

2018 年全国大学生数学建模竞赛暨美赛培训

回归分析

厦门大学2016 级各学院

数学建模团队：谭忠教授；助教：陈小伟，姜小蒙，姚瑶，余娇妍

要求：(1) 必须用TEX输入编辑后将TEXPDF以及图表一并发邮件提交给ztan85@163.com及sxjm004@163.com，压缩包及邮件主题为“编号+姓名+专业+第*次作业”；

(2) 必须抄题，以免判错。

一、回归分析练习

1. 某单位对50名女生测定血清蛋白含量(g/L)，数据如下：

74.3 78.8 68.8 78.0 70.4 80.5 80.5 69.7 71.2 73.5

79.5 75.6 75.0 78.8 72.0 72.0 72.0 74.3 71.2 72.0

75.0 73.5 78.8 74.3 75.8 65.0 74.3 71.2 69.7 68.0

73.5 75.0 72.0 64.3 75.8 80.3 69.7 74.3 73.5 73.5

75.8 75.8 68.8 76.5 70.4 71.2 81.2 75.0 70.4 68.0

计算均值、方差、标准差、极差、标准误、变异系数、偏度、峰度。

2. 设元件无故障工作时间 X 具有指数分布，取1000个元件工作时间的记录数据，经分组后得到它的频数分布为表2。

表2 元件工作时间频数

组中值	5	15	25	35	45	55	65
频数	365	245	150	100	70	45	25

如果各组中数据都取为组中值，试用极大似然估计求 λ 的点估计。

3. 甲乙两种水稻分别种在10块试验田中，每块试验田甲乙水稻各种一半，假设两种水稻产量 X, Y 均服从正态分布，且方差相等。收获后10块试验田的产量如下所示(单位： kg)。

表3 试验田产量

甲种	140	137	136	140	145	148	140	135	144	141
乙种	135	118	115	140	128	131	130	115	131	125

求出两种水稻产量的期望差 $\mu_1 - \mu_2$ 的置信区间($\alpha = 0.05$)。

4. 观察每分钟进入商店的人数 X ，任取200min，所得数据如下：

表4 每分钟进入商店人数

顾客人数	0	1	2	3	4	5
频数	92	68	28	11	1	0

试分析能否认为每分钟顾客人数 X 服从Poisson分布($\alpha = 0.1$)。

5. 调查某大学学生每周学习与得分的平均等级之间的关系，现抽查10个学生的资料如表5所示：

表5 某大学学生学习时间与平均等级

学习时间	24	17	20	41	52	23	46	18	15	29
学习等级	8	1	4	7	9	5	10	3	2	6

其中等级10表示最好，1表示最差. 试用秩相关检验分析学习与学习等级有无关系？

6. 为估计山上积雪融化后对下游灌溉的影响，在山上建立一个观测站，测量最大积雪深度 X 与当年的灌溉面积 Y ，现测得连续10年的数据如下表所示.

序号	X/m	Y/hm^2	序号	X/m	Y/hm^2
1	5.1	1907	6	7.8	3000
2	3.5	1287	7	4.5	1947
3	7.1	2700	8	5.6	2273
4	6.2	2373	9	8.0	3113
5	8.8	3260	10	6.4	2493

- (1) 试画出相应的散点图，判断 Y 与 X 是否有线性关系；
- (2) 求出 Y 关于 X 的一元线性回归方程；
- (3) 对方程作显著性检验；
- (4) 现测得今年的数据是 $X = 7m$ ，给出今年灌溉面积的预测值和相应的区间估计($\alpha = 0.05$)

7. 研究同一地区土壤所含可给态磷的情况 (Y)，得到18组数据如下表所示，表中 X_1 为土壤内所含无机磷浓度， X_2 为土壤内溶

于 K_2CO_3 溶液并受溴化物分解的有机磷， X_3 为土壤内溶于 K_2CO_3 但不溶于溴化物水解的有机磷.

序号	X_1	X_2	X_3	Y	序号	X_1	X_2	X_3	Y
1	0.4	52	158	64	10	12.6	58	112	51
2	0.4	23	163	60	11	10.9	37	111	76
3	3.1	19	37	71	12	23.1	46	114	96
4	0.6	34	157	61	13	23.1	50	134	77
5	4.7	24	59	54	14	21.6	44	73	93
6	1.7	65	123	77	15	23.1	56	168	95
7	9.4	44	46	81	16	1.9	36	143	54
8	10.1	31	117	93	17	26.8	58	202	168
9	11.6	29	173	93	18	29.9	51	124	99

- (1)、求 Y 关于 X 的多元线性回归方程；
 - (2)、对方程作显著性检验；
 - (3)、对变量作逐步回归分析；
8. 一位饮食公司的分析人员想调查自助餐馆中的自动咖啡售货机数量与咖啡销售量之间的关系，她选择了14家餐馆来进行实验. 这14家餐馆在营业额，顾客类型和地理位置方面都是相近的. 放在实验餐馆的自动售货机数量从0（这里咖啡由服务员端来）到6不等，并且随机分配到每个餐馆的，下表是关于实验结果的数据，其中咖啡销

售量为某一天的平均值，单位为杯.

估计($\alpha = 0.05$);

序号	售货机数量	咖啡销售量	序号	售货机数量	咖啡销售量
1	0	508.1	8	3	697.5
2	0	498.4	9	4	755.3
3	1	568.2	10	4	758.9
4	1	577.3	11	5	787.6
5	2	651.7	12	5	792.1
6	2	657.0	13	6	841.4
7	3	713.4	14	6	831.8

(1) 作线性回归模型

(2) 作多项式回归模型

(3) 画出数据的散点图和拟合曲线

9. 3个工厂生产同一种零件，现从各厂产品中分别抽取4件产品

做检测，其检测强度如下表所示：

工厂	零件强度			
甲	115	116	98	83
乙	103	107	118	116
丙	73	89	85	97

(1) 对数据作方差分析，判断3个厂生产的产品的零件强度是否

有显著差异；

(2) 求每个工厂生产的产品零件强度的均值，作出相应的区间

(3) 对数据作多重检验

9. 已知某种水泥在凝固时放出的热量Y与水泥中四种化学成分 X_1, X_2, X_3, X_4 有关，现测得13组数据，如下图所示. 希望从中选出主要的变量，建立Y 关于它们的线性回归方程. 并判断水泥数据是否存在多重共线性.

序号	X_1	X_2	X_3	X_4	Y	序号	X_1	X_2	X_3	X_4	Y
1	7	26	6	60	78.5	8	1	31	22	44	72.5
2	1	29	15	52	74.3	9	2	54	18	22	93.1
3	11	56	8	20	104.3	10	21	47	4	26	115.9
4	11	31	8	47	87.6	11	1	40	23	34	83.8
5	7	52	6	33	95.9	12	11	66	9	12	113.3
6	11	55	9	22	109.2	13	10	68	8	12	109.4
7	3	71	17	6	102.7						

二、方差分析练习

1. 正常男子血小板计数均值为 $225 \times 10^9/L$ ，今测得20名男性油漆作业工人的血小板计数值(单位： $10^9/L$)如下：

220 188 162 230 145 160 238 188 247 113

126 245 164 231 256 183 190 158 224 175

问油漆工人的血小板计数与成年男子有无显著差异？

2. 为研究某种新药对抗凝血酶活力的影响，随机安排新药组病人12例，对照组病人10例，分别测定其抗凝血酶活力(单位： mm^3)，其结果如下

新药组：126 125 136 128 123 138 142 116 110 108 115 140

对照组：162 172 177 170 175 152 157 159 160 162

试分析新药组和对照组病人的抗凝血酶活力有无差别($\alpha = 0.05$)

(1)检验两组数据是否服从正态分布；

(2)检验两组样本方差是否相同；

(3)选择最适合的检验方法检验新药组和对照组病人的抗凝血酶活力有无差别.

3. 3个工厂生产同一种零件，现从各厂产品中分别抽取4件产品做检测，其检测强度如下表所示

工厂	零件强度			
甲	115	116	98	83
乙	103	107	118	116
丙	73	89	85	97

(1) 对数据作方差分析，判断3个厂生产的产品的零件强度是否有显著差异；

(2) 求每个工厂生产的产品零件强度的均值，作出相应的区间估计($\alpha = 0.05$)；

(3) 对数据作多重检验

4. 有4种产品, $A_i, (i = 1, 2, 3)$ 分别为国内甲、乙、丙三个工厂生产的产品， A_4 为国外同类产品.现从各厂分别取10,6,6和2个产品做300h连续磨损老化实验，得变化率如下所示，假定各厂产品实验变化率服从等方差的正态分布

产品	变化率									
A_1	20	18	19	17	15	16	13	18	22	17
A_2	26	19	26	28	23	25				
A_3	24	25	18	22	27	24				
A_4	12	14								

(1) 试问4个厂生产的产品的变化率是否有显著差异

(2) 若有差异，请做进一步的检验，1国内产品与国外产品有无显著差异？2国内各厂家的产品有无显著差异？

5. 以小白鼠为对象研究正常肝核糖核酸(RNA)对癌细胞的生物作用，试验分别为对照组(生理盐水)，水层RNA组和酚层RNA组，分别用此3种不同处理方法诱导肝癌细胞的果糖二磷酸酯酶(FDP酶)活力，数据如下表所示，问3种不同处理的诱导作用是否相同？

处理方法	诱导结果							
对照组	2.79	2.69	3.11	3.47	1.77	2.44	2.83	2.52
水层RNA组	3.83	3.15	4.70	3.97	2.03	2.87	3.65	5.09
酚层RNA组	5.41	3.47	4.92	4.07	2.18	3.13	3.77	4.26

6. 将抗生素注入人体会产生抗生素与血浆蛋白质结合的现象,以致减少了药效. 下表列出5 种常用的抗生素注入到牛的体内时, 抗生素与血浆蛋白质结合的百分比. 试在水平 $\alpha = 0.05$ 下检验这些百分比的均值有无显著的差异. 设各总体服从正态分布, 且方差相同.

青霉素	四环素	链霉素	红霉素	氯霉素
29.6	27.3	5.8	21.6	29.2
24.3	32.6	6.2	17.4	32.8
28.5	30.8	11.0	18.3	25.0
32.0	34.8	8.3	19.0	24.2

7. 为分析4 种化肥(A_1, A_2, A_3, A_4)和3 个小麦品种(B_1, B_2, B_3)对小麦产量的影响, 把一块试验田等分成36 小块, 对种子和化肥的每一种组合种植3 小块田, 产量如下表所示 (单位公斤), 问品种、化肥

及二者的交互作用对小麦产量有无显著影响.

	A_1	A_2	A_3	A_4
B_1	173,172,173	174,176,178	177,179,176	172,173,174
B_2	175,173,176	178,177,179	174,175,173	172,173,174
B_3	177,175,176	174,174,175	174,173,174	169,169,170

8. 为研究人们在催眠状态下对各种情绪的反应是否有差异, 选取了8个受试者, 在催眠状态下, 要求每人按任意次序做出恐惧、愉快、忧虑、平静4种反应, 下表给出了个受试者在处于这4种情绪状态下的皮肤电位变化值. 试在 $\alpha = 0.05$ 下, 检验受试者在催眠状态下对这4种情绪的反应力是否有显著差异.

情绪状态	受试者							
	1	2	3	4	5	6	7	8
恐惧	23.1	57.6	10.5	23.6	11.9	54.6	21.0	20.3
愉快	22.7	53.2	9.7	19.6	13.8	47.1	13.6	23.6
忧虑	22.5	53.7	10.8	21.1	13.7	39.2	13.7	16.3
平静	22.6	53.1	8.3	21.6	13.3	37.0	14.8	14.8