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:

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

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:

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

Table B.3 The t Distribution

	Level of Significance for a One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for a Two-Tailed Test					
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.206	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Source: TABLE B-3 is adapted with permission from Table III of Fisher and Yates, *Statistical Tables for Biological, Agricultural and Medical Research* (6th ed.). Published by Longman Group UK Ltd., 1974.