Formula Sheet for Exam #3

1. Sample mean:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{\text{Sum of Scores}}{\text{Number of Scores}}$$

2. Sample proportion - if data are in 0/1 form (i.e., 0 means the absence of a characteristic and 1 means the presence of the characteristic) then you can use the same formula as the sample mean. If data are presented in 2-group form (i.e., group A and group B), then the proportion of cases in the sample belonging to group A is:

$$p(A) = \theta = \frac{\text{Number of cases in Group A}}{\text{Total Number of Cases}}$$

3. Standard deviation for a binary/dichotomous variable (leading to a proportion, θ):

$$\text{Standard Deviation} = s = \sqrt{\theta(1-\theta)}$$

4. Standard deviation of a continuous variable:

$$\text{Standard Deviation} = s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{X})^2}$$

5. Standard error of a sample mean or proportion:

$$\text{standard error} = \frac{\text{standard deviation}}{\sqrt{N}} = \frac{s}{\sqrt{N}}$$

6. Converting a raw score into a z-score:

$$z\text{-score} = \frac{x - \overline{X}}{s}$$

7. Calculating a single-sample hypothesis test statistic for a t-distribution:

$$t\text{-statistic} = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

with $n-1$ degrees of freedom.

8. Calculating a single sample hypothesis test statistic for a z-distribution (continuous variables):

$$z\text{-statistic} = \frac{\overline{X} - \mu}{s/\sqrt{n}}$$

9. Calculating a single-sample hypothesis test statistic for a z-distribution (sample proportion $= \hat{\theta}$):

$$z\text{-statistic} = \frac{\hat{\theta} - P}{\sqrt{P(1-P)/n}}$$

10. Confidence interval around a sample mean, $\overline{X}$, with large samples:

$$\text{Lower Confidence Limit} = \overline{X} - (z\text{-multiplier} \times \text{standard error})$$
$$\text{Upper Confidence Limit} = \overline{X} + (z\text{-multiplier} \times \text{standard error})$$

with z-multiplier chosen for the appropriate significance level.

11. Confidence interval around a sample mean, $\overline{X}$, with small samples:

$$\text{Lower Confidence Limit} = \overline{X} - (t\text{-multiplier} \times \text{standard error})$$
$$\text{Upper Confidence Limit} = \overline{X} + (t\text{-multiplier} \times \text{standard error})$$

with the t-multiplier chosen at the appropriate significance level and $N - 1$ degrees of freedom.

12. Confidence interval around a sample proportion, θ, with large samples:

$$\text{Lower Confidence Limit} = \theta - (z\text{-multiplier} \times \text{standard error})$$
$$\text{Upper Confidence Limit} = \theta + (z\text{-multiplier} \times \text{standard error})$$

with z-multiplier chosen for the appropriate significance level.

13. Law of total probability expression of partial identification for the probability (or proportion of times) that event A occurs:

$$p(A) = p(\text{obs}) \times p(A|\text{obs}) + p(\text{miss}) \times p(\text{A}|\text{miss})$$

14. Partial identification lower bound on the probability (or proportion of times) that event A occurs:

$$\text{Lower Bound}[p(A)] = p(\text{obs}) \times p(A|\text{obs}) + p(\text{miss}) \times 0$$

15. Partial identification upper bound on the probability (or proportion of times) that event A occurs:

$$\text{Upper Bound}[p(A)] = p(\text{obs}) \times p(A|\text{obs}) + p(\text{miss}) \times 1$$

16. Difference between 2 proportions (or percentages). Consider the following generic 2×2 contingency table:

	$y = 0$	$y = 1$	Total
$x = 0$	A	B	A+B
$x = 1$	C	D	C+D
Total	A+C	B+D	A+B+C+D

Based on the information in this table, we can calculate $p(y = 1|x = 0)$ as:

$$p(y = 1|x = 0) = \frac{B}{A + B}$$

And, we can calculate $p(y = 1|x = 1)$ as:

$$p(y = 1|x = 0) = \frac{D}{C + D}$$

So, the difference between these 2 probabilities (or proportions or percentages) is:

$$\text{Difference between proportions} = p(y = 1|x = 1) - p(y = 1|x = 0)$$

17. Relative Risk Statistic:

$$\text{Relative Risk Statistic} = \frac{p(y = 1|x = 1)}{p(y = 1|x = 0)}$$

18. Chi-Square Test of Independence for a 2×2 contingency table:

$$\text{Chi Square Test} = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B} + \frac{(O_C - E_C)^2}{E_C} + \frac{(O_D - E_D)^2}{E_D}$$

where O is the observed frequency in each cell, E is the frequency of cases we expect to see in each cell if the two variables are independent of each other. As an example, we calculate O_A by counting the number of cases in cell A. And, we calculate E_A as follows:

$$E_A = \frac{(A+C)\times(A+B)}{A+B+C+D} = \frac{(A+C)\times(A+B)}{N}$$

The critical values of the chi-square test for 1 degree of freedom (note: 2×2 tables always have 1 degree of freedom) is:

Significance Level	Critical Value of Chi-Square
0.10	2.706
0.05	3.841
0.01	6.635
0.001	10.828

You can use the chi-square test of independence as a test for the significance of the difference between two proportions (discussed in 13 above).

19. Yule's Q - Consider the following generic 2×2 contingency table:

	$Y=0$	$Y=1$	Total
$X=0$	A	B	
$X=1$	C	D	
Total			

Then Yule's Q is:

$$Q = \frac{AD - BC}{AD + BC}$$

Interpretation of Yule's Q is: a positive number indicates a positive correlation; a negative number indicates a negative correlation (i.e., when X goes from 0 to 1 it is an increase in X; conversely when X goes from 1 to 0 it is a decrease). A positive correlation means that when X increases, Y also tends to increase. A negative correlation means that when X increases, Y tends to decrease. A correlation of $+1$ or -1 indicates a perfect correlation; a correlation of 0 means the two variables are independent. Interpretation table:

Yule's Q Value	Interpretation
$Q = -1$	Perfect Negative Correlation
$Q \in \{-0.5, -1.0\}$	Strong Negative Correlation
$Q \in \{-0.2, -0.5\}$	Moderate Negative Correlation
$Q \in \{0, -0.2\}$	Weak Negative Correlation
$Q = 0$	Variables are Independent
$Q \in \{0, 0.2\}$	Weak Positive Correlation
$Q \in \{0.2, 0.5\}$	Moderate Positive Correlation
$Q \in \{0.5, 1.0\}$	Strong Positive Correlation
$Q = 1$	Perfect Positive Correlation

20. Approximate Confidence Interval for Yule's Q:

$$Q \pm z\text{-multiplier} \times \sqrt{\frac{(1-Q^2)^2(\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D})}{4}}$$

So, the lower confidence limit is:

$$Q - z\text{-multiplier} \times \sqrt{\frac{(1-Q^2)^2(\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D})}{4}}$$

and the upper confidence limit is:

$$Q + z\text{-multiplier} \times \sqrt{\frac{(1-Q^2)^2(\frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D})}{4}}$$