Section 5.5 How Big Is the Effect?: Confidence Intervals and Effect Sizes

Loaded needed packages

```
library(Stat2Data)
library(mosaic)
library(agricolae)     # needed for LSD.test function near the end of this file
```

EXAMPLE 5.14 Confidence interval for a group mean

 (a) Leafhoppers

Create a dataframe for **Leafhoppers** and look at the structure of the data.

```
data("Leafhoppers")
str(Leafhoppers)
```

```
## 'data.frame':    8 obs. of  3 variables:
##  $ Dish: int  1 2 3 4 5 6 7 8
##  $ Diet: Factor w/ 4 levels "Control","Fructose",..: 1 1 4 4 3 3 2 2
##  $ Days: num  2.3 1.7 3.6 4 2.9 2.7 2.1 2.3
```

Leafhopper ANOVA table (with P-value)

```
LHAnovaModel=aov(Days~Diet,data=Leafhoppers)
summary(LHAnovaModel)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## Diet          3   3.92   1.307   17.42 0.00925 **
## Residuals     4   0.30   0.075
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Compute a CI for the mean of the no sugar (Control) group, using the pieces needed for the formula "by hand."

```
ControlMean=as.numeric(mean(Days~Diet,data=Leafhoppers)["Control"])
ControlMean
```

```
## [1] 2
```

```
errordf=LHAnovaModel$df.residual
errordf
```

```
## [1] 4
```

```
tstar=qt(0.975,4)
tstar
```

```
## [1] 2.776445
```

```
MSE=summary(LHAnovaModel)[[1]]$"Mean Sq"[2]    #gets the MSE from aov summary
MSE
```

```
## [1] 0.075
```

```
SD=sqrt(MSE)    #gets sqauare root of the MSE from ANOVA
SD
```

```
## [1] 0.2738613
```

```
ME=tstar*SD/sqrt(2)
Lower=ControlMean-ME
Upper=ControlMean+ME
c(Lower,Upper)
```

```
## [1] 1.462344 2.537656
```

(b) Fruit fly lifetimes

Create a dataframe for **FruitFlies** and look at the structure of the data.

```
data("FruitFlies")
#order the Treatment categories
FruitFlies$Treatment=factor(FruitFlies$Treatment,levels=c("none","1 pregnant","8 pregnant","1 virgin","8
str(FruitFlies)
```

```
## 'data.frame':    125 obs. of  7 variables:
##  $ ID       : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Partners : int  8 8 8 8 8 8 8 8 8 8 ...
##  $ Type     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Longevity: int  35 37 49 46 63 39 46 56 63 65 ...
##  $ Thorax   : num  0.64 0.68 0.68 0.72 0.72 0.76 0.76 0.76 0.76 0.76 ...
##  $ Sleep    : int  22 9 49 1 23 83 23 15 9 81 ...
##  $ Treatment: Factor w/ 5 levels "none","1 pregnant",..: 3 3 3 3 3 3 3 3 3 3 ...
```

Fit the one-way ANOVA model.

```
FFmodel=aov(Longevity~Treatment,data=FruitFlies)
```

Compute a CI for the mean of the Control ("none") group, using the pieces needed for the formula "by hand."

```
NoneMean=as.numeric(mean(Longevity~Treatment,data=FruitFlies)["none"])
NoneMean
```

```
## [1] 63.56
```

```
errordf=FFmodel$df.residual
errordf
```

```
## [1] 120
```

```
tstar=qt(0.975,errordf)
tstar
```

```
## [1] 1.97993
```

```
MSE=summary(FFmodel)[[1]]$"Mean Sq"[2]    #gets the MSE from aov summary
MSE
```

```
## [1] 219.2793
```

```
SD=sqrt(MSE)    #gets sqauare root of the MSE from ANOVA
SD
```

```
## [1] 14.80808
```

```
ME=tstar*SD/sqrt(25)
Lower=NoneMean-ME
Upper=NoneMean+ME
c(Lower,Upper)
```

```
## [1] 57.69621 69.42379
```

EXAMPLE 5.15 Confidence interval for the difference of two group means

(a) For FruitFlies "none" (Control) versus "1 pregnant"
    Do the calculations brute force (see below for more an alternative method).

```
NoneMean=as.numeric(mean(Longevity~Treatment,data=FruitFlies)["none"])
NoneMean
```

```
## [1] 63.56
```

```
Preg1Mean=as.numeric(mean(Longevity~Treatment,data=FruitFlies)["1 pregnant"])
Preg1Mean
```

```
## [1] 64.8
```

```
Diff=NoneMean-Preg1Mean
Diff
```

```
## [1] -1.24
```

```
errordf=FFmodel$df.residual
tstar=qt(0.975,errordf)
MSE=summary(FFmodel)[[1]]$"Mean Sq"[2]    #gets the MSE from aov summary
SD=sqrt(MSE)    #gets sqauare root of the MSE from ANOVA
MEDiff=tstar*SD*sqrt(1/25+1/25)
MEDiff
```

```
## [1] 8.292658
```

```
c(Diff-MEDiff,Diff+MEDiff)
```

```
## [1] -9.532658  7.052658
```

(b) For FruitFlies "none" (Control) versus "8 virgins"

```
NoneMean=as.numeric(mean(Longevity~Treatment,data=FruitFlies)["none"])
NoneMean
```

```
## [1] 63.56
```

```
Virgin8Mean=as.numeric(mean(Longevity~Treatment,data=FruitFlies)["8 virgin"])
Virgin8Mean
```

```
## [1] 38.72
```

```
Diff=NoneMean-Virgin8Mean
Diff
```

```
## [1] 24.84
```

```
errordf=FFmodel$df.residual
tstar=qt(0.975,errordf)
MSE=summary(FFmodel)[[1]]$"Mean Sq"[2]    #gets the MSE from aov summary
SD=sqrt(MSE)    #gets sqauare root of the MSE from ANOVA
MEDiff=tstar*SD*sqrt(1/25+1/25)
MEDiff
```

```
## [1] 8.292658
```

```
c(Diff-MEDiff,Diff+MEDiff)
```

```
## [1] 16.54734 33.13266
```

EXAMPLE 5.17 Fruit flies: effect size

First, use direct compuation (for all treatments).

```
FFMeans=mean(Longevity~Treatment,data=FruitFlies)
FFalphas=FFMeans-mean(FruitFlies$Longevity)
FFalphas
```

```
##       none 1 pregnant 8 pregnant   1 virgin   8 virgin
##       6.12       7.36       5.92      -0.68     -18.72
```

```
MSE=summary(FFmodel)[[1]]$"Mean Sq"[2]   #gets the MSE from aov summary
SD=sqrt(MSE)
FFEffectSizes=FFalphas/SD
FFEffectSizes
```

```
##        none  1 pregnant  8 pregnant    1 virgin    8 virgin
##  0.41328779  0.49702583  0.39978165 -0.04592087 -1.26417441
```

Compare none (Alone) to 1 pregnant

```
Diff12=as.numeric(FFMeans[1]-FFMeans[2])
Diff12/SD
```

```
## [1] -0.08373805
```

Compare none (Alone) to 8 virgins

```
Diff15=as.numeric(FFMeans[1]-FFMeans[5])
Diff15/SD
```

```
## [1] 1.677462
```

FIGURE 5.17 What happens as the sample size increases?

First, we look at none versus 1 pregnant.

```
Ns=c(10,100,1000,10000)
SE=14.8*sqrt(1/Ns+1/Ns)
df=2*Ns-2
tstars=qt(0.975,df)
M95=tstars*SE
t=1.24/SE
pvalue=1-pt(t,df)
CvsP1=data.frame(n=Ns,SE,M95,t,pvalue)
CvsP1
```

```
##        n        SE        M95         t       pvalue
## 1     10 6.6187612 13.9055013 0.1873462 4.267424e-01
## 2    100 2.0930361  4.1275037 0.5924408 2.771157e-01
## 3   1000 0.6618761  1.2980397 1.8734624 3.057522e-02
## 4  10000 0.2093036  0.4102524 5.9244082 1.592793e-09
```

Now, we consider none versus 8 virgins.

```
Ns=c(10,100,1000,10000)
SE=14.8*sqrt(1/Ns+1/Ns)
df=2*Ns-2
tstars=qt(0.975,df)
M95=tstars*SE
t=24.84/SE
pvalue=1-pt(t,df)
CvsV8=data.frame(n=Ns,SE,M95,t,pvalue)
CvsV8
```

```
##        n        SE        M95         t       pvalue
## 1     10 6.6187612 13.9055013   3.752968 0.0007281407
## 2    100 2.0930361  4.1275037  11.867927 0.0000000000
## 3   1000 0.6618761  1.2980397  37.529681 0.0000000000
## 4  10000 0.2093036  0.4102524 118.679273 0.0000000000
```

Alternative Solutions to Example 5.15

Use a new function to compute the pairwise confidence intervals directly with the asbio package.

```
library(asbio)
```

```
lsdCI(FruitFlies$Longevity,FruitFlies$Treatment)
```

```
##
## 95% LSD confidence intervals
##
##                              LSD  Diff    Lower    Upper  Decision
## munone-mu1 pregnant      8.29266 -1.24 -9.53266  7.05266    FTR H0
## munone-mu8 pregnant      8.29266   0.2 -8.09266  8.49266    FTR H0
## mu1 pregnant-mu8 pregnant 8.29266  1.44 -6.85266  9.73266    FTR H0
## munone-mu1 virgin        8.29266   6.8 -1.49266 15.09266    FTR H0
## mu1 pregnant-mu1 virgin  8.29266  8.04 -0.25266 16.33266    FTR H0
## mu8 pregnant-mu1 virgin  8.29266   6.6 -1.69266 14.89266    FTR H0
## munone-mu8 virgin        8.29266 24.84 16.54734 33.13266 Reject H0
## mu1 pregnant-mu8 virgin  8.29266 26.08 17.78734 34.37266 Reject H0
## mu8 pregnant-mu8 virgin  8.29266 24.64 16.34734 32.93266 Reject H0
## mu1 virgin-mu8 virgin    8.29266 18.04  9.74734 26.33266 Reject H0
##                          Adj. p-value
## munone-mu1 pregnant            0.7677
## munone-mu8 pregnant           0.96199
## mu1 pregnant-mu8 pregnant     0.73159
```

```
## munone-mu1 virgin                     0.1071
## mu1 pregnant-mu1 virgin               0.05728
## mu8 pregnant-mu1 virgin               0.11771
## munone-mu8 virgin                           0
## mu1 pregnant-mu8 virgin                     0
## mu8 pregnant-mu8 virgin                     0
## mu1 virgin-mu8 virgin                  3e-05
```

Here is another method that we will use throughout the rest of Unit B. This method is called LSD.test and uses the agricolae package.

```
require(agricolae)
(LSD.test(FFmodel, "Treatment")) #Outer parentheses print the result.
```

```
## $statistics
##     MSerror  Df  Mean       CV t.value      LSD
##    219.2793 120 57.44 25.78009 1.97993 8.292658
##
## $parameters
##         test p.ajusted    name.t ntr alpha
##    Fisher-LSD      none Treatment   5  0.05
##
## $means
##             Longevity      std  r       se      LCL      UCL Min Max Q25 Q50 Q75
## 1 pregnant      64.80 15.65248 25 2.961617 58.93621 70.66379  42  97  50  65  72
## 1 virgin        56.76 14.92838 25 2.961617 50.89621 62.62379  21  81  48  56  68
## 8 pregnant      63.36 14.53983 25 2.961617 57.49621 69.22379  35  86  56  65  77
## 8 virgin        38.72 12.10207 25 2.961617 32.85621 44.58379  16  60  32  40  47
## none            63.56 16.45215 25 2.961617 57.69621 69.42379  37  96  47  62  75
##
## $comparison
## NULL
##
## $groups
##             Longevity groups
## 1 pregnant      64.80      a
## none            63.56      a
## 8 pregnant      63.36      a
## 1 virgin        56.76      a
## 8 virgin        38.72      b
##
## attr(,"class")
## [1] "group"
```

If you want to get the confidence intervals, rather than a summary of the results, you just add the option group=FALSE.

```
require(agricolae)
(LSD.test(FFmodel, "Treatment", group=FALSE)) #Outer parentheses print the result.
```

```
## $statistics
##     MSerror  Df  Mean       CV t.value      LSD
```

```
##    219.2793 120 57.44 25.78009 1.97993 8.292658
##
## $parameters
##        test p.ajusted   name.t ntr alpha
##   Fisher-LSD      none Treatment   5  0.05
##
## $means
##          Longevity      std  r       se      LCL      UCL Min Max Q25 Q50 Q75
## 1 pregnant    64.80 15.65248 25 2.961617 58.93621 70.66379  42  97  50  65  72
## 1 virgin      56.76 14.92838 25 2.961617 50.89621 62.62379  21  81  48  56  68
## 8 pregnant    63.36 14.53983 25 2.961617 57.49621 69.22379  35  86  56  65  77
## 8 virgin      38.72 12.10207 25 2.961617 32.85621 44.58379  16  60  32  40  47
## none         63.56 16.45215 25 2.961617 57.69621 69.42379  37  96  47  62  75
##
## $comparison
##                    difference pvalue signif.        LCL        UCL
## 1 pregnant - 1 virgin       8.04 0.0573      .  -0.2526583  16.332658
## 1 pregnant - 8 pregnant     1.44 0.7316         -6.8526583   9.732658
## 1 pregnant - 8 virgin      26.08 0.0000    ***  17.7873417  34.372658
## 1 pregnant - none           1.24 0.7677         -7.0526583   9.532658
## 1 virgin - 8 pregnant      -6.60 0.1177        -14.8926583   1.692658
## 1 virgin - 8 virgin        18.04 0.0000    ***   9.7473417  26.332658
## 1 virgin - none            -6.80 0.1071        -15.0926583   1.492658
## 8 pregnant - 8 virgin      24.64 0.0000    ***  16.3473417  32.932658
## 8 pregnant - none          -0.20 0.9620         -8.4926583   8.092658
## 8 virgin - none           -24.84 0.0000    *** -33.1326583 -16.547342
##
## $groups
## NULL
##
## attr(,"class")
## [1] "group"
```

The nice thing about LSD.test is that you can also plot the results to get a nice visual summary.

```
require(agricolae)
plot(LSD.test(FFmodel, "Treatment"))
```

Groups and Range