Chapter 1

1. Descriptive measures for samples:

Mean: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$;(n = sample size) &

Median: $\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{, n is odd} \\ \dfrac{x_{(n/2)} + x_{\left(n/2+1\right)}}{2} & \text{, n is even} \end{cases}$

Standard Deviation:

$$S = \sqrt{\frac{\sum(x - \bar{x})^2}{n-1}} = \sqrt{\frac{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)}{n-1}} = \sqrt{\frac{\sum x^2 - n(\bar{x})^2}{n-1}}$$

Coefficient of variation: $CV = \dfrac{s}{\bar{x}} \times 100\%$

Coefficient of skewness: $CS = \dfrac{\bar{x} - \tilde{x}}{s/3}$

2. Mean & Standard deviation of grouped data:

$$\bar{x} = \frac{\sum_{j=1}^{k} x_j f_j}{\sum_{j=1}^{k} f_j}; \ k = \text{number of groups}$$

$$s = \sqrt{\frac{\sum_{j=1}^{k}\left(x_j - \bar{x}\right)^2 f_j}{\left(\sum_{j=1}^{k} f_j\right) - 1}} = \sqrt{\frac{\sum_{j=1}^{k} x_j^2 f_j - n\bar{x}^2}{\left(\sum_{j=1}^{k} f_j\right) - 1}}$$

Chapter 2

1. Permutations: ${}_nP_r = \dfrac{n!}{(n-r)!}$
2. Combinations: ${}_nC_r = \dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$
3. $P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(A - B) = P(A \cap B') = P(A) - P(A \cap B)$
5. $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, $P(B) > 0$
6. $P(A \cap B) = P(A) \times P(B \mid A) = P(B) \times P(A|B)$
7. A & B indep. $\leftrightarrow P(A \cap B) = P(A) \times P(B)$ or $P(A \mid B) = P(A)$ or $P(B \mid A) = P(B)$
8. Baye's Rule

$$P(B_j \mid A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)} \quad \text{for } j = 1,2,\ldots,k$$

Chapter 3

1. Discrete probability distributions
   cdf of r.v. X is $F(x) = P(X \le x) = \sum_{\forall t \le x} f(t)$
   pdf for r.v. X is $f(x) = F(x) - F(x^-)$
2. Continuous probability distributions

$$P(a < X < b) = \int_a^b f(x)dx, \ f(x) = \frac{d}{dx}F(x) = F'(x)$$

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt, \ -\infty < x < \infty$$

Chapter 4

1. If X is discrete $\Rightarrow \mu = E(X) = \sum_{\forall x} x f(x)$ &
   $V(X) = \sigma^2 = E[(X - \mu)^2] = \sum_{\forall x} (x - \mu)^2 f(x)$
2. If X is continuous $\rightarrow \mu = E(X) = \int_{-\infty}^{+\infty} x f(x)dx$ &
   $V(X) = \sigma^2 = E[(X - \mu)^2] = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)dx$
3. $\sigma^2 = E(X^2) - (E(X))^2 = E(X^2) - E^2(X) = E(X^2) - \mu^2$
   $E(aX+b) = aE(X) + b$ & $V(aX+b) = a^2V(X)$
   $E[ag(X) \pm bh(X)] = aE[g(X)] \pm bE[h(X)]$

Chapter 5

1. **Binomial** distribution: If X:B(n , p) $\Rightarrow$
   $f(x) = b(x; n, p) = \dbinom{n}{x} p^x q^{n-x}; \ x = 0,1,\ldots,n$ &
   $\mu = E(X) = np$, $\sigma^2 = V(X) = npq$
2. **Hypergeometric** distribution: If X:HG(N,n,k) $\Rightarrow$

$$f(x) = h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}; \ \max\{0, n-(N-k)\} \le x \le \min\{k, n\}$$

$$\mu = E(X) = \frac{nk}{N}, \sigma^2 = V(X) = \frac{N-n}{N-1} n \frac{k}{N}\left(1 - \frac{k}{N}\right)$$

If X:HG(N, n , k) such that $\dfrac{n}{N} \le 0.05 \Rightarrow$

X~B($n, \dfrac{k}{N}$)

3. **Geometric** distribution: If X:G(p) $\Rightarrow$ $f(x) = g(x;p) = pq^{x-1}, x = 1,2,...$ &

$\mu_X = E(X) = \frac{1}{p}$ & $\sigma_X^2 = V(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$

4. **Poisson** distribution: If X:P(λt) $\Rightarrow$

$f(x) = p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}$; $x = 0, 1, 2, ...$ &

$\mu_X = E(X) = \lambda t = V(X) = \sigma_X^2$. If X:B(n , p) such that $n \to \infty, p \to 0 \Rightarrow$ If X~P(np)

Chapter 6

1. **Uniform** distribution: If X:U(A , B) $\Rightarrow$

$f(x) = \left\{ \begin{matrix} \frac{1}{B-A} & , A \le x \le B \\ 0 & , e.w. \end{matrix} \right\}$ &

$\mu_X = \frac{A+B}{2}$ & $\sigma_X^2 = \frac{(B-A)^2}{12}$

2. **Normal** distribution: If X:N(μ , σ^2) such that $E(X)= \mu$ & $V(X)= \sigma^2$ $\Rightarrow$ $Z = \frac{X-\mu}{\sigma}$:N(0,1)

If X:B(n,p); $np \ge 5$ & $nq \ge 5 \Rightarrow$ X~N(np,npq)

3. **Gamma, Exponential** & **Chi-squared** distributions:

If X:Γ(α , β) $\Rightarrow$

$f(x) = \left\{ \begin{matrix} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & , x > 0 \\ 0 & , e.w. \end{matrix} \right\}$ where

$\mu_X = E(X) = \alpha\beta$ & $\sigma_X^2 = V(X) = \alpha\beta^2$.

If $\alpha = 1$ then X:Exp(β) with

$f(x) = \left\{ \begin{matrix} \frac{1}{\beta} e^{-x/\beta} & , x > 0 \\ 0 & , e.w. \end{matrix} \right\}$.

If $\alpha = \frac{\nu}{2}$ & $\beta = 2$ then X: χ_ν^2 where ν = df.

Chapter 8

1. $E(\bar{X}) = \mu$ & $V(\bar{X}) = \frac{\sigma^2}{n}$, $\bar{X}$ ~N(μ , $\frac{\sigma^2}{n}$) if $n \ge 30$

2. $E(\bar{X}_1 - \bar{X}_2) = (\mu_1 - \mu_2)$ & $V(\bar{X}_1 - \bar{X}_2) = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

Chapter 9

1. If $n \ge 30$ OR σ is known then:

A (1 – α) 100 % C.I. for μ is $\left[\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$

2. If $n \le 30$ AND σ is unknown then:

A (1 – α) 100 % C.I. for μ is $\left[\bar{X} \pm t_{\alpha/2;n-1} \frac{s}{\sqrt{n}}\right]$

3. $e \le z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ & $n \ge \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2$ with a confidence level of (1 – α)100%

4. For large samples OR known σ's then:

A C.I. for $\mu_1 - \mu_2$ is $\left[(\bar{X}_1 - \bar{X}_2) \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$

5. For small samples AND unknown σ's, a C.I. for $\mu_1 - \mu_2$ is $\left[(\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2};n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$

assuming $\sigma_1^2 = \sigma_2^2$ with

$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$

6. A (1 – α) 100% C.I. for P is $\left[\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}\right]$

$e \le z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$ with conf. level (1 – α)100%.

$n \ge \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2}$ with conf. level (1 – α)100%.

$n = \frac{z_{\alpha/2}^2}{4e^2}$ with conf. level (**at least**) (1 – α)100%.

7. A C.I. for σ^2 is $\left[\frac{(n-1)S^2}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2}\right]$.

Chapter 10

1. $\alpha = P$(Type-I error) = P(Rejecting H_0 | H_0 true).

$\beta = P$(Type-II error) =P(Accepting H_0 | H_0 false).

2. If $n \ge 30$ OR σ is known then

$Z_0 = \frac{\bar{X} - \mu_0}{SE(\bar{X})} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ for testing

$H_0 : \mu = \mu_0$ vs. $H_1 : \mu (>, <, \text{or} \ne) \mu_0$.

3. If $n \le 30$ AND σ is unknown then $T_0 = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$ for testing $H_0: \mu = \mu_0$ vs. $H_1: \mu(>, <, \text{or} \ne)\, \mu_0$.
4. For large samples OR known σ's then $Z_0 = \dfrac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ for testing $H_0: \mu_1 - \mu_2 = D_0$ vs. $H_1: \mu_1 - \mu_2 (>, <, \text{or} \ne)\, D_0$.
5. For small samples AND unknown σ's, $Z_0 = \dfrac{(\bar{X}_1 - \bar{X}_2) - D_0}{s_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ for testing $H_0: \mu_1 - \mu_2 = D_0$ vs. $H_1: \mu_1 - \mu_2 (>, <, \text{or} \ne)\, D_0$.
6. For testing $H_0: p = p_0$ vs. $H_1: p(>, <, \text{or} \ne)\, p_0$ then test statistic is $Z_0 = \dfrac{x - np_0}{\sqrt{np_0q_0}} = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0q_0}{n}}}$.

Chapter 11

1. Sample correlation coefficient
$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}} = \frac{\sum xy - n\overline{xy}}{\sqrt{\left[\sum x^2 - n\bar{x}^2\right]\left[\sum y^2 - n\bar{y}^2\right]}}$$
$$= \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\left[n\sum x^2 - (\sum x)^2\right]\left[n\sum y^2 - (\sum y)^2\right]}} = \frac{s_{xy}}{\sqrt{s_{xx}s_{yy}}}$$
For testing $H_0: \rho = \rho_0$ vs. $H_1: \rho(>, <, \text{or} \ne)\, \rho_0$ statistic
$t_{n-2} = r\big/\sqrt{(1-r^2)/(n-2)}$ has df=n-2.
2. Estimated (fitted) regression model $\hat{y}_i = a + bx$ & the error (residual) $e_i = y_i - \hat{y}_i$
3. The Least Square Estimates are
$$b = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{\sum xy - (\sum x \sum y)/n}{\sum x^2 - (\sum x)^2/n}$$
$$= \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \text{ and } a = \bar{y} - b\bar{x}$$
4. Total, Regression, & Error Sum of Squares
$$SST = \sum(y-\bar{y})^2 = s_{yy},\ SSR = \sum(\hat{y}-\bar{y})^2 = bs_{xy}$$
$$SSE = \sum(y-\hat{y})^2 = SST - SSR = s_{yy} - b^2 s_{xx}$$
5. Coefficient of Determination (in %)
R-squared = $R^2 = \dfrac{SSR}{SST} = b^2\dfrac{s_{xx}}{s_{yy}} = \dfrac{s_{xy}^2}{s_{xx}s_{yy}} = r^2$
$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = r^2$$
6. Mean square error & Standard Error of the estimate
$$MSE = S_e^2 = \frac{SSE}{n-2} = \frac{SST - SSR}{n-2} = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-2} = \frac{s_{yy} - bs_{xy}}{n-2}$$
and $s_e = \sqrt{SSE/(n-2)} = \sqrt{MSE}$
7. Inference about β
A (1-α)100% C.I. for β is $b \pm t_{\alpha/2, n-2}\dfrac{s_e}{\sqrt{s_{xx}}}$
For testing H_0: $\beta = \beta_0$ vs. H_1: β (>, <, or $\ne$) β_0
$\Rightarrow$ test statistic $t_0 = \dfrac{B - \beta_0}{S_e/\sqrt{s_{xx}}}$ has df = n-2.
8. Inference about α
A (1-α)100% C.I. for α is $a \pm t_{\alpha/2, n-2}s_e\sqrt{\dfrac{\sum x^2}{ns_{xx}}}$
For testing H_0: $\alpha = \alpha_0$ vs. H_1: α (>, <, or $\ne$) α_0
$\Rightarrow$ test statistic $t_0 = \dfrac{A - \alpha_0}{S_e\sqrt{\dfrac{\sum x^2}{ns_{xx}}}}$ has df = n-2.
9. A(1 − α)100% C.I. for the **mean response** $\boldsymbol{\mu_{Y|X_\circ}}$ is $\hat{y}_\circ \pm t_{\alpha/2, n-2}s_e\sqrt{\dfrac{1}{n} + \dfrac{(x_\circ - \bar{x})^2}{s_{xx}}}$
10. A(1 − α)100% P.I. for a **single response $\boldsymbol{Y_o}$** is $\hat{y}_\circ \pm t_{\alpha/2, n-2}s_e\sqrt{1 + \dfrac{1}{n} + \dfrac{(x_\circ - \bar{x})^2}{s_{xx}}}$