

数据的概括性度量

名称	公式
中位数	$M_e = \begin{cases} x_{(\frac{n+1}{2})} & n \text{为奇数} \\ \frac{1}{2} \left\{ x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right\} & n \text{为偶数} \end{cases}$
简单样本平均数	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
加权样本平均数	$\bar{x} = \frac{\sum_{i=1}^k M_i f_i}{n}$
几何平均数	$G_m = \sqrt[n]{x_1 \times x_2 \times \cdots \times x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$
异众比率	$V_r = \frac{\sum f_i - f_m}{\sum f_i} = 1 - \frac{f_m}{\sum f_i}$
四分位差	$Q_d = Q_U - Q_L$
极差	$R = \max(x_i) - \min(x_i)$
简单平均差	$M_d = \frac{\sum_{i=1}^n x_i - \bar{x} }{n}$
加权平均差	$M_d = \frac{\sum_{i=1}^k M_i - \bar{x} f_i}{n}$
简单样本方差	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
简单样本标准方差	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$
加权样本方差	$s^2 = \frac{\sum_{i=1}^k (M_i - \bar{x})^2 f_i}{n-1}$
加权样本标准差	$s = \sqrt{\frac{\sum_{i=1}^k (M_i - \bar{x})^2 f_i}{n-1}}$

标准分数	$z_i = \frac{x_i - \bar{x}}{s}$
离散系数	$v_s = \frac{s}{\bar{x}}$
未分组数据的偏态系数	$SK = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$
分组数据的偏态系数	$SK = \frac{\sum_{i=1}^k (M_i - \bar{x})^3 f_i}{ns^3}$
未分组数据的峰态系数	$K = \frac{n(n+1) \sum (x_i - \bar{x})^4 - 3 \left(\sum (x_i - \bar{x})^2 \right)^2 (n-1)}{(n-1)(n-2)(n-3)s^4}$
分组数据的峰态系数	$K = \frac{\sum_{i=1}^k (M_i - \bar{x})^4 f_i}{ns^4} - 3$

概率与概率分布

名称	公式
概率的古典定义	$P(A) = \frac{\text{事件} A \text{所包含的基本事件个数}}{\text{样本空间所包含的基本事件个数}}$ $= \frac{m}{n}$
概率的统计定义	$P(A) = \frac{m}{n} = p$
两个互斥事件之和的概率	$P(A \cup B) = P(A) + P(B)$
n 个两两互斥事件 A_1, A_2, \dots, A_n 之和的概率	$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$
事件 A 与其逆事件 \bar{A} 之和的概率	$P(A) + P(\bar{A}) = 1$
两个任意事件之和的概率	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
概率的乘法公式	$P(AB) = P(B)P(A B)$
两个独立事件之积的概率	$P(AB) = P(A)P(B)$
n 个相互独立事件 A_1, A_2, \dots, A_n 之积的概率	$P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n)$
全概率公式	$P(B) = \sum_{i=1}^n P(A_i)P(B A_i)$
逆概率公式	$P(A_i B) = \frac{P(A_i)P(B A_i)}{\sum_{j=1}^n P(A_j)P(B A_j)}$
离散型随机变量的期望值	$E(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i$
离散型随机变量的方差	$\sigma^2 = D(X) = \sum_{i=1}^{\infty} [x_i - E(X)]^2 \cdot p_i$
二项分布的概率	$P\{X = x\} = C_n^x p^x q^{n-x}$
二项分布的期望值	$E(X) = np$
二项分布的方差	$E(X) = npq$
泊松分布的概率	$P(X) = \frac{\lambda^x e^{-\lambda}}{x!}$
连续型随机变量的期望值	$E(X) = \int_{-\infty}^{+\infty} x f(x) d(x)$
连续型随机变量的方差	$D(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f(x) d(x) = \sigma^2$
正态分布的概率密度函数	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

标准正态分布的概率密度函数	$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
标准正态分布的分布函数	$\Phi(x) = \int_{-\infty}^x \varphi(t) d(t) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$
标准化公式	$Z = \frac{X - \mu}{\sigma}$
正态随机变量的 $a \leq X \leq b$ 概率	$P = (a \leq X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$

抽样分布

名称	公式
\bar{X} 抽样分布的期望值	$E(\bar{X}) = \mu$
\bar{X} 抽样分布的方差	$D(\bar{X}) = \frac{\sigma^2}{n}$
两个样本均值之差抽样分布的期望值	$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$
两个样本均值之差抽样分布的方差	$D(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_1^2}{n_2}$

参数估计

名称		公式
一个总体参数的区间估计	总体均值的置信区间（正态总体， σ 已知）	$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$
	总体均值的置信区间（ σ 未知，大样本）	$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$
	总体均值的置信区间（正态总体， σ 未知，小样本）	$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$
	总体比例的置信区间	$p \pm z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}$
	总体方差的置信区间	$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$
两个总体参数的区间估计	均值之差的置信区间：独立大样本， σ_1^2 和 σ_2^2 已知	$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
	均值之差的置信区间：独立大样本， σ_1^2 和 σ_2^2 未知	$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
	均值之差的置信区间：独立小样本， σ_1^2 和 σ_2^2 已知	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(n_1 + n_2 - 2) \sqrt{s_p^2 \frac{1}{n_1} + \frac{1}{n_2}}$
	均值之差的置信区间：独立小样本， σ_1^2 和 σ_2^2 未知但相等	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(n_1 + n_2 - 2) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
	均值之差的置信区间：独立小样本， σ_1^2 和 σ_2^2 未知且不相等，两个样本的容量不相等	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}(v) \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
	均值之差的置信区间：匹配大样本	$\bar{d} \pm z_{\alpha/2} \frac{\sigma_d}{\sqrt{n}}$
	均值之差的置信区间：匹配小样本	$\bar{d} \pm t_{\alpha/2}(n-1) \frac{s_d}{\sqrt{n}}$
	两个总体比例之差的区间估计	$(p_1 - p_2) \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$
	两个总体方差比的置信区间	$\frac{s_1^2/s_2^2}{F_{\alpha/2}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2/s_2^2}{F_{1-\alpha/2}}$
样本量的确定	估计总体均值时的样本容量	$n = \frac{(z_{\alpha/2})^2 \sigma^2}{E^2}$
	估计总体比例时的样本容量	$n = \frac{(z_{\alpha/2})^2 \cdot \pi(1-\pi)}{E^2}$

假设检验

名称	公式
总体均值检验的统计量（正态总体， σ 已知）	$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
总体均值检验的统计量（ σ 未知，大样本）	$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$
总体均值检验的统计量（正态总体， σ 未知，小样本）	$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$
总体比例检验的统计量	$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$
总体方差检验的统计量	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
两个总体均值之差检验的统计量（ σ_1^2, σ_2^2 已知）	$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
两个总体均值之差检验的统计量（ σ_1^2, σ_2^2 未知但相等，小样本）	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$
两个总体均值之差检验的统计量（ σ_1^2, σ_2^2 未知且不相等，小样本）	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(f)$
两个样本比例之差检验的统计量（检验两个总体比例相等的假设）	$z = \frac{(p_1 - p_2)}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$
两个样本比例之差检验的统计量（检验两个总体比例之差不为零的假设）	$z = \frac{(p_1 - p_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$
两个样本方差比检验的统计量	$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$

列联分析

名称	公式
χ^2 统计量	$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$
φ 统计量	$\varphi = \sqrt{\chi^2 / n}$
列联相关系数	$c = \sqrt{\frac{\chi^2}{\chi^2 + n}}$
ν 相关系数	$\nu = \sqrt{\frac{\chi^2}{n \times \min[(R - 1), (C - 1)]}}$

方差分析

名称	公式
组间方差	$MSA = \frac{\text{组间平方和}}{\text{自由度}} = \frac{SSA}{k-1}$
组内方差	$MSE = \frac{\text{组内平方和}}{\text{自由度}} = \frac{SSE}{n-k}$
方差分析的检验统计量	$F = \frac{MSA}{MSE} \sim F(k-1, n-k)$
关系强度的测量	$R^2 = \frac{SSA(\text{组间SS})}{SST(\text{总SS})}$
多重比较的 LSD	$LSD = t_{\alpha/2} \sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})}$

相关与线性回归分析

名称	公式
相关系数	$r = \frac{n \sum xy - \sum x \times \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \times \sqrt{n \sum y^2 - (\sum y)^2}}$
相关系数检验的统计量	$t = r \sqrt{\frac{n-2}{1-r^2}} \sim t(n-2)$
一元线性回归模型	$y = \beta_0 + \beta_1 x + \varepsilon$
估计的一元线性回归模型	$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
一元线性回归方程的截距	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
多元线性回归模型	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$
估计的多元线性回归模型	$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$
回归方程的斜率（回归系数）	$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$
判定系数	$R^2 = \frac{SSR}{SST} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$
一元回归估计标准误差	$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{SSE}{n-2}} = \sqrt{MSE}$
多元回归估计标准误差	$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-k-1}} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{MSE}$
一元回归线性关系检验的统计量	$F = \frac{SSR/1}{SSE/(n-2)} = \frac{MSR}{MSE} \sim F(1, n-2)$
多元回归线性关系检验的统计量	$F = \frac{SSR/k}{SSE/(n-k-1)} \sim F(k, n-k-1)$
一元回归系数检验的统计的统计量	$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}}$
多元回归系数检验的统计的统计量	$t_i = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \sim t(n-k-1)$
y 的平均值的置信区间	$\hat{y}_0 \pm t_{\alpha/2} s_e \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$

y 的个别值的预测区间	$\hat{y}_0 \pm t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$
残差	$e_i = y_i - \hat{y}_i$
标准化残差	$z_{e_i} = \frac{e_i}{s_e} = \frac{y_i - \hat{y}_i}{s_e}$
修正的多重判定系数	$R_a^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k-1}$

时间序列分析和预测

名称	公式
环比增长率	$G_i = \frac{Y_i}{Y_{i-1}} - 1$
定基增长率	$G_i = \frac{Y_i - Y_0}{Y_0} = \frac{Y_i}{Y_0} - 1$
平均增长率	$\bar{G} = \sqrt[n]{\frac{Y_1}{Y_0} \times \frac{Y_2}{Y_1} \times \cdots \times \frac{Y_n}{Y_{n-1}}} - 1 = \sqrt[n]{\frac{Y_n}{Y_0}} - 1$
年度化增长率	$G_A = \left(\frac{Y_i}{Y_{i-1}}\right)^{m/n} - 1$
平均预测误差	$ME = \frac{\sum_{i=1}^n (Y_i - F_i)}{n}$
平均绝对预测误差	$MAD = \frac{\sum Y_i - F_i }{n}$
均方预测误差	$MSE = \frac{\sum_{i=1}^n (Y_i - F_i)^2}{n}$
平均百分比预测误差	$MPE = \frac{\sum \left(\frac{Y_i - F_i}{Y_i} \times 100 \right)}{n}$
简单平均法预测	$F_{t+1} = \frac{1}{t}(Y_1 + Y_2 + \cdots + Y_t) = \frac{1}{t} \sum_{i=1}^t Y_i$
移动平均法预测	$F_{t+1} = \bar{Y}_t = \frac{Y_{t-k+1} + Y_{t-k+2} + \cdots + Y_{t-1} + Y_t}{k}$
指数平滑法预测	$F_{t+1} = \alpha Y_t + (1 - \alpha) F_t$
线性趋势方程的截距和斜率	$\begin{cases} b = \frac{n \sum tY - \sum t \sum Y}{n \sum t^2 - (\sum t)^2} \\ a = \bar{Y} - b\bar{t} \end{cases}$
二次曲线的标准方程组	$\begin{cases} \sum Y = na + b \sum t + c \sum t^2 \\ \sum tY = a \sum t + b \sum t^2 + c \sum t^3 \\ \sum t^2 Y = a \sum t^2 + b \sum t^3 + c \sum t^4 \end{cases}$
指数曲线的标准方程组	$\begin{cases} \sum \lg Y = n \lg b_0 + \lg b_1 \sum t \\ \sum t \lg Y = \lg b_0 \sum t + \lg b_1 \sum t^2 \end{cases}$

修正指数曲线的未知数	$\begin{cases} b_1 = \left(\frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ b_0 = (S_2 - S_1) \frac{b_1 - 1}{b_1(b_1^m - 1)^2} \\ K = \frac{1}{m} \left(S_1 - \frac{b_0 b_1 (b_1^m - 1)}{b_1 - 1} \right) \end{cases}$
Gompertz 曲线的待定系数	$\begin{cases} b_1 = \left(\frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ \lg b_0 = (S_2 - S_1) \frac{b_1 - 1}{b_1(b_1^m - 1)^2} \\ \lg K = \frac{1}{m} \left(S_1 - \frac{b_1(b_1^m - 1)}{b_1 - 1} \times \lg b_0 \right) \end{cases}$
Logistic 曲线待定系数	$\begin{cases} b_1 = \left(\frac{S_3 - S_2}{S_2 - S_1} \right)^{\frac{1}{m}} \\ b_0 = (S_2 - S_1) \frac{b_1 - 1}{b_1(b_1^m - 1)^2} \\ K = \frac{1}{m} \left(S_1 - \frac{b_0 b_1 (b_1^m - 1)}{b_1 - 1} \right) \end{cases}$
k 阶曲线方程	$\hat{Y}_t = b_0 + b_1 t + b_2 t^2 + \cdots + b_k t^k$

指数

名称	公式
加权综合销售量指数	$I_q = \frac{\sum q_1 p}{\sum q_0 p}$
加权综合价格指数	$I_p = \frac{\sum q p_1}{\sum q p_0}$
拉式数量指标指数	$I_q = \frac{\sum q_1 p_0}{\sum q_0 p_0}$
拉式质量指标指数	$I_p = \frac{\sum q_0 p_1}{\sum q_0 p_0}$
帕式数量指标指数	$I_q = \frac{\sum q_1 p_1}{\sum q_0 p_1}$
帕式质量指标指数	$I_p = \frac{\sum q_1 p_1}{\sum q_1 p_0}$
基期总量加权的加权平均数量指标	$A_q = \frac{\sum \frac{q_1}{q_0} q_0 p_0}{\sum q_0 p_0}$
基期总量加权的加权平均质量指标	$A_p = \frac{\sum \frac{p_1}{p_0} q_0 p_0}{\sum q_0 p_0}$
报告期总量加权的加权平均数量指标	$H_q = \frac{\sum q_1 p_1}{\sum \frac{q_0}{q_1} q_1 p_1}$
报告期总量加权的加权平均质量指标	$H_p = \frac{\sum q_1 p_1}{\sum \frac{p_0}{p_1} q_1 p_1}$
总量指数	$v = \frac{\sum p_1 q_1}{\sum p_0 q_0}$
总量指数体系	$\frac{\sum q_1 p_1}{\sum q_0 p_0} = \frac{\sum q_1 p_0}{\sum q_0 p_0} \times \frac{\sum q_1 p_1}{\sum q_1 p_0}$