Sample mean $\overline{x} = \frac{x_1 + \cdots + x_n}{n}$.
Population mean $\mu_X = E(X) = \sum_{x \in \chi}(xp(x))$
$E(g(X)) = \sum_{x \in \chi}(g(x)p(x)) = \int_{-\infty}^{\infty} g(x)f(x)\,dx$
$E(aX + Y + b) = aE(X) + E(Y) + b$

Sample Variance: $s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$
Sample Standard Deviation: s_{n-1}
Population Variance $\sigma_X^2 = \mathrm{Var}(X) = \sum_{x \in \chi}((x-\mu)^2 p(x))$
$\mathrm{Var}(X) = E((X-\mu)^2) = E(X^2) - \mu^2$
$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$
If X and Y are independent, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Chapter 1 Descriptive Statistics

Data:

1. Categorical/ Qualitative
   Nominal (Cannot be ranked), Ordinal (Can be ranked)
2. Quantitative
   Discrete (Counting), Continuous (Measuring)

Sample median $\widetilde{x} = \begin{cases} x_{\frac{x+1}{2}}, & \text{if } n \text{ is odd} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & \text{if } n \text{ is even} \end{cases}$.
Trimmed mean: $\overline{x}_{\mathrm{tr}(10)}$ (Mean after eliminate top and bottom 10%)
Sample range: $x_{max} - x_{min}$ Inter-quartile range (IQR): $Q_3 - Q_1$

Ways of presenting:

1. Line Chart, Pie Chart, Bar Chart
   Histogram (Table with bars)
   Frequency Table (Arranged with columns and rows)
   Boxplot (Gives quartiles and outliers, left line $Q_1 - 1.5$IQR, right line $Q_3 + 1.5$IQR)
   Scatter plot (Data comes in pairs)

Skewness:
Left skewed (Mean < median), Right skewed (Mean > median)
Symmetric (Mean ≈ median)

Chapter 2 Probability

Difference: $A - B = \{s | s \in A \text{ and } s \notin B\} = A \cap B^c$.
Symmetric Difference: $A \Delta B = \{s | s \in A \cup B \text{ and } s \notin A \cap B\}$.
Commutative, associative, distributive
De Morgan's Laws: $(A \cap B)^c - A^c \cup B^c$, $(A \cup B)^c - A^c \cap B^c$
Permutation: Ordered arrangement of set of objects
No. of permutations of n distinct objects taken r once: $\frac{n!}{(n-r)!}$
No. of permutations of n objects arranged in a circle: $(n-1)!$
Combination: Unordered arrangement of selecting r from n: $\binom{n}{r}$
No. of combinations of n distinct objects taken r once: $\frac{n!}{(n-r)!r!}$
No. of ways that n distinct stuff grouped into k classes: $\frac{n!}{n_1! \cdots n_k!}$
If $P(AB) = P(A)P(B)$, A and B are independent.
Independent: not mutually effected. Disjoint: No overlap.
Mutually independent: $P(\bigcap_{k=i}^j A_k) = \prod_{k=i}^j P(A_k)$ for all $i < j$
Pairwise independent: $P(A_i A_j) = P(A_i)P(A_j)$ for all $i < j$

Probability: $0 \le P(E) \le 1$, $P(S) = 1$, $P(E^c) = 1 - P(E)$
Mutually exclusive (Disjoint): $P(\bigcup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i)$
Countable: $A = \bigcup_{k=1}^{\infty}\{a_k\}$ E.g. $\mathbb{N}$
Probability of empty set: $P(\Phi) = 0$
Exhaustive: $E_1 \cup \cdots \cup E_n = S$
Partition: Mutually Exclusive + Exhuastive
$P(A) \le P(B)$ if $A \subseteq B$
$P(A_1 \cup \cdots \cup A_n) = \sum_{j=1}^n (-1)^{j-1} \sum_{i_1 < \cdots i_j} \binom{n}{j} P(A_{i_1} \cdots A_{i_j})$
$P(A|B) = \frac{P(AB)}{P(B)} \ge P(AB) = P(A|B)P(B) = P(B|A)P(A)$
$P(A_1 \cdots A_n) = P(A_n | A_1 \cdots A_{n-1}) \cdots P(A_2|A_1)P(A_1)$
$P((A \cap B)|D) = P(A|(B \cap D))P(B|D)$
If B_i is partition of S, $P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$
Law of total probability: Let S be sample space.
Bayes' Theorem: If B_i is partition, $P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^n P(A|B_i)P(B_i)}$

Chapter 3 Random Variables

Random variable $X : S \to R$: $X(a)$ is assigned to outcome a in S
Probability mass function (pmf): $p(x)$
Cumulative distribution function (cdf): $F(x) = P(X \le x)$
$P(a < X \le b) = F(b) - F(a)$
$F(x)$ is non-decreasing and $0 \le F(x) \le 1$.

Bernoulli $X \sim \mathrm{Binomial}(n, p)$: $E(X) = np$ $\mathrm{Var}(X) = np(1-p)$
For n trials: $p(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, \cdots, n$
Poisson distribution $X \sim \mathrm{Poisson}(\lambda)$: $E(X) = \lambda$ $\mathrm{Var}(X) = \lambda$
Determine probability of counts of occurrence over line
λ is rate of occurrences of event per unit time or space
or average number of occurrences of event per unit time or space
$p(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, \cdots$
Count of occurrences for t units of time with rate λ: $Y_t \sim \mathrm{Poisson}(\lambda t)$
Poisson Limit Theorem: When $\lambda = np$,
$\lim_{n \to \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$
Closely approximates if n is large and p is small
Normal distribution $X \sim N(\mu, \sigma^2)$: $E(X) = \mu$ $\mathrm{Var}(X) = \sigma^2$
$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$
Standard normal distribution $X \sim N(0, 1)$ (Use z-table):
Distribution function $\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt$
If $X \sim N(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$

Discrete r.v.: Finite or countably infinite number of values
Continuous r.v.: Values continuously on an interval
$0 < p(x) \le 1$ for all $x \in \chi$, $p(x) = 0$ for all $x \notin \chi$.
Probability density function (pdf): $f(x)$
$f(x) > 0$ for all $x \in \chi$, and $f(x) = 0$ for all $x \notin \chi$
$\sum_{x \in \chi} p(x) = \int_{-\infty}^{\infty} f(x)\,dx = 1$
$P(X \in A) = \sum_A p(x) = \int_A f(x)\,dx$
$P(X = k) = 0$ for any k for continuous r.v..
$F(a) = P(X \le a) = \sum_{x \le a} p(x) = \int_{-\infty}^{a} f(x)\,dx$
$P(a < X \le b) = F(b) - F(a) = \int_a^b f(x)\,dx$
Chebystev's Inequality: For any $t > 0$: $P(|X - \mu| \ge t) \le \frac{\sigma^2}{t^2}$
Let $t = k\sigma$, $P(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}$
r-th moment about origin of X: $E(X^r)$ for $r \in \mathbb{N}$
r-th central moment: $E((X - E(X))^r)$ for $r \in \mathbb{N}$
Skewness: $E(Z^3) = \frac{\mu_3}{\sigma^3}$ where μ_3 is 3rd central moment
+ve skewness: Right long tail -ve skewness: Left long tail
Kurtosis: $E(Z^4) = \frac{\mu_4}{\sigma^4}$ where μ_4 is 4th central moment
Excess kurtosis = Kurtosis −3
0 excess kurtosis: Normal distribution
+ve ex. kurtosis: Thicker tails -ve ex. kurtosis: Thinner tails
Moment generation function (mgf): $M_X(t) = E(e^{tX})$
$M_X^k(0) = \frac{d^k}{dt^k} M_X(t)\big|_{t=0} = E(X^k)$
For Binomial distribution: $M_X(t) = E(e^{tX}) = (pe^t + 1 - p)^n$
For Normal distribution: $M_Z(t) = E(e^{tZ}) = e^{\frac{t^2}{2}}$

Chapter 4 Parameter Estimation

Unknown population: An unknown distribution of r.v. X
Sample: Collection of data of X
Parameter: μ_X, σ_X^2 Statistic: $\overline{x}, s_{n-1}^2, s_n^2$
Estimator: $X_1, \cdots, X_n$ Sample mean: $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$
k-th sample moment about origin: $\overline{X^k}$ Estimate: $x_1, \cdots, x_n$
If $E(X) = \mu$, $\text{Var}(X) = \sigma^2$,
$E(\overline{X}) = \mu$, $\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$, $E(s_{n-1}^2) = \sigma^2$
Point estimator of θ: $Y = T(X_1, \cdots, X_n)$ (estimate θ)
Point estimate of θ: $y = T(x_1, \cdots, x_n)$
cdf of X: $X \sim F(x;\theta)$
Method of moment estimator (MME): $(\hat{\theta}_1, \cdots, \hat{\theta}_m)$ of $(\theta_1, \cdots, \theta_m)$
$\hat{\theta}_i = g_i(\overline{X}, \cdots, \overline{X^m}, \cdots)$ where $\overline{X^k} = \frac{1}{n}\sum_{i=1}^n X_i^k$
When n is large, $\overline{X^k} \approx E(X^k)$.
E.g. MME of λ when $X \sim \text{Poisson}(\lambda)$ is $\hat{\lambda} = \overline{X}$.
If $E(\hat{\theta}) = \theta$, then $\hat{\theta}$ is unbiased estimator of θ.
Bias of $\hat{\theta}$: $b_n(\hat{\theta}) = E(\hat{\theta}) - \theta$.
If unbiased only at ∞, it is asymptotic.
If $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$ and $\text{Var}(\hat{\theta}_1) \le \text{Var}(\hat{\theta}_2)$, $\hat{\theta}_1$ is more efficient.

Point estimators cannot provide precision and reliability.
Range may be more meaningful.
R.v.: Random interval Numerical: Confidence interval
Given $T_1 \le T_2$, high $T_2 - T_1$ have high reliability, low precision
$P(T_1 \le \theta \le T_2) \ge 1 - \alpha$. $[T_1, T_2]$ is $1 - \alpha$ confidence interval.
$P(\mu_X - k_1 \le \overline{X} \le \mu_X + k_2) = P(\overline{X} - k_2 \le \mu_X \le \overline{X} + k_1) = 1 - \alpha$
If $X \sim N(0, 1)$ and $Y = X_1^2 + \cdots + X_n^2$, then $Y \sim \chi^2(n)$.
If $Z \sim N(0, 1)$, $Y \sim \chi^2(n)$, $W = \frac{Z}{\sqrt{Y/n}} \sim t(n)$.
If $X \sim N(\mu_X, \sigma_X^2)$, $\overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{n})$, $\frac{(n-1)S_{n-1}^2}{\sigma_X^2} \sim \chi^2(n-1)$
$\frac{\overline{X}-\mu_X}{S_{n-1}/\sqrt{n}} \sim t(n-1)$.
μ_X (known σ_X^2): $P(-z_{\frac{\alpha}{2}} \le \frac{\overline{X}-\mu_X}{\sigma_X/\sqrt{n}} \le z_{\frac{\alpha}{2}}) = 1 - \alpha$
μ_X with $1 - \alpha$ C.I.: $\left[\overline{X} - z_{\frac{\alpha}{2}}\sigma_X/\sqrt{n}, \overline{X} + z_{\frac{\alpha}{2}}\sigma_X/\sqrt{n}\right]$
μ_X (unknown σ_X^2): $P(-t_{n-1,\frac{\alpha}{2}} \le \frac{\overline{X}-\mu_X}{S_{n-1}/\sqrt{n}} \le t_{n-1,\frac{\alpha}{2}}) = 1 - \alpha$
σ_X^2 (unknown μ_X): $P(\chi^2_{n-1,1-\frac{\alpha}{2}} \le \frac{(n-1)S_{n-1}^2}{\sigma_X^2} \le \chi^2_{n-1,\frac{\alpha}{2}}) = 1 - \alpha$
σ_X^2 (known μ_X): $P(\chi^2_{n,1-\frac{\alpha}{2}} \le \sum_{i=1}^n \left(\frac{X_i-\mu_X}{\sigma_X}\right)^2 \le \chi^2_{n,\frac{\alpha}{2}}) = 1 - \alpha$

Chapter 5 Hypothesis Testing

Null hypothesis H_0: Tested to reject or not (With = sign)
Alternative hypothesis H_1: Accept if reject H_0.
$\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 \text{ if } H_0 \text{ is true})$
$\beta = P(\text{Type II error}) = P(\text{Not reject } H_0 \text{ if } H_0 \text{ is false})$
Critical value: c where $\overline{X}$ is a rare event under H_0 (Reject H_0)
Power of test statement: $1 - \beta = 1 - P(\overline{X} \le c \text{ if } \mu_X = \mu_2)$
At significance level a:
One-sided right [left] test: Let $H_0 : \sigma_X^2 = \sigma_0^2, H_1 : \sigma_X^2 > [<]\sigma_0^2$
Critical value (unknown μ_X): Reject if $\frac{(n-1)s_{n-1}^2}{\sigma_0^2} > [<]\chi^2_{n-1,a[1-a]}$
p-value (unknown μ_X): Reject if $P(U_{n-1} > [<]\frac{(n-1)s_{n-1}^2}{\sigma_0^2}) < a$
Two-sided test: Let $H_0 : \sigma_X^2 = \sigma_0^2, H_1 : \sigma_X^2 \ne \sigma_0^2$
Critical value (unknown μ_X):
Reject if $\frac{(n-1)s_{n-1}^2}{\sigma_0^2} < \chi^2_{n-1,1-\frac{a}{2}}$ or $\frac{(n-1)s_{n-1}^2}{\sigma_0^2} > \chi^2_{n-1,\frac{a}{2}}$
p-value (unknown μ_X):
Reject if $2\min(P(U_{n-1} < \frac{(n-1)s_{n-1}^2}{\sigma_0^2}), P(U_{n-1} > \frac{(n-1)s_{n-1}^2}{\sigma_0^2})) < a$

Simple test: Let $H_0 : \mu_X = \mu_1, H_1 : \mu_X = \mu_2$ for $\mu_1 < \mu_2$
$\alpha = P(\overline{X} > c \text{ if } \mu_X = \mu_1) = P(\frac{\overline{X}-\mu_1}{\sigma_X/\sqrt{n}} > \frac{c-\mu_1}{\sigma_X/\sqrt{n}} = z_a)$
Critical value: Reject H_0 if $\overline{x} > c$
p-value $= P(\overline{X} > \overline{x} \text{ if } \mu_X = \mu_1)$ Reject if p-value $< a$
One-sided right [left] test: Let $H_0 : \mu_X = \mu_0, H_1 : \mu_X > [<]\mu_0$
Critical value (known σ_X^2): Reject if $\overline{x} > [<]\mu_0 + [-]z_a\frac{\sigma_X}{\sqrt{n}}$
p-value (known σ_X^2): Reject if $P(Z > [<]\frac{\overline{x}-\mu_0}{\sigma_X/\sqrt{n}}) < a$.
t value (unknown σ_X^2): Reject if $\overline{x} > [<]\mu_0 + [-]t_{n-1,a}\frac{s_{n-1}}{\sqrt{n}}$
p-value (unknown σ_X^2): Reject if $P(T_{n-1} > [<]\frac{\overline{x}-\mu_0}{s_{n-1}/\sqrt{n}}) < a$
Two-sided test: Let $H_0 : \mu_X = \mu_0, H_1 : \mu_X \ne \mu_0$
Critical value (known σ_X^2): Reject if $\left|\frac{\overline{x}-\mu_0}{\sigma_X/\sqrt{n}}\right| > z_{\frac{\alpha}{2}}$
p-value (known σ_X^2): Reject if $2P(Z > \left|\frac{\overline{x}-\mu_0}{\sigma_X/\sqrt{n}}\right|) < a$
t value (unknown σ_X^2): Reject if $\left|\frac{\overline{x}-\mu_0}{s_{n-1}/\sqrt{n}}\right| > t_{n-1,\frac{a}{2}}$
p-value (unknown σ_X^2): Reject if $2P(T_{n-1} > \left|\frac{\overline{x}-\mu_0}{s_{n-1}/\sqrt{n}}\right|) < a$

Chapter 6 Simple linear regression model and Least squares

Scatter plot: Collection of paired data of x and y
Model: $Y = \beta_0 + \beta_1 x + \epsilon$ Regression coefficients: β_0, β_1
Least square method: $S(u, v) = \sum_{i=1}^n (y_i - (u + vx_i))$
Finding minimum of S at (a, b): $a = \overline{y} - b\overline{x}$
$b = \frac{\sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)/n}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2/n} = \frac{\sum_{i=1}^n (x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^n (x_i-\overline{x})^2} = \frac{S_{XY}}{S_{XX}}$
Fitted regression line: $\hat{y} = a + bx$ Substitute x_i, $e_i = y_i - \hat{y}_i$
Sum of Squared Errors (SSE) $= \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Pearson's correlation coefficient $r = \frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$
Population correlation coefficient $\rho = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X\sigma_Y}$
$0 < \rho < 1$: Positively correlated (Slope is +ve)
$-1 < \rho < 0$: Negatively correlated (Slope is -ve)
$\rho = 0$: Uncorrelated
$z_{\text{Fisher}} = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$. $\mu = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$ and $\sigma = \frac{1}{n-3}$
Regression Sum of Squares (RSS) $= \sum_{i=1}^n (\hat{y}_i - \overline{y})^2$
Total variability of response (SST) $= \sum_{i=1}^n (y_i - \overline{y})^2 = \text{RSS} + \text{SSE}$
Variation due to regression model and variation due to error
R-squared: $R^2 = \frac{\text{RSS}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$
$\text{Var}(\epsilon) = \sigma^2$ $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i-\overline{x})(Y_i-\overline{Y})}{S_{XX}}$, $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{x}$
$100(1 - \alpha)\%$ prediction interval for y_{new}
$= \hat{y}_{new} \pm t_{n-2,\frac{\alpha}{2}} s\sqrt{1 + \frac{1}{n} + \frac{(x_{new}-\overline{x})^2}{S_{XX}}}$

Assume ϵ are independent and normally distributed.
Assume $\text{Var}(\epsilon) = \sigma^2, E(\epsilon) = 0$
$\epsilon_i \sim N(0, \sigma^2)$, $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{XX}}\right)$, $\hat{\beta}_0 \sim N\left(\beta_0, \frac{\sigma^2\sum_{i=1}^n x_i^2}{nS_{XX}}\right)$
They are unbiased estimators. $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i-\overline{x})Y_i}{S_{XX}}$
$E(\hat{\beta}_0) = \frac{\sum_{i=1}^n (x_i-\overline{x})E(Y_i)}{S_{XX}} = \frac{\sum_{i=1}^n (x_i-\overline{x})(\beta_0+\beta_1 x_i)}{S_{XX}} = \beta_1$
$\text{Var}(\hat{\beta}_1) = \frac{\sum_{i=1}^n (x_i-\overline{x})^2\,\text{Var}(Y_i)}{(S_{XX})^2} = \frac{\sigma^2}{S_{XX}}$
Residual $e_i = y_i = \hat{y}_i$ is actual value of ϵ_i
Mean Square Error (MSE) $S^2 = \frac{\sum_{i=1}^n E_i^2}{n-2} = \frac{\sum_{i=1}^n (Y_i-\hat{Y}_i)^2}{n-2}$
MSE is unbiased estimator of σ^2. $s^2 = \frac{\text{SSE}}{n-2} = \frac{S_{YY}-bS_{XY}}{n-2}$
Replacing unknown σ^2 by MSE, $T_{n-2} = \frac{\hat{\beta}_1-\beta_1}{\sqrt{S^2/S_{XX}}} \sim t(n-2)$
$100(1 - \alpha)\%$ C.I. is $b \pm t_{n-2}\cdot\frac{\alpha}{2}\sqrt{s^2/S_{XX}}$
Replacing unknown σ^2 by MSE, $T_{n-2} = \frac{\hat{\beta}_0-\beta_0}{\sqrt{S^2\sum_{i=1}^n x_i^2/nS_{XX}}}$
$100(1 - \alpha)\%$ C.I. is $(\overline{y} - b\overline{x}) \pm t_{n-2,\frac{\alpha}{2}}\sqrt{s^2\sum_{i=1}^n x_i^2/nS_{XX}}$
One-sided right test: $H_0 : \beta_1 = b_1, H_1 : \beta_1 > b_1$
t value: $\frac{b-b_1}{s/\sqrt{S_{XX}}} > t_{n-2,\alpha}$ p-value: $P(T_{n-2} > \frac{b-b_1}{s/\sqrt{S_{XX}}}) < \alpha$
One-sided right test: $H_0 : \beta_0 = b_0.H_1 : \beta_0 > b_0$
t value: $\frac{a-b_0}{s\sqrt{\sum_{i=1}^n x_i^2/nS_{XX}}} > t_{n-2,\alpha}$
p-value: $P(T_{n-2} > \frac{a-b_0}{s\sqrt{\sum_{i=1}^n x_i^2/nS_{XX}}}) < \alpha$