ENGG 319

Probability & Statistics for Engineers

Section #09
One & Two-Sample Estimation Problems

Dr. Sameh Nassar

L01

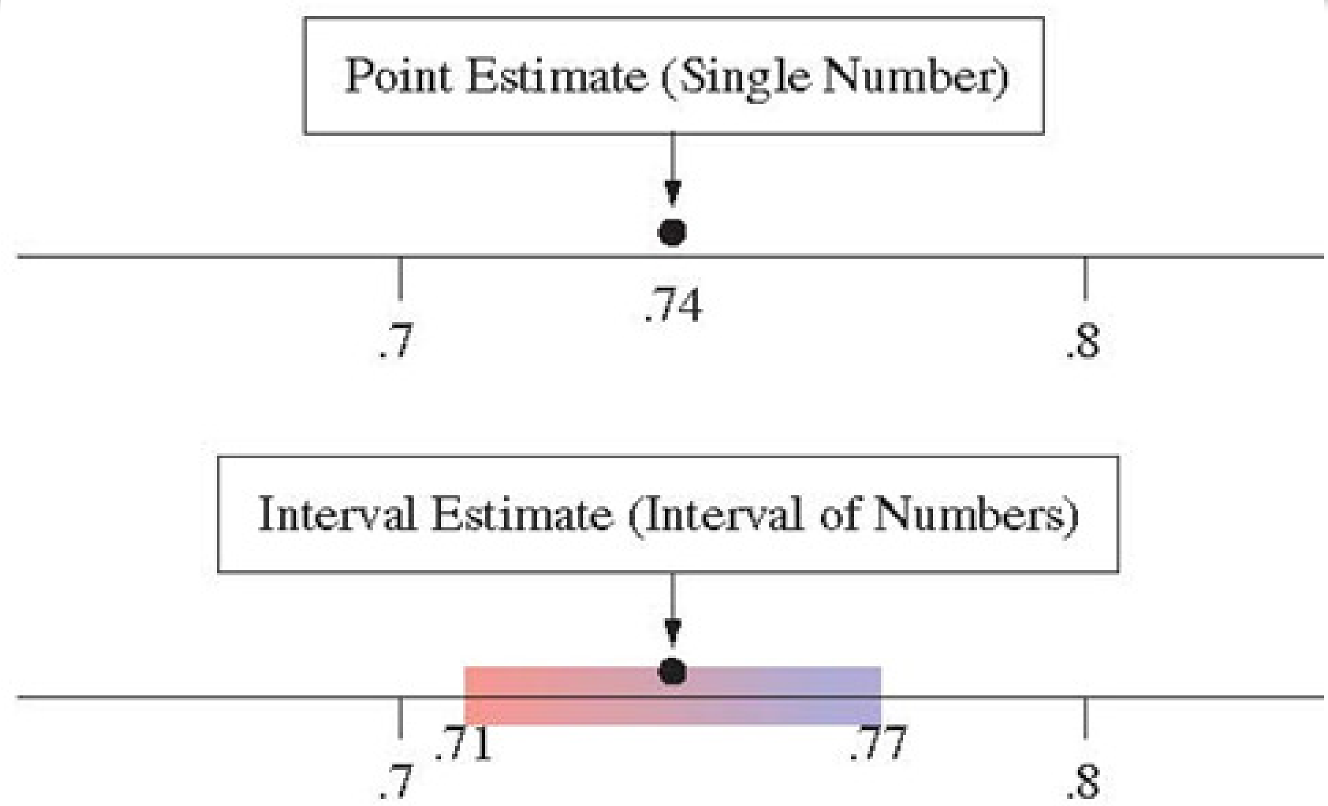
Fall 2016

Some Definitions

- A **point estimate** is a **single number** that is our "best guess" for the parameter:
  - Point estimates are the most common form of inference reported by the mass media.

- An **interval estimate** is an **interval of numbers** within which the parameter value is believed to fall.

Point vs. Interval Estimate

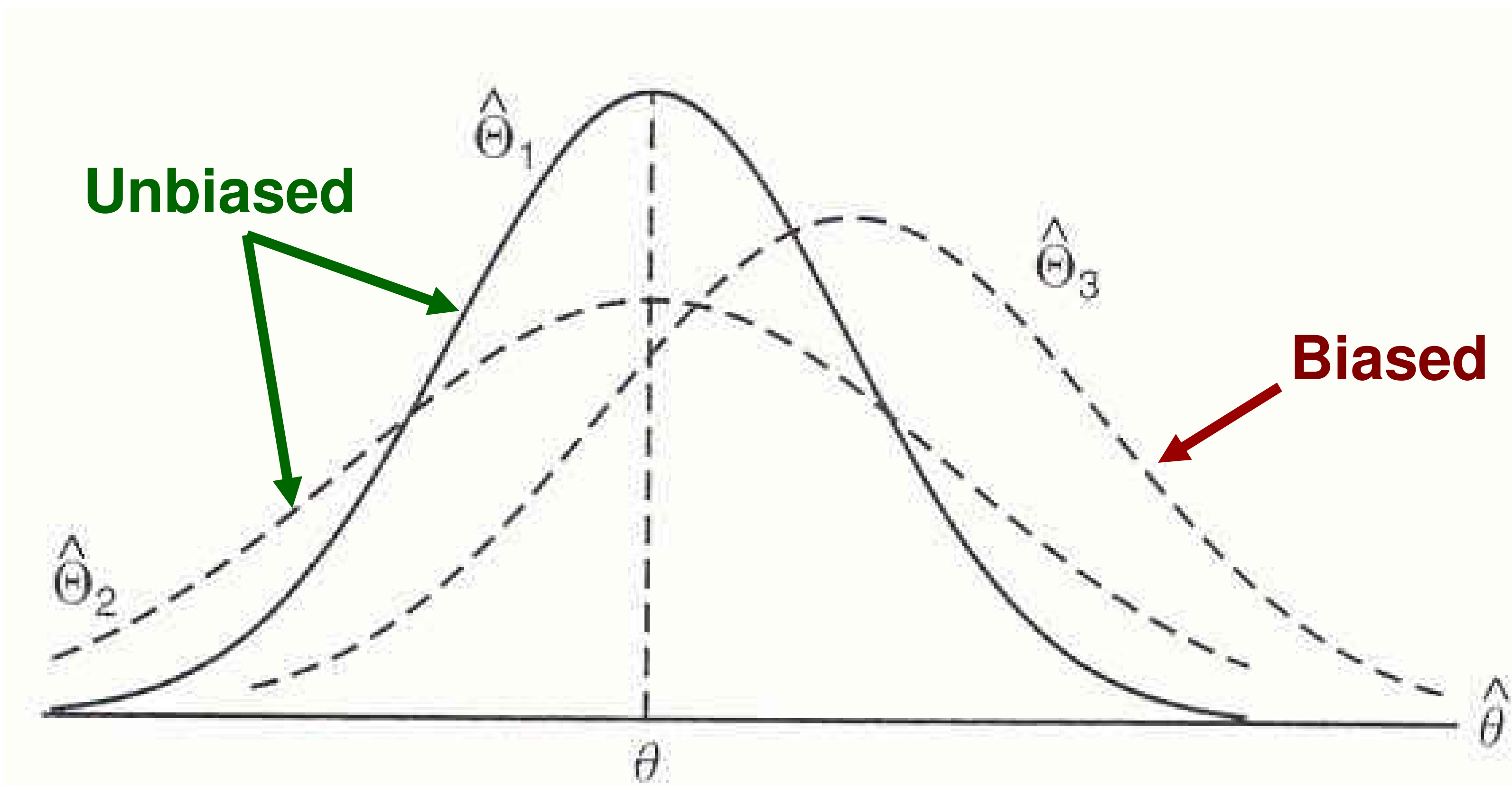


A **point estimate** predicts a parameter by a single number. An **interval estimate** is an interval of numbers that are believable values for the parameter.

From Agresti & Franklin: Statistics © 2006

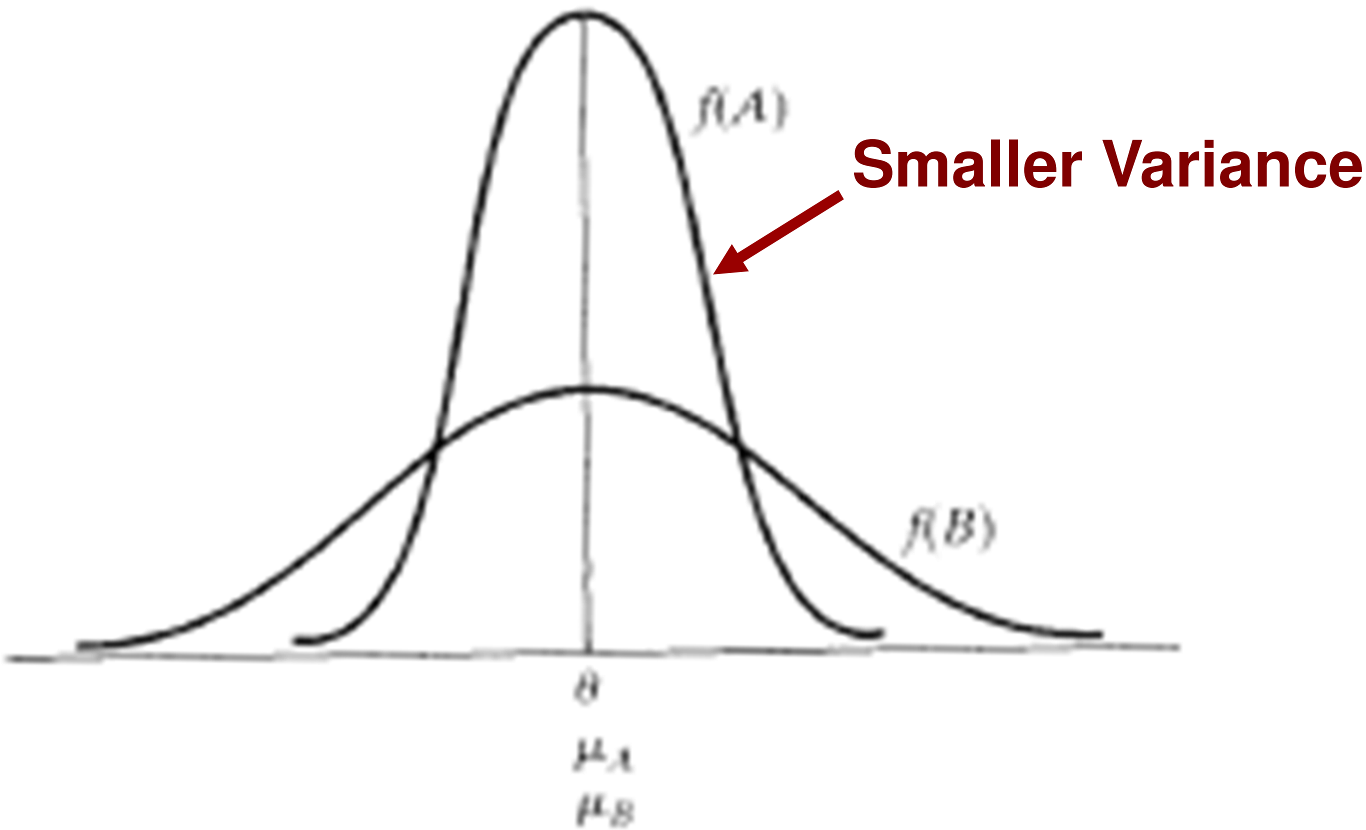
Estimation: Some Properties (1/2)

- A good estimator $\hat{\theta}$ has a sampling distribution that is centered at the parameter θ: the estimator is said to be **unbiased**.



Estimation: Some Properties (2/2)

- A good estimator has a **small variance** compared to other estimators.



Sampling distributions for 2 unbiased estimators of θ with different variances

Estimation: Some Definitions

- A point estimate doesn't tell us how close the estimate is likely to be to the parameter
- An interval estimate is more useful.
  - It incorporates a margin of error which helps us to gauge the accuracy of the point estimate
  - Interval Estimation: Constructing an Interval that Contains the Parameter (We Hope!): the Confidence Interval (**CI**).
  - The probability that this method produces an interval that contains the parameter is called the confidence coefficient.

Single Sample: Estimating the Mean (case 1)

If $\bar{x}$ is the mean of a random sample of size n from a population with **known** variance σ^2, a 100(1-α)% confidence interval for μ is given by:

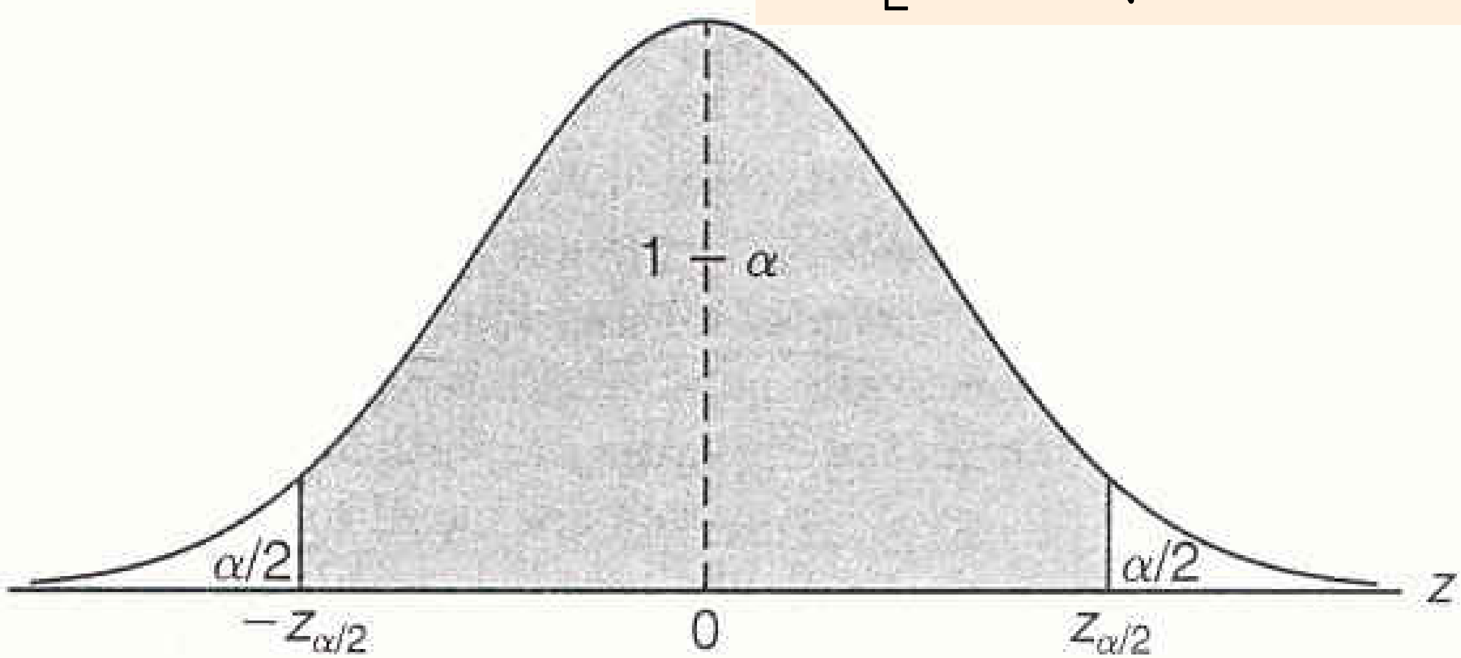
$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where $\mathbf{z_{\alpha/2}}$ is the Z-value leaving an area of $\alpha/2$ to the **right**.

How to Find the $z_{\alpha/2}$ for a $(1-\alpha)\%$ Confidence Interval ?

$$P\left[-z_{\alpha/2} < Z < z_{\alpha/2}\right] = 1-\alpha \quad \& \quad Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$\Rightarrow P\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$



z-scores ($z_{\alpha/2}$) for the most common CIs

Confidence Level	Error Probability	z-Score
0.90	0.10	1.645
0.95	0.05	1.96
0.99	0.01	2.58

Z	0	0.01	0.02	0.03	0.04	0.05
0.00	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.10	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.20	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.30	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.40	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.50	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088
0.60	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422
0.70	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734
0.80	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023
0.90	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289
1.00	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531
1.10	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749
1.20	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944
1.30	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115
1.40	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265
1.50	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394
1.60	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505

$\alpha = 0.10$, $\alpha/2 = 0.05$

$(1-\alpha/2) = 0.95$

Example #1 (part 1)

A pump powered by a 25 kW electric motor is operating at a constant speed. Maintenance staff have randomly sampled the power usage of the electric motor 6 times. The following statistic was then computed:

$$\bar{x} = 25$$

Assume that the power usage follows a normal distribution with a true standard deviation of 0.4 kW, what is the 95% confidence interval for the mean power usage of the 25 kW electric motor?

Example #1 (part 1 "Sol.")

Given: $\bar{x} = 25$ $\sigma = 0.40$ $n = 6$

Required: 95% CI of the mean

$\alpha = 0.05$ ➡ $\alpha/2 = 0.025$ ➡ $(1-\alpha/2) = 0.975$

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$z_{\alpha/2} = 1.96$$

$$25 - 1.96 * \frac{0.4}{\sqrt{6}} < \mu < 25 + 1.96 * \frac{0.4}{\sqrt{6}}$$

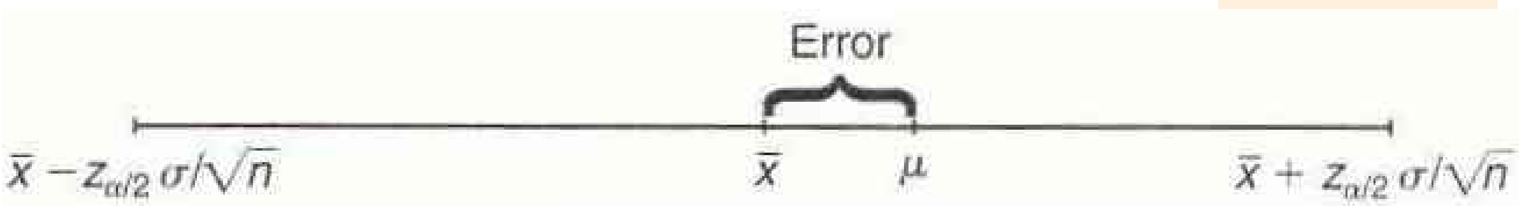
$$25 - 0.32 < \mu < 25 + 0.32$$

$$24.68 < \mu < 25.32$$

Single Sample: Estimating the Mean (Concept of Error)

- If $\bar{x}$ is used to estimate μ, we can be $100(1-\alpha)\%$ confident that the error **e** = $|\bar{x} - \mu|$ will not exceed:

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



- If $\bar{x}$ is used to estimate μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed a specified amount **e** if the sample size is:

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

Example #1 (part 2)

For the data given in Example #1 and from the solution obtained in part 1:

How large a sample is needed if we wish to be 95% confident that our sample mean will be within 0.20 kW of the true mean?

Example #1 (part 2 "Sol.")

Given: $e = 0.20$ $\sigma = 0.40$ $z_{\alpha/2} = 1.96$

Required: **n** for the 95% CI of the mean (satisfying e)

What to use:

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$$

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2 = \left(\frac{1.96 * 0.4}{0.2}\right)^2 = 15.37 \Rightarrow n = 16$$

Single Sample: Estimating the Mean (case 2)

If $\bar{x}$ and s are the mean and standard deviation of a random sample of size n from a **normal** population with **unknown** variance, a $100(1-\alpha)\%$ confidence interval for μ is given by:

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

where $t_{\alpha/2}$ is the t-value with $n-1$ degrees of freedom leaving an area of $\alpha/2$ to the right.

t-scores ($t_{\alpha/2}$) for the most common CIs

From Agresti & Franklin: Statistics © 2006

	Confidence Level					
	80%	90%	95%	98%	99%	99.8%
df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$
1	3.078	6.314	12.706	31.821	63.657	318.3
...						
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785

v	0.4	0.25	0.1	0.05	0.025
1	0.325	1.000	3.078	6.314	12.706
2	0.289	0.816	1.886	2.920	4.303
3	0.277	0.765	1.638	2.353	3.182
4	0.271	0.741	1.533	2.132	2.776
5	0.267	0.727	1.476	2.015	2.571
6	0.265	0.718	1.440	1.943	2.447
7	0.263	0.711	1.415	1.895	2.365

$\alpha = 0.05$

$\alpha/2 = 0.025$

Example #2

A plastic extrusion machine produces sheets with a nominal thickness of 7.5 mm.

A sample of 11 measurements has produces a mean thickness of 7.8 mm with a standard deviation of 0.8 mm.

What is the 99% confidence interval for the mean sheet thickness?

Assume a normal distribution.

Example #2 (Sol.)

Given: $\bar{x} = 7.8$ $S = 0.80$ $n = 11$

Required: 99% CI of the mean

$\alpha = 0.01 \Rightarrow \alpha/2 = 0.005$

$$t_{\alpha/2} = 3.169$$

df = 10

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$7.8 - 3.169 * \frac{0.8}{\sqrt{11}} < \mu < 7.8 + 3.169 * \frac{0.8}{\sqrt{11}}$$

$$7.8 - 0.76 < \mu < 7.8 + 0.76$$

$$7.04 < \mu < 8.56$$

Comment!

Single Sample: Estimating the Mean (case 3)

If $\overline{x}$ and s are the mean and standard deviation of a random sample of size **n ≥ 30** from a **normal** population with **unknown** variance, a 100(1-α)% confidence interval for μ is given by:

$$\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

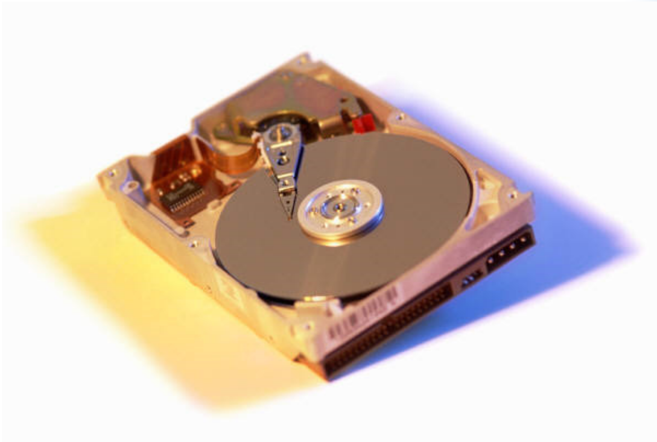
where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example #3

Suppose a PC manufacturer wants to evaluate the performance of its hard disk memory system. One measure of performance is the average time between failures of the disk drive. To estimate this value, a quality control engineer recorded the time between failures from a random sample of 45 disk-drive failures. The following sample statistics were computed:

$\bar{x}$ = 1762 hours and S = 215 hours

Estimate the true mean time between failures with a 90% confidence interval.



Example #3 (Sol.)

Given: $\bar{x} = 1762$ $s = 215$ $n = 45$ **(n>30)**

Required: 90% CI of the mean

$\alpha = 0.10$ ➡ $\alpha/2 = 0.05$ ➡ $(1-\alpha/2) = 0.95$

$$\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$z_{\alpha/2} = 1.645$$

$$1762 - 1.645 * \frac{215}{\sqrt{45}} < \mu < 1762 + 1.645 * \frac{215}{\sqrt{45}}$$

$$1762 - 52.7 < \mu < 1762 + 52.7$$

$$1709.3 < \mu < 1814.7$$

Two Samples: Estimating the Difference of the Means (case 1)

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of size n_1 and n_2 from populations with **known** variances σ_1^2 and σ_2^2, respectively, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by:

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example #4

We want to estimate at 95% CI the difference between the mean starting salaries for recent graduates with civil engineering and oil&gas engineering degrees in Alberta. The following information is available (June 2006 © APEGGA):

1. A random sample of 46 starting salaries for AB oil&gas engineering graduates produced a sample mean of $54,788 with a normal population standard deviation of $5,120.

2. A random sample of 37 starting salaries for AB civil engineering graduates produced a sample mean of $61,442 with a normal population standard deviation of $3,600.

Example #4 (Sol.)

Given:

$\bar{x}_1 = 54788$ $\quad \sigma_1 = 5120$ $\quad n_1 = 46$

$\bar{x}_2 = 61442$ $\quad \sigma_2 = 3600$ $\quad n_2 = 37$

Required: 95% CI of the difference between both means

$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \Rightarrow (1-\alpha/2) = 0.975 \rightarrow z_{\alpha/2} = 1.96$

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$\bar{x}_1 - \bar{x}_2 = 54788 - 61442 = -6654 \qquad \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2} = \sqrt{(5120)^2/46 + (3600)^2/37} = 959.2$$

$$-6654 - 1.96 * 959.2 < \mu_1 - \mu_2 < -6654 + 1.96 * 959.2$$

$$-8534.0 < \mu_1 - \mu_2 < -4773.7$$

Two Samples: Estimating the Difference of the Means (case 2)

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of size n_1 and n_2 from approximately **normal** populations with **unknown but equal** variances, a 100(1-α)% confidence interval for $\mu_1-\mu_2$ is given by:

$$(\bar{x}_1-\bar{x}_2)-t_{\frac{\alpha}{2}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\frac{\alpha}{2}}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

where:

$t_{\alpha/2}$ is the t-score with (n_1+n_2-2) df leaving an area of $\alpha/2$ to the right.

s_p is a pooled estimate of the population standard deviation:

$$s_p^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

Example #5

During the first week of May 2003, 25 of Palm M515 PDA were auctioned off on eBay, 7 of which had the "buy-it-now" option.

- **"Buy-it-now" option:**

  235 225 225 240 250 250 210

- **"Bidding only":**

  250 249 255 200 199 240 228 255 232 246
  210 178 246 240 245 225 246 225

Find the 95% CI for the difference of the means.

Assume both populations are normal with equal variances.

Example #5 (Sol.)

Given: $n_1 = 7$ $n_2 = 18$ $\sigma_1 = \sigma_2 = \sigma$

Computed: $\bar{x}_1 = 233.57$ $\bar{x}_2 = 231.61$ $s_1 = 14.64$ $s_2 = 21.94$ or $s_1^2 = 214.29$ $s_2^2 = 481.19$

Required: 95% CI of the difference between both means

What to Use:

$$(\bar{x}_1 - \bar{x}_2) - t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Example #5 (Sol.)

df = n_1+n_2-2 =7+18-2 =23

$$\bar{x}_1 - \bar{x}_2 = 233.57 - 231.61 = 1.96$$

α = 0.05 ➡ $\alpha/2$ = 0.025 → $t_{\alpha/2} = 2.069$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{6*214.29 + 17*481.19}{23} = 411.56$$

$$t_{\frac{\alpha}{2}} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 2.069 * \sqrt{411.56} * \sqrt{\frac{1}{7} + \frac{1}{18}} = 18.697$$

$$1.96 - 18.697 < \mu_1 - \mu_2 < 1.96 + 18.697$$

$$-16.74 < \mu_1 - \mu_2 < 20.66$$

Two Samples: Estimating the Difference of the Means (case 3)

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of size n_1 and n_2 from approximately **normal** populations with **unknown and unequal** variances, a $100(1-\alpha)\%$ confidence interval for $\mu_1-\mu_2$ is given by:

$$(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}<\mu_1-\mu_2<(\bar{x}_1-\bar{x}_2)+t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

where $t_{\alpha/2}$ is the t-score leaving an area of $\alpha/2$ to the right.

$$df=\upsilon=[s_1^2/n_1+s_2^2/n_2]^2\Bigg/\left[\frac{(s_1^2/n_1)^2}{n_1-1}+\frac{(s_2^2/n_2)^2}{n_2-1}\right]$$

If not Integer, round it **Down**

Example #6

(Ex. 9.43 Textbook): A taxi company is trying to decide whether to purchase brand **A** or brand **B** tires for its fleet of taxis. To estimate the difference in the two brands, an experiment is conducted using twelve of each brand. The tires are run until they wear out. The results are:

Brand A: mean=36,300 km & Stdev = 5,000 km

Brand B: mean=38,100 km & Stdev = 6,100 km

Compute the 95% CI for $\mu_A - \mu_B$ assuming the populations to be approximately normal.

You may not assume that the variances are equal.

Example #6 (Sol.)

Given:

$n_A = 12$ $\quad$ $n_B = 12$

$\bar{x}_A = 36300$ $\quad$ $\bar{x}_B = 38100$

$s_A = 5000$ $\quad$ $s_B = 6100$

$\sigma_A \neq \sigma_B$

Required: 95% CI of the difference between both means

What to Use:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

With:

$$df = \upsilon = [s_A^2/n_A + s_B^2/n_B]^2 \Bigg/ \left[\frac{(s_A^2/n_A)^2}{n_A - 1} + \frac{(s_B^2/n_B)^2}{n_B - 1}\right]$$

Example #6 (Sol.)

$$df = \upsilon = [s_A^2/n_A + s_B^2/n_B]^2 \Bigg/ \left[\frac{(s_A^2/n_A)^2}{n_A - 1} + \frac{(s_B^2/n_B)^2}{n_B - 1}\right]$$

$$df = [(5000)^2/12 + (6100)^2/12]^2 \Bigg/ \left[\frac{((5000)^2/12)^2}{11} + \frac{((6100)^2/12)^2}{11}\right] = 21.18$$

$$\Rightarrow \quad df = 21 = \upsilon$$

$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \rightarrow t_{\alpha/2} = 2.080$

$$(\bar{x}_A - \bar{x}_B) = 36300 - 38100 = -1800$$

Example #6 (Sol.)

$$\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}=\sqrt{\frac{(5000)^2}{12}+\frac{(6100)^2}{12}}=2276.876$$

$$\therefore 95\% \quad \text{CI} \quad \text{for} \quad \mu_A-\mu_B=-1800 \quad \pm \quad 2.080*2276.877$$

$$\text{or} \qquad \mu_A-\mu_B=-1800 \quad \pm \quad 4735.9$$

$$\text{or} \qquad -6535.9<\mu_A-\mu_B<2935.9$$

Two Samples: Estimating the Difference of the Means for Paired Observations (case 4)

- For samples that are **not** independent and the variances of the 2 populations are not necessarily equal, they are called: **paired observations**.
- If $\bar{d}$ and s_d are the mean and standard deviation of the **normally** distributed differences of **n** random pairs of measurements, a 100(1-α)% confidence interval for $\mu_D=\mu_1-\mu_2$ is given by:

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

where $t_{\alpha/2}$ is the t-score with df = n-1 degrees of freedom leaving an area of $\alpha/2$ to the right.

Example #8

The following table provides data on the modulus of elasticity obtained 1 min after loading in a certain configuration. The values were also obtained 4 weeks after loading for the same lumber specimens.

Observation	1 min	4 weeks	difference
1	10490	9110	1380
2	16620	13250	3370
3	17300	14270	2580
4	15480	12740	2740
5	12970	10120	2850

What is the 95% CI for the difference in modulus of elasticity?

Example #8 (Sol.)

Given: $n_d = 5$

Required: 95% CI of the difference in modulus of elasticity

Computed: $\bar{d} = 2584$ $s_d = 735.275$

What to Use:

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

$\alpha = 0.05$ → $\alpha/2 = 0.025$ → $t_{\alpha/2} = 2.776$ (df=4)

$$95\% \quad CI = 2584 \pm 2.776 * 735.275/\sqrt{5} = 2584 \pm 2.776 * 328.825$$

$$\text{or} \quad 1671.181 < \mu_D < 3496.818$$

Single Sample: Estimating a Population Proportion

If $\hat{p}$ is the proportion of successes in a random sample of size n and $\hat{q} = 1 - \hat{p}$, an approximate 100(1-α)% confidence interval for the binomial parameter **p** is given by:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Single Sample: Estimating a Population Proportion

- If $\hat{p}$ is used to estimate **p**, we can be 100(1-α)% confident that the error will not exceed:

$$z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

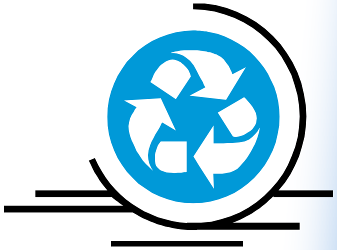
- If $\hat{p}$ is used to estimate **p**, we can be 100(1-α)% confident that the error will not exceed a specified amount **e** if the sample size is:

$$n = \frac{z^2_{\alpha/2}\hat{p}\hat{q}}{e^2}$$

Example #9

In 2000, the Americans were asked: “Are you willing to pay much higher prices in order to protect the environment?” Of 1154 respondents, 518 were willing to do so.

Find and interpret a 95% confidence interval for the population proportion of adult Americans willing to do so at the time of the survey.



Example #9 (Sol.)

Given: $n = 1154$ $n_p = 518$

Required: 95% CI of the yes population proportion

What to Use: $$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

α = 0.05 ⇒ α/2 = 0.025 ⇒ (1-α/2) = 0.975 → $z_{\alpha/2} = 1.96$

$$\hat{p} = n_p / n = 518/1154 = 0.45 \Rightarrow \hat{q} = 1 - \hat{p} = 1 - 0.449 = 0.55$$

$$95\% \quad CI = 0.45 \pm 1.96 * \sqrt{\frac{0.45 * 0.55}{1154}} = 0.45 \pm 1.96 * 0.0146$$

or $$0.42 < p < 0.48$$

Single Sample: Estimating the Variance

If S^2 is the variance of a random sample of size n from a **normal** population, a 100(1-α)% confidence interval for σ^2 is given by:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are the χ^2-values with df = n-1, leaving areas of $\alpha/2$ and (1- $\alpha/2$), respectively, to the right.

Example #10

A quality control supervisor in a cannery knows that the exact amount each can contains will vary, since there are certain uncontrollable factors that affect the amount of fill. The mean fill per can is important, but equally important is the variation, S^2, of the amount of fill. If S^2 is large, some cans will contain too little and others too much. To estimate the variation of fill at the cannery, the supervisor randomly selects 10 cans and weights the content of each.

The weights (in ounces) are:

7.96 7.90 7.98 8.01 7.97 7.96 8.03 8.02 8.04 8.02

Construct a 90% confidence interval for the true standard deviation of the can weights.

Example #10 (Sol.)

Given: $n = 10$

Required: 90% CI for the true σ

What to Use:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

Computed: $S^2 = 0.00185$

$\alpha = 0.10 \Rightarrow \alpha/2 = 0.05 \rightarrow$ $\chi^2_{\alpha/2} = 16.919$ $\chi^2_{1-\alpha/2} = 3.325$ (df=9)

$$\frac{9*0.00185}{16.919} < \sigma^2 < \frac{9*0.00185}{3.325} \Rightarrow 0.00099 < \sigma^2 < 0.005$$

$$0.031 < \sigma < 0.071$$

Two Samples: Estimating the Ratio of 2 Variances

If S_1 and S_2 are the variances of independent random samples of size n_1 and n_2 from **normal** populations, a 100(1-α)% confidence interval for σ_1^2/σ_2^2 is given by:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(df_1, df_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(df_2, df_1)$$

where $f_{\alpha/2}(df_1, df_2)$ is the f-score with $df_1=n_1-1$ and $df_2=n_2-1$ degrees of freedom leaving an area of $\alpha/2$ in the upper tail of the F-distribution, and $f_{\alpha/2}(df_2, df_1)$ is the f-score leaving an area of $\alpha/2$ in the upper tail of the F-distribution.

Example #11

A firm has been experimenting with 2 different physical arrangements of its assembly line. It has been determined that both arrangements yield approximately the same average number of finished units per day. To obtain an arrangement that produces greater process control, you suggest that the arrangement with the smaller variance in the number of finished units produced per day be permanently adopted. 2 independent random samples yield the following data:

Line #	n	Mean	Variance
1	21	491.095	1407.89
2	25	499.36	3729.41

Construct a 95% CI for σ_1^2/σ_2^2. *Based on the result, which of the 2 arrangements would you recommend?*

Example #11 (Sol.)

Given:

$$n_1 = 21 \Rightarrow df_1 = 20 \qquad \bar{x}_1 = 491.095 \qquad s_1^2 = 1407.89$$

$$n_2 = 25 \Rightarrow df_2 = 24 \qquad \bar{x}_2 = 499.36 \qquad s_2^2 = 3729.41$$

Required: 95% CI for σ_1^2/σ_2^2

What to Use:

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(df_1, df_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(df_2, df_1)$$

$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \rightarrow$ $f_{\alpha/2}(20,24) = 2.33$ $\quad f_{\alpha/2}(24,20) = 2.41$

$$\frac{1407.89}{3729.41} * \frac{1}{2.33} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1407.89}{3729.41} * 2.41 \Rightarrow 0.162 < \frac{\sigma_1^2}{\sigma_2^2} < 0.91$$

Textbook Sections

- 9.1
- 9.2
- 9.3
- 9.4
- 9.8
- 9.9
- 9.10
- 9.12
- 9.13