**Confidence Intervals and Sample Size Exercises**

* CL is the confidence level
* CI is the confidence interval
* is the mean estimate

Thus , if or 5%(CL is 95%),

* SB=Standard Bound
* CI, for sample mean
* CI=, for population variance
* The term ‘degrees of freedom’ may be thought of as the number of independent variables involved minus the number of constraints imposed.
* CI=, for the difference between means
* CI= , for the difference between proportions

1. **Confidence Intervals with Normal pdf**
2. Suppose we have collected data from a sample. We know the sample mean but we do not know the mean for the entire population. The sample mean is seven, and the error bound for the mean is 2.5:  =7 and Standard Bound=2.5. What is the confidence interval?

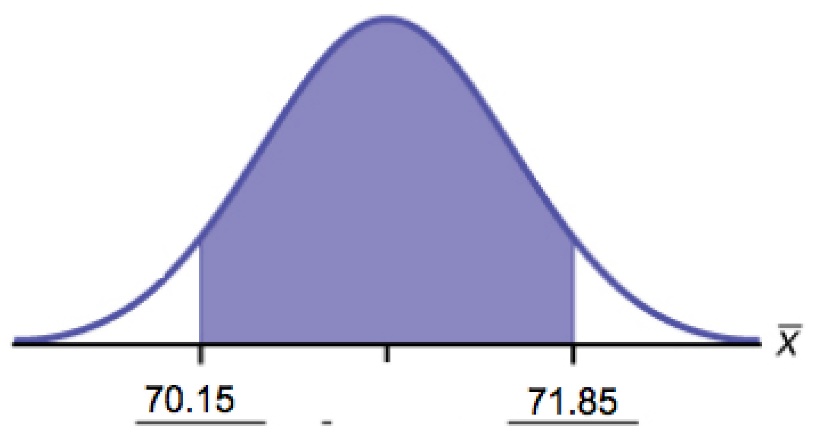
**Answer**:

The confidence interval is (7 – 2.5, 7 + 2.5) and calculating the values gives (4.5, 9.5). If the confidence level (CL) is 95%, then we say that, "We estimate with 95% confidence that the true value of the population mean is between 4.5 and 9.5."

1. Among various ethnic groups, the standard deviation of heights is known to be approximately three inches. We wish to construct a 95% confidence interval for the mean height of male Swedes. Forty-eight male Swedes are surveyed. The sample mean is 71 inches. The sample standard deviation is 3 inches.
   1. =
   2. σ=
   3. n=
2. In words, define the random variables x and .
3. Which distribution should you use for this problem? Explain your choice.
4. Construct a 95% confidence interval for the population mean height of male Swedes.
   * + 1. State the confidence interval.
       2. Sketch the graph.
       3. Calculate the error bound.
5. What will happen to the level of confidence obtained if 1,000 male Swedes are surveyed instead of 48? Why?

**Answer**:

* 1. 71
  2. 2.8
  3. n=48

1. x is the height of a Swiss male, and is the mean height from a sample of 48 Swiss males.
2. Normal. We know the standard deviation for the population, and the sample size is greater than 30. EB(48)=6.92, EB(1,000)=0.1735
   1. CI:
   2. 
3. The confidence interval will decrease in size, because the sample size increased. Recall, when all factors remain unchanged, an increase in sample size decreases variability. Thus, we do not need as large as interval to capture the true population mean
4. Suppose we have data from a sample. The sample mean is 15, and the error bound for the mean is 3.2. What is the confidence interval estimate for the population mean?

**Answer**:

1. Suppose scores on exams in statistics are normally distributed with an unknown population mean and a population standard deviation of three points. A random sample of 36 scores is taken and gives a sample mean of 68.
2. Find a 90% confidence interval for the true (population) mean of statistics exam scores.

**Answer**:

1. Find a 95% confidence interval for the true (population) mean statistics exam score.

**Answer**:

1. Why are they different?

**Answer**:

The 95% confidence interval can only exclude 5% of all possible results, therefor is wider than the 90% confidence interval which needs to exclude 10% of the mass.

1. ***Increasing the confidence level increases the error bound, making the confidence interval wider.***
2. ***Decreasing the confidence level decreases the error bound, making the confidence interval narrower.***
3. Suppose average pizza delivery times are normally distributed with an unknown population mean and a population standard deviation of six minutes. A random sample of 28 pizza delivery restaurants is taken and has a sample mean delivery time of 36 minutes.
4. Find a 90% confidence interval estimate for the population mean delivery time.

**Answer**:

Sample size n=28

With z

With t

1. Leave everything the same except the sample size. Use the original 90% confidence level. What happens to the error bound and the confidence interval if we increase the sample size and use n=100 instead of n=28? What happens if we decrease the sample size to n=25 instead of n=28?

**Answer**:

Sample size n=100

Samples size n=25

1. Discuss the difference
2. ***Increasing the sample size causes the error bound to decrease, making the confidence interval narrower.***
3. ***Decreasing the sample size causes the error bound to increase, making the confidence interval wider.***
4. The Specific Absorption Rate (SAR) for a cell phone measures the amount of radio frequency (RF) energy absorbed by the user’s body when using the handset. Every cell phone emits RF energy. Different phone models have different SAR measures. To receive certification from the Federal Communications Commission (FCC) for sale in the United States, the SAR level for a cell phone must be no more than 1.6 watts per kilogram. Table shows the highest SAR level for a random selection of cell phone models as measured by the FCC.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Phone Model** | **SAR** | **Phone Model** | **SAR** | **Phone Model** | **SAR** |
| Apple iPhone 4S | 1.11 | LG Ally | 1.36 | Pantech Laser | 0.74 |
| BlackBerry Pearl 8120 | 1.48 | LG AX275 | 1.34 | Samsung Character | 0.5 |
| BlackBerry Tour 9630 | 1.43 | LG Cosmos | 1.18 | Samsung Epic 4G Touch | 0.4 |
| Cricket TXTM8 | 1.3 | LG CU515 | 1.3 | Samsung M240 | 0.867 |
| HP/Palm Centro | 1.09 | LG Trax CU575 | 1.26 | Samsung Messager III SCH-R750 | 0.68 |
| HTC One V | 0.455 | Motorola Q9h | 1.29 | Samsung Nexus S | 0.51 |
| HTC Touch Pro 2 | 1.41 | Motorola Razr2 V8 | 0.36 | Samsung SGH-A227 | 1.13 |
| Huawei M835 Ideos | 0.82 | Motorola Razr2 V9 | 0.52 | SGH-a107 GoPhone | 0.3 |
| Kyocera DuraPlus | 0.78 | Motorola V195s | 1.6 | Sony W350a | 1.48 |
| Kyocera K127 Marbl | 1.25 | Nokia 1680 | 1.39 | T-Mobile Concord | 1.38 |

Find a 98% confidence interval for the true (population) mean of the Specific Absorption Rates (SARs) for cell phones. Assume that the population standard deviation is σ=0.337.

**Answer**:

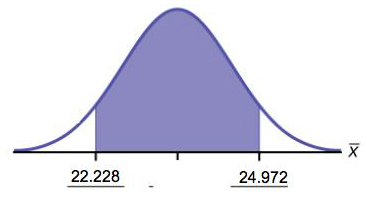
1. Suppose that an accounting firm does a study to determine the time needed to complete one person’s tax forms. It randomly surveys 100 people. The sample mean is 23.6 hours. There is a known standard deviation of 7.0 hours. The population distribution is assumed to be normal.
2. =
3. σ=
4. n=
5. In words, define the random variables x and   .
6. Which distribution should you use for this problem? Explain your choice.
7. Construct a 95% confidence interval for the population mean height of male Swedes.
   * + 1. State the confidence interval.
       2. Sketch the graph.
       3. Calculate the error bound.
8. If the firm wished to increase its level of confidence and keep the error bound the same by taking another survey, what changes should it make?
9. If the firm did another survey, kept the error bound the same, and only surveyed 49 people, what would happen to the level of confidence? Why?
10. Suppose that the firm decided that it needed to be at least 96% confident of the population mean length of time to within one hour. How would the number of people the firm surveys change? Why?

**Answer**:

=23.6

σ=7

n=100

1. X is the time needed to complete an individual tax form.   is the mean time to complete tax forms from a sample of 100 customers.
2. N because we know sigma.
   * + 1. (22.228, 24.972)
       2. 
       3. Standard Bound=1.372
3. It will need to change the sample size. The firm needs to determine what the confidence level should be, then apply the error bound formula to determine the necessary sample size.
4. The confidence level would increase as a result of a larger interval. Smaller sample sizes result in more variability. To capture the true population mean, we need to have a larger interval.
5. According to the error bound formula, the firm needs to survey 206 people. Since we increase the confidence level, we need to increase either our error bound or the sample size.
6. **Sample Size**
7. The population standard deviation for the age of Foothill College students is 15 years. If we want to be 95% confident that the sample mean age is within two years of the true population mean age of Foothill College students, how many randomly selected Foothill College students must be surveyed?

**Answer**:

1. The population standard deviation for the height of high school basketball players is three inches. If we want to be 95% confident that the sample mean height is within one inch of the true population mean height, how many randomly selected students must be surveyed?

**Answer**:

1. The average height of young adult males has a normal distribution with standard deviation of 2.5 inches. You want to estimate the mean height of students at your college or university to within one inch with 93% confidence. How many male students must you measure?

**Answer**:

1. **Using t-Student**
2. A pharmaceutical company makes tranquilizers. It is assumed that the distribution for the length of time they last is approximately normal. Researchers in a hospital used the drug on a random sample of nine patients. The effective period of the tranquilizer for each patient (in hours) was as follows: 2.7; 2.8; 3.0; 2.3; 2.3; 2.2; 2.8; 2.1; and 2.4.
3. =
4. =
5. n=
6. n-1=
7. In words, define the random variable x
8. and   .
9. Which distribution should you use for this problem? Explain your choice.
10. Construct a 95% confidence interval for the population mean length of time.
    * + 1. State the confidence interval.
        2. Sketch the graph.
        3. Calculate the error bound.
11. What does it mean to be “95% confident” in this problem?

**Answer**:

* + 1. =2.511
    2. =0.318
    3. n=9
    4. n-1=8

1. The effective length of time for a tranquilizer
2. the mean effective length of time of tranquilizers from a sample of nine patients
3. We need to use a Student’s-t distribution because we do not know the population standard deviation.
   * + 1. (2.267,2.755)
       2. Check plot
       3. Standard Bound==0.244
4. If we were to sample many groups of nine patients, 95% of the samples would contain the true population mean length of time
5. The Federal Election Commission (FEC) collects information about campaign contributions and disbursements for candidates and political committees each election cycle. A political action committee (PAC) is a committee formed to raise money for candidates and campaigns. A Leadership PAC is a PAC formed by a federal politician (senator or representative) to raise money to help other candidates’ campaigns.

The FEC has reported financial information for 556 Leadership PACs that operating during the 2011–2012 election cycle. The following table shows the total receipts during this cycle for a random selection of 30 Leadership PACs.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| $46,500.00 | $0 | $40,966.50 | $105,887.20 | $5,175.00 |
| $29,050.00 | $19,500.00 | $181,557.20 | $31,500.00 | $149,970.80 |
| $2,555,363.20 | $12,025.00 | $409,000.00 | $60,521.70 | $18,000.00 |
| $61,810.20 | $76,530.80 | $119,459.20 | $0 | $63,520.00 |
| $6,500.00 | $502,578.00 | $705,061.10 | $708,258.90 | $135,810.00 |
| $2,000.00 | $2,000.00 | $0 | $1,287,933.80 | $219,148.30 |

=$251,854.23

=$521,130.41

Use this sample data to construct a 96% confidence interval for the mean amount of money raised by all Leadership PACs during the 2011–2012 election cycle. Use the Student's t-distribution.

**Answer**:

1. Unoccupied seats on flights cause airlines to lose revenue. Suppose a large airline wants to estimate its mean number of unoccupied seats per flight over the past year. To accomplish this, the records of 225 flights are randomly selected and the number of unoccupied seats is noted for each of the sampled flights. The sample mean is 11.6 seats and the sample standard deviation is 4.1 seats.
2. =
3. =
4. n=
5. n-1=
6. In words, define the random variables x and   .
7. Which distribution should you use for this problem? Explain your choice.
8. Construct a 95% confidence interval for the population mean length of time.
   * + 1. State the confidence interval.
       2. Sketch the graph.
       3. Calculate the error bound.

**Answer**:

1. =
2. =
3. n=
4. n-1=
5. X is the number of unoccupied seats on a single flight. X¯ is the mean number of unoccupied seats from a sample of 225 flights.
6. We will use a Student’s t-distribution, because we do not know the population standard deviation.
7. Construct a 95% confidence interval for the population mean length of time.
   * + 1. (11.12,12.08)
       2. Sketch the graph.
       3. Standard Bound=0.48
8. **Population Proportion**
9. According to a recent survey of 1,200 people, 61% feel that the president is doing an acceptable job. We are interested in the population proportion of people who feel the president is doing an acceptable job.
   1. Define the random variables X and P' in words.
   2. Which distribution should you use for this problem? Explain your choice.
   3. Construct a 90% confidence interval for the population proportion of people who feel the president is doing an acceptable job.
      1. State the confidence interval.
      2. Sketch the graph.
      3. Calculate the error bound.

**Answer:**

1. X= the number of people who feel that the president is doing an acceptable job

P'=the proportion of people in a sample who feel that the president is doing an acceptable job.

* + 1. CI:(0.5868,0.6331)
    2. Check student’s solution
    3. Error==0.0232

1. **Sample Size Proportions**
2. You plan to conduct a survey on your college campus to learn about the political awareness of students. You want to estimate the true proportion of college students on your campus who voted in the 2012 presidential election with 95% confidence and a margin of error no greater than five percent. How many students must you interview?

**Answer**:

You need to interview at least 385 students to estimate the proportion to within 5% at 95% confidence.

1. **Population Variance**
2. A large candy manufacturer produces, packages and sells packs of candy targeted to weigh 52 grams. A quality control manager working for the company was concerned that the variation in the actual weights of the targeted 52-gram packs was larger than acceptable. That is, he was concerned that some packs weighed significantly less than 52-grams and some weighed significantly more than 52 grams. In an attempt to estimate σ, the standard deviation of the weights of all of the 52-gram packs the manufacturer makes, he took a random sample of n = 10 packs off of the factory line. The random sample yielded a sample variance of 4.2 grams. Use the random sample to derive a 95% confidence interval for σ.

**Answer:**

Since the degrees of freedom are n-1=9, then and . Thus

And taking square root:

That is, we can be 95% confident that the standard deviation of the weights of all of the packs of candy coming off of the factory line is between 1.41 and 3.74 grams.

1. A random sample of 20 nominally measured 2mm diameter steel ball bearings is

taken and the diameters are measured precisely. The measurements, in mm, are

as follows:

2.02 1.94 2.09 1.95 1.98 2.00 2.03 2.04 2.08 2.07

1.99 1.96 1.99 1.95 1.99 1.99 2.03 2.05 2.01 2.03

Assuming that the diameters are normally distributed with unknown mean, µ, and

unknown variance σ2:

* + - * 1. find a two-sided 95% confidence interval for the variance, σ2
        2. find a two-sided confidence interval for the standard deviation, σ.

**Answer:**

From the data, we calculate = 40.19 and = 80.7977. Hence

There are 19 degrees of freedom and the critical values of the -distribution are:

and

(a) the confidence interval for σ2 is

≡ 1.0927 × 10−3mm < σ2 ≤ 4.0286 × 10−3mm

(b) the confidence interval for σ is

≡ 0.033mm < σ < 0.063 mm

1. **Difference in Mean**
2. Take as an example the data from the "Animal Research" case study. In this experiment, students rated (on a 7-point scale) whether they thought animal research is wrong. The sample sizes, means, and variances are shown separately for males and females in Table 1.

Table 1. Means and Variances in Animal Research study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Condition** | **n** | **Mean** | **Variance** |
| Females | 17 | 5.353 | 2.743 |
| Males | 17 | 3.882 | 2.985 |

Find the confidence interval for the difference in opinion of male and females and interpret the result

Answer:

Since both samples are the size:

We estimate  by pooling the sample variances

This analysis provides evidence that the mean for females is higher than the mean for males, and that the difference between means in the population is likely to be between 0.29 and 2.65.

1. Consider the following small example:

|  |  |
| --- | --- |
| **Group 1** | **Group 2** |
| 3 | 2 |
| 4 | 4 |
| 5 |  |

Find the confidence interval for the difference in the mean

**Answer:**

= 4 and  = 3

Then,  is computed by:

where the degrees of freedom (df) is computed as before:

df = (n-1) + (m-1) = (3-1) + (2-1) = 3+2-2= 3.

1. **Difference in Proportion**
2. What is the prevalence of anemia in developing countries?

|  |  |  |
| --- | --- | --- |
| **Data** | **African Women** | **Women from Americas** |
| Sample size | 2100 | 1900 |
| Number with anemia | 840 | 323 |
| Sample proportion | 840/2100 = 0.40 | 323/1900 = 0.17 |

Find a 95% confidence interval for the difference in proportions of all African women with anemia and all women from the Americas with anemia.

**Answer:**

Substituting in the numbers that we know into the formula for a 95% confidence interval for p1−p2, we get:

which simplifies to:

We can be 95% confident that there are between 20.3% and 25.7% more African women with anemia than women from the Americas with anemia.

1. A social experiment conducted in 1962 involved n = 123 three- and four-year-old children from poverty-level families in Ypsilanti, Michigan. The children were randomly assigned either to (1) a treatment group receiving two years of preschool instruction, or to (2) a control group receiving no preschool instruction. The participants were followed into their adult years. Here is a summary of the data:

Table 1. People arrested for some crime

|  |  |  |
| --- | --- | --- |
| **Data** | **Yes** | **No** |
| Control | 32 | 30 |
| Preschool | 19 | 42 |

Find a 95% confidence interval for p1−p2, the difference in the two population proportions.

**Answer:**

Of the n = 62 children serving as the control group, 32 were later arrested for some crime, yielding a sample proportion of:

And, of the m = 61 children receiving preschool instruction, 19 were later arrested for some crime, yielding a sample proportion of:

A 95% confidence interval for p1−p2 is therefore:

which simplifies to:

We can be 95% confident that between 3.5% and 37.5% more children not having attended preschool were arrested for a crime by age 19 than children who had received preschool instruction.