2020级数据科学基础期末考回忆版

(考试时间: 2022.1.2)

(PS:回忆版为纯人肉记忆,可能存在错误,见谅)

一、

变量W=(aX+3Y)²,已知E(X)=E(Y)=0,D(X)=4,D(Y)=16,求a取何值时E(W)最小及E(W)的最小值

二、

X, Y的联合概率密度为f(x,y)=12e^(-3x-4y), x>0,y>0

(1)求P{0<X<1, 0<Y<2}

(2)X,Y是否独立

三、

(秩和检验的题,数据忘了呜呜)

(就是单纯求个秩和,简单的)

四、

 X_1 , X_2 ,, X_n 是来自分布 $N(\mu,\sigma^2)$ 的样本, μ 已知, σ 未知,试证明 $\frac{\sum_{i=1}^n(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$,并求, σ^2 的置信水平为 $1-\alpha$ 的置信空间

五、

试证明:

- (1)信息熵的最大值为logn
- (2)当p₁=p₂=...=p_n=1/n时,信息熵取到最大值logn

六、

使用PCA降维以下数据

t1	1	1	2	4	2
t2	1	3	3	4	4

七、

X~N(0,1), 求Y=2X²+1的概率密度

八、

证明在单因素方差分析中,对原数据 x_{ij} 进行线性变换 $x_{ij} = \frac{x_{ij} - a}{b}$ 后(其中a, $b \neq 0$),不影响最终F的值九、

(回归方程的题,直接代数据就行,具体数据忘了呜呜)

貌似是下面的数据:

 $I_{xx} = 4$, $I_{xy} = 8$, $\overline{y} = 20$, $\overline{x} = 0.2$,求一元线性回归方程 $y = \beta_1 + \beta_2 x$

+、

已知X的概率密度为 $f(x)=\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}$, x>0, 求E(X)和D(X)