1 9.3 Classical Methods of Estimation

Point Estimation

A **point estimate** of some population parameter θ is a single value $\hat{\theta}$ of a statistic $\widehat{\Theta}$ calculated based on observed values $x_1, \ldots, x_n$ of sample $X_1, \ldots, X_n$.
Examples:

- Sample mean $\bar{x}$ is observed value of statistic $\overline{X}$ and point estimate of population mean μ.
- Sample proportion $\hat{p} = x/n$ is a point estimate of population proportion p.
- Sample variance s^2 is a point estimate of population variance σ^2.

1.1 Unbiasedness of a Point Estimate

Unbiased Estimator
Definition 1. A statistic $\widehat{\Theta}$ is said to be an **unbiased** estimator of the parameter θ if the expected value of $\widehat{\Theta}$ equals θ, that is

$$\mu_{\widehat{\Theta}} = E(\widehat{\Theta}) = \theta.$$

Example 1. Sample mean $\overline{X}$ is an unbiased estimator of population mean μ.

Unbiased Estimator If $\hat{\Theta}$ is an unbiased estimator of θ, then $a\hat{\Theta} + b$ is an unbiased estimator of $a\theta + b$. But $\hat{\Theta}^r$ is a biased estimator of θ^r, $r \neq 1$.

Example 2. Sample variance S^2 is an unbiased estimator of population variance σ^2. However, sample standard deviation S is not an unbiased estimator of population standard deviation σ.
In fact, if sample is drawn from normal population, then

$$E(S) = \frac{\sqrt{2}\,\Gamma(n/2)}{\sqrt{n-1}\,\Gamma[(n-1)/2]}\sigma < \sigma.$$

So

$$\hat{\sigma} = \frac{\sqrt{n-1}\,\Gamma[(n-1)/2]}{\sqrt{2}\,\Gamma(n/2)} S$$

is an unbiased estimator of σ.

1.2 Efficiency

Variance of an Unbiased Estimator
The **most efficient unbiased estimator** of θ is such an unbiased estimator, $\widehat{\Theta}$, of θ that has the smallest variance among all possible unbiased estimators of θ, that is,

$$\mathrm{Var}(\widehat{\Theta}) \leqslant \mathrm{Var}(\widetilde{\Theta}), \quad \text{for any unbiased estimator } \widetilde{\Theta} \text{ of } \theta.$$

Example 1. A manufacturer of car batteries guarantees that his batteries will last, on the average, 3 years with a standard deviation of 1 year. If five of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5, and 4.2 years, is the manufacturer still convinced that his batteries have a standard deviation of 1 year? Assume that the battery lifetime follows a normal distribution.

Solution: $\bar{x} = \frac{1.9+2.4+3.0+3.5+4.2}{5} = 3,$

$$\sum x_i^2 = 1.9^2 + 2.4^2 + 3.0^2 + 3.5^2 + 4.2^2 = 48.26$$

$$s^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \frac{1}{4}[48.26 - (5)(3^2)] = 0.815$$

Since $\sigma = 1$ $\chi^2 = \frac{(n-1)S^2}{\sigma^2} = (5-1)(0.815) = 3.26.$
Using Table A.5,

$$P(\chi^2 \leqslant 3.26) = 1 - P(\chi^2 > 3.26) > 1 - P(\chi^2 > 2.195) = 0.3$$

Using Excel

$$P(\chi^2 \leqslant 3.26) = 0.485$$

So it is likely to have an observed χ^2 as small as 3.26. There is no strong evidence against the hypothesis $\sigma = 1$.

1.3 Interval Estimation

Example 1 Let x equal the weight in grams of 52-gram snack pack of candies. Assume that the distribution of X is normal with $\sigma = 2$ but an unknown μ. If the quality inspector decides to select a random sample of size $n = 10$, then he will get 10 observations $X_1, X_2, \ldots, X_{10}$ and use $\bar{X}$ as a point estimate of μ. Can he get an interval for μ?

$$\bar{X}, \quad Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}, \quad P(-1.65 < Z < 1.65) = 0.90$$

where $\sigma_{\bar{x}} = \sigma/\sqrt{10}$.

$$-1.65 < \frac{\bar{X} - \mu}{\sigma/\sqrt{10}} < 1.65$$

is equivalent to

$$\bar{X} - 1.65\frac{2}{\sqrt{10}} < \mu < \bar{X} + 1.65\frac{2}{\sqrt{10}}$$

So

$$P\Big(\bar{X} - 1.65\frac{2}{\sqrt{10}} < \mu < \bar{X} + 1.65\frac{2}{\sqrt{10}}\Big) = 0.90$$

Denote $E = 1.65\frac{2}{\sqrt{10}} = 1.04$

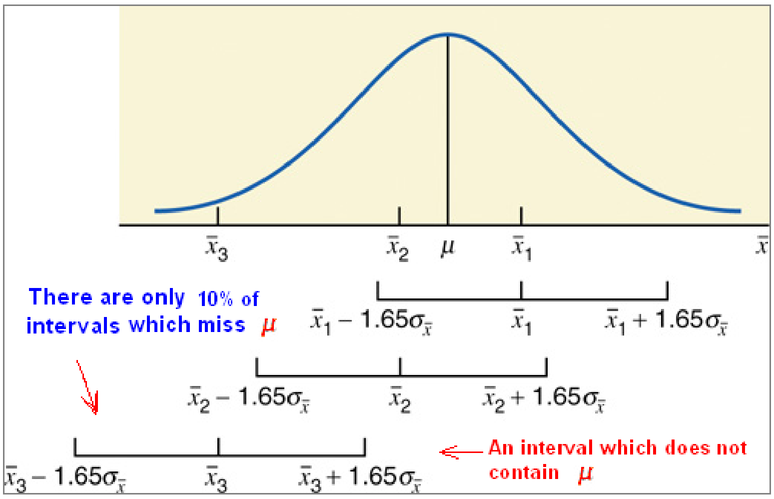
$$\hat{\mu}_L = \bar{X} - E, \quad \text{and} \quad \hat{\mu}_U = \bar{X} + E$$

Then

$$P(\hat{\mu}_L < \mu < \hat{\mu}_U) = 0.90$$

Interval $(\hat{\mu}_L, \hat{\mu}_U)$ is called a confidence interval for μ with 90% confidence level.

What does the confidence level mean?



Confidence Interval for θ

Definition. Let θ be a population parameter and $\hat{\Theta}_L$ and $\hat{\Theta}_U$ be two statistics and $0 < \alpha < 1$. If

$$P(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1 - \alpha$$

then interval $(\hat{\Theta}_L, \hat{\Theta}_U)$ is called a $100(1-\alpha)\%$ confidence interval for θ.

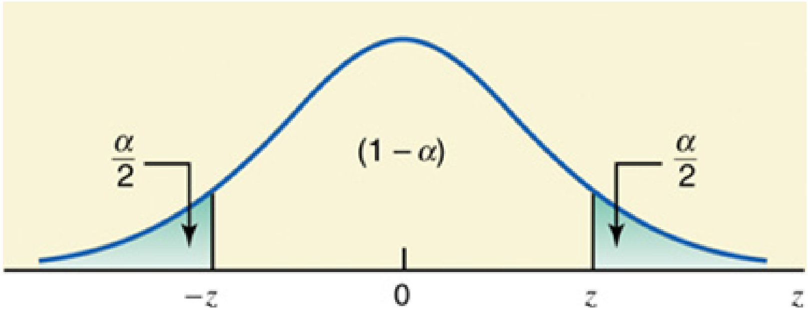
9.4 Single Sample: Estimating the Mean

Confidence Interval for μ with Known σ

Confidence Interval for μ The $(1-\alpha)100\%$ *confidence interval* for μ is

$$\bar{x} \pm z_{\alpha/2}\sigma_{\hat{\mu}}$$

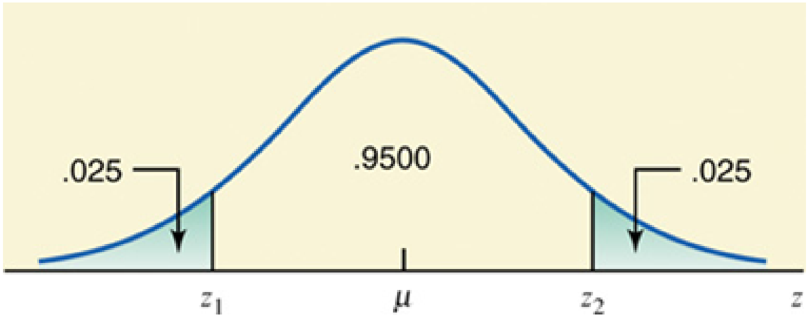
where $E = z_{\alpha/2}\sigma_{\hat{\mu}}$ is the margin of error $\sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}}$ and $z_{\alpha/2}$ satisfies



z value in the confidence interval for μ with $\alpha = 0.05$

$$1-\alpha = 1-0.05 = 0.95, \quad \frac{\alpha}{2} = 0.025$$

$$z_{\alpha/2} = z_{0.025} = 1.96$$

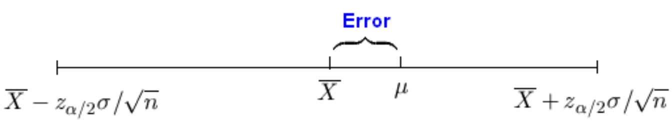


Commonly used z values

Table 8.1 **z Values for Commonly Used Confidence Levels**

Confidence Level	Areas to Look for in Table IV	z Value
90%	.0500 and .9500	1.64 or 1.65
95%	.0250 and .9750	1.96
96%	.0200 and .9800	2.05
97%	.0150 and .9850	2.17
98%	.0100 and .9900	2.33
99%	.0050 and .9950	2.57 or 2.58

Error in Estimating μ by $\overline{X}$



Error in estimating μ by $\overline{X}$

The $100(1-\alpha)\%$ confidence interval

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

also means that when we use $\bar{x}$ to estimate μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed the margin of error $E = z_{\alpha/2}\sigma/\sqrt{n}$

Example 1(cont.) Find the 99% confidence interval for μ based on the data of Example 1:

```
55.95, 56.54, 57.58, 55.13, 57.48
56.06, 59.93, 58.30, 52.57, 58.46
```

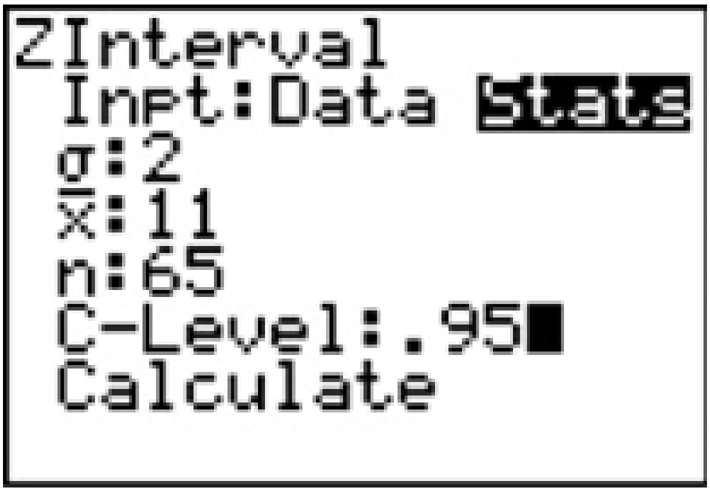
Solution

$$\bar{x} = 56.8, \quad \alpha = 0.01,$$
$$z_{\alpha/2} = z_{0.005} = 2.57$$

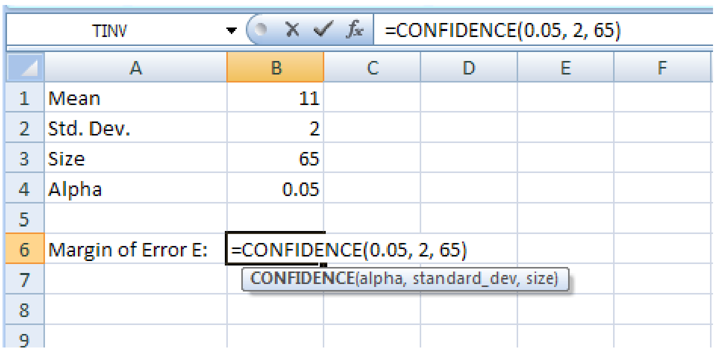
The 99% confidence interval for μ is

$$[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}] = [55.17, 58.43]$$

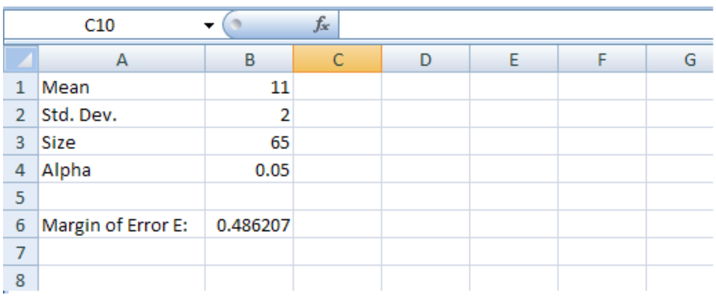
Using TI-8x to Obtain CI: ZInterval



Using Excel to Obtain CI: CONFIDENCE



Using Excel to Obtain CI: CONFIDENCE



1.4 Sample Size

Sample size, confidence level and width of confidence interval The CI $\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ is centered at $\bar{x}$ with width $w_n(\alpha) = 2E = 2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

(a) when α is fixed, as $n \nearrow \infty$, $w_n(\alpha) \searrow 0$.

(b) when n is fixed, there is a **trade-off** between confidence level and width of confidence interval. As $1-\alpha \nearrow 1$ ($\alpha \searrow 0$), then $z_{\alpha/2} \nearrow \infty$, so $w_n(\alpha) \nearrow \infty$.

Given margin of error, determine sample size

Because

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \longrightarrow \quad E^2 = z_{\alpha/2}^2 \frac{\sigma^2}{n}$$

In order to obtain an CI with a given margin of error E, the sample size should be

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

What if σ is unknown?
1. Take preliminary sample of any size(better be 30 or greater). Use sample standard deviation s to estimate σ.
2. Estimate σ based on other study done earlier.
3. $\sigma \approx$range/4.

Example 2 Let x be the excess weight of soap in a 1000-gram bottle. Assume that the distribution of x is normal with μ and standard deviation of $\sigma = 13$. What sample size is required so that we have 95% CI for μ with margin of error 1.5?
Solution $\sigma^2 = 169$. $E = 1.5$. $\alpha = 0.05$, $z_{\alpha/2} = 1.96$.

$$n = \left(\tfrac{z_{\alpha/2}\sigma}{E}\right)^2 = \tfrac{1.96^2(169)}{1.5^2} = 288.5468.$$

So we should choose $n = 289$.

Example 3: Price of Textbook If we know that the maximum price is \$170 and the cheapest textbook is \$10. We want to estimate the mean(average) price μ of the college textbooks. Confidence level is 99% with margin of error E =\$5. What is the sample size?
Solution Since we know that the maximum price is \$170 and the cheapest textbook is \$10, we can estimate σ by

$$\sigma = \frac{\text{range}}{4} = (170-10)/4 = 40.$$

$\alpha = 0.99$, $z_{\alpha/2} = 2.575$, $E = 5$,

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 = (2.575(40)/5)^2 = 424.36 \approx 425.$$

Example 4: A light bulb manufacturer sells a light bulb that has a mean life of 1450 hours with standard deviation of 33.7 hours. A new manufacturing process is being tested and there is interest in knowing the mean life μ of the new bulbs. How large a sample is required so that $\bar{x} \pm 5$ is a 95% CI for μ? You may assume that the standard deviation is still the same.
Solution $\sigma = 33.7$. $E = 5$. $\alpha = 0.05$, $z_{\alpha/2} = 1.96$.

$$n = \left(\tfrac{z_{\alpha/2}\sigma}{E}\right)^2 = \tfrac{1.96^2(33.7)^2}{5^2} = 174.5147.$$

So we should choose $n = 175$.

1.5 One-Sided Confidence Bounds with Known σ

Confidence Bounds (Limits)

- Sometimes, we are only interested in an upper/lower limits for μ.
- A $100(1-\alpha)\%$ **lower confidence bound** for μ with known σ^2 is

$$\overline{X} - z_\alpha \tfrac{\sigma}{\sqrt{n}}$$

- A $100(1-\alpha)\%$ **upper confidence bound** for μ with known σ^2 is

$$\overline{X} + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Example 5: During the Friday night shift, $n = 28$ mints were selected at random from a production line and weighted. They had an average weight of $\bar{x} = 21.45$ grams. Assume that the weights of all the mints have a normal distribution with a unknown mean μ and a known standard deviation $\sigma = 0.30$. Give a 90% lower confidence bound for μ.
Solution $n = 28$, $\bar{x} = 21.45$, $\sigma = 0.30$, $\alpha = 0.1$. A $100(1-\alpha)\% = 90\%$ lower confidence bound for μ is

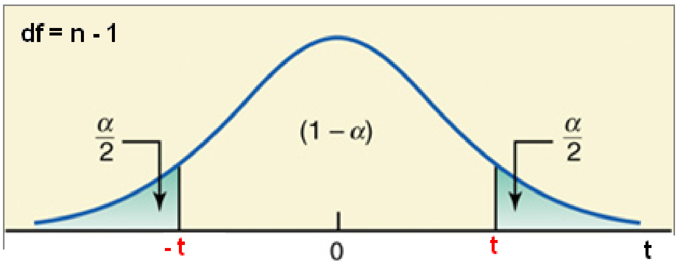
$$\bar{x} - z_\alpha \frac{\sigma}{\sqrt{n}} = 21.45 - 1.28\frac{0.30}{\sqrt{28}} = 21.45 - 0.073 = 21.38.$$

Confidence Interval for μ with Unknown σ

Replace σ and $z_{\alpha/2}$ with s and $t_{\alpha/2}$ respectively
If σ is not know, the $(1-\alpha)100\%$ *confidence interval* for μ is

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

where $E = t_{\alpha/2} \frac{s}{\sqrt{n}}$ is the margin of error and $t_{\alpha/2}$ satisfies



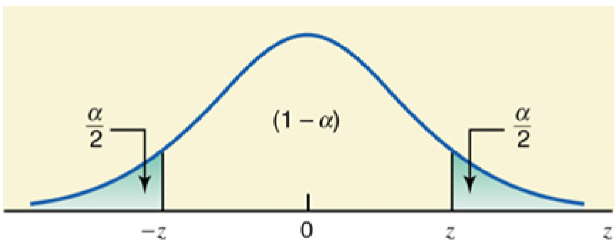
Large Sample Confidence Interval for μ w/ Unknown σ

If $n \geqslant 30$, replace σ with s

The $(1-\alpha)100\%$ *large sample confidence interval* for μ is

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

where $E = z_{\alpha/2} \frac{s}{\sqrt{n}}$ is the margin of error and $z_{\alpha/2}$ satisfies



Example 6: As a clue to the amount of organic waste in Lake Macatawa, a count was made of the number of bacteria colonies in 100 milliliters of water. The number of colonies, in hundreds, for $n = 30$ samples of water from the east basin yield

93	140	8	120	3	120	33	70	91	61
7	100	19	98	110	23	14	94	57	9
66	53	28	76	58	9	73	49	37	92

Find a 90% exact and large sample confidence interval for the mean number μ_E of colonies in 100 milliliters of water in the east basin.

Solution $\bar{x} = 60.37$, $s = 39.62$, Choose $\alpha = 0.1$.
The exact 90% confidence interval is

$$\bar{x} \pm t_{\alpha/2}\tfrac{s}{\sqrt{n}} = 60.37 \pm 1.70\tfrac{39.62}{\sqrt{30}} = 60.37 \pm 12.29 = [48.08, 72.66].$$

A 90% large sample confidence interval for μ_E is

$$\bar{x} \pm z_{\alpha/2}\tfrac{s}{\sqrt{n}} = 60.37 \pm 1.65\tfrac{39.62}{\sqrt{30}} = 60.37 \pm 11.899 = [48.468, 72.265].$$

Homework
Pages 285-288. Sections 9.3-9.4 Exercise 9.4, 9.7, 9.8, 9.10, 9.13.
Study the textbook Examples 9.1-9.5 in Sections 9.3-9.4, pages 271-280.

2 9.5 Standard Error of a Point Estimate

2.1 Standard Error of $\bar{X}$

When σ is Known. If σ is known, the **standard error of point estimator $\bar{X}$** is the standard deviation of the sampling distribution of $\overline{X}$:

$$\text{s.e.}(\overline{X}) = \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

In this case, the $100(1-\alpha)\%$ **confidence interval (limits) of μ is written as:**

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = z_{\alpha/2}\text{s.e.}(\bar{x})$$

When σ is Not Known. If σ is not known, the **standard error of point estimator $\bar{X}$** is the standard deviation of the sampling distribution of $\overline{X}$:

$$\text{s.e.}(\overline{X}) = \sigma_{\overline{X}} = \frac{s}{\sqrt{n}}$$

In this case, the $100(1-\alpha)\%$ **confidence interval (limits) of μ is written as:**

$$\bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = t_{\alpha/2}(n-1)\text{s.e.}(\bar{x})$$

3 9.6 Prediction Intervals

Example 1: Predict a Future Value Due to the decrease in interest rates, the First Citizen Bank received a lot of mortgage applications. A recent sample of 50 mortgage loans resulted in an average of \$257, 300. Assume a population standard deviation of \$25,000, If the next customer called in for mortgage loan application, what would be the next customer's loan amount?

Solution Denote sample mean of of 50 mortgage loans by $\bar{X}$ and the loan amount of the next customer by X_0, then X_0 and $\bar{X}$ are independent normal random variables.

$$X_0 \sim N(\mu, \sigma^2), \quad \overline{X} \sim N\Big(\mu, \frac{\sigma^2}{n}\Big)$$

So

$$X_0 - \overline{X} \sim N\Big(0, \sigma^2 + \frac{\sigma^2}{n}\Big)$$

$$Z = \frac{X_0 - \overline{X}}{\sqrt{\sigma^2 + \frac{\sigma^2}{n}}} \sim N(0,1)$$

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1-\alpha$$

$$P\left(-z_{\alpha/2} < \frac{X_0 - \overline{X}}{\sigma\sqrt{1+1/n}} < z_{\alpha/2}\right) = 1-\alpha$$

Solving inequalities for X_0, we have

$$P\left(\overline{X} - z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}} < X_0 < \overline{X} + z_{\alpha/2}\sigma\sqrt{1+\frac{1}{n}}\right) = 1-\alpha$$

$\bar{x} = \$257,300$, $n = 50$, and $\sigma = \$25,000$. Choose $\alpha = 0.05$, so that $1-\alpha = 0.95$ and $z_{\frac{\alpha}{2}} = 1.96$.

$$\bar{x} \pm z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} = (\$207,812, \$306,787)$$

$$P(\$207,812 < X_0 < \$306,787) = 95\%$$

Prediction Intervals

- If σ is known, given the observed mean $\bar{x}$ of a sample of size n, the $100(1-\alpha)\%$ **prediction interval** for the next value x_0 is

$$\bar{x} - z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + z_{\frac{\alpha}{2}}\sigma\sqrt{1+\frac{1}{n}}$$

- If σ is unknown, given the observed mean $\bar{x}$ of a sample of size n, the $100(1-\alpha)\%$ **prediction interval** for the next value x_0 is

$$\bar{x} - t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}} < x_0 < \bar{x} + t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}}$$

Example with Unknown σ

Example 2. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed. If a sample of 30 bulbs has an average life of 780 hours and a standard deviation of 41 hours, find a 99% prediction interval for the next observed life of a bulb manufactured by this firm.
Solution $\bar{x} = 780$, $n = 30$, and $s = 41$. $\alpha = 0.01$, $1-\alpha = 0.99$ and $t_{\frac{\alpha}{2}}(n-1) = 2.756$.

$$\bar{x} \pm t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}} = (665.12, 894.88)$$

So the $100(1-\alpha)\%$ **prediction interval** for the next observed life is

$$(665.12, 894.88)$$

3.1 Use Prediction Interval to Detect Outlier

Use Prediction Interval to Detect Outlier Given a observed mean $\bar{x}$ of a sample of size n from a normal population, the next observed value x_0 is an **outlier**, if x_0 falls outside the $100(1-\alpha)\%$ **prediction interval**

$$\left(\bar{x} - t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}}, \bar{x} + t_{\frac{\alpha}{2}}(n-1)s\sqrt{1+\frac{1}{n}}\right)$$

Note: This method is a parametric or large sample method because we assume that either the population is normal or the sample size is large. The outlier detection method using Box-and-Whisker plot is a nonparametric method.

4 9.7 Tolerance Interval(Limits)

Example 1. A manufacturer is producing metal pieces that are cylindrical in shape. A sample of these pieces is taken and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. Find an interval so that we are 99% confidence that this interval contains at least 95% of the metal pieces produced by this machine, assuming an approximate normal distribution.
Solution: If both μ and σ are known, then we are 100% confident that 95% of the metal pieces produced by this machine are contained in

$$(\mu - 1.96\sigma, \quad \mu + 1.96\sigma)$$

This is a tolerance interval.

Now, both μ and σ were unknown, we have to estimate them by $\bar{x}$ and s:

$$\bar{x} = \frac{1.01 + 0.97 + 1.03 + 1.04 + 0.99 + 0.98 + 0.99 + 1.01 + 1.03}{9} = \frac{9.05}{9} = 1.005556,$$

$$\begin{aligned}\sum x_i^2 &= 1.01^2 + 0.97^2 + 1.03^2 + 1.04^2 + 0.99^2 + 0.98^2 + 0.99^2 + 1.01^2 + 1.03^2 \\ &= 9.1051 \\ s^2 &= \tfrac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right) = \tfrac{1}{8}[9.1051 - (9)(1.005556^2)] = 0.0006017722\end{aligned}$$

$$s = \sqrt{0.0006017722} = 0.02453105$$

If μ and σ unknown, then the tolerance interval (limits) is

$$\bar{x} \pm ks,$$

where k is determined (See Table A.7) so that one can be $100(1-\gamma)\%$ confident that the given interval contains at least $100(1-\alpha)\%$ of all the measurements.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by Table A.7, $k = 4.55$.
So the 99% tolerance limits are

$$1.005556 \pm (4.55)(0.0246) = (0.978, 1.033)$$

A tolerance interval is an estimate of the "middle" 95% of the normal population.
Approximation of Tolerance Factor The tolerance factor k can also be approximated by

$$k = \sqrt{\frac{(n-1)\left(1+\frac{1}{n}\right) z_{\alpha/2}^2}{\chi_\gamma^2(n-1)}}$$

If $n \to \infty$, then $k \to z_{\alpha/2}$.
Since $1-\alpha = 0.95$, $\gamma = 0.01$, and $n = 9$, by the above formula we have, $k = 4.554$.
So the 99% tolerance limits are

$$1.005556 \pm (4.554)(0.0246) = (0.978, 1.033)$$

Homework Pages 285-288. Sections 9.3-9.4 Exercise 9.4, 9.7, 9.8, 9.10, 9.13.
Study the textbook Examples 9.1-9.5 in Sections 9.3-9.4, pages 271-280.

5 9.8 Estimating the Differences of Two Means

5.1 Known Variances

An Example Example 1

- Comparing gas mileage between two types of engines, A and B.
- Fifty experiments were conducted using engine type A: average gas mileage is 36 mpg. Population standard deviation is $\sigma_A = 6$.
- Seventy five experiments were conducted using engine type B: average gas mileage is 42 mpg. Population standard deviation is $\sigma_B = 8$.

- Find a 95% confidence interval for the difference of the mean gas mileage, $\mu_A - \mu_B$.

Solution of Example 1

- Point estimate of $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B = 36 - 42 = -6$.
- 95% confidence interval for $\mu_A - \mu_B$: $\bar{X}_A - \bar{X}_B \pm E$.
- $E = ?$

$$95\% = P(\bar{X}_A - \bar{X}_B - E < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + E)$$

Equivalently

$$95\% = P(-E < (\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B) < E)$$

- By the CLT,

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} \quad \text{is standard normal.}$$

Solution of Example 1(cont.)

- Because $P(-1.96 < Z < 1.96) = 0.95$

$$\frac{E}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} = 1.96$$

$$\rightarrow \quad E = 1.96\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}$$
$$= 1.96\sqrt{6^2/50 + 8^2/70} = 2.506$$

- a 95% confidence interval for $\mu_A - \mu_B$ is

$$[-6 - 2.506, -6 + 2.506] = [-8.506, -3.494]$$

Two Normal Independent Samples

- Let $X_1, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples of sizes m and n from $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$ respectively.
- Since sample means $\overline{X}$ and $\overline{Y}$ are independent and

$$\overline{X} \sim N\big(\mu_X, \tfrac{\sigma_X^2}{m}\big), \quad \overline{Y} \sim N\big(\mu_Y, \tfrac{\sigma_Y^2}{n}\big).$$

- So

$$W = \overline{X} - \overline{Y} \sim N\big(\mu_X - \mu_Y, \tfrac{\sigma_X^2}{m} + \tfrac{\sigma_Y^2}{n}\big)$$

- $$P\Big(-z_{\alpha/2} \leqslant \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}} \leqslant z_{\alpha/2}\Big) = 1 - \alpha.$$

Confidence Interval for $\mu_x - \mu_y$: σ_X^2 and σ_Y^2 known

$$P\Big[(\overline{X} - \overline{Y}) - z_{\alpha/2}\sigma_W \leqslant \mu_X - \mu_Y \leqslant (\overline{X} - \overline{Y}) + z_{\alpha/2}\sigma_W\Big] = 1 - \alpha,$$

where

$$\sigma_W = \sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}.$$

So we have a $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$:

$$\big[(\overline{X} - \overline{Y}) - z_{\alpha/2}\sigma_W, \quad (\overline{X} - \overline{Y}) + z_{\alpha/2}\sigma_W\big]$$

Example 2. If $m = 15$, $n = 8$, $\bar{x} = 70.1$, $\bar{y} = 75.3$, $\sigma_X = 60$, $\sigma_Y = 40$, and $1 - \alpha = 0.95$, then from Table A-3, $z_{\alpha/2} = z_{0.025} = 1.96$, $\bar{x} - \bar{y} = -5.2$ and

$$1.96\sigma_W = 1.96\sqrt{\frac{60^2}{15} + \frac{40^2}{8}} = 5.88.$$

Thus a 95% confidence interval for $\mu_X - \mu_Y$ is

$$[-5.2 - 5.88, -5.2 + 5.88] = [-11.08, 0.68].$$

5.2 Equal But Unknown Variances

Confidence Interval for $\mu_x - \mu_y$: $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ unknown A $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$:

$$(\overline{X} - \overline{Y}) \pm E$$

$$E = t_{\frac{\alpha}{2}}(m+n-2)S_p$$

$$S_p = \sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)}$$

Example 3: The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)					
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Assume the samples are from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ with equal variances.
Give a 95% confidence interval for the difference $\mu_1 - \mu_2$ of the mean running times of the two companies.

Solution of Example 3:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- The pooled standard error

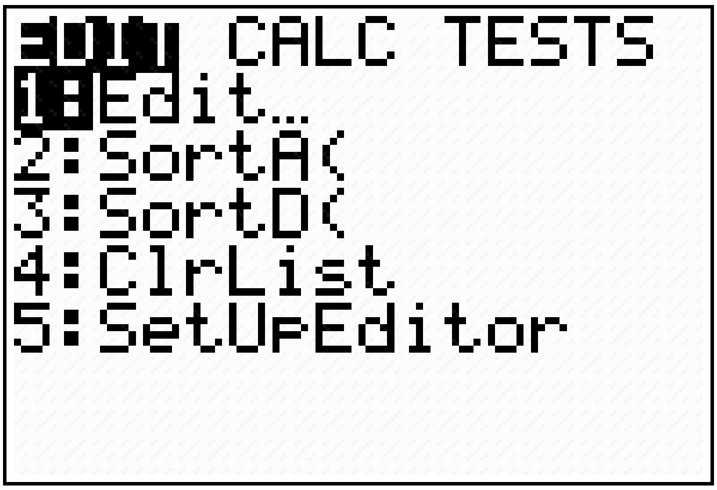
$$S_p = \sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)}$$

$$= \sqrt{\frac{(5-1)78.8 + (7-1)913.3}{5+7-2}\left(\frac{1}{5} + \frac{1}{7}\right)} = 14.10$$

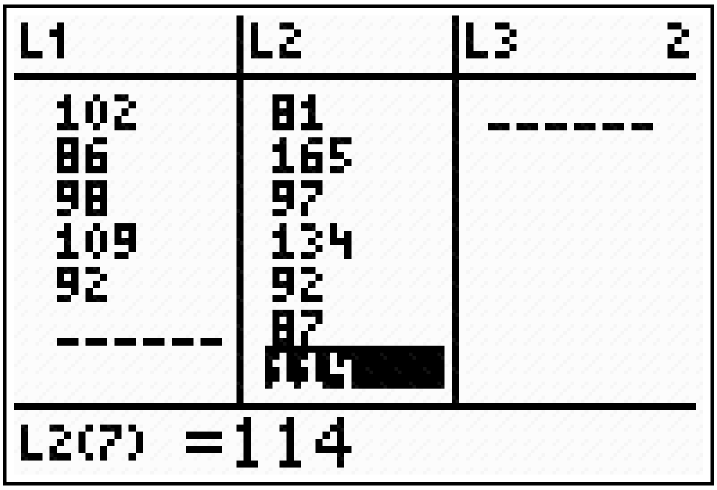
- $E = t_{\frac{\alpha}{2}}(m+n-2)S_p = t_{0.025}(10)S_p = 2.228(14.10) = 31.41$
- The 95% CI:

$$(\overline{X} - \overline{Y}) \pm E = (-44.01, 18.81)$$

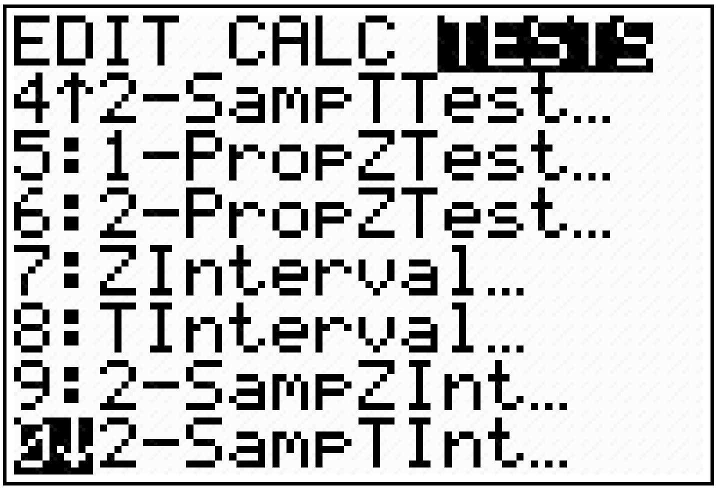
Solution of Example 3 using TI8x



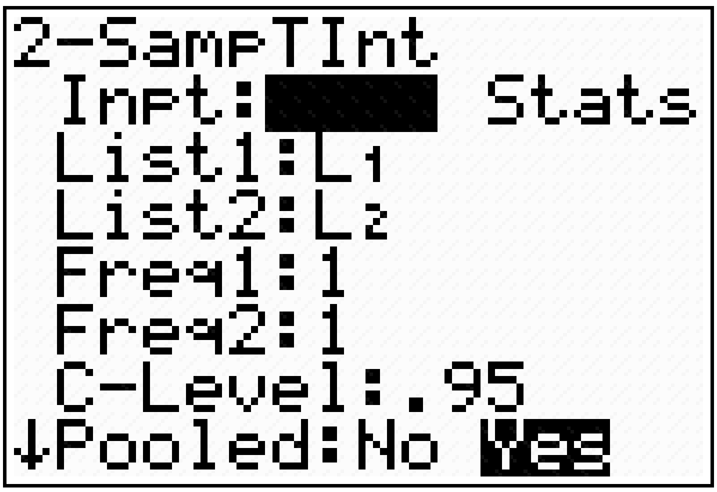
Solution of Example 3 using TI8x



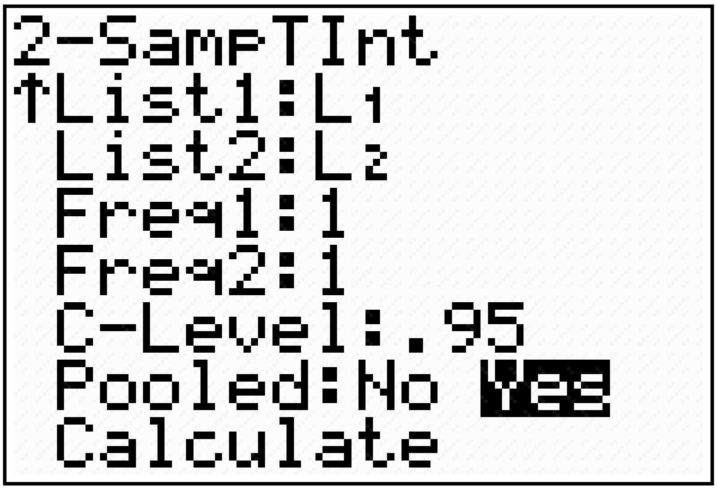
Solution of Example 3 using TI8x



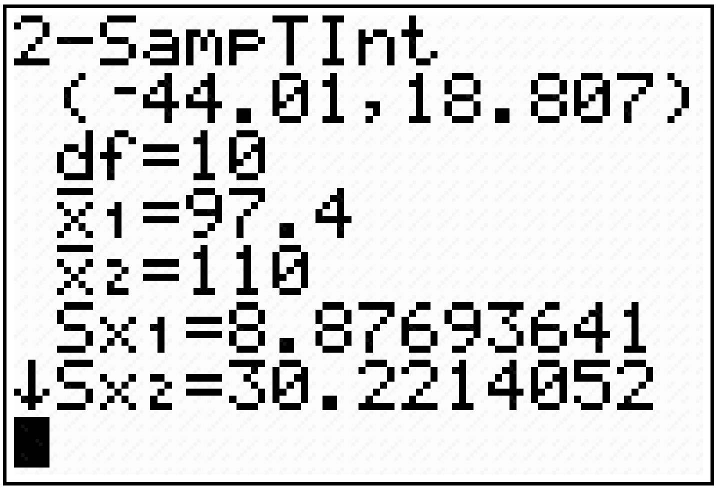
Solution of Example 3 using TI8x



Solution of Example 3 using TI8x



Solution of Example 3 using TI8x



Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances		
m	5	
n	7	
alpha	0.05	
Xbar1	=AVERAGE(C5:G5)	
Xbar2	110	
Sp	14.09584	
CI:	-44.0075	18.80748

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances		
m	5	
n	7	
alpha	0.05	
Xbar1	97.4	
Xbar2	=AVERAGE(C6:I6)	
Sp	14.09584	
CI:	-44.0075	18.80748

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances		
m	5	
n	7	
alpha	0.05	
Xbar1	97.4	
Xbar2	110	
Sp	=SQRT((((C9-1)*VAR(C5:G5)+(C10-1)*VAR(C6:I6))/(C9+C10-2))*(1/C9+1/C10))	
CI:	-44.0075	18.80748

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances	
m	5
n	7
alpha	0.05
Xbar1	97.4
Xbar2	110
Sp	14.09584
CI:	=C12-C13-TINV(C11, C9+C10-2)*C14

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances		
m	5	
n	7	
alpha	0.05	
Xbar1	97.4	
Xbar2	110	
Sp	14.09584	
CI:	-44.0075	=C12-C13+TINV(C11, C9+C10-2)*C14

Solution of Example 3 using Excel

Company	Times(minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Confidence Interval: Unknown equal Variances		
m	5	
n	7	
alpha	0.05	
Xbar1	97.4	
Xbar2	110	
Sp	14.09584	
CI:	-44.0075	18.80748

5.3 Unknown and Unequal Variances

Unknown and Unequal Variances

(a) If $m, n \geqslant 30$, then by the C.L.T. **approximately**

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_X - \mu_Y)}{\sqrt{S_X^2/m + S_Y^2/n}} \sim N(0,1),$$

(b) If $m < 30$ or $n < 30$, then **approximately**,

$$W = \frac{\overline{X}-\overline{Y}-(\mu_X-\mu_Y)}{\sqrt{S_X^2/m+S_Y^2/n}} \sim t(\lfloor r \rfloor),$$

where $\lfloor r \rfloor$, "floor" of r, is the integer part of r and

$$\frac{1}{r} = \frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}, \quad c = \frac{S_X^2/m}{S_X^2/m+S_Y^2/n}.$$

$$r = \frac{\left(S_X^2/m+S_Y^2/n\right)^2}{\frac{1}{m-1}\left(\frac{S_X^2}{m}\right)^2+\frac{1}{n-1}\left(\frac{S_Y^2}{n}\right)^2}.$$

C.I. for $\mu_x - \mu_y$: both σ_X^2 and σ_Y^2 unknown

(a) If $m, n \geqslant 30$, then an approximate $100(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is

$$(\overline{X}-\overline{Y}) \pm z_{\alpha/2}\sqrt{\frac{S_X^2}{m}+\frac{S_Y^2}{n}}$$

(b) If $m < 30$ or $n < 30$, then a $1-\alpha$ confidence interval for $\mu_X - \mu_Y$ is

$$(\overline{X}-\overline{Y}) \pm E$$

$$E = t_{\alpha/2}(\lfloor r \rfloor)\sqrt{\frac{S_X^2}{m}+\frac{S_Y^2}{n}}$$

Example 4: The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)						
1	102	86	98	109	92		
2	81	165	97	134	92	87	114

Assume the samples are from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ with unequal variances.
Give a 95% confidence interval for the difference $\mu_1 - \mu_2$ of the mean running times of the two companies.

Solution of Example 4:

- $m = 5$, $n = 7$, $\alpha = 0.05$;
- $\bar{X} = 97.4$, $\bar{Y} = 110$, $S_X^2 = 78.8$, $S_Y^2 = 913.33$,
- $c = \frac{S_X^2/m}{S_X^2/m+S_Y^2/n} = 0.11$,
- $r = 1/\left[\frac{c^2}{m-1} + \frac{(1-c)^2}{n-1}\right] = 6.26$
- $$E = t_{.025}(6)\sqrt{\frac{78.8}{5}+\frac{913.3}{7}} = 34.49$$
- The 95% CI:
$$(\overline{X}-\overline{Y}) \pm E = (-47.09, 21.89)$$

6 9.9 Paired Observations

6.1 Based on Paired Data

Example 5. (Reaction Time to Red and Green) Eight subjects were involved in an experiment was conducted to compare people's reaction times to a red light versus a green light. The reaction times in seconds were recorded as follows.

Subject	Red(X)	Green(Y)	$D = X - Y$
1	0.30	0.43	−0.13
2	0.23	0.32	−0.09
3	0.41	0.58	−0.17
4	0.53	0.46	0.07
5	0.24	0.27	−0.03
6	0.36	0.41	−0.05
7	0.38	0.38	0.00
8	0.51	0.61	−0.10

Find a 90% confidence interval for $\mu_X - \mu_Y$.

The CI for $\mu_X - \mu_Y$ based Paired Data

- Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n independent pairs of (dependent) r.v.'s.
- Assume that $D_i = X_i - Y_i$, $i = 1, 2, \ldots, n$, is a random sample from $N(\mu_D, \sigma_D^2)$, $\mu_D = \mu_X - \mu_Y$.
- A $1-\alpha$ confidence interval for $\mu_X - \mu_Y$:

$$\overline{D} \pm t_{\alpha/2}(n-1)\frac{S_D}{\sqrt{n}},$$

where

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i, \quad S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \overline{D})^2.$$

Solution of Example 5:

- Statistics: $n = 8$, $\overline{D} = -0.0625$,$S_D = 0.0765$

$$t_{\alpha/2}(n-1) = t_{.05}(7) = 1.895$$

- $E = 0.0512$
- The 90% confidence interval for $\mu_X - \mu_Y$:

$$-0.0625 \pm 0.0512 = [-0.1137, -0.0113].$$

- The CI does not include 0. We conclude that people react a red light faster.

7 9.10 Estimating a Proportion

7.1 Estimating Single Proportion

Do you know how to read the polls' results?

- If the the opinion poll says candidate A leads B 3%: 45% vs 42%, what does it mean?
- The poll gives sample proportions not the population proportions.
- You need to find the margin of error.

How does the poll work?

- The poll collects the opinions of n voters.
- Y of the n voters will vote for A.
- The sample proportion is $\hat{p} = Y/n$ which is a point estimate of the population proportion p.
- Y has a binomial distribution $b(n, p)$.
- How to construct a confidence interval for p?

Normal Approximation of Sampling Distribution of $\hat{p}$:

- By the C.L.T., if n is large, then

$$P\left[-z_{\alpha/2} \leqslant \frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \leqslant z_{\alpha/2}\right] \approx 1-\alpha.$$

- Since $p \approx \hat{p}$,

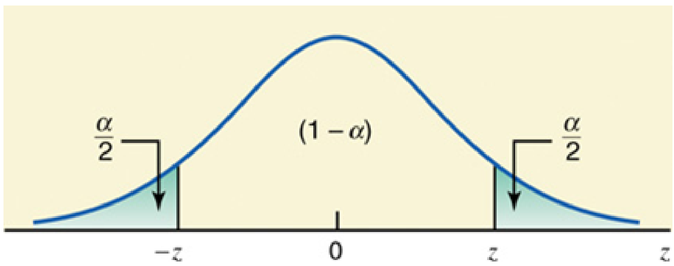
$$P\left[-z_{\alpha/2} \leqslant \frac{\overline{X}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \leqslant z_{\alpha/2}\right] \approx 1-\alpha.$$

Confidence Interval for Population Proportion p:

- If sample size n is large, i.e. $n\hat{p} > 5$ and $n\hat{q} = n(1-\hat{p}) > 5$,
- the $100(1-\alpha)\%$ confidence interval for the population proportion p is

$$\overline{X} \pm E, \quad E = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}.$$

- E is called the margin of error, and $z_{\alpha/2}$ satisfies

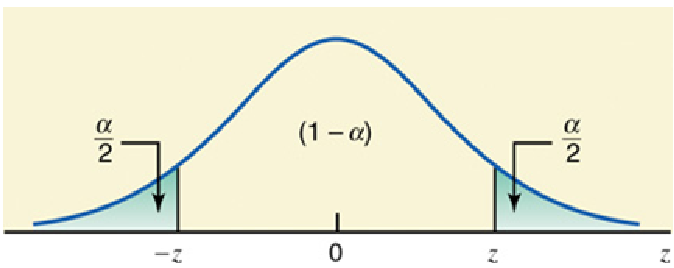


Confidence Interval for Population Proportion p:

- If sample size n is large, i.e. $n\hat{p} > 5$ and $n\hat{q} = n(1-\hat{p}) > 5$,
- the $100(1-\alpha)\%$ confidence interval for the population proportion p is

$$\overline{X} \pm E, \quad E = z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}.$$

- E is called the margin of error, and $z_{\alpha/2}$ satisfies



A Little Better Confidence Interval for p:

- Since $-z_{\alpha/2} \leqslant \frac{\overline{X}-p}{\sqrt{p(1-p)/n}} \leqslant z_{\alpha/2}$ is equivalent to

$$\frac{|\overline{X}-p|}{\sqrt{p(1-p)/n}} \leqslant z_{\alpha/2} \quad \text{or} \quad \frac{(\overline{X}-p)^2}{p(1-p)/n} \leqslant z_{\alpha/2}^2$$

- That is $H(p) = (\overline{X}-p)^2 - z_{\alpha/2}^2 p(1-p)/n \leqslant 0$
- Let $a = 1 + \frac{z_{\alpha/2}^2}{n}$, $b = -2\overline{X} - \frac{z_{\alpha/2}^2}{n}$, $c = \overline{X}^2$. $H(p) = ap^2 + bp + c$.
- The solutions of $H(p) = 0$ are

$$\hat{p}_L = \frac{-b-\sqrt{b^2-4ac}}{2a}, \quad \hat{p}_U = \frac{-b+\sqrt{b^2-4ac}}{2a}.$$

So an approximate $100(1-\alpha)\%$ CI for p is $[\hat{p}_L, \quad \hat{p}_U]$.

Example 1: Poll Results in The News **Example 1: USA Today/Gallup Poll:** http://www.usatoday.com/

(a) For a poll result $\hat{p}$ of interest, find the sample size n, the margin of error, and calculate the 95% confidence interval for p.
(b) Find the margin of error for 99% confidence level and confidence interval.

Example 2: A machine shop manufactures toggle levers. A lever is flawed if a standard nut cannot be screwed onto the threads. Let p equal the proportion of flawed toggle levers that the shop manufactures. If there were 24 flawed levers out of a sample of 642 that were selected randomly from the production line.
(a) Give a point estimate of p.
(b) Find a 95% confidence interval for p.
Solution of Example 2 Solution
(a) The point estimate is $\hat{p} = 24/642 = 0.037$.
(b) $\alpha = 0.05$, $z = 1.96$, $s_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n} = 0.0075$.
$E = 0.0147$ and 95% CI for p is $[0.0223, 0.0517]$.
If the machine shop wants to be 99% confident with a margin of error as small as 0.001, how large a sample should be selected?
What is the most conservative sample size?

Sample Size for Estimating p

Determine Sample Size by Margin of Error

- If $\hat{p}$ is available,

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

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

$$n = \left(\frac{z_{\frac{\alpha}{2}}}{E}\right)^2\hat{p}(1-\hat{p}) = \left(\frac{z_{\frac{\alpha}{2}}\sqrt{\hat{p}(1-\hat{p})}}{E}\right)^2$$

- If $\hat{p}$ is not available, use the *most conservative* estimate

$$n = \left(\frac{z_{\frac{\alpha}{2}}}{2E}\right)^2$$

Example 3: A machine shop manufactures toggle levers. A lever is flawed if a standard nut cannot be screwed onto the threads. Let p equal the proportion of flawed toggle levers that the shop manufactures. If there were 24 flawed levers out of a sample of 642 that were selected randomly from the production line.
In §6.7 Example 2, have found a point estimate of p: $\hat{p} = 0.037$ and the margin of error $E = 0.0147$ with 95% confidence.
(a) If the machine shop wants to be 99% confident with a margin of error as small as 0.001, how large a sample should be selected?
(b) What is the most conservative sample size?
Solution of Example 3 Solution
(a) $\alpha = 0.01$, $z_{\frac{\alpha}{2}} = 2.575$

$$n = \left(\frac{z}{E}\right)^2\hat{p}(1-\hat{p}) = \left(\frac{2.575}{0.001}\right)^2(0.037)(1-0.037) = 236,256$$

(b) The most conservative sample size is

$$n = \left(\frac{z_{\frac{\alpha}{2}}}{2E}\right)^2 = \left(\frac{2.575}{2(0.001)}\right)^2 = 1,657,657$$

8 9.11 Difference of Two Proportions

Example 4:

- Consider two groups of women:

  Group 1: women who spend less than \$500 annually on clothes.

  Group 2: women who spend over \$1000 annually on clothes.

- Let p_1 and p_2 be the proportions of women in the two groups who believe that the clothes are too expensive, respectively.

- If $Y_1 = 1009$ out of $n_1 = 1230$ women from Group 1 and $Y_2 = 207$ out of $n_2 = 340$ from Group 2 believe that the clothes are too expensive.

(a) Find a point estimate of $p_1 - p_2$.

(b) Find an approximate 95% CI for $p_1 - p_2$.

Solution of Example 4:

(a) Calculate the sample proportions

$$\hat{p}_1 = \frac{Y_1}{n_1} = 0.82, \quad \hat{p}_2 = \frac{Y_2}{n_2} = 0.61$$

The point estimate of $p_1 - p_2$ is $\hat{p}_1 - \hat{p}_2 = 0.21$.

(b) Since Y_1 and Y_2 are independent r.v.'s. $Y_1 \sim b(n_1, p_1)$ and $Y_2 \sim b(n_2, p_2)$,by the C.L.T.,

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}} \sim N(0,1)$$

From this we an approximate $100(1-\alpha)\%$ CI for $p_1 - p_2$ is $\hat{p}_1 - \hat{p}_2 \pm E$, where

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0.056.$$

Thus $\hat{p}_1 - \hat{p}_2 \pm E = [0.154, 0.266]$.

9 9.12 Estimating a Variance

Assumptions and Point Estimator

- Assume that $X_1, \ldots, X_n$ is a random sample from normal population with mean μ and variance σ^2.
- Both μ and σ^2 are unknown.
- The point estimator is the sample variance $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$. This is an unbiased and consistent estimator.

Confidence Interval

- $X^2 = (n-1)S^2/\sigma^2$ has a chi-squared distribution with $n-1$ degrees of freedom.
- There exist a and b such that

$$\begin{aligned} 1-\alpha = P(a \leqslant X^2 \leqslant b) &= P(a \leqslant (n-1)S^2/\sigma^2 \leqslant b) \\ &= P((n-1)S^2/b \leqslant \sigma^2 \leqslant (n-1)S^2/a) \end{aligned}$$

- For convenience we choose $a = \chi^2_{1-\alpha/2}$ and $b = \chi^2_{\alpha/2}$. Thus we have a $100(1-\alpha)\%$ confidence interval for σ^2: $[(n-1)S^2/\chi^2_{\alpha/2}, (n-1)S^2/\chi^2_{1-\alpha/2}]$.

Find Critical Value χ^2_α

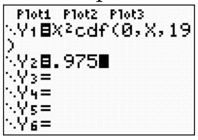
- Using Table A.5.

- Using TI 8x:
  (1) graph cdf curve $y = \chi^2 cdf(0, X, df)$ and horizontal line $y = 1 - \alpha$;
  (2) the x-coordinate of the intersection is χ^2_α.

Use TI-83/84 to Find Critical χ^2 Value For example, $n = 20$, df $=n - 1$=19, to find $\chi^2_{.025}$:

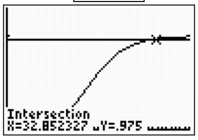
- Enter equation for χ^2 cdf : $\boxed{\text{Y=}}$ -¿ Y1= χ^2cdf(0, X, 19)

- Enter equation: Y2=1-.025=0.975.



- Press $\boxed{\text{GRAPH}}$ and adjust $\boxed{\text{WINDOW}}$ if needed.

- Press $\boxed{\text{2ND}}$ and $\boxed{\text{CALC}}$ to find intersection of the two graphs. We can obtain $\chi^2_{.025} = 32.85$.



Challenging Problems

- Find a and b so that
$$1 - \alpha = P((n-1)S^2/b \leqslant \sigma^2 \leqslant (n-1)S^2/a)$$
  and $\frac{1}{a} - \frac{1}{b}$ is minimized.

- If μ is known, how do you estimate σ^2 using point estimator and confidence interval?

10 9.13 Estimating Ratio of Variances

Assumptions and Point Estimator

- Assume that $X_{11}, \ldots, X_{1n_1}$ is a random sample from normal population with unknown mean μ_1 and variance σ_1^2, and

- $X_{21}, \ldots, X_{2n_2}$ is a random sample from normal population with unknown mean μ_2 and variance σ_2^2, and that

- the two samples are independent.

- The sample variances and sample means are

$$S_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2,$$

$$\bar{X}_i = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}, \quad i = 1, 2.$$

- The point estimator of σ_1^2/σ_2^2 is S_1^2/S_2^2, This is a consistent but biased estimator.

Confidence Interval

- $X_1^2 = (n_1 - 1)S_1^2/\sigma_1^2$ and $X_2^2 = (n_2 - 1)S_2^2/\sigma_2^2$ are independent and have chi-squared distributions with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom, respectively.

- So $F = \frac{X_1^2/v_1}{X_2^2/v_2} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$ has an F distribution with numerator and denominator degrees of freedom $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$, respectively.

Confidence Interval(Cont.)

- There exist a and b such that

$$1-\alpha = P(a \leqslant F \leqslant b) = P(a \leqslant \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} \leqslant b)$$

$$= P(\frac{S_1^2}{bS_2^2} \leqslant \frac{\sigma_1^2}{\sigma_2^2} \leqslant \frac{S_1^2}{aS_2^2})$$

- For convenience we choose $a = F_{1-\alpha/2}$ and $b = F_{\alpha/2}$. Thus we have a $100(1-\alpha)\%$ confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$:

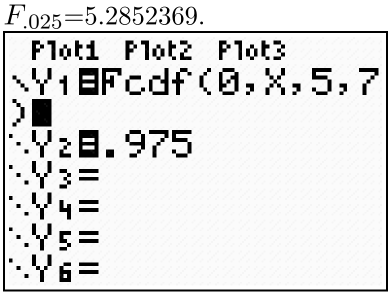
$$\left[\frac{S_1^2}{F_{\alpha/2}S_2^2}, \frac{S_1^2}{F_{1-\alpha/2}S_2^2}\right]$$

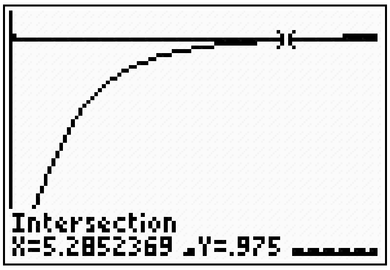
Find Critical Value F_α

- Using Table A.6.
- Using TI 8x:
  (1) graph Fcdf $y = Fcdf(0, X, v_1, v_2)$ and horizontal line $y = 1-\alpha$;
  (2) the x-coordinate of the intersection is F_α.

The F Critical Value using Technology

- In Excel: $F_{.025}$=F.INV.RT(0.025,5,7)=5.285236852, or =F.INV(1-0.025,5,7)=5.285236852.
- In TI-8x: Graph Y1=Fcdf(0, X, 5,7) and Y2=1-.025=0.975, the x-coordinate of the intersection is $F_{.025}$=5.2852369.





- In R: $F_{.975}$= "qf(.975,5,7)=5.285237".

Challenging Problems

- Find a and b so that

$$1-\alpha = P(\frac{S_1^2}{bS_2^2} \leqslant \frac{\sigma_1^2}{\sigma_2^2} \leqslant \frac{S_1^2}{aS_2^2})$$

  and $\frac{1}{a} - \frac{1}{b}$ is minimized.

- If μ_1 and μ_2 are known, how do you estimate σ_1^2/σ_2^2 using point estimator and confidence interval?