lab5-Chen

Weixuan Chen

3/1/2023

data preparation

```
d1 <- read.csv("http://faraway.neu.edu/biostats/lab5_dataset1.csv")
d2 <- read.csv("http://faraway.neu.edu/biostats/lab5_dataset2.csv")
```

```
d1
```

```
##      treatment  mortality
## 1      control  1.4528777
## 2      control  3.2662526
## 3      control  1.1786520
## 4      control 13.4003497
## 5      control  3.7791827
## 6      control  1.1966567
## 7      control  4.4257026
## 8      control  5.6878067
## 9      control  4.8345176
## 10     control  2.0029310
## 11     control 12.3268667
## 12     control  4.0142207
## 13     control  1.4604716
## 14     control  0.2967991
## 15     control  8.3723191
## 16     control  2.5988431
## 17     control  2.6746265
## 18     control  6.9854975
## 19     control  6.1794001
## 20     control  4.9229174
## 21     control  6.8139867
## 22     control  5.9425379
## 23     control  2.9287186
## 24     control  0.3718177
## 25     control  5.0522099
## 26     control  2.5699110
## 27     control  2.3261266
## 28     control  0.6245322
## 29     control  1.6851422
## 30     control  4.1286132
## 31          t1 14.2774240
## 32          t1  3.3108742
```

```
## 33        t1  5.4068767
## 34        t1  3.4770873
## 35        t1  0.9258347
## 36        t1  2.4229976
## 37        t1  2.4736877
## 38        t1  3.4579869
## 39        t1 11.0234561
## 40        t1  7.8709262
## 41        t1  3.1126561
## 42        t1  2.8480607
## 43        t1  7.3666523
## 44        t1  6.4023378
## 45        t1  1.8427229
## 46        t1  1.8085128
## 47        t1  5.2834641
## 48        t1  7.9132053
## 49        t1  3.2793781
## 50        t1  8.8561110
## 51        t1  5.4635889
## 52        t1  1.9896796
## 53        t1  5.1609449
## 54        t1  1.1860600
## 55        t1 15.3793193
## 56        t1 26.5864024
## 57        t1  2.5415612
## 58        t1  1.2915788
## 59        t1  6.4864775
## 60        t1  3.2057478
## 61        t2 81.5827433
## 62        t2  7.1047246
## 63        t2 14.7278368
## 64        t2  7.5988898
## 65        t2  3.5139009
## 66        t2  8.9244286
## 67        t2  1.2153613
## 68        t2 31.9942072
## 69        t2  8.6128333
## 70        t2 64.8846884
## 71        t2 11.8877627
## 72        t2  3.6329812
## 73        t2 13.6089322
## 74        t2  2.9034578
## 75        t2  2.1093221
## 76        t2  9.8892295
## 77        t2  4.7431816
## 78        t2  7.3972281
## 79        t2  7.9593021
## 80        t2  4.0979181
## 81        t2  4.1842659
## 82        t2  6.4547829
## 83        t2 24.0007960
## 84        t2  1.6103205
## 85        t2 13.3824773
## 86        t2 10.3083101
```

```
## 87           t2 21.3937717
## 88           t2  5.4510927
## 89           t2 10.6975935
## 90           t2  9.6513595
## 91           t3  1.5800871
## 92           t3  9.0963006
## 93           t3  8.6746295
## 94           t3  5.4751170
## 95           t3 13.2876290
## 96           t3  4.7516239
## 97           t3  0.7583637
## 98           t3  1.5322459
## 99           t3  0.7988256
## 100          t3  1.6931647
## 101          t3  1.4617485
## 102          t3  2.8352096
## 103          t3  1.0931663
## 104          t3  3.1836514
## 105          t3  1.4125765
## 106          t3 15.9154012
## 107          t3  5.5661715
## 108          t3  6.7542655
## 109          t3  3.9915729
## 110          t3 14.6168662
## 111          t3  1.4394535
## 112          t3  1.7131868
## 113          t3 11.3848354
## 114          t3  1.4180797
## 115          t3  2.2091753
## 116          t3  1.8352708
## 117          t3  1.9738918
## 118          t3  2.0562557
## 119          t3  4.4557185
## 120          t3  2.2765691
## 121          t4  7.3451669
## 122          t4 46.6670723
## 123          t4  9.8298197
## 124          t4 10.1801879
## 125          t4 11.0210740
## 126          t4 24.8452430
## 127          t4 11.3184665
## 128          t4 11.7325359
## 129          t4  6.1616187
## 130          t4  8.8086107
## 131          t4 12.9378929
## 132          t4  6.7605586
## 133          t4 20.7282228
## 134          t4  2.6687386
## 135          t4 16.5528429
## 136          t4  2.6209849
## 137          t4  9.0162082
## 138          t4  7.1830214
## 139          t4  6.3465110
## 140          t4 11.5086994
```

```
## 141        t4  1.7961412
## 142        t4 39.5111659
## 143        t4  2.3048776
## 144        t4  7.6635061
## 145        t4  3.9911519
## 146        t4  5.7498916
## 147        t4 98.2157457
## 148        t4 12.3962700
## 149        t4  3.3659137
## 150        t4  2.3617302
```
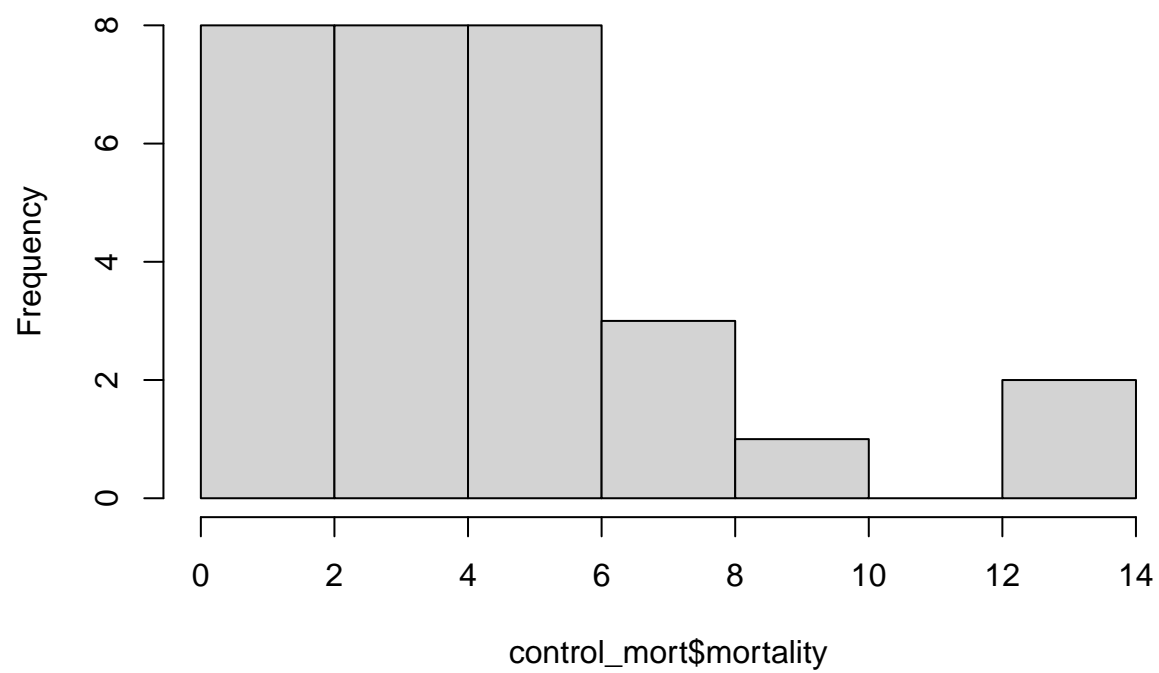
Task1

1

H0: There is no differemce in means among all groups H1: Not H0 (there is difference)

2

```
control_mort <- subset(d1, treatment == 'control')
t1_mort <- subset(d1, treatment == 't1')
t2_mort <- subset(d1, treatment == 't2')
t3_mort <- subset(d1, treatment == 't3')
t4_mort <- subset(d1, treatment == 't4')
hist(control_mort$mortality)
```
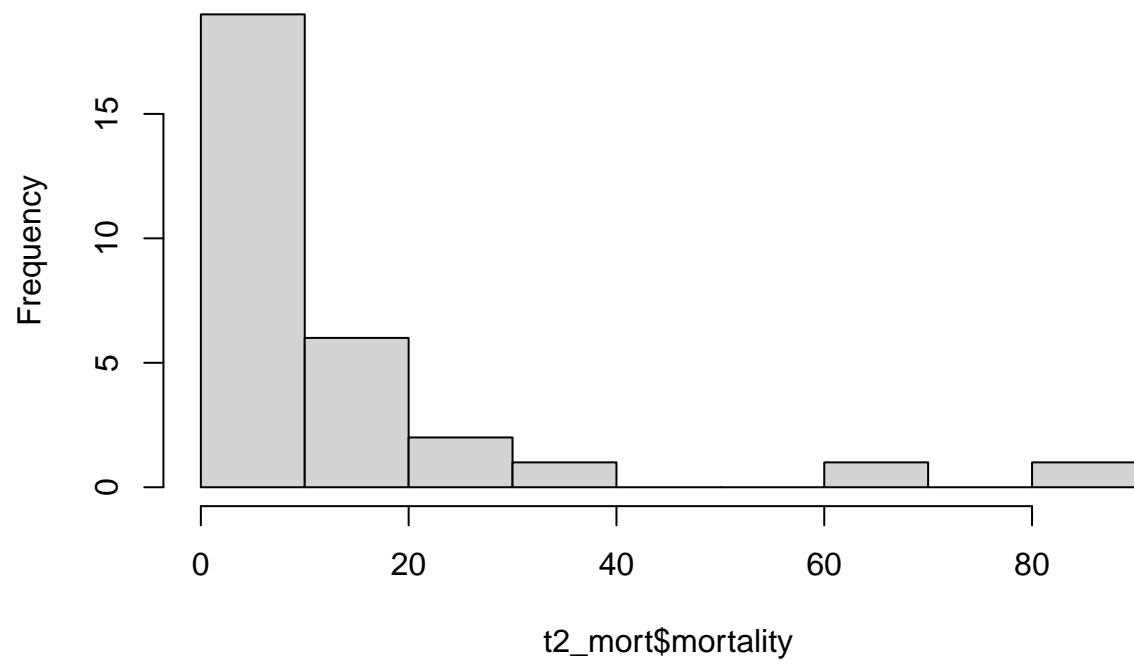
Histogram of control_mort$mortality


```
hist(t1_mort$mortality)
```
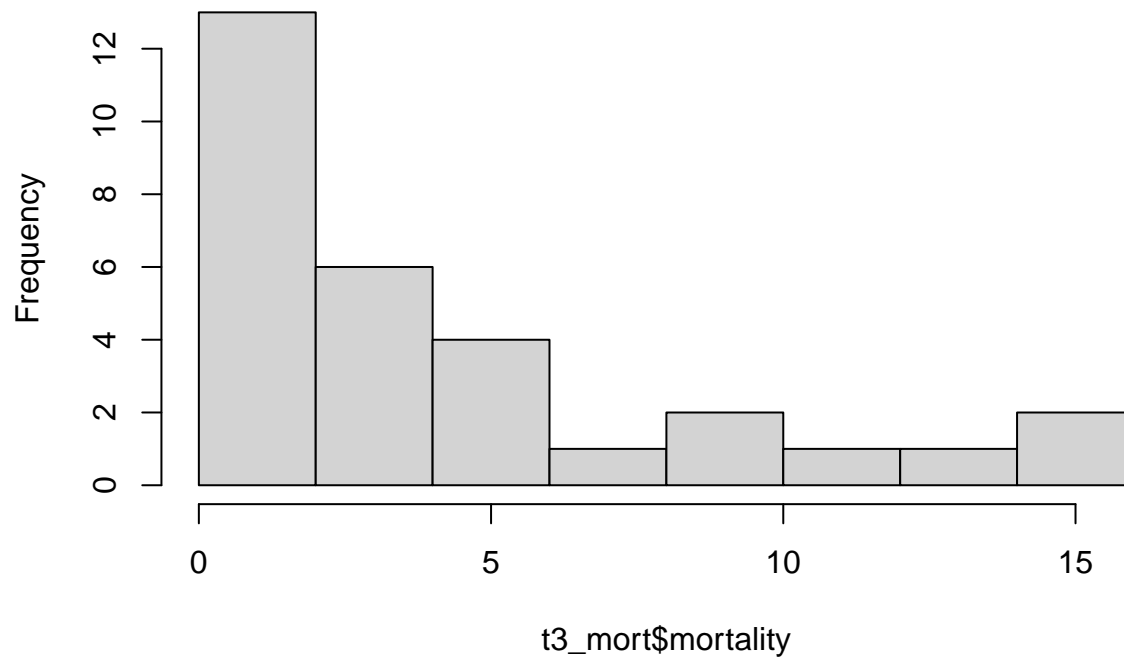
Histogram of t1_mort$mortality
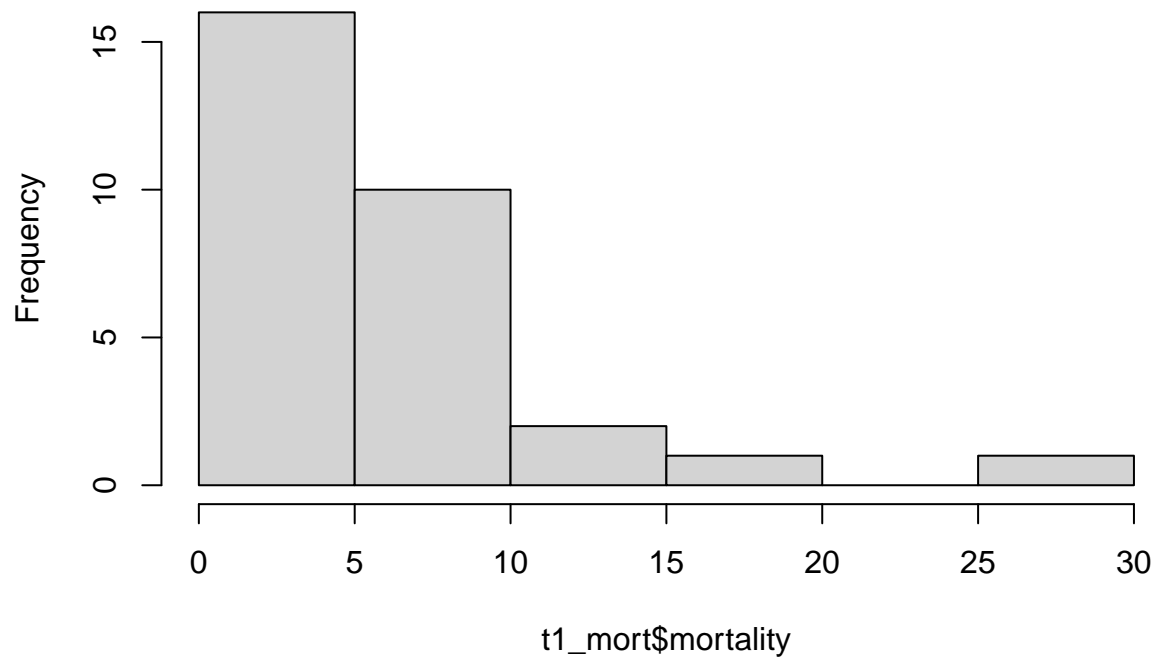

```
hist(t2_mort$mortality)
```
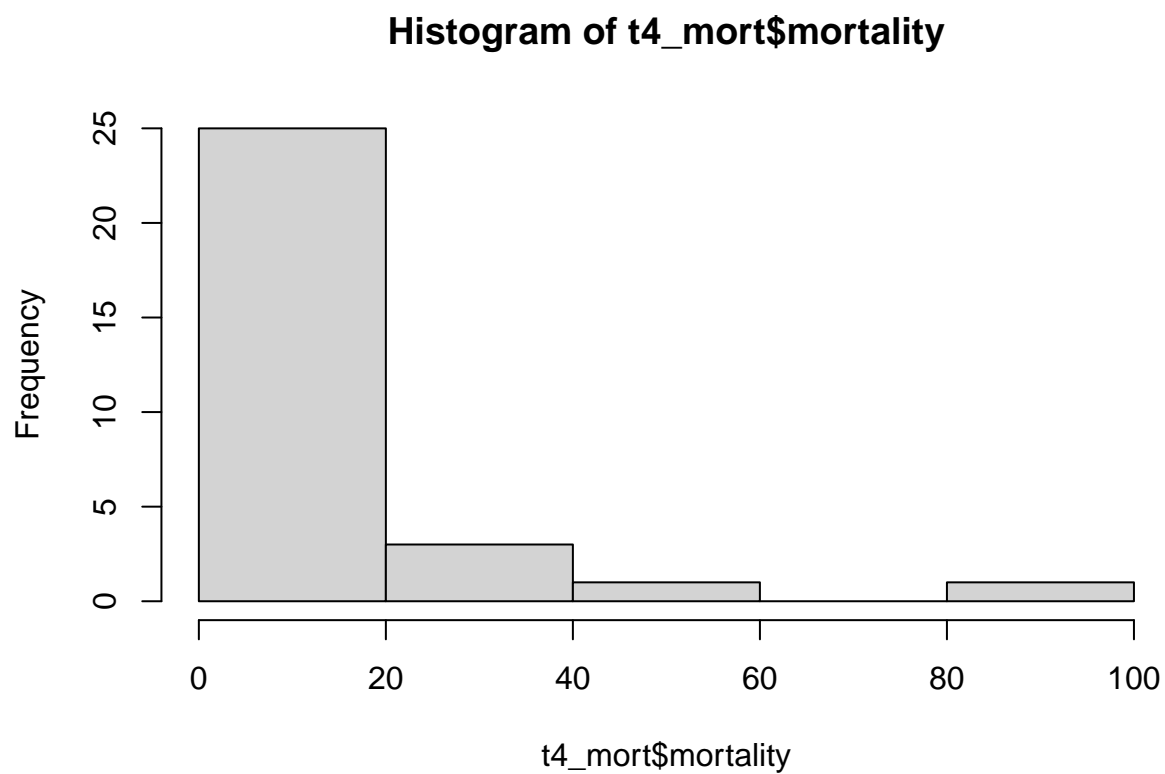
Histogram of t2_mort$mortality


```
hist(t3_mort$mortality)
```

Histogram of t3_mort$mortality


```
hist(t4_mort$mortality)
```
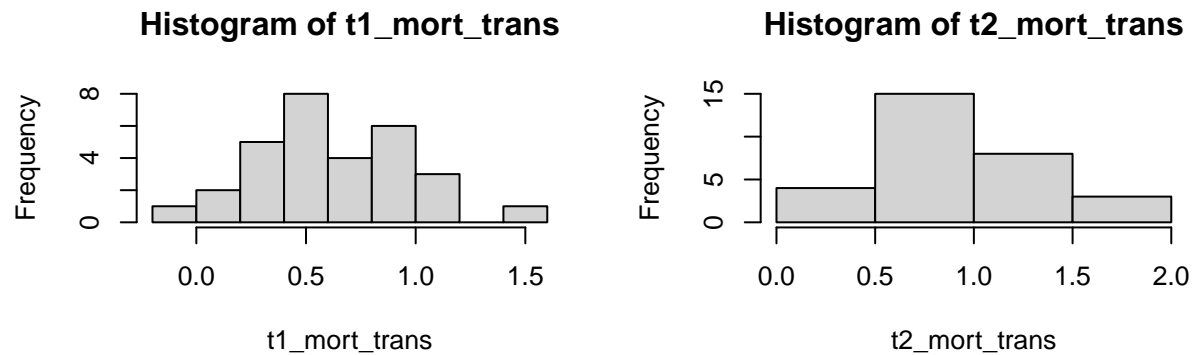
Histogram of t4_mort$mortality

No, they do not seem normally distributed.

3

```
t1_mort_trans <- log10(t1_mort$mortality)
t2_mort_trans <- log10(t2_mort$mortality)
```

```
par(mfrow=c(2,2))
hist(t1_mort_trans)
hist(t2_mort_trans)
```

Histogram of t1_mort_trans

Histogram of t2_mort_trans

Now they seem to be normally distributed. The transformation is necessary. I would use transformed data.

4

```
d1['mortality_transformed'] <- log10(d1$mortality)
mod.aov.mort <- aov(mortality ~ treatment, data = d1)
summary(mod.aov.mort)
```

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## treatment      4   2959   739.8   5.069 0.000749 ***
## Residuals    145  21162   145.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5

From te table, we can see the p-value are significant. We can reject the null hypothesis. There is difference between means among groups.

6

Because Tukey's HSD controls the family-wise error while t-tests don't.

7

```
thsd <- TukeyHSD(mod.aov.mort)
thsd
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = mortality ~ treatment, data = d1)
##
## $treatment
##                 diff        lwr         upr     p adj
## t1-control  1.6383709  -6.9782430 10.2549847 0.9846614
## t2-control  9.4007737   0.7841599 18.0173876 0.0249811
## t3-control  0.3913522  -8.2252617  9.0079661 0.9999435
## t4-control  9.9363128   1.3196989 18.5529266 0.0150140
## t2-t1       7.7624029  -0.8542110 16.3790168 0.0988879
## t3-t1      -1.2470187  -9.8636325  7.3695952 0.9945629
## t4-t1       8.2979419  -0.3186720 16.9145558 0.0650864
## t3-t2      -9.0094215 -17.6260354 -0.3928077 0.0355977
## t4-t2       0.5355390  -8.0810749  9.1521529 0.9998032
## t4-t3       9.5449606   0.9283467 18.1615744 0.0218409
```

8

```
plot(thsd)
```

95% family–wise confidence level
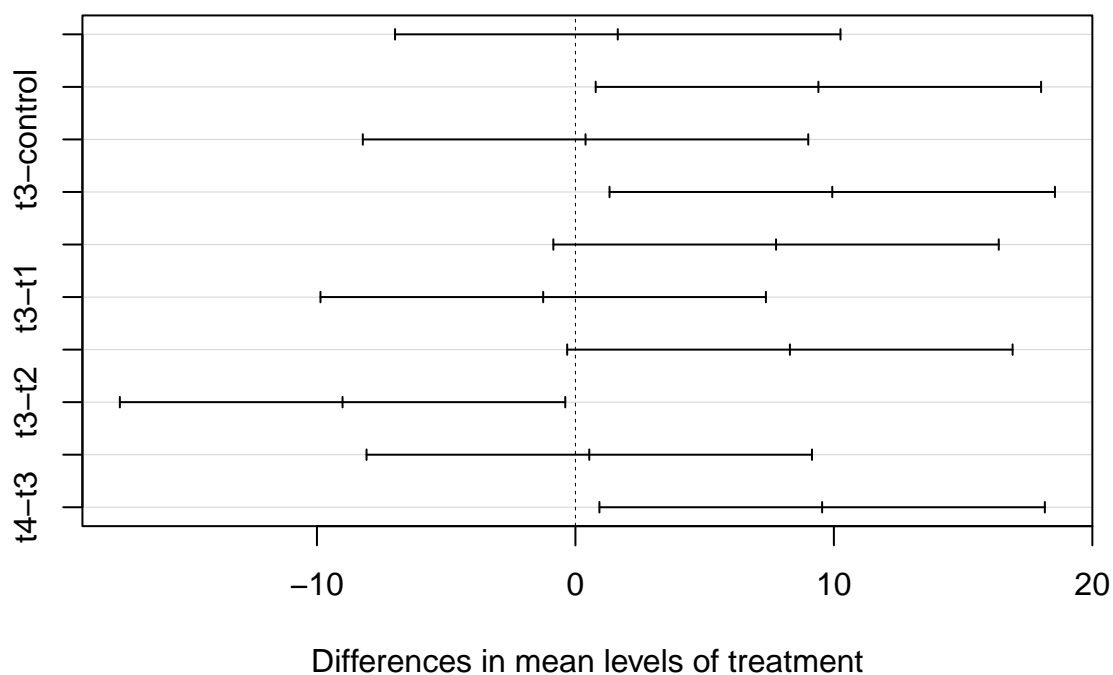
t3-t2, and t4-t3 are different from the mean

9

```r
##install.packages("multcompView")
require(multcompView)
```

```
## Loading required package: multcompView
```

10

```r
mortality_lables <-multcompLetters(thsd$treatment [, "p adj"])$Letters


mortality_lables_order <- mortality_lables[order(names(mortality_lables))]

print (mortality_lables_order)
```

```
## control      t1      t2      t3      t4
##     "b"    "ab"     "a"     "b"     "a"
```

11

```r
mean.mortality <- aggregate(mortality ~ treatment, data = d1, FUN = mean)

stderr.mortality <- aggregate(mortality ~ treatment, FUN = function(x) sd(x)/sqrt(length(x)),data = d1)

ci.upp <- mean.mortality$mortality + 1.96 * stderr.mortality$mortality

ci.low <- mean.mortality$mortality - 1.96 * stderr.mortality$mortality

bp <- barplot(mean.mortality$mortality, names = mean.mortality$treatment,
              main = "Mortality Rate for Treatments",
              ylim = c(0, max(ci.upp + 1.9)), ylab = "Mortality")

arrows(y0 = ci.low, y1 = ci.upp, x0 = bp, x1 = bp, angle = 90,
       code = 3, length = 0.1)

text(x = bp, y = mean.mortality$mortality + stderr.mortality$mortality,
     mortality_lables, pos = 3)
```
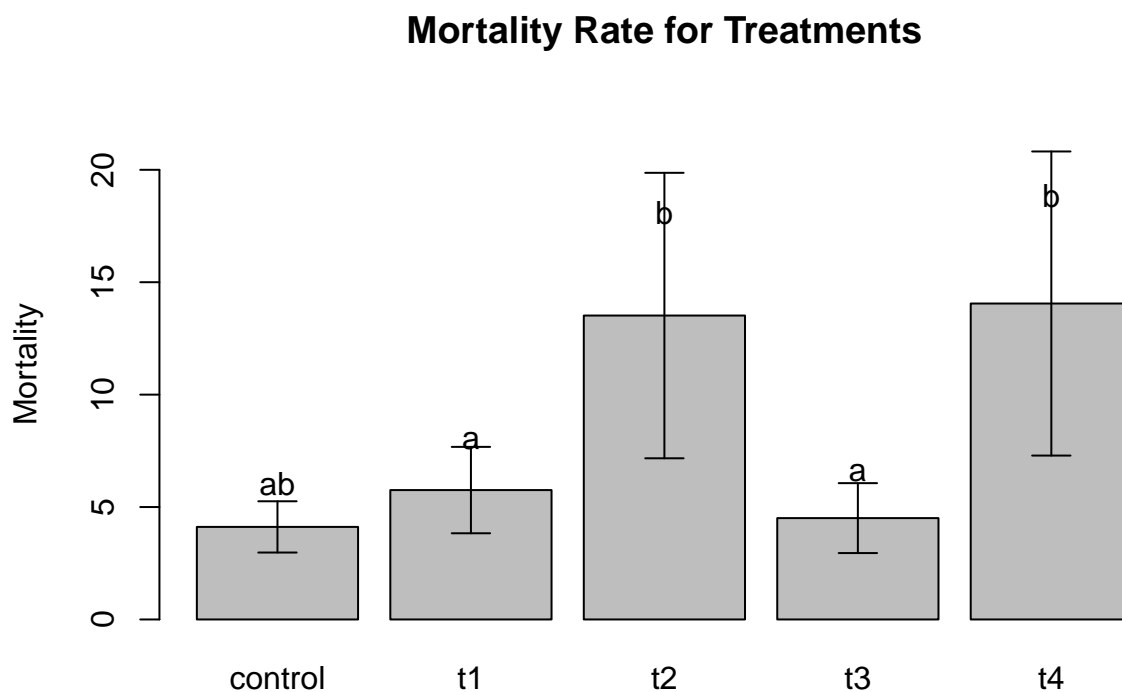

12

t4 seems most effective in increasing pest mortality because it has the hightest mortality rate and highest reaching error bar

Task 2

1

```
mortality_anova2 = aov(log10(mortality) ~ treatment, d2)

summary(mortality_anova2)
```

```
##                Df Sum Sq Mean Sq F value    Pr(>F)
## treatment       4  5.745   1.4362    9.426 8.42e-07 ***
## Residuals     145 22.093   0.1524
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA table suggests that the mean mortality is different accross the treatments

2

```
mortality_tukey2 = TukeyHSD(mortality_anova2)

print(mortality_tukey2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = log10(mortality) ~ treatment, data = d2)
##
## $treatment
##                   diff         lwr         upr       p adj
## t1-control  0.23899938 -0.03941002  0.51740878 0.1291508
## t2-control  0.44637650  0.16796710  0.72478590 0.0001783
## t3-control  0.01340866 -0.26500074  0.29181806 0.9999286
## t4-control -0.09228160 -0.37069100  0.18612780 0.8905013
## t2-t1       0.20737711 -0.07103229  0.48578651 0.2443410
## t3-t1      -0.22559072 -0.50400012  0.05281868 0.1716297
## t4-t1      -0.33128098 -0.60969038 -0.05287158 0.0109690
## t3-t2      -0.43296783 -0.71137723 -0.15455843 0.0003031
## t4-t2      -0.53865809 -0.81706749 -0.26024869 0.0000034
## t4-t3      -0.10569026 -0.38409966  0.17271914 0.8321379
```

3

```
library(multcompView)

mortality_lables2 <-multcompLetters(mortality_tukey2$treatment [, "p adj"])$Letters

mortality_lables_order2 <- mortality_lables2[order(names(mortality_lables2))]

print (mortality_lables_order2)
```

```
## control      t1       t2       t3       t4
##     "bc"     "ab"      "a"     "bc"      "c"
```

```r
mean.mortality2 <- aggregate(mortality ~ treatment, data = d2, FUN = mean)

stderr.mortality2 <- aggregate(mortality ~ treatment, FUN = function(x) sd(x)/sqrt(length(x)), data = d2

ci.upp2 <- mean.mortality2$mortality + 1.96 * stderr.mortality2$mortality

ci.low2 <- mean.mortality2$mortality - 1.96 * stderr.mortality2$mortality

bp2 <- barplot(mean.mortality2$mortality, names = mean.mortality2$treatment,
               main = "Mortality Rate for Treatments",
               ylim = c(0, max(ci.upp2 + 1.9)), ylab = "Mortality")

arrows(y0 = ci.low2, y1 = ci.upp2, x0 = bp2, x1 = bp2, angle = 90,
       code = 3, length = 0.1)

text(x = bp2, y = mean.mortality2$mortality + stderr.mortality2$mortality,
     mortality_lables2, pos = 3)
```
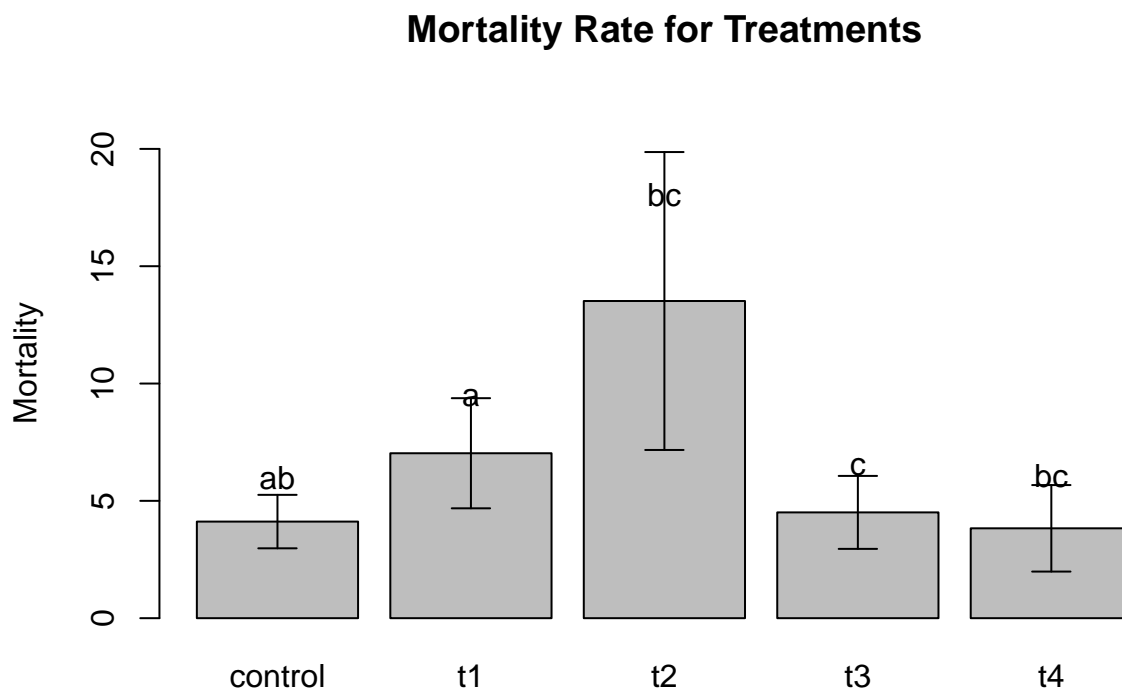
Mortality Rate for Treatments

4

t2-control, t4-t1,t3-t2, and t4-t2 are all different from the mean t4 has the lowest mortality rate for the enemy of the pest

5

t4 is the best treatment because it has the highest mortality rate for the pest, and lowest mortality rate for the enemy of the pest.