Bonus assignment 3

Ludwig Tranheden
ludtra@student.chalmers.se
9406238791
MVE155

2016/03/03

Contents

1 Problem 11.22

In this problem we are to compare two methods for measuring the calcium content of animal feed, oxalate precipitation and flame photometry. The main question to be investigated is if there is any systematical difference between the two methods.

1.1 Parametric test

First we assume the samples are independent and investigate normality. Plotting the probability plots of each sample we see that the samples most likely are not normal. The skewness and kurtosis for oxalate is 1.1799 and 3.7526. The skewness and kurtosis for flame is 1.1569 and 3.6780. This together with the result of the χ^2-test for goodness of fit further justifies our conclusion (neighter the $\log(\cdot)$ or $1/\cdot$ transform yields any different result).

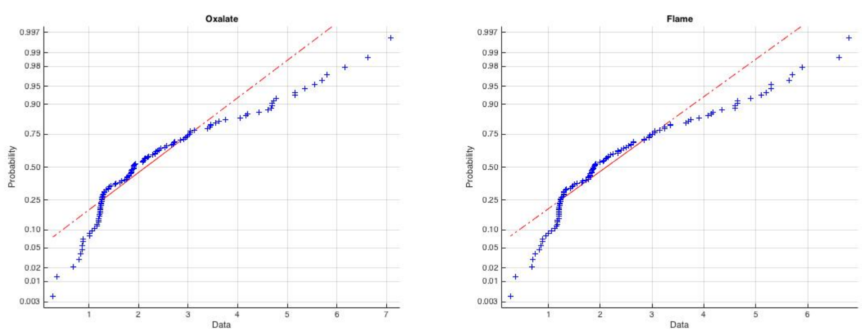

Figure 1: Normal probability plot of oxalate precipitation and flame photometry

If X is the oxalate sample and Y is the flame sample and also μ_1 and μ_2 are the true means of oxalate and flame respectively. Then the large sample test yields $(\bar{X} - \bar{Y} \sim^a N(\mu_1 - \mu_2, S^2_{\bar{X}} + S^2_{\bar{Y}}))$. We want to test the null hypothesis

$$H_0 : \mu_1 = \mu_2 \qquad \text{Both means are equal}$$

against the alternative that there are a systematic difference. The test statistic is $T_{obs} = \frac{\bar{X}-\bar{Y}}{\sqrt{S^2_{\bar{X}}+S^2_{\bar{Y}}}} = 0.1902$. This gives a P-value of 0.8491, we do not reject the null-hypothesis.

Listing 1: Code for parametric test

```
load Data

figure(1)
normplot((Oxalate))
title('Oxalate')
```

```
figure(2)
normplot((Flame))
title('Flame')
[h1,p1] = chi2gof(Oxalate);
[h2,p2] = chi2gof(Flame);

mean = mean(Oxalate) - mean(Flame);
st = sqrt((std(Oxalate)^2)/length(Oxalate) + (std(Flame)^2)/length(Flame));
T2 = mean/st;
[h,p] = ztest(T2,0,1);
```

1.2 Nonparametric test

As in the previous subsection we assume that the samples are independent. The test we will use is the "Rank sum test". First we pool the samples and replace the data by their ranks. Next we sum up the X and Y ranks yielding R_1 and R_2 respectively. Because we have two large samples (118 each) we can use the normal approximation for the null distribution of R_1 and R_2 (Matlab does this by default for large samples as ours). Now we are testing the null hypothesis:

$$H_0 : M_1 = M_2 \qquad \text{The medians are equal}$$

Doing this in matlab requires only one line of code diplayed below. Our P-value is 0.7559 and we again do not reject the null hypothesis.

Listing 2: Code for nonparametric test

```
[p,h] = ranksum(Oxalate,Flame)
```

1.3 Graphical method

To compare the two samples with eachother in a graphical way you can use a boxplot or a QQ-plot, both shown below.

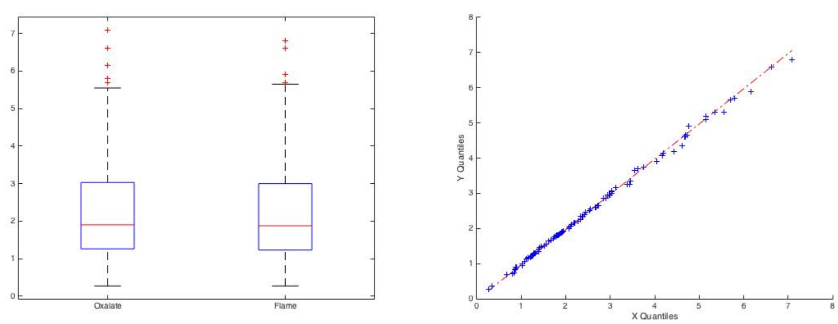

Figure 2: Boxplot and qq-plot of the two samples

As we see from the boxplot the two samples looks very similar and the medians are roughly the same as we concluded in the nonparametric case. Further the qq-plot show a very linear relationship and the samples thereby most likely come from the same distribution.

Listing 3: Code for nonparametric test

```
figure(4);
boxplot([Oxalate', Flame'],'labels',{'Oxalate','Flame'})
figure(5)
qqplot(Oxalate, Flame)
```

1.4 Conclusion

All the test supported the hypothesis that there are no systematic difference, with rather large marginals as well. So the conclusion is that there are no systematic difference between the two methods and flame photometry should be used since it's faster.