Lab 9 Handout Solution

STAT 3011

In this lab session, we will revisit the fruit fly data set from Lab handout 1. The data set is from Hanley, J. A. and S. H. Shapiro (1994) "Sexual activity and the lifespan of male fruit flies: a data set that gets attention." Journal of Statistics Education 2.

Hanley and Shapiro (1994) report on a study conducted by Partridge and Farquhar (1981) about the sexual behavior of fruit flies. It was already known that increased reproduction leads to shorter life spans for female fruit flies. But the question remained *whether an increase in sexual activity would also reduce the life spans of male fruit flies.* The researchers designed an experiment to answer this question. They had a total of 125 male fruit flies to use and they randomly assigned each of the 125 to one of the following five groups: 1 pregnant female fruit fly, 1 virgin female fruit fly, 8 pregnant female fruit fly, 8 virgin female fruit fly, none.

We are interested in the following variables:

- Longevity: Longevity of male fruit flies in days.
- Partners: Number of female partners, 0, 1, 8.
- Treatment: Type of treatment (different type and number of partners) each male fruit fly received.

You may use one of the following commands to import the data set.

```
dat<-read.table("http://stat2.org/datasets/FruitFlies.csv", sep=',', header=TRUE)
## OR
dat<-read.csv("http://stat2.org/datasets/FruitFlies.csv")
attach(dat)
```

0. To explore the data set, use `summary(dat)`. What is the overall average longevity based on all 125 male fruit flies?

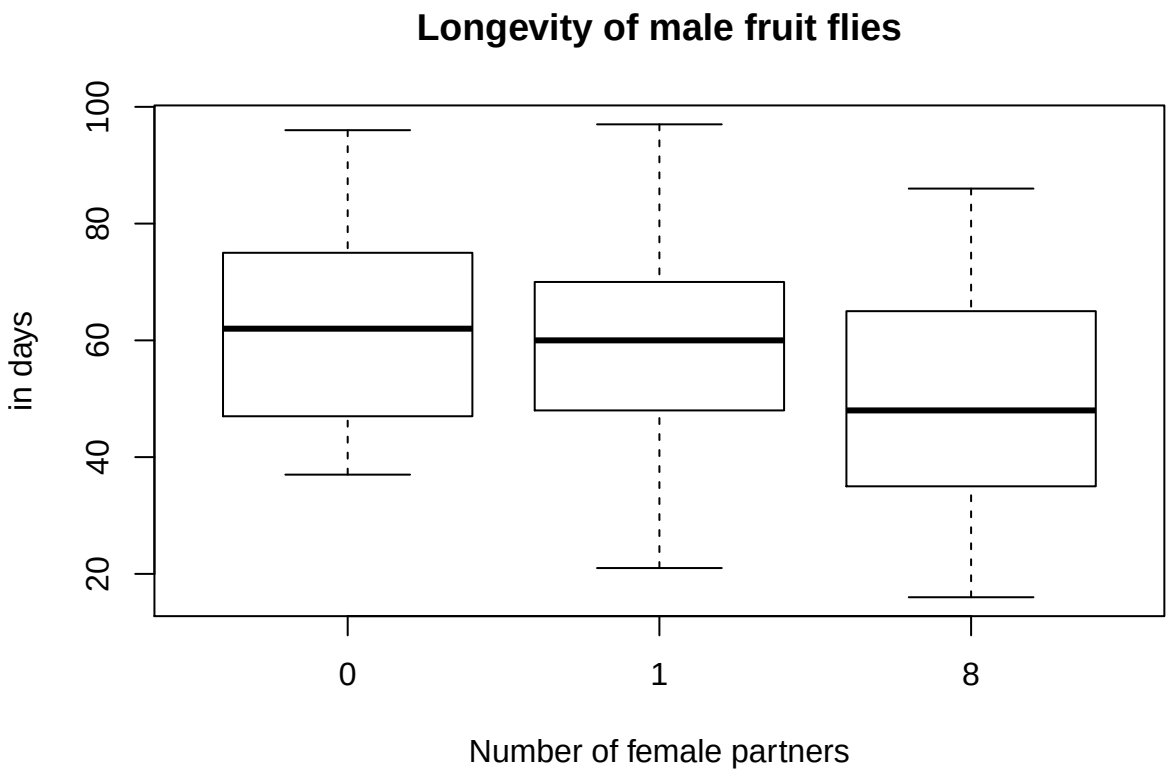
```
summary(dat)
```

```
##        ID        Partners        Type       Longevity         Thorax
##  Min.   : 1   Min.   :0.0   Min.   :0.0   Min.   :16.00   Min.   :0.640
##  1st Qu.: 7   1st Qu.:1.0   1st Qu.:0.0   1st Qu.:46.00   1st Qu.:0.760
##  Median :13   Median :1.0   Median :1.0   Median :58.00   Median :0.840
##  Mean   :13   Mean   :3.6   Mean   :2.2   Mean   :57.44   Mean   :0.821
##  3rd Qu.:19   3rd Qu.:8.0   3rd Qu.:1.0   3rd Qu.:70.00   3rd Qu.:0.880
##  Max.   :25   Max.   :8.0   Max.   :9.0   Max.   :97.00   Max.   :0.940
##      Sleep            Treatment
##  Min.   : 1.00   1 pregnant:25
##  1st Qu.:13.00   1 virgin  :25
##  Median :20.00   8 pregnant:25
##  Mean   :23.46   8 virgin  :25
##  3rd Qu.:29.00   none      :25
##  Max.   :83.00
```

The mean longevity of all male fruitflies is 57.44 days.

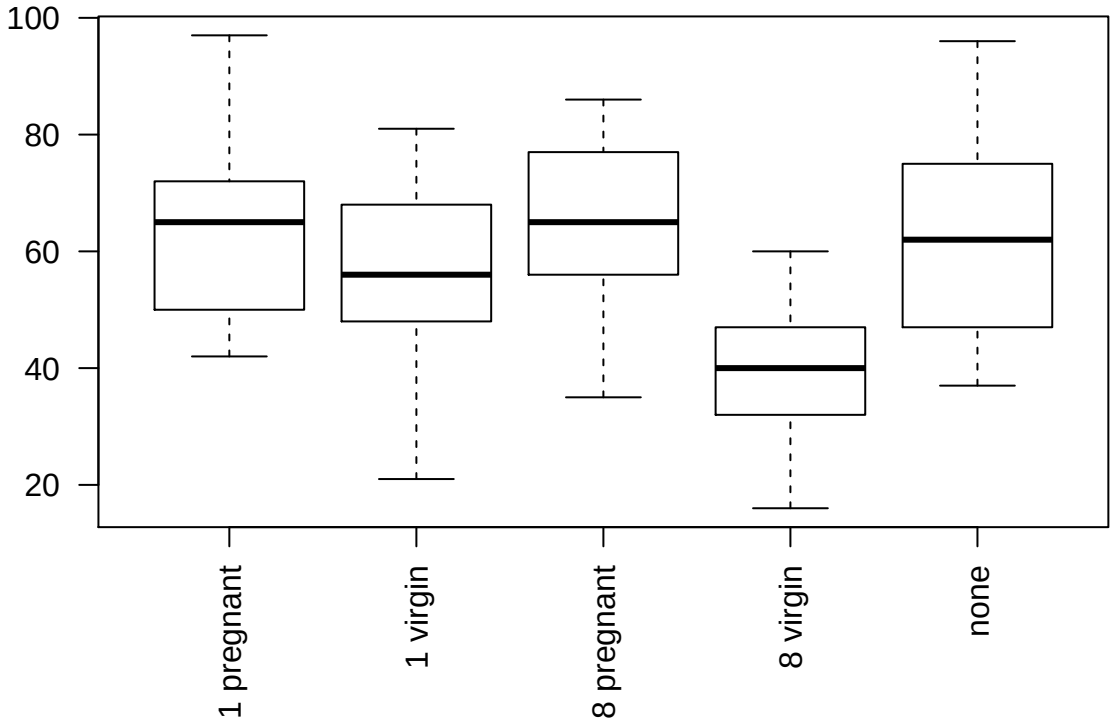
1. Construct a side by side boxplot of longevity by the number of female partners. Use the following command. Which group has the shortest longevity based on the plot?

```
##boxplot(NumericVariable~GroupVariable)
boxplot(Longevity~Partners, xlab="Number of female partners",
main="Longevity of male fruit flies", ylab= "in days")
```



2. Modify the command above to construct a side by side boxplot of longevity by **Treatment** variable. Which group of male fruit flies have the shortest / longest lifespan? Based on IQR, which group of male fruit flies have the smallest / largest variability within group?

```
boxplot(Longevity~Treatment, las=2)
```




```
##las =2 to make all labels perpendicular to axis.
```

Male fruit flies with 8 virgin female partners have the shortest longevity whereas either the group with 8 pregnant partners or no partner group have the longest longevity. Based on IQR, 8 virgin partner group has the smallest variability within group and no partner group (none) has the largest variability within group. We can verify our guesses in Problem 3 and 4.

3. Use the following commands to explore the summary statistics of the data. `tapply()` command can be used to apply a function for each group. To find sample mean `Longevity` of each five treatment, use the following command.

```
##tapply(QuantitativeVar, GroupVar, function)
tapply(Longevity, Treatment, mean)
```

```
## 1 pregnant   1 virgin 8 pregnant   8 virgin       none
##      64.80      56.76      63.36      38.72      63.56
```

4. To find the **standard deviation** and **sample size** of each treatment group, modify the following commands.

```
tapply(Longevity, Treatment, sd) ##group sample standard deviation
```

```
## 1 pregnant   1 virgin 8 pregnant   8 virgin       none
##   15.65248   14.92838   14.53983   12.10207   16.45215
```

```
tapply(Longevity, Treatment, length) ##group sample size
```

```
## 1 pregnant   1 virgin 8 pregnant   8 virgin       none
##         25         25         25         25         25
```

5. Before conducting ANOVA F-test, state the null and alternative hypotheses.

$$H_0 : \mu_{1p} = \mu_{1v} = \mu_{1v} = \mu_{8p} = \mu_{none}$$
$$H_a : \text{At least two means are different.}$$

where μ_{1v} represents the mean longevity of male fruit fly with 1 virgin partner μ_{1p} for 1 pregnant partner group, etc.

6. Before running `aov()` command in R, print the guide for this function.

```
?aov
```

We will concern ourselves with the first two arguments of this function, namely: `formula` and `data`. The `formula` argument says we need to tell R the response variable (quantitative variable) and the group variable (categorical variable). The format for `formula` is `response variable` $\sim$ `group variable`.

The next argument is `data` where we need to tell R which data set we are using. Combining both, we use the following commands

```
aov(Longevity~Treatment, data=dat)
```

```
## Call:
##    aov(formula = Longevity ~ Treatment, data = dat)
##
## Terms:
##                 Treatment Residuals
## Sum of Squares   11939.28  26313.52
## Deg. of Freedom         4       120
##
## Residual standard error: 14.80808
## Estimated effects may be unbalanced
```

```
summary(aov(Longevity~Treatment, data=dat))
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Treatment     4  11939  2984.8   13.61 3.52e-09 ***
## Residuals   120  26314   219.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Note that `aov()` output does not give us the value of F-test statistic nor the p-value whereas `summary(aov())` output provides both. The following commands produces the same output as `summary(aov())`

```
mymodel<-aov(Longevity~Treatment, data=dat)
summary(mymodel)
```

7. Based on R summary table, conduct a 5 step hypothesis test (i. Assumptions, ii. Hypotheses, iii. Test-statistic, iv. P-value, v. Conclusion and interpretation).

i) Random sample, approximately normally distributed, constant variance : not violated
ii) Hypotheses from Problem 5
iii) F-test statistic : $F = \frac{MSG}{MSE} \sim F_{df1=(g-1),df2=(N-g)}$ **The value of F-test statistic = 13.61**
iv) P-value =3.52e-09 ≈ 0**. If the null hypothesis is true, the probability that we obtain the test statistic of 13.61 and larger is nearly 0. v) Conclusion: We reject the null hypotheses and conclude that at least two treatment groups have statistically significantly different mean longevity.**

8. Based on your answer from 7, do we need to any follow up procedure? If so, use `TukeyHSD(yourAovModel)` and interpret the result.

Yes, we do to find out which two or more groups have different means and how much the difference exist.

```
mymodel<-aov(Longevity~Treatment, data=dat)
TukeyHSD(mymodel)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Longevity ~ Treatment, data = dat)
##
## $Treatment
##                           diff        lwr        upr     p adj
## 1 virgin-1 pregnant    -8.04 -19.640468   3.560468 0.3126549
## 8 pregnant-1 pregnant  -1.44 -13.040468  10.160468 0.9969591
## 8 virgin-1 pregnant   -26.08 -37.680468 -14.479532 0.0000001
## none-1 pregnant        -1.24 -12.840468  10.360468 0.9983034
## 8 pregnant-1 virgin     6.60  -5.000468  18.200468 0.5157692
## 8 virgin-1 virgin     -18.04 -29.640468  -6.439532 0.0003240
## none-1 virgin           6.80  -4.800468  18.400468 0.4854206
## 8 virgin-8 pregnant   -24.64 -36.240468 -13.039532 0.0000004
## none-8 pregnant         0.20 -11.400468  11.800468 0.9999988
## none-8 virgin          24.84  13.239532  36.440468 0.0000003
```

Male fruitflies with 8 virgin female partners have significantly short mean longevity compared to the other 4 groups (1 pregnant, 1 virgin, 8 pregnant, none) We are confident that male fruitflies with 8 virgin female partner group has the mean longevity 14 to 37 days less than the mean longevity of male fruitflies with 1 pregnant partner, etc.