

# One-Way ANOVA

---

## Preliminaries

### Load Necessary Packages

```
> library(car)      # leveneTest()
> library(multcomp)  # for glht(), mcp() DO BEFORE dplyr!!
> library(FSA)       # for filterD(), hist(), Summarize()
> library(dplyr)     # for mutate(), select()
> library(plotrix)   # for plotCI()
```

### Load Data

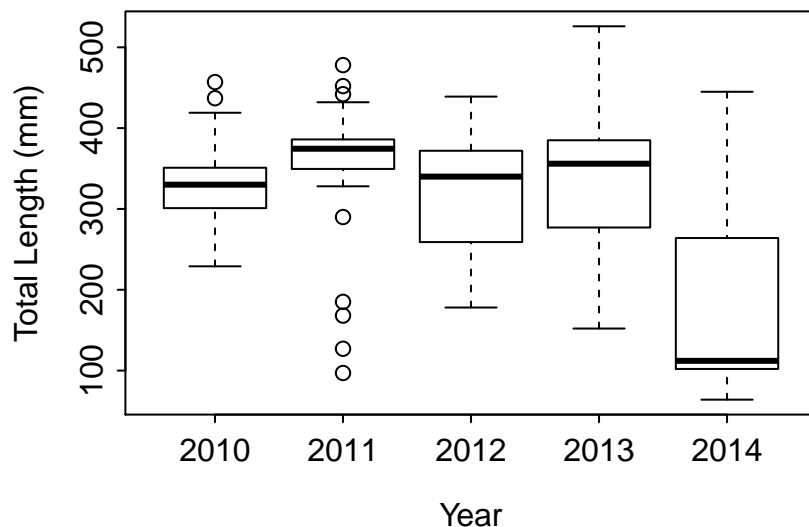
```
> # Set your working directory to where your external data files (and scripts) are located.
> setwd("C:/aaaWork/Web/GitHub/RcourseNunavut2016/Handouts")
> dSC <- read.csv("SawyerCo_reduced.csv")
> dSC <- mutate(dSC,sex=mapvalues(sex,from="",to="ND"),fyear=factor(year))
> LChip_LMB <- filterD(dSC,waterbody=="LAKE CHIPPEWA",species=="Largemouth Bass")
```

### Quick Summaries

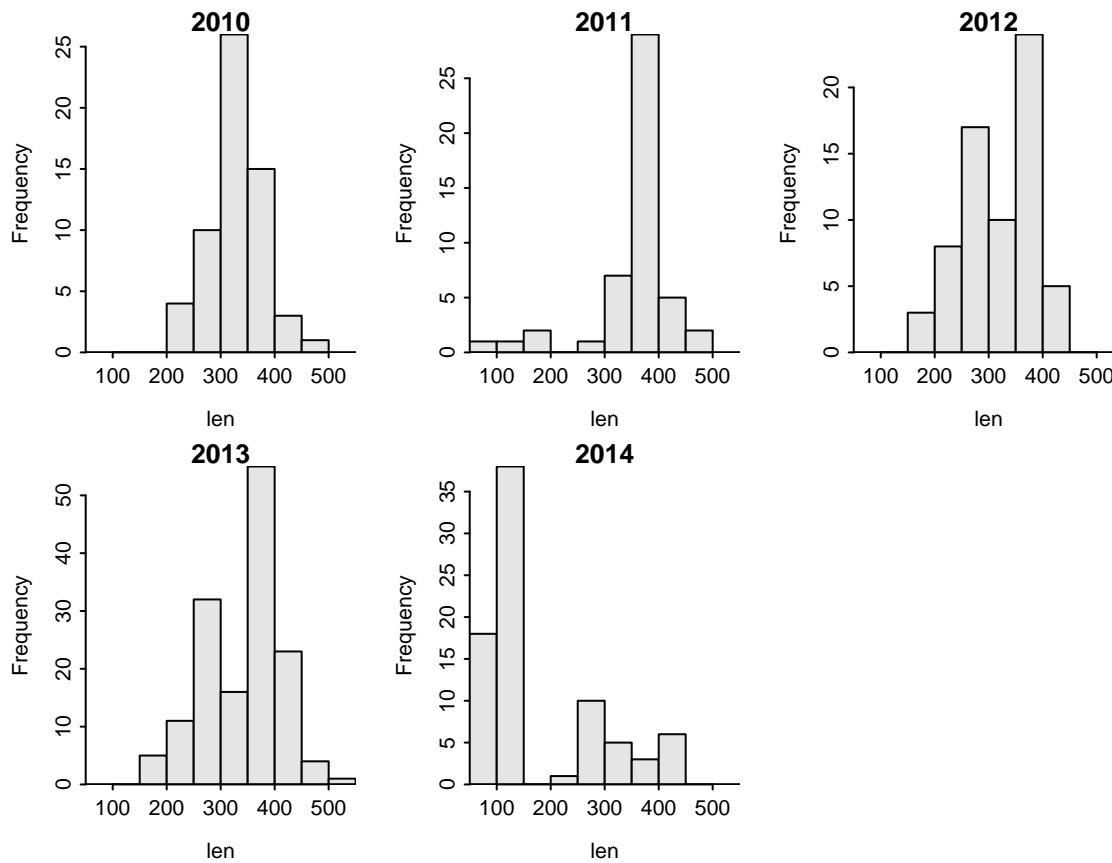
```
> Summarize(len~fyear,data=LChip_LMB)
```

	fyear	n	nvalid	mean	sd	min	Q1	median	Q3	max	percZero
1	2010	59	59	326.9831	45.84400	229	301.0	330.0	351	457	0
2	2011	48	48	357.1458	73.51769	97	350.2	374.5	386	478	0
3	2012	67	67	317.9851	67.17041	178	259.0	340.0	372	439	0
4	2013	147	147	338.8503	70.55430	152	277.0	356.0	385	526	0
5	2014	81	81	173.4321	111.22151	64	102.0	112.0	264	445	0

```
> boxplot(len~fyear,data=LChip_LMB,xlab="Year",ylab="Total Length (mm)")
```

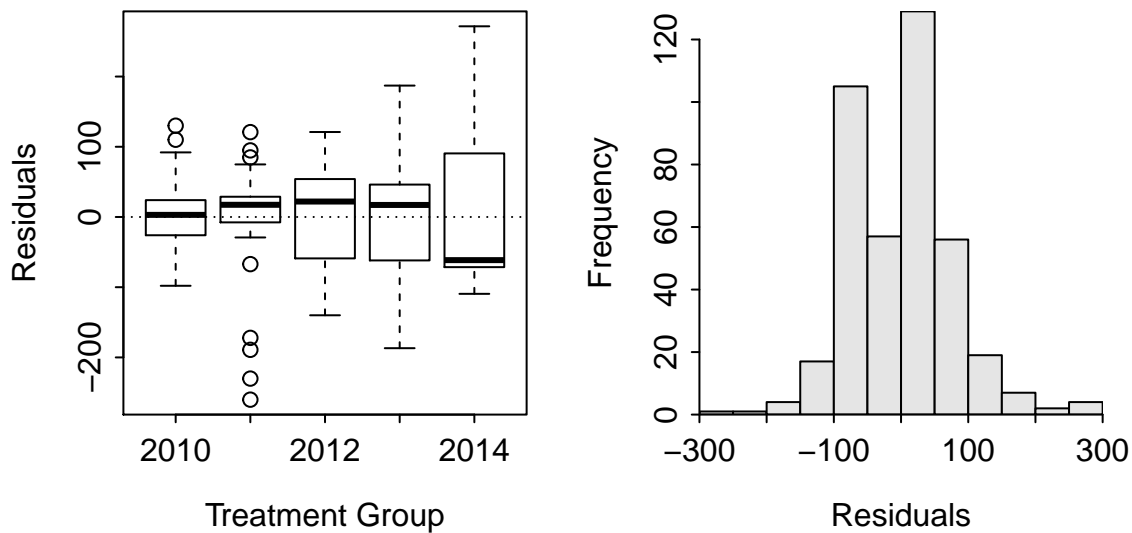


```
> hist(len~fyear,data=LChip_LMB,same.ylim=FALSE)
```



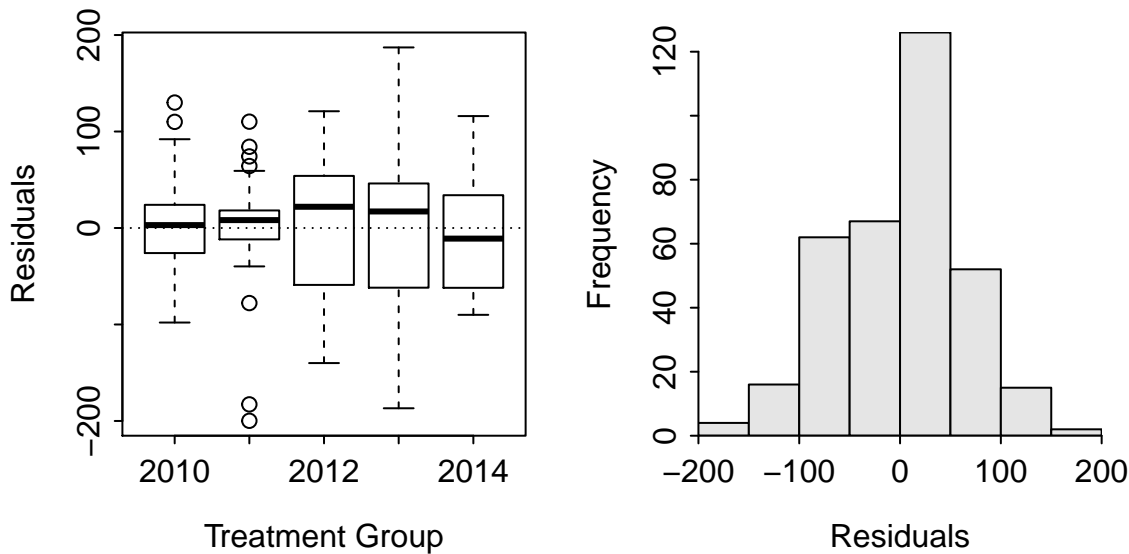
## 1-Way ANOVA

```
> aov1 <- lm(len~fyear,data=LChip_LMB)
> residPlot(aov1)
```



```
> leveneTest(aov1)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  4  4.8125 0.0008431
      397
```

```
> LChip_LMB2 <- filterD(LChip_LMB, len > 150)
> aov2 <- lm(len ~ fyear, data = LChip_LMB2)
> residPlot(aov2)
```



```
> leveneTest(aov2)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  4  6.1649 8.467e-05
      339
```

```
> anova(aov2)
Analysis of Variance Table

Response: len
      Df Sum Sq Mean Sq F value    Pr(>F)
fyear   4  75487 18871.8   4.6436 0.001161
Residuals 339 1377714  4064.1
```

```
> mc2 <- glht(aov2, mcp(fyear = "Tukey"))
> summary(mc2)
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

Fit: `lm(formula = len ~ fyear, data = LChip_LMB2)`

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t )
2011 - 2010 == 0	40.821	12.539	3.255	0.010253
2012 - 2010 == 0	-8.998	11.382	-0.791	0.930321

```

2013 - 2010 == 0    11.867      9.825    1.208 0.738535
2014 - 2010 == 0     2.017     15.213    0.133 0.999926
2012 - 2011 == 0   -49.819     12.207   -4.081 0.000469
2013 - 2011 == 0   -28.954     10.770   -2.688 0.055009
2014 - 2011 == 0   -38.804     15.840   -2.450 0.100257
2013 - 2012 == 0    20.865      9.397    2.220 0.167789
2014 - 2012 == 0    11.015     14.941    0.737 0.945268
2014 - 2013 == 0    -9.850     13.792   -0.714 0.951032
(Adjusted p values reported -- single-step method)

```

```

> cld(mc2)
2010 2011 2012 2013 2014
"a"  "b"  "a"  "ab" "ab"

```

```

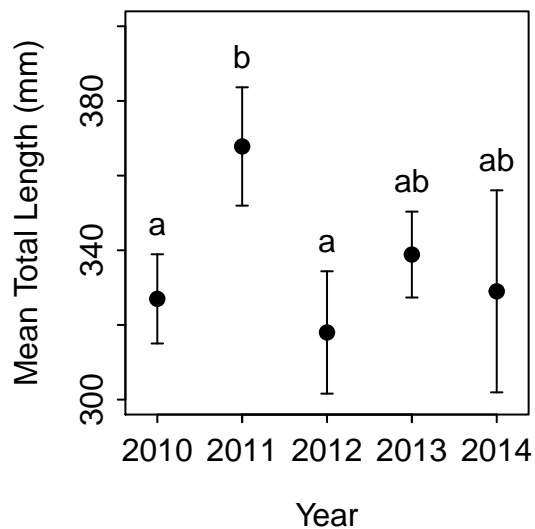
> sum_LMB2 <- Summarize(len=fyear,data=LChip_LMB2)
> sum_LMB2 <- select(sum_LMB2,fyear,n,mean,sd)
> sum_LMB2 <- mutate(sum_LMB2,year=fact2num(fyear),se=sd/sqrt(n),
                     LCI=mean-qt(0.975,df=n-1)*se,UCI=mean+qt(0.975,df=n-1)*se)
> sum_LMB2
  fyear   n   mean    sd year      se    LCI    UCI
1  2010  59 326.9831 45.84400 2010  5.968380 315.0360 338.9301
2  2011  46 367.8043 53.36650 2011  7.868461 351.9565 383.6522
3  2012  67 317.9851 67.17041 2012  8.206171 301.6009 334.3692
4  2013 147 338.8503 70.55430 2013  5.819220 327.3495 350.3511
5  2014  25 329.0000 65.54261 2014 13.108521 301.9453 356.0547

```

```

> with(sum_LMB2,plotCI(year,mean,li=LCI,ui=UCI,pch=19,ylim=c(300,400),xlim=c(2009.8,2014.2),
                      xlab="Year",ylab="Mean Total Length (mm)"))
> with(sum_LMB2,text(year,UCI,c("a","b","a","ab","ab"),pos=3))
> axis(1,c("2011","2013"))

```



## Kruskal-Wallis Test

```
> kruskal.test(len~fyear,data=LChip_LMB2)

Kruskal-Wallis rank sum test

data: len by fyear
Kruskal-Wallis chi-squared = 23.447, df = 4, p-value = 0.0001031
```

```
> dunnTest(len~fyear,data=LChip_LMB2)
Dunn (1964) Kruskal-Wallis multiple comparison
p-values adjusted with the Holm method.
```

	Comparison	Z	P.unadj	P.adj
1	2010 - 2011	-4.1681228	3.071185e-05	0.0003071185
2	2010 - 2012	-0.1623240	8.710507e-01	0.8710507178
3	2011 - 2012	4.1302625	3.623493e-05	0.0003261144
4	2010 - 2013	-2.1841883	2.894841e-02	0.1736904732
5	2011 - 2013	2.8602657	4.232862e-03	0.0296300373
6	2012 - 2013	-2.0870383	3.688467e-02	0.1844233645
7	2010 - 2014	-0.4310591	6.664254e-01	1.0000000000
8	2011 - 2014	2.8855142	3.907748e-03	0.0312619867
9	2012 - 2014	-0.3152718	7.525553e-01	1.0000000000
10	2013 - 2014	1.0804842	2.799266e-01	1.0000000000