Lab 9 Solutions

Veerasak Kritsanapraphan

We’ll begin by loading some packages.

library(MASS)  
library(plyr)  
library(ggplot2)  
library(reshape)

##   
## Attaching package: 'reshape'

## The following objects are masked from 'package:plyr':  
##   
## rename, round\_any

library(e1071)

Let’s form our favourite birthwt data set.

# Rename the columns to have more descriptive names  
colnames(birthwt) <- c("birthwt.below.2500", "mother.age", "mother.weight",   
 "race", "mother.smokes", "previous.prem.labor", "hypertension", "uterine.irr",   
 "physician.visits", "birthwt.grams")  
  
# Transform variables to factors with descriptive levels  
birthwt <- transform(birthwt,   
 race = as.factor(mapvalues(race, c(1, 2, 3),   
 c("white","black", "other"))),  
 mother.smokes = as.factor(mapvalues(mother.smokes,   
 c(0,1), c("no", "yes"))),  
 hypertension = as.factor(mapvalues(hypertension,   
 c(0,1), c("no", "yes"))),  
 uterine.irr = as.factor(mapvalues(uterine.irr,   
 c(0,1), c("no", "yes")))  
 )

### ANOVA with birthwt data

**(a)** Create a new factor that categorizes the number of physician visits into three levels: 0, 1, 2, 3 or more.

pvb <- unique(birthwt$physician.visits)  
repvb <- as.character(pvb)  
for (i in 4:length(repvb)) {  
 repvb[i] = "3 or more"  
}  
birthwt <- transform(birthwt, phys.visit.binned1 = as.factor(mapvalues(physician.visits, pvb, repvb )))  
birthwt$phys.visit.binned1

## [1] 0 3 1 3 or more 0 0 1   
## [8] 1 1 0 0 1 0 3 or more  
## [15] 0 0 0 3 0 1 3 or more  
## [22] 3 1 0 3 or more 0 0 3 or more  
## [29] 0 1 1 1 1 1 0   
## [36] 3 or more 3 or more 0 3 or more 1 3 or more 3 or more  
## [43] 1 0 0 0 3 or more 0 3 or more  
## [50] 0 1 0 0 3 or more 0 0   
## [57] 0 0 0 0 0 3 or more 0   
## [64] 0 0 1 3 or more 3 or more 1 3 or more  
## [71] 0 3 or more 1 0 0 0 1   
## [78] 3 or more 0 0 1 0 0 0   
## [85] 0 0 0 0 0 1 0   
## [92] 3 or more 0 0 0 1 1 0   
## [99] 0 1 1 0 0 1 0   
## [106] 0 1 0 3 or more 3 or more 3 or more 1   
## [113] 3 or more 1 0 1 0 0 3 or more  
## [120] 1 1 0 1 0 3 or more 3 or more  
## [127] 1 0 1 1 0 3 or more 0   
## [134] 0 0 0 1 1 0 1   
## [141] 0 0 0 1 0 3 or more 3 or more  
## [148] 0 0 0 1 3 or more 0 0   
## [155] 0 0 3 1 0 0 0   
## [162] 1 0 0 0 0 3 or more 0   
## [169] 1 0 1 0 0 0 0   
## [176] 0 1 3 0 3 or more 1 3   
## [183] 0 0 3 or more 3 or more 0 0 3   
## Levels: 0 1 3 3 or more

phys.visit.binned <- birthwt$physician.visits  
phys.visit.binned[phys.visit.binned >= 3] <- "3.or.more"  
birthwt <- transform(birthwt, phys.visit.binned = as.factor(phys.visit.binned))  
birthwt$phys.visit.binned

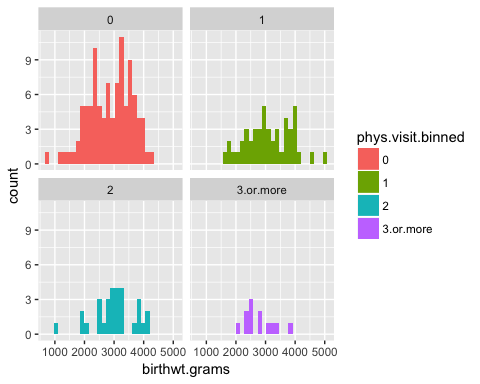
## [1] 0 3.or.more 1 2 0 0 1   
## [8] 1 1 0 0 1 0 2   
## [15] 0 0 0 3.or.more 0 1 2   
## [22] 3.or.more 1 0 2 0 0 2   
## [29] 0 1 1 1 1 1 0   
## [36] 2 2 0 2 1 2 2   
## [43] 1 0 0 0 3.or.more 0 2   
## [50] 0 1 0 0 2 0 0   
## [57] 0 0 0 0 0 2 0   
## [64] 0 0 1 2 3.or.more 1 2   
## [71] 0 2 1 0 0 0 1   
## [78] 3.or.more 0 0 1 0 0 0   
## [85] 0 0 0 0 0 1 0   
## [92] 2 0 0 0 1 1 0   
## [99] 0 1 1 0 0 1 0   
## [106] 0 1 0 2 3.or.more 2 1   
## [113] 2 1 0 1 0 0 2   
## [120] 1 1 0 1 0 2 2   
## [127] 1 0 1 1 0 2 0   
## [134] 0 0 0 1 1 0 1   
## [141] 0 0 0 1 0 2 2   
## [148] 0 0 0 1 2 0 0   
## [155] 0 0 3.or.more 1 0 0 0   
## [162] 1 0 0 0 0 3.or.more 0   
## [169] 1 0 1 0 0 0 0   
## [176] 0 1 3.or.more 0 2 1 3.or.more  
## [183] 0 0 2 2 0 0 3.or.more  
## Levels: 0 1 2 3.or.more

**Hint**: One way of doing this is with mapvalues, by mapping all instances of 3, 4,… etc, to “3 or more”.

Histogram of Birthweight

qplot(data=birthwt, x = birthwt.grams, facets = ~phys.visit.binned, geom = "histogram", fill = phys.visit.binned)

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## Check for Skewness

with(data=birthwt, skewness(birthwt.grams))

## [1] -0.205337

with(data=birthwt, skewness(birthwt.grams[phys.visit.binned=="0"]))

## [1] -0.3924435

with(data=birthwt, skewness(birthwt.grams[phys.visit.binned=="1"]))

## [1] 0.07103918

with(data=birthwt, skewness(birthwt.grams[phys.visit.binned=="2"]))

## [1] -0.5105783

with(data=birthwt, skewness(birthwt.grams[phys.visit.binned=="3.or.more"]))

## [1] 0.5874174

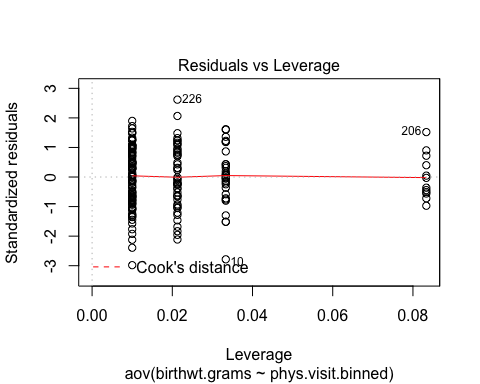
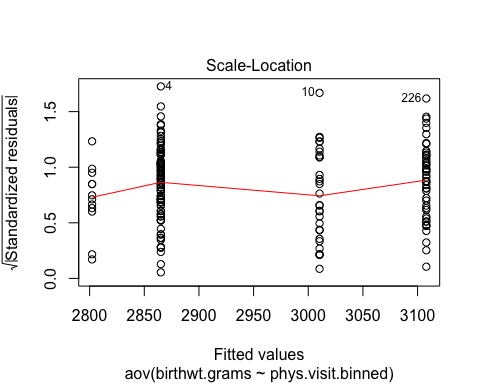
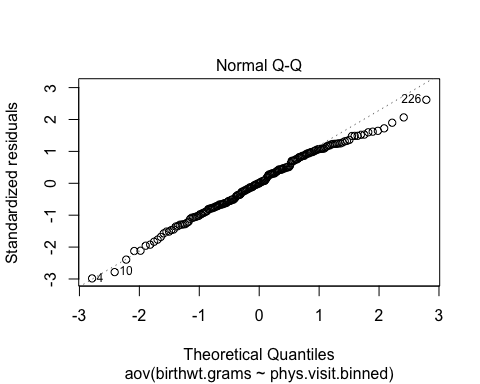
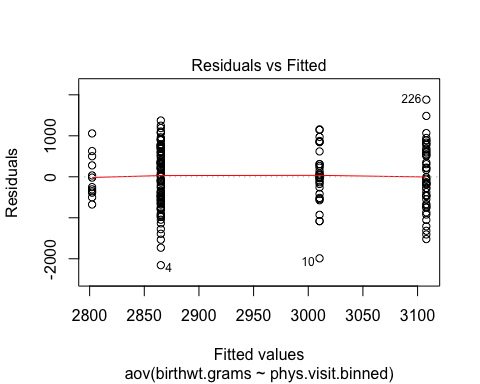
**(b)** Run an ANOVA to determine whether the average birth weight varies across number of physician visits.

aov.birthwt <- aov(birthwt.grams ~ phys.visit.binned, data = birthwt)  
summary(aov.birthwt)

## Df Sum Sq Mean Sq F value Pr(>F)  
## phys.visit.binned 3 2259057 753019 1.426 0.237  
## Residuals 185 97710599 528165

The p-value is greater than 0.05, so the variation in birthweight across number of physician visits is not statistically significant.

plot(aov.birthwt)



## Three-way ANOVA

twaov.birthwt <- aov(birthwt.grams ~ race+mother.smokes+phys.visit.binned, data = birthwt)  
summary(twaov.birthwt)

## Df Sum Sq Mean Sq F value Pr(>F)   
## race 2 5015725 2507863 5.227 0.006205 \*\*   
## mother.smokes 1 7322575 7322575 15.262 0.000132 \*\*\*  
## phys.visit.binned 3 311098 103699 0.216 0.885120   
## Residuals 182 87320257 479782   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1