数学实验·实验报告9

计算机系　计01　容逸朗　2020010869

12-5 供货问题

问题分析与模型假设

首先，以随机变量 X 表示产品的质量，其中 $X=0$ 代表产品合格，若 $X=1$ 则产品不合格。记 p 为合格率，那么有 $\mu=p$ ，$\sigma^2=p(1-p)$ 。

由于 X 不是正态分布，所以需要根据中心极限定理作假设：对于足够大的 n，我们可以认为 $z=\frac{\mu-p}{\sqrt{p(1-p)/n}}$ 服从标准正态分布 $N(0,1)$。

模型建立

由于甲方承诺合格率达 $p_0=90\%$ ，而且对于乙方而言，更高的合格率显然是更好的，故只需要判断产品的合格率是否符合如下假设检验即可：

$$H_0: p \geq p_0 = 0.9; \quad H_1: p < p_0 = 0.9$$

由模型假设知，我们认为 z 服从标准正态分布，故可以利用正态总体分布的 z 检验，找到 $\alpha=0.05$ 时的分位数 u_α。记样本的均值为 $\bar{x}$ ，那么有 $z=\frac{\bar{x}-p_0}{\sqrt{p_0(1-p_0)/n}}$ 满足 $P(z\geq u_\alpha)=1-\alpha,$ 此时检验规则变为：

$$z \geq u_\alpha \text{时接受} H_0; \quad \text{否则拒绝} H_0$$

算法设计

由于 `MATLAB` 没有关于 `0-1` 分布的总体均值检验函数，因此我们需要自编程序解决这一问题。

对于第一问，只需要直接代入飆数即可。

对于第二问，则可以尝试更改 z 中参数（显著性水平 α，样本数 n，合格率 p_0）并求解。

程序

1. `0-1` 分布总体均值检验

```matlab
% 0-1 分布总体均值检验
function [h, sig] = ztestbin(xbar, p0, n, alpha, tail)
z = (xbar - p0) / sqrt(p0 * (1 - p0) / n);
if tail == 0
    sig = 2 * (1 - normcdf(abs(z)));
    if abs(z) < norminv(1 - alpha / 2)
        h = 0;
    else
```

```
            h = 1;
        end
end
if tail == 1
    sig = 1 - normcdf(z);
    if z < norminv(1 - alpha)
        h = 0;
    else
        h = 1;
    end
end
if tail == -1
    sig = normcdf(z);
    if z > norminv(alpha)
        h = 0;
    else h = 1;
    end
end
```

2. 主函数

```
format long

xbar = 43 / 50;
p0 = 0.9;
n = 50;
alpha = 0.05;

% 第一问
[h, sig] = ztestbin(xbar, p0, n, alpha, -1)

% 第二问: 更改 alpha
for alpha_ = 0.05: 0.00001: 0.5
    [h, sig] = ztestbin(xbar, p0, n, alpha_, -1);
    if h == 1
        alpha_
        break
    end
end

% 第二问: 更改 p0
for p0_ = 0.9: 0.00001: 1
    [h, sig] = ztestbin(xbar, p0_, n, alpha, -1);
    if h == 1
        p0_
        break
    end
end

% 第二问: 更改 n
```

```
for n_ = 50: 1: 1000
    [h, sig] = ztestbin(xbar, p0, n_, alpha, -1);
    if h == 1
        n_
        break
    end
end
```

计算结果

对于第一问，我们有 $h=0, sig=0.172889$ ，这说明了乙方应当接受这批货物。

对于第二问，可以调整参数为：

1. 将显著性水平提升至 $\alpha=0.17289$ 以上（也就是将置信区间降至 82.71%），这时恰有 $sig<\alpha$ ；
2. 将合格率要求提升至 $p_0=92.23\%$ 以上；
3. 将样本数提升至 $n=153$ 个以上。

分析

通过上面的讨论可以看出，我们可以通过改变不同的参数（显著性水平、样本数、合格率等）来改变假设检验的结论。但是，这样调整参数会带来一定问题，例如提高显著性水平（即置信空间减少），会让货物落入接收区间的概率减少，相对地，也更容易导致第一类错误（拒绝不该拒绝的货物）出现。

结论

1. 在题目给定的条件下，乙方应当接受这批货物。
2. 乙方可以采取下述任意一种参数调整方法来拒绝货物：
    - 降低置信空间至 82.71%；
    - 提升合格率到 92.23%；
    - 增加样本个数至 153 个以上。

12-6 身高体重估计

问题分析与模型建立

题目给出了 $n=100$ 名学生的身高和体重的数据，为了检验正态性，可以运用 Jarque-Bera 或 Lilliefors 检验来判断数据的正态性，若数据符合正态分布，则可以利用参数估计的方法来得到数据的区间估计结果。

除此之外，题目还给出了十年前数据的均值，为了检测数据是否有变，可以构造如下假设检验：

$$H_0: \mu=\mu_0; \quad H_1: \mu \neq \mu_0$$

由于方差未知，由此可以利用 t 分布来检验，我们有：$t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t(n-1)$ ，这时可以找到 $t_{1-\alpha/2}$ 使得 $P(|t| \leq t_{1-\alpha/2})=1-\alpha$ ，因此更改假设检验方式为：

$$\text{当}|t| \leq t_{1-\alpha/2}\text{时接受}H_0; \quad \text{否则拒绝}H_0$$

算法设计

1. 可以利用 **MATLAB** 中的 **jbtest** 函数和 **lillietest** 来检验数据是否符合正态分布；
2. 然后利用 **normfit** 来估计正态分布的参数和区间；
3. 最后利用 **ttest** 来检验数据即可。

程序

```
format short

height = [
    172; 169; 169; 171; 167;
    171; 168; 165; 169; 168;
    166; 168; 164; 170; 165;
    160; 175; 173; 172; 168;
    155; 176; 172; 169; 176;
    173; 168; 169; 167; 170;
    166; 161; 173; 175; 158;
    170; 169; 173; 164; 165;
    167; 171; 166; 166; 172;
    173; 178; 163; 169; 169;
    178; 177; 170; 167; 169;
    173; 170; 160; 179; 172;
    163; 173; 165; 176; 162;
    165; 172; 177; 182; 175;
    170; 170; 169; 186; 174;
    163; 172; 176; 166; 167;
    172; 177; 177; 169; 166;
    182; 176; 172; 173; 174;
    171; 175; 165; 169; 168;
    177; 184; 166; 171; 170;];

weight = [
    75; 55; 64; 65; 47;
    62; 67; 52; 62; 65;
    62; 65; 59; 58; 64;
    55; 67; 74; 64; 57;
    57; 64; 69; 58; 57;
    58; 50; 52; 72; 57;
    55; 49; 57; 76; 51;
    63; 63; 61; 59; 62;
    53; 61; 70; 63; 53;
    60; 64; 57; 54; 66;
    60; 66; 56; 54; 58;
    73; 58; 65; 62; 50;
    47; 67; 58; 63; 52;
```

```
    66; 59; 66; 69; 75;
    60; 62; 63; 77; 66;
    50; 59; 60; 76; 63;
    57; 58; 67; 72; 50;
    63; 68; 56; 59; 64;
    59; 68; 56; 65; 62;
    64; 70; 49; 71; 59;];

% 第一问：正态检验
h1_h = jbtest(height)
h2_h = lillietest(height)
h1_w = jbtest(weight)
h2_w = lillietest(weight)

% 第二问：参数估计
[h, sig, muci, sigmaci] = normfit(height, 0.05)
[h, sig, muci, sigmaci] = normfit(weight, 0.05)

% 第三问：检验数据
[h, sig, ci] = ttest(height, 167.5)
[h, sig, ci] = ttest(weight, 60.2)
```

计算结果与分析

- 两种检验（Jarque-Bera, Lilliefors）对两种不同的数据（身高、体重）均给出了 $h=0$ 的结论，也就是说数据符合正态分布。
- 取显著性水平 $\alpha=0.05$ ，得到的数据估计值如下表示：

	身高/cm	体重/kg
均值点估计	170.25	61.27
均值区间估计	(169.1782, 171.3218)	(59.9023, 62.6377)
标准差点估计	5.4018	6.8929
标准差区间估计	(4.7428, 6.2751)	(6.0520, 8.0073)

- 对于身高，t 检验给出了 $h=1$ 的结果，这代表与十年前相比，学生的平均身高出现了明显变化，从上表中也可以直接看到 $167.5 \notin (169.1782, 171.3218)$。
- 对于体重，t 检验给出了 $h=0$ 的结果，这代表与十年前相比，学生的平均体重没有明显变化。

结论

1. 学生的身高和体重有正态性；
2. 显著性水平 $\alpha=0.05$ 时，平均身高和体重的估计如下表所示；

	身高/cm	体重/kg
均值点估计	170.25	61.27
均值区间估计	(169.1782, 171.3218)	(59.9023, 62.6377)
标准差点估计	5.4018	6.8929
标准差区间估计	(4.7428, 6.2751)	(6.0520, 8.0073)

3. 与十年前相比，学生的平均身高出现了明显变化，而平均体重则没有明显变化。

11-7 胃液检验

问题分析与模型假设

题设给出了病人的数据和正常人的溶菌酶含量数据各 $n=30$ 条，我们需要检查这两种数据的相关性。记病人数据的均值、标准差为 μ_1,σ_1 ，正常人的数据均值、标准差为 μ_2,σ_2。

在此需要假设两组数据均具有正态性，且两组数据的方差满足 $\sigma_1^2=\sigma_2^2$ 。

模型建立

在数据满足上述假设的情况下，我们可以作如下假设检验：

$$H_0:\mu_1=\mu_2,\quad H_1:\mu_1\neq\mu_2$$

又因为我们假设数据具有正态性，且满足 $\sigma_1^2=\sigma_2^2$ ，由此可以利用 **t** 分布来检验。记 $t=\frac{\mu_1-\mu_2}{\sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}}}\sim t(n_1+n_2-2)$，其中 $s^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ 。那么可以取到 $t_{1-\alpha/2}$ 使得 $P(|t|\leq t_{1-\alpha/2})=1-\alpha$ ，此时假设检验变为：

$$\text{当}|t|\leq t_{1-\alpha/2}\text{时接受}H_0;\quad \text{否则拒绝}H_0$$

算法设计

由于样本数较少，因此先采用 `Lilliefors` 检验题给数据的正态性（对应 **MATLAB** 的 **`lillietest`** 函数），然后利用 **`vartest2`** 检查数据的方差是否相等，最后利用 **`ttest2`** 来计检测两组数据的差异性便可。

对于第二问只需要去掉病人组的后 `5` 个数据即可。

程序

```
format short

a = [0.2; 10.4; 0.3; 0.4; 10.9; 11.3; 1.1; 2.0; 12.4; 16.2;
    2.1; 17.6; 18.9; 3.3; 3.8; 20.7; 4.5; 4.8; 24.0; 25.4;
    4.9; 40.0; 5.0; 42.2; 5.3; 50.0; 60.0; 7.5; 9.8; 45.0;];

```

```
b = [0.2; 5.4; 0.3; 5.7; 0.4; 5.8; 0.7; 7.5; 1.2; 8.7;
    1.5; 8.8; 1.5; 9.1; 1.9; 10.3; 2.0; 15.6; 2.4; 16.1;
    2.5; 16.5; 2.8; 16.7; 3.6; 20.0; 4.8; 20.7; 4.8; 33.0;];

% 检查正态性
ha1 = lillietest(a)
ha2 = lillietest(a(1:25, 1))
hb = lillietest(b)

% 检查数据标准差是否相同
hs1 = vartest2(a, b)
hs2 = vartest2(a(1:25, 1), b)

% 检查数据的差别
[h, sig, ci] = ttest2(a, b)
[h, sig, ci] = ttest2(a(1:25, 1), b)
```

计算结果与分析

数据检查结果：

- 对于三组数据（病人原数据、病人更正、正常人），检验后均得到 $h = 1$ ，即数据不具有正态性。
- 调用 `vartest2` 后，得到 $h = 1$ ，说明拒绝 $\sigma_1^2 = \sigma_2^2$ 的假设。

虽然数据完全不符合检验的要求，但为了完成推论，我们将继续使用 **t** 检验（ 取显著性水平 $\alpha = 0.05$ ）：

- 对于第一问，使用全集数据检验给出 $h = 1$ ，$sig = 0.0251, ci = (0.9886, 14.3114)$ ；

  这说明病人与正常人在溶菌酶含量上有明显区别。
- 对于第二问，去掉错误数据后检验给出 $h = 0, sig = 0.1558, ci = (-1.5035, 9.1528)$

  这说明病人与正常人在溶菌酶含量上无明显区别。

结论

在显著性水平 $\alpha = 0.05$ 的情况下：

- 使用全集数据可以给出病人与正常人在溶菌酶含量上有明显区别。
- 去掉 5 个错误的数据后给出病人与正常人在溶菌酶含量上无明显区别。