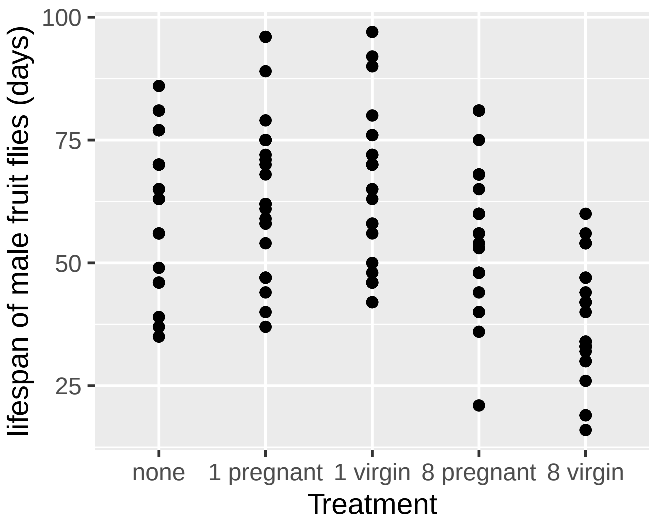
STAT 222 Spring 2022 HW2

Matthew Zhao

Question 1

```
pr3.2 = read.table("http://users.stat.umn.edu/~gary/book/fcdae.data/pr3.2",h=T)
pr3.2$trt = factor(pr3.2$trt, labels=c("none", "1 pregnant", "1 virgin",
                    "8 pregnant", "8 virgin"))
```

```
library(ggplot2)
ggplot(pr3.2, aes(x=trt, y=days))+ geom_point() +
  ylab("lifespan of male fruit flies (days)")+xlab("Treatment")
```



Q1a — 5 points

```
library(mosaic)
mean(days ~ trt, data=pr3.2)
##       none 1 pregnant   1 virgin 8 pregnant   8 virgin
##      63.36      63.56      64.80      56.76      38.72
```

```
anova(lm(days ~ as.factor(trt), data = pr3.2))
## Analysis of Variance Table
##
## Response: days
```

```
##                   Df Sum Sq Mean Sq F value       Pr(>F)
## as.factor(trt)    4  11939    2985    13.6 0.0000000035
## Residuals       120  26314     219
```

LSD $= t_{N-g,\frac{\alpha}{2}} * \sqrt{\text{MSE}(\frac{1}{n} + \frac{1}{n})}$ We find the value for $t_{N-g,\frac{\alpha}{2}}$ here:

```
qt(0.05/2,125-5,lower.tail=F)
## [1] 1.9799
```

For $\sqrt{\text{MSE}(\frac{1}{n} + \frac{1}{n})}$, each group has $n = 25$ observations and the MSE is taken from the ANOVA table above and is equal to 219 Therefore, $\sqrt{219(\frac{1}{25} + \frac{1}{25})} = 15.453$ and so LSD $= 1.9799 * 4.1857 = 8.2873$. Below we calculate the group means and :

```
## [1] "1-2 0.2   Not Significant"
## [1] "1-3 1.44   Not Significant"
## [1] "1-4 6.6   Not Significant"
## [1] "1-5 24.64   Significant"
## [1] "2-3 1.24   Not Significant"
## [1] "2-4 6.8   Not Significant"
## [1] "2-5 24.84   Significant"
## [1] "3-4 8.04   Not Significant"
## [1] "3-5 26.08   Significant"
## [1] "4-5 18.04   Significant"
```

Underline Diagram

```
 5 (8v)      4 (8p)       1 (none)      2 (1p)      3 (1v)
  38.72       56.76         63.36         63.56       64.80
            ------------------------------------------------
```

Q1b — 6 points

Construct a 95% confidence for each of the contrasts below;

- $C_1 = \mu_0 - (\mu_{8p} + \mu_{8v})/2$ .... (none v.s. 8 female companion)
- $C_2 = (\mu_{1v} + \mu_{8v})/2 - (\mu_{1p} + \mu_{8p})/2$ .... (virgin v.s. pregnant)

A 95% confidence interval for a contrast equals $\hat{C}_{k.} \pm t_{N-g,\frac{\alpha}{2}} \times SE(\hat{C})$ where $SE(\hat{C}) = \sqrt{MSE * \sum_{i=1}^{g} \frac{c_i^2}{n_i}}$

For the first contrast: $\hat{C}_1 = \sum_{i=1}^{g} c_i \bar{y}_{i.} = 1 * \bar{y}_{0.} + -0.5 * \bar{y}_{8p.} + -0.5 * \bar{y}_{8v.} = 1 * 63.36 + -0.5 * 56.76 + -0.5 * 38.72 =$ 15.62

$$SE(\hat{C}_1) = \sqrt{MSE * \sum_{i=1}^{g} \frac{c_i^2}{n_i}} = \sqrt{219 * (\frac{1^2}{25} + \frac{(-0.5)^2}{25} + \frac{(-0.5)^2}{25})} = 3.6249$$

CI: $15.62 \pm 1.9799 \times 3.6249 = C_1 \in (8.4431, 22.797)$

For second contrast: $\hat{C}_2 = \sum_{i=1}^{g} c_i \bar{y}_{i\cdot} = 0.5 * \bar{y}_{1v\cdot} + 0.5 * \bar{y}_{8v\cdot} + -0.5 * \bar{y}_{1p\cdot} + -0.5 * \bar{y}_{8p\cdot} = 0.5 * 64.80 + 0.5 * 38.72 + -0.5 * 63.56 + -0.5 * 56.76 = -8.4$

$$SE(\hat{C}_2) = \sqrt{MSE * \sum_{i=1}^{g} \frac{c_i^2}{n_i}} = \sqrt{219 * (\frac{0.5^2}{25} + \frac{0.5^2}{25} + \frac{(-0.5)^2}{25} + \frac{(-0.5)^2}{25})} = 2.9597$$

CI: $-8.4 \pm 1.9799 \times 2.9597 = C_2 \in (-14.26, -2.5401)$

From the confidence intervals, we can conclude that having 8 female companions reduces a male fly's lifespan and that having unreceptive companions positively impacts a fly's lifespan.

Q1c — 4 points

$C = (\mu_{1p} + \mu_{1v})/2 - (\mu_{8p} + \mu_{8v})/2$ ....(effect of receptive/not receptive across number of companions)

$$t = \frac{\hat{C}}{\mathrm{SE}(\hat{C})} = \frac{\sum_{i=1}^{g} c_i \bar{y}_{i\cdot}}{\sqrt{MSE * \sum_{i=1}^{g} \frac{c_i^2}{n_i}}}$$

$\hat{C} = \sum_{i=1}^{g} c_i \bar{y}_{i\cdot} = 0.5*\bar{y}_{1p\cdot} + 0.5*\bar{y}_{1v\cdot} + -0.5*\bar{y}_{8p\cdot} + -0.5*\bar{y}_{8v\cdot} = 0.5*63.56 + 0.5*64.80 + -0.5*56.76 + -0.5*38.72 = 16.44$

$$SE(\hat{C}) = \sqrt{MSE * \sum_{i=1}^{g} \frac{c_i^2}{n_i}} = \sqrt{219 * (\frac{0.5^2}{25} + \frac{0.5^2}{25} + \frac{(-0.5)^2}{25} + \frac{(-0.5)^2}{25})} = 2.9597$$

$$t = \frac{16.44}{2.9597} = 5.5546$$

```
2*pt(5.5546,125-5,lower.tail = F)
## [1] 0.0000001701
```

We can conclude at the 95% confidence level that the effect of receptive vs nonreceptive companions on flies' lifespan does change with the number of companions since $p < 0.05$.

Q1d — 4 points

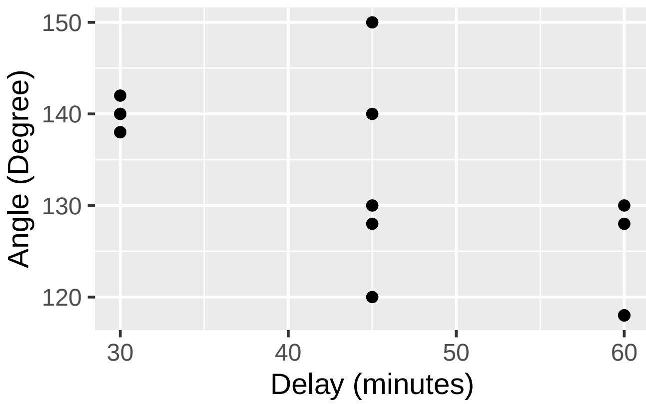
No, we cannot conclude that reproductive activities shorten males' mean lifespan because we also increase the number of companions between the 'none' and '1 virgin' group to the '8 virgin' group. This means that we could be confounding the effect of reproductive activities with increasing the number of companions. As a result, we need to have control groups, namely '1 pregnant' and '8 pregnant' so we can control for number of companions and isolate the effect of reproductive activities on males' mean lifespan.

Question 2 — 6 points

```
ex3.5 = read.table("http://users.stat.umn.edu/~gary/book/fcdae.data/ex3.5", h=T)
ex3.5$trt = 15+15*ex3.5$trt
```

Here is a plot of the data.

```r
ggplot(ex3.5, aes(x=trt, y=deg))+ geom_point() +
  ylab("Angle (Degree)")+xlab("Delay (minutes)")
```



We define the contrast as $C = (\mu_{30} - \mu_{45})/15 - (\mu_{45} - \mu_{60})/15$, comparing the slopes to see if the relationship is the same. This simplifies to $C = (\mu_{30} + \mu_{60} - 2\mu_{45})/15$

Here we find the group means:

```r
ex3.5 %>% group_by(trt) %>% summarise(means = mean(deg))
## # A tibble: 3 x 2
##     trt means
##   <dbl> <dbl>
## 1    30  140.
## 2    45  134.
## 3    60  122.
```

$$\hat{C} = \sum_{i=1}^{g} c_i \bar{y}_{i\cdot} = \frac{1}{15} * \bar{y}_{30} + \frac{1}{15} * \bar{y}_{60} - \frac{2}{15} * \bar{y}_{45} = \frac{1}{15} * 139.6 + \frac{1}{15} * 122.4 - \frac{2}{15} * 133.6 = -0.34667$$

$SE(\hat{C}) = \sqrt{MSE * \sum_{i=1}^{g} \frac{c_i^2}{n_i}}$ We get the MSE from the ANOVA table below:

```r
anova(lm(deg ~ as.factor(trt), data = ex3.5))
## Analysis of Variance Table
##
## Response: deg
##                Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(trt)  2    762     381    6.56  0.012
## Residuals      12    698      58
```

$$SE(\hat{C}) = \sqrt{55 * (\frac{(\frac{1}{15})^2}{5} + \frac{(\frac{1}{15})^2}{5} + \frac{(\frac{-2}{15})^2}{5})} = 0.5416$$

$$t = \frac{-0.34667}{0.5416} = -0.64008$$

```r
2*pt(abs(-0.64008),15-3,lower.tail = F)
## [1] 0.53415
```

Since $p > 0.05$, we fail to reject the null hypothesis that the relationship is linear, meaning that it is likely that the relationship is indeed linear.