

Exercise – One-Way ANOVA

Answer the following questions with R code by creating (*and editing if you make a mistake*) an R script and iteratively running the code in RStudio.

1. Load the data in the **InchLake2.csv** file into a data frame in R. Isolate the Bluegill data and then isolate the Bluegill data from 2007 (i.e., should result in two data frames – Bluegills and Bluegills in 2007). Note that length is inches and weight is in grams.

```
> setwd("C:/aaaWork/Web/fishR/courses/Vermont2014/CourseMaterial/Exercises")
> IL <- read.csv("Data/InchLake2.csv")
> IL$lenmm <- IL$length*25.4
> bg <- Subset(IL,species=="Bluegill")
> bg07 <- Subset(bg,year==2007)
> bg07 <- Subset(bg07,complete.cases(bg07[,c("lenmm","weight"))))
```

2. Construct relative weight and five-cell length class variables.

```
> ( wsBG <- wsVal("Bluegill") )
      species units   type ref      int slope quad min.len max.len measure method
20 Bluegill metric linear  75 -5.374 3.316   NA      80      NA      TL      Other
      comment      source
20      none Hillman, 1982
> ( psdBG <- psdVal("Bluegill") )
      zero      stock      quality preferred memorable      trophy
      0          80          150          200          250          300
> bg07 <- Subset(bg07,lenmm>=80)
> bg07 <- within(bg07,{
  Ws <- 10^(wsBG$int)*lenmm^wsBG$slope
  Wr <- weight/Ws*100
})
> bg07 <- lencat(~lenmm,data=bg07,breaks=psdBG)
> view(bg07)
```

	netID	fishID	species	length	weight	year	lenmm	Wr	Ws	LCat
3	15	203	Bluegill	5.6	40	2007	142.24	68.65	58.27	80
6	16	210	Bluegill	9.0	280	2007	228.60	99.65	280.99	200
12	21	217	Bluegill	3.7	14	2007	93.98	94.96	14.74	80
19	28	230	Bluegill	5.4	50	2007	137.16	96.81	51.65	80
42	9	161	Bluegill	8.1	175	2007	205.74	88.32	198.14	200
75	4	119	Bluegill	7.6	145	2007	193.04	90.40	160.40	150

3. Compute summary statistics of relative weight of Bluegill by five-cell length classes.

```
> sumWr <- Summarize(Wr~LCat,data=bg07,digits=1)
> sumWr <- within(sumWr, {
  LCI <- mean-1.96*sd/sqrt(n)
  UCI <- mean+1.96*sd/sqrt(n)
})
> sumWr
```

	LCat	n	mean	sd	min	Q1	median	Q3	max	percZero	UCI	LCI
1	80	24	80.0	18.8	45.8	68.5	84.8	92.5	116	0	87.52	72.48
2	150	38	89.8	7.5	71.2	85.0	90.9	95.3	102	0	92.18	87.42
3	200	21	96.3	5.0	88.3	92.0	97.3	99.7	103	0	98.44	94.16

4. Statistically determine if there is a difference in relative weight among length classes. If so, which length classes differ? [*For simplicity, ignore assumption violations at this point.*]

```

> lm1 <- lm(Wr~LCat,data=bg07)
> residPlot(lm1)
> outlierTest(lm1)

No Studentized residuals with Bonferonni p < 0.05
Largest |rstudent|:
      rstudent unadjusted p-value Bonferonni p
11      3.393          0.001082      0.08984
> leveneTest(lm1)

Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2    15.8 1.7e-06
      80
> anova(lm1)

Analysis of Variance Table

Response: Wr
      Df Sum Sq Mean Sq F value  Pr(>F)
LCat    2   3064    1532    11.4 4.3e-05
Residuals 80  10727     134
> mc1 <- glht(lm1,mcp(LCat="Tukey"))
> summary(mc1)

Simultaneous Tests for General Linear Hypotheses

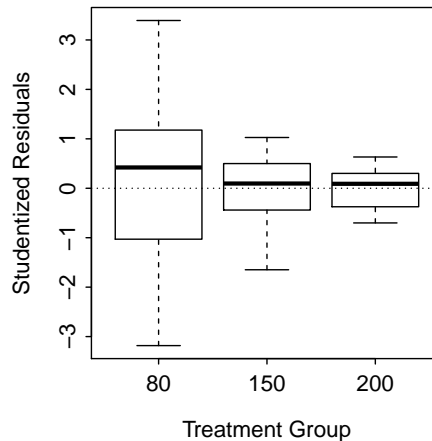
Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = Wr ~ LCat, data = bg07)

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
150 - 80 == 0      9.82      3.02   3.25  0.0046
200 - 80 == 0     16.26      3.46   4.70 <1e-04
200 - 150 == 0      6.43      3.15   2.04  0.1079
(Adjusted p values reported -- single-step method)
> cld(mc1)

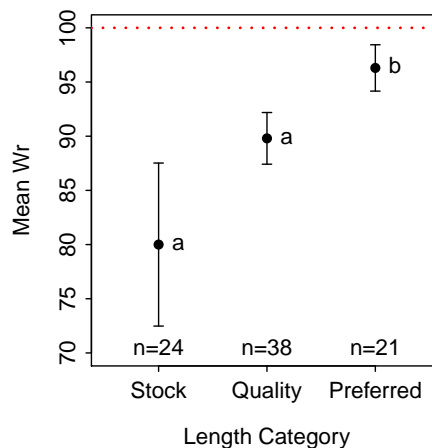
      80 150 200
"a" "b" "b"

```



5. Construct a plot of mean relative weights with symbols that represent significant differences.

```
> with(sumWr, plotCI(1:3, mean, ui=UCI, li=LCI, pch=16, xlim=c(0.5, 3.5), xaxt="n",
  ylim=c(70, 100), xlab="Length Category", ylab="Mean Wr"))
> axis(1, 1:3, c("Stock", "Quality", "Preferred"))
> abline(h=100, col="red", lty=3, lwd=2)
> text(1:3, sumWr$mean, c("a", "a", "b"), pos=c(4, 4, 4))
> text(1:3, 70.5, paste("n=", sumWr$n, sep=""))
```



6. If time permits ... repeat the above analysis for Bluegills captured in 2008.

```
> bg08 <- Subset(bg, year==2008)
> bg08 <- Subset(bg08, complete.cases(bg07[, c("lenmm", "weight")]))
> bg08 <- Subset(bg08, lenmm >= 80)
> bg08 <- within(bg08, {
  Ws <- 10^(wsBG$int) * lenmm^wsBG$slope
  Wr <- weight / Ws * 100
})
> bg08 <- lencat(~lenmm, data=bg08, breaks=psdBG)
>
> sumWr2 <- Summarize(Wr ~ LCat, data=bg08, digits=1)
> sumWr2 <- within(sumWr2, {
  LCI <- mean - 1.96 * sd / sqrt(n)
  UCI <- mean + 1.96 * sd / sqrt(n)
})
>
> lm2 <- lm(Wr ~ LCat, data=bg08)
> outlierTest(lm2)
```

```

No Studentized residuals with Bonferonni p < 0.05
Largest |rstudent|:
      rstudent unadjusted p-value Bonferonni p
67    -3.076          0.002948      0.227
> leveneTest(lm2)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value  Pr(>F)
group  2    18.6 2.9e-07
      74
> anova(lm2)
Analysis of Variance Table

Response: Wr
      Df Sum Sq Mean Sq F value Pr(>F)
LCat    2   4224    2112    20.1 1e-07
Residuals 74   7762     105
> mc2 <- glht(lm2,mcp(LCat="Tukey"))
> summary(mc2)

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: lm(formula = Wr ~ LCat, data = bg08)

Linear Hypotheses:
      Estimate Std. Error t value Pr(>|t|)
150 - 80 == 0    13.75      2.85   4.83 <1e-04
200 - 80 == 0    17.67      2.93   6.04 <1e-04
200 - 150 == 0     3.92      2.82   1.39  0.35
(Adjusted p values reported -- single-step method)
> cld(mc2)
      80 150 200
"a" "b" "b"

> with(sumWr2,plotCI(1:3,mean,ui=UCI,li=LCI,pch=16,xlim=c(0.5,3.5),xaxt="n",
                    ylim=c(60,100),xlab="Length Category",ylab="Mean Wr"))
> axis(1,1:3,c("Stock","Quality","Preferred"))
> abline(h=100,col="red",lty=3,lwd=2)
> text(1:3,sumWr2$mean,c("a","b","b"),pos=c(4,4,4))
> text(1:3,60.5,paste("n=",sumWr2$n,sep=""))

```

