Final Exam: Cumulative Review (Math 115 / Statistics)

1.1 statistics, descriptive *v.* inferential, quantitative *v.* qualitative, population *v.* sample

1.2 simple random sampling (use Table 1, Appendix II), identify random, stratified, systematic, cluster & convenience samples, sampling *v.* non-sampling errors (see 1.3/p.26 regarding "potential pitfalls")

1.3 census, observational *v.* experimental, placebo, control group, double-blind

2.1 raw *v.* group data, frequency, relative frequency, frequency distribution, class width, class limits, histogram, distribution shapes (p.49), outlier, cumulative frequency & ogive

2.2 bar graphs (cluster & Pareto); circle/pie graphs, time series

2.3 stem-and-leaf display

3.1 find the mode, median (MD), mean ($\bar{x}$ or μ), and 5% trimmed mean for either raw data or grouped data; find a weighted average

3.2 find the standard deviation (s or σ) for either raw data or grouped data (see p.117); find the coefficient of variation (CV); use Chebyshev's Theorem (formula for "k" provided)

3.3 find the five-number summary values;
box-and-whisker plots not covered

4.1 Sample space, n(s), event/outcome notation and terminology, probability notation, find the probability values for events

4.2 find probability values for compound events, independence, mutually exclusive, conditional probability – see formula summary, p.168

4.3 **multiplication rule for counting**, factorial, permutations (${}_nP_r$), and combinations (${}_nC_r$)

5.1 discrete *vs* continuous; probability distribution (table-graph depicting P(***x***) for each outcome, sum equals one); "expected value" is the mean, $\mu = \sum \boldsymbol{x} \cdot P(\boldsymbol{x})$

5.2 binomial distribution characteristics (p.212); probability formula (p.216): $\mathbf{P(r) = {}_nC_r \times p^r \times (1-p)^{n-r}}$ (provided on test)

5.3 binomial distribution with "n" trials has mean ($\mu = np$) and standard deviation ($\sigma = \sqrt{np(1-p)}$); skewed or symmetric

6.1 normal curve/distribution characteristics (p.273); empirical rule (p.274 / percent values provided on test); graph & interpret a control chart (p.279 / graphics-box provided on test ~ signals I-III)

6.2 standard normal distribution ($\mu=0$, $\sigma=1$); z-score; left-tail distribution table (provided on test) for $P(z < b)$ values

6.3 use distribution table to find $P(a < x < b)$ for any normal distribution given μ & σ; determine z-score(s) when $P(z < b)$ is known; determine raw score "x" given μ, σ, and $P(x < b)$

6.4 population *vs* sample; sample statistics, mean sampling distribution (*i.e.*, for $\bar{x}$)

6.5 Central Limit Theorem (p.321): $\mu_{\bar{x}} = \mu$ & $\sigma_{\bar{x}} = \sigma \div \sqrt{n}$ standard deviation of $\bar{x}$ is *a.k.a.* the "standard error"; find $P(\bar{x} < b)$ using the standard normal distribution table

6.6 use the normal distribution as an approximation to the binomial distribution by applying a continuity correction(s)

7.1 Estimate μ when σ is known with a confidence interval where the margin of error, $E = z_c \times \sigma \div \sqrt{n}$

7.2 Estimate μ when σ is unknown with a confidence interval where the margin of error, $E = t_c \times s \div \sqrt{n}$ (d.f. = n − 1)

7.3 Estimate p with a confidence interval where the margin of error, $E = z_c \times \sigma \div \sqrt{n}$ ($\sigma = \sqrt{p(1-p)}$, using $\bar{p}$ when p is unknown)

7.4 Estimate a confidence interval for difference between means ($\mu_1 - \mu_1$) or population percentages ($p_1 - p_2$), interpret results

8.1 determine null hypothesis (H_0) & alternate hypothesis (H_1)

8.2 apply hypothesis testing for μ at significance level α by finding the P-value using z_c or t_c (as required)

8.3 apply hypothesis testing for p at significance level α by finding the P-value using z_c

9.1 graph a scatter plot, compute the correlation coefficient "r," $r \approx \pm 1 \Leftrightarrow$ strong positive/negative linear correlation, whereas $r \approx 0 \Leftrightarrow$ no/weak linear correlation between the variables

9.2 find, graph and/or use the least-square line, $y = a + bx$

Reference Formulas & Information (provided on final examination):

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \qquad \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2 \div n}{n}} \qquad s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \qquad \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2 \div n}{n-1}}$$

Permutations: ${}_nP_r = \frac{n!}{(n-r)!}$ Combinations: ${}_nC_r = \frac{n!}{r!(n-r)!}$

Probability of "r" success in "n" trials of a binomial experiment: $P(r) = {}_nC_r \times p^r \times (1-p)^{n-r}$

Central Limit Theorem:
in a sampling distribution of means (*i.e.*, for $\bar{x}$) when $n \geq 30$, or if "x" is normally distributed, $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \sigma \div \sqrt{n}$

Minimum sample size "n" for estimating the population mean (μ): $n = \left(\frac{z_c \times \sigma}{E}\right)^2$

Standard Deviation for the Difference of...

Means ($\mu_1 - \mu_2$) Testing:

$\overline{\sigma} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ use s_1 & s_2 when σ_1 & σ_2 are unknown

Percentages/Proportions ($p_1 - p_2$) Testing:

$\overline{\sigma} = \sqrt{\frac{\overline{p_1}(1-\overline{p_1})}{n_1} + \frac{\overline{p_2}(1-\overline{p_2})}{n_2}}$

Regression:

Correlation Coefficient, $r = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{n \cdot \sum_{i=1}^{n}(x_i)^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \sqrt{n \cdot \sum_{i=1}^{n}(y_i)^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}$

Least-Squares Line, $\hat{y} = a + bx$

$b = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n \cdot \sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2}$ and $a = \bar{y} - b\bar{x}$