MA 207 General Statistics Formula Sheet

Topic	Assumptions			
Sample Mean $\bar{x}$	Observations $x_1, x_2, \ldots, x_n$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}$	The balance point of the distribution	
Sample Variance s^2	Observations $x_1, x_2, \ldots, x_n$	$s^2 = \frac{\sum (x-\bar{x})^2}{n-1}$	The sum of the squared deviations from the mean divided by one less than the sample size	
Sample Standard Deviation s	Observations $x_1, x_2, \ldots, x_n$	$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$	A typical distance or a type of average distance of an observation from the mean	
***z*-score** z		$z = \frac{\text{observation} - \text{mean}}{\text{standard deviation}}$	The number of standard deviations an observation falls above or below the mean	
Binomial Distribution of the Count of Successes / Sampling Distribution of the Count of Successes X	1. Each of the *n* trials has two possible outcomes. 2. Each of the *n* trials has the same probability of success, *p*. 3. The *n* trials are independent.	$\mu_X = np$ $\sigma_X = \sqrt{np(1-p)}$	The binomial distribution is approximately normal if $np \geq 15$ and $n(1-p) \geq 15$. Sampling without replacement is close to independent if $n < 10\%$ of the population.	Because the value of a statistic varies from sample to sample, we get an entire distribution of possible values for it; the sampling distribution describes probabilities for all possible values of the statistic.
Sampling Distribution of the Sample Proportion $\hat{p} = \frac{X}{n}$	Random samples of size n from a population with proportion p of successes	$\mu_{\hat{p}} = p$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	The sampling distribution of the sample proportion is approximately normal if $np \geq 15$ and $n(1-p) \geq 15$ and $n < 10\%$ of the population.	
Sampling Distribution of the Sample Mean $\bar{x}$	Random samples of size n from a population with mean μ and standard deviation σ	$\mu_{\bar{x}} = \mu$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	The mean of the sampling distribution of the sample mean $\bar{x}$ is the same as the mean of the population distribution x. The variability of sample means decreases as the sample size increases (averages are less variable than individuals).	
Central Limit Theorem	Random samples of size n from a population with mean μ and standard deviation σ	$\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$	If the population is normal, the sampling distribution of the sample mean is also normal. Even if the population is not normal, the sampling distribution of the sample mean is approximately normal if the sample size is large enough (at least $n \geq 30$ is usually enough).	

Statistical Inference Formulas

Topic	Assumptions	**Confidence Interval** estimate ± margin of error estimate ± (*z** or *t**)(standard error) (Use the critical value *z** or *t** appropriate for the confidence level; do NOT use the test statistic)	Null Hypothesis	**Test Statistic** $\dfrac{\text{estimate} - \text{null hypothesis value}}{\text{standard error of the estimate}}$
One sample population proportion	Data obtained by randomization (such as a random sample or randomized experiment) CI: $n\hat{p} \geq 15$ and $n(1-\hat{p}) \geq 15$ Test: $np_0 \geq 15$ and $n(1-p_0) \geq 15$	$\hat{p} \pm z^* \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$	$H_0 : p = p_0$	$z = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$
One sample population mean	Data obtained by randomization Approximately normal population *or* $n \geq 30$	$\bar{x} \pm t^* \dfrac{s}{\sqrt{n}}$ $df = n-1$	$H_0 : \mu = \mu_0$	$t = \dfrac{\bar{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$ $df = n-1$
Two sample population proportion	Data obtained by randomization Two independent samples $n_1\hat{p}_1 \geq 10$ and $n_1(1-\hat{p}_1) \geq 10$ $n_2\hat{p}_2 \geq 10$ and $n_2(1-\hat{p}_2) \geq 10$ (5 instead of 10 if two-sided test)	$(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	$H_0 : p_1 = p_2$	$z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ where $\hat{p} = \dfrac{X_1 + X_2}{n_1 + n_2}$ is the pooled proportion
Two sample population mean	Data obtained by randomization Two independent samples Approximately normal populations *or* $n_1, n_2 \geq 30$	$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$ $df = \min(n_1 - 1, n_2 - 1)$	$H_0 : \mu_1 = \mu_2$	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $df = \min(n_1 - 1, n_2 - 1)$