APPENDIX A

Selected Formulas

$Range = Max - Min$

$IQR = Q3 - Q1$

Outlier Rule-of-Thumb: $y < Q1 - 1.5 \times IQR$ or $y > Q3 + 1.5 \times IQR$

$$\bar{y} = \frac{\sum y}{n}$$

$$s = \sqrt{\frac{\sum (y - \bar{y})^2}{n - 1}}$$

$$z = \frac{y - \mu}{\sigma} \text{ (model based)}$$

$$z = \frac{y - \bar{y}}{s} \text{ (data based)}$$

$$r = \frac{\sum z_x z_y}{n - 1}$$

$$\hat{y} = b_0 + b_1 x \qquad \text{where } b_1 = \frac{r s_y}{s_x} \text{ and } b_0 = \bar{y} - b_1 \bar{x}$$

$$P(\mathbf{A}) = 1 - P(\mathbf{A}^C)$$

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \cap \mathbf{B})$$

$$P(\mathbf{A} \cap \mathbf{B}) = P(\mathbf{A}) \times P(\mathbf{B}|\mathbf{A})$$

$$P(\mathbf{B}|\mathbf{A}) = \frac{P(\mathbf{A} \cap \mathbf{B})}{P(\mathbf{A})}$$

If **A** and **B** are independent, $P(\mathbf{B}|\mathbf{A}) = P(\mathbf{B})$

$E(X) = \mu = \sum x \cdot P(x)$	$Var(X) = \sigma^2 = \sum (x - \mu)^2 P(x)$
$E(X \pm c) = E(X) \pm c$	$Var(X \pm c) = Var(X)$
$E(aX) = aE(X)$	$Var(aX) = a^2 Var(X)$
$E(X \pm Y) = E(X) \pm E(Y)$	$Var(X \pm Y) = Var(X) + Var(Y)$, if X and Y are independent

Geometric:	$P(x) = q^{x-1} p$	$\mu = \frac{1}{p}$	$\sigma = \sqrt{\frac{q}{p^2}}$
Binomial:	$P(x) = \binom{n}{x} p^x q^{n-x}$	$\mu = np$	$\sigma = \sqrt{npq}$
	$\hat{p} = \frac{x}{n}$	$\mu(\hat{p}) = p$	$SD(\hat{p}) = \sqrt{\frac{pq}{n}}$

Sampling distribution of $\bar{y}$:
(CLT) As n grows, the sampling distribution approaches the Normal model with

$$\mu(\bar{y}) = \mu_y \qquad SD(\bar{y}) = \frac{\sigma}{\sqrt{n}}$$

Inference:

Confidence interval for parameter = ***statistic* ± *critical value* × *SD(statistic)***

$$\text{Test statistic} = \frac{Statistic - Parameter}{SD(statistic)}$$

Parameter	Statistic	*SD*(statistic)	*SE*(statistic)
p	$\hat{p}$	$\sqrt{\frac{pq}{n}}$	$\sqrt{\frac{\hat{p}\hat{q}}{n}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$	$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$
μ	$\bar{y}$	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$
$\mu_1 - \mu_2$	$\bar{y}_1 - \bar{y}_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
μ_d	$\bar{d}$	$\frac{\sigma_d}{\sqrt{n}}$	$\frac{s_d}{\sqrt{n}}$
σ_ε	$s_e = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}}$		
β_1	b_1		$\frac{s_e}{s_x\sqrt{n - 1}}$
$^*\mu_\nu$	$\hat{y}_\nu$		$\sqrt{SE^2(b_1) \cdot (x_\nu - \bar{x})^2 + \frac{s_e^2}{n}}$
$^*y_\nu$	$\hat{y}_\nu$		$\sqrt{SE^2(b_1) \cdot (x_\nu - \bar{x})^2 + \frac{s_e^2}{n} + s_e^2}$

Pooling: For testing difference between proportions: $\hat{p}_{pooled} = \frac{y_1 + y_2}{n_1 + n_2}$

For testing difference between means: $s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$

Substitute these pooled estimates in the respective SE formulas for both groups when assumptions and conditions are met.

Chi-square: $\chi^2 = \sum \frac{(obs - exp)^2}{exp}$