

习题复习

课堂测试时间

- 1、使用数据集mpg,完成绘制hwy与cty关系图。要求：
 - ✱1) 绘制散点图颜色根据class变量显示, 点的大小根据displ变量。透明度为0.5, position抖动
 - ✱2) 画出hwy与cty关系的平滑曲线
 - ✱3) 控制点大小范围为4~10
 - ✱4) 根据变量year分面绘图, 要求一列
 - ✱5) 标题为'汽车油耗与型号', x轴"每加仑高速公路行驶距离", y轴"每加仑城市公路行驶距离"
 - ✱6) 添加图例, 要求颜色图例的标题为"车型", 大小图例的标题为"排量"
- 2、使用vcd包内的数据集Arthritis画条形图。要求：
 - ✱1) 条形图表示Improved变量下Sex变量的统计值, 条形图宽度为0.5, position为"fill"
 - ✱2) 添加数据标签。要求统计属性为"stat", 颜色为白色, 大小为3.5, 位置位于中间, 标签内容为统计属性值
 - ✱3) 标度填充色人为指定, 颜色为'#999999','#E69F00'
 - ✱4) 简约主题

- 3、(学生身体4项指标的主成分分析)
随机抽取30名某年级中学生,测量其身高(X_1),体重(X_2),胸围(X_3),坐高(X_4),数据如下表所示,试对这30名学生身体四项指标作主成分分析

序号	X_1	X_2	X_3	X_4	序号	X_1	X_2	X_3	X_4
1	148	41	72	78	2	139	34	71	76
3	160	49	77	86	4	149	36	67	79
5	159	45	80	86	6	142	31	66	76
7	153	43	76	83	8	150	43	77	79
9	151	42	77	80	10	139	31	68	74
11	140	29	64	74	12	161	47	78	84
13	158	49	78	83	14	140	33	67	77
15	137	31	66	73	16	152	35	73	79
17	149	47	82	79	18	145	35	70	77
19	160	47	74	87	20	156	44	78	85
21	151	42	73	82	22	147	38	73	78
23	157	39	68	80	24	147	30	65	75
25	157	48	80	88	26	151	36	74	80
27	144	36	68	76	28	141	30	67	76
29	139	32	68	73	30	148	38	70	78

- 4、27名糖尿病人的血清总胆固醇(X_1)、甘油三酯(X_2)、空腹胰岛素(X_3)、糖化血红蛋白(X_4)、空腹血糖(Y)的测量值列于下表中, 试建立血糖与其它指标的多元线性回归方程, 并作进一步分析.

i	X_1	X_2	X_3	X_4	Y
1	5.68	1.90	4.53	8.2	11.2
2	3.79	1.64	7.32	6.9	8.8
3	6.02	3.56	6.95	10.8	12.3
4	4.85	1.07	5.88	8.3	11.6
5	4.60	2.32	4.05	7.5	13.4
6	6.05	0.64	1.42	13.6	18.3
7	4.90	8.50	12.60	8.5	11.1
8	7.08	3.00	6.75	11.5	12.1
9	3.85	2.11	16.28	7.9	9.6
10	4.65	0.63	6.59	7.1	8.4
11	4.59	1.97	3.61	8.7	9.3
12	4.29	1.97	6.61	7.8	10.6
13	7.97	1.93	7.57	9.9	8.4
14	6.19	1.18	1.42	6.9	9.6
15	6.13	2.06	10.35	10.5	10.9
16	5.71	1.78	8.53	8.0	10.1
17	6.40	2.40	4.53	10.3	14.8
18	6.06	3.67	12.79	7.1	9.1
19	5.09	1.03	2.53	8.9	10.8
20	6.13	1.71	5.28	9.9	10.2
21	5.78	3.36	2.96	8.0	13.6
22	5.43	1.13	4.31	11.3	14.9
23	6.50	6.21	3.47	12.3	16.0
24	7.98	7.92	3.37	9.8	13.2
25	11.54	10.89	1.20	10.5	20.0
26	5.84	0.92	8.61	6.4	13.3
27	3.84	1.20	6.45	9.6	10.4

习题和大作业讲解

图形类

统计类

- 1、某公司想要了解消费者购买牙膏时更追求什么样的目标,于是通过商场拦访对30个人进行访谈,用7级里克特量表询问他们对以下陈述的认同程度(即1表示非常不同意,7表示非常同意,V1:购买预防蛀牙的牙膏是重要的;V2:我喜欢使牙齿亮泽的牙膏; v3:牙膏应当保护牙龈; V4:我喜欢使口气清新的牙膏; V5:预防坏牙不是牙膏提供的一项重要功效; V6:购买牙膏时最重要的考虑是富有魅力的牙齿:
 - ✱ 将调查样本存储于文本文件 yagao.txt。请使用R函数factanal对数据进行因子分析,根据载荷系数矩阵,写出因子和原变量之间的线性关系式。
- 2、某地区农业生态经济系统的各区域单元相关指标数据在文本文件agriculture.txt中,使用R中的主成分分析的函数princomp选取更少的指标来描述该地区的农业生态经济系统。写出主成分和原变量之间的线性关系式。

- 3、下表是一个一个村庄儿童年龄和平均身高的统计数据
 - ✱ (1) 画出平均身高height和年龄age关系的散点图
 - ✱ (2) 建立回归模型并提取结果输出，在(1)中的图中表示生成的模型

平均身高与年龄			
年龄 (月)	平均身高 (厘米)	年龄 (月)	平均身高 (厘米)
18	76.1	24	79.9
19	77	25	81.1
20	78.1	26	81.2
21	78.2	27	81.8
22	78.8	28	82.8
23	79.7	29	83.5

- 4、 revenue.txt中记录了财政收入(y)和第一产业GDP X_1 、第二产业GDP X_2 、第三产业GDP x_3 、人口数 X_4 、社会消费品零售总额 X_5 、受灾面积 X_6 、等情况的统计数据。要求:写出多元线性回归模型。

- 利用美国60个商学院的数据（33_bschool.txt），包括的变量由GMAT分数、学费、进入MBA前后的工资等，其中有四个定量变量，试图对这4个变量用主成分分析进行降维，得到结果后，再对该数据做因子分析，比较这两个结果，得出你的结论
-
- 有48位应聘者应聘公司某职位，公司为这些应聘者的15个指标打分，分数从0到10，0最低，10最高，具体分数见33_applicant.txt，公司要录用其中优秀的8名，写一个程序来选择：
 - 对这15个变量进行分组，并按照分组后的指标来计算总分
 - 进行主成分分析和因子分析的计算

- 100个学生的数学、物理、化学、语文、历史、英语的成绩如下表，见34_student.txt
- 进行主成分分析和因子分析的计算，解释说明他们的区别

学生代码	数学	物理	化学	语文	历史	英语
1	65	61	72	84	81	79
2	77	77	76	64	70	55
3	67	63	49	65	67	57
4	80	69	75	74	74	63
5	74	70	80	84	81	74
6	78	84	75	62	71	64
7	66	71	67	52	65	57
8	77	71	57	72	86	71
9	83	100	79	41	67	50
...

练习 - 0035

- 一个城市工业部门的十三个行业，分别是冶金、电力、煤炭、化学、机械、建材、森工、食品、纺织、缝纫、皮革、造纸、文教艺术用品，8个指标分别是年末固定资产净值 X_1 （万元），职工工人数 X_2 （人人）、工业总产值 X_3 （万元）、全员劳动生产率 X_4 （元 / 人人年），百元固定原值实现产值 X_5 （万元）、全员劳动生产率 X_6 （%），标准燃料消费量 X_7 （吨）和能源利用效果 X_8 （万元 / 吨），用主成分分析确定8个指标的主成分，并对主成分进行行解释(35_industry.txt)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	90342	52455	101091	19272	82.0	16.1	197435	0.172
2	4903	1973	2035	10313	34.2	7.1	592077	0.003
3	6735	21139	3767	1780	36.1	8.2	726396	0.003
4	49454	36241	81557	22504	98.1	25.9	348226	0.985
5	139190	203505	215898	10609	93.2	12.6	139572	0.628
6	12215	16219	10351	6382	62.5	8.7	145818	0.066
7	2372	6572	8103	12329	184.4	22.2	20921	0.152
8	11062	23078	54935	23804	370.4	41.0	65486	0.263
9	17111	23907	52108	21796	221.5	21.5	63806	0.276
10	1206	3930	6126	15586	330.4	29.5	1840	0.437
11	2150	5704	6200	10870	184.2	12.0	8913	0.274
12	5251	6155	10383	16875	146.4	27.5	78796	0.151
13	14341	13203	19396	14691	94.6	17.8	6354	1.574

- 对305个女中学生测量八个体形指标，相应的相关矩阵如表3（数据见36_student.txt）所示，用因子分析方法对这八个体形指标进行分析，找出公共因子，并给出合理解释

	身高 x_1	手臂长 x_2	上肢长 x_3	下肢长 x_4	体重 x_5	颈围 x_6	胸围 x_7	胸宽 x_8
身高	1.000							
手臂长	0.846	1.000						
上肢长	0.805	0.881	1.000					
下肢长	0.859	0.826	0.801	1.000				
体重	0.473	0.376	0.380	0.436	1.000			
颈围	0.398	0.326	0.319	0.329	0.762	1.000		
胸围	0.301	0.277	0.237	0.327	0.730	0.583	1.000	
胸宽	0.382	0.277	0.345	0.365	0.629	0.577	0.539	1.000

- 33_bschoool.txt是美国60个商学院的数据集，包括读MBA之前的工资（SalaPreMBA）、读MBA之后的工资（SalaPostMBA），学费（Tuition），这三个变量的单位均是千美元，GMAT（进商学院之前的考试成绩）等变量，
 - 建立一个模型，反映读MBA后工资和其余几个变量之间的关系，
 - 找到SalaPreMBA和SalaPostMBA之间的回归直线
 - 把剩下的三个变量都作为自变量进行回归
- pairs(), cor(), plot(), abline()

- 下表（见38_coffee.txt）是14家餐厅中自动咖啡售货机和咖啡销售量之间的关系，顾客类型和地址位置是相近的，放在餐厅的自动咖啡售货机的数量随机从0到6不等，要求：
 - 做线性回归模型
 - 做多项式回归模型
 - 画出数据的散点图、两种回归的拟合曲线

餐馆	售货机数量	咖啡销售量	餐馆	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

- 自变量有定性变量的线性回归（自学）
- 39_areif2.txt数据有三个变量：x, y, u, 其中u为定性变量（有A和B两个水平）
 - 做出三张散点图，一个是全部数据，一个是 $u=A$ 的，一个是 $u=B$ 的，并放在一行中
 - 做线性回归，画回归直线
- `par()`, `plot()`, `lm()`

练习 – 0041

例 7.1 利用四种不同配方的材料 A_1 、 A_2 、 A_3 、 A_4 生产出来的元件，测得其使用寿命如表 7.1 所示. 问：四种不同配方下元件的使用寿命有无显著的差异？

表 7.1: 元件寿命数据

材料	使用寿命							
A_1	1600	1610	1650	1680	1700	1700	1780	
A_2	1500	1640	1400	1700	1750			
A_3	1640	1550	1600	1620	1640	1600	1740	1800
A_4	1510	1520	1530	1570	1640	1600		

异？

例 7.9 在一个农业试验中，考虑四种不同的种子品种 A_1, A_2, A_3, A_4 和三种不同的施肥方法 B_1, B_2, B_3 得到产量数据如表 7.10 所示 (单位: kg). 试分析种子与施肥对产量有无显著影响?

表 7.10: 农业试验数据

	B_1	B_2	B_3
A_1	325	292	316
A_2	317	310	318
A_3	310	320	318
A_4	330	370	365

例 7.11 研究树种与地理位置对松树生长的影响, 对四个地区的三种同龄松树的直径进行测量得到数据如下表 7.15 所示 (单位: cm). A_1, A_2, A_3 表示三个不

表 7.15: 三种同龄松树的直径测量数据

	B_1	B_2	B_3	B_4
A_1	23 25 21	20 17 11	16 19 13	20 21 18
	14 15	26 21	16 24	27 24
A_2	28 30 19	26 24 21	19 18 19	26 26 28
	17 22	25 26	20 25	29 23
A_3	18 15 23	21 25 12	19 23 22	22 13 12
	18 10	12 22	14 13	22 19

同树种, B_1, B_2, B_3, B_4 表示四个不同地区. 对每一种水平组合, 进行了 5 次测量, 对此试验结果进行方差分析.

7.1 三个工厂生产同一种零件. 现从各厂产品中分别抽取 4 件产品作检测, 其检测强度如表 7.25 所示.

表 7.25: 产品检测数据

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, 判断三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。

7.5 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、忧虑和平静 4 种反应。表 7.29 给出了各受试者在处于这 4 种情绪状态下皮肤的电位变化值。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。

表 7.29: 4 种情绪状态下皮肤的电位变化值 (单位: mV)

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

大作业0004

图形复习

6月14日开始包汇报

提前一周把PPT发给我： 6月7日

谢谢！

孙惠平

sunhp@ss.pku.edu.cn