DESCRIPTIVE STATISTICS

Median:

n= odd	n= even
$Me = X_{\frac{n+1}{2}}$	$Me = \dfrac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

Mean:

Population	Sample
$\mu = \dfrac{\sum_{i=1}^{N} X_i}{N}$	$\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

Central value: Central value = $(X_{max}+X_{min})/2$

Skewness	Kurtosis
$M_3 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{n}$	$\alpha_4 = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^4}{S^4} - 3$

Amplitude (Range): $R = X_{max} - X_{min}$

Mean of deviation:

From the mean	From the median
$R_{\overline{X}} = \dfrac{\sum_{i=1}^{n} \lvert X_i - \overline{X} \rvert}{n}$	$R_{Me} = \dfrac{\sum_{i=1}^{n} \lvert X_i - Me \rvert}{n}$

Variance:

Population	Sample
$\sigma^2 = \dfrac{SS}{N} = \dfrac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N}$	$s^2 = \dfrac{SS}{n-1} = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

Standard deviation:

$$s = \sqrt{s^2} = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

Standard error: $ES = \dfrac{s}{\sqrt{n}}$

Coefficient of variance: $CV = \dfrac{s}{\overline{X}}$

PROBABILITIES

Pr(A) = (number of favourable cases)/(number of possible cases)

Sensibility: Se = Pr(B|A)

where Pr(B) = probability of a positive test; Pr(nonB) = probability of a no positive test; Pr(A) = probability of existence of a certain disease; Pr(nonA) = probability of non existence of a certain disease.

Specificity: Sp = Pr(nonB|nonA)

Positive predictive value: PPV = Pr(A|B)

Negative predictive value: NPV = Pr(nonA|nonB)

False positive rate: RFP = Pr(B|nonA)

False negative rate: RFN = Pr(nonA|B)

Addition rules: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

Addition rule for mutually exclusive events:

$Pr(A \cup B) = Pr(A) + Pr(B)$

Multiplication rules: $Pr(A \cap B) = Pr(A) \cdot Pr(B|A)$

Multiplication rule for independent events:

$Pr(A \cap B) = Pr(A) \cdot Pr(B)$

BAYES Formula:

$$Pr(A|B) = \frac{Pr(B|A) \cdot Pr(A)}{Pr(B|A) \cdot Pr(A) + Pr(B|\overline{A}) \cdot Pr(\overline{A})}$$

RANDOM VARIABLE

Mean = expected value:

$$M(X) = \sum_{i=1}^{n} X_i \cdot Pr(X_i)$$

Variance:

$$V(X) = \sum_{i=1}^{n}(X_i - M(X))^2 \cdot Pr(X_i)$$

Standard deviation:

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\sum_{i=1}^{n}(X_i - M(X))^2 \cdot Pr(X_i)}$$

$$\chi^2 = \sum_{i=1}^{r \cdot c} \frac{(f_i^0 - f_i^t)^2}{f_i^t}$$

BINOMIAL RANDOM VARIABLES

Binomial DISTRIBUTION:

$$Pr(X = k) = C_n^k p^k q^{n-k} \text{ where } C_n^k = \frac{n!}{k! \cdot (n-k)!}$$

Mean: $M(X) = n \cdot p$

Variance: $V(X) = n \cdot p \cdot q$

Standard: $\sigma(X) = \sqrt{V(X)} = \sqrt{n \cdot p \cdot q}$

POISSON RANDOM VARIABLE

$$X : \begin{pmatrix} k \\ e^{-\theta} \cdot \dfrac{\theta^k}{k!} \end{pmatrix} \quad Pr(X = k) = \frac{e^{-\theta} \cdot \theta^k}{k!}$$

e = 2.718281828; $\theta = n \cdot p$ (n = sample size, p = probability of the apparition of an event).

CONFIDENCE LEVELS

Mean: $\left[\overline{X} - Z_\alpha \dfrac{s}{\sqrt{n}}, \overline{X} + Z_\alpha \dfrac{s}{\sqrt{n}}\right]$

Frequency: $\left[f - Z_\alpha \sqrt{\dfrac{f(1-f)}{n}}; f + Z_\alpha \sqrt{\dfrac{f(1-f)}{n}}\right]$

PEARSON CORRELATION COEFICIENT

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 \sum (Y - \overline{Y})^2}}$$

CHI-SQUARE TEST (χ^2)

$$\chi^2 = \sum_{i=1}^{r \cdot c} \frac{(f_i^0 - f_i^t)^2}{f_i^t}$$

where χ^2 = χ^2 test statistic; f_i^o = observed frequency; f_i^t = theoretical frequency.

Critical region for α = 0.05 is $[3.84, \infty)$.

- If $\chi^2 \in [3.84; \infty)$ H_0 is not accepted with a type I risk of error (α).
- If $\chi^2 \notin [3.84; \infty)$ H_0 is accepted with a type II risk of error (β).

Z TEST FOR PROPORTIONS

Comparison of an observed frequency with a theoretical frequency

$$z = \frac{p - \pi}{\sqrt{\dfrac{\pi(1-\pi)}{n}}}$$

where π = theoretical frequency (population frequency); p = observed frequency, n = sample size.

Testing the equality of two proportions

$$z = \frac{(p_1 - p_2)}{\sqrt{p(1-p)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}, \quad p = \frac{p_1 n_1 + p_2 n_2}{n_1 + n_2}$$

where p_1 = frequency for 1st sample; n_1 = sample size of 1st sample; p_2 = frequency of 2nd sample; n_2 = sample size for 2nd sample.

RISKS AND ODDS ON 2×2 CONTINGENCY TABLE

	Disease+	Disease -	Total
Test +	TP	FP	= TP+FP
Test -	FN	TN	= FN+TN
Total	= TP+FN	=FP+TN	= TP+FP+FN+TN = n

where TP = true positive, TN = true negative, FP = false positive, FN = false negative

Name	Formula
False positive rate	=FP/(FP+TN)
False negative rate	=FN/(FN+TP)
Sensibility	=TP/(TP+FN)
Specificity	=TN/(TN+FP)
Accuracy	=(TP+TN)/n
Predictive positive value	=TP/(TP+FP)
Predictive negative value	=TN/(TN+FN)
Relative risk	=TP(FP+TN)/FN(TP+FP)
Odds ratio	=(TP·TN)/(FN·FP)
Attributable risk	=TP/(TP+FP)-FN/(FN+TN)

Z TEST FOR COMPARING OF A SAMPLE MEAN WITH A POPULATION MEAN (EQUAL VARIANCES)

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

where μ = population mean; $\overline{X}$ = sample mean; σ = population standard deviation; n = sample size

- **Critical region for α=** 0.05 (two-tailed test):

$(-\infty, -1.96] \cup [1.96, \infty)$.

STUDENT (T) TEST FOR COMPARING A MEAN WITH A KNOWN MEAN (UNKNOWN VARIANCES)

$$t = \frac{\overline{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

where μ = population mean; $\overline{X}$ = sample mean; s = sample standard deviation; n = sample size.

- **Degree of freedom (df):** df = n-1
- **Critical region** for α = 0.05 (two-tailed test): $(-\infty, -t_{n-1,0.025}] \cup [t_{n-1,0.025}, +\infty)$

STUDENT (T) TEST FOR COMPARING TWO MEANS (EQUAL VARIANCES)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

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

$\overline{X}_1$ = mean of 1st sample; n_1 = sample size of 1st sample; s_1^2 = variance of 1st sample; $\overline{X}_2$ = mean of second sample; n_2 = sample size of 2nd sample; s_2^2 = variance of 2nd sample.

STUDENT (T) TEST FOR COMPARING TWO PAIRED MEANS

$$t = \frac{\overline{d}}{\frac{s}{\sqrt{n}}}$$, Where $$\overline{d} = \frac{(d_1 + d_2 + ... + d_n)}{n}$$; d_i = difference between first and second/last determination (*1 ≤ i ≤ n*); s = standard deviation of differences; n = sample size

Z TEST FOR COMPARIND THE MEANS OF TWO POPULATIONS (KNOWN AND UNEQUAL VARIANCES)

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\overline{X}_1$ = mean of 1st sample; n_1 = sample size of first sample; s_1^2 = variance of first sample; $\overline{X}_2$ = mean of 2nd sample; n_2 = sample size of 2nd sample; s_2^2 = variance of 2nd sample.

SAMPLE SIZE ESTIMATION

Comparison of a mean with population mean (normal distributed data)

$$n = \frac{(z_{1-\alpha} - z_{1-\beta})^2 \sigma^2}{(m - \mu)^2}$$

Critical value for bilateral test: $z_{1-5\%}$ = 1.960, $z_{1-\beta(\beta=20\%)}$ = 0.842

Comparison of two means (data normal distributed)

$$n_1 = n_2 = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 (\sigma_1^2 + \sigma_2^2)}{d^2}$$

Means (data are not normal distributed)

$$n = \frac{\sigma^2}{\alpha k^2}$$

where k = the difference to be identified (clinical value choose arbitrary)

Means: there is no objective a priori of experiment

$$n = \frac{(z_{1-\alpha} - z_{1-\beta})^2 \sigma^2}{(m - \mu)^2}$$

Proportions

Central proportion (the value is near to 0.5):

$$n = \left[\frac{(z_{1-\alpha/2}\sqrt{\pi(1-\pi)} + z_{1-\beta}\sqrt{p(1-p)}}{p - \pi}\right]^2$$

where π = theoretical proportion; p = desired proportion

Extreme proportion (values near 0 or 1):

$$n = \left[\frac{(z_{1-\alpha/2}\sqrt{\pi} + z_{1-\beta}\sqrt{p}}{p - \pi}\right]^2$$