14.1.2: Formulas

Table B4

Basic probability rules	
$P(A \cap B) = P(A \mid B) \cdot P(B)$	Multiplication rule
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	Addition rule
$P(A \cap B) = P(A) \cdot P(B)$ or $P(A \mid B) = P(A)$	Independence test

Hypergeometric distribution formulae	
$nCx = \binom{n}{x} = \frac{n!}{x!(n-x)!}$	Combinatorial equation
$P(x) = \frac{\binom{A}{x}\binom{N-A}{n-x}}{\binom{N}{n}}$	Probability equation
$E(X) = \mu = np$	Mean
$\sigma^2 = \left(\frac{N-n}{N-1}\right) np(q)$	Variance

Binomial distribution formulae	
$P(x) = \frac{n!}{x!(n-x)!} p^x (q)^{n-x}$	Probability density function
$E(X) = \mu = np$	Arithmetic mean
$\sigma^2 = np(q)$	Variance

Geometric distribution formulae			
$P(X = x) = (1-p)^{x-1}(p)$	Probability when x is the first success.	Probability when x is the number of failures before first success	$P(X = x) = (1-p)^x(p)$
$\mu = \frac{1}{p}$	Mean	Mean	$\mu = \frac{1-p}{p}$
$\sigma^2 = \frac{(1-p)}{p^2}$	Variance	Variance	$\sigma^2 = \frac{(1-p)}{p^2}$

Poisson distribution formulae	
$P(x) = \frac{e^{-\mu}\mu^x}{x!}$	Probability equation
$E(X) = \mu$	Mean
$\sigma^2 = \mu$	Variance

Uniform distribution formulae	
$f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$	PDF
$E(X) = \mu = \frac{a+b}{2}$	Mean
$\sigma^2 = \frac{(b-a)^2}{12}$	Variance

Exponential distribution formulae	
$P(X \leq x) = 1 - e^{-mx}$	Cumulative probability
$E(X) = \mu = \frac{1}{m}$ or $m = \frac{1}{\mu}$	Mean and decay factor
$\sigma^2 = \frac{1}{m^2} = \mu^2$	Variance

Table B5

The following page of formulae requires the use of the "Z", "t", "χ^2" or "F" tables.	
$Z = \frac{x-\mu}{\sigma}$	Z-transformation for normal distribution
$Z = \frac{x-np'}{\sqrt{np'(q')}}$	Normal approximation to the binomial
Probability (ignores subscripts) **Hypothesis testing**	**Confidence intervals** [bracketed symbols equal margin of error] (subscripts denote locations on respective distribution tables)
$Z_c = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	*Interval for the population mean when sigma is known* $\bar{x} \pm \left[Z_{(\alpha/2)} \frac{\sigma}{\sqrt{n}}\right]$
$Z_c = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$	*Interval for the population mean when sigma is unknown but $n > 30$* $\bar{x} \pm \left[Z_{(\alpha/2)} \frac{s}{\sqrt{n}}\right]$
$t_c = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$	*Interval for the population mean when sigma is unknown but $n < 30$* $\bar{x} \pm \left[t_{(n-1),(\alpha/2)} \frac{s}{\sqrt{n}}\right]$
$Z_c = \frac{p'-p_0}{\sqrt{\frac{p_0q_0}{n}}}$	*Interval for the population proportion* $p' \pm \left[Z_{(\alpha/2)} \sqrt{\frac{p'q'}{n}}\right]$
$t_c = \frac{\bar{d}-\delta_0}{s_d}$	*Interval for difference between two means with matched pairs* $\bar{d} \pm \left[t_{(n-1),(\alpha/2)} \frac{s_d}{\sqrt{n}}\right]$ *where s_d is the deviation of the differences*
$Z_c = \frac{(\bar{x}_1-\bar{x}_2)-\delta_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	*Interval for difference between two means when sigmas are known* $(\bar{x}_1 - \bar{x}_2) \pm \left[Z_{(\alpha/2)} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$
$t_c = \frac{(\bar{x}_1-\bar{x}_2)-\delta_0}{\sqrt{\left(\frac{(s_1)^2}{n_1}+\frac{(s_2)^2}{n_2}\right)}}$	*Interval for difference between two means with equal variances when sigmas are unknown* $(\bar{x}_1 - \bar{x}_2) \pm \left[t_{df,(\alpha/2)} \sqrt{\left(\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}\right)}\right]$ where $df = \frac{\left(\frac{(s_1)^2}{n_1}+\frac{(s_2)^2}{n_2}\right)^2}{\left(\frac{1}{n_1-1}\right)\left(\frac{(s_1)^2}{n_1}\right)+\left(\frac{1}{n_2-1}\right)\left(\frac{(s_2)^2}{n_2}\right)}$
$Z_c = \frac{(p_1'-p_2')-\delta_0}{\sqrt{\frac{p_1'(q_1')}{n_1}+\frac{p_2'(q_2')}{n_2}}}$	*Interval for difference between two population proportions* $(p_1' - p_2') \pm \left[Z_{(\alpha/2)} \sqrt{\frac{p_1'(q_1')}{n_1} + \frac{p_2'(q_2')}{n_2}}\right]$
$\chi_c^2 = \frac{(n-1)s^2}{\sigma_0^2}$	*Tests for GOF, Independence, and Homogeneity* $\chi_c^2 = \sum \frac{(O-E)^2}{E}$ *where $O =$ observed values and $E =$ expected values*
$F_c = \frac{s_1^2}{s_2^2}$	*Where s_1^2 is the sample variance which is the larger of the two sample variances*

The next 3 formule are for determining sample size with confidence intervals. (note: E represents the margin of error)		
$n = \frac{Z^2\left(\frac{a}{2}\right)^{\sigma^2}}{E^2}$ *Use when sigma is known* $E = \bar{x} - \mu$	$n = \frac{Z^2\left(\frac{a}{2}\right)^{(0.25)}}{E^2}$ *Use when p' is unknown* $E = p' - p$	$n = \frac{Z^2\left(\frac{a}{2}\right)^{[p'(q')]}}{E^2}$ *Use when p'p' is uknown* $E = p' - p$

Table B6

Simple linear regression formulae for $y = a + b(x)$

$r=\frac{\Sigma[(x-\bar{x})(y-\bar{y})]}{\sqrt{\Sigma(x-\bar{x})^2*\Sigma(y-\bar{y})^2}}=\frac{S_{xy}}{S_xS_y}=\sqrt{\frac{SSR}{SST}}$	Correlation coefficient
$b=\frac{\Sigma[(x-\bar{x})(y-\bar{y})]}{\Sigma(x-\bar{x})^2}=\frac{S_{xy}}{SS_x}=r_{y,x}\left(\frac{s_y}{s_x}\right)$	Coefficient b (slope)
$a=\bar{y}-b(\bar{x})$	y-intercept
$s_e^2=\frac{\Sigma(y_i-\hat{y}_i)^2}{n-k}=\frac{\sum_{i=1}^n e_i^2}{n-k}$	Estimate of the error variance
$S_b=\frac{s_e^2}{\sqrt{(x_i-\bar{x})^2}}=\frac{s_e^2}{(n-1)s_x^2}$	Standard error for coefficient b
$t_c=\frac{b-\beta_0}{s_b}$	Hypothesis test for coefficient β
$b\pm\left[t_{n-2,\alpha/2}S_b\right]$	Interval for coefficient β
$\hat{y}\pm\left[t_{\alpha/2}*s_e\left(\sqrt{\frac{1}{n}+\frac{(x_p-\bar{x})^2}{s_x}}\right)\right]$	Interval for expected value of y
$\hat{y}\pm\left[t_{\alpha/2}*s_e\left(\sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{s_x}}\right)\right]$	Prediction interval for an individual y
ANOVA formulae	
$SSR=\sum_{i=1}^n(\hat{y}_i-\bar{y})^2$	Sum of squares regression
$SSE=\sum_{i=1}^n(\hat{y}_i-\bar{y}_i)^2$	Sum of squares error
$SST=\sum_{i=1}^n(y_i-\bar{y})^2$	Sum of squares total
$R^2=\frac{SSR}{SST}$	Coefficient of determination

The following is the breakdown of a one-way ANOVA table for linear regression.

Source of variation	Sum of squares	Degrees of freedom	Mean squares	F-ratio
Regression	SSR	1 or $k-1$	$MSR=\frac{SSR}{df_R}$	$F=\frac{MSR}{MSE}$
Error	SSE	$n-k$	$MSE=\frac{SSE}{df_E}$	
Total	SST	$n-1$		