List of statistical formulas with description

Data sampling & sorting

$f_i = \frac{n_i}{n} \cdot 100[\%]$	relative frequency
$N_1 = n_1; N_2 = N_1 + n_2; N_3 = N_2 + n_3; ...$	cumulative (absolute) frequency
$F_1 = f_1; F_2 = F_1 + f_2; F_3 = F_2 + f_3; ...$	cumulative relative frequency
$h = \frac{(\max - \min)}{m}$	width of interval (class); grouped frequency distribution
$m = \sqrt{n}$	number of intervals (classes); grouped frequency distribution

Descriptive statistics

$\mu = \frac{\sum x_i}{N}$ (simple) or $\mu = \frac{\sum x_i \cdot n_i}{N}$ (weighted)	population mean
$\bar{x} = \frac{\sum x_i}{n}$ (simple) or $\bar{x} = \frac{\sum x_i \cdot n_i}{n}$ (weighted)	sample mean
IQR = Q3 – Q1	interquartile range
R = max – min	range
$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$ (simple) or $\sigma^2 = \frac{\sum (x_i - \mu)^2 \cdot n_i}{N}$ (weighted)	population variance
$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ (simple) or $s^2 = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{n-1}$ (weighted)	sample variance
$\sigma = \sqrt{\sigma^2}$	population standard deviation

$s = \sqrt{s^2}$	sample standard deviation
$CV_s = \frac{s}{\bar{x}} \cdot 100[\%]$	coefficient of variation
$\gamma_1 = \frac{\sum (x_i - \bar{x})^3}{s^3 \cdot n}$ (simple) or $\gamma_1 = \frac{\sum (x_i - \bar{x})^3 \cdot n_i}{s^3 \cdot n}$ (weighted)	coefficient of skewness
$\gamma_2 = \frac{\sum (x_i - \bar{x})^4}{s^4 \cdot n} - 3$ (simple) or $\gamma_2 = \frac{\sum (x_i - \bar{x})^4 \cdot n_i}{s^4 \cdot n} - 3$ (weighted)	coefficient of kurtosis

Theory of probability

no formulas

Point and interval estimate

$est\, \mu = \bar{x}$	point estimate of mean
$est\ \sigma^2 = s_1^2$	point estimate of variance
$est\ \sigma = s_1$	point estimate of standard deviation
$P(\bar{x} - \Delta < \mu < \bar{x} + \Delta) = 1 - \alpha$	interval estimate of mean
$\Delta = u_{(1-\alpha/2)} \cdot \frac{s_1}{\sqrt{n}}$	sampling error if n > 30
$\Delta = t_{(\alpha;n-1)} \cdot \frac{s_1}{\sqrt{n}}$	sampling error if n <= 30
$P\left(\frac{(n-1) \cdot s_1^2}{\chi^2_{(\alpha/2;n-1)}} < \sigma^2 < \frac{(n-1) \cdot s_1^2}{\chi^2_{(1-\alpha/2;n-1)}}\right) = 1 - \alpha$	interval estimate of variance
$P\left(\sqrt{\frac{(n-1) \cdot s_1^2}{\chi^2_{(\alpha/2;n-1)}}} < \sigma^2 < \sqrt{\frac{(n-1) \cdot s_1^2}{\chi^2_{(1-\alpha/2;n-1)}}}\right) = 1 - \alpha$	interval estimate of standard deviation

$$n = u^2_{(1-\alpha/2)} \cdot \frac{s_1^2}{\Delta^2}$$

sample size

Hypothesis testing (tests for mean)

$$u = \frac{\bar{x} - \mu_0}{\frac{s_1}{\sqrt{n}}}$$

hypothesis test about a population mean if n > 30

$$t = \frac{\bar{x} - \mu_0}{\frac{s_1}{\sqrt{n}}}$$

hypothesis test about a population mean if n <= 30

$$u = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{n_2 s_{11}^2 + n_1 s_{12}^2}{n_1 . n_2}}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_2 s_{11}^2 + n_1 s_{12}^2}{n_1 . n_2}}}$$

hypothesis test about the difference between means of two populations (independent samples) if n > 30

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(n_1 - 1)s_{11}^2 + (n_2 - 1)s_{12}^2}{n_1 + n_2 - 2}}} . \sqrt{\frac{n_1 . n_2}{n_1 + n_2}}$$

hypothesis test about the difference between means of two populations (independent samples) if n <= 30

$$t = \frac{\bar{d}}{\sqrt{\frac{\sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2}{n \cdot (n-1)}}}$$

hypothesis test about the difference between means of two populations (matched samples)

Hypothesis testing (tests for variance)

$$\chi^2 = \frac{(n-1) \cdot s_1^2}{\sigma_0^2}$$

hypothesis test about a population variance

$$F = \frac{s_{11}^2}{s_{12}^2}$$

hypothesis test about variances of two populations

Analysis of variance (ANOVA)

no formulas

Chi-square test for independence

$$\chi^2_{(\alpha;(m-1)\cdot(k-1))} = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{((a_i b_j) - (a_i b_j)_0)^2}{(a_i b_j)_0}$$

chi-square test for independence

$$\chi^2_{(\alpha;(m-1)\cdot(k-1))} = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{(O - E)^2}{E}$$

$$(a_i b_j)_0 = \frac{a_i \cdot b_j}{n}$$

expected frequencies

$$V = \sqrt{\frac{\chi^2}{n(\min((m,k)-1)}}$$

Cramer's V coefficient

$$C = \sqrt{\frac{\chi^2}{n+\chi^2}}$$

Pearson's C coefficient

$$T^2 = \frac{\chi^2}{n(\sqrt{(m-1)(k-1)}}$$

Chuprov's T^2 coefficient

Regression and correlation analysis

$$t_{(\alpha;n-2)} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

test the significance of a correlation coefficient in simple regression