Lab 8 Handout

STAT 3011

Topics covered : Ch 14 & Ch 11

Problem 1

In this problem, we consider a fruit fly data set, which is from Hanley, J. A. and S. H. Shapiro (1994) "Sexual activity and the lifespan of male fruit flies: a dataset 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?

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.
```

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)
```

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

```
tapply(Longevity, Treatment, **function1**)

tapply(Longevity, Treatment, **function2**)
```

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

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

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). Use $\alpha = 0.05$.

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

Problem 2: Gender and happiness

A 2 by 3 contingency table below shows gender and happiness of randomly selected 1964 respondents from the General Social Survey 2012.

	Happiness		
Gender	**Not**	**Pretty**	**Very**
Female	154	592	336
Male	123	502	257

a. Identify explanatory and response variables.
b. State the assumptions of Chi-squared test of independence. Determine whether each is met.
c. State the null and alternative hypothesis.
d. Test statistic is 1.04. Use the following R command to find the p-value and draw a conclusion in context. Use $\alpha = 0.05$

```
# entering all 6 numbers from Problem 2 table into R
prob2<-matrix(c(154, 592, 336, 123, 502, 257), nrow=2, byrow=2)
chisq.test(prob2) # test for independence
```