Sampling Distributions & One Sample Tests Notes[1]

Stanford GSE Math Camp

Do Not Distribute Outside GSE

1 Sampling Distributions for Means

The Sampling Distribution for Means is similar to that of proportions but a bit more complex.

- The mean for the sampling distribution of means is given by:

$$\mu_{\bar{x}} = \mu$$

- The standard deviation for the sampling distribution of means is given by:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

  - *Note: This formula requires that we know the population standard deviation. If we are trying to make predictions about the population mean, it's incredibly unlikely we'll know the population standard deviation.*

- If we don't know the population standard deviation, then we need to estimate it. When we are forced to estimate a parameter, the calculation is called a **standard error** rather than a standard deviation. The formula for **standard error of the mean (SEM)** is:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

Note: We are using the sample standard deviation in the numerator to estimate the population standard deviation.

- Just like in the sampling distribution for proportions, we must check conditions. Most conditions are the same.
  - **Randomization Condition**(random sample)
  - **10% Condition**, which says that the sample size n should be no more than 10% of the population
  - **Nearly Normal Condition**, which can be met if (1) the population is normal, (2) you are given a dataset and its graph is normal, or (3) the Large Enough Sample Condition is satisfied
    - **Large Enough Sample Condition**, which states that, because of the Central Limit Theorem, large enough samples will result in a sampling distribution that fits a normal model. There is no rule for how large a sample you'll need, but a typical rule of thumb is $n \geq 30$. There are also other ways to check for normality, if your sample is small.

- Thus,if conditions are met, the sampling distribution for means is given by:

Known σ	Unknown σ
$\sim N(\mu, \frac{\sigma}{\sqrt{n}})$	$\sim N(\mu, \frac{s}{\sqrt{n}})$

[1]Contributor(s): Kelly Boles & Erin Fahle. If you find errors, please let us know so that we may correct them. Thanks!

2 1 Sample Z or T Test for Means (See handouts)

To conduct a 1 Sample Z or T Test for Means, we must follow these steps:

1. Name test, write hypothesis & describe parameter
2. Check assumptions/conditions
   - Independence Assumption
   - Randomization Condition
   - Nearly Normal Condition
   - 10% Condition
3. Do calculations to get test statistic & p-value.
   - We'll have to determine whether a z-test or t-test is appropriate. We use a z-test when the population standard deviation is known (σ) or the sample size is large. Otherwise, t-tests work best for small samples.
     - While you calculate a t-score just like you would a z-score (but using a t table), the t distribution has an additional parameter called degrees of freedom. For one sample tests, degrees of freedom are given by n-1.
4. Interpret results & draw conclusions
   - If the p-value is smaller than some *α-level* or *significance level*, we will reject the null hypothesis and say the mean is *statistically significantly* different/greater than/less than from the null hypothesis.
   - Typically, we use $\alpha = 0.001, 0.01, 0.05, 0.10$.

3 1 Sample T Interval for Means (See handout)

- A confidence interval provides us a range of values that the true population mean could take for a given confidence level. For example, we can construct a 95% confidence interval. We then can say that we are 95% confident the true mean falls in our interval.
- To make a confidence interval for a mean (a one-sample t-interval), we use the following equation:

$$\bar{x} \pm t^*_{n-1} \times SE(\bar{x})$$

where $\alpha = 1 - \frac{confidence\ \%}{100}$.