Investigating if glaucoma affects corneal thickness in R

Jack Hodgkinson
1st Year Mathematics

To determine whether glaucoma affects the corneal thickness, measurements were made in 8 people affected by glaucoma in one eye but not in the other. The corneal thickness (in microns) were as follows:

Person	Eye affected by glaucoma	Eye not affeced by glaucoma
1	488	484
2	478	478
3	480	492
4	426	444
5	440	436
6	410	398
7	458	464
8	460	476

Throughout the coursework, I have defined x = the data set for the eye affected by glaucoma and y = the data set for the eye not affected by glaucoma.

a) My R command to calculate the statistics in the question for eye affected by glaucoma is as follows:

```
>  x <- c(488,478,480,426,440,410,458,460)
> c(mean(x),median(x),sd(x),max(x),min(x))
[1] 455.00000 459.00000  27.69219 488.00000 410.00000
```

This gives Mean = 455, Median = 459, Standard Deviation = 27.69219, Maximum = 488, Minimum = 410.

b) My R command to calculate the statistics in the question for eye not affected by glaucoma is as follows:

```
> y <- c(484,478,492,444,436,398,464,478)
> c(mean(y),median(y),sd(y),max(y),min(y))
[1] 459.25000 471.00000  31.47675 492.00000 398.00000
```

This gives Mean = 459.25, Median = 471, Standard Deviation = 31.47675, Maximum = 492, Minimum = 398.

c) From the sample of 8 people given, the mean, median, standard deviation and maximum all appear to have a larger corneal thickness for the eye not affected by glaucoma however the minumum corneal thickness is larger for the eye affected by glaucoma.

d) My R command is as below:

```
> boxplot(x,y,names=c("Eye affected by glaucoma","Eye not affected
by glaucoma"),ylab="Corneal Thickness (in microns)")
```

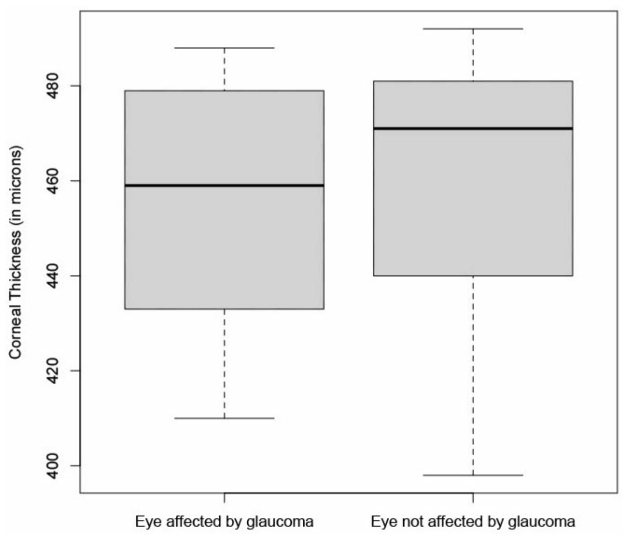

Figure 1: The boxplots of the corneal thickness of eye affected by glaucoma and eye not affected by glaucoma

Looking at Figure 1, we can say that the median, range, upper quartile and lower quartile is larger for the eye not affected by glaucoma. However, the interquartile range is larger for the eye affected by glaucoma.

e) Looking at the eye affected by glaucoma, let $\hat{\mu}$ denote our estimate for the mean, and let $\hat{\sigma}$ denote our estimate for the standard deviation.

$$\hat{\mu} = \frac{1}{8}\sum_{i=1}^{8} x_i = 455 \text{ and } \hat{\sigma} = \sqrt{\frac{1}{8}\sum_{i=1}^{8}(x_i - \hat{\mu})} = 27.69219$$

Using the Kolmogorov Smirnov test, I used the following command in R:

```
> ks.test (x,"pnorm", 455, 27.69219)
```

and the following was outputted:

```
One-sample Kolmogorov-Smirnov test

data:  x
D = 0.17189, p-value = 0.9416
alternative hypothesis: two-sided
```

As the p-value = 0.9416 > 0.05, by the Kolmogorov Smirnov test we can conclude that the normal distribution is an adequate fit to the data.

f) Looking at the eye not affected by glaucoma, let $\hat{\mu}$ denote our estimate for the mean, and let $\hat{\sigma}$ denote our estimate for the standard deviation.

$$\hat{\mu} = \frac{1}{8}\sum_{i=1}^{8} y_i = 459.25 \text{ and } \hat{\sigma} = \sqrt{\frac{1}{8}(y_i - \hat{\mu})^2} = 31.47675$$

Using the Kolmogorov Smirnov test, I used the following command in R:

```
> ks.test(y,"pnorm",459.25,31.47675)
```

and the following was outputted:

```
One-sample Kolmogorov-Smirnov test

data:  y
D = 0.2243, p-value = 0.8157
alternative hypothesis: two-sided
```

As the p-value = 0.8157 > 0.05, by the Kolmogorov Smirnov test we can conclude that the normal distribution is an adequate fit to the data.

g) In this question, we have been asked to find a 95% confidence interval for the difference between the population mean on data on eye affected by glaucoma and the population mean of the data on eye not affected by glaucoma. We are told to assume that the standard deviations calculated in a) and b) are the true population standard deviations. Using the information given, I deduced that the formula needed calculate the 95% C.I. for $\mu_x - \mu_y$ is

$$\bar{x} - \bar{y} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{n}}$$

where n denotes the sample size. From the data, we know that $n = 8$. As this is a 95% confidence interval, then we can calculate $\alpha = 0.05$ hence $1 - \frac{\alpha}{2} = 0.975$. This is the value we will use in the qnorm command in R, as this command gives us the population 0.975 quantile Q(0.975), which is also known as the z-value. I inputted the following commands into R:

```
> lower <- mean(x)-mean(y)-qnorm(0.975)*sqrt(var(x)/8+var(y)/8)
> lower
[1] -33.30149
> upper <- mean(x)-mean(y)+qnorm(0.975)*sqrt(var(x)/8+var(y)/8)
> upper
[1] 24.80149
```

which gave me the confidence interval:

$$[-33.30149, 24.80149]$$

From the interval calculated, there is a 95% probability that the difference between the population mean of the data on eye affected by glaucoma and the population mean of the data on eye not affected by glaucoma is between -33.30149 and 24.80149.

h) Let μ_x = the population mean of the data on eye affected by glaucoma and μ_y = the poplation mean of the data on eye not affected by glaucoma. We are tasked with testing the following hypothesis:

$$H_0 : \mu_x = \mu_y$$

vs

$$H_1 : \mu_x < \mu_y$$

This is at the five percent level of significance, hence we know that $\alpha = 0.05$. We take the standard deviations $\sigma_x = 27.69219$ and $\sigma_y = 31.47675$ from a) and b) to be the true population standard deviations.

We are told to reject H_0 if

$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2 + \sigma_y^2}{n}}} < -z_{1-\alpha}$$

with n being the sample size. Using our value of alpha, we know we need to find $-z_{0.95}$. I found this using the following command in R:

```
> -qnorm(0.95)
[1] -1.644854
```

I calculated the value of $\frac{\bar{x}-\bar{y}}{\sqrt{\frac{\sigma_x^2+\sigma_y^2}{n}}}$ in R too, as seen below:

```
> (mean(x)-mean(y))/sqrt((var(x)+var(y))/8)
[1] -0.7882634
```

As $-0.7882634 > -1.644854$, there is insufficient evidence from the data provided to reject H_0 at the 5% level of significance. Therefore we have insufficient evicence to reject the claim that $\mu_x = \mu_y$.