MATH 254: Introduction to Statistics **Mintaek Lee**

Glossary of Statistical Terms

Corresponding Workbook Modules: 3 – 6

FOR EXAM 2:

Following terms are usually associated with hypothesis testing and/or confidence intervals for both means and proportions.

- H_0: null hypothesis (always has '=' sign, like $\mu = 10$)
- H_A: alternative hypothesis (has '$\neq$', '$<$', or '$>$' sign based on problems, like $\mu > 10$)
- α: significance level (for hypothesis testing), usually $\alpha = 0.01$ or 0.05, but it can vary
- z: z-test statistic (calculate it using values like: n, σ, etc.)
- t: t-test statistic (calculate it using values like: n, s, etc.)
- $d.f.$: degrees of freedom (only used for t). For one-sample and paired two sample, it is $n - 1$. For independent two-sample (not paired), it is $\min\{n_1 - 1, n_2 - 1\}$.
- z^*: z critical value (used in the z-confidence interval, find it from the t table)
- t^*: t critical value (used in the t-confidence interval, find it from the t table)
- m: margin of error (the one to the right of '$\pm$' in the confidence interval)

Following terms are usually associated with inference for a single mean (Module 3).

- N: population size
- n: sample size
- μ: population mean
- μ_0: hypothesized population mean for hypothesis testing (values from hypotheses)
- $\bar{x}$: sample mean
- σ: population standard deviation (if known, use z. otherwise, use t.)
- s: sample standard deviation (calculate it from sample)

Following terms are usually associated with inference for a single proportion (Module 4).

- p: population proportion
- $\hat{p}$: sample proportion $\left(\frac{x}{n}\right)$ where x is the number of successes.
- p_0: hypothesized population proportion for hypothesis testing (values from hypotheses)
- p^*: given population proportion for the sample size calculation for a desired margin of error (usually one of 0.5, p_0, or p from the context. it depends on problems)

FOR EXAM 3:

Note that n, μ, $\bar{x}$, σ, and s can have subscripts for two-sample means cases (Activity 5-1).

- If they have 1 or 2 as subscripts, it means they are for the two independent groups. For example, μ_1 and μ_2 would be the population means of first and second group.
- If they have D as subscripts, it means they are for the paired (dependent) groups. For example, $\bar{x}_D$ would be the sample mean of differences of observations between two paired groups.
- For definitions of n, μ, $\bar{x}$, σ, and s, see the front page.

Following terms are usually associated with two-sample proportions cases (Activity 5-3).

- p_1: population proportion for the first group
- p_2: population proportion for the second group
- $\hat{p}_1$: sample proportion for the first group ($\frac{x_1}{n_1}$)
- $\hat{p}_2$: sample proportion for the second group ($\frac{x_2}{n_2}$)
- $\hat{p}$: pooled sample proportion ($\frac{x_1+x_2}{n_1+n_2}$) (used only in the hypothesis testing. notice that this is not the same as $\hat{p}$ from the one sample proportion cases)

Following terms are associated with regressions (Module 6).

- x: the explanatory (or independent) variable
- y: the observed response (or dependent) variable (obtained from the data)
- $\hat{y}$: the predicted response variable (calculated from the regression model)
- $y - \hat{y}$: residual (the vertical distance between y and $\hat{y}$)
- β_0: population slope of the regression line (usually unknown)
- β_1: population y-intercept of the regression line (usually unknown)
- b_0: sample slope of the regression line
- b_1: sample y-intercept of the regression line (*Excel outputs under "Coefficients" give both b_0 and b_1*)
- r: correlation coefficient, measures the strength and direction of a linear relationship (see workbook page 148 for more information) (*Excel output "Multiple R" is the same as $|r|$, but be careful when $r < 0$!*)
- r^2: r squared, interpretation of this is: the percent of variation in y that is due to (or explained by) the variation in x. (*Excel output "R Square" is the same as r^2*)