innovate achieve lead

BITS Pilani
Pilani Campus

# Course No: MATH F113

# Probability and Statistics

BITS Pilani
Pilani Campus

# Chapter 7: Statistical Intervals Based on a Single Sample


**Sumanta Pasari**

**sumanta.pasari@pilani.bits-pilani.ac.in**

# Interval Estimation

- A point estimate cannot be expected to provide the exact value (close value) of the population parameter.
- **Usually**, an interval estimate can be obtained by adding and subtracting a margin of error to the point estimate. Then,

***Interval Estimate = Point Estimate + / – Margin of Error***

- Interval estimation provides us information about **how close** the point estimate is to the value of the parameter.
- Why we use the term ***confidence interval?***

# Interval (CI) Estimation

– Instead of considering a statistic as a point estimator, we may use *random intervals* to trap the parameter.

– In this case, the end points of the interval are RVs and we can talk about the probability that it traps the parameter value.

**Confidence Interval** : A $100(1-\alpha)\%$ confidence interval for a parameter is a **random interval** $[L_1, L_2]$ such that

$$P[L_1 \le \theta \le L_2] = 1-\alpha \text{, regardless the value of } \theta.$$

# Theorem 7.1: Interval estimation for μ: σ known

Let $X_1,\ X_2,\ \cdots, X_n$ be a random sample from a normal population with mean $\mu$ (unknown) and the variance $\sigma^2$ (known). Then, we know,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0,1)$$

Taking two points $\pm z_{\alpha/2}$ symmetrically about the origin, we get

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

Here $(1-\alpha)$ is known as confidence level, and $\alpha$ is the level of significance.

# Interval estimation for μ: σ known

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

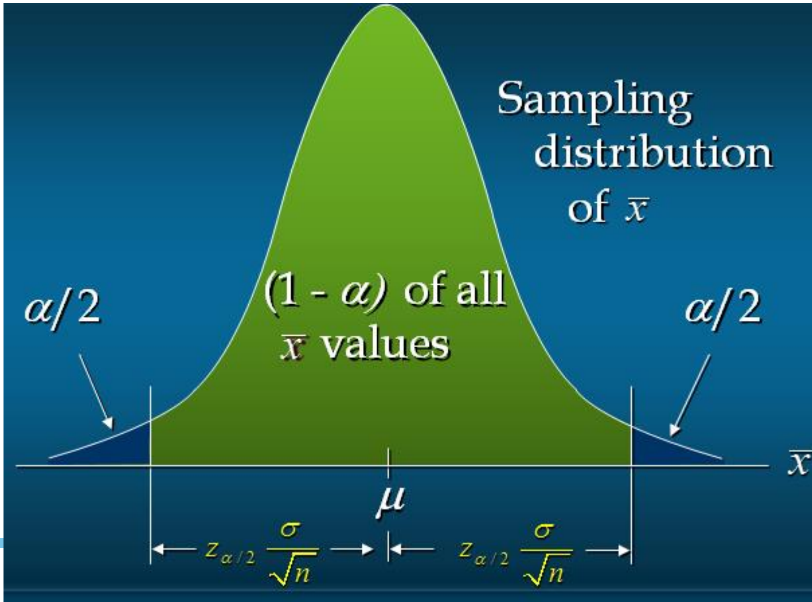
Hence, the confidence interval for population mean $\mu$ having confidence level $100\times(1-\alpha)\%$ is given as $\left(\bar{x}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{x}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$.

The endpoints of the confidence interval is called confidence limits.

Interval Estimate of $\mu$

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

where:
- $\bar{x}$ is the sample mean
- $1-\alpha$ is the confidence coefficient
- $z_{\alpha/2}$ is the z value providing an area of $\alpha/2$ in the upper tail of the standard normal probability distribution
- $\sigma$ is the population standard deviation
- $n$ is the sample size



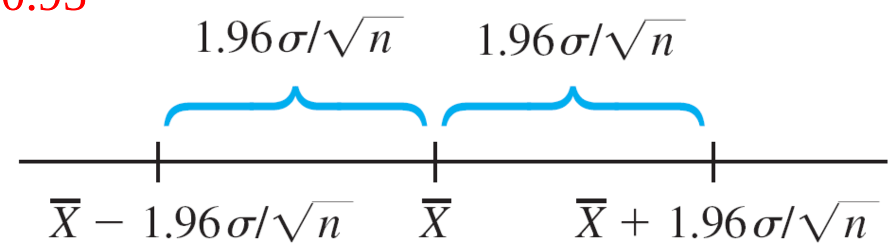
# Interval estimation for μ: σ known

## Most commonly used confidence levels:

| Confidence Level | $\alpha$ | $\alpha/2$ | Table Look-up Area | $z_{\alpha/2}$ |
|---|---|---|---|---|
| 90% | .10 | .05 | .9500 | 1.645 |
| 95% | .05 | .025 | .9750 | 1.960 |
| 99% | .01 | .005 | .9950 | 2.576 |

Hence, 95% CI for $\mu$ is given as $\left(\bar{x} - 1.96\dfrac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\dfrac{\sigma}{\sqrt{n}}\right)$.

That is, $P\left(\bar{X} - 1.96\dfrac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\dfrac{\sigma}{\sqrt{n}}\right) = 0.95$



# Practice Problems

**Ex.1.** The mean of a sample size 50 from a normal population is observed to be 15.68. If the s.d. of the population is 3.27, find (a) 80% (b) 95%, (c) 99% confidence interval for the population mean. Can you find out the respective margin of errors? What is the length of CI for each case?

**Sol. (b) First check the two assumptions : (i) normality (ii)** $\sigma$ known

Step 1: Here $n = 50$, $\bar{x} = 15.68$, $\sigma = 3.27$, and $\alpha = 0.05$. We need CI for $\mu$.

Step 2: As $\alpha = 0.05$, we need to find $z_{\alpha/2}$ such that $P\left(Z < z_{\alpha/2}\right) = 0.975$.

From cumulative normal distribution table, we see $z_{\alpha/2} = 1.96$.

Step 3: The CI for $\mu\left(\sigma \text{ known}\right)$ is $\left(\bar{x} - \frac{\sigma}{\sqrt{n}} z_{0.025}, \bar{x} + \frac{\sigma}{\sqrt{n}} z_{0.025}\right) = \left(14.77, 16.59\right)$

# Interval estimation for $\mu$: $\sigma$ known

***Confusions and confusions? ☹ ☹***

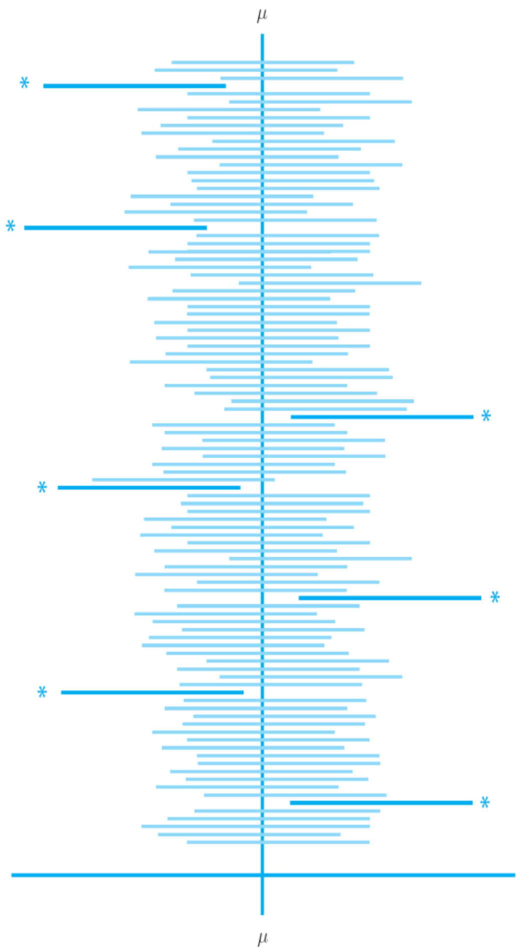
In the above example, we found 95% CI for $\mu$ is (14.77, 16.59). This means, the unknown $\mu$ lies within the fixed interval with probability 0.95. That is,

$$P\,[\mu \text{ lies in } (14.77, 16.59)] = 0.95$$ **– right?**

**If not, then what is the interpretation of "95% confidence"?**

- Long run relative frequency?
- A single replication/realization of random interval is not enough! Not satisfactory, at least.

# 95% CIs for population mean




One hundred 95% CIs (asterisks identify intervals that do not include $\mu$).

# Practice Problems

HW 1. Studies have shown that the random variable X, the processing time required to do a multiplication on a new 3-D computer, is normally distributed with mean μ and standard deviation 2 microseconds. A random sample of 16 observations is to be taken

(a) These data are obtained

| | | | |
|---|---|---|---|
| 42.65 | 45.15 | 39.32 | 44.44 |
| 41.63 | 41.54 | 41.59 | 45.68 |
| 46.50 | 41.35 | 44.37 | 40.27 |
| 43.87 | 43.79 | 43.28 | 40.70 |

Based on these data, find an unbiased estimate for μ.

(b) Find a 95% confidence interval for μ.

# Confidence Level, Precision, and Sample Size

**Ex.2.** 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 $\mu$ 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?

# Confidence Level, Precision, and Sample Size

The sample size *n* must satisfy

$$10 = 2 \cdot (1.96)(25/\sqrt{n})$$

Rearranging this equation gives

$$\sqrt{n} = 2 \cdot (1.96)(25)/10 = 9.80$$

So

$$n = (9.80)^2 = 96.04$$

Since *n* must be an integer, a sample size of 97 is required.

# Deriving a CI: Example 7.5

A theoretical model suggests that the time to breakdown of an insulating fluid between electrodes at a particular voltage has an exponential distribution with parameter $\lambda$.

A random sample of $n$ = 10 breakdown times yields the following sample data (in min): $x_1 = 41.53$, $x_2 = 18.73$, $x_3 = 2.99$, $x_4 = 30.34$, $x_5 = 12.33$, $x_6 = 117.52$, $x_7 = 73.02$, $x_8 = 223.63$, $x_9 = 4.00$, $x_{10} = 26.78$.

A 95% CI for $\lambda$ and for the true average breakdown time are desired.

# Example 7.5

Let $h(X_1, X_2, \ldots, X_n; \lambda) = 2\lambda\Sigma X_i$. It can be shown that this random variable has a probability distribution called a chi-squared distribution with $2n$ degrees of freedom (df) ($\nu = 2n$, where $\nu$ is the parameter of a chi-squared distribution).

(https://en.wikipedia.org/wiki/Relationships_among_probability_distributions)

Appendix Table A.7 pictures a typical chi-squared density curve and tabulates critical values that capture specified tail areas. The relevant number of df here is 2(10) = 20.

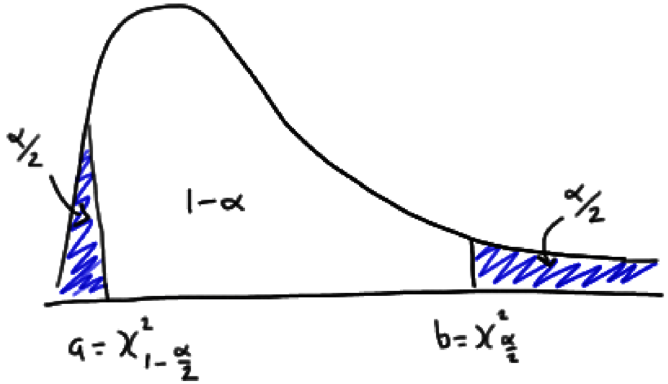
The $\nu = 20$ row of the table shows that 34.170 captures upper-tail area .025 and 9.591 captures lower-tail area .025 (upper-tail area .975).

# Example 7.5

Thus for $n$ = 10,

$$P(9.591 < 2\lambda\Sigma X_i, 34.170) = .95$$

Division by




$2\Sigma X_i$ isolates $\lambda$, yielding

$$P(9.591/(2\Sigma X_i) < \lambda < (34.170/(2\Sigma X_i)) = .95$$

The lower limit of the 95% CI for $\lambda$ is $9.591/(2\Sigma X_i)$, and the upper limit is $34.170/(2\Sigma X_i)$.

# Example 7.5

For the given data, $\Sigma x_i = 550.87$, giving the interval (.00871, .03101).

The expected value of an exponential rv is $\mu = 1/\lambda$.

Since

$$P((2\Sigma X_i/34.170 < 1/\lambda < (2\Sigma X_i/9.591) = .95$$

the 95% CI for true average breakdown time is

$$(2\Sigma x_i/34.170,\ 2\Sigma x_i/9.591) = (32.24,\ 114.87).$$

This interval is obviously quite wide, reflecting substantial variability in breakdown times and a small sample size.

# Impracticality of Assumptions in CI

innovate achieve lead

In practice, we usually face mainly two problems in application of previous C.I. formula.

- What if the population is not normal? **(large sample size is needed) - *Can we take help from CLT?***
- What if the population variance is unknown?

# 7.2: Large Sample CI for μ

Let $X_1,\ X_2,\ \cdots, X_n$ (large sample) be a random sample from a with mean $\mu$ (unknown) and the variance $\sigma^2$ (known). Then, using CLT,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0,1)$$

$100\times(1-\alpha)\%$ CI for $\mu$ is given as $\left(\bar{x}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{x}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$.

# 7.2: Large Sample CI for μ

Let $X_1,\ X_2,\ \cdots, X_n$ be a random sample from a large sample $(n \geq 40)$ with mean $\mu$ (unknown) and sample variance $S^2$. Then,

$$\frac{\bar{X}-\mu}{\left(\frac{S}{\sqrt{n}}\right)} \sim N(0,1) \quad \text{[approximately, a standard normal distribution]}$$

$$\Rightarrow \quad P\left(\bar{X}-\frac{S}{\sqrt{n}} z_{\alpha/2} < \mu < \bar{X}+\frac{S}{\sqrt{n}} z_{\alpha/2}\right) = 1-\alpha$$

Hence, the large-sample confidence interval for population mean $\mu$ having confidence level $100\times(1-\alpha)\%$ is approximately given as $\left(\bar{x}-\frac{s}{\sqrt{n}} z_{\alpha/2},\ \bar{x}+\frac{s}{\sqrt{n}} z_{\alpha/2}\right)$, $n \geq 40$ is needed.

# HW 2

Udaipur in Rajasthan is one of the most attractive tourist destinations for foreigners. The average cost per night of a deluxe hotel room in Udaipur city is Rs. 2500. Assuming that this estimate is based on a sample of 144 hotels and that the sample standard deviation is Rs. 600, what is the 95% confidence interval estimate of the population mean? You may use $P(Z < -1.96) = 0.025$. **[4]**

# One Sided CI for μ

The confidence intervals discussed thus far give both a lower confidence bound *and* an upper confidence bound for the parameter being estimated.

In some circumstances, an investigator will want only one of these two types of bounds.

For example, a psychologist may wish to calculate a 95% upper confidence bound for true average reaction time to a particular stimulus, or a reliability engineer may want only a lower confidence bound for true average lifetime of components of a certain type.

# Sample/Population Proportion



Let us draw a random sample $X_1, X_2, \ldots, X_n$ of size $n$ from population, where

$X_i$ = 1, if the i-th member of the sample has the trait

= 0, if i-th sample does not have the trait

Then, $X = \sum_{i=1}^{n} X_i$ gives the number of objects in the sample with the trait and the **statistic** $\boldsymbol{X/n}$ gives the proportion of the sample with the trait. Note that $\boldsymbol{X}$ is a binomial RV with parameters $\boldsymbol{n}$ (known) and $\boldsymbol{p}$.

# Sample Proportion

The statistic that estimates the parameter $p$, a proportion of a population that has some property, is the sample proportion

$$\hat{p} = \frac{\text{number in sample with the trait (success)}}{\text{sample size}} = \frac{X}{n}$$

Properties:

(i) As the sample size increases ($n$ large), the sampling distribution of $\hat{p}$ becomes approximately normal (WHY?)

(ii) The mean of $\hat{p}$ is $p$, and variance of $\hat{p}$ is $\dfrac{p(1-p)}{n}$ (WHY?)

(iii) Can we get estimators of $p$? Point and interval estimator

# Sample Proportion

Note that $\hat{p} = \dfrac{X}{n} = \dfrac{\sum_{1}^{n} X_i}{n} = \bar{X}$,

where, each $X_i$ is an independent point binomial (Bernoulli RV),

that is, $P\left(X_i = 1\right) = p$ and $P\left(X_i = 0\right) = 1 - p$

| $x_i$ | 1 | 0 |
|---|---|---|
| $f(x_i)$ | $p$ | $1-p$ |

$$E[X_i] = 1(p)+0(1-p)=p$$

$$Var(X_i) = E[X_i^2]-(E[X_i])^2 = p(1-p)$$

# Form of Sampling Distribution of Sample Proportion

If $np \geq 10$ and $n(1-p) \geq 10$, sampling distribution of $\hat{p}$ can be approximated by a normal distribution with mean $p$ and s.d. $\sqrt{\dfrac{p(1-p)}{n}}$.

Knowing, $E(\hat{p}) = 0.60$ and $\sigma_{\hat{p}} = 0.0894$, can we find

(i) $P(0.55 < \hat{p} < 0.65) = ?$

(ii) $P(0.50 < \hat{p} < 0.65) = ?$

# (Traditional) Confidence Interval on p

**Interval Estimate = Point Estimate + / − Margin of Error**

Note that for large $n$ (using CLT),

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right) \Rightarrow \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

Taking two points $\pm z_{\alpha/2}$ symmetrically about the origin, we get

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

Here $(1-\alpha)$ is known as confidence level.

# Confidence Interval on p



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

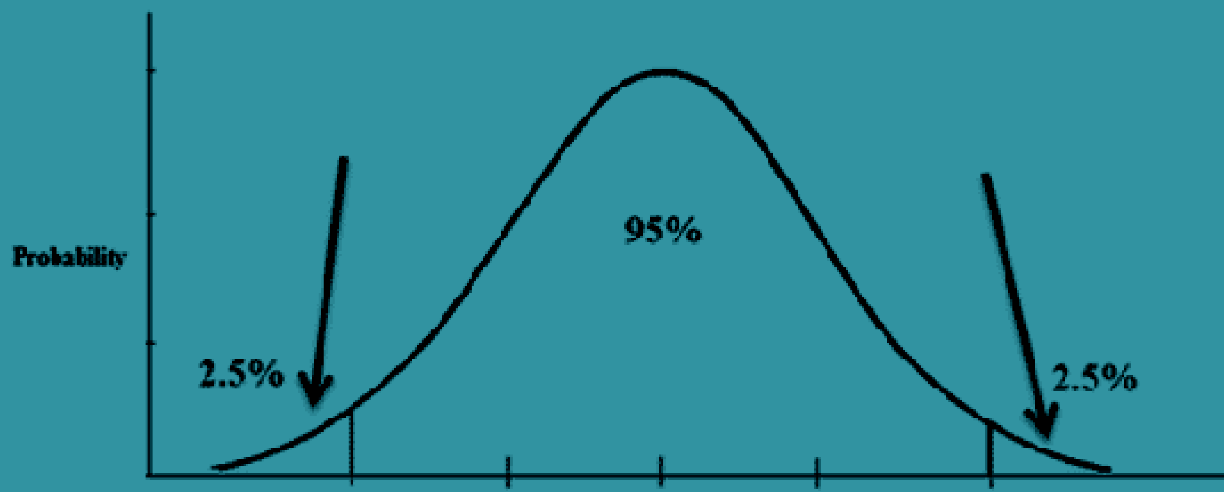
As $p$ is unknown, above confidence bounds are not statistics. So replace $p$ by unbiased estimator $\hat{p}$, and then the CI on $p$ having confidence level $(1-\alpha)$ is

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

The endpoints of the confidence interval is called confidence limits.

# Confidence Interval on p

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




$\alpha = 0.05$ $\quad$ $z_{0.05/2} = z_{0.025}$

# Sample Size for Estimating p

We can be 100(1-α)% sure that $\hat{p}$ and $p$ differ by at most $d$ , where $d$ is given by

$$d = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Thus, sample size for estimating $p$, when prior estimate available is

$$n = z_{\alpha/2}^{2}\frac{\hat{p}(1-\hat{p})}{d^{2}}$$

# Sample Size for Estimating p

It can be shown that the value of $\hat{p}(1-\hat{p}) \le \frac{1}{4}.$

Thus, sample size for estimating $p$, when

prior estimate is not available is $n = \frac{z_{\alpha/2}^{2}}{4d^{2}}.$

# Problem Solving

## *Ex 3*

A study of electromechanical protection devices used in electrical power systems showed that of 193 devices that failed when tested, 75 were due to mechanical part failures.

a) Find a point estimate for p, the proportion of failures that are due to mechanical failures.

b) Find a 95% confidence interval on p.

c) How large a sample is required to estimate p to within 0.03 with 95% confidence.

# Problem Solving

Random variable X= number of failed devices which were due to mechanical failure among 193 failed devices.

X has approx. normal dist with mean = 193p, variance=193p(1-p).

a) Point estimate for p = $\hat{p}_{obs}$ = *x/n* = 75/193 = 0.3886.

b) 95% confidence interval on p is

$$\frac{75}{193} \pm z_{0.025}\sqrt{\frac{75}{193}\left(1-\frac{75}{193}\right)\Big/193} = (0.3198, 0.4574)$$

$$\text{(c) (with using prior estimate)} \;\; n = \frac{1.96^2(0.389)(0.611)}{(0.03^2)} \sim 1015$$

$$\text{(without using prior estimate)} \; n = \frac{z_{\alpha/2}^2}{4d^2} = \frac{1.96^2}{(4)(0.03^2)} \sim 1068.$$

# Score CI for p

Proceeding as suggested in the subsection "Deriving a Confidence Interval", the confidence limits result from replacing each < by = and solving the resulting quadratic equation for *p*. This gives the two roots

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

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

# Score CI for p

**Proposition**

Let $\tilde{p} = \dfrac{\hat{p} + z^2_{\alpha/2}/2n}{1 + z^2_{\alpha/2}/n}$. Then a **confidence interval for a population proportion *p*** with confidence level approximately $100(1 - \alpha)$ % is

$$\tilde{p} \pm z_{\alpha/2} \frac{\sqrt{\hat{p}\hat{q}/n + z^2_{\alpha/2}/4n^2}}{1 + z^2_{\alpha/2}/n} \qquad \textbf{(7.10)}$$

It is thus not necessary to check the conditions $n\hat{p} \geq 10$ and $n(1 - \hat{p}) \geq 10$ that would be required were the traditional interval employed.

# Score CI for p (Example 8; Page 281)



The article "Repeatability and Reproducibility for Pass/Fai Data" (*J. of Testing and Eval.*, 1997: 151–153) reported th 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. A point estimate for $p$ is $\hat{p}$ = 16/48 = .333. A confidence interval for $p$ with a confidence level of approximately 95% is

$$\frac{.333 + (1.96)^2/96}{1 + (1.96)^2/48} \pm (1.96)\frac{\sqrt{(.333)(.667)/48 + (1.96)^2/9216}}{1 + (1.96)^2/48}$$

# Score CI for p (Example 8; Page 281)

$$= .345 \pm .129$$

$$= (.216, .474)$$

his interval is quite wide because a sample size of 48 is iot at all large when estimating a proportion.

he traditional interval is

$$.333 \pm 1.96\sqrt{(.333)(.667)/48} = .333 \pm .133$$

$$= (.200, .466)$$

# 7.3: CI based on a normal population, Interval estimation for μ: σ unknown

innovate achieve lead

*Interval Estimate for $\mu$ = $\overline{X}$ + / – Margin of Error*

- If **sample size is small** and the population standard deviation ***σ*** is unknown, then?
  - Like previously in large-sample case, we can use sample standard deviation ***S*** to calculate the **margin of error.**
- With **small sample**, what is the distribution of RV obtained by replacing σ by S? Can it be a normal distribution, as earlier?
  - In this case, ***assuming that the population is normal,*** we would obtain a ***T-distribution – HOW?***

# Interval estimation for μ: σ unknown

innovate achieve lead

**Note :**

Let $Z \sim N(0,1)$ and $\chi_\gamma^2$ be an independent chi-squared RV with $\gamma$ degrees of freedom, then $T_\gamma = \dfrac{Z}{\sqrt{\chi_\gamma^2 / \gamma}}$ follows a $T$-distribution with $\gamma$ dof.

**Theorem :**

Let $X_1, X_2, \cdots, X_n$ be a random sample of size $n$ from a normal distribution with mean $\mu$ and variance $\sigma^2$. Then, $\dfrac{\bar{X} - \mu}{S / \sqrt{n}} \sim T_{n-1}$

# T-Distribution

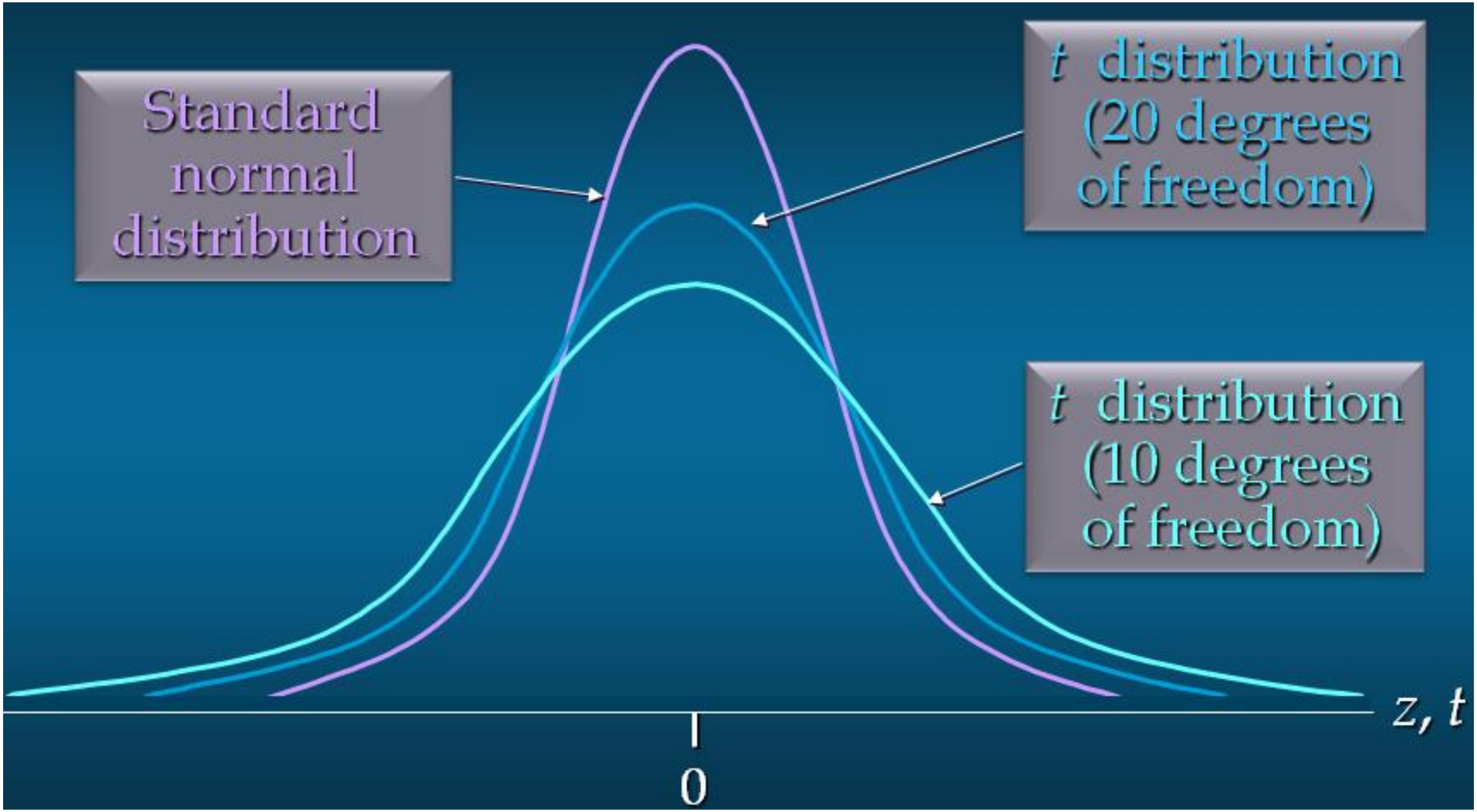


- Random variable T with γ degrees of freedom (called parameter) is a continuous r.v. with density

$$f(t) = \frac{\Gamma(\gamma+1)/2}{\Gamma(\gamma/2)\sqrt{\pi\gamma}}\left(1+\frac{t^2}{\gamma}\right)^{-(\gamma+1)/2} ;-\infty < t < \infty.$$

- Density plot is bell shaped, symmetric about 0.
- Variance of T decreases as γ increases. In fact T approximates standard normal for large γ.

# T-Distribution



# Properties of T-distribution

Let $t_\nu$ denote the $t$ distribution with $\nu$ df.

**1.** Each $t_\nu$ curve is bell-shaped and centered at 0.

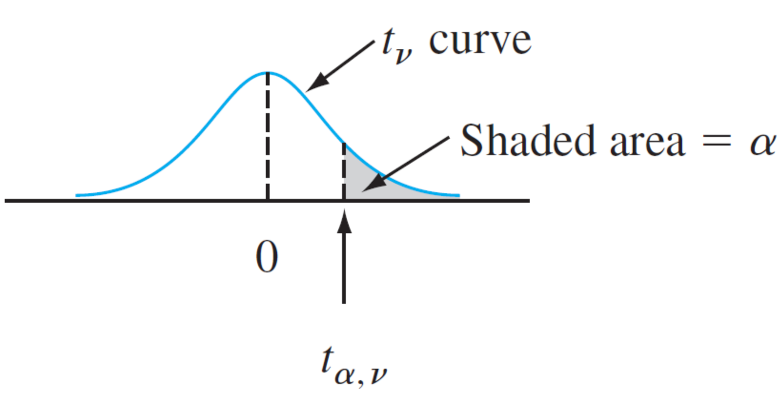
**2.** Each $t_\nu$ curve is more spread out than the standard normal ($z$) curve.

**3.** As $\nu$ increases, the spread of the corresponding $t_\nu$ curve decreases.

**4.** As $\nu \to \infty$, the sequence of $t_\nu$ curves approaches the standard normal curve (so the $z$ curve is often called the $t$ curve with df $= \infty$).

# CDF of T-distribution

- Critical values for t-distribution are given in Table A.5.
- By $t_{\alpha,v}$ we denote the value of the t-variable such that area under its density to its right is $\alpha$. (degrees of freedom $v$ must be mentioned separately).



# Interval estimation for μ: σ unknown

The $T$-distribution is symmetric, and becomes approx. std. normal for large $n$.

Taking two points $\pm t_{\alpha/2}$ symmetrically about the origin, we get

$$P\left(-t_{\alpha/2,n-1} < \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} < t_{\alpha/2,n-1}\right) = 1-\alpha$$

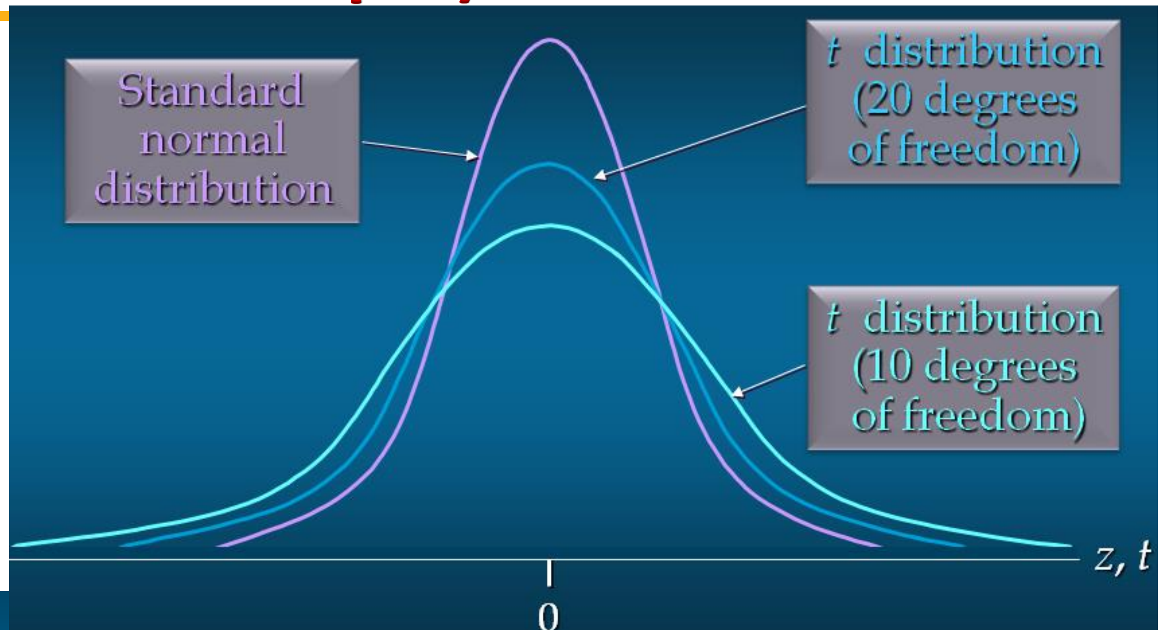
Here $(1-\alpha)$ is known as confidence level, and $\alpha$ is the level of significance. So,

$$P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2,n-1} < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2,n-1}\right) = 1-\alpha$$

Hence, the CI for $\mu$ $(\sigma$ unknown$)$ having confidence level $100\times(1-\alpha)\%$

is given as $\left(\bar{x} - \frac{S}{\sqrt{n}} t_{\alpha/2,n-1},\ \bar{x} + \frac{S}{\sqrt{n}} t_{\alpha/2,n-1}\right)$.

# Interval estimation for μ: σ unknown (small sample)




Interval Estimate

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

where: $1 - \alpha$ = the confidence coefficient

$t_{\alpha/2}$ = the $t$ value providing an area of $\alpha/2$ in the upper tail of a $t$ distribution with $n - 1$ degrees of freedom

$s$ = the sample standard deviation

# Examples

**Ex.4.** Seven laboratory experiments of the value of $g$ (accelearation due to gravity that follows normal distribution) at Pilani gave a mean 977.51 cm/s$^2$ and a s.d. 4.42 cm/s$^2$. Find 95% CI for the true value of g (i.e., population mean).

**Sol.**

Step 1: Here $n = 7$, $\bar{x} = 977.51$, $s$=4.42, and $\alpha = 0.05$. We need CI for $\mu$. Population is also known to be normal dist.

Step 2: As $\alpha = 0.05$, we need to find $t_{\alpha/2}$ from t-distribution with $(n\text{-}1) = 6$ degree of freedom, such that $P\left(T < t_{\alpha/2}\right)$=0.975.

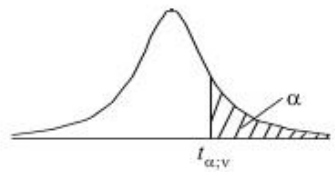
From t-distribution table, we see $t_{0.025,6} = 2.447$.

Step 3: The CI for $\mu\left(\sigma \text{ unknown}\right)$ is $\left(\bar{x} - \dfrac{s}{\sqrt{n}} t_{0.025}, \bar{x} + \dfrac{s}{\sqrt{n}} t_{0.025}\right)$

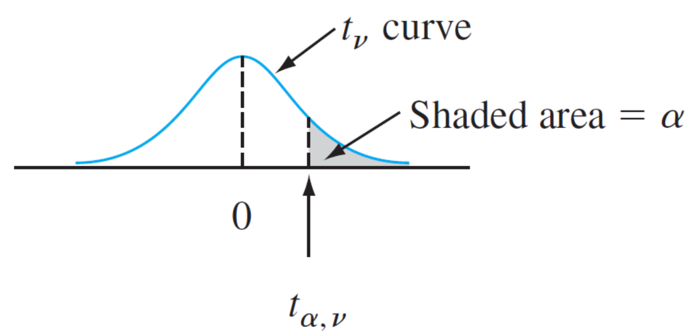
$$= \left(973.09, 981.93\right)$$

# Table of the Student's $t$-distribution

The table gives the values of $t_{\alpha;\nu}$ where
$\Pr(T_\nu > t_{\alpha;\nu}) = \alpha$, with $\nu$ degrees of freedom

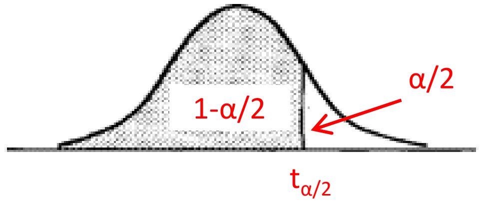



| ν \ α | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
|---|---|---|---|---|---|---|---|
| 1 | 3.078 | 6.314 | 12.076 | 31.821 | 63.657 | 318.310 | 636.620 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.326 | 31.598 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.213 | 12.924 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 | 3.767 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 | 3.725 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.690 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 | 3.396 | 3.659 |
| 30 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 | 3.385 | 3.646 |
| 40 | 1.303 | 1.684 | 2.021 | 2.423 | 2.704 | 3.307 | 3.551 |
| 60 | 1.296 | 1.671 | 2.000 | 2.390 | 2.660 | 3.232 | 3.460 |
| 120 | 1.289 | 1.658 | 1.980 | 2.358 | 2.617 | 3.160 | 3.373 |
| ∞ | 1.282 | 1.645 | 1.960 | 2.326 | 2.576 | 3.090 | 3.291 |




https://en.wikipedia.org/wiki/Student%27s_t-distribution

**Similar to the table A.5**




Degree of freedom

| | Cumulative Probability | | | | | | |
|---|---|---|---|---|---|---|---|
| $\nu$ | .60 | .70 | .80 | .85 | .90 | .95 | .975 |
| 1 | 0.325 | 0.727 | 1.376 | 1.963 | 3.078 | 6.314 | 12.706 |
| 2 | 0.289 | 0.617 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 |
| 3 | 0.277 | 0.584 | 0.978 | 1.250 | 1.638 | 2.353 | 3.182 |
| 4 | 0.271 | 0.569 | 0.941 | 1.190 | 1.533 | 2.132 | 2.776 |
| 5 | 0.267 | 0.559 | 0.920 | 1.156 | 1.476 | 2.015 | 2.571 |
| 6 | 0.265 | 0.553 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 |
| 7 | 0.263 | 0.549 | 0.896 | 1.119 | 1.415 | 1.895 | [illegible] |
| 8 | 0.262 | 0.546 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 |
| 9 | 0.261 | 0.543 | 0.883 | 1.100 | 1.383 | 1.833 | 2.262 |
| 10 | 0.260 | 0.542 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 |
| 11 | 0.260 | 0.540 | 0.876 | 1.088 | 1.363 | 1.796 | 2.201 |
| 12 | 0.259 | 0.539 | 0.873 | 1.083 | 1.356 | 1.782 | 2.179 |
| 13 | 0.259 | 0.537 | 0.870 | 1.079 | 1.350 | 1.771 | 2.160 |
| 14 | 0.258 | 0.537 | 0.868 | 1.076 | 1.345 | 1.761 | 2.145 |
| 15 | 0.258 | 0.536 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 |
| 16 | 0.258 | 0.535 | 0.865 | 1.071 | 1.337 | 1.746 | 2.120 |
| 17 | 0.257 | 0.534 | 0.863 | 1.069 | 1.333 | 1.740 | 2.110 |
| 18 | 0.257 | 0.534 | 0.862 | 1.067 | 1.330 | 1.734 | 2.101 |
| 19 | 0.257 | 0.533 | 0.861 | 1.066 | 1.328 | 1.729 | 2.093 |
| 20 | 0.257 | 0.533 | 0.860 | 1.064 | 1.325 | 1.725 | 2.086 |
| 21 | 0.257 | 0.532 | 0.859 | 1.063 | 1.323 | 1.721 | 2.080 |
| 22 | 0.256 | 0.532 | 0.858 | 1.061 | 1.321 | 1.717 | 2.074 |
| 23 | 0.256 | 0.532 | 0.858 | 1.060 | 1.319 | 1.714 | 2.069 |
| 24 | 0.256 | 0.531 | 0.857 | 1.059 | 1.318 | 1.711 | 2.064 |
| 25 | 0.256 | 0.531 | 0.856 | 1.058 | 1.316 | 1.708 | 2.060 |
| 26 | 0.256 | 0.531 | 0.856 | 1.058 | 1.315 | 1.706 | 2.056 |
| 27 | 0.256 | 0.531 | 0.855 | 1.057 | 1.314 | 1.703 | 2.052 |
| 28 | 0.256 | 0.530 | 0.855 | 1.056 | 1.313 | 1.701 | 2.048 |
| 29 | 0.256 | 0.530 | 0.854 | 1.055 | 1.311 | 1.699 | 2.045 |
| 30 | 0.256 | 0.530 | 0.854 | 1.055 | 1.310 | 1.697 | 2.042 |
| 40 | 0.255 | 0.529 | 0.851 | 1.050 | 1.303 | 1.684 | 2.021 |
| 60 | 0.254 | 0.527 | 0.848 | 1.045 | 1.296 | 1.671 | 2.000 |
| 120 | 0.254 | 0.526 | 0.845 | 1.041 | 1.289 | 1.658 | 1.980 |
| $\infty$ | 0.253 | 0.524 | 0.842 | 1.036 | 1.282 | 1.645 | 1.960 |

**Similar to the table A.5**

# Example 11 (Page 288)

- 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.

# Example 11 (Page 288)

cont'd

- Here is data on the modulus of rupture (psi; the article contained summary data expressed in MPa):

- 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

# Example 11 (Page 288)

cont'd

- Let's now calculate a confidence interval for true average MOR using a confidence level of 95%. The CI is based on $n - 1 = 29$ degrees of freedom, so the necessary $t$ critical value is $t_{.025,29} = 2.045$. The interval estimate is now

$$\bar{x} \pm t_{.025,29} \cdot \frac{s}{\sqrt{n}} = 7203.191 \pm (2.045) \cdot \frac{543.5400}{\sqrt{30}}$$

$$= 7203.191 \pm 202.938$$

$$= (7000.253, 7406.129)$$

- We estimate $7000.253 < \mu < 7406.129$ that with 95% confidence.

# Example 11 (Page 288)



cont'd

- If we use the same formula on sample after sample, in the long run 95% of the calculated intervals will contain $\mu$. Since the value of $\mu$ is not available, we don't know whether the calculated interval is one of the "good" 95% or the "bad" 5%.

- A lower 95% confidence bound would result from retaining only the lower confidence limit (the one with –) and replacing 2.045 with $t_{.05,29} = 1.699$.

# Examples

**HW.3.** Seven laboratory experiments of the value of $g$ (accelearation due to gravity that follows normal distribution) at Pilani gave a mean 977.51 cm/s$^2$ and a s.d. 4.42 cm/s$^2$. Find 80%, 90% and 95% CIs for the population mean. Can you find out the respective margin of errors (bound on the the error of estimation)?

**HW.4.** A sample of size 15 taken from a larger population (normally dist.) has a sample mean 12, and sample variance 25. Construct 95% CI for population mean when population s.d. is 5. What is the length of CI?

**HW .5.** To estimate the average number of pounds of copper recovered per ton of ore mine, a sample of 150 tons of ores is monitored. A sample mean of 11 pounds with a sample s.d. of 3 pounds was obtained. Construct a 95% confidence interval on the mean number of pounds of copper recovered per ton of ore mined. Assume normaility of the population.

# Examples

**HW 6.** A credit card company wants to determine the mean income of its card holders. It also wants to find if there are any differences in mean income between males and females. A random sample of 225 male card holders and 190 female card holders was drawn. The mean and s.d. for male card holders are Rs. 16450 and Rs. 3675, resp. For female card holders, these are Rs. 13220 and Rs. 3050.
(i) Calculate 90% and 95% CIs for mean income of male and female card holders.
(ii) Do you think that, on average, males' and females' incomes differ?

**HW.7.** A reporter for a student newspaper is writing an article on the cost of off-campus housing. A sample of 16 one-bedroom apartments within a half-mile of campus resulted in a sample mean rent (in rs.) of 750 per month and a sample s.d. of 55. Find 70%, 95%, and 99% CIs estimate of the mean rent per month for the population. Any assumtion must be stated clearly.

# One Sided CI

HW 8. One sided confidence interval can be used to approximate the maximum and minimum value of the population mean.

An interval $(-\infty, L_1]$ such that $P(\mu \le L_1) = 1-\alpha$ allows us to place bounds on the maximum value of population mean

$$L_1 = \bar{X} + t_{\alpha, n-1} S / \sqrt{n}$$

An interval $[L_2, \infty)$ such that $P(\mu \ge L_2) = 1-\alpha$ allows us to place bounds on the minimum value of population mean

$$L_2 = \bar{X} - t_{\alpha, n-1} S / \sqrt{n}$$

# One Sided CI

Use the following data on X, the time that a commercial airliner stays at the gate during a through flight, to find a 95% one sided confidence interval that puts a bound on the minimum time in minutes for μ:

25 29 32 37 40 27 30 35 38 41 42 45 45 47 49 50 55 53 60 (Assume normality of population)

**Solution :** $\bar{x} = 41.05, s^2 = 98.61, \; s = 9.93$

$$L_{obs} = \bar{x} - t_{\alpha/2} \frac{s}{\sqrt{n}} = 41.05 - 1.734 \times \frac{9.93}{\sqrt{19}} = 37.10$$

$$CI = [37.10, \infty)$$

# A Prediction Interval for a Single Future Value

- In many applications, the objective is to *predict* a single value of a variable to be observed at some future time, rather than to *estimate* the mean value of that variable.

# Example 12 (Page 289)

- Consider the following sample of fat content (in percentage) of $n$ = 10 randomly selected hot dogs ("Sensory and Mechanical Assessment of the Quality of Frankfurters," *J. of Texture Studies,* 1990: 395–409):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 25.2 | 21.3 | 22.8 | 17.0 | 29.8 | 21.0 | 25.5 | 16.0 | 20.9 | 19.5 |

- Assuming that these were selected from a normal population distribution, a 95% CI for (interval estimate of) the population mean fat content is

$$\bar{x} \pm t_{.025,9} \cdot \frac{s}{\sqrt{n}} = 21.90 \pm 2.262 \cdot \frac{4.134}{\sqrt{10}} = (18.94, 24.86)$$

# Example 12

cont'd

- Suppose, however, you are going to eat a single hot dog of this type and want a *prediction* for the resulting fat content.

- A *point* prediction, analogous to a *point* estimate, is just = 21.90. This prediction unfortunately gives no information about reliability or precision.

# A Prediction Interval for a Single Future Value

- The general setup is as follows: We have available a random sample $X_1, X_2, \ldots, X_n$ from a normal population distribution, and wish to predict the value of $X_{n+1}$, a single future observation (e.g., the lifetime of a single lightbulb to be purchased or the fuel efficiency of a single vehicle to be rented).

# A Prediction Interval for a Single Future Value

- A point predictor is $\overline{X}$, and the resulting prediction error is $\overline{X} - X_{n+1}$. The expected value of the prediction error is

$$E(\overline{X} - \overline{X}_{n+1}) = E(\overline{X}) - E(X_{n+1}) = \mu - \mu = 0$$

Since $X_{n+1}$ is independent of $X_1, \ldots, X_n$, it is independent of $\overline{X}$, so the variance of the prediction error is $V(\overline{X} - X_{n+1}) = V(\overline{X}) + V(X_{n+1}) = \dfrac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(1 + \dfrac{1}{n}\right)$

# A Prediction Interval for a Single Future Value

- The prediction error is a linear combination of independent, normally distributed rv's, so itself is normally distributed.

- Thus

$$Z = \frac{(\overline{X} - X_{n+1}) - 0}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}} = \frac{\overline{X} - X_{n+1}}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}}$$

has a standard normal distribution.

# A Prediction Interval for a Single Future Value

- It can be shown that replacing $\sigma$ by the sample standard deviation $S$ (of $X_1, \ldots, X_n$) results in

$$T = \frac{\overline{X} - X_{n+1}}{S\sqrt{1 + \frac{1}{n}}} \sim t \text{ distribution with } n - 1 \text{ df}$$

- Manipulating this $T$ variable as $T = (\overline{X} - \mu)/(S/\sqrt{n})$ was manipulated in the development of a CI gives the following result.

# A Prediction Interval for a Single Future Value

- A **prediction interval** (PI) for a single observation to be selected from a normal population distribution is

$$\bar{x} \pm t_{\alpha/2,n-1} \cdot s\sqrt{1 + \frac{1}{n}} \qquad \textbf{(7.16)}$$

- The *prediction level* is 100(1 – $\alpha$)%. A lower prediction bound results from replacing $t_{\alpha/2}$ by $t_{\alpha}$ and discarding the + part of (7.16); a similar modification gives an upper prediction bound.