5–Comparing\_groups\_tests.R

rstudio-user

2022-02-18

# R code snippets from slides for Chapman & Feit 2015  
# Slide file: Chapter6/Chapter6-ChapmanFeit  
  
# All code is (c) 2015, Springer. http://r-marketing.r-forge.r-project.org/  
  
# ==========  
  
  
# Load the data (same as Chapter 5)  
# ==========  
seg.df <- read.csv("http://goo.gl/qw303p")  
summary(seg.df)

## age gender income kids   
## Min. :19.26 Length:300 Min. : -5183 Min. :0.00   
## 1st Qu.:33.01 Class :character 1st Qu.: 39656 1st Qu.:0.00   
## Median :39.49 Mode :character Median : 52014 Median :1.00   
## Mean :41.20 Mean : 50937 Mean :1.27   
## 3rd Qu.:47.90 3rd Qu.: 61403 3rd Qu.:2.00   
## Max. :80.49 Max. :114278 Max. :7.00   
## ownHome subscribe Segment   
## Length:300 Length:300 Length:300   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##

# Chi-square test  
# ==========  
tmp.tab <- table(rep(c(1:4), times=c(25,25,25,20)))  
tmp.tab

##   
## 1 2 3 4   
## 25 25 25 20

chisq.test(tmp.tab)

##   
## Chi-squared test for given probabilities  
##   
## data: tmp.tab  
## X-squared = 0.78947, df = 3, p-value = 0.852

# chisq.test "significant" and "not significant"  
# ==========  
tmp.tab <- table(rep(c(1:4), times=c(25,25,25,20)))  
chisq.test(tmp.tab)

##   
## Chi-squared test for given probabilities  
##   
## data: tmp.tab  
## X-squared = 0.78947, df = 3, p-value = 0.852

tmp.tab <- table(rep(c(1:4), times=c(25,25,25,10)))  
tmp.tab

##   
## 1 2 3 4   
## 25 25 25 10

chisq.test(tmp.tab)

##   
## Chi-squared test for given probabilities  
##   
## data: tmp.tab  
## X-squared = 7.9412, df = 3, p-value = 0.04724

# chisq.test with segment data  
# ==========  
table(seg.df$Segment)

##   
## Moving up Suburb mix Travelers Urban hip   
## 70 100 80 50

chisq.test(table(seg.df$Segment))

##   
## Chi-squared test for given probabilities  
##   
## data: table(seg.df$Segment)  
## X-squared = 17.333, df = 3, p-value = 0.0006035

# chisq.test with segment data  
# ==========  
table(seg.df$subscribe, seg.df$ownHome)

##   
## ownNo ownYes  
## subNo 137 123  
## subYes 22 18

chisq.test(table(seg.df$subscribe, seg.df$ownHome))

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(seg.df$subscribe, seg.df$ownHome)  
## X-squared = 0.010422, df = 1, p-value = 0.9187

chisq.test(table(seg.df$subscribe, seg.df$ownHome), correct=FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: table(seg.df$subscribe, seg.df$ownHome)  
## X-squared = 0.074113, df = 1, p-value = 0.7854

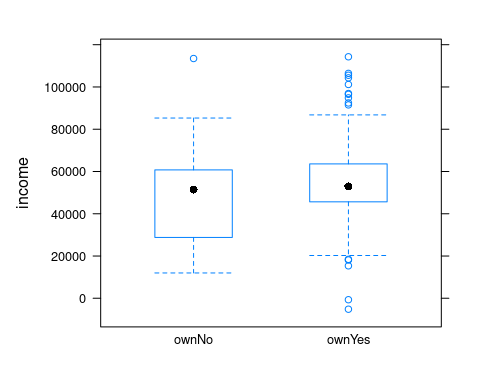
# Proportions: binomial test  
# ==========  
binom.test(12, 20, p=0.5)

##   
## Exact binomial test  
##   
## data: 12 and 20  
## number of successes = 12, number of trials = 20, p-value = 0.5034  
## alternative hypothesis: true probability of success is not equal to 0.5  
## 95 percent confidence interval:  
## 0.3605426 0.8088099  
## sample estimates:  
## probability of success   
## 0.6

# Proportions: binomial test continued  
# ==========  
# binom.test(12, 20, p=0.5)  
binom.test(120, 200, p=0.5)

##   
## Exact binomial test  
##   
## data: 120 and 200  
## number of successes = 120, number of trials = 200, p-value = 0.005685  
## alternative hypothesis: true probability of success is not equal to 0.5  
## 95 percent confidence interval:  
## 0.5285357 0.6684537  
## sample estimates:  
## probability of success   
## 0.6

# t-tests  
# ==========  
library(lattice)  
bwplot(income ~ ownHome, data=seg.df)



# t.test()  
# ==========  
t.test(income ~ ownHome, data=seg.df)

##   
## Welch Two Sample t-test  
##   
## data: income by ownHome  
## t = -3.2731, df = 285.25, p-value = 0.001195  
## alternative hypothesis: true difference in means between group ownNo and group ownYes is not equal to 0  
## 95 percent confidence interval:  
## -12080.155 -3007.193  
## sample estimates:  
## mean in group ownNo mean in group ownYes   
## 47391.01 54934.68

# t.test() for a subset() of data  
# ==========  
t.test(income ~ ownHome, data=subset(seg.df, Segment=="Travelers"))

##   
## Welch Two Sample t-test  
##   
## data: income by ownHome  
## t = 0.26561, df = 53.833, p-value = 0.7916  
## alternative hypothesis: true difference in means between group ownNo and group ownYes is not equal to 0  
## 95 percent confidence interval:  
## -8508.993 11107.604  
## sample estimates:  
## mean in group ownNo mean in group ownYes   
## 63188.42 61889.12

# ANOVA basics  
# ==========  
seg.aov.own <- aov(income ~ ownHome, data=seg.df)  
anova(seg.aov.own)

## Analysis of Variance Table  
##   
## Response: income  
## Df Sum Sq Mean Sq F value Pr(>F)   
## ownHome 1 4.2527e+09 4252661211 10.832 0.001118 \*\*  
## Residuals 298 1.1700e+11 392611030   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# ANOVA: Multiple groups  
# ==========  
aggregate(income ~ Segment, mean, data=seg.df)

## Segment income  
## 1 Moving up 53090.97  
## 2 Suburb mix 55033.82  
## 3 Travelers 62213.94  
## 4 Urban hip 21681.93

seg.aov.seg <- aov(income ~ Segment, data=seg.df)  
anova(seg.aov.seg)

## Analysis of Variance Table  
##   
## Response: income  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Segment 3 5.4970e+10 1.8323e+10 81.828 < 2.2e-16 \*\*\*  
## Residuals 296 6.6281e+10 2.2392e+08   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# ANOVA with Segment + Ownership  
# ==========  
anova(aov(income ~ Segment + ownHome, data=seg.df))

## Analysis of Variance Table  
##   
## Response: income  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Segment 3 5.4970e+10 1.8323e+10 81.6381 <2e-16 \*\*\*  
## ownHome 1 6.9918e+07 6.9918e+07 0.3115 0.5772   
## Residuals 295 6.6211e+10 2.2444e+08   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(aov(income ~ ownHome, data=seg.df))

## Analysis of Variance Table  
##   
## Response: income  
## Df Sum Sq Mean Sq F value Pr(>F)   
## ownHome 1 4.2527e+09 4252661211 10.832 0.001118 \*\*  
## Residuals 298 1.1700e+11 392611030   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# ANOVA with interaction  
# ==========  
anova(aov(income ~ Segment \* ownHome, data=seg.df))

## Analysis of Variance Table  
##   
## Response: income  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Segment 3 5.4970e+10 1.8323e+10 81.1305 <2e-16 \*\*\*  
## ownHome 1 6.9918e+07 6.9918e+07 0.3096 0.5784   
## Segment:ownHome 3 2.6329e+08 8.7762e+07 0.3886 0.7613   
## Residuals 292 6.5948e+10 2.2585e+08   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

anova(aov(income ~ Segment + ownHome + Segment:ownHome, data=seg.df))

## Analysis of Variance Table  
##   
## Response: income  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Segment 3 5.4970e+10 1.8323e+10 81.1305 <2e-16 \*\*\*  
## ownHome 1 6.9918e+07 6.9918e+07 0.3096 0.5784   
## Segment:ownHome 3 2.6329e+08 8.7762e+07 0.3886 0.7613   
## Residuals 292 6.5948e+10 2.2585e+08   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Model Comparison  
# ==========  
anova(aov(income ~ Segment, data=seg.df),  
 aov(income ~ Segment + ownHome, data=seg.df))

## Analysis of Variance Table  
##   
## Model 1: income ~ Segment  
## Model 2: income ~ Segment + ownHome  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 296 6.6281e+10   
## 2 295 6.6211e+10 1 69918004 0.3115 0.5772

# Visualization: ANOVA Group means  
# ==========  
# install.packages("multcomp") # if needed  
library(multcomp)

## Loading required package: mvtnorm

## Loading required package: survival

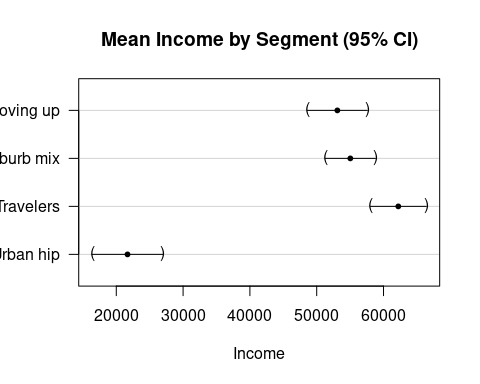
## Loading required package: TH.data

## Loading required package: MASS

##   
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':  
##   
## geyser

seg.aov <- aov(income ~ -1 + Segment, data=seg.df) # model w/o int.  
by.seg <- glht(seg.aov) # means and CIs  
plot(by.seg, xlab="Income", main="Mean Income by Segment (95% CI)")



# Exercises (Basic)  
# ==========  
library(car) # install.packages("car") if needed

## Loading required package: carData

data(Salaries)  
  
  
# Answers (1)  
# ==========  
with(Salaries, prop.table(table(discipline, sex), margin=1))

## sex  
## discipline Female Male  
## A 0.09944751 0.90055249  
## B 0.09722222 0.90277778

with(Salaries, chisq.test(table(discipline, sex)))

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table(discipline, sex)  
## X-squared = 2.0875e-29, df = 1, p-value = 1

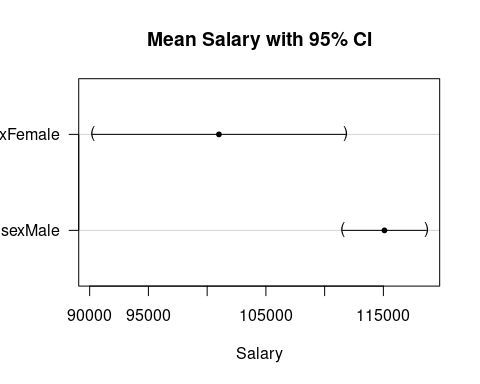
# Answers (2)  
# ==========  
aggregate(salary ~ sex, data=Salaries, mean)

## sex salary  
## 1 Female 101002.4  
## 2 Male 115090.4

anova(aov(salary ~ sex, data=Salaries))

## Analysis of Variance Table  
##   
## Response: salary  
## Df Sum Sq Mean Sq F value Pr(>F)   
## sex 1 6.9800e+09 6980014930 7.7377 0.005667 \*\*  
## Residuals 395 3.5632e+11 902077538   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Answers (3)  
# ==========  
# install.packages("multcomp") # if needed  
library(multcomp)  
salary.aov <- aov(salary ~ -1 + sex, data=Salaries)  
by.sex <- glht(salary.aov)   
plot(by.sex, xlab="Salary", main="Mean Salary with 95% CI")



# Optional: Stepwise ANOVA  
# ==========  
seg.aov.step <- step(aov(income ~ ., data=seg.df))

## Start: AIC=5779.17  
## income ~ age + gender + kids + ownHome + subscribe + Segment  
##   
## Df Sum of Sq RSS AIC  
## - age 1 4.7669e+06 6.5661e+10 5777.2  
## - ownHome 1 1.0337e+08 6.5759e+10 5777.6  
## - kids 1 1.3408e+08 6.5790e+10 5777.8  
## - subscribe 1 1.5970e+08 6.5816e+10 5777.9  
## - gender 1 2.6894e+08 6.5925e+10 5778.4  
## <none> 6.5656e+10 5779.2  
## - Segment 3 1.9303e+10 8.4959e+10 5850.5  
##   
## Step: AIC=5777.19  
## income ~ gender + kids + ownHome + subscribe + Segment  
##   
## Df Sum of Sq RSS AIC  
## - ownHome 1 1.0159e+08 6.5762e+10 5775.7  
## - kids 1 1.3205e+08 6.5793e+10 5775.8  
## - subscribe 1 1.5794e+08 6.5819e+10 5775.9  
## - gender 1 2.7009e+08 6.5931e+10 5776.4  
## <none> 6.5661e+10 5777.2  
## - Segment 3 4.9044e+10 1.1470e+11 5938.6  
##   
## Step: AIC=5775.66  
## income ~ gender + kids + subscribe + Segment  
##   
## Df Sum of Sq RSS AIC  
## - kids 1 1.0707e+08 6.5869e+10 5774.1  
## - subscribe 1 1.6370e+08 6.5926e+10 5774.4  
## - gender 1 2.5520e+08 6.6017e+10 5774.8  
## <none> 6.5762e+10 5775.7  
## - Segment 3 5.2897e+10 1.1866e+11 5946.7  
##   
## Step: AIC=5774.15  
## income ~ gender + subscribe + Segment  
##   
## Df Sum of Sq RSS AIC  
## - subscribe 1 1.6226e+08 6.6032e+10 5772.9  
## - gender 1 2.4390e+08 6.6113e+10 5773.3  
## <none> 6.5869e+10 5774.1  
## - Segment 3 5.3005e+10 1.1887e+11 5945.3  
##   
## Step: AIC=5772.88  
## income ~ gender + Segment  
##   
## Df Sum of Sq RSS AIC  
## - gender 1 2.4949e+08 6.6281e+10 5772.0  
## <none> 6.6032e+10 5772.9  
## - Segment 3 5.4001e+10 1.2003e+11 5946.2  
##   
## Step: AIC=5772.02  
## income ~ Segment  
##   
## Df Sum of Sq RSS AIC  
## <none> 6.6281e+10 5772.0  
## - Segment 3 5.497e+10 1.2125e+11 5947.2

# Stepwise ANOVA: Result  
# ==========  
seg.aov.step <- step(aov(income ~ ., data=seg.df))

## Start: AIC=5779.17  
## income ~ age + gender + kids + ownHome + subscribe + Segment  
##   
## Df Sum of Sq RSS AIC  
## - age 1 4.7669e+06 6.5661e+10 5777.2  
## - ownHome 1 1.0337e+08 6.5759e+10 5777.6  
## - kids 1 1.3408e+08 6.5790e+10 5777.8  
## - subscribe 1 1.5970e+08 6.5816e+10 5777.9  
## - gender 1 2.6894e+08 6.5925e+10 5778.4  
## <none> 6.5656e+10 5779.2  
## - Segment 3 1.9303e+10 8.4959e+10 5850.5  
##   
## Step: AIC=5777.19  
## income ~ gender + kids + ownHome + subscribe + Segment  
##   
## Df Sum of Sq RSS AIC  
## - ownHome 1 1.0159e+08 6.5762e+10 5775.7  
## - kids 1 1.3205e+08 6.5793e+10 5775.8  
## - subscribe 1 1.5794e+08 6.5819e+10 5775.9  
## - gender 1 2.7009e+08 6.5931e+10 5776.4  
## <none> 6.5661e+10 5777.2  
## - Segment 3 4.9044e+10 1.1470e+11 5938.6  
##   
## Step: AIC=5775.66  
## income ~ gender + kids + subscribe + Segment  
##   
## Df Sum of Sq RSS AIC  
## - kids 1 1.0707e+08 6.5869e+10 5774.1  
## - subscribe 1 1.6370e+08 6.5926e+10 5774.4  
## - gender 1 2.5520e+08 6.6017e+10 5774.8  
## <none> 6.5762e+10 5775.7  
## - Segment 3 5.2897e+10 1.1866e+11 5946.7  
##   
## Step: AIC=5774.15  
## income ~ gender + subscribe + Segment  
##   
## Df Sum of Sq RSS AIC  
## - subscribe 1 1.6226e+08 6.6032e+10 5772.9  
## - gender 1 2.4390e+08 6.6113e+10 5773.3  
## <none> 6.5869e+10 5774.1  
## - Segment 3 5.3005e+10 1.1887e+11 5945.3  
##   
## Step: AIC=5772.88  
## income ~ gender + Segment  
##   
## Df Sum of Sq RSS AIC  
## - gender 1 2.4949e+08 6.6281e+10 5772.0  
## <none> 6.6032e+10 5772.9  
## - Segment 3 5.4001e+10 1.2003e+11 5946.2  
##   
## Step: AIC=5772.02  
## income ~ Segment  
##   
## Df Sum of Sq RSS AIC  
## <none> 6.6281e+10 5772.0  
## - Segment 3 5.497e+10 1.2125e+11 5947.2

anova(seg.aov.step)

## Analysis of Variance Table  
##   
## Response: income  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Segment 3 5.4970e+10 1.8323e+10 81.828 < 2.2e-16 \*\*\*  
## Residuals 296 6.6281e+10 2.2392e+08   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# More Advanced: Bayesian ANOVA  
# ==========  
# install.packages("BayesFactor") # if needed  
library(BayesFactor)

## Loading required package: coda

## Loading required package: Matrix

## \*\*\*\*\*\*\*\*\*\*\*\*  
## Welcome to BayesFactor 0.9.12-4.3. If you have questions, please contact Richard Morey (richarddmorey@gmail.com).  
##   
## Type BFManual() to open the manual.  
## \*\*\*\*\*\*\*\*\*\*\*\*

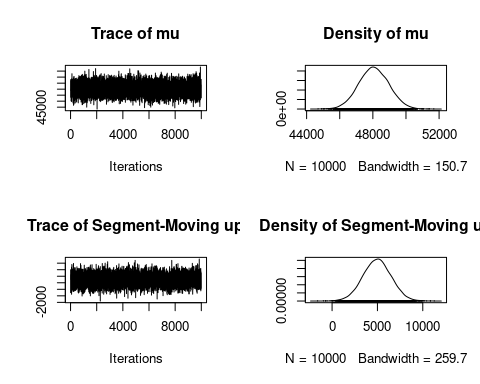
set.seed(96761) # optional, for replication  
seg.bf1 <- lmBF(income ~ Segment, data=seg.df)  
seg.bf2 <- lmBF(income ~ Segment + ownHome, data=seg.df)  
  
seg.bf1 / seg.bf2

## Bayes factor analysis  
## --------------  
## [1] Segment : 6.579729 ±1.62%  
##   
## Against denominator:  
## income ~ Segment + ownHome   
## ---  
## Bayes factor type: BFlinearModel, JZS

# Bayesian ANOVA: Under the hood  
# ==========  
seg.bf.chain <- posterior(seg.bf1, 1, iterations = 10000)  
head(seg.bf.chain[, 1:4])

## Markov Chain Monte Carlo (MCMC) output:  
## Start = 1   
## End = 7   
## Thinning interval = 1   
## mu Segment-Moving up Segment-Suburb mix Segment-Travelers  
## [1,] 48055.75 4964.3516 6909.129 13983.40  
## [2,] 47706.52 6524.7816 7909.242 12359.19  
## [3,] 48361.79 5608.6227 7553.687 14379.18  
## [4,] 49396.99 5468.3964 7568.978 12856.49  
## [5,] 48176.00 5779.9796 5168.127 13859.16  
## [6,] 49446.75 967.5255 8146.482 17934.65  
## [7,] 46642.84 3881.8313 7745.062 15994.53

# Bayesian ANOVA: Plotting the Draws  
# ==========  
plot(seg.bf.chain[, 1:2]) # overall mean + first segment



# Bayesian ANOVA: Segment Estimates  
# ==========  
seg.bf.chain[1:4, 1:4]

## mu Segment-Moving up Segment-Suburb mix Segment-Travelers  
## [1,] 48055.75 4964.352 6909.129 13983.40  
## [2,] 47706.52 6524.782 7909.242 12359.19  
## [3,] 48361.79 5608.623 7553.687 14379.18  
## [4,] 49396.99 5468.396 7568.978 12856.49

seg.bf.chain[1:4, 2:4] + seg.bf.chain[1:4, 1]

## Segment-Moving up Segment-Suburb mix Segment-Travelers  
## [1,] 53020.10 54964.87 62039.15  
## [2,] 54231.30 55615.76 60065.71  
## [3,] 53970.41 55915.48 62740.98  
## [4,] 54865.39 56965.97 62253.49

# Bayesian ANOVA: Segment CIs  
# ==========  
seg.bf.chain.total <- seg.bf.chain[, 2:5] + seg.bf.chain[, 1]  
seg.bf.chain.total[1:4, 1:3]

## Segment-Moving up Segment-Suburb mix Segment-Travelers  
## [1,] 53020.10 54964.87 62039.15  
## [2,] 54231.30 55615.76 60065.71  
## [3,] 53970.41 55915.48 62740.98  
## [4,] 54865.39 56965.97 62253.49

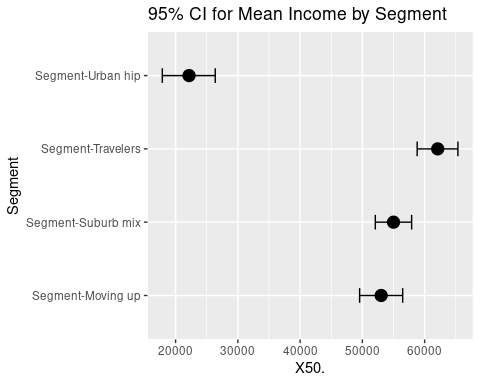
seg.bf.ci <- t(apply(seg.bf.chain.total, 2,   
 quantile, pr=c(0.025, 0.5, 0.975)))  
seg.bf.ci

## 2.5% 50% 97.5%  
## Segment-Moving up 49558.48 53017.70 56472.80  
## Segment-Suburb mix 52077.24 55001.33 57911.31  
## Segment-Travelers 58798.46 62101.16 65345.16  
## Segment-Urban hip 17868.25 22161.13 26370.48

# Bayesian ANOVA: Plot the CIs  
# ==========  
seg.bf.df <- data.frame(seg.bf.ci)  
seg.bf.df$Segment <- rownames(seg.bf.df)  
  
library(ggplot2)

## Warning in register(): Can't find generic `scale\_type` in package ggplot2 to  
## register S3 method.

# basic plot object with CIs on Y axis by Segment on X  
p <- ggplot(seg.bf.df, aes(x=Segment,   
 y=X50., ymax=X97.5., ymin=X2.5.))  
  
# add points for the Y var and error bars for ymax, ymin  
p <- p + geom\_point(size=4) + geom\_errorbar(width=0.2)  
  
# add a title and rotate the plot to horizontal  
p <- p +   
 ggtitle("95% CI for Mean Income by Segment") + coord\_flip()  
  
  
# Plot it  
# ==========  
p



# Exercises (Advanced)  
# ==========  
library(car) # install.packages("car") if needed  
data(Salaries)  
  
  
# Answers (Advanced, 1)  
# ==========  
salary.step <- step(aov(salary ~ ., data=Salaries)) # output hidden

## Start: AIC=7965.19  
## salary ~ rank + discipline + yrs.since.phd + yrs.service + sex  
##   
## Df Sum of Sq RSS AIC  
## - sex 1 7.8068e+08 1.9890e+11 7964.8  
## <none> 1.9812e+11 7965.2  
## - yrs.since.phd 1 2.5041e+09 2.0062e+11 7968.2  
## - yrs.service 1 2.7100e+09 2.0083e+11 7968.6  
## - discipline 1 1.9237e+10 2.1735e+11 8000.0  
## - rank 2 6.9508e+10 2.6762e+11 8080.6  
##   
## Step: AIC=7964.75  
## salary ~ rank + discipline + yrs.since.phd + yrs.service  
##   
## Df Sum of Sq RSS AIC  
## <none> 1.9890e+11 7964.8  
## - yrs.since.phd 1 2.5001e+09 2.0140e+11 7967.7  
## - yrs.service 1 2.5763e+09 2.0147e+11 7967.9  
## - discipline 1 1.9489e+10 2.1839e+11 7999.9  
## - rank 2 7.0679e+10 2.6958e+11 8081.5

anova(salary.step)

## Analysis of Variance Table  
##   
## Response: salary  
## Df Sum Sq Mean Sq F value Pr(>F)   
## rank 2 1.4323e+11 7.1616e+10 140.7855 < 2.2e-16 \*\*\*  
## discipline 1 1.8430e+10 1.8430e+10 36.2303 4.039e-09 \*\*\*  
## yrs.since.phd 1 1.6565e+08 1.6565e+08 0.3256 0.56857   
## yrs.service 1 2.5763e+09 2.5763e+09 5.0646 0.02497 \*   
## Residuals 391 1.9890e+11 5.0869e+08   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Answers (Advanced, 2)  
# ==========  
library(BayesFactor)  
set.seed(96761) # optional for replication  
  
salary.b <- lmBF(salary ~ rank + discipline + yrs.service,   
 data=Salaries)  
salary.mc <- posterior(salary.b, 1, iterations=10000)  
  
t(apply(salary.mc[, 1:7], 2, quantile, pr=c(0.025, 0.5, 0.975)))

## 2.5% 50% 97.5%  
## mu 96140.5552 99183.0465 102160.5734  
## rank-AsstProf -25606.7816 -20875.8574 -16296.1832  
## rank-AssocProf -10971.6824 -6811.8484 -2680.7953  
## rank-Prof 23840.8543 27721.1831 31590.4752  
## discipline-A -8918.5337 -6635.0952 -4386.1522  
## discipline-B 4386.1522 6635.0952 8918.5337  
## yrs.service-yrs.service -273.9851 -60.5864 147.7961

# Answers (Advanced, 3)  
# ==========  
salary.b2 <- lmBF(salary ~ rank + discipline + yrs.service + sex,   
 data=Salaries)  
  
salary.b2 / salary.b

## Bayes factor analysis  
## --------------  
## [1] rank + discipline + yrs.service + sex : 0.3426499 ±8.08%  
##   
## Against denominator:  
## salary ~ rank + discipline + yrs.service   
## ---  
## Bayes factor type: BFlinearModel, JZS

# Answers (Advanced, 4)  
# ==========  
aov1 <- aov(salary ~ rank + discipline + yrs.service,   
 data=Salaries)  
aov2 <- aov(salary ~ rank + discipline + yrs.service + sex,   
 data=Salaries)  
anova(aov1, aov2)

## Analysis of Variance Table  
##   
## Model 1: salary ~ rank + discipline + yrs.service  
## Model 2: salary ~ rank + discipline + yrs.service + sex  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 392 2.0140e+11   
## 2 391 2.0062e+11 1 776686259 1.5137 0.2193

# Answers (Advanced, 5)  
# ==========  
salary.cidf <- data.frame(t(apply(salary.mc[, 2:4] + salary.mc[ , 1], 2,   
 quantile, pr=c(0.025, 0.5, 0.975))))  
salary.cidf$rank <- rownames(salary.cidf)  
library(ggplot2)  
p <- ggplot(salary.cidf, aes(x=rank,   
 y=X50., ymax=X97.5., ymin=X2.5.))  
p <- p + geom\_point(size=4) + geom\_errorbar(width=0.2)  
p + ggtitle("95% Credible Intervals for Mean Salary by Rank") +   
 coord\_flip()

