

数学模型与数学软件

第 12 次作业

1907402030

熊 雄^{*}



2022 年 6 月 6 日

^{*}mrxiong@foxmail.com 苏州大学数学科学学院本科生

Problem 1

(P297 Ex.2.)

据说某地汽油的价格是 115 美分/gal (加仑, 1gal=3.785412 立方分米), 为了验证这种说法, 一位司机开车随机选择了一些加油站, 得到某年 1 月和 2 月的数据如下:

1 月 119 117 115 116 112 121 115 122 116 118 109 112 119 112 117 113 114 109 109 118

2 月 118 119 115 122 118 121 120 122 128 116 120 123 121 119 117 119 128 126 118 125

- 分别用两个月的数据验证这种说法的可靠性;
- 分别给出 1 月和 2 月汽油价格的置信区间 ($\alpha = 0.05$);
- 如何给出 1 月和 2 月汽油价格差的置信区间 ($\alpha = 0.05$).

Solution.

设该地汽油的 1 月价格为随机变量 X , 2 月价格为随机变量 Y , 设总体 X 的均值和方差为 μ_1 和 σ_1^2 , 设总体 Y 的均值和方差为 μ_2 和 σ_2^2 .

- 先进行正态性检验, 然后在方差 σ^2 未知的情况下进行假设检验

$$H_0: \mu = \mu_0 = 115 \leftrightarrow H_1: \mu \neq \mu_0 = 115.$$

输入以下 *Matlab* 代码:

```

1 mu = 115;
2 First = [119 117 115 116 112 121 115 122 116 118 109 112 119 112 117 113 114
           109 109 118];
3 Second = [118 119 115 122 118 121 120 122 128 116 120 123 121 119 117 119
            128 126 118 125];
4 %先对样本做正态性检验
5 h(1) = jbtest(First);
6 h(2) = lillietest (First);
7 h(3) = jbtest(Second);
8 h(4) = lillietest (Second);
9 h,
10 h1 = ttest(First, mu), %假设检验一月
11 h2 = ttest(Second, mu) %假设检验二月

```

运行后得到输出为:

$$h = [0, 0, 0, 0], \quad h0 = 0, \quad h1 = 1.$$

这表明在显著性水平 $\alpha = 0.05$ 的情况下样本数据通过正态性检验, 且在显著性水平 $\alpha = 0.05$ 的情况下, 第一个月的数据可靠, 第二个月的数据不可靠.

- 在方差 σ^2 未知的情况下进行总体均值 μ 的区间估计. 输入以下 *Matlab* 代码:

```

1 First = [119 117 115 116 112 121 115 122 116 118 109 112 119 112 117 113 114
           109 109 118];
2 Second = [118 119 115 122 118 121 120 122 128 116 120 123 121 119 117 119
            128 126 118 125];
3 [mu1 sigma1 muc1 sigmac1] = normfit(First);
4 [mu2 sigma2 muc2 sigmac2] = normfit(Second);
5 muc1, %1 月汽油价格的置信区间
6 muc2  %2 月汽油价格的置信区间

```

运行后得到输出为:

$\text{muc1} = [113.3388, 116.9612], \quad \text{muc2} = [119.0129, 122.4871].$

这表明在 $\alpha = 0.05$ 的情况下, 一月汽油价格的置信区间为 $[113.3388, 116.9612]$, 二月汽油价格的置信区间为 $[119.0129, 122.4871]$.

- c) 在方差 σ_1^2 和 σ_2^2 未知的情况下进行均值差 $\mu_1 - \mu_2$ 的区间估计, 首先我们先需要检验方差齐性, 即

$$H_0 : \sigma_1^2 = \sigma_2^2 \leftrightarrow H_1 : \sigma_1^2 \neq \sigma_2^2.$$

然后再求双正态总体均值差的置信区间.

我们构造的枢轴量为

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

$\mu_1 - \mu_2$ 的置信度为 $1 - \alpha$ 的置信区间为:

$$[(\bar{X} - \bar{Y}) - t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2) S_W, (\bar{X} - \bar{Y}) + t_{1-\frac{\alpha}{2}}(n_1 + n_2 - 2) S_W,]$$

其中 $S_W = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$

输入以下 *Matlab* 代码:

```

1 First = [119 117 115 116 112 121 115 122 116 118 109 112 119 112 117 113 114
           109 109 118];
2 Second = [118 119 115 122 118 121 120 122 128 116 120 123 121 119 117 119
            128 126 118 125];
3 h = vartest2(First, Second), %先检验方差齐性
4 %计算置信区间
5 alpha = 0.05;
6 t = tinv(1-alpha/2, length(First)+length(Second)-2);
7 s = sqrt(((length(First)-1)*var(First) + (length(Second)-1)*var(Second)) /
           (length(First)+length(Second)-2)) ;

```

```
8 d(1) = (mean(First) - mean(Second)) - t*s* (sqrt(1/length(First) + 1/  
length(Second))); %置信下  
限  
9 d(2) = (mean(First) - mean(Second)) + t*s* (sqrt(1/length(First) + 1/  
length(Second))); %置信上  
限  
10 d
```

运行后得到输出为:

$$h = 0, \quad d = [-8.0273, -3.1727],$$

这表明在 $\alpha = 0.05$ 的情况下, 可以认为 $\sigma_1^2 = \sigma_2^2$, 然后再求双正态总体均值差的置信区间, 可以得到 $\mu_1 - \mu_2$ 的置信区间为 $d = [-8.0273, -3.1727]$. ■

Problem 2

(P297 Ex.3.)

某校 60 名学生的一次考试成绩如下:

93 75 83 93 91 85 84 82 77 76 77 95 94 89 91 88 86 83 96 81 79 97 78 75 67
 69 68 84 83 81 75 66 85 70 94 84 83 82 80 78 74 73 76 70 86 76 90 89 71 66
 86 73 80 94 79 78 77 63 53 55

- 作直方图, 计算均值、标准差、极差、偏度、峰度;
- 检验分布的正态性.

Solution.

a) 输入以下 *Matlab* 代码:

```
1 Grade = [93 75 83 93 91 85 84 82 77 76 77 95 94 89 91 88 86 83 96 81 79 97 78
           75 67 69 68 84 83 81 75 66 85 70 94 84 83 82 80 78 74 73 76 70 86 76
           90 89 71 66 86 73 80 94 79 78 77 63 53 55];
2 hist(Grade),
3 x1 = mean(Grade),
4 x2 = std(Grade),
5 x3 = range(Grade),
6 x4 = skewness(Grade),
7 x5 = kurtosis(Grade)
```

运行后输出为:

```
1 x1 = 80.1000 x2 = 9.7106 x3 = 44 x4 = -0.4682 x5 = 3.1529
```

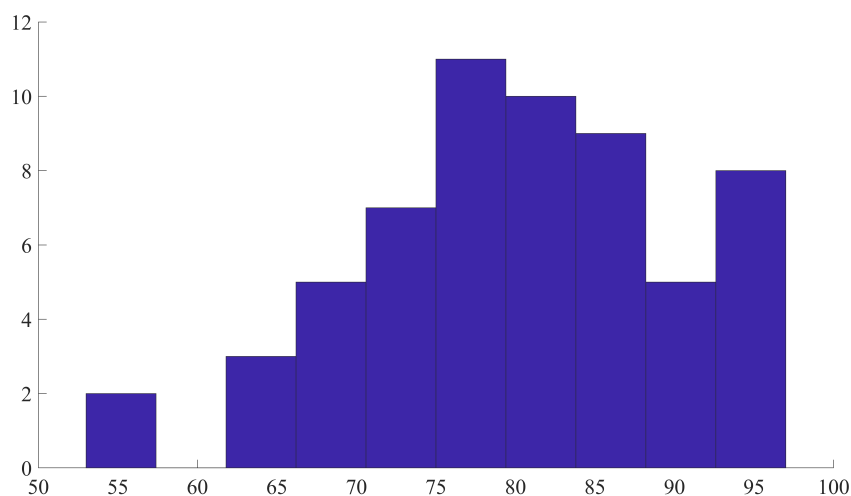


图 1: 考试成绩直方图

综上, 均值 $x_1 = 80.1$ 分, 标准差 $x_2 = 9.7106$, 极差 $x_3 = 44$ 分, 偏度 $x_4 = -0.4682$, 峰度 $x_5 = 3.1529$.

b) 新建脚本输入以下 *Matlab* 代码:

```
1 Grade = [93 75 83 93 91 85 84 82 77 76 77 95 94 89 91 88 86 83 96 81 79 97 78  
           75 67 69 68 84 83 81 75 66 85 70 94 84 83 82 80 78 74 73 76 70 86 76  
           90 89 71 66 86 73 80 94 79 78 77 63 53 55];  
2 h(1) = jbtest(Grade);  
3 h(2) = lillietest (Grade);  
4 h
```

得到输出为 $h = [0, 0]$, 这表明全班成绩通过了正态性检验, 成绩的分布在显著性水平为 $\alpha = 0.05$ 的情况下符合正态分布. ■

Problem 3

(P297 Ex.5.)

甲方向乙方成批供货, 甲方承诺合格率为 90%, 双方商定置信区间为 95%. 现从一批货中抽取 50 件, 43 件为合格品, 问乙方应否接受这批货物? 你能为乙方不接受它出谋划策吗?

Solution.

设随机变量 X 表示是否合格, 则 X 服从 0-1 分布. 设总体 X 的不合格率为 p , 期望为 $\mu = p$, 方差为 $\sigma^2 = p(1-p)$.

对总体的不合格率作如下右侧假设检验:

$$H_0 : p \leq p_0 = 0.1 \leftrightarrow H_1 : p > p_0.$$

由中心极限定理知, 无论总体是正态或非正态, 只要其均值与方差存在, 其样本均值 \bar{x} 在大样本场合就近似服从正态分布, 即

$$u = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} \sim AN(0, 1).$$

故建立检验统计量

$$U = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} \sim AN(0, 1).$$

对于给定的显著性水平 $\alpha = 1 - 0.95 = 0.05$, 其拒绝域为 $W = u > u_{1-\alpha}$, 查表得 $u_{1-0.05} = 1.65$, 则拒绝域为

$$W = [1.65, +\infty].$$

由于 $\bar{x} = \frac{7}{50}$, $\mu_0 = 0.1$, $\sigma = 0.3$, 故 U 统计量的观测值为

$$u = \frac{\bar{x} - \mu}{\sigma} \sqrt{n} = \frac{\sqrt{50}(\frac{7}{50} - 0.1)}{0.3} = 0.9428 \notin W.$$

从而应当接受原假设 H_0 , 即乙方应接受这批货物.

若要使乙方不接受这批货物, 可以升高合格率达到 95%, 或者增大显著性水平, 提高抽检次数等等.



Problem 4

(P297 Ex.6.)

学校随机抽取 100 名学生, 测量他们的身高与体重, 所得数据见下表.

身高	体重	身高	体重	身高	体重	身高	体重	身高	体重
172	75	169	55	169	64	171	65	167	47
171	62	168	67	165	52	169	62	168	65
166	62	168	65	164	59	170	58	165	64
160	55	175	67	173	74	172	64	168	57
155	57	176	64	172	69	169	58	176	57
173	58	168	50	169	52	167	72	170	57
166	55	161	49	173	57	175	76	158	51
170	63	169	63	173	61	164	59	165	62
167	53	171	61	166	70	166	63	172	53
173	60	178	64	163	57	169	54	169	66
178	60	177	66	170	56	167	54	169	58
173	73	170	58	160	65	179	62	172	50
163	47	173	67	165	58	176	63	162	52
165	66	172	59	177	66	182	69	175	75
170	60	170	62	169	63	186	77	174	66
163	50	172	59	176	60	166	76	167	63
172	57	177	58	177	67	169	72	166	50
177	63	176	68	172	56	173	59	174	64

- 对数据给出直观的图形描述, 检验分布的正态性;
- 根据这些数据对全校学生的平均身高和体重做出估计, 并给出估计的误差范围;
- 学校 10 年前作过普查, 学生的平均身高为 167.5cm, 平均体重为 60.2kg, 根据抽查的数据, 对学生的平均身高和体重有无明显变化做出结论.

Solution.

- 利用 *Matlab* 输入以下代码:

```

1 Height = [172,171,166,160,155,173,166,170,167,173,178,173,163,165,170,163,...
2           172,182,171,177,169,168,168,175,176,168,161,169,171,178,177,170,173,...
3           172,170,172,177,176,175,184,169,165,164,173,172,169,173,173,166,163,...
4           170,160,165,177,169,176,177,172,165,166,171,169,170,172,169,167,175,...
5           164,166,169,167,179,176,182,186,166,169,173,169,171,167,168,165,168,...
6           176,170,158,165,172,169,169,172,162,175,174,167,166,174,168,170];
7 Weight = [75 62 62 55 57 58 55 63 53 60 60 73 47 66 60 50 ...
8           57 63 59 64 55 67 65 67 64 50 49 63 61 64 66 58 67 ...
9           59 62 59 58 68 68 70 64 52 59 74 69 52 57 61 70 57 ...
10          56 65 58 66 63 60 67 56 56 49 65 62 58 64 58 72 76 ...

```



```

11      59 63 54 54 62 63 69 77 76 72 59 65 71 47 65 64 57 ...
12      57 57 51 62 53 66 58 50 52 75 66 63 50 64 62 59];
13 histogram(Height);
14 figure
15 histogram(Weight);
16 h(1) = jbtest(Height);
17 h(2) = lillietest (Height);
18 h(3) = jbtest(Weight);
19 h(4) = lillietest (Weight);

```

运行后输出图像如下:

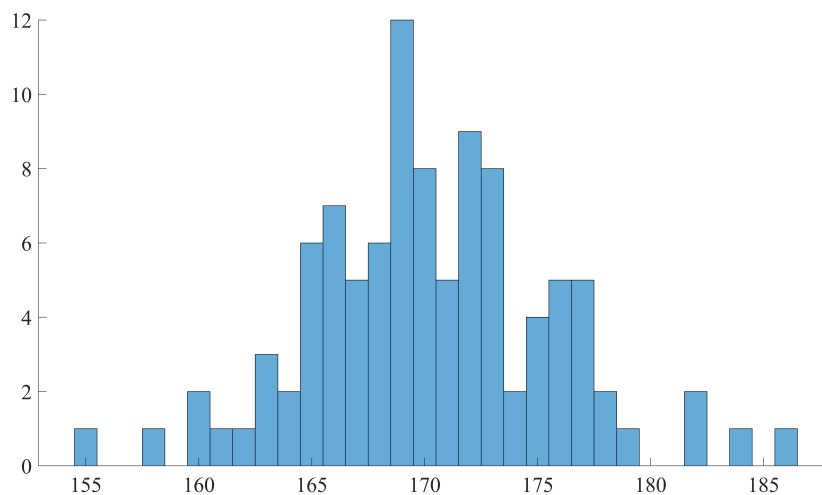


图 2: 身高直方图

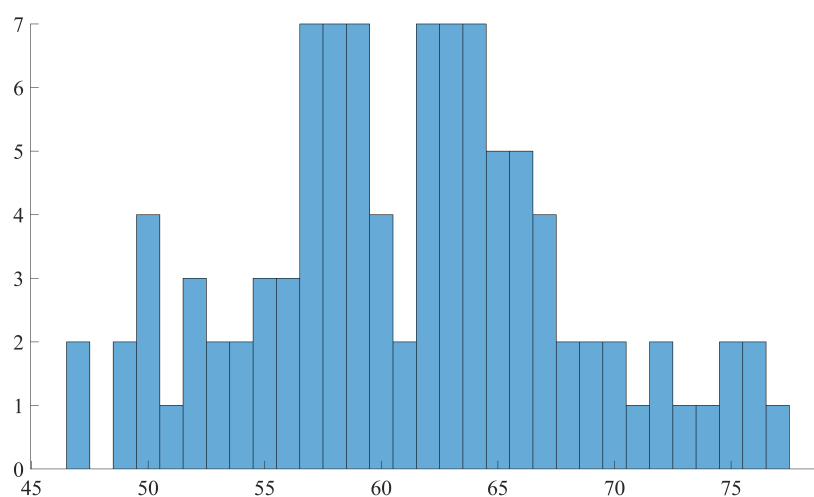


图 3: 体重直方图

$h = [0 \ 0 \ 0 \ 0]$, 表明通过了正态性检验, 身高、体重的分布在显著性水平为 $\alpha = 0.05$ 的情况下均符合正态分布.

b) 分别计算样本中身高、体重的平均值与标准差:

```
1 A = [mean(Height),std(Height),mean(Weight),std(Weight)]
```

可以得到: 由 a) 可知体重与身高均服从正态分布, 因此我们可以对身高、体重进行

	身高 (Height)	体重 (Weight)
平均值	170.25	61.27
方差	29.1793	47.5122

估计:

```
1 [mu sigma muci sigmaci] = normfit(Height);
2 [mu sigma muci sigmaci] = normfit(Weight);
```

可以得到身高与体重的参数估计如下表所示:

身高均值的点估计	身高标准差的点估计	身高均值的区间估计	身高标准差的区间估计
170.25	5.4018	[169.1782, 171.3218]	[4.7428, 6.2751]
体重均值的点估计	体重标准差的点估计	体重均值的区间估计	体重标准差的区间估计
61.27	6.8929	[59.9023, 62.6377]	[6.0520, 8.0073]

由以上的两个表格我们可以看出, 身高与体重的平均值与标准差的点估计与用样本计算出的数据相等, 且均被包含在区间估计的区间中.

c) 设随机变量 X 表示身高, 均值为 μ_1 , 方差为 σ_1^2 ; 随机变量 Y 表示体重, 均值为 μ_2 , 方差为 σ_2^2 ,

假设检验

$$H_{01} : \mu_1 = 167.5 \leftrightarrow H_{11} : \mu_1 \neq 167.5;$$

$$H_{02} : \mu_2 = 60.2 \leftrightarrow H_{12} : \mu_2 \neq 60.2.$$

利用 *Matlab* 执行以下语句:

```
1 [h1] = ttest(Height, 167.5)
2 [h2] = ttest(Weight, 60.2)
```

可以得到 $h1 = 1, h2 = 0$.

综上所述,

- 对于身高, 拒绝原假设 H_{01} , 表明学生身高在 10 年间发生了明显的变化;
- 对于体重, 接受原假设 H_{11} , 表明学生体重没有在 10 年间发生明显的变化. ■