Formulas Exam I

Standard Deviation

Samples: $s_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$ or $s_x = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}}$

Populations: $\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$ or $\sigma = \sqrt{\frac{\sum x^2}{N} - \mu^2}$

Sampling Distribution of the Sample Mean and Central Limit Theorem (CTL)

The distribution of all sample means of size n from a random variable with mean μ_x and standard deviation σ_x has the following properties:

i. $\mu_{\bar{x}} = \mu_x$

ii. $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$

iii. If the population from where the samples are drawn is normal, then the distribution of the $\bar{x}$'s is normal regardless of the sample size *n*.

iv. CTL: If the population from where the samples are drawn in not normal, you need at least *n=30* for the distribution of the $\bar{x}$'s to be approximately normal.

Confidence Intervals

For Population Means:

Known σ :

If $n \geq 30$ use $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ If $n < 30$ use $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ **only if X is normal**

Unknown σ :

If $n \geq 30$ use $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$ If $n < 30$ use $\bar{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$ **only if X is normal**

For the Difference Between Means of Two Independent Random Variables

Known σ_1 and σ_2 (no need to worry about them being equal or not):

If both n_1 and $n_2 \geq 30$:

$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}}$$

otherwise the random variable where the samples come from need to be normally distributed for the above confidence interval to be valid.

Unknown σ_1 and σ_2:

If there is evidence that the standard deviations σ_1 and σ_2 of the random variables are the same (rule of thumb: if the largest of the sample standard deviation s_1 and s_2 is about twice as large as the smallest):

$$(\bar{x} - \bar{y}) \pm t_{\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where $s_p = \sqrt{\frac{\sum(x_i-\bar{x})^2 + \sum(y_i-\bar{y})^2}{n_1+n_2-2}} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$

If σ_1 and σ_2 cannot be assumed to be the same:

$$(\bar{x} - \bar{y}) \pm t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_1^2/n_2)^2}{n_2-1}}$$

Interpretation of C.I's:

I am 95% confident the population mean of ______ " add subject matter" is between______ and _______.

Interpretation ofTest of Hypothesis:

The data shows evidence to reject/not reject Ho that the population mean of mean of ______ " add subject matter) is equal to _____ in favor of the alternative Ha that the population mean of _____"add subject matter" is (not) larger/smaller/different from ____.