Statistical Inference: Problem 3

Sebastian Nilsson

March 2019

a)

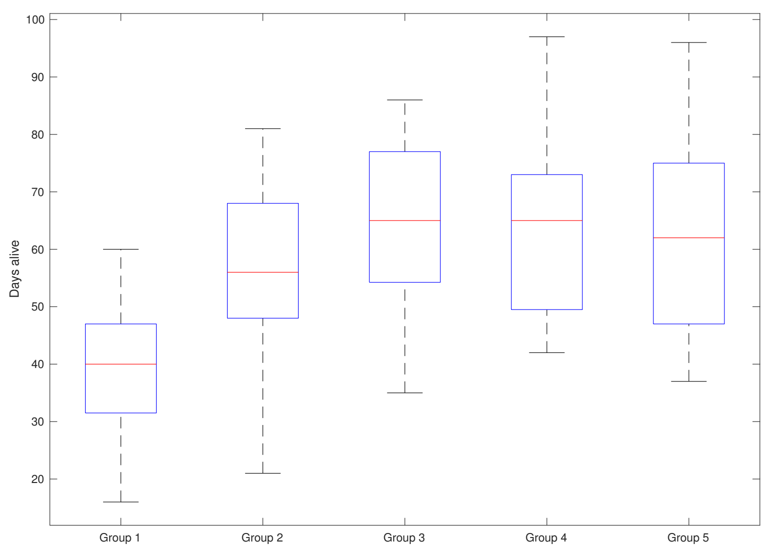

Figure 1: Box plot of the days alive between each group

Group 1 is the males living with 8 virgin females. Group 2 lives with one virgin female. Group 3 lives with 8 newly pregnant females, group 4 lives with 1 newly pregnant. Group 5 is housed alone.

Looking at the box plot one can see that the male fruit flies living with virgin females have a reduced lifespan. It seems that the number of females affect the life span as well. Living with 8 reduces the lifespan more than living with one.

b)

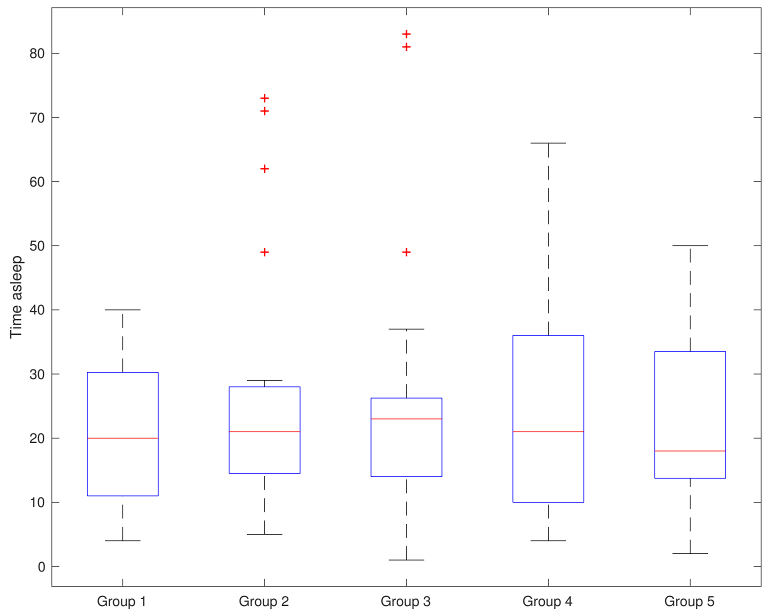

Figure 2: Box plot of the time asleep between each group

There is no obvious difference of time asleep between the groups.

c)

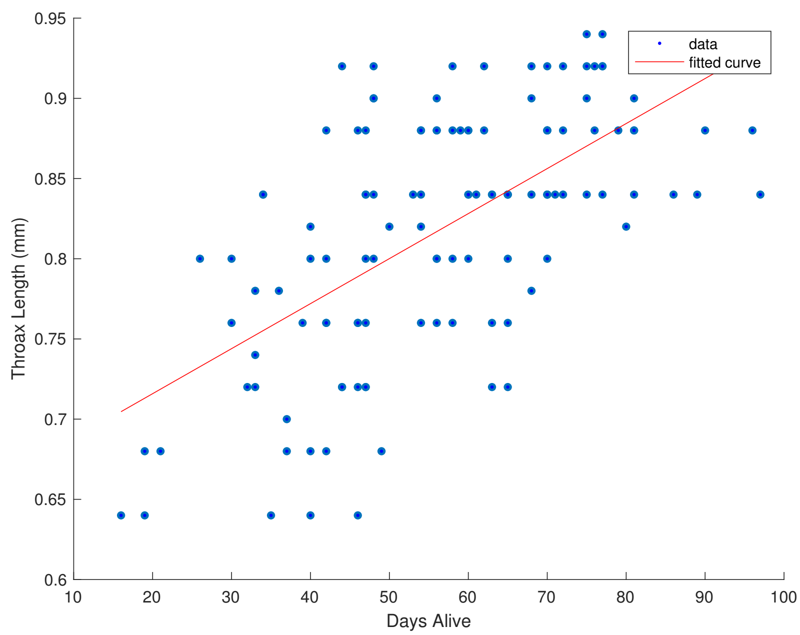

Figure 3: Scatter plot of lifespan versus thorax length

Yes, there seems to be a correlation between thorax length and lifespan. Fitting a line to the data we can see that the days alive increases linearly with thorax length.

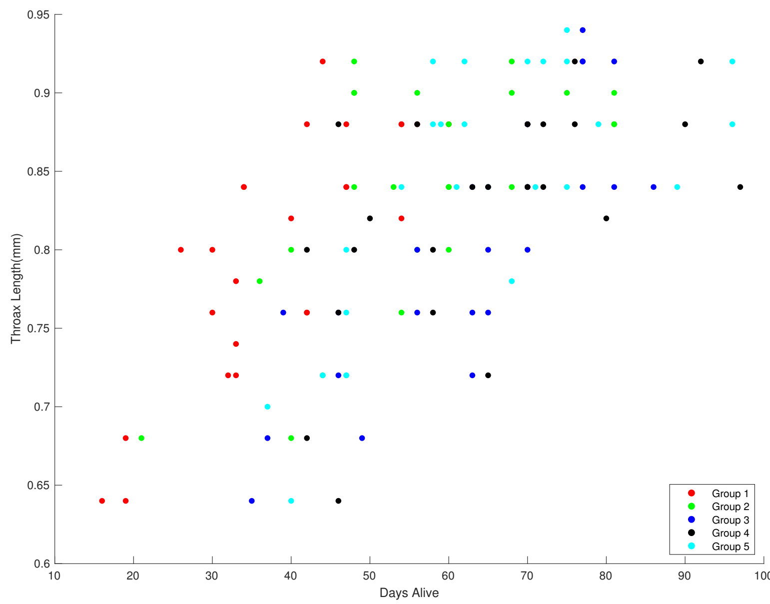

Figure 4: Scatter plot of lifespan versus thorax length, where each group has been colored

Looking at the scatter plot, where the data points has been coloured by group number, one can notice that a larger proportion of the smaller male fruit flies has been placed in group 1. Hence, the randomisation did not fully balance the thorax length between the groups.

d)

To calculate the F statistic, we make use of the ANOVA table:

Source	SS	df	MS	F
Groups	11939	4	2984.75	13.56
Error	26314	120	219.28	
Total	38253	124		

Let our null hypothesis claim that there is no difference of longevity between the groups. If our null hypothesis was true, then the F statistic should be close to 1. It's 13.56, so we see that we have differences of lifespan between our groups, thus we can reject the null hypothesis.

Tukey's Method

We set $\alpha = 0.05$ and we look up a matching percentile in the studentized range distribution. We a limit for the maximum confidence interval width:

$$q_{I,I(J-1)}(\alpha)\frac{s_p}{J} = q_{5,120}(0.05)\frac{s_p}{\sqrt{J}} = 3.92 * \frac{14.80}{5} = 11.60$$

Where we get S_p from

$$s_p = \sqrt{\frac{SS_W}{I(J-1)}} = 14.80$$

If any difference of mean between two groups is greater than 2.32 the difference is significant at a 95 % confidence level.

$$|\bar{Y_1} - \bar{Y_2}| = 18.04$$
$$|\bar{Y_1} - \bar{Y_3}| = 24.64$$
$$|\bar{Y_1} - \bar{Y_4}| = 26.08$$
$$|\bar{Y_1} - \bar{Y_5}| = 24.84$$
$$|\bar{Y_2} - \bar{Y_3}| = 6.60$$
$$|\bar{Y_2} - \bar{Y_4}| = 8.04$$
$$|\bar{Y_2} - \bar{Y_5}| = 6.80$$
$$|\bar{Y_3} - \bar{Y_4}| = 1.44$$
$$|\bar{Y_3} - \bar{Y_5}| = 0.20$$
$$|\bar{Y_4} - \bar{Y_5}| = 1.24$$

The results follow the trend seen in the first box plot in task a). Between the control groups there is no significant difference in lifespan, while group 1 and 2 show a significant difference in comparison with the other groups.

The Bonferroni Method

For the bonferroni method we set $k = \binom{5}{2} = 10$ and create the maximum confidence interval width:

$$s_p\frac{t_{I(J-1)}(1-\alpha/k)}{\sqrt{J/}} = s_p\frac{t_{120}(0.025)}{\sqrt{12.5}} = 14.80\frac{2.617}{3.54} = 10.95$$

With this method we claim that there is a significant difference between group 1 and the other groups.

e)

We get a new maximum confidence interval

$$s_{p_{kw}} \frac{t_{120}(0.025)}{\sqrt{12.5}} = 30.65 \frac{2.617}{3.54} = 22.59$$

Using the average rank as the measurement instead of the average lifespan we get the following differences between the groups:

$$|\bar{R_1} - \bar{R_2}| = 37.70$$
$$|\bar{R_1} - \bar{R_3}| = 51.98$$
$$|\bar{R_1} - \bar{R_4}| = 52.76$$
$$|\bar{R_1} - \bar{R_5}| = 49.56$$
$$|\bar{R_2} - \bar{R_3}| = 14.28$$
$$|\bar{R_2} - \bar{R_4}| = 15.06$$
$$|\bar{R_2} - \bar{R_5}| = 11.86$$
$$|\bar{R_3} - \bar{R_4}| = 0.78$$
$$|\bar{R_3} - \bar{R_5}| = 2.42$$
$$|\bar{R_4} - \bar{R_5}| = 3.20$$

Using the maximum confidence interval width for significance, 22.59, we can see that the differences between the control groups and between the control groups and group 2 are negligible. While group significant differences in comparison with the other groups.

f)

Looking at the box plot diagram there doesn't seem to be any significant difference of sleep time between the groups. Creating an ANOVA table and looking at the F statistic we can see that F is not close to 1, rather it's below 1.

Source	SS	df	MS	F
Groups	486	4	121.5	0.47
Error	30778	120	256.4	
Total	31265	124		

Let our null hypothesis be that there is no difference of sleep between each group. With a significance level $\alpha = 0.95$ we say that we will reject the null hypothesis if our $F > F_{4,120}(0.95)$. We find the tabulated F to be $F_{4,120}(0.95) = 2.45$, thus we cannot reject the null hypothesis since our F is 0.47.