5

1): Find the equilibrium point.

$\frac{df}{dt}=0$

$\frac{1}{10}f(1-\frac{f}{1000})-h=0$

$1000f-f^2-10000h=0$

$f=\frac{1000\pm\sqrt{10000(100-4h)}}{2}$

$=500\pm50\sqrt{100-4h}=50(10\pm\sqrt{100-4h})$

$100-4h\geqslant0$ $h\leq25$ Real solution

根解出来. 2) $\boxed{\text{Equilibrium at } 100}$

$50(10\pm\sqrt{100-4h})=100$

$10+\sqrt{100-4h}=2$

$\sqrt{100-4h}=-8$

$100-4h=64$

$h=9$

3) remain constant means the initial population must at the equilibrium point

$f_0=50(10-\sqrt{100-4h})$ since $h=21$

$=50(10-\sqrt{100-4\times21})$

$=50\times6$

$=300$

Summary

5

Particular Integral

类型1. $f(x)=e^{\lambda x}P_n(x)$

$P_n(x)=a_0x^n+a_1x^{n-1}+\cdots+a_{n-1}x+a_n$

特解形式: $y=e^{\lambda x}Q(x)$ 代入 $\Rightarrow$

$Q''(x)+(2\lambda+p)Q'(x)+(\lambda^2+p\lambda+q)Q(x)=P_n(x)$

① 若λ不是齐次特征方程的根. $2\lambda+p\neq0$. $\lambda^2+p\lambda+q\neq0$

$\therefore P_n(x)$ n次. $Q(x)$ 也是 n次

$\therefore Q(x)=b_0x^n+b_1x^{n-1}+\cdots+b_n$.

② 若λ是齐次特征方程的单特征根 $\Rightarrow \lambda^2+p\lambda+q=0$

$2\lambda+p\neq0$ $Q'(x)$ 应该是 n次 多项式

$\therefore Q(x)=x\cdot Q_n(x)$.

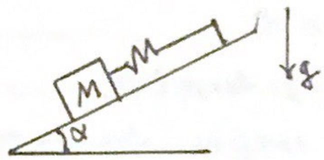
③ 若λ是齐次特征方程的重特征根 $\lambda^2+p\lambda+q=0$.

$2\lambda+p=0$ $Q''(x)$ 是 n次 多项式

$Q(x)=x^2Q_n(x)$.

Summary

5



(1.5) A block of mass M lies on a smooth flat plane inclined at an angle α to the horizontal. It is attached to a light spring with elastic modulus Mg and natural length L that is tethered to a fix point so that it is parallel to the plane:
If the block is released from rest with the spring stretched to twice its natural length, show that the shortest length of the spring in the subsequent oscillation is $2L\sin\alpha$.

隐藏了两个初始条件 $x(0)=2L$. $\dot{x}(0)=0$.

total length: x

$$N\text{II}) \Rightarrow \quad M\ddot{x} = Mg\sin\alpha - Mg\frac{(x-L)}{L}$$

Try $x_p = C$ (constant)

$$\ddot{x} + \frac{g}{L}x = g(1+\sin\alpha)$$

$$x(t) = L(1+\sin\alpha) + A\sin\sqrt{\frac{g}{L}}\,t + B\cos\sqrt{\frac{g}{L}}\,t$$

$x(0)=0$. $\dot{x}(0)=0$

$$x(t) = L(1+\sin\alpha) + L(1-\sin\alpha)\cos\sqrt{\frac{g}{L}}\,t$$

when $\cos\sqrt{\frac{g}{L}}\,t = -1$ $\quad x_{min} = 2L\sin\alpha$.

Summary

Logistic Growth:

$$\frac{dP}{dt} = f(P)\, P.$$

$f(0) = r$. $f(K) = 0$

let $f(P) = c_1 P + c_2$.

$\Rightarrow$ $c_2 = r$. $c_1 = -r/K$

$\therefore$ $\frac{dP}{dt} = P(r - \frac{r}{K}P)$

$\frac{dP}{dt} = rP(1 - \frac{P}{K})$

Predator-Prey Model:

$x(t)$: fox $y(t)$: rabbit.

$\frac{dx}{dt} = -ax$. $a>0$ (no rabbit) $\frac{dy}{dt} = dy$. $d>0$ (no fox)

$\frac{dx}{dt} = -ax + bxy$ $\frac{dy}{dt} = dy - cxy$

it's reasonable that. the interaction between these two species per unit time is jointly proportionaly to their population x. y.

$\therefore$ Lotka-Volterra Predator-Prey Model

$$\begin{cases} \frac{dx}{dt} = -ax + bxy \\ \frac{dy}{dt} = dy - cxy \end{cases}$$

Competition Models:

$\frac{dx}{dt} = ax$ $\frac{dy}{dt} = cy$.

$\downarrow$ $\downarrow$

$\frac{dx}{dt} = ax - by$ $\frac{dy}{dt} = cy - dx$

rate is diminished by the influence. or existence of the other population.

$\downarrow$ $\downarrow$

$\frac{dx}{dt} = ax - bxy$ $\frac{dy}{dt} = cy - dxy$ consider the interaction between them.

$\downarrow$ $\downarrow$

$\frac{dx}{dt} = a_1 x - b_1 x^2 - c_1 xy$ $\frac{dy}{dt} = a_2 y - b_2 y^2 - c_2 xy$.

More realistic: Grow Logistically.

Summary

$$\begin{cases} \frac{dx}{dt} = a_1 x - b_1 x^2 - c_1 xy \\ \frac{dy}{dt} = a_2 y - b_2 y^2 - c_2 xy \end{cases}$$

SIR Model

S: Susceptible

I: Infective

R: Removal.

$$\begin{cases} \frac{dS}{dt} = -kSI \\ \frac{dI}{dt} = kSI - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases}$$

$\frac{d}{d} + \frac{d}{d} + \frac{d}{d} = 0.$

系统不变.

$$\boxed{S} \xrightarrow{k} \boxed{I} \xrightarrow{\gamma} \boxed{R}$$

$N = S + I + R$

Summary

Rate Equation:

$\frac{dx}{dt} = f(x, t)$, $x(0) = x_0$. non-autonomous

$\frac{dx}{dt} = f(x)$ $x(0) = x_0$ autonomous.

Introduction: Newton's Law

$\frac{dx}{dt} = V$. $\frac{dp}{dt} = F$ $p = mv$

$m\frac{d^2x}{dt^2} = F$.

Chemical Reaction Kinetics

reactants $\xrightarrow{k}$ products.

$\frac{dA}{dt} = + \sum_{n=1}^{N} (\text{creation rate})_n - \sum_{n=1}^{N} (\text{consumption})_n$.

Reaction law:

A is pumped into the system at a constant rate K.

① Constant supply: (source) $\xrightarrow{K}$ A

$\frac{dA}{dt} = K$

② Decay: $A \xrightarrow{K} (\text{waste})$ $\frac{dA}{dt} = -KA$

A is consumed with B produced from A.

③ Transformation: $A \xrightarrow{K} B$ $\frac{dA}{dt} = -KA$. $\frac{dB}{dt} = KA$

④ Reversible transformation: $A \underset{K_2}{\overset{K_1}{\rightleftharpoons}} B$

$\frac{dA}{dt} = -K_1A + K_2B$; $\frac{dB}{dt} = K_1A - K_2B$

⑤ Compound formation: $A + B \xrightarrow{K} C$

$\frac{dA}{dt} = -KAB$ $\frac{dB}{dt} = -KAB$ $\frac{dC}{dt} = KAB$

$A + A \xrightarrow{K} C = 2A \xrightarrow{K} C$

$\frac{dC}{dt} = KA^2$ $\boxed{\frac{dA}{dt} = -2KA^2}$

next page.

Summary

① Multiple product:

$$nA + mB \xrightarrow{K} pC + qD.$$

$$\text{Rate} = \frac{1}{p}\frac{dC}{dt} = \frac{1}{q}\frac{dD}{dt} = -\frac{1}{n}\frac{dA}{dt} = -\frac{1}{m}\frac{dB}{dt} = KA^nB^m$$

reactants: $\frac{dA}{dt} = -nKA^nB^m \quad \frac{dB}{dt} = -mKA^nB^m$

products: $\frac{dC}{dt} = pKA^nB^m \quad \frac{dD}{dt} = qKA^nB^m$

Summary

5

Note 1

range: Max − min

Sample variance:

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

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

$$= \frac{1}{n-1}\left\{\sum f_i x_i^2 - \frac{1}{n}\left(\sum f_i x_i\right)^2\right\}$$

Statistics Revision

more robust

Measures of location: Mean Mode Median

Measures of Spread: range variance standard deviation

not robust too sensitive

interquartile range.

most robust

Min Q_1 Q_2 Q_3 Max

Lower quartile Median Upper quartile

Calculate Quartile:

~~$x_p = \begin{cases} X_{[np]+1} & np \neq \text{integer} \\ \frac{1}{2}[X_{(np)} + X_{(np+1)}] & np = \text{integer} \end{cases}$~~

~~p is the quartile. ie. lower quartile p = 0.25~~

~~Upper quartile p = 0.75~~

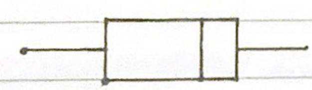
Interquartile range = | Lower − Upper |

location: Low: $\frac{n+1}{4}$ ~~$\frac{1}{4}x = \frac{3}{4}$~~

Upper: $\frac{3(n+1)}{4}$ ~~$\frac{3}{4}x + \frac{1}{4}$~~

↳ This is the exact position.

Box plot: due to distance and proportions



Min Q_1 M Q_3 Max

Outlier: $[Q_1 - 1.5IQR,\ Q_3 + 1.5IQR]$ beyond

Histogram: ⇒ Area proportional to frequency.

Stem-leaf: number cut instead of round.

Summary

5

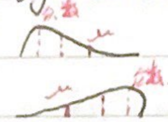
Skewness: $g = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^3$

s: standard deviation $\bar{x}$: sample mean

skewness $= 0$ symmetric

> 0

< 0



Robustness.

切尾均值

- Trimmed Mean: delete the $\alpha\%$ smallest observation and $\alpha\%$ largest observation

缩尾均值

- Winsorised Mean: set the $\alpha\%$ smallest observation equal to the next largest observations and the $\alpha\%$ largest obs. equal to the next smallest obs.

$F(x) = \int f(x)\,dx$

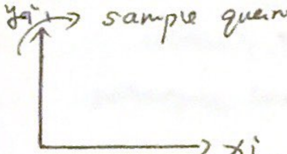
QQ plot

$F(x_i) = \frac{i-\frac{1}{2}}{n}$ $F(\,)$ is $\boxed{\text{cdf}}$ of the model.

plot y_i vs x_i

straight line through (0,0) with gradient 1

QQ Plot with known mean and variance.



$y_i \to$ sample quantile

$\to x_i$ Theoretical quantile

1. Rank $z_{(i)} = \frac{y_{(i)}-\mu}{\sigma}$.

2. $F(x_i) = \frac{i-\frac{1}{2}}{n}$ $F(x) = P(X \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} du$.

3. Plot.

QQ Plot with unknown mean and variance.

if data comes from normal distribution

$z_{(i)} \approx x_{(i)}$ so $y_{(i)} \approx \sigma x_i + \mu$.

① straight line

② Gradient 1

Summary

5 . .

[illegible]: iid:
independent and identically distributed

$S_n = X_1 + \cdots + X_n$
$S_n \sim N(n\mu, n\sigma^2)$
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

iid. independent and identically distributed

$X_1 \cdots X_n$

Mutually Independent: $N(\mu_i, \sigma_i^2)$

$\sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$

Furthermore, if iid. $(E(X_i) = \mu$. $var(X_i) = \sigma^2)$.

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Central Limit Theorem

If $X_1 \cdots X_n$ are i.i.d. with μ, σ^2

$\bar{X} \sim N(\mu, \sigma^2/n)$ for large n, no matter what the distribution of the X_i.

Note 2:

Good estimate: point estimate

- $E(T) = \theta$
- min $Var(T)$ minimum variance unbiased estimator (MVUE)

Unbiased Estimation of population mean μ.

$E[\bar{X}] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu$.

Unbiased Estimation of population variance σ^2.

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

$E[\Sigma(X_i - \bar{X})^2]$
$= E[\Sigma(X_i^2 - 2X_i\bar{X} + \bar{X}^2)]$
$= E[\Sigma X_i^2 - n\bar{X}^2]$
$= \sum_{i=1}^{n} E[X_i^2] - nE(\bar{X}^2)$

$var(X_i) = \sigma^2 = E(X_i^2) - E(X_i)^2$

$var(\bar{X}) = \frac{\sigma^2}{n} = E(\bar{X}^2) - E(\bar{X})^2$

$E[\Sigma(X_i - \bar{X})^2] = (n-1)\sigma^2$

$E[S^2] = \frac{1}{n-1}E[\Sigma(X_i - \bar{X})^2] = \sigma^2$.

Summary

either $n \uparrow$ CLT. holds

or data normally dist.

width: $Z = 2 Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Interval Estimate:

$P(\theta_L(X_1 \dots X_n) < \theta < \theta_U(X_1 \dots X_n)) = 1-\alpha$

Confidence Interval: for μ. σ known.

$P(-Z_{\alpha/2} < Z \le Z_{\alpha/2}) = 1-\alpha \quad Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Confidence Interval for μ σ. unknown

$t = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t_{n-1}$

$P(-t_{n-1,\alpha/2} \le t \le t_{n-1,\alpha/2}) = 1-\alpha$

$\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$.

$P(\le Z_{\alpha/2}) \Rightarrow \Phi(Z_{\alpha/2})$

$Z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$

$Z_{1-\alpha/2} = \Phi^{-1}(\alpha/2)$

Hypothesis Testing.

Null hypothesis H_0

Alternative hypothesis H_1

significant level α: $\alpha = P(\text{Reject } H_0 \text{ in favor of } H_1 / H_0 \text{ true})$

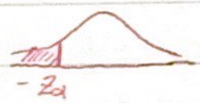
Test for mean μ σ known

one side test: α

two sides test: $\frac{\alpha}{2}$

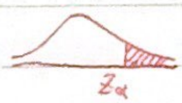
- $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$

  Reject H_0 if $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -\boxed{Z_\alpha}$



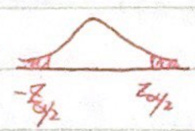
- $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$

  Reject H_0 if $Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > \boxed{Z_\alpha}$



- $H_0: \mu = \mu_0$ vs $H_1: \mu \ne \mu_0$

  Reject H_0 if $|Z| = \left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right| > \boxed{Z_{\alpha/2}}$



Summary

Interpretation of C.I.

confident interval.

repeated trials

Every C.I: 要么有 contain μ，要么不 contain μ.

$\downarrow$

Under repeated trial

95% of interval will cover that μ.

原假设可能被拒绝的最小显著性水平

做法: ① only use $\alpha = ..$ as reference in exam.

p-values. ② 我们的 p 可以找一个范围 (table)

$$\text{p-value} = P\left(\begin{array}{l}\text{obtaining a test statistic}\\ \text{at least as extreme}\\ \text{as that observed}\end{array} \middle| H_0 \text{ is true}\right)$$

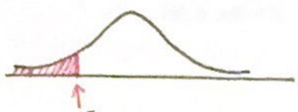
one-side

$$\text{p-value} = P\left(\begin{array}{l}\text{obtaining a test statistic}\\ \text{at least as extreme in}\\ \text{either direction}\end{array} \middle| H_0 \text{ is true}\right)$$

two-side

$$= 2P\left(\begin{array}{l}\text{p-value for one side}\\ \text{test for symmetrical problem}\end{array}\right)$$

$p \le 0.01$ strong evident against H_0



$z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$

$0.01 < p \le 0.05$ evidence against H_0

$0.05 < p \le 0.1$ weak evidence against H_0

$p > 0.1$ no evidence against H_0

Require: $X_1 \cdots X_n \sim \text{iid } N(\mu, \sigma^2)$

Test for means: σ unknown

- Test: $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{n-1, \boxed{\alpha}}$$

- Test: $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{n-1, \boxed{\alpha}}$$

- Test: $H_0: \mu = \mu_0$ vs $H_1: \mu \ne \mu_0$

$$|t| = \left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| > t_{n-1, \boxed{\alpha/2}}$$

Type I Type II error and Power

	H_0 True	H_0 FALSE
Reject H_0	Type I (错杀好人) α	
Accept H_0		Type II (放走坏人) β

$P(\text{Type I}) = P(\text{reject } H_0 \mid H_0 \text{ true}) = \alpha$

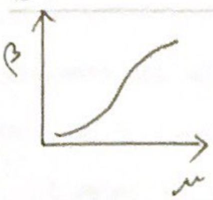
$P(\text{Type II}) = P(\text{accept } H_0 \mid \underset{H_1 \text{ true}}{H_0 \text{ false}}) = \beta$

Power: $1 - P(\text{Type II error}) = P(\text{reject } H_0 \mid H_0 \text{ false}) = 1-\beta$

Summary

1. 先算 critical value.
2. 求的拒绝范围 (rejection region)
3. 求 $1-\beta$. ($\mu_1 =$) as a new parameter.
4. 正态化 $z = \frac{\bar{x}-\mu_1}{\sigma/\sqrt{n}}$
5. 得出概率

5



Compare means of two population

size n_1: $N(\mu_1, \sigma_1^2)$

size n_2: $N(\mu_2, \sigma_2^2)$ $\qquad \mu_1 - \mu_2$

$\bar{x}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$ $\qquad \bar{x}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$

$E[\bar{x}_1 - \bar{x}_2] = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2$

$Var(\bar{x}_1 - \bar{x}_2) = Var(\bar{x}_1) + Var(\bar{x}_2) - 2Cov(\bar{x}_1, \bar{x}_2)$

$= Var(\bar{x}_1) + Var(\bar{x}_2)$ (since independent)

$= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$$\boxed{\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}$$

Standardise: $Z = \frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

Case 1: σ_1^2 σ_2^2 known [or n_1 $n_2 \geq 30$]

100α% reject H_0 if Z lies in the critical region

H_1	Critical Region
$\mu_1 < \mu_2$	$Z < -z_\alpha$
$\mu_1 > \mu_2$	$Z > z_\alpha$
$\mu_1 \neq \mu_2$	$\lvert Z \rvert > z_{\alpha/2}$

require independent normal observations with known variance

General form of C.I.s.: estimate ± table value × standard deviation of estimate.

C.I.: $\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Case 2: σ_1^2 σ_2^2 unknown, assume equal.

$\bar{x}_1 - \bar{x}_2 \sim N(\mu_1 - \mu_2, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$

$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\boxed{\sigma^2}(\frac{1}{n_1} + \frac{1}{n_2})}} \sim N(0,1)$

unknow

Require independent & normal observations with same unknown variance

pooled estimator $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ (weight average)

$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{S_p^2(\frac{1}{n_1} + \frac{1}{n_2})}} \sim t_{n_1 + n_2 - 2}$

Summary

$100\alpha\%$ reject H_0 if t lies in the critical region

H_1	Critical region
$\mu_1 < \mu_2$	$t < -t_{n_1+n_2-2,\alpha}$
$\mu_1 > \mu_2$	$t > t_{n_1+n_2-2,\alpha}$
$\mu_1 \neq \mu_2$	$\lvert t \rvert > t_{n_1+n_2-2,\boxed{\alpha/2}}$

C.I. $\bar{x}_1 - \bar{x}_2 \pm t_{n_1+n_2-2,\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

Case 3: σ_1^2 σ_2^2 unknown & unequal.

Behrens-Fisher problem.

Paired Example.

100 ~~%~~: same person ← before and after. Before vs after

$d_i = x_{1i} - x_{2i}$

Assume $D_i \sim N(\mu_d, \sigma_d^2)$

$$\bar{D} \sim \frac{1}{n}\sum_{i=1}^{n} D_i \sim N(\mu_d, \sigma_d^2/n)$$

σ_d^2 is estimated by $S_d^2 = \frac{1}{n-1}\left\{\left(\sum_{i=1}^{n} d_i^2\right) - \frac{1}{n}\left(\sum_{i=1}^{n} d_i\right)^2\right\}$

$$t = \frac{\bar{D} - \mu_d}{S_d/\sqrt{n}} \sim t_{n-1}$$

Test. $H_0: \mu_d = 0$. →

$100\alpha\%$ level ~~test~~. reject H_0 if t lies in the critical region.

H_1	Critical Region
$\mu_d < 0$	$t < -t_{n-1,\alpha}$
$\mu_d > 0$	$t > t_{n-1,\alpha}$
$\mu_d \neq 0$	$\lvert t \rvert > t_{n-1,\alpha/2}$

C.I.: $\bar{d} \pm t_{n-1,\alpha/2}\frac{S_d}{\sqrt{n}}$.

Summary

5

Relationship between confidence interval & hypothesis tests.

hypothesis test at α% level

confidence interval $(100(1-\alpha))$% level.

A two-sided hypothesis test at the 100α% level will accept $H_0: \mu = \mu_0$ iff μ_0 lies in the $(100(1-\alpha))$% confidence interval for μ.

Inference about Population Variance

One population.

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

$$Z_1^2 + \cdots + Z_n^2 \sim \chi_n^2$$

$$\frac{(X_1-\mu)^2}{\sigma^2} + \cdots + \frac{(X_n-\mu)^2}{\sigma^2} \sim \chi_n^2$$

$$\frac{\sum (X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$$

$$s^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

Chi-squared distribution with n-1 degrees of freedom

$$\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

[($X_1 \ldots X_i$ iid $N(\mu, \sigma^2)$.)]

$$P(a \le \frac{(n-1)s^2}{\sigma^2} \le b) = 1-\alpha$$

$a = \chi^2_{n-1,\alpha/2}$, $b = \chi^2_{n-1,\alpha/2}$

$$100(1-\alpha)\%\ \text{CI}: \left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

Hypothesis Test:

$H_0: \sigma = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$

Reject H_0 if $\frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1,\alpha/2}$ or $\frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1,1-\alpha/2}$

Ratio of two population variances

$$\frac{(n_X-1)S_X^2}{\sigma_X^2} \sim \chi^2_{n_X-1} \qquad \frac{(n_Y-1)S_Y^2}{\sigma_Y^2} \sim \chi^2_{n_Y-1}$$

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \sim F_{n_X-1,\, n_Y-1}.$$

S^2 is sample variance.

Summary

Testing $H_0: \sigma_x^2 = \sigma_y^2$ vs $H_1: \sigma_x^2 \neq \sigma_y^2$

Under H_0 $F = \frac{S_x^2}{S_y^2} \sim F_{n_x-1, n_y-1}$

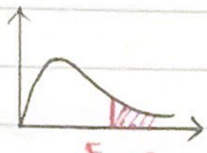
Reject H_0 at the $100\alpha\%$ level if

$F > F_{n_x-1, n_y-1, \alpha/2}$ or

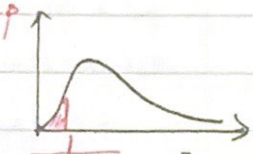
$F < F_{n_x-1, n_y-1, 1-\alpha/2} = \frac{1}{F_{n_y-1, n_x-1, \alpha/2}}$

$F_{v_1, v_2, \beta} = \frac{1}{F_{v_2, v_1, 1-\beta}}$

swap



$F_{v_1, v_2, \beta}$



$\frac{1}{F_{v_2, v_1, \beta}} = F_{v_1, v_2, 1-\beta}$

Note 3:

Correlation.

Sample covariance.

$S_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{n-1}\left(\sum x_i y_i - \frac{1}{n}(\sum x_i)(\sum y_i)\right)$

[estimation ~~from~~ for $\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X\mu_Y$]

Pearson's correlation coeffirient. ←

Sample correlation: $r = \frac{S_{xy}}{S_x S_y} \in [-1, 1]$

[estimate for $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$]

$\boxed{\text{T - test}}$

(Assume X. Y. linear)

((X. Y. normal dis.)

Step:

$S_{xy} = \frac{1}{n-1}(\sum x_i y_i - \frac{\sum x_i \sum y_i}{n})$

$S_x^2 = \frac{1}{n-1}(\sum x_i - \frac{(\sum x_i)^2}{n})$

$S_y^2 = \frac{1}{n-1}(\sum y_i - \frac{(\sum y_i)^2}{n})$

Hypothesis test for correlation

$H_0: \rho = 0$ vs $H_1: \rho \neq 0$

Assume $(x_1, y_1) \ldots (x_n, y_n)$ are normally distributed.

$t = r\sqrt{\frac{n-2}{1-r^2}} \sim t_{n-2}$ $r = \frac{S_{xy}}{S_x S_y}$

Reject H_0 if $|t| = \left| r\sqrt{\frac{n-2}{1-r^2}} \right| > t_{\boxed{n-2}, \boxed{\alpha/2}}$

(Requires X Y to have a linear relationship).

如果是 one side test, use $t_{n-2, \alpha}$

Summary

$r = \frac{S_{xy}}{\boxed{\sqrt{S_x^2}} \cdot \boxed{\sqrt{S_y^2}}}$

5

H_0: no correlation

H_1: some correlation

Non-parametric Test

(suitable for nonlinear)

(no assumption about dist.)

Rank

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

Spearman's rank order correlation coefficient.

d_i is the difference between the rank of x_i and y_i.

Linear Regression.

$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, \dots n.$

Assumption:

(not need to be normally dist.)

但是假设检验需要

1. $E(\epsilon_i) = 0$ the errors have zero mean
2. $cov(\epsilon_i, \epsilon_j) = 0$ the errors are uncorrelated
3. $var(\epsilon_i) = \sigma^2$ ~~the~~ have the same variances

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$S_{xy} = \frac{1}{n-1}\left(\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i\right) \qquad \hat{\beta}_1 = \frac{S_{xy}}{S_x^2}$$

$$S_x^2 = \frac{1}{n-1}\left(\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right)$$

(estimation for β_0, β_1)

Property: $E(\hat{\beta}_0) = \beta_0$

$E(\hat{\beta}_1) = \beta_1$.

$$var(\hat{\beta}_0) = \frac{\sigma^2 \sum x_i^2}{n\sum(x_i - \bar{x})^2}$$

$$var(\hat{\beta}_1) = \frac{\sigma^2}{\sum(x_i - \bar{x})^2}$$

$$cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\sigma^2 \bar{x}}{\sum(x_i - \bar{x})^2}$$

Summary

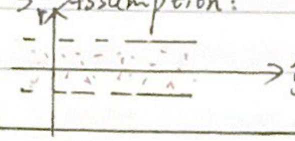
Estimate σ^2

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=1}^{n} r_i^2$$

where $r_i = y_i - \hat{y}_i$ is the residual.

Quality of fit broad horizontal band of points.

3. Assumption:



Inference in Regression.

Inferences β_1.

$$\hat{\beta}_1 = \frac{S_{xy}}{S_x^2} \sim N\left(\beta_1, \frac{\sigma^2}{(n-1)S_x^2}\right)$$

$$\hat{\sigma}^2 = \frac{(n-1)}{(n-2)}\left[s_y^2 - \hat{\beta}_1^2 s_x^2\right]$$

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{(n-1)S_x^2}} \sim N(0,1)$$

↳ unknown.

$$t = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}/\sqrt{(n-1)S_x^2}} \sim t_{n-2}.$$

test for existence of a regression

$100(1-\alpha)\%$ C.I. : $\hat{\beta}_1 \pm t_{n-2,\alpha/2}\dfrac{\hat{\sigma}}{\sqrt{(n-1)S_x^2}}$

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

Under H_0 reject H_0 if $|t| = \left|\dfrac{\hat{\beta}_1 - 0}{\hat{\sigma}/\sqrt{(n-1)S_x^2}}\right| > t_{n-2,\alpha/2}$

In general: $H_0: \beta_1 = \beta^*$ vs $H_1: \beta_1 \neq \beta^*$ we reject H_0 if

$$|t| = \left|\frac{\hat{\beta} - \beta^*}{\hat{\sigma}/\sqrt{(n-1)S_x^2}}\right| > t_{n-2,\alpha/2}$$

Summary

5

Inference for mean value of Y

A $100(1-\alpha)\%$ C.I. for mean value of Y when $x = x_0$ is

at $X = X_0$.

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{n-2,\alpha/2}\, \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{(n-1)s_x^2}}$$

Inference for individual Y

A $100(1-\alpha)\%$ C.I. for an individual value of Y when $X = X_0$ is

$$(\hat{\beta}_0 + \hat{\beta}_1 x_0) \pm t_{n-2,\alpha/2}\, \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0-\bar{x})^2}{(n-1)s_x^2}}$$

Analysis of variance (ANOVA)

Total Sum of Squares: $\sum_{i=1}^{n}(y_i-\bar{y})^2 = (n-1)s_y^2$

$$\sum(y_i-\bar{y})^2 = \underbrace{\sum(y_i-\hat{y}_i)^2}_{\text{residual sum of squares}} + \underbrace{\sum(\hat{y}_i-\bar{y})^2}_{\text{regression sum of squares}}$$

$\uparrow R^2$. Good fit of a straight line.

$R^2 = \frac{\text{Regression SS}}{\text{Total SS}}$ $\quad 0 \le R^2 \le 1$

$(R^2 = r^2.)$

residual: $r_i = y_i - \hat{y}_i$

TSS: $\sum(y_i-\bar{y})^2$

RSS: $\sum(y_i-\hat{y})^2$

ESS: $\sum(\hat{y}_i-\bar{y})^2$ $\quad R^2 = \frac{\text{ESS}}{\text{TSS}}$

Summary

George Willsons

5

Predict X from Y

$\beta_0^* = \bar{x} - \beta_1^* \bar{y}$ $\beta_1^* = \frac{S_{xy}}{S_y^2}$ (not symmetric)

换成了 S_y^2

Familiar formula: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, $\hat{\beta}_1 = \frac{S_{xy}}{S_x^2}$

Inference about proportion

Single proportion

π is the probability of having some specific attitude

$\begin{cases} \mu = n\pi \\ \sigma^2 = n\pi(1-\pi) \end{cases}$

Binomial Model. $X \sim Binomial(n, \pi)$

$X \sim Binomial(n, \pi) \approx N(n\pi, n\pi(1-\pi))$

Hypothesis test $H_0: \pi = \pi_0$ vs $H_1: \pi \neq \pi_0$

Under H_0, $Z = \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} \sim N(0,1)$

$100\alpha\%$ level test reject H_0 if

$$|Z| = \left| \frac{X - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} \right| > Z_{\alpha/2}$$

sample porportion

$$p = \frac{X}{n} \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

$100\alpha\%$ level test reject H_0 if

$$|Z| = \left| \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \right| > Z_{\alpha/2}$$

CI: $100(1-\alpha)\%$ $p \pm Z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}$ margin of error

π_0 was replaced by p.

For small n, a slightly improved statistic is

$$\frac{|X - n\pi_0| - \frac{1}{2}}{\sqrt{n\pi_0(1-\pi_0)}} \sim N(0,1)$$

Summary

5

Two population

two samples are independent n_1 n_2 so large.

$p_1 \sim N(\pi_1, \frac{\pi_1(1-\pi_1)}{n_1})$ $p_2 \sim N(\pi_2, \frac{\pi_2(1-\pi_2)}{n_2})$

$E(p_1-p_2) = E(p_1) - E(p_2) = \pi_1 - \pi_2$

$var(p_1-p_2) = var(p_1) + var(p_2) = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$

$p_1 - p_2 \sim N(\pi_1 - \pi_2, \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2})$

Hypothesis test:

$H_0: \pi_1 = \pi_2$ vs $H_1: \pi_1 \neq \pi_2$

Under H_0. $\pi_1 = \pi_2 = \pi$ $p_1 - p_2 \sim N(0, \pi(1-\pi)(\frac{1}{n_1} + \frac{1}{n_2}))$

Since we don't know π, so we estimate it

$\hat{\pi} = \frac{X_1 + X_2}{n_1 + n_2}$

$100\alpha\%$ level reject H_0 if

$|Z| = \left| \frac{p_1 - p_2}{\sqrt{\hat{\pi}(1-\hat{\pi})(\frac{1}{n_1} + \frac{1}{n_2})}} \right| > Z_{\alpha/2}$

Goodness of fit

$\chi^2_{obs} = \sum_{\text{all group } i} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2_v$

$df = v$ = numbers of groups − number of constraints

= number of groups − number of estimated parameter −1

$e_i \geq 5$: valid.

if it is not $e_i \geq 5$. merge some groups.

Summary

5

Contingency tables and independence.

df = $(R-1)(C-1)$

Summary

Seminar Sta. Revision

Stem-leaf diagram: interval should be suiteble

always 1. 2.5 2.

Quantile: $\frac{n+1}{4}$ exact position → number

$\frac{3(n+1)}{4}$ by the porportion of the distance between 2 numbers.

Outlier: $Q_1 - 1.5IQR$ $Q_3 + 1.5IQR$

Average: $\bar{x} = \frac{x_1 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_kx_k}{f_1 + f_2 + \cdots + f_k} = \frac{\sum f_i x_i}{\sum f_i}$$

if it is in interval: individual observations are not available. We assume that all the observations in a class interval fall at the mid-point of the interval.

m_1 m_2 .. m_k. frequencies f_1 f_2 .. f_k

$$\bar{x} = \frac{\sum f_i m_i}{\sum f_i}$$

for ungrouped data

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

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

$$= \frac{1}{n-1}(\sum x_i^2 - \frac{1}{n}(\sum x_i)^2)$$

Summary

for frequences data

$$\cancel{s^2 = \frac{1}{n-1}\left\{\sum f_i x\right.}$$

$$s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{(\sum f_i) - 1} = \frac{1}{n-1}\left\{\sum f_i x_i^2 - \frac{1}{n}\left(\sum f_i x_i\right)^2\right\}$$

$\boxed{\text{指的是} \sum f_i}$

QQ plot

① straight line ② intercept 0 ③ slope = 1

Confidence interval:

σ^2 is known: $\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

σ^2 is unknown: $\cancel{\bar{x} \pm Z_{\alpha/2}}$ $\bar{x} \pm t_{n-1, \alpha/2} \frac{s}{\sqrt{n}}$

width: $2 Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ $\quad 2 Z_{\alpha/2} \frac{s}{\sqrt{n}}$

Sta

176 A sample size of 64 is drawn by simple random sampling from a normal population which has vanture 4. The sample mean is 0.45.

(a) Test the hypothesis $H_0: \mu = 0$ vs $H_1: \mu \neq 0$ at 5% level of significance.

(b) Test the hypothesis $H_0: \mu = 0$ vs $H_1: \mu > 0$ at 5% significance level.

(c) Obtain the power of the test in (a) when $\mu = -1, -\frac{1}{2}, \frac{1}{2}, 1$ and so give a sketch of the power function.

相当于告诉了分布 $\sim N(\mu, \frac{1}{16})$

$-1 \quad -\frac{1}{2} \quad \frac{1}{2} \quad 1$

消同标准化.

Summary