$\frac{1}{2}(x_{\lfloor np+0.5 \rfloor} + x_{\lceil np+0.5 \rceil})$

$x \notin (Q_1 - 1.5 \cdot IQR, Q_3 + 1.5 \cdot IQR)$

$\binom{n}{k_1, \dots, k_r} = \frac{n!}{k_1! k_2! \dots k_r!}$

$F_X(x) = P(X \le x)$

$P(a < X \le b) = F_X(b) - F_X(a)$

$E[g(x)] = \sum_{x \in X} g(x) p(x)$

$= \int_{-\infty}^{+\infty} g(x)\, p(x)\, dx$

independent: $\forall x, y$

$P(X \le x, Y \le y) = P(X \le x) P(Y \le y)$

$E[X^2] = (E[X])^2 + Var(X)$

$P(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}$

Binomial (n, p) $p \in [0,1]$ $n \in \mathbb{N}_0$

- $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ $k \ge 0$
- $E[X] = np$
- $Var(X) = np(1-p)$

Poisson (λ) $\lambda > 0$

- $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ $k \in \mathbb{N}_0$
- $E[X] = \lambda$
- $Var(X) = \lambda$

Poisson limit theorem

$\lim_{n \to \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$

$p_n \in [0,1]^{\mathbb{N}} \to \lambda$

continuous uniform distribution

$f(x) = \frac{1}{b-a}$ $a \le x \le b$

- $E[X] = \frac{a+b}{2}$
- $Var(X) = \frac{(b-a)^2}{12}$

$U(a,b)$ $a<b$

normal distribution

$N(\mu, \sigma^2)$

- $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
- $E[X] = \mu$
- $Var(X) = \sigma^2$

$\int_{-\infty}^{\infty} e^{t^2} dx = \sqrt{\pi}$

$\Phi(-z) = 1 - \Phi(z)$

$E[X^k]$ $E[|X^k|] < \infty$

n ~~th~~ moment ~~at~~ any point

$\to$ all lower moment about every point

$E[(X - E[X])^k]$

both exists including $E[|X|] < \infty$

k-th standardized

$E\left[\left(\frac{X-\mu}{\sigma}\right)^k\right]$

skewness 3rd,

kurtosis 4th

MGF

$M_X(t) = E[e^{tX}]$

if exists for open interval of 0

check t range before use

$E[X^k] = M_X^{(k)}(0)$

binomial

$M_X(t) = (pe^t + 1 - p)^n$

normal

$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

$SST = RSS + SSE$

$\Sigma(y - \bar{y})^2 = \Sigma(\hat{y} - \bar{y})^2 + \Sigma(y - \hat{y})^2$

$R^2 = \frac{RSS}{SST} = 1 - \frac{SSE}{SST} = r^2$

$(\hat{\beta}_1 - \beta_1)\sqrt{\frac{S_{xx}}{\sigma_\varepsilon^2}} \sim N(0,1)$

$(\hat{\beta}_0 - \beta_0)\sqrt{\frac{S_{xx}}{\sigma^2 \overline{x^2}}} \sim N(0,1)$

$(\hat{\beta}_1 - \beta_1)\sqrt{\frac{S_{xx}}{S_{\varepsilon,n-2}^2}} \sim t(n-2)$

$(\hat{\beta}_0 - \beta_0)\sqrt{\frac{S_{xx}}{\overline{x^2} S^2}} \sim t(n-2)$

$b \pm t_{n-2,\alpha/2}\sqrt{\frac{S^2}{S_{xx}}}$

$a \pm t_{n-2,\alpha/2}\sqrt{\frac{\overline{x^2} S^2}{S_{xx}}}$

$\hat{y}_{new} = a + b x_{new}$

CI

$\hat{y}_{new} \pm t_{n-2,\alpha/2}\sqrt{S_{\varepsilon,n-2}^2\left(1 + \frac{1}{n} + \frac{(x_{new} - \bar{x})^2}{S_{xx}}\right)}$

$S_\varepsilon^2 = \frac{S_{yy} - b S_{xy}}{n-2}$

$= \frac{\Sigma^n (y - \hat{y})^2}{n-2} = \frac{SSE}{n-2}$

$\frac{(n-2) S_{n-2}^2}{\sigma_\varepsilon^2} \sim \chi^2(n-2)$

Sample mean

$$\bar{X} = \frac{1}{N}\sum_{k=1}^{N} X_k \qquad E[\bar{X}] = \mu_X \qquad var(\bar{X}) = \frac{\sigma_X^2}{N}$$

sample covariance

$$Cov(X,Y) = \frac{1}{N-1}\sum_{k=1}^{n}(X_k - \bar{X})(Y_k - \bar{Y})$$

$$S_{Z,n-1}^2 = Cov(Z,Z) = \frac{1}{N-1}\sum_{k=1}^{N}(Z_k - \bar{Z})^2$$

$$E[S_{X,n-1}^2] = \sigma_X^2$$

$$\overline{X^k} = \frac{1}{N}\sum_{i=1}^{N} X_i^k \qquad E[\overline{X^k}] = E[X^k]$$

Method of Moments

$$E[X] = \bar{X}$$

$$E[X^k] = \overline{X^k}$$

non-unique, existence of k moments, sample size large, then biased

biased, asymptotic unbiased

$$P(T_1 \le \theta \le T_2) \ge 1-\alpha$$

$\chi^2(k)$

$$f(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2} \quad y>0$$

$E[Y] = k \quad Var(Y) = 2k$

$t(v)$

$$f(w) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n}\,\Gamma(n/2)}\left(1+\frac{w^2}{n}\right)^{-(n+1)/2}$$

William Sealy Gosset "Student" 1908

$$(\bar{X}-\mu_X)\sqrt{\frac{n}{\sigma_X^2}} \sim N(0,1)$$

$\bar{X}$ and $S_{X,n-1}^2$ independent

$$\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma_X}\right)^2 = \frac{(n-1)S_{X,n-1}^2}{\sigma_X^2} \sim \chi^2(n-1)$$

$$(\bar{X}-\mu_X)\sqrt{\frac{n}{S_{X,n-1}^2}} \sim t(n-1)$$

var $\to$ mean

$$(\bar{X}-\mu_X)\sqrt{n/\sigma_X^2} \sim N(0,1)$$

$$\bar{X} \pm z_{\alpha/2}\sqrt{\frac{\sigma_X^2}{n}}$$

$\to$ mean

$$(\bar{X}-\mu_X)\sqrt{n/S_{X,n-1}^2} \sim t(n-1)$$

$$\bar{X} \pm t_{n-1,\alpha/2}\sqrt{\frac{S_{X,n-1}^2}{n}}$$

mean $\to$ var

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu_X}{\sigma_X}\right)^2 \sim \chi^2(n)$$

$$\frac{\sum_{i=1}^{n}(X_i-\mu_X)^2}{\chi^2_{n,\alpha/2}} \le \sigma_X^2 \le \frac{\sum_{i=1}^{n}(X_i-\mu_X)^2}{\chi^2_{n,1-\alpha/2}}$$

$\to$ var

$$\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma_X}\right)^2 \sim \chi^2(n-1)$$

$$\frac{(n-1)S_{n-1}^2}{\chi^2_{n-1,\alpha/2}} \le \sigma_X^2 \le \frac{(n-1)S_{n-1}^2}{\chi^2_{n-1,1-\alpha/2}}$$

contains equal never contains equal

simple H_0 $\theta=\theta_0$ against H_1 $\theta=\theta_1$

left-tailed H_0 $\theta=\theta_0$ against H_1 $\theta<\theta_0$

right H_0 $\theta=\theta_0$ against H_1 $\theta>\theta_0$

two H_0 $\theta=\theta_0$ against H_1 $\theta\ne\theta_0$

$\alpha = Pr(\text{reject } H_0 \mid H_0 \text{ true})$

$\beta = Pr(\text{not reject } H_0 \mid H_0 \text{ false})$

power $= 1-\beta$

level of significance α

(not) reject H_0 at significance level α

hypothesis testing

$$(\bar{X}-\mu_0)\sqrt{\frac{n}{\sigma_X^2}} \sim N(0,1)$$

$-z_{\alpha/2} \quad z_{\alpha/2}$

$$(\bar{X}-\mu_0)\sqrt{\frac{n}{S_{X,n-1}^2}} \sim t(n-1)$$

$-t_{n-1,\alpha/2} \quad t_{n-1,\alpha/2}$

$$\frac{\sum_{i=1}^{n}(X_i-\mu_X)^2}{\sigma_0^2} \sim \chi^2(n)$$

$\chi^2_{n,\alpha} \quad \chi^2_{n,1-\alpha}$

$\chi^2_{n,\alpha/2} \quad \chi^2_{n,1-\alpha/2}$

$$\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma_0^2} \sim \chi^2(n-1)$$

$\chi^2_{n-1,\alpha} \quad \chi^2_{n-1,1-\alpha}$

$$\frac{(n-1)S_{n-1}^2}{\sigma_0^2}$$

$\chi^2_{n-1,\alpha/2} \quad \chi^2_{n-1,1-\alpha/2}$

$Y = \beta_0 + \beta_1 x + \varepsilon_i$

$\varepsilon_i \sim N(0,\sigma^2)$ iid

$$b = \frac{Cov(X,Y)}{Cov(X,X)} = \frac{E[XY]-E[X]E[Y]}{E[X^2]-(E[X])^2}$$

$$= \frac{S_{XY}}{S_{XX}} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$$

$$= \frac{\sum xy - \sum x \sum y/n}{\sum x^2 - (\sum x)^2/n}$$

$$a = \bar{y} - b\bar{x}$$

$$r_{xy} = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

$$= \frac{\sum xy - \sum x \sum y/n}{\sqrt{\sum x^2 - (\sum x)^2/n}\sqrt{\sum y^2 - (\sum y)^2/n}}$$

$\rho_{X,Y}$

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \operatorname{arctanh}(r)$$

$$\left(z - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)\right)\sqrt{n-3} \sim N(0,1)$$