Implicit differention

$F(x,y)=0$ find $\frac{dy}{dx}$.

let $w = F(x,y)$

$$\frac{dw}{dx} = \frac{\partial F}{\partial x}\frac{dx}{dx} + \frac{\partial F}{\partial y}\frac{dy}{dx} = 0$$

$$\frac{dy}{dx} = -\frac{\partial F}{\partial x}\Big/\frac{\partial F}{\partial y}, \quad -\frac{F_x}{F_y}$$

Transformation partial differential equation

One-dimension wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2\frac{\partial^2 u}{\partial x^2}$$

Surface

Curve $r(t)$. $F(x,y,z) = F(x(t), y(t), z(t)) = 0$ $\frac{dF}{dt} = 0$

$$\frac{dF}{dt} = \frac{\partial F}{\partial x}\cdot\frac{dx}{dt} + \frac{\partial F}{\partial y}\cdot\frac{dy}{dt} + \frac{\partial F}{\partial z}\cdot\frac{dz}{dt} = 0$$

Scalar product.

$$\left(\frac{\partial F}{\partial x}i + \frac{\partial F}{\partial y}j + \frac{\partial F}{\partial z}k\right)\cdot\left(\frac{dx}{dt}i + \frac{dy}{dt}j + \frac{dz}{dt}k\right) = 0$$

Define a vector ∇F.

$$\nabla F = \left(\frac{\partial F}{\partial x}i + \frac{\partial F}{\partial y}j + \frac{\partial F}{\partial z}k\right)$$

gradient.

→ tangent vector

$$\boxed{\nabla F}\cdot\boxed{\frac{dr}{dt}} = 0$$

∇F is a normal vector to S at p.

$\boxed{\text{Tangent plane}}$: one plane through p perpendicular to ∇F.

整个推导类似2.12中

- tangent line $\frac{dy}{dx}$
- normal line perpendicular to the tangent

☆

批注：题目中的向量看清楚是不是unit单位向量.

Summary

5

Surfaces in parametric coordinates

Surface $r(u,v)$.

$r_u \times r_v$ is a normal vector to S at P

(cross product)

Taylor theorem $f(x,y)$

$$f(a+h,\ b+h) = \sum_{r=0}^{n} \frac{1}{r!} \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^r f(x,y)\Big|_{a,b} + R_n$$

Remainder: $R_n = \frac{1}{(n+1)!} \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^{n+1} f(x,y)\Big|_{a+th,\ b+tk}$

$(0 \le t \le 1)$

$$f(a+h,\ b+h) = f(a,b) + (hf_x + kf_y) + \frac{1}{2!}(h^2 f_{xx} + 2hk f_{xy} + k^2 f_{yy})$$

Linear approximation

$f(x,y) \simeq f(a,b) + (x-a)f_x(a,b) + (y-b)f_y(a,b) = L(x,y)$

$z = L(x,y)$ is linear in x y and is a plane

↳ tangent plane

Quadratic Approximation

$f(x,y) \simeq f + (x-a)f_x + (x-b)f_y + \frac{1}{2}[(x-a)^2 f_{xx} + 2(x-a)(x-b)f_{xy} + (y-b)^2 f_{yy}] = Q(x,y)$

Stationary point

Definition: $\boxed{f_x = f_y = 0}$, (a,b) is a stationary point of $f(x,y)$

Summary

5

Local Maximum and Local minima

maximum if $f(x,y) \le f(a,b)$

minima if $f(x,y) \ge f(a,b)$

Saddle point

(a,b) is a saddle point of $f(x,y)$ if there is a point (x_1,y_1) in D where $f(x_1,y_1) > f(a,b)$ and a point (x_0,y_0) in D where $f(x_0,y_0) < f(a,b)$

A saddle point is a stationary point which is not a maximum nor a local minimum

$\boxed{\text{Hessian}}$

$H = f_{xx}f_{yy} - f_{xy}^2$

If $H<0$, f has a saddle point

If $H>0$. if $f_{xx}>0$, f has a $\boxed{\text{local}}$ minimum at (a,b)

if $f_{xx}<0$, f has a $\boxed{\text{local}}$ maximum at (a,b)

$H=0$. no conclusion.

扣：Hessian 是用来判断 stationary point 类型的 而不是求 stationary point 的.

求 stationary point $f_x=0$. $f_y=0$.

Lagrange multiplier

$F(x,y,\lambda) = f(x,y) - \lambda g(x,y)$ $\quad \nabla f = \lambda \nabla g(x,y)$

solve $F_x = F_y = F_\lambda = 0$

Generalisation

$F(x_1, \dots x_n, \lambda_1, \dots \lambda_m) = f - \sum_{k=1}^{m} \lambda_k g_k$

- 先求 stationary point let $f_x=0$ $f_y=0$
- Hessian 判断.

Double integral

$\iint_R f(x,y)\,dx\,dy$

$A = \iint_R 1\,dx\,dy$ Area of R

Summary

Average and centroids

$$\bar{f} = \frac{1}{n}\sum_{i=1}^{n} f_i$$

$$\frac{\sum_{i=1}^{n} f_i w_i}{\sum_{i=1}^{n} w_i}$$

$$\bar{f} = \frac{\iint_R f(x,y)\,dx\,dy}{\iint_R dx\,dy}$$

$$\bar{x} = \frac{\iint_R x\,dx\,dy}{\iint_R dx\,dy} \qquad \bar{y} = \frac{\iint_R y\,dx\,dy}{\iint_R dx\,dy}$$

$(\bar{x}, \bar{y})$ centroid.

Change of variable and Jacobian

$x = x(u,v) \qquad y = y(u,v)$

$$\frac{\partial(x,y)}{\partial(u,v)} = \left(\frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}\right)$$

$$= \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

$$I = \iint_S f(x(u,v), y(u,v)) \left|\frac{\partial(x,y)}{\partial(u,v)}\right| du\,dv$$

Useful Property.

$$\frac{\partial(u,v)}{\partial(x,y)} = 1 \Big/ \frac{\partial(x,y)}{\partial(u,v)}.$$

Summary

Triple integral

$$\iiint_V f(x,y,z)\,dx\,dy\,dz$$

$$\text{Volumn} = \iiint_D dV$$

$$J = \frac{\partial(x,y,z)}{\partial(u,v,w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$$

given $x = x(u,v,w)$ $y = y(u,v,w)$, $z = z(u,v,w)$

Integration Technicals:

DI Method:

Theorem: $\int f(x)g(x)\,dx = \sum_{j=0}^{n-1} (-1)^j f^{(j)}(x)\, g^{(-(j+1))}(x) + (-1)^n \int f^{(n)}(x)\, g^{(-n)}(x)\,dx$

Collary: Suppose that $f^{(n)}(x) \equiv 0$

$$\int f(x)g(x)\,dx = \sum_{j=0}^{n-1} (-1)^j f^{(j)}(x)\, g^{(-(j+1))}(x) + C$$

D	I
f	g
$f^{(1)}$	$g^{(-1)}$
$f^{(2)}$	$g^{(-2)}$
$\vdots$	
$0 = f^{(n)}$	g^{-n}

$+ \; - \; + \; - \; + \; -$

Summary

eg: $\int x^3 \cos x \, dx$

D	I
x^3	$\cos x$
$3x^2$	$\sin x$
$6x$	$-\cos x$
6	$-\sin x$
0	$\cos x$

$$\int x^3 \cos x \, dx = x^3 \sin x - 3x^2(-\cos x) + 6x(-\sin x) - 6\cos x + C.$$

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$$

$[x]$

$[2.5] = 2$

$[3.3] = 3.$

$$\int_1^\infty e^{-[x]}dx = \int_1^2 e^{-[x]}dx + \int_2^3 e^{-[x]}dx + \dots +$$

$$= \int_1^2 e^{-1}dx + \int_2^3 e^{-2}dx + \int_3^4 e^{-3}dx + \dots$$

$$= e^{-1}\int_1^2 dx + e^{-2}\int_2^3 dx$$

$$= e^{-1} + e^{-2} + e^{-3} + \dots$$

$$= \lim_{n\to\infty} \frac{e^{-1}(1-(\frac{1}{e})^n)}{1-(\frac{1}{e})} = \frac{e^{-1}}{1-e^{-1}} = \frac{1}{e-1}$$

Advanced substitution

$$I = \int \frac{2\sin x - \cos x}{\sin x + 2\cos x}\,dx$$

$$= A\int \frac{\sin x + 2\cos x}{\sin x + 2\cos x}\,dx + B\int \frac{\cos x - 2\sin x}{\sin x + 2\cos x}\,dx$$

$$2\sin x - \cos x = A(\sin x + 2\cos x) + B(\cos x - 2\sin x)$$

$$(A-2B)\sin x + (2A+B)\cos x$$

$$\begin{cases} A - 2B = 2 \\ 2A + B = -1 \end{cases} \quad \begin{cases} A = 0 \\ B = -1 \end{cases}$$

Summary

Useful facts:

$$\lim_{x\to 0} \frac{\sin x}{x} = 1 \qquad \lim_{x\to 0} \frac{e^x - 1}{x} = 1$$

$$\lim_{x\to 0} \frac{1-\cos x}{x^2} = \frac{1}{2} \qquad \lim_{x\to a} \frac{x^n - a^n}{x-a} = n a^{n-1}$$

$$\lim_{x\to 0} \frac{\ln(1+x)}{x} = 1 \qquad \lim_{x\to\infty} (1+\frac{1}{x})^x = e$$

Trapezium Rule. with error

$$\int_a^b f(x)\,dx \simeq \frac{h}{2}[f_0 + 2[f_1 + \dots + f_{n-1}] + f_n]$$

$h = (b-a)/n$

$$|E_T| \le \frac{M(b-a)^3}{12n^2} \quad M = \max_{a\le\xi\le b} |f''(\xi)|$$

n^{th} derivative and stationary point

Theorem: lowest nonvanishing derivative of f at stationary point is $f^{(k)}(x_0)$, $k \geq 2$, $f^{(k+1)}(x_0)$ exist.

(i) k is even $f^{k}(x_0) < 0$, $\Rightarrow x_0$ is maximum

(ii) k is even $f^{k}(x_0) > 0$ $\Rightarrow$ x_0 is minimum

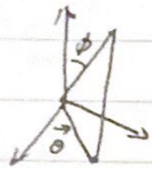
(iii) k is odd $\Rightarrow$ x_0 is a point of inflection.

Summary

注意顺序

课堂上和试卷答的顺序是(ρ. θ. φ)

Spherical coordinate



$(\rho,\ \phi,\ \theta)$

$\phi \in [0, \pi]$, $\theta \in [0, 2\pi)$

eg: $(0,\ 2\sqrt{3},\ -2)$ in rectangular coordinate.

$$\begin{cases} \rho = \sqrt{x^2+y^2+z^2} & \phi \in [0, \pi] \\ \rho \sin\phi \cos\theta = x & \theta \in [0, 2\pi) \\ \rho \sin\phi \sin\theta = y \\ \rho \cos\phi = z \end{cases}$$

$$\rho = \sqrt{0^2 + (2\sqrt{3})^2 + (-2)^2} = 4$$

$$\cos\phi = \frac{z}{\rho} = \frac{-2}{4} = -\frac{1}{2}. \quad \phi = \frac{2\pi}{3}$$

$$2\sqrt{3} = 4 \times \frac{\sqrt{3}}{2} \times \sin\theta. \quad \sin\theta = 1 \quad \theta = \frac{\pi}{2}$$

$$\left(4,\ \frac{2\pi}{3},\ \frac{\pi}{2}\right)$$

注意代数式的导数几次求导之后会为0的情况

Leibniz's Theorem

$$f^{(n)} = \sum_{r=0}^{n} \binom{n}{r} u^{r} v^{n-r}$$

右边两可拆看

$$y'' - 2y' + y = e^{x} + e^{-x}$$

通解: $y_{cp} = Ae^{x} + Bxe^{-x}$.

特解:
$$\begin{cases} e^{x} \to \text{重根} & y_{p1} = Ax^2e^{x} \\ e^{-x} \to \text{不是根} & y_{p2} = Be^{-x} \end{cases}$$

$$y = y_{cf} + y_{p2}^{①} + y_{p2}^{②}$$
$$= Ae^{x} + Bxe^{-x} + \frac{1}{2}x^2e^{x} + \frac{1}{4}Be^{-x}$$

Summary

formal definition to get partial derivative

$$f(x,y) = \begin{cases} \dfrac{x^3+x^4-y^3}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

$$\frac{\partial f}{\partial x}(a,b) = \lim_{h\to 0} \frac{f(a+h,b)-f(a,b)}{h}$$

$$\frac{\partial f}{\partial y}(a,b) = \lim_{h\to 0} \frac{f(a,b+h)-f(a,b)}{h}$$

想计算除 (0,0) 以外的点，用公式

但想求 $\frac{\partial f}{\partial x}(0,0)$，必须用定义 $f(0,0)=0$

$$f(h,0) = \frac{h^3+h^4-0^3}{h^2+0^2} = \frac{h^3+h^4}{h^2} = h+h^2$$

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0} \frac{f(0+h,0)-f(0,0)}{h} = \lim_{h\to 0} \frac{f(h,0)-f(0,0)}{h}$$

$$= \lim_{h\to 0} \frac{h+h^2-0}{h}$$

$$= \lim_{h\to 0} (1+h)$$

$$= 1$$

Hession: $= f_{xx}f_{yy} - f_{xy}^2$

$\boxed{H<0}$, saddle point

$H>0$, $f_{xx}<0$ local maxim

$f_{xx}>0$ local mini

$H=0$ no conclusion.

Summary

5

Lagrange Multipliers 考虑间值
代回原函数找最值.

二重积分互换 dx dy. 表示方向也要换.

eg. $\int_0^3 \int_{\sqrt{x/3}}^1 e^{y^3}\,dy\,dx$

$\int_0^1 \int_0^{3y^2} e^{y^3}\,dx\,dy$ 最好画个图

三重积分. z, y. 分别在3D图和2D中寻找积分域
· 最后一个积分元变量
· from inside to outside.

Jacobian 代进去时是绝对值 |J|.

画 domain 时, 如果是圆, 指出圆半径和圆心

判断方程形状, let z=0. 俯视图. x, y=0. 侧视图
eg. $z=\sqrt{x^2+y^2}$. cone. 圆锥.

Taylor 的误差
$$R_n = \frac{1}{(n+1)!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\right)^{n+1} f(x,y)\Big|_{a+th,\, b+th} \quad (0\le t\le 1)$$

Lagrange 算距离记得最后开平方

Summary

5

积分常数不要忘记

换元，积分域也要变，记得代回原值

Euler homogenous

$y = vx \qquad \frac{dy}{dx} = \frac{dv}{dx} x + v$

记得注明取值范围；分母不为0 etc.

三角积分，用的是θ，原来是无θ的.

最终将其换回去

$\sqrt{a^2-x^2} \qquad x = a\sin\theta$

$x^2+a^2 \qquad x = a\tan\theta$

$\sqrt{x^2-a^2} \qquad x = a\sec\theta$

$\sqrt{\frac{a-x}{a+x}} \qquad x = a\cos 2\theta$

技法：三角换元
常用.

$$\int \sec dx = \int \sec x \cdot \left(\frac{\sec x + \tan x}{\sec x + \tan x}\right) dx$$

$$= \ln|\sec x + \tan x| + c$$

$$\int \csc dx = \int \csc x \cdot \left(\frac{\csc x - \cot x}{\csc x - \cot x}\right) dx$$

$$= \ln|\csc x - \cot x| + c$$

$$\int \frac{1}{\sqrt{x^2+a^2}} dx = \ln|x + \sqrt{x^2+a^2}| + c$$

Summary

$$\frac{1}{(x+a)(x+b)} = \frac{A}{x+a} + \frac{B}{x+b}$$

$$\frac{1}{(x+a)^2(x+b)} = \frac{A}{x+a} + \frac{B}{(x+a)^2} + \frac{C}{(x+b)}$$

$$\frac{1}{(x^2+a)(x+b)} = \frac{Ax+B}{x^2+a} + \frac{C}{x+b}$$

form $\quad \frac{1}{a+b\cos x+c\sin x} \quad \frac{1}{a+b\cos x} \quad \frac{1}{a+b\sin x}$

$\tan(\frac{x}{2}) = t \qquad \sin x = \frac{2t}{1+t^2}$

$dx = \frac{2dt}{1+t^2} \qquad \cos x = \frac{1-t^2}{1+t^2}$

$\tan x = \frac{2t}{1-t^2}$

$\int p(x) f(x)\, dx$

1). 2. Method.

$\downarrow \quad + \quad \downarrow \qquad + \; - \; + \; - \; +.$

$$\int \frac{1}{x^2+a^2}\, dx = \frac{1}{a}\tan^{-1}(\frac{x}{a}) + c$$

$$\int \frac{1}{\sqrt{a^2-x^2}}\, dx = \sin^{-1}(\frac{x}{a}) + c$$

$$\int \frac{1}{|x|\sqrt{x^2-a^2}} = \frac{1}{a}\sec^{-1}\left|\frac{x}{a}\right| + c$$

$$\int \frac{1}{\sqrt{x^2-a^2}} = \ln\left|x+\sqrt{x^2-a^2}\right| + c$$

Summary

5

$$\int \frac{1}{x^2-a}\,dx = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + c$$

$$\int \frac{1}{a^2-x^2}\,dx = \frac{1}{2a}\ln\left|\frac{x+a}{x-a}\right| + c$$

$$e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$\cos x = \sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

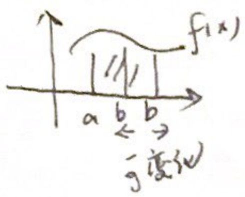
$$\sin x = \sum_{n=0}^{\infty}\frac{(-1)^n x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

可整体代换

$$\frac{1}{1+x} = \sum_{n=0}^{\infty}(-1)^n x^n = 1 - x + x^2 - x^3 + x^4 - \cdots$$

$$\ln(1+x) = \sum_{n=1}^{\infty}\frac{(-1)^{n+1} x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} +$$

Summary



变限积分.

$\int_a^b f(x)dx = F(b)-F(a)$

面积是区间的函数

积分值与积分变量无关.

$\varphi(x)=\int_a^x f(x)dx$

↓便于区分

$\varphi(x)=\int_a^x f(t)dt = F(x)-F(a)$

$\varphi'(x)=[F(x)-F(a)]' = F'(x) = f(x)$

推广 复合了 x^2

$\varphi(x)=\int_a^{x^2} f(t)dt = F(x^2)-F(a)$

$\varphi'(x)=[F(x^2)]' = 2x f(x^2)$

一般地:

$\Phi(x)=\int_{\varphi_1(x)}^{\varphi_2(x)} f(t)dt$

$\Phi'(x) = \varphi_2'(x)f(\varphi_2(x)) - \varphi_1'(x)f(\varphi_1(x))$

eg: $\frac{d}{dx}\left(\int_{3x}^{x^2} \sin t\, dt\right)$

$= 2x\sin x^2 - 3\sin 3x$

Summary

5

Statisits Summary:

Central limit theorem:

If $x_1, \cdots x_n$ are independent and identically distributed, with same mean μ and finial variance σ^2. Then $\bar{x}$ is approximately $N(\mu, \frac{\sigma^2}{n})$ for large n no matter what the distribution of the x_i.

iid

泊松的期望方差相等

Poisson distribution: $p\{x=k\} = \frac{\lambda^k}{k!} e^{-\lambda}$, $k=0, 1, 2 \cdots$

$E(x) = \lambda$

$var(x) = \lambda$

Exponetial distribution: $f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

$F_x(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

$E(x) = \frac{1}{\lambda}$

$var(x) = \frac{1}{\lambda^2}$

significance & p-value & power

type I

- significance level of a hypothesis $\alpha = p($reject H_0 | H_0 true$)$
- p-value is the probability of observing an outcome "more extreme" than what was actually observed given H_0 is true

$p($obtaining a test statistic at least as extreme as that observed | H_0 is true$)$

$1-\beta$

- power of a hypothesis test is $p($reject H_0 | H_0 false$)$.

Summary

Assumptions of least squre linear regression

(i) $E(\varepsilon_i) = 0$ for all $i = 1, \cdots n$

the error has zero mean

(ii) $var(\varepsilon_i) = \sigma^2$ for all $i = 1 \cdots n$

errors have the same variances

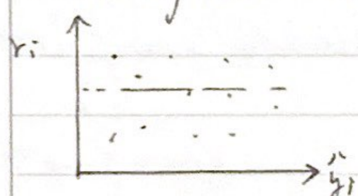
(iii) $Cor(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$

the errors are uncorrelated

residual plot

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

r_i against fitted values. $\hat{y}_i$



Contingency table

H_0: independent vs H_1: not independent

Exp: 不用再查。但和比较一样

Normal Approximation to Binomial Distribution

$n \geq 20$. $np \geq 5$ $n(1-p) \geq 5$

$X \approx N(np, np(1-p))$

$X = X_1 + \cdots + X_n$

Summary

proportion 用的是 z test.

$$\chi^2_{obs} = \sum \frac{(O_i - e_i)^2}{e_i} = \sum \frac{O_i^2}{e_i} - n$$ 公式技巧

Explaination for 0.2:

If the experiment were to be repeated multiple times and a 90% confidence interval was calculated in this way for each experiment, then 90% of the conficene interval would contain the true value of the parameter.

lower the confidence <=> increasing α

$\alpha\, z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

α↑	width ↑
n↑	↓
α↑	↓
1-α↑	↑
↓confidene	

QQ plot:

① Straight line ② Throught the origin ③ gradient 1

计算 degrees of fredom 是 计算整合之后的列.

列出上下表格的数据, 计算自由度看清用的 two sample 还是 difference. 如果不看清会导致巨大的错误.

Normal QQ plot

known μ, σ^2

1. $y_1, y_2 \cdots < y_n$
2. Standarlize y_i $z_{(i)} = \frac{y-\mu}{\sigma}$ 标准化 y_i
3. Find $F(x_i) = \frac{i-\frac{1}{2}}{n}$ (查表)
4. Plot z_i against x_i

① (0,0)
② straight line
③ Grad 1

Summary

unknow μ, σ^2

1. $y_1, \cdots y_n$
2. Find $F(x_i) = \frac{i-\frac{1}{2}}{n}$
3. plot
4. 观察图上的 interception 和 slope 即是 → μ 和 σ

5

看到题目中有 difference in the variance between … 想到 F-test.

Test for regression 本质为 test model 的斜率

$H_0: \beta_1 = \beta^* \text{ vs } H_1\ \beta_1 \neq \beta^*$

reject if

$$|t| = \left| \frac{\hat{\beta}_1 - \beta^*}{\hat{\sigma}/\sqrt{(n-1)S_x^2}} \right| > t_{n-2, \alpha/2}$$

题中可能对 β^* 换个表述: increasing by 1 …

χ^2 ~ test 公式推导

$$\sum_i \frac{(o_i - e_i)^2}{e_i} = \sum_i \left(\frac{o_i^2}{e_i} - 2o_i + e_i \right) = \sum \frac{o_i^2}{e_i} - 2\sum_i o_i + \sum_i e_i$$

$$= \sum_i \frac{o_i^2}{e_i} - 2n + n$$

$$= \sum_i \frac{o_i^2}{e_i} - n \rightarrow \text{总频数 不是 格子个数}$$

核心是抓住 o_i e_i 都是频数. 总和为 n

p-value 定义:

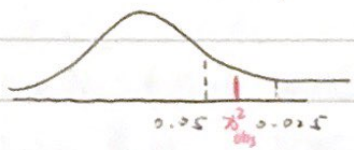
is the probability of observing a test statistic more extreme than observed assuming the null hypothesis is true.

p-value = p(obtaining a test statistic more extreme than observed | Ho is true)

Small p-value, implies there is evidence against the null hypothesis being true.

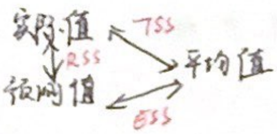
Summary

因为 p 值判断是一个具体范围
所以需要自己测试在多少 $\alpha\%$ significance level
会或不会拒绝 p. 可以得出 p 的范围.



$\Rightarrow 0.05 < p < 0.05$

⇒ 得出结论



Total Sum of squares: $TSS = \sum (y_i - \bar{y})^2 = \underset{\sum r_i^2}{(n-1) s_y^2}$ 总平方和

The residual sum of square: $RSS = \sum (y_i - \hat{y}_i)^2 = (n-2)\hat{\sigma}^2$ 残差平方和

The regression sum of square: $ESS = \sum (\hat{y}_i - \bar{y})^2$ 回归平方和

(explained sum of square)

$TSS = RSS + ESS$

$\sum r_i^2 = (n-1)[s_y^2 - \hat{\beta}_1^2 s_x^2]$ $= (n-1) \cdot (n-2)\hat{\sigma}^2 / (n-1)$ $= (n-2)\hat{\sigma}^2$	或 $\hat{\sigma}^2 = \frac{1}{n-2} \sum r_i^2$

proportion of the variance explained is $R^2 = \frac{ESS}{TSS}$

$\therefore r = \sqrt{R^2}$

批注: 有时间 use ← the proportion to determine the sample correlation between x and y.

Summary

TSS: how much variance there is in the dependent variable

RSS: how much of the variation your model did not explain

ESS: how much of the variation you model did explain

5

χ^2 goodness of fit / Contingency table 都是卡方检验

适用条件: valid for large sample.

$\boxed{e_i \geq 5}$

做假设是因为 sample size 太小了 未知 σ^2.

two population 检验中. 给出了 SD 不代表方差已知

仍需假设 σ^2 相等 $\rightarrow S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

$\Downarrow$

接着 检验 σ^2 相等这个假设. 用 F-test

这个题可谓完美的过渡

2015:

Standard deviation / appropriate no outlier, robust
not appropriate outlier

$(Q_1 - 1.5IQR,\ Q_3 + 1.5IQR)$

补充: 样本太小, 无法用正态分布.

二项分布:

$X \sim Bin(n, p)$

$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$

可以用计算器算, 也可以用表 但是要注意有没有等号

$P(X>5)$

$= 1 - P(X \leq 5)$ 因为是

$= 1 - P(X \leq 4)$ 离散的

Summary

interception: μ

slope: σ

$y(x) = \sigma z + \mu$

QQ plot 画正态分布（未知方差和均值）

利用公式 + table　　不已知方差和均值，先将样本排序并标准化，进行对比判断

$F(x_i) = \frac{i-\frac{1}{2}}{n}$

$\Phi(x_i) = \frac{i-\frac{1}{2}}{n}$　　① (0,0) ② straight line ③ slope 1

$x_i = \Phi^{-1}(\frac{i-\frac{1}{2}}{n})$　　during 取值对称性，Z值是对称的

加个加负号的问题

画出一个大致直线 估计即可

做假设检验需要的假设前提

Test for means: σ know (Z-test)

Assume: n is large ($n \geq 30$) OR Data are normally distributed

Test for means: σ unknow (t-test)

Assume: $X_1 \cdots X_n \sim iid\ N(\mu, \sigma^2)$

Two populations

Case I: Assume $\sigma_1^2\ \sigma_2^2$ known [OR $n_1\ n_2$ large]　　选择方差就 $\approx SD^2$

Case II: Assume $\sigma_1^2\ \sigma_2^2$ unknown but equal

Case III: Assume $\sigma_1^2\ \sigma_2^2$ unknown unequal.

可进一步引入 ⇐ F-test σ 是否 valid

⇓

If $n_1\ n_2$ large ($n_1 \geq 30$, $n_2 \geq 30$) treat $\sigma_1\ \sigma_2$ as known using $s_1\ s_2$ ⇒ Case I

If $\sigma_1\ \sigma_2$ too different ⇒ Case II.

Summary

5

Pair samples

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

normally distributed and matched pairs

If there is a positive relationship between the scores in a paired situation, then matched pairs test has a $\boxed{\text{higher power}}$ than the independent sample test. (i.e. has higher probability of correctly rejecting H_0 when H_0 is ~~true~~ false.

Inference about Variance σ^2

· Chi-square test

Assume: $X_1, \cdots X_n \sim iid\ N(\mu, \sigma^2)$

· F-test

Assume: $\left.\begin{array}{l} X_1\ X_2 \cdots X_{n_x} \sim N(\mu_x, \sigma_x^2) \\ Y_1\ Y_2 \cdots Y_{n_y} \sim N(\mu_y, \sigma_y^2) \end{array}\right\}$ independent

Correlation not means causality

(i.e. do not imply that one has any direct or indirect effect on the other)

Summary

T-test for linear correlation

Assume: $(x_1, y_1) \cdots (x_n, y_n)$ are normally distributed

Require X, Y have a linear relationship

Non-parametric test for linear correlation

Assume: non-linear H_0: no correlation H_1: some correlation

Least squares linear regression

1. $E(\varepsilon_i) = 0$ $i=1 \cdots n$ errors have zero mean
2. $cor(\varepsilon_i, \varepsilon_j) = 0$ $i \neq j$ errors are uncorrelated
3. $var(\varepsilon_i) = \sigma^2$ $i = 1 \cdots n$ have same variances

Assume: Data do not have to be normally distributed

· Test for the existence of a regression $H_0: \beta = 0$

· Inference for individual / mean $H_1: \beta \neq 0$

Assume: Data are normally distributed

Why individual + 1

(An individual observation will have greater variability than a mean value)

extra error should be taken into account

Proportion two samples

Assume: two samples are independent and n_1 n_2 are large

Summary

5

Spearman rank order

H_0: no correlation H_1: some correlation

线性方程 最小二乘法 的公式推导

$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i$

$S = \sum \varepsilon_i^2 = \sum (y_i - (\beta_0 + \beta_1 x_i))^2$

Minimum S

$$\frac{\partial S}{\partial \beta_0} = \frac{\partial}{\partial \beta_0} \sum_{i=1}^{n} -2 (y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial S}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_{i=1}^{n} -2 x_i (y_i - \beta_0 - \beta_1 x_i) = 0$$

残差

$$\begin{cases} \sum (y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow \sum \hat{e}_i = 0 \\ \sum x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \Rightarrow \sum \hat{e}_i x_i = 0 \end{cases}$$

$$\rightarrow \begin{cases} \sum y_i - n\beta_0 - \beta_1 \sum x_i = 0 \\ \sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 = 0 \end{cases}$$

先推出 β_0

$$\beta_0 = \frac{1}{n} (\sum y_i - \beta_1 \sum x_i)$$
$$= \bar{y} - \hat{\beta}_1 x_i$$

$$\sum x_i y_i - (\bar{y} - \hat{\beta}_1 x_i) \sum x_i - \beta_1 \sum x_i^2 = 0$$

$$\beta_1 (\sum x_i^2 - \bar{x} \sum x_i) = \sum x_i y_i - \bar{y} \sum x_i$$

$$\beta_1 = \frac{\sum x_i y_i - \bar{y} \sum x_i}{\sum x_i^2 - \bar{x} \sum x_i}$$

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

Summary