## 实验十二作业

杜国豪 2018011346 计03

1. 实验目的
2. 根据参数估计和假设经验的基本理论模型建立模型
3. 确定参数，使用MATLAB求解；
4. 实验内容

1.甲方向乙方成批供货,甲方承诺合格率为 90%,双方商定置信概率为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?

在这个问题中，我们需要考虑的是二项分布，用于估计一个二项随机变量（即合格或不合格的货物）的结果概率。我们可以使用Matlab来解决这个问题。使用binocdf函数，它可以计算在二项分布中观察到特定数量成功的累积概率。我们将使用这个函数来计算至少43件合格品的概率。

1. % 问题参数
2. n = 50; % 样本大小
3. p = 0.90; % 甲方承诺的合格率
4. x = 43; % 观察到的合格产品数量
5. % 计算在x-1或更少成功的概率，然后用1减去它来得到至少x成功的概率
6. prob\_x\_or\_more = 1 - binocdf(x-1, n, p);
7. % 检查乙方是否应该接受这批货物，如果prob\_x\_or\_more大于或等于置信概率，则接受
8. confidence\_level = 0.95;
9. if prob\_x\_or\_more >= confidence\_level
10. fprintf('乙方应接受这批货物。');
11. else
12. fprintf('乙方不应接受这批货物。');
13. end

在上述代码中，我们首先设定了问题参数。然后，我们计算了在x-1或更少成功的累积概率，并用1减去它来得到至少x成功的概率。最后，我们比较这个概率和置信概率，以决定乙方是否应接受这批货物。

如果乙方想要找到不接受货物的理由，他们可以要求更高的置信水平（例如99%），或者要求更高的产品质量（例如95%的合格率），或者要求在样本中看到更多的合格产品。因此，稳妥起见，乙方最好拒绝。

1. 学校随机抽取 100 名学生,测量他们的身高和体重,所得数据见表 12.6，

(1)对这些数据给出直观的图形描述,检验分布的正态性;

(2)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;

(3)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,根据这次抽查的数据,对学生的平均身高和体重有无明显变化做出结论.

首先导入数据，然后进行可视化和正态性检验：

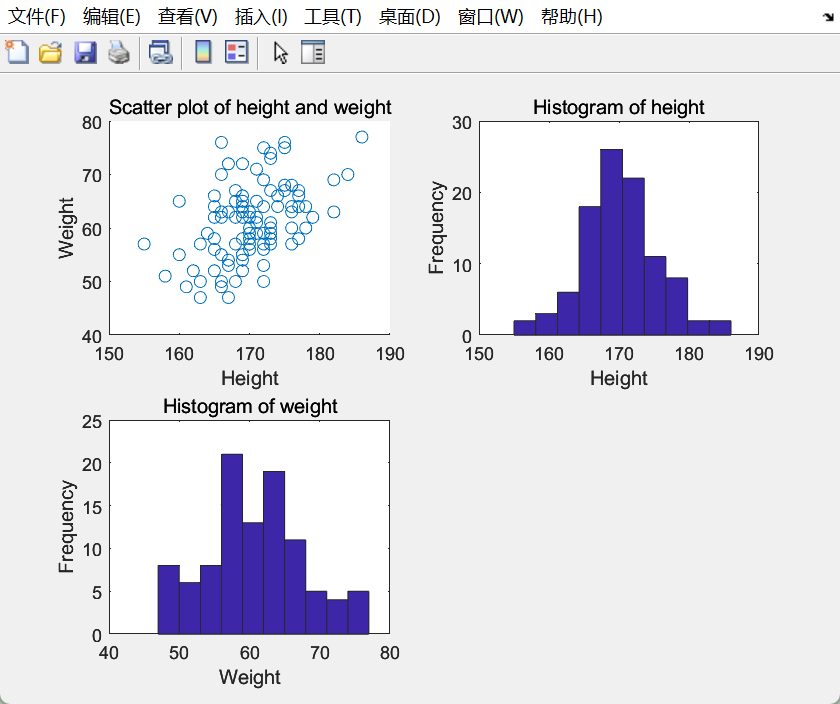
1. % 导入数据
2. height = [172 171 166 160 155 173 166 170 167 173 178 173 163 165 170 163 ...
3. 172 182 171 177 169 168 168 175 176 168 161 169 171 178 177 170 173 172 170 ...
4. 172 177 176 175 184 169 165 164 173 172 169 173 173 166 163 170 160 165 ...
5. 177 169 176 177 172 165 166 171 169 170 172 169 167 175 164 166 169 167 ...
6. 179 176 182 186 166 169 173 169 171 167 168 165 168 176 170 158 165 172 169 ...
7. 169 172 162 175 174 167 166 174 168 170];
8. weight = [75 62 62 55 57 58 55 63 53 60 60 73 47 66 60 50 57 63 59 64 55 ...
9. 67 65 67 64 50 49 63 61 64 66 58 67 59 62 59 58 68 68 70 64 52 59 ...
10. 74 69 52 57 61 70 57 56 65 58 66 63 60 67 56 56 49 65 62 58 64 58 ...
11. 72 76 59 63 54 54 62 63 69 77 76 72 59 65 71 47 65 64 57 57 57 51 ...
12. 62 53 66 58 50 52 75 66 63 50 64 62 59];
13. % 可视化数据：散点图和直方图
14. subplot(2,2,1); scatter(height, weight); xlabel('Height'); ylabel('Weight'); title('Scatter plot of height and weight');
15. subplot(2,2,2); hist(height); xlabel('Height'); ylabel('Frequency'); title('Histogram of height');
16. subplot(2,2,3); hist(weight); xlabel('Weight'); ylabel('Frequency'); title('Histogram of weight');
17. % 正态性检验
18. [hHeight, pValueHeight] = kstest((height-mean(height))/std(height));
19. [hWeight, pValueWeight] = kstest((weight-mean(weight))/std(weight));
20. if hHeight == 0
21. disp('Height data follows a normal distribution');
22. else
23. disp('Height data does not follow a normal distribution');
24. end
25. if hWeight == 0
26. disp('Weight data follows a normal distribution');
27. else
28. disp('Weight data does not follow a normal distribution');
29. end

然后，可以估计全校学生的平均身高和体重，并给出估计的误差范围。我们可以使用t分布的置信区间来做这个估计：

1. % 计算平均值和标准差
2. meanHeight = mean(height);
3. stdDevHeight = std(height);
4. meanWeight = mean(weight);
5. stdDevWeight = std(weight);
6. % 计算置信区间
7. confidence = 0.95;
8. df = length(height) - 1; % 自由度
9. tValue = tinv((1+confidence)/2, df); % t值
10. lowerHeight = meanHeight - tValue\*stdDevHeight/sqrt(length(height));
11. upperHeight = meanHeight + tValue\*stdDevHeight/sqrt(length(height));
12. lowerWeight = meanWeight - tValue\*stdDevWeight/sqrt(length(weight));
13. upperWeight = meanWeight + tValue\*stdDevWeight/sqrt(length(weight));
14. fprintf('The estimated average height is between %0.2f cm and %0.2f cm.\n', lowerHeight, upperHeight);
15. fprintf('The estimated average weight is between %0.2f kg and %0.2f kg.\n', lowerWeight, upperWeight);

最后，我们可以进行假设检验，以决定学生的平均身高和体重是否有明显变化：

1. % 旧的平均身高和体重
2. oldMeanHeight = 167.5;
3. oldMeanWeight = 60.2;
4. % 进行t检验
5. [hHeightChange, pValueHeightChange] = ttest(height, oldMeanHeight);
6. [hWeightChange, pValueWeightChange] = ttest(weight, oldMeanWeight);
7. if hHeightChange == 0
8. disp('There is no significant change in average height.');
9. else
10. disp('There is a significant change in average height.');
11. end
12. if hWeightChange == 0
13. disp('There is no significant change in average weight.');
14. else
15. disp('There is a significant change in average weight.');
16. end
17. 画图



（2）

The estimated average height is between 169.18 cm and 171.32 cm.

The estimated average weight is between 59.90 kg and 62.64 kg.

（3）

输出为：  
There is a significant change in average height.

There is no significant change in average weight.

证明平均身高发生了显著变化

7.为研究胃溃疡的病理医院作了两组人胃液成分的试验,患胃溃疡的病人组与无胃溃疡的对照组各取 30 人,胃液中溶菌酶含量见表 12.7(溶菌酶是一种能破坏某些细菌的细胞

壁的酶).

1. 根据这些数据判断患胃疡病人的溶菌酶含量与“正常人”有无显著差别;
2. 若表 12.7 患胃溃疡病人组的最后 5 个数据有误,去掉后再作判断.

对于(1)，我们将进行独立样本t检验来看胃溃疡病人组和正常人组的溶菌酶含量是否有显著差异

1. % 导入数据
2. patients = [0.2 10.4 0.3 0.4 10.9 11.3 1.1 2.0 12.4 16.2 ...
3. 2.1 17.6 18.9 3.3 3.8 20.7 4.5 4.8 24.0 25.4 ...
4. 4.9 40.0 5.0 42.2 5.3 50.0 60.0 7.5 9.8 45.0];
5. normal = [0.2 5.4 0.3 5.7 0.4 5.8 0.7 7.5 1.2 8.7 ...
6. 1.5 8.8 1.5 9.1 1.9 10.3 2.0 15.6 2.4 16.1 ...
7. 2.5 16.5 2.8 16.7 3.6 20.0 4.8 20.7 4.8 33.0];
8. % 进行独立样本t检验
9. [h,p] = ttest2(patients, normal);
10. % 显示结果
11. if h == 0
12. fprintf('No significant difference in lysozyme levels between patients with gastric ulcers and normal people (p = %0.3f).\n', p);
13. else
14. fprintf('Significant difference in lysozyme levels between patients with gastric ulcers and normal people (p = %0.3f).\n', p);
15. end

对于(2)，我们将去掉患胃溃疡病人组的最后 5 个数据，然后再进行独立样本t检验：

1. % 去掉患胃溃疡病人组的最后 5 个数据
2. patients\_corrected = patients(1:end-5);
3. % 进行独立样本t检验
4. [h,p] = ttest2(patients\_corrected, normal);
5. % 显示结果
6. if h == 0
7. fprintf('No significant difference in lysozyme levels between corrected patients with gastric ulcers and normal people (p = %0.3f).\n', p);
8. else
9. fprintf('Significant difference in lysozyme levels between corrected patients with gastric ulcers and normal people (p = %0.3f).\n', p);
10. end

我们首先导入数据，然后进行独立样本t检验，最后显示检验结果。如果 p 值小于0.05，我们认为两组之间存在显著差异。对于第二部分，我们去掉患胃溃疡病人组的最后 5 个数据，然后再进行相同的检验。

（1）

输出为：Significant difference in lysozyme levels between patients with gastric ulcers and normal people (p = 0.025).

因此有显著差别

（2）

输出为：No significant difference in lysozyme levels between corrected patients with gastric ulcers and normal people (p = 0.156).

因此没有显著差别