Elementary Statistics: Math 080

Jordan Hanson

August 4, 2020

Whittier College Department of Physics and Astronomy

Summary

Summary

Unit 2 and 3

1. Central Limit Theorem: 7.1
2. Confidence Intervals and Hypothesis Testing
   - Confidence intervals and data interpretation: 8.1 - 8.4
   - Rejecting the null hypothesis, types of error, underlying distributions: 9.1 - 9.3, 9.6

The Central Limit Theorem

The Central Limit Theorem

The Central Limit Theorem

Central Limit Theorem: Let X be a continuous random variable, with mean μ_x and standard deviation σ_x. The average $\bar{x}$ of n values of X is normally distributed like $N(\mu_x, \sigma_x/\sqrt{n})$.

The Central Limit Theorem

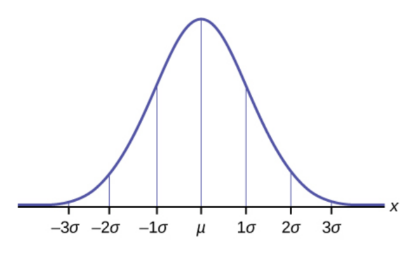

Figure 1: The normal distribution about a mean μ with the units of standard deviations shown.

Example: An unknown distribution has a mean of $\mu = 90$ and a standard deviation of $\sigma = 15$. Samples of size $n = 25$ are drawn randomly from the population. Find the probability that the *sample mean* is between 87 and 93.

Interactive Questions

Interactive Questions: Central Limit Theorem

Suppose we take samples of size $n = 16$ from a large data set and compute the averages and standard deviations of the samples. Suppose we repeat the whole process, but change $n = 100$. Which of the following is true?

- A: The means of our samples will shift upwards by a factor of $100/16$.
- B: The means of our samples will shift downwards by a factor of $100/16$.
- C: The standard deviations of our samples will shift downwards by a factor of $\sqrt{100/16}$.
- D: The standard deviations of our samples will shift upwards by a factor of $\sqrt{100/16}$.

Interactive Questions: Central Limit Theorem

Suppose we take samples of size $n = 100$ from a large data set that has mean μ and standard deviation σ, and compute the averages and standard deviations of the *samples*. Which of the following is true?

- A: Each standard deviation of each sample we collect will be $\sigma/10$.
- B: The standard deviation of the means of our samples will be $\sigma/10$.
- C: Each mean of each sample we collect will be $\mu/10$.
- D: The mean of the means of our samples will be μ.

Confidence Intervals

Confidence Intervals

Suppose we are attempting to measure a population parameter, such as the mean, μ. This is distinct from the *sample mean*, $\bar{x}$, of a set of data $\{x_i\}$.

Figure 2: Let the gray curve refer to a *population* of data. Let the red curve refer to the *distribution of averages* we find if we repeatedly sample. The standard deviation of the red curve is smaller because of the Central Limit Theorem: $\sigma_{\bar{x}} = \sigma_{pop}/\sqrt{n}$.

Confidence Intervals

Figure 3: Suppose we decide to communicate the range in which the true mean μ falls, using our *point estimates* $\bar{x}$. How should we do that?

A good range: (Mean of $\bar{x}$ distribution minus two standard deviations), to (mean of $\bar{x}$ distribution plus two standard deviations). **This interval has a 90 percent chance of containing μ.**

Confidence Intervals

(Mean of $\bar{x} - 2s/\sqrt{n}$), to (Mean of $\bar{x} + 2s/\sqrt{n}$). **This interval has a 90 percent chance of containing μ.**

- The error bound for the population mean (EBM) is the "two standard deviations" in this example, in which we know the population σ.
- Recall the 68 - 95 - 99.7 rule, and how sample means are normally distributed regardless of the underlying distribution.
- We choose the confidence level (CL), and construct the confidence interval for the given CL.
- $\alpha = 1 - CL$, probability the confidence interval *does not contain* μ.

Professor example, next slide.

Confidence Intervals

Professor example.

PhET: Illustration of Confidence Intervals

PhEt: Confidence Intervals

Go to this link: `https://digitalfirst.bfwpub.com/stats_applet/stats_applet_4_ci.html`

1. Select a confidence level (CL) of 95 percent using the slide bar at left.
2. Set your sample size to 100. The sample means will be distributed like $N(\mu, \sigma/\sqrt{n})$. This example assumes we know σ, the population standard deviation, but we do not know the mean μ.
3. Hit the green Sample 25 button to collect 25 samples of size 100, to get 25 means.
4. How many of these means fall within two standard deviations $\sigma/\sqrt{100}$ (CL: 95 percent) of the true mean μ?
5. Repeatedly hit Sample 25, and notice what happens to the percentage at left of sample means that fall within 2 $\sigma/\sqrt{100}$ of μ.

Interactive Questions

Interactive Questions

Suppose you wanted to measure the average age of inhabitants of a large neighborhood. The population mean μ is unknown, but we can be reasonably certain that our sample standard deviations, s_1, s_2, ... are approximately σ, the standard deviation of the population. So we know that the $\sigma_{age} = 20$ years. Suppose we collect 100 samples of 16 people apiece. This means the standard error in our mean $20/\sqrt{16} = 5$ years.

1. Suppose the mean of all our data is 70 years. What is the EBM for two standard deviations in the mean?
2. How many of our samples should fall within 1 EBM of the true mean?

Solve via Breakout Rooms! Go.