A. **Descriptive Statistics**

A.2 Average and variance are

$$\bar{x} \equiv \frac{1}{n}\sum x \text{ and } s^2 \equiv \frac{s_{xx}}{n-1},$$

$$s_{xx} \equiv \sum (x-\bar{x})^2 = \sum x^2 - \frac{1}{n}\left(\sum x\right)^2;$$

A.3 Mean and the variance for grouped data:

$$\bar{x} \equiv \frac{1}{n}\sum xf;\ s_{xx} \equiv \sum x^2 f \ - \frac{1}{n}\left(\sum xf\right)^2;$$

$$s^2 \equiv \frac{s_{xx}}{n-1}.$$

where $x's$ are the mid values of each class and the sum is over the number of classes.

A.4 Standardized score of an observation x is $z(x) \equiv (x-\bar{x})/s.$

A.5 Coefficient of Variation : $CV \equiv s/\bar{x}.$

A.6 Coefficient of Skewness :

$CS \equiv 3(\bar{x}-\tilde{x})/s,$ where $\tilde{x}$ is the sample median.

A. **Probability of Set Events (Two Sets A and B)**

B.1a $P(A\cup B) = P(AB') + P(A'B) + P(AB).$

B.1b $P(A\cup B) = P(A) + P(B) - P(AB).$

B.1c $P(A\cup B) = 1 - P(A\cup B)'.$

B.1c $P(A\cup B) = 1 - P(A'B').$

B.2 $P(AB)' = P(A'\cup B').$

B.3 $P(A\mid B) = \dfrac{P(AB)}{P(B)}, \quad P(B) \neq 0;$

$P(AB) = P(A)P(B\mid A) = P(B)P(A\mid B).$

B.4 Independence:

$P(A\mid B') = P(A) = P(A\mid B),$ or,
$P(AB) = P(A)P(B).$

C. Discrete Distributions

C.1 $P(a \le X \le b) = \sum_x f(x);$

$$P(X \le b) = \sum_{x\le b} f(x).$$

C.2 $\mu \equiv E(X) = \sum_x x\, f(x).$

C.3 $E(X^2) = \sum x^2 f(x),$

$\sigma^2 \equiv E(X-\mu)^2 = E(X^2) - \mu^2.$

C.4 The Binomial Distribution $B(n, p)$:

$$f(x) = \binom{n}{x} p^x q^{n-x};$$

$x = 0,1,\ldots,n;\ \ 0 < p < 1;$

$q = 1-p;\ \mu = np,\ \sigma^2 = npq.$

C.5 The Geometric Distribution:

$f(x) = q^x p,\ \ x = 0,1,2,\cdots;\ q = 1-p;$

$\mu = 1/p,\ \sigma^2 = q/p^2.$

C.7 The Hypergeometric Distribution

$$f(x) = \binom{K}{x}\binom{N-K}{n-x} \div \binom{N}{n},$$

$\max\{0, n-(N-K)\} \le x \le \min\{n, K\};$
$\mu = np,$

$\sigma^2 = (1-c)\ npq,$
$(N-1)\ c = n-1,\ p = (K/N),\ q = 1-p.$

C.8 The Poisson Distribution

$$f(x) = \frac{(\lambda t)^x}{x!} e^{-\lambda t};\ x = 0,1,\ldots;\ \mu = \lambda t,$$

$\sigma^2 = \lambda t.$

D. Continuous Distributions

D.1 $P(a < X < b) = \int_a^b f(x)dx;$

$F(k) = \int_{-\infty}^{k} f(x)dx$ where k is a particular value of x.

D.2 $\mu \equiv E(X) = \int_{-\infty}^{\infty} x\, f(x)dx.$

D.3 $E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)dx,$

$\sigma^2 \equiv V(X) = E(X^2) - \mu^2.$

D.4 The Normal Distribution $N(\mu, \sigma^2)$. The 100$(1-\alpha)\%$ percentile of a random variable $Z \sim N(0,1)$ is denoted by ${}_{1-\alpha}z_{\alpha}$ or often by z_{α}.

D.5 The Exponential Distribution:

$f(x) = \dfrac{1}{\beta}\, e^{-x/\beta}, \quad 0 \le x;\ \mu = \beta,$

$\sigma^2 = \beta^2.$

D.6 Waiting Time Distribution:

$f(t) = \lambda e^{-\lambda t}, \quad 0 \le t;\ \mu = 1/\lambda,$

$\sigma^2 = 1/\lambda^2,$

E. Sampling Distributions

E.1 Reproductive Theorem: If $X \sim N(\mu, \sigma^2)$, then $\dfrac{\overline{X} - \mu}{\sqrt{\sigma^2 / n}} = Z$ where $Z \sim N(0,1)$.

E.2 : If $X \sim (\mu, \sigma^2)$ and $30 \le n$, then $\dfrac{\overline{X} - \mu}{\sqrt{\sigma^2 / n}} = Z$ has an approximate (or weak) $N(0,1)$ distribution. This is known as Central Limit Theorem (**CLT**). In case σ^2 is unknown,

$\dfrac{\overline{X} - \mu}{\sqrt{S^2 / n}} = Z$ has an approximate (or weaker) $N(0,1)$ distribution.

E.3 The Student *T*- statistic is defined by $T = \dfrac{\overline{X} - \mu}{\sqrt{S^2 / n}}$, with degrees of freedom $\nu = n - 1.$

E.4 Approximate Sampling Distribution of the Proportion

$$\frac{\hat{p} - p}{\sqrt{pq/n}} = Z;\ \frac{\hat{p} - p}{\sqrt{\hat{p}\hat{q}/n}} = Z.$$

E.5 The binomial event $X = x$ can be inflated to $x - 0.50 \le X \le x + 0.50$ if $4 \le np$ and $4 \le nq.$

F. Statistical Estimation

F.1 Confidence Interval Estimates for the Mean μ

F.1.1 CI for μ, (σ known, any n, normal): $\overline{x} \mp z_{\alpha/2}\left(\dfrac{\sigma}{\sqrt{n}}\right).$

F.1.2 ACI for μ, (σ known, large n, nonnormal): $\overline{x} \mp z_{\alpha/2}\left(\dfrac{\sigma}{\sqrt{n}}\right).$

F.1.2a sample size for estimating μ :

$n = \dfrac{z_{\alpha/2}^2}{e^2}\sigma^2,$ where

$P(|\overline{X} - \mu| \le e) = 1 - \alpha.$

F.1.3 ACI for μ, (σ unknown, large n, nonnormal): $\overline{x} \mp z_{\alpha/2}\left(\dfrac{s}{\sqrt{n}}\right).$

F.1.3a sample size for estimating μ :

$n = \dfrac{z_{\alpha/2}^2}{e^2}s^2,$ where

$P(|\overline{X} - \mu| \le e) = 1 - \alpha.$

F.1.4 CI for μ, (σ unknown, $n \ge 2$, normal): $\overline{x} \mp t_{\alpha/2}\left(\dfrac{s}{\sqrt{n}}\right)$ for any $n \ge 2$.

F.2 Confidence Interval for $\mu_1 - \mu_2$

F.2.1 CI for $\mu_1 - \mu_2$, (σ_i^2 known, any n_i, normal):

$$(\bar{x}_1 - \bar{x}_2) \mp z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

F.2.2 ACI for $\mu_1 - \mu_2$, (σ_i^2 known, large n_i, nonnormal):

$$(\bar{x}_1 - \bar{x}_2) \mp z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

F.2.3 CI for $\mu_1 - \mu_2$,(σ_i^2 unknown, large n_i, nonnormal):

$$(\bar{x}_1 - \bar{x}_2) \mp z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

F.2.4 CI for $\mu_1 - \mu_2$, (Any $n_i \geq 2$, unknown $\sigma_1^2 = \sigma_2^2$ but equal, normal):

$$(\bar{x}_1 - \bar{x}_2) \mp t_{\alpha/2} \sqrt{\frac{s_w^2}{n_1} + \frac{s_w^2}{n_2}},$$

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

$$\nu = (n_1 - 1) + (n_2 - 1).$$

F.3 Confidence Interval for p

F.3.1 CI for p when n large:

$$\hat{p} \mp z_{\alpha/2} \sqrt{\hat{p}\hat{q}/n}.$$

F.3.1 a Large sample size for estimating p: $n = \frac{z_{\alpha/2}^2}{e^2} \hat{p}\hat{q}.$

F.3.2 CI for $p_1 - p_2$ with large sample sizes :

$$\hat{p}_1 - \hat{p}_2 \mp z_{\alpha/2} \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}.$$

G. Testing of Hypotheses

Reject H_0 for $p\text{-}value \leq \alpha$; Don't reject H_0 for $0 \leq \alpha < p\text{-}value$.

G.1 Testing of a Mean μ

G.1.1 σ known, normal: $z = \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}}.$

G.1.2 σ known, large n, nonnormal:

$$z = \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}}.$$

G.1.3 σ unknown, large n, nonnormal: $z = \frac{\bar{y} - \mu_0}{\sqrt{s^2/n}}$

G.1.4 σ unknown, normal population:

$$t = \frac{\bar{x} - \mu_0}{\sqrt{s^2/n}}, \quad (\nu = n - 1 \geq 1).$$

G.2 Testing $\mu_1 - \mu_2 = \delta$

G.2.1 known σ_i, normal:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\left(\sigma_1^2/n_1\right) + \left(\sigma_2^2/n_2\right)}}.$$

G.2.2 known σ_i, large n_i, nonnormal:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\left(\sigma_1^2/n_1\right) + \left(\sigma_2^2/n_2\right)}}.$$

G.2.3 unknown σ_i, large n_i, nonnormal:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\left(s_1^2/n_1\right) + \left(s_2^2/n_2\right)}}.$$

G.2.3 Any $n_i \geq 2$, unknown $\sigma_1^2 = \sigma_2^2$, normal)

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\left(s_w^2/n_1\right) + \left(s_w^2/n_2\right)}},$$

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

$$\nu = (n_1 - 1) + (n_2 - 1),$$

G.3.1 Testing of a proportion

Large sample: $z = \frac{\hat{p} - p_0}{\sqrt{p_0 q_0/n}},$

where $q_0 = 1 - p_0$.

H. Regression Analysis
($\nu = n-2)$

H.1 Line of Best Fit

H.1.0 $y = \beta_0 + \beta_1 x + \varepsilon.$

$Y \sim N(\mu(x),\ \sigma^2),\ \ \mu(x) = \beta_0 + \beta_1 x.$

Estimate: $\hat{y} = \hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1 x.$

H.1.1 $\hat{\beta}_1 = \dfrac{s_{xy}}{s_{xx}}$,

$$s_{xy} = \sum xy - \frac{1}{n}\left(\sum x\right)\left(\sum y\right), ,$$

$$s_{xx} = \sum x^2 - \frac{1}{n}\left(\sum x\right)^2,\ \ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\ \bar{x}\ .$$

H.1.2 Coefficient of correlation:

$$r = \frac{s_{xy}}{\sqrt{s_{xx}\ s_{yy}}}.\ \left(r\sqrt{s_{yy}} = \hat{\beta}_1\sqrt{s_{xx}}\right)$$

H.1.3 $s_{yy} = \sum y^2 - \dfrac{1}{n}\left(\sum y\right)^2,$

$SSR = \hat{\beta}_1 s_{xy},\ SSE = s_{yy} - SSR.$

H.1.4 Estimate of σ^2:

$$s_e^2 = MSE = \frac{SSE}{n-2}.$$

H.1.5 $R^2 = SSR / s_{yy}.$

H.2 Inference for Regression Coefficients

H.2.3 $100(1-\alpha)\%$ confidence interval for β_1 : $\hat{\beta}_1 \mp t_{\alpha/2}\sqrt{MSE / s_{xx}}.$

H.2.4 Testing the hypothesis

$$H_0 : \beta_1 = c :\ t = \frac{\hat{\beta}_1 - c}{\sqrt{MSE / s_{xx}}}.$$

H2.5 Testing the hypothesis $H_0 : \rho = 0$:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}.\qquad \left(\rho\sigma_y = \beta_1\sigma_x\right)$$

H.3 Inference for Response Variable

H.3.1 $100(1-\alpha)\%$ Confidence Interval of $\mu(x)$:

$$\hat{\mu}(x)\ \pm\ t_{\alpha/2}\sqrt{\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{s_{xx}}\right)MSE}.$$

H.3.2 $100(1-\alpha)\%$ Prediction Interval for an Individual Y for a given x:

$$\hat{\mu}(x) \pm\ t_{\alpha/2}\ \sqrt{\left(1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{s_{xx}}\right)MSE}.$$