

1 Inference for Quantitative Data: Means

1.1 Constructing a One Sample t-Interval

Sampling Distribution for Means:

- Shape - normal population indicates a normal sampling distribution. $n < 30$ then data needs to be stated as approximately normal. $n \geq 30$ satisfies Central Limit Theorem (CLT) which guarantees approximately normal
- Center - as long as you are taking a random sample or using random assignment, $\mu_{\bar{x}} = \mu$
- When sampling, as long as 10% condition is met, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

As long as the above conditions are met, the sampling distribution of $\bar{x} : \bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

When we do not know the population standard deviation (which we usually don't), we must estimate it from our sample using the sample standard deviation, s . However, when we do so, the test statistic (z) that we previously used changes. The new test statistic is now called the t-statistic and has a new distribution associated with it.

The new t-distribution is not exactly like the standard normal curve, but it is very close:

- It is still centered at 0
- It is bell shaped
- Its spread is slightly greater than the standard normal distribution
- It has more area in the tails

The special thing about t-distributions is the fact that it is a family of distributions. There is a unique density curve for each sample size (and the dependence on sample size is taken care of by degrees of freedom: $n - 1$.)

The tails on a t-distribution are "fatter" than a standard normal distribution. This is true because the smaller our sample size, the more variation we have, hence the fatter tails. The larger our sample size gets, the closer the t-distribution will have towards the standard normal distribution.

Constructing a confidence interval

- Define the parameter. μ = true mean of {population parameter in context}
- Check the assumptions and conditions:
 - Random: The sample should be a random sample of the population or random assignment in an experiment.
 - Independence (10% Condition): The sample size, n , must be no larger than 10% of the population.
 - Normality: There are multiple ways to verify this condition.
 - * Stated in problem: It may be stated in the problem that the sampling distribution is approx. normal
 - * Central Limit Theorem: When n is large, the sampling distribution of the sample means is approx. normal.
 - * Visual representation: You are given a graphical representation (histogram or boxplot) that depicts a shape that is approximately normal. You may also be given data that you have to graph yourself (include a sketch). You are looking for no strong skew or outliers.
- Name the inference method: One Sample t-Interval

- Calculate the interval: point estimate \pm margin of error: $\bar{x} \pm t_{n-1}^* \left(\frac{s}{\sqrt{n}} \right)$. Use 2nd-Vars-4:invT() in the calculator to calculate the critical value.
- Write your conclusion in context

You can determine the sample size from the formula $ME = t_{n-1}^* \left(\frac{s}{\sqrt{n}} \right)$.

1.2 Constructing a One Sample t-Test

Constructing a significance test:

1. Define the parameter: μ = true mean of {population parameter in context}.
2. State the hypothesis: the null and alternative.
3. Check the assumptions and conditions (same as last topic)
4. Method: One Sample t-Test
5. Test statistic is statistic-parameter divided by standard deviation of statistic and $t = \frac{\bar{x} - \mu_0}{\left(\frac{s}{\sqrt{n}} \right)}$.
6. Obtain the p-value from tcdf.
7. Make a decision.
8. State your conclusion on context.

1.3 Inference for Paired Data

Comparative studies are more convincing than single-sample investigations. For that reason, one-sample inference is less common than comparative inference. Study designs that involve making two observations on the same individual or one observation on each of two similar individuals, result in paired data.

When paired data result from measuring the same quantitative variable twice, we can make comparisons by analyzing the differences in each pair. If the conditions for inference are met, we can use one-sample t-procedures to perform inference about the mean difference: μ_D . These methods are called matched pairs procedures.

1.4 Inference for Comparing Two Sample Means

Same PANIC and PHANTOM as before.

Note that the degrees of free dom $n - 1$ is where n is the smaller sample size (this is known as the conservative df).

The confidence interval formula is $(\bar{x}_1 - \bar{x}_2) \pm t_{n-1}^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

The test statistic in a two sample t-test can be calculated by $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$.