

华东理工大学 2014 - 2015 学年第一学期

《数理统计》课程期末考试试卷 B 2015.01

开课学院：理学院，专业：数学, 信计，考试形式：闭卷，所需时间 120 分钟

考生姓名：_____ 学号：_____ 班级 _____ 任课教师 朱坤平

题序	一	二	三	四	五	六	七	总分
得分								
评卷人								

附表： $\chi_{0.95}^2(4) = 9.488$, $t_{0.95}(13) = 1.7709$, $t_{0.975}(4) = 2.776$, $t_{0.975}(13) = 2.1604$,

$F_{0.975}(6, 7) = 5.12$, $F_{0.975}(7, 6) = 5.70$, $F_{0.95}(2, 6) = 5.14$, $F_{0.95}(2, 17) = 3.59$

一. 单选题 (每小题 3 分, 共 39 分)

1. 设总体 $X \sim N_m(\mu, \Sigma)$, 样本协方差矩阵为 S , 则 (C)
- A. S 是 Σ 的矩法估计 B. S 是 Σ 的极大似然估计
- C. S 是 Σ 的无偏估计 D. $S = \Sigma$

2. 在正交试验设计中, 若有 2 个因子, 每个因子各有 3 个水平, 不考虑交互作用, 应该选取的正交表为 (A)

- A. $L_9(3^4)$ B. $L_{27}(3^{13})$
- C. $L_4(2^3)$ D. $L_{12}(3 \times 2^4)$

3. 以下可能导致复共线性问题的选项是 (B)

- A. 观测次数 n 大大超过自变元个数 m B. 自变元间线性相关
- C. 自变元间相互独立 D. 回归模型不显著

4. 设 (X_1, X_2, \dots, X_n) 为取自正态总体 $\xi \sim N(\mu, \sigma^2)$ 的样本. 错误的选项是 (C)

- A. 样本 (X_1, X_2, \dots, X_n) 服从 n 维正态分布 B. 样本均值 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

- C. $P\{X_1 > \mu, X_2 > \mu, \dots, X_n > \mu\} = 0.5$ D. $P\{\bar{X} \geq \mu\} = 0.5$

5. 设 Y_1, Y_2, Y_3, Y_4 均取自于总体 $Y \sim N(0,1)$ 的样本, 则有 (A)

A. $\frac{Y_2^2 + Y_3^2 + Y_4^2}{3Y_1^2} \sim F(3,1)$

B. $\frac{3Y_1^2}{Y_1^2 + Y_2^2 + Y_3^2} \sim F(1,3)$

C. $\frac{Y_1}{Y_2} \sim t(1)$

D. $\frac{Y_1}{\sqrt{(Y_1^2 + Y_2^2 + Y_3^2)/3}} \sim t(3)$

6. 对同一个总体参数的区间估计, 算得的置信区间长度最大的是 (D)

A. $n=100$, 置信水平=0.90

B. $n=100$, 置信水平=0.95

C. $n=90$, 置信水平=0.90

D. $n=90$, 置信水平=0.95

7. 关于变元 X 与 Y 的线性回归问题, 若总离差平方和 $SST=100$, 残差平方和 $SSE=19$, 则下列选项可能**错误**的是 (C)

A. 回归平方和 $SSR=81$

B. 判定系数 R^2 为 0.81

C. 变元 X 与 Y 的样本相关系数为 0.9

D. 因变元 Y 与其回归估计值 \hat{Y} 的样本相关系数为 0.9

8. 主成分的下列选项,**错误**的是 (D)

A. 主成分的个数不超过原变量的个数

B. 各主成分互不相关

C. 所有主成分的贡献率之和等于 1

D. 所有主成分的修正样本方差之和等于 1

9. 假设检验中的显著性水平 α 一般选取为 0.01 或 0.05, 如果你是厂家的老板, 质检局要对你厂产品是否符合国家标准进行检验, 对你有利的是 (A)

A. 选取 α 为 0.01

B. 选取 α 为 0.05

C. 都一样

D. 不能确定

10. 在显著性水平 0.01 下要检验数字 0-9 在圆周率 $\pi = 3.1415926 \dots$ 中出现的概率是否相同, 可统计圆周率的前 n 位中, 数字 0-9 分别出现的次数, 并据此作分布的拟合检验, 那么该检验的原假设 H_0 及 H_0 的接受域分别为 (C)

A. H_0 : 各数字出现的概率相同, $W_0 = [0, \chi_{0.995}^2(9)]$

B. H_0 : 各数字出现的概率不同, $W_0 = [0, \chi_{0.99}^2(n-1)]$

C. H_0 : 各数字出现的概率相同, $W_0 = [0, (\chi_{0.99}^2(9))]$

D. H_0 : 各数字出现的概率不同, $W_0 = [0, (\chi_{0.995}^2(n-1))]$

11. 在一个确定的假设检验中, 与判断结果有关的因素是 (D)
- A. 样本值与样本容量 B. 显著性水平 α
- C. 检验统计量 D. 三个选项 A,B,C 都有关

12. 总体 $X \sim N(\mu, \sigma^2)$, σ^2 已知, 要使总体均值 μ 的置信水平为 0.95 的置信区间长不大于 L , 则 $n \geq$ _____ (B)
- A. $15 \sigma^2 / L^2$ B. $15.3664 \sigma^2 / L^2$
- C. $16 \sigma^2 / L^2$ D. 16

13. 设总体 $X \sim B(1, p)$, (X_1, X_2, \dots, X_n) 是来自 X 的样本, 则有 (C)
- A. 当 n 充分大时, 近似有 $\bar{X} \sim N\left(p, \frac{p(1-p)}{\sqrt{n}}\right)$
- B. $P\{\bar{X} = k\} = C_n^k p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n$
- C. $P\{\bar{X} = \frac{k}{n}\} = C_n^k p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n$
- D. $P\{X_i = k\} = C_n^k p^k (1-p)^{n-k}, 1 \leq i \leq n$

二.(6 分) 设总体 X 服从参数为 λ 的泊松分布, $P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \dots)$

(X_1, \dots, X_n) 是 X 的简单随机样本, \bar{X} 为样本均值, S^{*2} 为修正样本方差. 试证明:

$\frac{1}{2}(\bar{X} + S^{*2})$ 是参数 λ 的无偏估计

证明:

$$E\left[\frac{1}{2}(\bar{X} + S^{*2})\right] = \frac{1}{2}E(\bar{X} + S^{*2}) = \frac{1}{2}[E(\bar{X}) + E(S^{*2})] = \frac{1}{2}(\lambda + \lambda) = \lambda$$

故有 $\frac{1}{2}(\bar{X} + S^{*2})$ 是参数 λ 的无偏估计

二 “设 $\hat{\theta}$ 是参数 θ 的无偏估计, 试证明: 若 $D(\hat{\theta}) \neq 0$, 则 $(\hat{\theta})^2$ 不是 θ^2 的无偏估计

证明: 反证法, 若 $(\hat{\theta})^2$ 不是 θ^2 的无偏估计, 即 $E(\hat{\theta})^2 \neq \theta^2$

又 $\hat{\theta}$ 是参数 θ 的无偏估计, 即 $E(\hat{\theta}) = \theta$

于是有 $D(\hat{\theta}) = E(\hat{\theta})^2 - [E(\hat{\theta})]^2 = (\theta)^2 - [\theta]^2 = 0$, 与 $D(\hat{\theta}) \neq 0$ 矛盾.

故 $(\hat{\theta})^2$ 不是 θ^2 的无偏估计

三. (10 分) 用三台电压表来测量某线路的电压,测得数据如下:

电压表	电 压 读 数			
A1	100.5	101.2	100.4	
A2	100.2	100.4	100.5	100.4
A3	100.2	99.8		

试通过方差分析说明在显著性水平 $\alpha = 0.05$ 下,三个电压表的读数有无显著差异?

解: 三个电压表的读数互相独立, 都服从正态分布, 且有相同的方差

设三个电压表的读数分别为 $\xi_i \sim N(\mu_i, \sigma^2) \quad i = 1, 2, 3$

$$H_0: \mu_1 = \mu_2 = \mu_3 \quad \text{-----}2'$$

$$SS_A = \sum_{i=1}^r m_i (\bar{X}_i - \bar{X})^2 = 0.5925;$$

$$SS_e = \sum_{i=1}^r SS_i = \sum_{i=1}^r \sum_{j=1}^t (X_{ij} - \bar{X}_i)^2 = 0.5075 \quad \text{-----}4'$$

方差分析: 单因素方差分析						
SUMMARY						
组	观测数	求和	平均	方差		
行 1	3	302.1	100.7	0.19		
行 2	4	401.5	100.375	0.015833		
行 3	2	200	100	0.08		
方差分析						
差异源	SS	df	MS	F	P-value	F crit
组间	0.5925	2	0.29625	3.502463	0.098204	5.143253
组内	0.5075	6	0.084583			
总计	1.1	8				

因 $F = 3.5025 < F_{1-\alpha}(r-1, n-r) = F_{0.95}(2, 6) = 5.14$

故接受 H_0 , 认为三个电表的读数没有显著性差异 -----4'

四. (10 分) 我们学过的许多统计方法(比如方差分析, 回归分析 等)都是以正态性, 独立性和方差齐性为前提的. 简述如何根据样本数据来判断这些前提是否成立?

解: 可按如下答案要点展开, 每个性质答对一个要点得 1 分, 过程说明得 2 分:

- 1) 正态性: 分布拟合检验; 概率纸检验; P-P 图 等
- 2) 独立性: 独立性检验; (残差的) 散点图
- 3) 方差齐性: 样本方差比较; F 检验; 残差图

五. (11 分) 设 (X_1, X_2, \dots, X_n) 为总体 X 的一个样本, X 的密度函数为

$$f(x) = \begin{cases} \beta x^{\beta-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}, \quad \beta > 0.$$

试求参数 β 的矩估计量和极大似然估计量.

解: (1) $E\xi = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x\beta x^{\beta-1}dx = \frac{\beta}{\beta+1}.$

解方程 $\frac{\beta}{\beta+1} = \bar{X}$, 得到矩法估计 $\hat{\beta} = \frac{\bar{X}}{1-\bar{X}}$ ----5'

(2) 先求似然函数:

$$L = \prod_{i=1}^n f(x_i) = \begin{cases} \prod_{i=1}^n \beta x_i^{\beta-1} = \beta^n \prod_{i=1}^n x_i^{\beta-1} & 0 < x_i < 1 \ (i=1, 2, \dots, n) \\ 0 & \text{其他} \end{cases}$$

当 $L \neq 0$ 时, 对 L 取对数, 得到 $\ln L = n \ln \beta + (\beta - 1) \sum_{i=1}^n \ln x_i$

解方程 $\frac{d \ln L}{d \beta} = \frac{n}{\beta} + \sum_{i=1}^n \ln x_i = 0$, 得到极大似然估计:

$$\hat{\beta} = \frac{-n}{\sum_{i=1}^n \ln x_i} = \frac{-1}{\frac{1}{n} \sum_{i=1}^n \ln x_i} = \frac{-1}{\ln \bar{X}} \quad \text{----6'}$$

六. (11 分) 顺通出租汽车公司为确定合理的管理费用, 需要研究出租车司机每天的收入

(元) 与他的行驶时间 (小时) 和行驶里程 (千米) 之间的关系, 为此随机调查了 20 个出租车司机, 根据每天的收入 (y)、行驶时间 (x_1) 和行驶里程 (x_2) 的有关数据进行线性回归, 得到下面的有关结果:

方程的截距 $b_0 = 42.38$	截距的标准差 $s_{b_0} = 36.59$	回归平方和 SSR=29882
回归系数 $b_1 = 9.16$	回归系数的标准差 $s_{b_1} = 4.78$	剩余平方和 SSE=5205
回归系数 $b_2 = 0.46$	回归系数的标准差 $s_{b_2} = 0.14$	

根据上面的部分计算结果回答下面的问题:

- 1). 写出每天的收入 (y) 与行驶时间 (x_1) 和行驶里程 (x_2) 的线性回归方程
- 2). 求判定系数 R^2 和 估计标准误差 $\hat{\sigma}$
- 3). 问在显著性水平 $\alpha = 0.05$ 下, 能否认为回归模型的线性关系是显著的?

解: 1) 回归方程为: $y = 42.38 + 9.16x_1 + 0.46x_2$. ---2'

2) 判定系数 $R^2 =$ 回归平方和/总离差平方和. 即:

$$R^2 = \frac{SSR}{SST} = \frac{SSR}{SSR + SSE} = \frac{29882}{29882 + 5205} = 0.8517 \quad \text{---2'}$$

估计标准误差为:

$$\hat{\sigma} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - m - 1}} = \sqrt{\frac{SSE}{n - m - 1}} = \sqrt{\frac{5205}{20 - 2 - 1}} \approx 17.4979 \quad \text{----2'}$$

3). 这是一个线性关系的显著性检验问题.

原假设 H_0 : 线性关系不显著, 备选假设 H_1 : 线性关系显著

$$F = \frac{SSR / k}{SSE / (n - k - 1)} = \frac{29882 / 2}{5205 / (20 - 2 - 1)} = \frac{29882 / 2}{5205 / 17} = \frac{14941}{306.1765} = 48.80$$

原假设的拒绝域为 $F > F_{1-\alpha}(2, 17) = 3.59$. 因 $48.8 > 3.59$,

故拒绝原假设, 即线性关系显著. -----5'

七. (13分) 甲、乙两台机床加工同一种零件, 从这两台机床加工的零件中, 随机抽取一些样品, 测得它们的外径 (单位: mm) 如下:

机床甲	20.5,	19.8,	19.7,	20.4,	20.1,	20.0,	19.0,	19.9
机床乙	19.7,	20.8,	20.5,	19.8,	19.4,	20.6,	19.2	

根据抽样数据算得: 甲已两机床零件的样本均值分别为 $\bar{X} = 19.925$, $\bar{Y} = 20.000$; 样本修

正标准差分别为 $S_x^* = 0.46522$, $S_y^* = 0.62981$. 假定零件的外径服从正态分布, 问:

- (1) 是否可以认为两台机床加工零件外径的方差相等? (显著水平 $\alpha = 0.05$)
- (2) 能否认为甲机床加工零件的外径显著地小于乙机床加工零件的外径? ($\alpha = 0.05$)
- (3) 求甲已两机床加工的零件外径之差水平为 95% 的置信区间.

解 设甲、乙两台机床加工零件的外径分别为 $\xi \sim N(\mu_1, \sigma_1^2)$ 和 $\eta \sim N(\mu_2, \sigma_2^2)$ 。

$$m = 8, \quad \bar{X} = 19.925, \quad S_x^* = 0.46522, \quad S_x^{*2} = 0.21643;$$

$$n = 7, \quad \bar{Y} = 20.000, \quad S_y^* = 0.62981, \quad S_y^{*2} = 0.39667。$$

(1) 问题相当于要检验 $H_0: \sigma_1^2 = \sigma_2^2$ 。

$$F = \frac{S_x^{*2}}{S_y^{*2}} = \frac{0.21643}{0.39667} = 0.5456。$$

对 $\alpha = 0.05$, 查 F 分布表, 可得

$$F_{1-\alpha/2}(m-1, n-1) = F_{0.975}(7, 6) = 5.70,$$

$$F_{\alpha/2}(m-1, n-1) = \frac{1}{F_{1-\alpha/2}(n-1, m-1)} = \frac{1}{F_{0.975}(6, 7)} = \frac{1}{5.12} = 0.195 ,$$

因为 $0.195 < F = 0.5456 < 5.70$, 所以接受 $H_0: \sigma_1^2 = \sigma_2^2$,

可以认为两台机床加工零件的外径的方差相等。 -----4'

(2) 问题相当于要检验 $H_0: \mu_1 \geq \mu_2$, $H_1: \mu_1 < \mu_2$

$$S_w = \sqrt{\frac{(m-1)S_x^{*2} + (n-1)S_y^{*2}}{m+n-2}} = \sqrt{\frac{(8-1) \times 0.21643 + (7-1) \times 0.39667}{8+7-2}} = 0.54737 ,$$

$$T = \frac{\bar{X} - \bar{Y}}{S_w \sqrt{\frac{1}{m} + \frac{1}{n}}} = \frac{19.925 - 20.000}{0.54737 \times \sqrt{\frac{1}{8} + \frac{1}{7}}} = -0.2647$$

对 $\alpha = 0.05$, 查 t 分布表可得 $t_{1-\alpha}(m+n-2) = t_{0.95}(13) = 1.7709$

因为 $T = -0.2647 > -t_{0.95}(13)$, 所以接受 $H_0: \mu_1 \geq \mu_2$,

故不能认为甲机床加工零件的外径显著地小于乙机床加工零件的外径. -----5'

$$(3) [\bar{X} - \bar{Y} - t_{1-\alpha/2}(m+n-2)S_w \sqrt{\frac{1}{m} + \frac{1}{n}}, \quad \bar{X} - \bar{Y} + t_{1-\alpha/2}(m+n-2)S_w \sqrt{\frac{1}{m} + \frac{1}{n}}]$$

$$= [19.925 - 20.000 - t_{0.975}(13) \times 0.2833, \quad 19.925 - 20.000 + t_{0.975}(13) \times 0.2833]$$

$$= [19.925 - 20.000 - 2.1604 \times 0.2833, \quad 19.925 - 20.000 + 2.1604 \times 0.2833]$$

$$= [-0.687, \quad 0.537] \quad \text{-----4'}$$