

统计与数据科学学院本科生2020 — 2021学年第一学期《回归分析》期末考试试卷(A卷)

草稿区

任课教师: 专业: 年级: 学号: 姓名: 成绩:

得分

一、叙述题(本题共15分, 每小题3分)

- (1) 设 X 服从 $N_n(\mu, \Sigma)$, A 是 $n \times n$ 的实对称矩阵, 且 $\text{rank}(A) = f$. 已知 $A\Sigma A = A$, 则 $(X - \mu)'A(X - \mu)$ 服从什么分布, 自由度是多少?
- (2) 设 θ 为 $p \times 1$ 的未知参数向量($p > 1$), $\hat{\theta}$ 为 θ 的一个估计, 写出该估计的均方误差的定义;
- (3) 设 $y_i \sim N(i\theta, i^2\sigma^2)$, $i = 1, \dots, n$, 且相互独立, 请给出 θ 的最小方差无偏估计 $\hat{\theta}$ 及 $\text{Var}(\hat{\theta})$;
- (4) 试写出Box-Cox变换的表达式;
- (5) 当存在中等程度复共线关系时, 回归参数的最小二乘估计有很大的均方误差, 此时Gauss-Markov定理是否还成立? 简单说明原因.

草稿区

得分

二、计算题(本题共12分)
已知 $(X_1, X_2)'$ 的密度函数为

$$f(x_1, x_2) = k^{-1} \exp \left\{ -\frac{1}{2} (x_1^2 + 2x_2^2 - x_1x_2 - 3x_1 - 2x_2 + 4) \right\}.$$

- (1) 求出 $(X_1, X_2)'$ 的数学期望和协方差阵;
- (2) 求出 $2X_1 + X_2$ 服从的分布.

草稿区

得分

三、计算题(本题共15分，每小题3分)
香烟的一氧化碳释放量(CO)对焦油含量(Tar)、尼古丁含量(Nicotine)和香烟重量(Weight)的回归结果如下.

	<i>Estimate</i>	<i>Std. Error</i>	<i>t-value</i>	<i>P-value</i>
<i>(Intercept)</i>	3.202	3.462	0.93	0.365
<i>Tar</i>	0.9626	0.2422	3.97	0.001
<i>Nicotine</i>	-2.632	3.901	-0.67	0.507
<i>Weight</i>	-0.130	3.885	-0.03	0.974
<i>Multiple R-squared: ???, Adjusted R-squared: 0.907</i>				

<i>Analysis of Variance</i>					
<i>Source</i>	<i>Degree of Freedom</i>	<i>Sum of Squares</i>	<i>Mean Squares</i>	<i>F-value</i>	<i>P-value</i>
<i>Regression</i>	3	???	165.09	???	0.000
<i>Residual Error</i>	21	???	2.09		
<i>Total</i>	24	???			

- 基于以上这些回归结果, 请回答如下问题(1)-(5).
- (1) 请指出显著的解释变量;

(2) 给出 β_{Tar} 的99%置信区间;

(3) 给出模型中误差项方差 σ^2 的估计值;

(4) 该回归的 R^2 是多少?

(5) 已知在Tar=14.1, Nicotine=0.860和Weight=0.985时的预测值为14.383, 95%预测区间为(13.142, 15.624), 试给出该参数下一氧化碳释放量的90%预测区间.
- (可能用到的临界值: $t_{21}(0.05) = 1.72$, $t_{21}(0.025) = 2.08$, $t_{21}(0.005) = 2.83$)

草稿区

得分

四、计算题(本题共12分)

- (1) 对给定的 $k > 0$, 求使得 $\|Y - X\beta\|^2 + k\|\beta\|^2$ 达到最小的 β , 其中 Y 为给定的 $n \times 1$ 向量, X 为给定的 $n \times p$ 数据阵;
- (2) 说明(1)中求解的 β 是我们学过的哪种估计形式? 该估计量是否无偏?

草稿区

得分

五、叙述题(本题共12分)

- (1) 针对普通线性回归模型, 请给出 RMS_q , C_p , AIC 和 BIC 这四种准则中任意两个的表达式;
- (2) 请叙述逐步回归的步骤.

草稿区

得分

六、计算题(本题共10分)

在天文学中, 对天空中三个星位点构成的三角形的三个内角 $\theta_1, \theta_2, \theta_3$ 各进行一次观测, 其观测值分别为 y_1, y_2, y_3 (假设观测误差 $e_i \overset{iid}{\sim} N(0, \sigma^2), i = 1, 2, 3$).

- (1) 将此问题表示成线性模型的形式;
- (2) 对三个内角分别给出合适的估计量, 并说明该估计所具有的性质.

草稿区

得分

七、证明题(本题共12分)

考虑线性模型 $\begin{cases} y = X\beta + e, \\ e \sim (0, \sigma^2 I_n), \end{cases}$ 其中 X 为 $n \times p$ 的矩阵.

- (1) 写出参数 $c'\beta$ 的最小二乘估计, 并求其期望与方差, 其中 c 为 $p \times 1$ 的常数向量;
- (2) 若 $E(a'y) = c'\beta$, 试说明 $\text{Var}(a'y)$ 会不会比 $c'\beta$ 的最小二乘估计的方差小, 为什么(写出证明过程)? 两者什么时候相等? (其中 a 为 $n \times 1$ 的常数向量)

草稿区

得分

八、计算题(本题共12分)

某试验欲研究酒精对人体反应时间的影响. 假设试验中有4位同学没有喝酒, 6位同学每人喝了两瓶, 8位同学每人喝了四瓶. 现测量这些同学对某视觉刺激的反应时间, 得到如下数据:

不喝	4	5	3	7				
两瓶	6	4	7	5	8	9		
四瓶	8	12	6	11	7	13	14	9

- (1) 将数据表示成方差分析模型的形式;
- (2) 列出方差分析表;
(可能要参考的数据 $\sum\sum y_{ij} = 138, \sum\sum y_{ij}^2 = 1230, \bar{y}_{1.} = 4.75, \bar{y}_{2.} = 6.5, \bar{y}_{3.} = 10$)
- (3) 在显著性水平 $\alpha = 0.05$ 下检验酒精含量对反应时间的影响有无显著差异.
(可能用到的常数: $F_{2,15}(0.05) = 3.68, F_{3,15}(0.05) = 3.29, F_{2,16}(0.05) = 3.63, F_{3,16}(0.05) = 3.24$)