Chapter 7. Statistical Intervals Based on a Single Sample

Math 3670 Summer 2024

Georgia Institute of Technology

Section 1.
Basic Properties of Confidence Intervals

Point estimate

μ (mean), $\hat{\mu} = \overline{X}$

$[a, b]$ contains μ with high probability.

Q: How can we find a, b? Interval Estimation.

Confidence Interval for Mean

In this section, we assume that $\begin{cases} \mu \leftarrow \overline{X} \\ \sigma \end{cases}$

1. The population distribution is normal.
2. The value of the population standard deviation σ is known.

Our goal is to find an interval $[a, b]$ such that $\mathbb{P}(\mu \in [a, b]) = 0.95$.

Let $z_{\frac{\alpha}{2}} > 0$ be such that $\mathbb{P}(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$. (Here, Z is standard normal.)

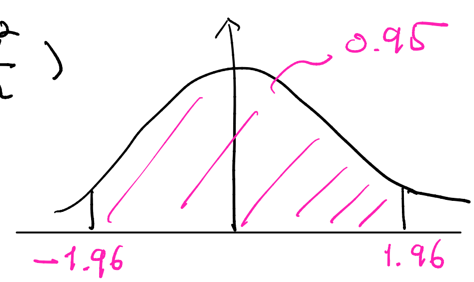
$$\mathbb{P}(a \leq \mu \leq b) = 0.95$$

in terms of $\overline{X}$

Recall $X_1, X_2, \cdots, X_n \sim \text{i.i.d. } N(\mu, \sigma^2)$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$



$$\mathbb{P}\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \in [-1.96, 1.96]\right) = 0.95$$

$$\mathbb{P}\left(-1.96 \leq \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \leq 1.96\right)$$

$$P\left(-1.96\cdot\frac{\sigma}{\sqrt{n}} \le \bar{X}-\mu \le 1.96\cdot\frac{\sigma}{\sqrt{n}}\right)$$

$$\mu \le \bar{X} + 1.96\frac{\sigma}{\sqrt{n}} \qquad \bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \le \mu$$

$$P\Big(\underbrace{\bar{X}-1.96\frac{\sigma}{\sqrt{n}}}_{=a} \le \mu \le \underbrace{\bar{X}+1.96\frac{\sigma}{\sqrt{n}}}_{=b}\Big) = 0.95$$

This means that

$$\mu \in \left[\bar{X}-1.96\frac{\sigma}{\sqrt{n}},\ \bar{X}+1.96\frac{\sigma}{\sqrt{n}}\right]$$

with prob. 0.95.

95% Confidence interval for μ.

under assumptions that

① $X_i \sim N(\mu, \sigma^2)$

② σ is known.

Confidence Interval for Mean

Since $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ is standard normal, we have

$$\mathbb{P}(-1.96 \leq \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96) = 0.95.$$

Solving for μ, we get

$$\mathbb{P}\left(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95.$$

Confidence Interval for Mean

Definition

The interval $\left[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$ is called the 95% confidence interval for unknown μ.

Confidence Interval for Mean

Example

Industrial engineers who specialize in ergonomics are concerned with designing workspace and worker-operated devices so as to achieve high productivity and comfort.

A sample of $n = 31$ trained typists was selected, and the preferred keyboard height was determined for each typist.

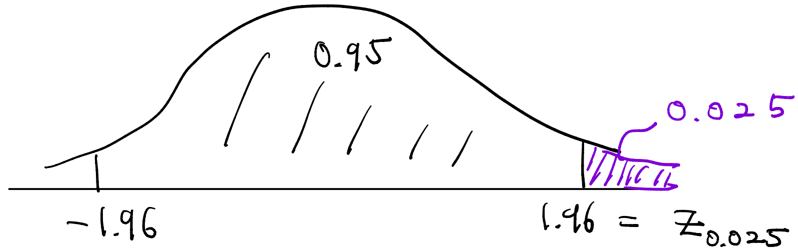
The resulting sample average preferred height was $\bar{x} = 80.0$ cm.

Assuming that the preferred height is normally distributed with $\sigma = 2.0$ cm, find a Confidence interval for μ.

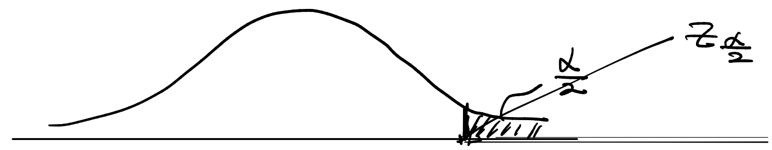
$$\left[\ \bar{X} - 1.96\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\ \right]$$

$$= \left[\ 80 - 1.96 \cdot \frac{2}{\sqrt{31}}\ ,\ 80 + 1.96\frac{2}{\sqrt{31}}\ \right]$$

$$= [\ 79.296\ ,\ 80.704\ ]$$



$$\mathbb{P}\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \in [\overset{-z_{\frac{\alpha}{2}}}{\cancel{-1.96}}, \overset{z_{\frac{\alpha}{2}}}{\cancel{1.96}}]\right) = \overset{1-\alpha}{\cancel{0.95}}$$



Confidence Interval for Mean

From

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\alpha/2}\right) = 1-\alpha,$$

we have

Definition

The interval $\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$ is called the $100(1-\alpha)\%$ **confidence interval** for unknown μ.

Ex

$\alpha = 0.05$ $\quad z_{\frac{\alpha}{2}} = 1.96$ $\quad$ 95% CI

$\alpha = 0.1$ $\quad z_{\frac{\alpha}{2}} = 1.645$ $\quad$ 90% CI

$\alpha = 0.01$ $\quad z_{\frac{\alpha}{2}} = 2.576$ $\quad$ 99% CI

CI width $= 2 \cdot z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} =: l$

$\alpha \uparrow$ $\quad \left\{ \begin{array}{l} \text{Confidence level} \downarrow \\ z_{\frac{\alpha}{2}} \downarrow \\ l \downarrow \end{array} \right.$

$n \uparrow$ $\quad l \downarrow$

Confidence Interval for Mean

Example

The production process for engine control housing units of a particular type has recently been modified.

Prior to this modification, historical data had suggested that the distribution of hole diameters for bushings on the housings was normal with a standard deviation of .100 mm.

It is believed that the modification has not affected the shape of the distribution or the standard deviation, but that the value of the mean diameter may have changed.

A sample of 40 housing units is selected and hole diameter is determined for each one, resulting in a sample mean diameter of 5.426 mm.

Calculate a confidence interval for true average hole diameter using a confidence level of 90%.

$\alpha = 0.1 \qquad z_{\frac{\alpha}{2}} = 1.645$

$$\left[\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

$$= \left[5.426 - 1.645 \cdot \frac{0.1}{\sqrt{40}}, \quad 5.426 + 1.645 \cdot \frac{0.1}{\sqrt{40}} \right]$$

$$= [\, 5.400 \,, \quad 5.452 \,]$$

Width of CI = $2 \cdot Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq 10$

$\Longrightarrow$ Solve for n $\quad n \geq \left(\frac{2 Z_{\alpha/2} \cdot \sigma}{10}\right)^2$

Confidence Interval for Mean

Example

Extensive monitoring of a computer time-sharing system has suggested that response time to a particular editing command is normally distributed with standard deviation 25 millisec.

A new operating system has been installed, and we wish to estimate the true average response time μ for the new environment.

Assuming that response times are still normally distributed with $\sigma = 25$, what sample size is necessary to ensure that the resulting 95% CI has a width of (at most) 10?

$\alpha = 0.05$

$Z_{\frac{\alpha}{2}} = 1.96$

$$\left[\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

$$\text{width} = 2 \cdot 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq 10 \qquad \sigma = 25$$

$$\frac{50 \cdot 1.96}{\sqrt{n}} \leq 10$$

$$\sqrt{n} \geq 5 \cdot (1.96)$$

$$n \geq 96.04$$

$$\underline{\underline{n \geq 97}}$$

Exercise

(7.1-4) A CI is desired for the true average stray-load loss μ (watts) for a certain type of induction motor when the line current is held at 10 amps for a speed of 1500 rpm. Assume that strayload loss is normally distributed with $\sigma = 3.0$.

1. Compute a 95% CI for μ when $n = 25$ and $\bar{x} = 58.3$. ← 1.96
2. Compute a 95% CI for μ when $n = 100$ and $\bar{x} = 58.3$.
3. Compute a 99% CI for μ when $n = 100$ and $\bar{x} = 58.3$.
4. How large must n be if the width of the 99% interval for μ is to be 1.0?

1. $\left[58.3 - 1.96 \cdot \frac{3}{\sqrt{25}},\ 58.3 + 1.96 \frac{3}{\sqrt{25}}\right]$

$= [57.124,\ 59.476]$ width 2.352 ↓ $\frac{1}{2}$

2. $[57.712,\ 58.888]$ 1.176

3. $[57.527,\ 59.073]$ 1.546

4. $\text{width} = 2 \cdot z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 2 \cdot 2.576 \cdot \frac{3}{\sqrt{n}} = 1$

$n \geq 238.89$

$\underline{n = 239}$ //

Final Exam

1. Date, Time: Monday, July 29 11:20 AM - 2:10 PM
2. Location: J. Erskine Lo 183
3. Coverage: Cumulative
4. No book, no notes are allowed. Only calculators are okay.
5. 3-4 problems From Midterm 1 and 2 (similar to exam problems with some modification)
6. 3-4 problems From new topics (From 5.4 to 7.2) - Mostly from HW and lecture slides with some modification

PDF = (⋮)

Find marginal
condition
Exp. . . .

Confidence Interval

Assuming ① $X_1, \cdots, X_n \sim$ i.i.d. Normal

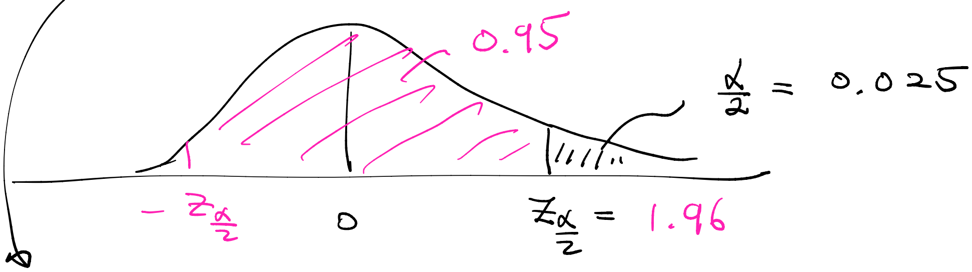
② σ is known

Goal : $\mathbb{P}(a \le \mu \le b) = 0.95$

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$P\left(-z_{\frac{\alpha}{2}} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

Solve for μ



$$P\left(\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

if $\alpha = 0.05$, $= 0.95$

$\left[\overline{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$: Confidence Interval with CL $(100)\cdot(1-\alpha)\%$

i.i.d. = independent and identically distributed

= "random sample"

Section 2.
Large-Sample Confidence Intervals for a Population Mean and ~~Proportion~~

Final.

- $X_1, \cdots, X_n \sim$ i.i.d.
  possibly Non-Normal

"n is large"

- σ is unknown

Large-Sample Confidence Intervals

Let $X_1, X_2, \cdots, X_n$ be a random sample from a population having a mean μ and standard deviation σ.

Provided that n is large, the Central Limit Theorem (CLT) implies distribution.

It then follows that

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx N(0,1)$$

has approximately a standard normal distribution, so that

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_{\alpha/2}\right) \approx 1 - \alpha,$$

$$\mathbb{P}\left(\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \approx 1-\alpha$$

$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} , \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$: CI with Confidence level $(100)\cdot(1-\alpha)\%$.

Large-Sample Confidence Intervals

Even though σ is unknown, if n is large enough, then

$S = \text{Sample s.d.}$

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

$$\frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has approximately a standard normal distribution, so that

Proposition

The interval

$$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right]$$

is a large-sample confidence interval with confidence level $100(1-\alpha)\%$ for unknown μ.

Generally speaking, $n > 40$ will be sufficient to justify the use of this interval.

framebreak

Large-Sample Confidence Intervals

Example

The article "Gas Cooking, Kitchen Ventilation, and Exposure to Combustion Products" (Indoor Air, 2006: 65–73) reported that for a sample of 50 kitchens with gas cooking appliances monitored during a one-week period, the sample mean CO2 level (ppm) was 654.16, and the sample standard deviation was 164.43.

μ

Calculate a 95% (two-sided) confidence interval for true average CO2 level in the population of all homes from which the sample was selected.

$n = 50$, $\bar{X} = 654.16$, $s = 164.43$

$\alpha = 0.05$, $z_{\frac{\alpha}{2}} = 1.96$

$$\left[\bar{X} - 1.96 \cdot \frac{164.43}{\sqrt{50}},\ \bar{X} + \text{''}\right]$$

$n > 40$ is large enough.

estimation

$$n \geq \left(\frac{2 \cdot z_{\frac{\alpha}{2}} \cdot s}{\text{width}} \right)^2$$

Large-Sample Confidence Intervals

Unfortunately, the choice of sample size to yield a desired interval width is not as straightforward here as it was for the case of known σ.

The only option for an investigator who wishes to specify a desired width is to make an educated guess as to what the value of s might be.

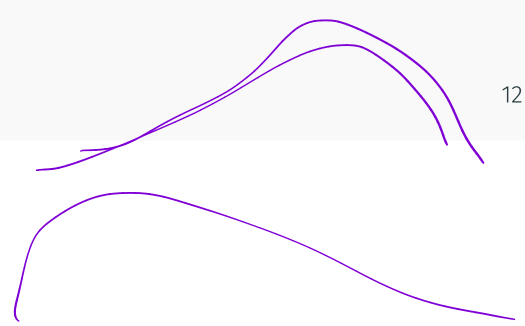
By being conservative and guessing a larger value of σ, an n larger than necessary will be chosen.

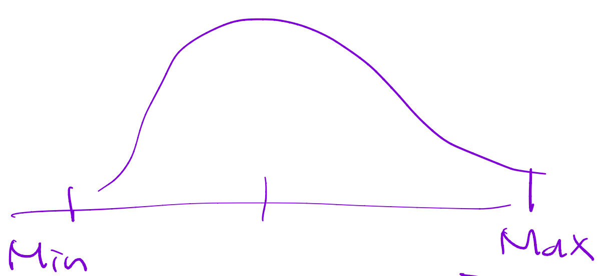
The investigator may be able to specify a reasonably accurate value of the population range (the difference between the largest and smallest values).

Then if the population distribution is not too skewed, dividing the range by 4 gives a ballpark value of what σ might be.

Suggestion to choose s.

① Not too skewed





② Known.

$$s = \frac{\overbrace{\text{Max} - \text{Min}}^{= \text{Range}}}{4}.$$

Large-Sample Confidence Intervals

Example

The charge-to-tap time (min) for carbon steel in one type of open hearth furnace is to be determined for each heat in a sample of size n.

If the investigator believes that almost all times in the distribution are between 320 and 440, what sample size would be appropriate for estimating the true average time to within 5 min. with a confidence level of 95%?

$$\text{Max} = 440 \qquad \text{Range} = 120$$

$$\text{Min} = 320 \qquad s \approx \frac{\text{Range}}{4} = 30$$

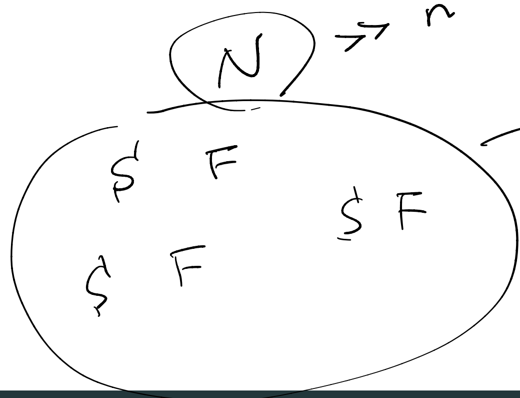
$$\text{Width} = 2 \cdot z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

$$= \frac{60 \cdot 1.96}{\sqrt{n}} \leq 5$$

$$n = (\sqrt{n})^2 \geq \left(\frac{60 \cdot 1.96}{5}\right)^2$$

μ

Final



$$\begin{cases}1\\0\end{cases}$$

almost i.i.d

$$X_1, \ldots, X_n \sim \cancel{\text{i.i.d}}\ \text{Ber}(p)$$

$$\# \text{ of } \$'\text{s} = X_1 + \cdots + X_n = X \sim \cancel{\text{HyperGeom}}$$

almost binomial

A Confidence Interval for a Population Proportion

Let p denote the proportion of "successes" in a population, where success identifies an individual or object that has a specified property.

A random sample of n individuals is to be selected, and X is the number of successes in the sample.

Provided that n is small compared to the population size, X can be regarded as a binomial RV with

$$\mathbb{E}[X] = np$$
$$\text{Var}(X) = n \cdot p \cdot (1-p)$$

$$X = \sum_{i=1}^{n} x_i \approx \text{Bin}(\underbrace{n}_{\text{sample size}}, \underbrace{p}_{\text{parameter to be estimated}})$$

If n large

$$X \underset{\text{CLT}}{\approx} \text{Normal}(np, np(1-p))$$

(Normal App. to Binomial)

$$\frac{X - np}{\sqrt{np(1-p)}} \approx N(0,1)$$

$$P\left(-z_{\frac{\alpha}{2}} \le \frac{X-np}{\sqrt{np(1-p)}} \le z_{\frac{\alpha}{2}}\right) \approx 1-\alpha$$

$$\Downarrow$$

Confidence Interval for $\underline{\underline{p}}$

$-z_{\frac{\alpha}{2}} \le \frac{X-np}{\sqrt{np(1-p)}}$ Solve for p?

$$-z_{\frac{\alpha}{2}}\sqrt{np(1-p)} = X-np$$

$$z_{\frac{\alpha}{2}}^2 \, n \cdot p(1-p) = (X-np)^2$$

$$\vdots$$

A Confidence Interval for a Population Proportion

assumption for CLT

Furthermore, if both $np > 10$ and $n(1-p) > 10$, X has approximately a normal distribution. Let $\hat{p} = X/n$, then

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{\sqrt{n}}}} \leq z_{\alpha/2}\right) \approx 1 - \alpha.$$

If we solve for p, then

$$\hat{p} = \frac{X}{n}$$

A Confidence Interval for a Population Proportion

Definition

A confidence interval for a population proportion p with confidence level approximately $100(1-\alpha)\%$ is

$$\frac{\hat{p} + z_{\alpha/2}^2/(2n)}{1 + z_{\alpha/2}^2/n} \pm z_{\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2/(4n^2)}}{1 + z_{\alpha/2}^2/n}.$$

If the sample size n is very large, the confidence interval can be simplified as

$$\underline{\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}}.$$

$$p \in \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + \text{"}\ \right]$$

with prob. $(1-\alpha)$.

A Confidence Interval for a Population Proportion

CLT works when $np \geq 10$ $\quad n(1-p) \geq 10$

$n \cdot \hat{p} = 16 \quad 32 \geq 10$

Example

The article "Repeatability and Reproducibility for Pass/Fail Data" (J. of Testing and Eval., 1997: 151–153) reported that in $n = 48$ trials in a particular laboratory, 16 resulted in ignition of a particular type of substrate by a lighted cigarette.

Let p denote the long-run proportion of all such trials that would result in ignition.

Find a confidence interval for p with a confidence level of approximately 95%?

\# of success $= 16$

$n = 48$

Point estimate for $p = \hat{p} = \frac{16}{48} = \frac{1}{3}$

Confidence interval for p w/ 95%

$$= \frac{\hat{p} + \frac{(z_{\alpha/2})^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} \pm \frac{z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

$$= \frac{\frac{1}{3} + \frac{(1.96)^2}{96}}{1 + \frac{(1.96)^2}{48}} \pm \frac{1.96\sqrt{\frac{\frac{1}{3}\cdot\frac{2}{3}}{48} + \left(\frac{1.96}{96}\right)^2}}{1 + \frac{1.96}{48}}$$

Exercise

(7.2-19) The article "Limited Yield Estimation for Visual Defect Sources" (IEEE Trans. on Semiconductor Manuf., 1997: 17–23) reported that, in a study of a particular wafer inspection process, 356 dies were examined by an inspection probe and 201 of these passed the probe.

Assuming a stable process, calculate a 95% (two-sided) confidence interval for the proportion of all dies that pass the probe.

7.1

$$\begin{cases} \text{Normal} \\ \sigma \quad \text{known} \end{cases} \Rightarrow \quad \text{CI} \quad \left[\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

7.2

$$\begin{cases} \text{Non-Normal} \\ \sigma \quad \text{known / unknown} \\ n \quad \text{large} \end{cases} \Rightarrow \left[\bar{X} \pm z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$$

7.3

$$\begin{cases} \text{Normal} \\ \sigma \quad \text{unknown} \\ n \quad \text{large or small} \end{cases} \Rightarrow \quad ??$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \approx \text{Normal}$$

What is this presizely if $X_1, \cdots, X_n \sim$ Normal

$$\frac{\frac{1}{n}(X_1 + \cdots + X_n) - \mu}{\frac{1}{\sqrt{n}}\sqrt{\frac{1}{n-1}\sum_1 (X_i - \bar{x})^2}}$$

$\Downarrow$ simplified

"t distribution"

$$\frac{Z}{\sqrt{\frac{V}{n-1}}}$$

$Z \sim N(0,1)$

$V \sim$??

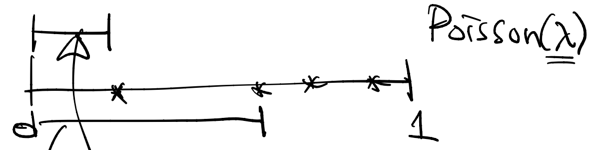
Section 3.
Intervals Based on a Normal Population Distribution

Recall $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$

$\Gamma(n) = (n-1)!$

$\mathrm{Gamma}(\alpha, \beta)$ PDF $f(x) = C \cdot x^{\alpha-1} \cdot e^{-\frac{x}{\beta}}$

Generalization of Exp.



waiting time = Exp(λ)

$\mathrm{Gamma}(2, \frac{1}{\lambda})$

$\chi^2(\nu) = \mathrm{Gamma}\left(\frac{\nu}{2}, 2\right)$

ν = degree of freedom.

Student's t distribution

Let Z be a standard normal RV and V be a χ^2 distribution with degree of freedom ν.

Assume Z and V are independent and define

$$T = \frac{Z}{\sqrt{V/\nu}}.$$

T is called **a t distribution with degree of freedom ν**.

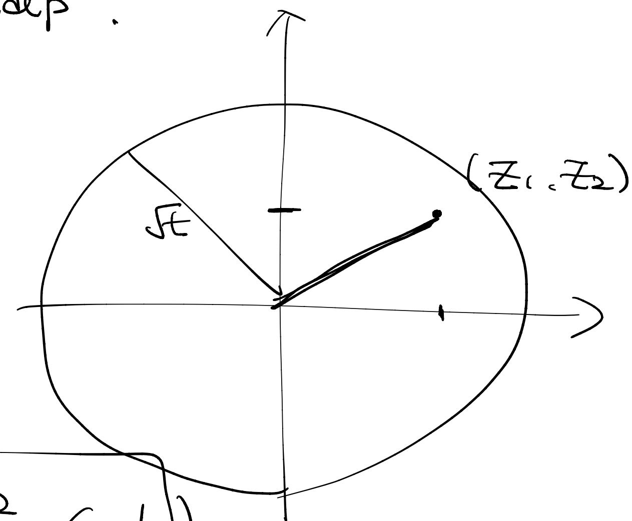
The PDF of T is

$$f(t) = C_\nu \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}.$$

$Z_1, Z_2 \sim N(0,1)$ indep.

$$\underline{V = Z_1^2 + Z_2^2}$$

↑ distribution?



CDF of V

$$= F(t) = P\left(\boxed{Z_1^2 + Z_2^2 \le t}\right)$$

$$= P\left((Z_1, Z_2) \in \underline{A}\right)$$

$$= \iint_A \underbrace{\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{z_1^2}{2}}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{z_2^2}{2}}\right)}_{\text{joint PDF}} dz_1\, dz_2$$

$$= \frac{1}{2\pi} \iint_{\oplus} e^{-\frac{1}{2}(z_1^2 + z_2^2)}\, dz_1\, dz_2$$

↑ centered disc of radius $\sqrt{t}$

$$= \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\sqrt{t}} e^{-\frac{r^2}{2}}\, r\,dr\, d\theta$$

$$= \int_0^{\sqrt{t}} r e^{-\frac{r^2}{2}}\, dr = G(\sqrt{t}) - \cancel{G(0)}$$

$$f(t) = F'(t) = \cancel{\sqrt{t}}\, e^{-\frac{t}{2}} \frac{1}{2\cancel{\sqrt{t}}}$$

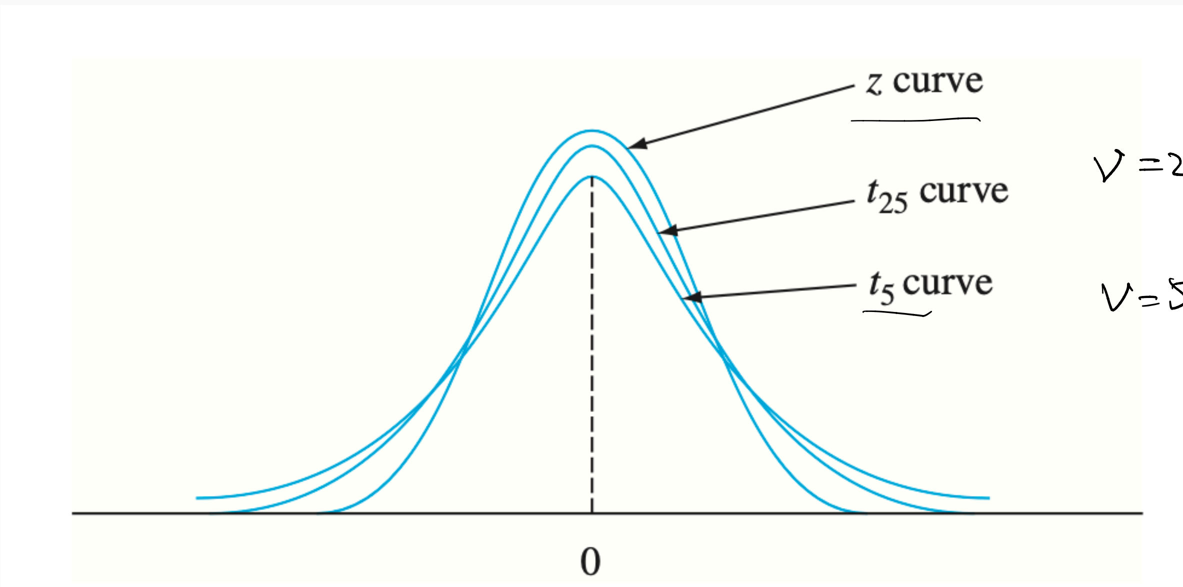
$$= \frac{1}{2} e^{-\frac{t}{2}}$$

$$V \sim \text{Exp}\left(\frac{1}{2}\right) = \chi^2(2) = \text{Gam}\left(\frac{2}{2}, 2\right)$$

$$\mathrm{Exp}(\lambda) = \mathrm{Gamma}\left(1, \frac{1}{\lambda}\right)$$

$$\chi^2(\nu) = \mathrm{Gamma}\left(\frac{\nu}{2}, 2\right)$$

Student's t distribution



$\nu = 25$

$\nu = 5$

Student's t distribution

Let T denote the t distribution with ν.

1. Each t_ν curve is bell-shaped and centered at 0.
2. Each t_ν curve is more spread out than the standard normal curve.
3. As ν increases, the spread of the corresponding t_ν curve decreases.
4. As $\nu \to \infty$, the sequence of t_ν curves approaches the standard normal curve.

Confidence interval for a Normal Population Distribution

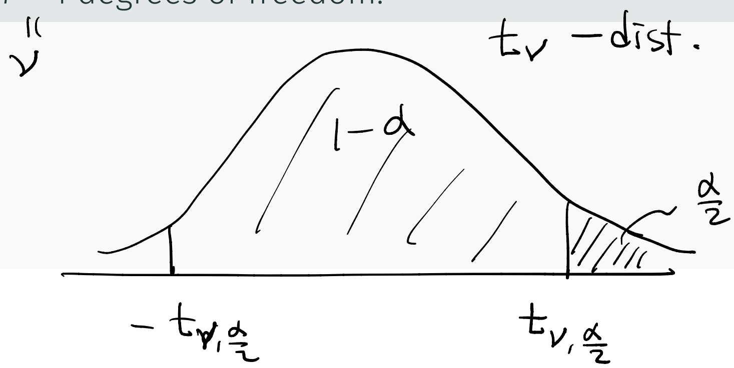
The population of interest is normal, so that $X_1, X_2, \cdots, X_n$ constitutes a random sample from a normal distribution with both μ and σ unknown.

Theorem

When X is the mean of a random sample of size n from a normal distribution with mean μ, the RV

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t distribution with $n-1$ degrees of freedom.



$$\mathbb{P}\left(\overline{X} - t_{n-1,\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{n-1,\frac{\alpha}{2}} \frac{S}{\sqrt{n}}\right) = 1-\alpha$$

Confidence interval.

Confidence interval for a Normal Population Distribution

Proposition

Let $\bar{x}$ and s be the sample mean and sample standard deviation computed from the results of a random sample from a normal population with mean μ.

Then a $100(1-\alpha)\%$ confidence interval for μ is

$$\left[\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right]$$

Confidence interval for a Normal Population Distribution

Example

Even as traditional markets for sweetgum lumber have declined, large section solid timbers traditionally used for construction bridges and mats have become increasingly scarce.

The article “Development of Novel Industrial Laminated Planks from Sweetgum Lumber” (J. of Bridge Engr., 2008: 64–66) described the manufacturing and testing of composite beams designed to add value to low-grade sweetgum lumber.

6807.99	7637.06	6663.28	6165.03	6991.41	6992.23
6981.46	7569.75	7437.88	6872.39	7663.18	6032.28
6906.04	6617.17	6984.12	7093.71	7659.50	7378.61
7295.54	6702.76	7440.17	8053.26	8284.75	7347.95
7422.69	7886.87	6316.67	7713.65	7503.33	7674.99

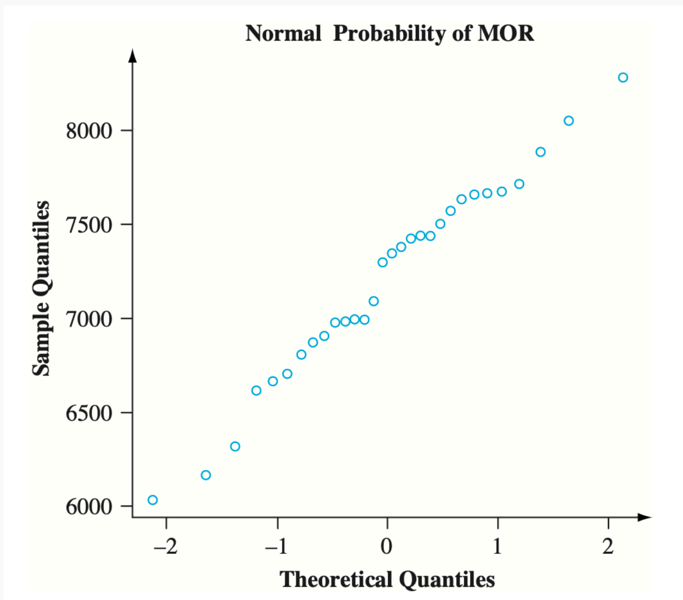
$x_1 < x_2 < \cdots < x_{30}$

: sample percentile

i-th per. $= \dfrac{(i - \frac{1}{2}) \cdot 100}{n}$

zvalue

Confidence interval for a Normal Population Distribution



Exercise

(7.3-33) The article “Measuring and Understanding the Aging of Kraft Insulating Paper in Power Transformers” (IEEE Electrical Insul. Mag., 1996: 28–34) contained the following observations on degree of polymerization for paper specimens for which viscosity times concentration fell in a certain middle range:

418	421	421	422	434	437	439
446	454	463	465	425	427	431
447	448	453				

1. Construct a boxplot of the data and comment on any interesting features.
2. Is it plausible that the given sample observations were selected from a normal distribution?

Exercise

3. Calculate a two-sided 95% confidence interval for true average degree of polymerization (as did the authors of the article). Does the interval suggest that 440 is a plausible value for true average degree of polymerization? What about 450?

Exercise