DESCRIPTIVE STATISTICS

	Individual data	Aggregated data

MEAN	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	$\overline{x} = \frac{\sum_{i=1}^{k} {}_c x_i f(x_i)}{\sum_{i=1}^{k} f(x_i)}$ k = # class
		i = 1,…, k
		$f(x_i)$ = absolute freq.

VARIANCE	$s^2 = \frac{1}{n-1}\sum_{1=1}^{n}(x_i - \overline{x})^2$	$s^2 = \frac{1}{n-1}\sum_{1=1}^{k}({}_c x_i - \overline{x})^2 f(x_i)$
	$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}$	$s^2 = \frac{\sum_{i=1}^{k} x_i^2 f(x_i) - \frac{\left(\sum_{i=1}^{k} x_i f(x_i)\right)^2}{n}}{n-1}$

COEFFIENT OF VARIATION $CV = \frac{s}{\overline{x}} \cdot 100$

PROBABILITY

P(A or B) = P(A) + P(B) – P(A and B)

P(A and B) = P(A)P(B|A)

$$P(A|B) = \frac{P(A\ and\ B)}{P(B)}$$

DIAGNOSTIC TEST

Sn = Pr(T+|D+)

Sp = Pr(T-|D-)

BAYES THEOREM

$$PPV = \frac{P(D+)P(T+|D+)}{P(D+)P(T+|D+)+P(D-)P(T+|D-)}$$

$$NPV = \frac{P(D-)P(T-|D-)}{P(D-)P(T-|D-)+P(D+)P(T-|D+)}$$

BINOMIAL DISTRIBUTION

$$P(x|n) = \frac{n!}{x! \cdot (n-x)!} \cdot \pi^x \cdot (1-\pi)^{n-x} =$$

$$= \binom{n}{x} \cdot \pi^x \cdot (1-\pi)^{n-x}$$

GAUSSIAN DISTRIBUTION

Given $X \sim N(\mu, \sigma^2)$, can be standardised in $Z \sim N(0,1)$ by:

$$z = \frac{(x-\mu)}{\sigma}$$

SAMPLE SIZE

…needed to get a confidence interval (1-α) of a proportion with length 2E

$$n = \frac{z_{\alpha/2}^2 p(1-p)}{E^2}$$

…needed to get a confidence interval (1-α) of a mean with length 2E

$$n = \left[\frac{z_{\alpha/2}\sigma}{E}\right]^2$$

…needed to compare two means

$$n = 2(z_{\alpha/2} + z_{\beta})^2 \frac{\sigma^2}{\Delta^2}$$

TEST t

one sample

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

$$t_{n-1} = \frac{\overline{x} - \mu_0}{ES(\overline{x})} = \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$$

$$IC : \overline{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

two samples

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

Assuming equal variances

$$t_{n_1+n_2-2} = \frac{\overline{x}_1 - \overline{x}_2}{ES(\overline{x}_1 - \overline{x}_2)} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_p^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

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

$$IC : (\overline{x}_1 - \overline{x}_2) \pm t_{n_1+n_2-2,\alpha/2} \sqrt{s_p^2 \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}$$

Not assuming equal variances

$$t_{\text{MIN}(n_1-1,n_2-1)} = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$IC_{1-\alpha}: (\overline{x_1} - \overline{x_2}) \pm t_{df,\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$df = \text{MIN}(n_1 - 1, n_2 - 1)$$

TEST on PROPORTION

one sample

$$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi \neq \pi_0 \end{cases}$$

$$z = \frac{p - \pi_0}{ES(p)} = \frac{p - \pi_0}{\sqrt{\pi_0 (1 - \pi_0)/n}}$$

$$IC : p \pm z_{\alpha/2} \sqrt{p(1-p)/n}$$

two samples

$$\begin{cases} H_0 : \pi_1 = \pi_2 \\ H_1 : \pi_2 \neq \pi_2 \end{cases}$$

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}}$$

$$p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2} = \frac{f_1 + f_2}{n_1 + n_2}$$

$$\text{C.I.95\%:} p_1 - p_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$