统计与数据科学学院

《数据采集方法》期末测试

姓　　名：蒋贵豪

年　　级：2021 级

专　　业：应用统计学

学　　号：B+X9bo

完成日期：2021 年 12 月 29 日

题目 1. (5 分) 在简单随机抽样中，以下命题是否正确。请将正确命题的序号写在答题纸上。

(a) 给定显著水平 α 和绝对误差，总体方差越大，所需要的样本量越大。

(b) 给定相对误差、总体方差和显著水平，总体均值越大，所需要的样本量越大。

(c) 给定总体方差和显著水平，绝对误差越小，需要的样本量就越大。

(d) 总体中所有单元的入样概率 π_i 必须一样。

(e) 样本量为 n, 则总体中任意 n 个单元的组合成为样本的概率必须一样。

解答. (a)、(c)、(d)、(e) 正确

(a) **正确**，在只有绝对误差的时候，确定样本量的公式如式1所示：

$$n = \frac{z_{\frac{\alpha}{2}}^2 S^2}{d^2 + \frac{z_{\frac{\alpha}{2}}^2 S^2}{N}} = \frac{z_{\frac{\alpha}{2}}^2}{\frac{d^2}{S^2} + \frac{z_{\frac{\alpha}{2}}^2}{N}} \tag{1}$$

那么，给定 α 和 d 后，S 越大，式1中右边的分母越小，n 越大。

(b) **错误**，在只有相对误差的时候，确定样本量的公式如式2所示：

$$n = \frac{z_{\frac{\alpha}{2}}^2 S^2}{(r\bar{y})^2 + \frac{z_{\frac{\alpha}{2}}^2 S^2}{N}} \tag{2}$$

那么，我们给定 r, S, α 后，$\bar{y}$ 越大，式2中分母越大，n 越小。

(c) **正确**，由式1，给定 S, α 后，一般来说总体 N 是固定的。d 越小，式1分母越小，n 越大。

(d) **正确**，$\pi_i = n/N$。

(e) **正确**，由 SRS 定义的第三条可得。

题目 2. (5 分) 关于不同的抽样方法，以下命题是否正确。请将正确命题的序号写在答题纸上。

(a) 整群抽样设计总是比简单随机抽样效率低。

(b) 分层抽样中，层间差异越大，估计量的精度越高。

(c) 分层抽样中，层内单元差异越大，估计量的精度越高。

(d) 按比例分配的分层抽样中，总体中每个单元的入样概率是相等的。

(e) 在分层抽样中，同一层的单元的入样概率一样。

解答. (b)、(d)、(e) 正确

(a) **错误**，整群抽样与 SRS 效率之比为:

$$deff = \frac{\mathrm{V}(\bar{y})}{\mathrm{V}_{SRS}(\bar{y})} \approx 1 + (M-1)\rho \tag{3}$$

只要在式3中取 $\rho < 0$，那么整群抽样估计量的方差更小，其效率更高。

(b) **正确**，分层抽样要求层间差异越大越好。

(c) **错误**，分层抽样要求层内差异越小越好。

(d) **正确**，分层抽样，按比例分配时，每一一个单元被抽到的概率为：

$$p = W_h \times \frac{n_h}{N_h} = \frac{N_h}{N} \times \frac{n_h}{N_h} = f \tag{4}$$

于是，每个单元的入样概率是相同的。

(e) **正确**，分层时，先对分好的层进行抽取，抽取后，再对每一层进行 SRS。因为是 SRS，所以同一层的单元入样概率一样。

题目 3. (4 分) 工程师计划用 8 次试验来研究三个因子 A, B 和 C 的效应，进行了 4 次试验之后，工程师出去休息了一段时间，回来后发现丢失了原始的试验计划。前四次的试验设计方案如下:

run	A	B	C	Response
1	$+$	$+$	$+$	80
2	$+$	$-$	$-$	45
3	$-$	$+$	$-$	75
4	$-$	$-$	$+$	60

工程师想使用一个完全因析设计, 他们能吗? 若能, 请填上后四次试验安排中的因子水平。

解答. **能**，有 3 个因子，每个因子有 2 个水平。因为 $2^3 = 8$，恰好为 8 次。后四次的设计如表1所示：

表 1: 后四次因子水平设计

run	A	B	C	Response
5	−	−	−	
6	−	+	+	
7	+	−	+	
8	+	+	−	

题目 4. (4 分) 工程师计划用不大于 10 次的试验来筛选 6 个因子 A, B, C, D, E 和 F 的主效应，请给出合适的设计表。

解答. 因为我们计划不大于 10 次试验，因子个数为 6，每个因子有高低 2 个水平。因此我们采用正交表 $L_8(2^7)$ 的前六列，给出的设计表如表2所示：

表 2: 正交设计表

因子	A	B	C	D	E	F
1	−	−	−	−	−	−
2	−	−	−	+	+	+
3	−	+	+	−	−	+
4	−	+	+	+	+	−
5	+	−	+	−	+	−
6	+	−	+	+	−	+
7	+	+	−	−	+	+
8	+	+	−	+	−	−

题目 5. (4 分) 假定如下两总体方差相同，下面哪种简单随机抽样设计可以给出总体均值更精确的估计？为什么？

(1) 总体规模为 2000，从中抽取容量为 200 的简单随机样本;

(2) 总体规模为 300000000，从中抽取容量为 3000 的简单随机样本。

解答. 我们知道，均值估计量的方差为：

$$\mathrm{V}(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{S^2}{n} \tag{5}$$

将式5代入第一种情况：

$$\mathrm{V}(\bar{y}) = \left(1 - \frac{200}{2000}\right)\frac{S^2}{200} = \frac{9}{2000}S^2 \tag{6}$$

将式5代入第二种情况：

$$\mathrm{V}(\bar{y}) = \left(1 - \frac{3000}{300000000}\right)\frac{S^2}{3000} \approx \frac{1}{3000}S^2 \tag{7}$$

第二种情况的方差更小，因此，**第二种可以得到总体均值更准确的估计**。

题目 6. (25 分) 校图书馆在年底调查南开大学学生 (假设总数为 10000) 在 2021 年从图书馆借书的情况。请回答以下问题:

1. (5 分) 考虑简单随机抽样，给定标准方差 $\sigma = 1$ (本)，显著水平 $\alpha = 0.05$，绝对误差 $d = 1$ (本)，请计算估计人均借书量所需的样本量。

2. (5 分) 我们采用 SRS 调查了 100 位学生，其中有 10 位学生在 2021 年从末借过书。那么，有多少学生从图书馆借过书？请给出一个估计。

3. (5 分) 给出上题中估计量的 95% 置信区间。

4. (5 分) 现在按学科分层，采用分层抽样。请根据下一页表中的数据估计出全校学生人均借书量及 95% 置信区间。

	理工	商科	医学	文科	艺术
N_h	10000	2000	2000	5000	1000
n_h	100	100	100	100	100
层内方差 σ_h^2	4	4	4	4	4
人均借书量 $\bar{y}_h$	2	5	4	3	5

5. (5 分) 在分层抽样中考虑按比例分配，请重新计算全校人均借书量估计值的方差。

解答. 1. 我们将已有的数据代入式1，得到：

$$n = \frac{z_{\frac{\alpha}{2}}^2 \sigma^2}{d^2 + \frac{z_{\frac{\alpha}{2}}^2 \sigma^2}{N}} = \frac{1.96^2 \times 1^2}{1^2 + \frac{1.96^2 \times 1^2}{10000}} = 3.8401 \tag{8}$$

于是，**估计人均借书量所需的样本量为 4**。

2. 我们可以得到，从图书馆借过书的人的比例估计为：

$$\hat{p} = \frac{100-10}{100} = 0.9 \tag{9}$$

因此，从图书馆借过书的人数的估计为：

$$\hat{Y} = N \times \hat{p} = 10000 \times 0.9 = 9000 \tag{10}$$

即**从图书馆借过书的人数的估计为 9000 人**。

3. 我们先给出 $\hat{p}$ 的置信区间，首先，计算 $\hat{p}$ 的方差：

$$\mathrm{V}(\hat{p}) = \frac{1}{n-1}\left(1-\frac{n}{N}\right)\hat{p}(1-\hat{p}) = \frac{1}{99} \times \left(1-\frac{100}{10000}\right) \times 0.9 \times 0.1 = 9 \times 10^{-4} \tag{11}$$

那么 $\hat{p}$ 的置信区间为：

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\mathrm{V}(\hat{p})} = 0.9 \pm 1.96 \times 0.03 = [0.8412, 0.9588] \tag{12}$$

那么，**去图书馆借过书人数估计的 95% 置信区间为**：

$$N \times [0.8412, 0.9588] = [8412, 9588] \tag{13}$$

4. 全校学生人均借书量的估计为：

$$\bar{y} = \frac{1}{N}\sum_{h=1}^{L} N_h \bar{y}_h = 2.9 \tag{14}$$

该估计的方差为：

$$\mathrm{V}(\bar{y}) = \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h \sigma_h^2}{N} = 0.0132 \tag{15}$$

从而全校人均借书量的 95% 置信区间为：

$$\bar{y} \pm z_{\frac{\alpha}{2}}\sqrt{\mathrm{V}(\bar{y})} = [2.6748, 3.1252] \tag{16}$$

综上，**全校学生人均借书量为 2.9 本，约为 3 本；其 95% 置信区间为 $[2.6748, 3.1252]$**。

5. 如果是分层抽样中的按比例分配，我们可以得到 n_h 的分配为：$n_1 = 250, n_2 = n_3 = 50, n_4 = 125, n_5 = 25$。**全校人均借书量估计值的方差为**：

$$\mathrm{V}(\bar{y}) = \sum_{h=1}^{L} \frac{W_h^2 \sigma_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h \sigma_h^2}{N} = 0.0078 \tag{17}$$

题目 7. 三. (15 分) 下表给出了一个整群的总体。

无放回地选择两个 PSU 单元，入样概率正比于 M_i。使用布鲁尔方法构建所有可能样本的 π_{ij} 的表格。计算霍维茨-汤姆森估计量的方差。

PSU	M_i	y_{ij}	t_i
1	6	$3,5,6,4,2,3$	23
2	5	$7,5,6,8,5$	31
3	8	$7,2,8,5,4,6,3,6$	41
4	6	$4,6,7,6,4,5$	32
5	4	$9,8,6,4$	27

解答. 由题意知：$M_0=\sum_{i=1}^{5} M_i=29,\ Z_i=\dfrac{M_i}{M_0}$。令 $\pi_i=2Z_i$, 布鲁尔方法的包含概率为：

$$\pi_{ij}=\frac{4Z_iZ_j\left(1-Z_i-Z_j\right)}{\left(1-2Z_i\right)\left(1-2Z_j\right)\left(1+\sum_{i=1}^{N}\frac{Z_i}{1-2Z_i}\right)} \tag{18}$$

霍维茨-汤普森估计量的方差为：

$$v\left(\hat{Y}_{\pi ij}\right)=\sum_{i=1}^{N}\sum_{j>i}^{N}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2=293.9742 \tag{19}$$

使用布鲁尔方法构建所有可能样本的 π_{ij} 如表3所示:

表 3: 所有可能样本的 π_{ij}

样本	π_{ij}
1,2	0.083101
1,3	0.161940
1,4	0.105261
1,5	0.063491
2,3	0.128795
2,4	0.083101
2,5	0.049831
3,4	0.161940
3,5	0.099049
4,5	0.063491
求和	1

题目 8. 四.（13 分）以下是一个因子的完全随机设计的方差分析表，

Source	DF	SS	MS	F
Factor	–	–	14.18	–
Error	–	37.75	–	
Total	23	108.63		

请根据方差分析表回答一下问题：

1. 请补充完整方差分析表中的剩下的 5 项；(5 分)

2. 如果显著性水平为 0.05，你的结论是接收还是拒绝原假设；(2 分)

3. 这个试验中因子有几个水平？试验重复了几次？(2 分)

4. 如果假设这个试验在实施的时候做的一个完全随机化区组设计，区组的平方和是 12。请给出这时的方差分析表。(4 分)

解答. 1. 补充后的方差分析表如表4所示：

表 4: 补充后的方差分析表

来源	自由度	平方和	均方	F
组间	5	70.88	14.18	6.7613
组内	18	37.75	2.0972	
总计	23	108.63		

2. 我们计算表4中 F 对应的 $p = 0.0010 < 0.05 = \alpha$，因此，我们**拒绝原假设**。

3. **从表4中可以看出，因子水平数为** $5 + 1 = 6$，**重复次数为** $\dfrac{23+1}{6} = 4$。

4. 此时方差分析表如表5所示：

表 5: 完全随机化区组方差分析表

来源	自由度	平方和	均方	F	p
区组	3	12	4	2.33	0.1157
处理	5	70.88	14.18	8.26	6.4038×10^{-4}
残差	15	25.75	1.7167		
总计	23	108.63			

题目 9. (10 分) 下面是一组中心复合设计的数据。

run	x_1	x_2	y
1	-1	-1	76.5
2	-1	1	77.0
3	1	-1	78.0
4	1	1	79.5
5	0	0	79.9
6	0	0	80.3
7	0	0	80.0
8	0	0	79.7
9	0	0	79.8
10	1.414	0	78.4
11	-1.414	0	75.6
12	0	1.414	78.5
13	0	-1.414	77.0

1. 利用前 9 次试验检验曲度。(5 分)
2. 用 1-13 次试验的数据构建二阶模型，找到稳定点及计算出响应的最值。(5 分)

解答. 1. 为了检验曲度，使用的模型为：

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i<j}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_1^2 \tag{20}$$

得到的结果如表6所示：

表 6: 初始一阶试验的最小二乘估计、标准误、t 统计量和 p 值

效应	估计	标准误	t	p 值
截距	79.94	0.102956	776.4459	1.65×10^{-11}
β_1	1	0.115109	8.687445	0.000966
β_2	0.5	0.115109	4.343722	0.012217
β_{12}	0.25	0.115109	2.171861	0.095611
$\beta_{11} + \beta_{22}$	-2.19	0.154434	-14.1808	0.000144

我们取 $\alpha = 0.1$，从表6中效应 β_{12} 和 $\beta_{11} + \beta_{22}$ 的 p 值可以看出，**存在明显的交互作用和曲度效应**。

2. 我们构建的二阶模型为：

$$y=\beta_0+\sum_{i=1}^{k}\beta_i x_i+\sum_{i<j}^{k}\beta_{ij}x_i x_j+\sum_{i=1}^{k}\beta_{ii}x_i^2 \tag{21}$$

得到的结果如表7所示：

表 7: 二阶试验的最小二乘估计、标准误、t 统计量和 p 值

效应	估计	标准误	t	p 值
截距	79.93995	0.119089	671.2644	4.3×10^{-18}
β_1	0.99505	0.094155	10.56822	1.48×10^{-5}
β_2	0.515203	0.094155	5.471862	0.000934
β_{12}	0.25	0.133145	1.87765	0.102519
β_{11}	-1.37645	0.100984	-13.6303	2.69×10^{-6}
β_{22}	-1.00134	0.100984	-9.91577	2.26×10^{-5}

取 $\alpha=0.1$，最终我们得到的模型为：

$$\hat{y}=79.94+0.995x_1+0.515x_2-1.376x_1^2-1.001x_2^2 \tag{22}$$

对式子22中的 x_1,x_2 分别求偏导，令偏导等于 0，联立后可以**解得稳定点为：**$x_1=995/2752=0.3616, x_2=515/2002=0.2572$，**对应的响应的最大值为：80.1861。**

题目 10. 一个半导体工厂做了 16 次试验来研究 4 个因子对产品曲度的的影响。每个试验重复四次，设计表及试验结果见下表：

1. 试验人员用的是那种设计？请写出设计表。（5 分）
2. 哪些因子对响应（曲度）的影响较大？(5 分)
3. 如果目标是尽可能的降低曲度，请给出你推荐的因子水平设置。（5 分）

Run	A	B	C	D	1	2	3	4
1	55	10	5	1580	0.0167	0.0128	0.0149	0.0185
2	75	10	5	1580	0.0062	0.0066	0.0044	0.0020
3	55	25	5	1580	0.0041	0.0043	0.0042	0.0050
4	75	25	5	1580	0.0073	0.0081	0.0039	0.0030
5	55	10	10	1580	0.0047	0.0047	0.0040	0.0089
6	75	10	10	1580	0.0219	0.0258	0.0147	0.0296
7	55	25	10	1580	0.0121	0.0090	0.0092	0.0086
8	75	25	10	1580	0.0255	0.0250	0.0226	0.0169
9	55	10	5	1620	0.0032	0.0023	0.0077	0.0069
10	75	10	5	1620	0.0078	0.0158	0.0060	0.0045
11	55	25	5	1620	0.0043	0.0027	0.0028	0.0028
12	75	25	5	1620	0.0186	0.0137	0.0158	0.0159
13	55	10	10	1620	0.0110	0.0086	0.0101	0.0158
14	75	10	10	1620	0.0065	0.0109	0.0126	0.0071
15	55	25	10	1620	0.0155	0.0158	0.0145	0.0145
16	75	25	10	1620	0.0093	0.0124	0.0110	0.0133

解答. 1，用的是 2^4 完全因析设计，设计表如所示：

表 8: 设计表

Run	A	B	C	D
1	−	−	−	−
2	+	−	−	−
3	−	+	−	−
4	+	+	−	−
5	−	−	+	−
6	+	−	+	−
7	−	+	+	−
8	+	+	+	−
9	−	−	−	+
10	+	−	−	+
11	−	+	−	+
12	+	+	−	+
13	−	−	+	+
14	+	−	+	+
15	−	+	+	+
16	+	+	+	+

2. 首先为因子的主效应，例如 A 的主效应，我们的计算方法如下：

$$ME(A) = \bar{z}(A+) - \bar{z}(A-) \tag{23}$$

其中 $A+$ 和 $A-$ 分别表示 A 的高水平和低水平，$\bar{z}(A+)$ 为在 $A+$ 上所观测到的 z_i 的平均，$\bar{z}(A-)$ 类似。然后我们来考虑交互效应，例如我们计算 AB 的交互效应，计算方式如下：

$$INT(A,B) = \bar{z}(AB=+) - \bar{z}(AB=-) \tag{24}$$

其中，AB 对应 A,B 两个因子水平高低设置之积。类似的我们还有三个水平和四个水平的交互效应：

$$INT(A,B,C) = \bar{z}(ABC=+) - \bar{z}(ABC=-) \tag{25}$$

$$INT(A,B,C,D) = \bar{z}(ABCD=+) - \bar{z}(ABCD=-) \tag{26}$$

使用原始数据，得到的因子效应如表9所示：

表 9: 因子效应 (原始响应数据)

效应	$\bar{y}$	$\ln s^2$
A	0.003890625	1.52401
B	0.000578125	-1.6613
C	0.005603125	0.548741
D	-0.001421875	-0.18544
AB	0.001915625	0.78801
AC	0.002240625	-0.10377
AD	-0.001221875	-0.278411
BC	0.001815625	0.227669
BD	0.002303125	-0.44817
CD	-0.001971875	-0.87434
ABC	-0.003446875	-0.358494
ACD	-0.007746875	-0.244959
ABD	0.000053124	-0.122657
BCD	-0.001734375	-0.17134
$ABCD$	0.000740625	0.69595

从表9可以看出，由于题目中数据的响应值过小，得到的位置效应的值也很小，在后续操作中，我们可能难以从半正态图中找出显著因子。因此我们对得到的响应数据先进行了取以 e **为底的对数，再进行因子效应的检验。**最后得到的因子效应如表10所示：

表 10: 因子效应 (原始响应数据取对数)

效应	$\bar{y}$	$\ln s^2$
A	0.35398	0.94865
B	0.052267	-1.7268
C	0.641333	-0.72031
D	-0.03913	-0.20998
AB	0.23393	0.33703
AC	0.08995	-0.185034
AD	0.0656	-0.540266
BC	0.234622	-0.15833
BD	0.179307	-0.72767
CD	-0.05994	-0.63476
ABC	-0.399655	0.35816
ACD	-0.757482	1.2547
ABD	0.02648	-0.063869
BCD	-0.19045	0.140764
$ABCD$	0.10825	0.39669

事实上，我们也可以通过回归和模型矩阵计算因子效应，将表10中的第一列作为自变量 X。A, B, C, D 分别取高低不同的水平对应的 X 记作 $X_1, ..., X_{16}$，考虑如下线性模型：

$$z_i = \beta_0 + \sum_{j=1}^{15} \beta_j x_{ij} + \varepsilon_i \tag{27}$$

用最小二乘法求解式27，有：

$$\hat{\beta}_j = \frac{1}{1-(-1)}(\bar{z}(x_{ij}=+1) - \bar{z}(x_{ij}=-1)) \tag{28}$$

通过最小二乘法求得的系数如表11所示：

表 11: 因子效应回归系数 (原始响应数据取对数)

效应	$\bar{y}$	$\ln s^2$
$\hat{\beta}_0$	-4.74395	-2.94273
$\hat{\beta}_1$	0.17699	0.47432
$\hat{\beta}_2$	0.026133	-0.8634
$\hat{\beta}_3$	0.320667	-0.36016
$\hat{\beta}_4$	-0.01956	-0.10499
$\hat{\beta}_5$	0.11697	0.16851
$\hat{\beta}_6$	0.04497	-0.092517
$\hat{\beta}_7$	0.0328	-0.270133
$\hat{\beta}_8$	0.117311	-0.07916
$\hat{\beta}_9$	0.089653	-0.36384
$\hat{\beta}_{10}$	-0.02997	-0.31738
$\hat{\beta}_{11}$	-0.199827	0.17908
$\hat{\beta}_{12}$	-0.378741	0.62735
$\hat{\beta}_{13}$	0.01324	-0.031934
$\hat{\beta}_{14}$	-0.09523	0.070382
$\hat{\beta}_{15}$	0.05412	0.19834

下面进行上述这些效应的分析：我们分别绘制 $\bar{y}$ 和 $\ln s^2$ 的半正态概率图。设 $|\hat{\theta}|_{(1)} \leq \cdots \leq |\hat{\theta}|_{(I)}$ 表示无符号效应估计 $|\hat{\theta}|_{(i)}$ 的顺序值，半正态概率表图由以下点给出：

$$\left(\Phi^{-1}(0.5+0.5[i-0.5]/I), |\hat{\theta}|_{(i)}\right), \quad i=1,\ldots,I \tag{29}$$

$\bar{y}$ 和 $\ln s^2$ 的半正态概率图如图1和图2所示。从图1和图2中可以看出，**主效应中，B 和 C 对响应影响较大，交互效应中，ACD 对响应影响较大。**

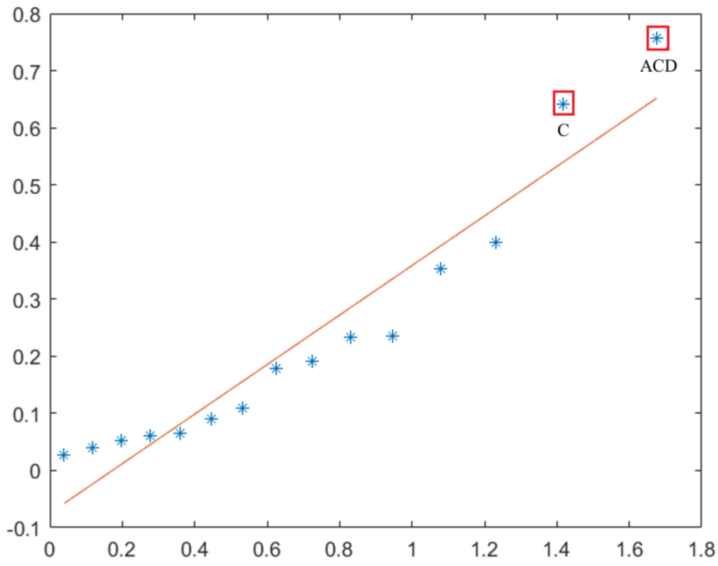

图 1: 位置半正态图

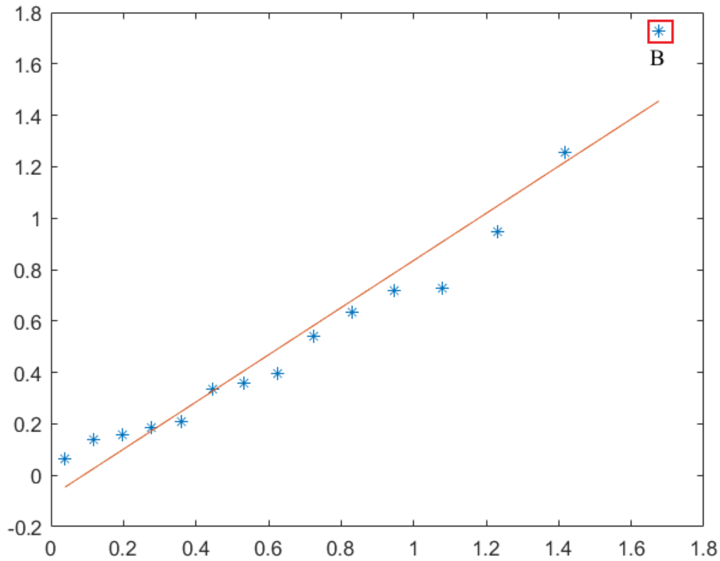

图 2: 散度半正态图

3. 我们最终得到的**响应回归方程为：**

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_3 x_C + \hat{\beta}_{12} x_{ACD} = -4.7440 + 0.3207 x_C - 0.3787 x_{ACD} \tag{30}$$

$$\hat{z} = \hat{\gamma}_0 + \hat{\gamma}_2 x_B = -2.9427 - 0.8634 x_B \tag{31}$$

为了尽可能降低曲度，也就是对应我们的 $\hat{y}$ 尽可能小，同时最好我们的方差也尽可能的小，也就是 $\hat{z}$ 尽可能的小。因此，**我们可以取** $x_A = -1, x_B = 1, x_C = -1, x_D = 1$。

附录

本文用到的所有代码均有 **Matlab** 编写

题目 7 的代码

```
%sum就是我们的所求方差
%pi_ij对应ij位置的元素值就是我们所要求的$\pi_{i j}$
M = [6 5 8 6 4];
Y = [23 31 41 32 27];
M_0 = sum(M);
Z = M/M_0;
A = sum(Z./(1-2*Z));
for i = 1:length(M)
    for j = 1:length(M)
        pi_ij(i,j)= ...
        (4*Z(i)*Z(j)*(1-Z(i)-Z(j)))/(1-2*Z(i))/(1-2*Z(j))/(1+A);
    end
end
pi_i = 2*Z;
sum = 0;
for i = 1:length(M)
    j = length(M);
    while j>i
        sum = ...
        sum+(pi_i(j)*pi_i(i)-pi_ij(i,j))*(Y(i)/pi_i(i)-Y(j)/pi_i(j))^2;
        j = j-1;
    end
end
```

题目 9 的代码

```
%用于构建X, 得到X后直接放入SPSS软件进行最小二乘回归
x=[-1   -1
-1  1
```

```
1   -1
1   1
0   0
0   0
0   0
0   0
0   0];
N = size(x);
X(:,1) = ones(N(1),1);
X(:,2:(1+N(2))) = x;
for i = 1:N(2)
    for j = 1:N(2)
        if i < j
            X = [X';(x(:,i).*x(:,j))']';
        end
    end
end
%X = [X';(x(:,1).^2)']';
X = [X';(x.^2)']';
%求稳定点
syms a b;
z=79.94+0.995*a+0.515*b-1.376*a^2-1.001*b^2;
[a,b]=solve(diff(z,a),diff(z,b),a,b)
```

题目 10 的代码

```
A_1 = [0.0167   0.0128  0.0149  0.0185
0.0062  0.0066  0.0044  0.002
0.0041  0.0043  0.0042  0.005
0.0073  0.0081  0.0039  0.003
0.0047  0.0047  0.004   0.0089
0.0219  0.0258  0.0147  0.0296
```

```
0.0121  0.009   0.0092  0.0086
0.0255  0.025   0.0226  0.0169
0.0032  0.0023  0.0077  0.0069
0.0078  0.0158  0.006   0.0045
0.0043  0.0027  0.0028  0.0028
0.0186  0.0137  0.0158  0.0159
0.011   0.0086  0.0101  0.0158
0.0065  0.0109  0.0126  0.0071
0.0155  0.0158  0.0145  0.0145
0.0093  0.0124  0.011   0.0133
]';
A = log(A_1);
Size = size(A);
y_bar = mean(A);
s_2 = var(A);
ln_s_2 = log(s_2);
D_plus = 9:16;
D_minus = 1:8;
C_plus = [5 6 7 8 13 14 15 16];
C_minus = C_plus-4;
B_plus = [3 4 7 8 11 12 15 16];
B_minus = B_plus-2;
A_minus = linspace(1,15,8);
A_plus =A_minus+1;
ME_A = mean(y_bar(A_plus))-mean(y_bar(A_minus));
ME_B = mean(y_bar(B_plus))-mean(y_bar(B_minus));
ME_C = mean(y_bar(C_plus))-mean(y_bar(C_minus));
ME_D = mean(y_bar(D_plus))-mean(y_bar(D_minus));
INT_AB = 0.5*((mean(y_bar(intersect(A_plus, B_plus)))-...
    mean(y_bar(intersect(A_plus, B_minus))))-...
```

```
    ((mean(y_bar(intersect(A_minus, B_plus)))-...
    mean(y_bar(intersect(A_minus, B_minus))))));
INT_AC = 0.5*((mean(y_bar(intersect(A_plus, C_plus)))-...
    mean(y_bar(intersect(A_plus, C_minus))))-...
    ((mean(y_bar(intersect(A_minus, C_plus)))-...
    mean(y_bar(intersect(A_minus, C_minus))))));
INT_AD = 0.5*((mean(y_bar(intersect(A_plus, D_plus)))-...
    mean(y_bar(intersect(A_plus, D_minus))))-...
    ((mean(y_bar(intersect(A_minus, D_plus)))-...
    mean(y_bar(intersect(A_minus, D_minus))))));
INT_BC = 0.5*((mean(y_bar(intersect(B_plus, C_plus)))-...
    mean(y_bar(intersect(B_plus, C_minus))))-...
    ((mean(y_bar(intersect(B_minus, C_plus)))-...
    mean(y_bar(intersect(B_minus, C_minus))))));
INT_BD = 0.5*((mean(y_bar(intersect(B_plus, D_plus)))-...
    mean(y_bar(intersect(B_plus, D_minus))))-...
    ((mean(y_bar(intersect(B_minus, D_plus)))-...
    mean(y_bar(intersect(B_minus, D_minus))))));
INT_CD = 0.5*((mean(y_bar(intersect(C_plus, D_plus)))-...
    mean(y_bar(intersect(C_plus, D_minus))))-...
    ((mean(y_bar(intersect(C_minus, D_plus)))-...
    mean(y_bar(intersect(C_minus, D_minus))))));
a = ones(1,Size(2));
b = a;
c = a;
d = a;
a(A_minus) = -1;
b(B_minus) = -1;
c(C_minus) = -1;
d(D_minus) = -1;
```

```
abc = a.*b.*c;
INT_ABC = mean(y_bar.*abc)*2
abd = a.*b.*d;
INT_ABD = mean(y_bar.*abd)*2
acd = a.*c.*d;
INT_ACD = mean(y_bar.*acd)*2
bcd = b.*c.*d;
INT_BCD = mean(y_bar.*bcd)*2
abcd = a.*b.*c.*d;
INT_ABCD = mean(y_bar.*abcd)*2
%%%分割线，以下是最小二乘回归
X(:,1) = a;
X(:,2) = b;
X(:,3) = c;
X(:,4) = d;
X(:,5) = X(:,1).*X(:,2);
X(:,6) = X(:,1).*X(:,3);
X(:,7) = X(:,1).*X(:,4);
X(:,8) = X(:,2).*X(:,3);
X(:,9) = X(:,2).*X(:,4);
X(:,10) = X(:,3).*X(:,4);
X(:,11) = X(:,1).*X(:,3).*X(:,2);
X(:,12) = X(:,1).*X(:,3).*X(:,4);
X(:,13) = X(:,1).*X(:,2).*X(:,4);
X(:,14) = X(:,4).*X(:,2).*X(:,3);
X(:,15) = X(:,1).*X(:,2).*X(:,3).*X(:,4);
X(:,16) = ones(16,1);
Y = y_bar';
beta = regress(Y,X);
INT = beta(1:length(beta)-1)*2;
```

```
I = 1:length(beta)-1;
qq_x = 0.5+0.5.*(I-0.5)/length(Y);
qq_y = abs(INT);
QQ_y = sort(qq_y);
QQ_x = norminv(qq_x,0,1);
plot(QQ_x,QQ_y,'*')
hold on
RE_X(:,1) = QQ_x';
RE_X(:,2) = ones(length(QQ_x'),1);
RE= regress(QQ_y,RE_X);
plot_re_x = linspace(min(RE_X(:,1)),max(RE_X(:,1)),1000);
plot_re_y = polyval(RE,plot_re_x);
plot(plot_re_x ,plot_re_y);

Y_s = ln_s_2';
beta_s = regress(Y_s,X);
INT_s = beta_s(1:length(beta_s)-1)*2;
I_s = 1:length(beta_s)-1;
qq_x_s = 0.5+0.5.*(I_s-0.5)/length(Y_s);
qq_y_s = abs(INT_s);
QQ_y_s = sort(qq_y_s);
QQ_x_s = norminv(qq_x_s,0,1);
plot(QQ_x_s,QQ_y_s,'*')
hold on
RE_X(:,1) = QQ_x_s';
RE_X(:,2) = ones(length(QQ_x_s'),1);
RE= regress(QQ_y_s,RE_X);
plot_re_x = linspace(min(RE_X(:,1)),max(RE_X(:,1)),1000);
plot_re_y = polyval(RE,plot_re_x);
plot(plot_re_x ,plot_re_y);
```