Variance

$Var(Y) = E[Y^2] - E[Y]^2$
$= E[(Y - E[Y])^2]$

Population : $s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$

coeff. of variation $v = \frac{\sigma}{\bar{x}}$

Measurements

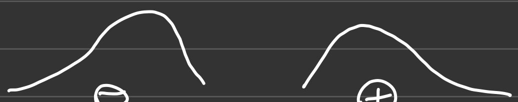
Quantitative

- Ratio scale (abs 0)
- Interval scale

Qualitative

- Ordinal
- Nominal

Skewedness

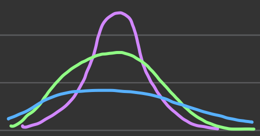


$\ominus$ $\beta_1 < 0$ $\oplus$ $\beta_1 > 0$

$\beta_1 = \frac{\mu_3}{(\mu_2)^{3/2}}$

$b_1 = \frac{m_3}{(m_2)^{3/2}}$

Kurtosis



leptokurtic $\beta_2 > 3$
mesokurtic $\beta_2 = 3$
platykurtic $\beta_2 < 3$

$\beta_2 = \frac{\mu_4}{(\mu_2)^2}$ spiky if $\beta_2 > 3$

Quantiles

$p_i = x_{\left(\frac{i(n+1)}{100}\right)}$ $i = 1..99$

Trimmed mean

- remove $\lfloor \alpha n \rfloor$ on left and right
- $w = (\lfloor \alpha n \rfloor + 1) - \alpha n$

number : $\lfloor \alpha n \rfloor$ 1 $n - 2\alpha n - 2$ 1 $\lfloor \alpha n \rfloor$

weight : 0 w 1 w 0

Mode

- not defined if more than 1

Outliers

- outside $K = [q_1 - 1.5\,I.Q.R.,\ q_3 + 1.5\,I.Q.R]$
- boxplot whiskers within K
- Outliers indicated with •

$\Gamma(\alpha)$

$$\boxed{\Gamma(\alpha+1) = \alpha\Gamma(\alpha)}$$

$$= \int_0^\infty y^{\alpha-1} e^{-y} dy$$

if $\alpha \in \mathbb{N}$: $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

Gamma dist (gamma(α, β))

α : shape

β : scale ↕

Chi-square

$$\chi^2_\nu = gamma(\tfrac{\nu}{2}, 2)$$

Beta-dist

$$\boxed{B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}}$$

$$B(\alpha, \beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1} dy$$

$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

$Y \sim beta(\alpha, \beta)$

$$F_Y(y) = \sum_{i=\alpha}^{\alpha+\beta-1} bi(\alpha+\beta-1, y)$$

Moments

k^{th} non-central

$$\mu_k' = E[Y^k]$$

k^{th} central

$$\mu_k = E[(Y-\mu)^k]$$

Moment generating

$$m(t) = E[e^{tY}]$$

- exists if $m(t)$ is finite for any $-b \le t \le b$
- $e^{tY} = \sum_{i=0}^{\infty} \frac{(ty)^i}{i!}$

$$\boxed{e^\alpha = \frac{\alpha^0}{0!} + \frac{\alpha^1}{1!} + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!} = \sum_{i=0}^{\infty} \frac{\alpha^i}{i!}}$$

- μ_k is central, μ_k' is non-central

$$\mu_2 = \mu_2' - \mu^2$$

$$Var(Y) = E[Y^2] - E[Y_1]^2$$

Tschebbyscheff's thm

$$P(|Y-\mu| < k\sigma) > 1 - \frac{1}{k^2} \quad \text{or} \quad P(|Y-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Tschebyshefd's theorem

$$P(|Y-\mu| < k\sigma) \geq 1 - \frac{1}{k^2} \quad , k>0$$

$$1 - P(|Y-\mu| > k\sigma) \geq 1 - \frac{1}{k^2}$$

$$P(|Y-\mu| > k\sigma) \leq \frac{1}{k^2}$$

Baye's rule

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

$$P(y_1, y_2) = P(Y_1 = y_1 \mid Y_2 = y_2) * P(Y_2 = y_2)$$

Independence

$$f_{Y_1,Y_2}(y_1, y_2) = f_{Y_1}(y_1) * f_{Y_2}(y_2)$$

$$P(y_1, y_2) = P(y_1) * P(y_2)$$

$$E[g(Y_1) f(Y_2)] = E[g(Y_1)]\,E[f(Y_2)]$$

Covariance

$$\text{cov}(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)]$$

$$= E[Y_1 Y_2] - E[Y_1]\,E[Y_2]$$

if indep, $= 0$

correlation coeff.

$$\rho = \frac{\text{cov}(Y_1, Y_2)}{\sigma_1 \sigma_2} \quad , -1 \leq \rho \leq 1$$

Linear combinations

$$U = \sum_{i=1}^{n} a_i Y_i \qquad W = \sum_{j=1}^{m} b_j X_j$$

Variance

$$\text{Var}(U) = \sum_i a_i^2 \,\text{var}(Y_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j \,\text{cov}(Y_i, Y_j)$$

Covariance

$$\text{Cov}(U, W) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \,\text{cov}(Y_i, X_j)$$

Finding p.d.f. of random vars.

- Method of dist Fns

$$F_U(u) = P(U \le u)$$

- Method of transformation

$U = h(Y_1, Y_2)$ N.B. h increasing or decreasing for $Y_1, Y_2 > 0$

$$F_U(u) = P(U < u)$$
$$= P(h^{-1}(U) < h^{-1}(u))$$
$$= P(Y < h^{-1}(u))$$
$$= F_Y(h^{-1}(u))$$
$$f_U(u) = f_Y(h^{-1}(u)) \times \left| \frac{d}{du} h^{-1}(u) \right|$$

- Method of transformations for multivariable

$u_1 = h_1(y_1, y_2)$ $u_2 = h_2(y_1, y_2)$

$y_1 = h_1^{-1}(u_1, u_2)$ $y_2 = h_2^{-1}(u_1, u_2)$

N.B. $u_1 = h_1(y_1, y_2)$
$u_2 = h_2(y_1, y_2)$
must be one-to-one

$$J = \begin{bmatrix} \frac{du_1}{dy_1} & \frac{du_1}{dy_2} \\ \frac{du_2}{dy_1} & \frac{du_2}{dy_2} \end{bmatrix}$$

$$f_{U_1, U_2}(u_1, u_2) = f_{Y_1, Y_2}(h_1^{-1}(u_1, u_2), h_2^{-1}(u_1, u_2)) \, |J|$$

Sampling distributions

$\bar{Y} \sim n(\mu, \frac{\sigma^2}{n})$ $\quad$ (σ known)

$$\frac{\bar{Y} - \hat{\mu}}{\frac{\sigma}{\sqrt{n}}} \sim n(0,1)$$

t-dist

$Y \sim n(\mu, \frac{\sigma^2}{n})$ $\quad$ $s \sim \chi^2$

$$\frac{Y - \mu}{\frac{s}{\sqrt{n}}} \sim t_{(n-1)}$$

Hypergeometric

$$P_Y(y) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$$

n = sample
N = population
r = # y in population

$$E[Y] = \frac{nr}{N} = E[bi(n, \frac{r}{N})]$$

$$Var(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$= Var(bi(n, \frac{r}{N})) \times \left(\frac{N-n}{N-1}\right)$$