**Hypotheses of the test:**

Null hypothesis: the means of the average weights of tusks from 20 seizure sites are the same for year 1970, 1990, and 2010.

Alternative hypothesis: at least one mean of the average weights of tusks is different from one of the others.

**Check the assumptions:**

Check normality assumption with qqplot and Shapiro test. None of the tests of demonstrate that the samples are significantly different from the null hypothesis of a normal distribution.

Check homogeneity of variance with following functions:

sd(Tusk.kg[Year==1970])/sd(Tusk.kg[Year==1990])

sd(Tusk.kg[Year==1970])/sd(Tusk.kg[Year==2010])

sd(Tusk.kg[Year==1990])/sd(Tusk.kg[Year==2010])

The result is 1.18~2.42, which does not match our assumption very well.

**Conduct one-way ANOVA:**

> mod1 <- aov(Tusk.kg ~ factor(Year))

> summary(mod1)

Df Sum Sq Mean Sq F value Pr(>F)

factor(Year) 2 1125.8 562.9 238.1 <2e-16 \*\*\*

Residuals 57 134.7 2.4

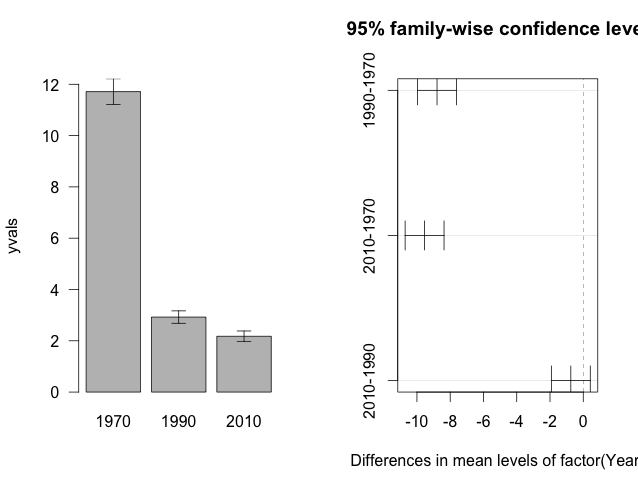
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

We have a large F-statistic and a very small p value. These tell us that we should reject the null hypothesis that the average tusk weights of three year have the same mean weight value in favor of the alternative hypothesis that there is a difference among the mean values of the average tusk weight.

**Post-hoc test:**

There is not significant difference between year 1990 and 2010. But there is significant difference between year 1970 and 1990, 1970 and 2010.



**Appendix.**

setwd("/Users/xuhong/Documents/Duke/Term 1 Courses/ENVIRON 710 Applied data analysis/Lab/Lab 7")

adat <- read.csv("TuskData.csv", header=T)

attach(adat)

# Test for normality

qqnorm(Tusk.kg[Year==1970])

qqline(Tusk.kg[Year==1970])

shapiro.test(Tusk.kg[Year==1970])

qqnorm(Tusk.kg[Year==1990])

qqline(Tusk.kg[Year==1990])

shapiro.test(Tusk.kg[Year==1990])

qqnorm(Tusk.kg[Year==2010])

qqline(Tusk.kg[Year==2010])

shapiro.test(Tusk.kg[Year==2010])

# Test for homogeneity of variance

sd(Tusk.kg[Year==1970])/sd(Tusk.kg[Year==1990])

sd(Tusk.kg[Year==1970])/sd(Tusk.kg[Year==2010])

sd(Tusk.kg[Year==1990])/sd(Tusk.kg[Year==2010])

# one way ANOVA

mod1 <- aov(Tusk.kg ~ factor(Year))

summary(mod1)

#To determine which of the means are significantly different from one another

#we conduct a post-hoc test.

TukeyHSD(mod1)

plot(TukeyHSD(mod1))

error.bars <- function(yvalues, se, nm){

xv <- barplot(yvalues, ylim=c(0, (max(yvalues)+max(se))),

names=nm, ylab=deparse(substitute(yvalues)), las=1)

g <- (max(xv)-min(xv))/50

for (i in 1:length(xv)){

arrows(xv[i], yvalues[i] + se[i], xv[i],

yvalues[i]-se[i], length=0.1, angle=90,

code=3)

}

}

year.mean <- tapply(Tusk.kg, list(Year), mean)

year.n <- tapply(Tusk.kg, list(Year), length)

year.sd <- tapply(Tusk.kg, list(Year), sd)

year.se <- site.sd/sqrt(site.n)

year.se[is.na(year.se)==T] <- 0

labls <- as.character(levels(factor(Year)))

yvals <- as.vector(year.mean)

error.bars(yvalues = yvals, se = year.se, nm = labls)