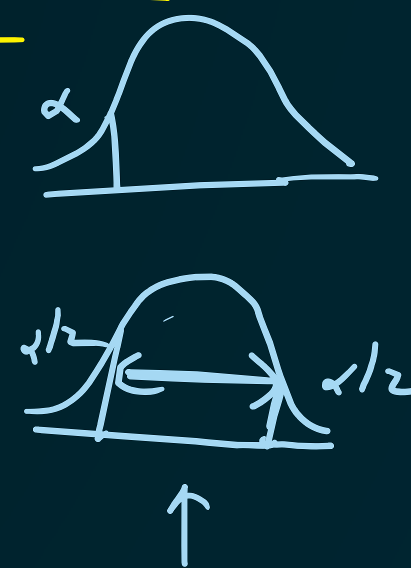
$$\left(-\infty, \frac{(n-1)s^2}{\chi^2_{(df,1-\alpha)}}\right)$$

"0"



$\chi^2_{7, 0.95}$

2.167



$$\left(0, \frac{7 \cdot 16}{2.167}\right)$$

$$= \boxed{(0, 51.67)}$$

^ The last 10 seconds got cut out of the video, but I just typed those numbers in my calculator