chapter 1

formulas:

- Relative frequency: $\frac{f}{\sum f}$
- Percent frequency: $r.f \times 100$
- Measure angle: $r.f \times 360$
- IQR: $Q_3 - Q_1$
- # of classes: $K = \lfloor 3.322 \log n \rfloor$
- Range: $x_l - x_s$
- class limit: R + one measurement unit / K
- Mid point: upper limit + lower limit / 2

coordinates:

- bar chart: (class, f_i)
- histogram: (b_i, f_i)
- polygon: (x_m, f_i)
- less than ogive → ACF: (upper b_i, F_i)
- greater than ogive → DCF: (lower b_i, Φ_i)

central Tendencies:

1] Mean $\bar{x}$:

- for raw data: $\frac{\sum x_i}{n}$
- frequency: $\frac{\sum x_i f_i}{\sum f}$
- distribution: $\frac{\sum x_m f_i}{\sum f_i}$
- weighted: $\frac{\sum w_i x}{\sum w_i}$

2] Median $\tilde{x}$:

- raw data: $x_{n+1/2}$ (odd), $\frac{x_{n/2} + x_{n/2+1}}{2}$ (even)
- frequency: $\sum f/2$ (even), $\sum f+1/2$ (odd)
- distribution: $\tilde{l}_b + \left(\frac{\sum f/2 - (\tilde{F} - \tilde{f})}{\tilde{f}}\right) c$

3) Mode $\hat{x}$:

- raw & frequency: data with the higist frequency.
- distribution: $\hat{L}_b + \left(\frac{d_1}{d_1+d_2}\right) c$

dispersion:

- $q_r = \frac{r(n+1)}{4}$
- $d_r = \frac{r(n+1)}{10}$
- $P_r = \frac{r(n+1)}{100}$
- $Q, D, P = x_k + S(x_{k+1} - x_k)$
- lower fence: $LF = Q_1 - 1.5(Q_3 - Q_1)$
- higher fence: $HF = Q_3 + 1.5(Q_3 - Q_1)$

1) Variance S^2:

- raw: $\frac{\Sigma(x-\bar{x})^2}{n-1}$
- frequency: $\frac{\Sigma f(x-\bar{x})^2}{\Sigma f - 1}$
- distribution: $\frac{\Sigma f(x_n-\bar{x})^2}{\Sigma f - 1}$

2) Standard deviation S:

- $S = \sqrt[+]{var}$
- coefficient of variation: $CV = \frac{S}{\bar{x}} \times 100$
- Z-Score: $Z_{x_n} = \frac{x_i - \bar{x}}{S}$

chapter 2

- factorial notation:

$n! = n(n-1)(n-2)\ldots 1$

- permutation:

$^nP_r = \frac{n!}{(n-r)!}$

- combination:

$\binom{n}{r} = \frac{n!}{(n-r)!\,r!}$

- circular permutation:

$(n-1)!$

- exactly one of an event:

$A \Delta B = (A \setminus B) \cup (B \setminus A)$

- laplace's principle:

$P(A) = \frac{|A|}{|\Omega|}$; if ω are equally likely

- compliment:

$P(\bar{A}) = 1 - P(A)$

- difference:

$P(A \setminus B) = P(A) - P(A \cap B)$

- Probability:

$P(A) = \sum_i P(\{\omega_i\}) ; \forall \omega \in A$

- De-morgan's law:

$P(\bar{A} \cup \bar{B}) = P(\overline{A \cap B})$

$P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B})$

- additive rule:

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

$P(A \cup B) = P(A) + P(B)$ → mutually exclusive events

- conditional probability:

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

$P(A \mid B) = P(A)$ → Independent

- multiplicative rule:

$P(A \cap B) = P(B)\,P(A \mid B)$

$P(A \cap B) = P(A)\,P(B \mid A)$

Independent → $P(A \cap B) = P(A)\,P(B)$

- total probability:

$\sum P(Z_i)\,P(B \mid Z_i) = P(B)$

- Baye's Theorem:

$$P(Z_i \mid B) = \frac{P(Z_i)\,P(B \mid Z_i)}{\sum P(Z_i)\,P(B \mid Z_i)}$$

examples:

- regular experiment:

analysis of pure water with electric current.

- random experiment:

tossing a coin three times.

- finite countable space:

rolling a die once. $\Omega = \{1,2,3,4,5,6\}$

- uncountable space:

choosing a number randomly from $\mathbb{R}$

- casting stones in a well randomly.

- infinite countable space:

rolling a die until a six is obtained.

chapter three

- definition of a random variable:

$$X^{-1}(x) = \{\omega_i\} \in \mathcal{A}, \forall x \in \mathbb{R}$$

- distribution function:

$$F_X(x) = P(X \leq x)$$

i. $0 \leq F_X(x) \leq 1$

ii. $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$

iii. $P(a < X < b) = F_X(b) - F_X(a)$

- probability mass function of a discrete variable:

i. $P(X = x) \geq 0$

ii. $\Sigma P(X = x) = 1$

- distribution function of a d.r.v:

$$F_X(x) = \Sigma P(X \leq x)$$

- mean of a discrete random variable:

$$\mu = E(X) = \Sigma x_i \, P(X = x_i)$$

- variance and standard deviation:

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

$$\sigma = \pm\sqrt{\sigma^2}$$

Scanned by CamScanner

- probability density function for a c.r.v:

$$f_X(x) = \int_{-\infty}^{\infty} f(x)\,dx$$

i. $f_X(x) \geq 0$

ii. $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx = 1$

iii. $P(X = x) = 0$ since its continuous

- distribution function:

$$F_X(x) = \int_{-\infty}^{x} f(x)\,dx$$

- mean of a continuous random variable:

$$\mu = \int_{-\infty}^{\infty} x f(x)\,dx$$

- variance and standard deviation:

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$

$$\sigma = \sqrt[+]{\sigma^2}$$

- derivatives:

$$\frac{d}{dx} e^{f(x)} = f'(x)\, e^{f(x)}$$

$$\frac{d}{dx} uv = u'v + uv'$$

$$\frac{d}{dx} \frac{u}{v} = \frac{u'v - uv'}{v^2}$$

Special Distributions:

i. Discrete:

- Binomial: $X \sim B(n,p)$

$P_X(x) = \binom{n}{k} p^k (1-p)^{n-k}$

$\mu = np$ $\sigma^2 = npq$ $\sigma = +\sqrt{\sigma^2}$

- Geometric: $X \sim G(p)$

$P_X(x) = pq^{k-1}$

- Poissons: $X \sim P_o(\lambda)$

$\mu = \lambda$ $\sigma^2 = \lambda$ $\sigma = +\sqrt{\sigma^2}$

- Uniform: $X \sim DU(k)$

ii. Continuous:

- Exponential: $X \sim E_p(\lambda)$

$f_X(x) = \lambda e^{-\lambda x}$ $\mu = \frac{1}{\lambda}$

$F_X(x) = 1 - e^{-\lambda x}$ $\sigma^2 = \frac{1}{\lambda^2}$ $\sigma = \frac{1}{\lambda}$

- Normal: $X \sim (\mu, \sigma)$

$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

- Standard: $Z \sim (0,1)$

$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2}$

Proofs:

1] $var(x)$ is expressed as $E(X^2)-(E(x))^2$:

$E[(X-E(X))^2] \rightarrow E[X^2-2XE(X)+(E(X))^2]$

$\rightarrow E(X^2)-2\overbrace{E(X)E(X)}^{(E(X))^2}+(E(X))^2$

$\rightarrow E(X^2)-(E(X))^2$ } Steiner's formula

2] $\sum_{0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}=1$:

$e^{-\lambda}\sum_{0}^{\infty}\frac{\lambda^k}{k!}$ } $\rightarrow [1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}\ldots]$ Taylor's expansion

$\rightarrow$ based on Taylor's expansion we have $\sum_{0}^{\infty}\frac{\lambda^k}{k!}=e^{\lambda}$

Thus, $e^{-\lambda}\sum_{0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=e^0=1$.

Examples:

- Discrete random variable: number of heads in 5 tosses.
- Continuous random variable: life expectancy of a lamp.
- Binomial Dist.: number of correct guesses in a true or false quiz.
- Poisson Dist.: number of cars during rush hour.
- Normal Dist.: IQ measures.
- Expon. Dist.: life length of an electric device.

chapter four

	Mean	Proportion
Centoral limit theorem	$\mu_{\bar{x}} = \mu$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$\mu_p = p$ $\sigma_p = \sqrt{p(1-p)/n}$
Z-Score	$\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$	$\frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}}$
Confidence interval	$\mu \in \bar{x} \mp Z_{1-\alpha/2}\, \sigma/\sqrt{n}$	$p \in \hat{p} \mp Z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n}$
Margin error	$\delta_\mu = Z_{1-\alpha/2}\, \sigma/\sqrt{n}$	$\delta_p = Z_{1-\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n}$
Sample Size	$n = (\sigma Z_{1-\alpha/2} / \delta_\mu)^2$	$n = \underbrace{\hat{p}(1-\hat{p})}_{0.25} (Z_{1-\frac{\alpha}{2}} / \delta_p)^2$
Length		$n = (Z_{1-\frac{\alpha}{2}} / L)^2, L = 2\delta_p$

Hypothesis testing for the population mean and proportion:

1st identify the null and alternate hypothesises:

$H_0 : \mu = \mu_0$

$$H_1 : \begin{cases} \mu \neq \mu_0 \\ \mu > \mu_0 \\ \mu < \mu_0 \end{cases}$$

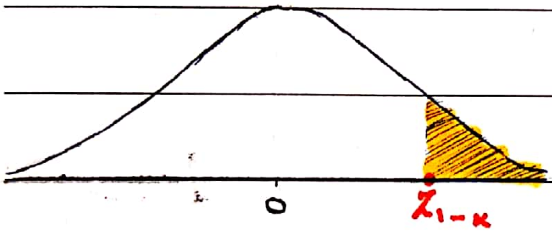
2nd test the statistic:

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \qquad\qquad Z_0 = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

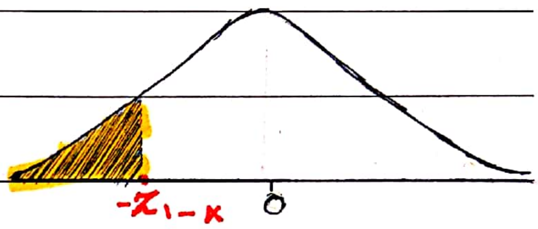
3rd critical region or P-value:

i. critical region:

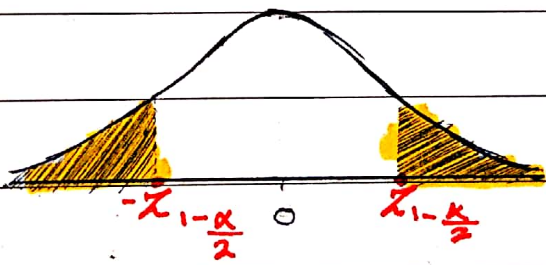
a) $H_1: \mu > \mu_0$



b) $H_1: \mu < \mu_0$



c) $H_1: \mu \neq \mu_0$



ii. p-value:

a) $H_1: \mu > \mu_0$ $\quad P(Z > Z_0)$

b) $H_1: \mu < \mu_0$ $\quad P(Z < -|Z_0|)$

c) $H_1: \mu \neq \mu_0$ $\quad 2P(Z > |Z_0|)$

4th decision:

we reject H_0 only if it falls in the critical region or α is greater than or equal to the p-value.

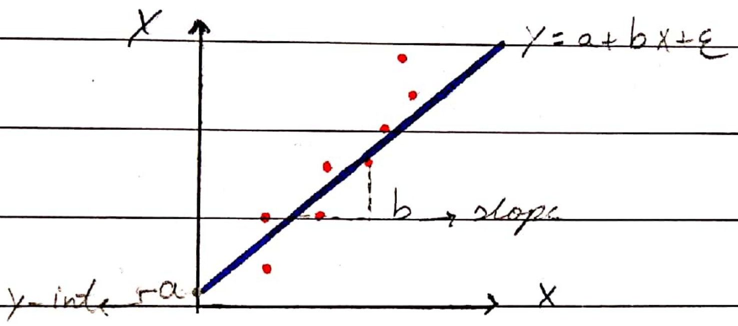
chapter five

Pearsons coorelation coefficient:

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\cdot\sum(y_i-\bar{y})^2}} \quad , \quad r = \frac{n\sum x_i y_i - \sum x_i \cdot \sum y_i}{\sqrt{n\sum x_i^2-(\sum x_i)^2}\sqrt{n\sum y_i^2-(\sum y_i)^2}}$$

Simple linear regression line of a population:

$$Y = a + bX + \varepsilon$$



The estimated regression line of a sample:

$$\hat{y} = \hat{a} + \hat{b}X$$

$$\hat{b} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} \quad , \quad \hat{b} = \frac{n\sum x_i y_i - \sum x_i \cdot \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$\hat{a} = \hat{y} = \bar{y} - \hat{b}\bar{x} \quad , \quad \hat{a} = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

Coefficient of determination:

$$r^2 = SSR / SS_{tot} \quad , \quad r^2 = 1 - SSE / SS_{tot}$$

Total sum of squared deviations (Total var): $SS_{tot} = \sum(y_i-\bar{y})^2$

Sum of squared regression error (exp. var): $SSR = \sum(\bar{y}-\hat{y}_i)^2$

Sum of squared error / residuals (unexp. var): $SSE = \sum(y_i-\hat{y})^2$

$$SS_{tot} = SSR + SSE$$