CHAPTER 9 ANALYSIS EXAMPLES REPLICATION-R SURVEY PACKAGE 3.22

GENERAL NOTES ABOUT ANALYSIS EXAMPLES REPLICATION

These examples are intended to provide guidance on how to use the commands/procedures for analysis of complex sample survey data and assume all data management and other preliminary work is done. The relevant syntax for the procedure of interest is shown first along with the associated output for that procedure(s). In some examples, there may be more than one block of syntax and in this case all syntax is first presented followed by the output produced.

In some software packages certain procedures or options are not available but we have made every attempt to demonstrate how to match the output produced by Stata 10+ in the textbook. Check the ASDA website for updates to the various software tools we cover.

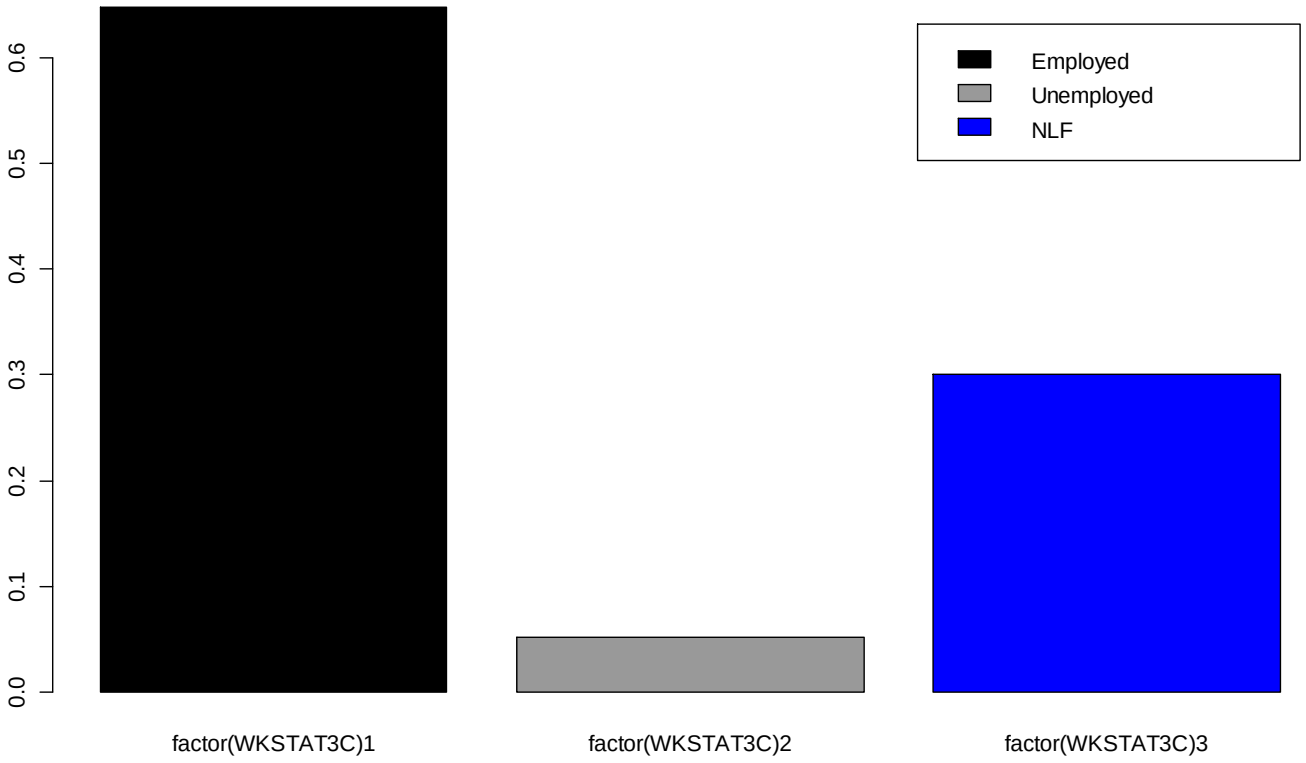
NOTES ABOUT CHAPTER 9 ANALYSES IN R SURVEY PACKAGE 3.22 (WITH R 2.7)

The R survey package used in these examples is 3.22 and was run under R 2.7 on a PC.

The R survey package offers a good range of svy commands for the generalized linear models of this chapter: use of the barplot (with svymean output) and svyhist commands provide weighted plots and histograms of survey data. For models, svyolr is used for ordinal regression and svyglm with a log link for Poisson type models. Other features demonstrated include use of a factor statement for categorical variables, how to obtain odds ratios from an object from a modeling command and testing of predictors for models using the svychisq command. Multinomial logit and negative binomial models are not included as part of the survey package and are therefore omitted in this chapter's output. See the recent Lumley book mentioned on the ASDA website for information about how to extend the R Survey package for multinomial logit regression.

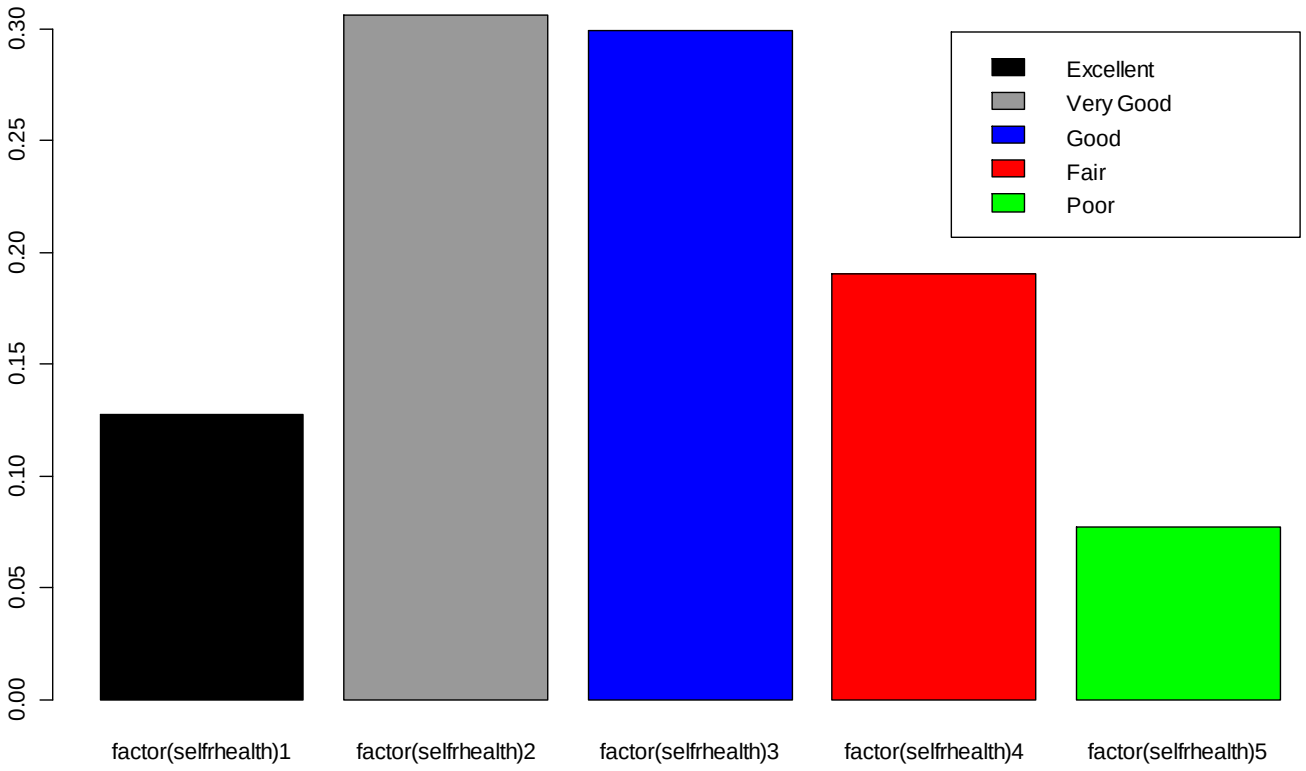
```
#Data production and set up of design objects
#remember to load package first survey package

#NHANES
nhanesdata <- read.table(file = "f:/applied_analysis_book/r/nhanes_final.txt", sep = "\t", header = T, as.is=T)

#create factor variables
nhanesdata$racec <- factor(nhanesdata$RIDRETH1, levels = 1: 5 , labels =c("Mexican", "Other Hispanic", "White",
"Black", "Other"))
nhanesdata$marcatc <- factor(nhanesdata$marcat, levels = 1: 3, labels =c("Married", "Previously Married", "Never
Married"))
nhanesdata$edcatc <- factor(nhanesdata$edcat, levels = 1: 4, labels =c("0-11", "12", "13-15","16+"))
nhanesdata$bp_catc <- factor(nhanesdata$bp_cat, levels = 1: 4, labels =c("Normal", "Pre-HBP", "Stage 1
HBP","Stage 2 HBP"))
nhanesdata$agesq <- (nhanesdata$agecent * nhanesdata$agecent )
names(nhanesdata)

nhanessvy2 <- svydesign(strata=~SDMVSTRA, id=~SDMVPSU, weights=~WTMEC2YR, data=nhanesdata, nest=T)
subnhanes <- subset(nhanessvy2 , RIDAGEYR >= 18)

#NCS-R
ncsr <- read.table(file = "f:/applied_analysis_book/r/ncsr2010.txt", sep = "\t", header = T, as.is=T)
names(ncsr)

#create factor versions with labels
ncsr$racec <- factor(ncsr$racecat, levels = 1: 4, labels =c("Other", "Hispanic", "Black", "White"))
ncsr$marcatc <- factor(ncsr$MAR3CAT, levels = 1: 3, labels =c("Married", "Previously Married", "Never Married"))
ncsr$edcatc <- factor(ncsr$ED4CAT, levels = 1: 4, labels =c("0-11", "12", "13-15","16+"))
ncsr$sexc <- factor(ncsr$SEX, levels = 1:2, labels=c("Male","Female"))
ncsr$agcatc <- factor(ncsr$ag4cat, levels = 1:4, labels=c("18-29", "30-44", "45-59", "60+"))

ncsrsvyp1 <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~NCSRWTSH, data=ncsr, nest=T)
ncsrsvyp2 <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~NCSRWTLG, data=ncsr, nest=T)
ncsrsvypop <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~popweight, data=ncsr, nest=T)

#HRS
#both hh and r weights are needed plus financial respondent for hh level analysis
hrs <- read.table(file = "f:/applied_analysis_book/r/hrs2010.txt", sep = "\t", header = T, as.is=T)
hrssvyhh <- svydesign(strata=~STRATUM, id=~SECU, weights=~KWGTHH , data=hrs, nest=T)
summary(hrssvyhh)
hrssvysub <-subset(hrssvyhh, KFINR==1)

hrssvyr <- svydesign(strata=~STRATUM, id=~SECU, weights=~KWGTR , data=hrs, nest=T)
summary(hrssvyr)
```

```
#FIGURE 9.1 BAR CHART OF WORK STATUS NCS-R DATA
fig91 <- svymean( ~factor(WKSTAT3C), ncsrsvyp2, na.rm=T)

 barplot(fig91, legend=c("Employed", "Unemployed", "NLF") , col=c("black", "grey60", "blue"))
```



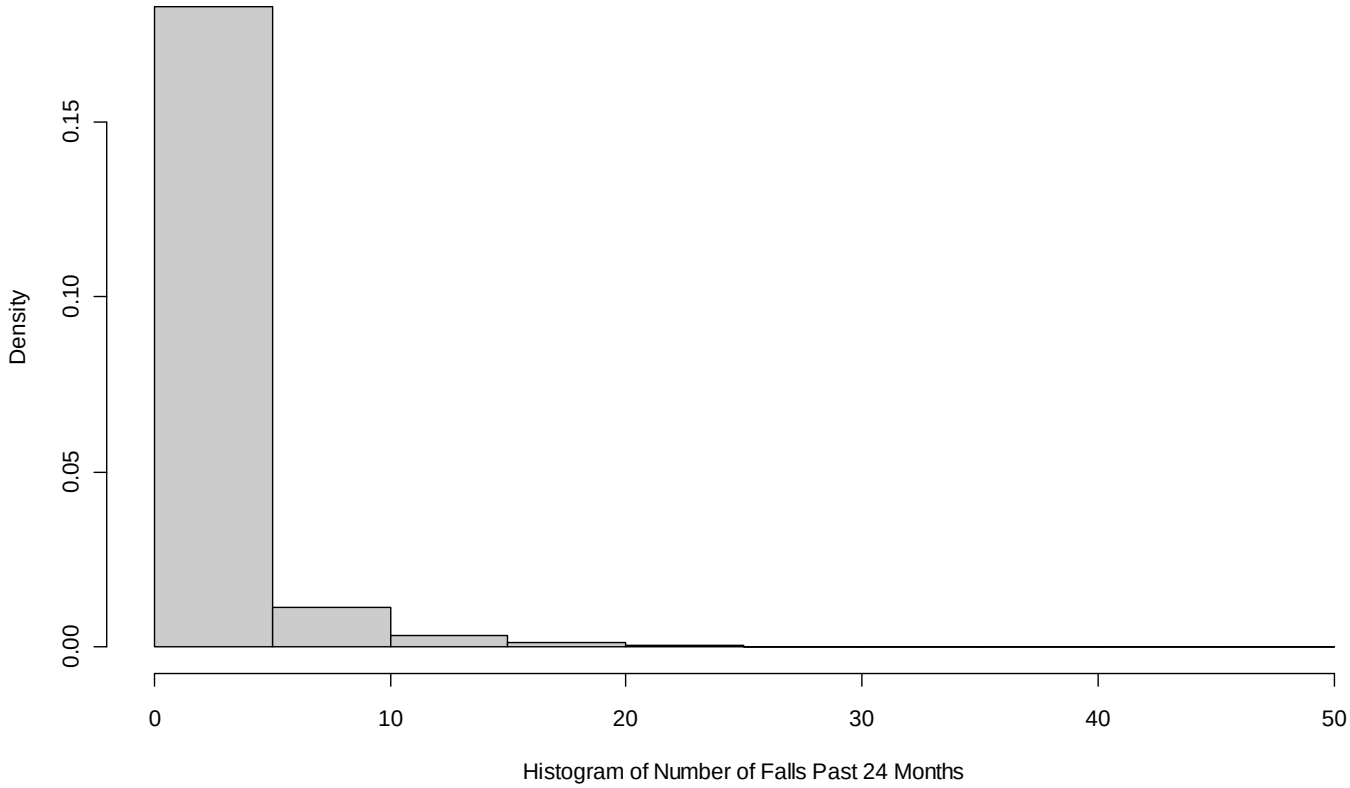
```
#BIVARIATE TESTING PRIOR TO MULTINOMIAL LOGIT
>svychisq(~WKSTAT3C+SEX, ncsrsvyp2, statistic="F")

        Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~WKSTAT3C + SEX, ncsrsvyp2, statistic = "F")
F = 27.3292, ndf = 1.875, ddf = 78.748, p-value = 2.171e-09


> svychisq(~WKSTAT3C+ald, ncsrsvyp2, statistic="F")

        Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~WKSTAT3C + ald, ncsrsvyp2, statistic = "F")
F = 3.1249, ndf = 1.725, ddf = 72.441, p-value = 0.05716



> svychisq(~WKSTAT3C+mde, ncsrsvyp2, statistic="F")

        Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~WKSTAT3C + mde, ncsrsvyp2, statistic = "F")
F = 4.6693, ndf = 1.735, ddf = 72.861, p-value = 0.01605



> svychisq(~WKSTAT3C+ED4CAT, ncsrsvyp2, statistic="F")

        Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~WKSTAT3C + ED4CAT, ncsrsvyp2, statistic = "F")
F = 27.6404, ndf = 5.146, ddf = 216.118, p-value < 2.2e-16



> svychisq(~WKSTAT3C+MAR3CAT, ncsrsvyp2, statistic="F")

        Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~WKSTAT3C + MAR3CAT, ncsrsvyp2, statistic = "F")
F = 23.1237, ndf = 3.198, ddf = 134.337, p-value = 1.229e-12


> svychisq(~WKSTAT3C+ag4cat, ncsrsvyp2, statistic="F")

        Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~WKSTAT3C + ag4cat, ncsrsvyp2, statistic = "F")
F = 113.4945, ndf = 4.965, ddf = 208.513, p-value < 2.2e-16


# MULTINOMIAL LOGISTIC IS NOT AVAILABLE IN SURVEY PACKAGE OF R, SEE THE EXTENSION OF THE SURVEY R PACKAGE IN THE
APPENDIX OF LUMLEY'S BOOK ABOUT HOW TO POSSIBLY EXTEND THE PACKAGE
```

```
#FIGURE 9.2 BAR CHART OF SELF RATED HEALTH HRS DATA
> fig92 <- svymean( ~factor(selfrhealth),hrssvyr, na.rm=T)
> barplot(fig92, legend=c("Excellent", "Very Good", "Good", "Fair", "Poor") , col=c("black", "grey60", "blue",
"red", "green"))
```



```
#ORDINAL LOGISTIC REGRESSION USING HRS DATA
> summary(ex92_ordinal <- svyolr (factor(selfrhealth) ~ male + KAGE, design=hrssvyr))
Call:
svyolr(factor(selfrhealth) ~ male + KAGE, design = hrssvyr)

Coefficients:
           Value  Std. Error  t value
male -0.07067736 0.032327977 -2.18626
KAGE  0.02881077 0.002178849 13.22293

Intercepts:
    Value   Std. Error t value
1|2 -0.0709  0.1531    -0.4632
2|3  1.6142  0.1529    10.5578
3|4  2.9167  0.1588    18.3634
4|5  4.4053  0.1653    26.6432
(25 observations deleted due to missingness)

#Odds Ratios from Ordinal Logistic
> exp(ex92_ordinal$coef)
     male      KAGE
0.9317625 1.0292298
```

```
#HISTOGRAM OF NUMBER OF FALLS DURING PAST 24 MONTHS HRS DATA
svyhist(~numfalls24 , subset (hrssvyr, KAGE >=65), main="", col="grey80", xlab ="Histogram of Number of Falls
Past 24 Months")
```



```
#EXAMPLE 9.3 POISSON MODEL NUMBER OF FALLS DURING PAST 24 MONTHS HRS DATA

> ex93_poisson <- svyglm(numfalls24 ~ male + factor(age3cat) + arthritis + DIABETES + bodywgt + totheight,
design=hrssvyr, family=quasipoisson(log))
Warning messages:
1: In summary.glm(g) :
  observations with zero weight not used for calculating dispersion
2: In summary.glm(glm.object) :
  observations with zero weight not used for calculating dispersion
> summary(ex93_poisson)

Call:
svyglm(numfalls24 ~ male + factor(age3cat) + arthritis + DIABETES +
    bodywgt + totheight, design = hrssvyr, family = quasipoisson(log))

Survey design:
svydesign(strata = ~STRATUM, id = ~SECU, weights = ~KWGTR, data = hrs,
    nest = T)


#NOTE: CODES FOR AGE3CAT 1=65-74 2=75-84 3=85+ YEARS OF AGE
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.4938245  0.6359374   0.777 0.441499
male              0.1831258  0.1073485   1.706 0.094921 .
factor(age3cat)2  0.2383983  0.0534633   4.459 5.43e-05 ***
factor(age3cat)3  0.5838654  0.0899710   6.489 5.84e-08 ***
arthritis         0.4867153  0.0824179   5.905 4.31e-07 ***
DIABETES          0.2596115  0.0689276   3.766 0.000478 ***
bodywgt           0.0009237  0.0008851   1.044 0.302200
totheight        -0.0224337  0.0110317  -2.034 0.047917 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be 3.052147)


Number of Fisher Scoring iterations: 6

> exp(ex93_poisson$coef)
 (Intercept)       male       factor(age3cat)2  factor(age3cat)3    arthritis          DIABETES          bodywgt
  1.638571         1.200965   1.269215          1.792955           1.626963         1.296426         1.000924

totheight
0.977816
```

```
# NEGATIVE BINOMIAL (NOT AVAILABLE WITH SURVEY CORRECTION, DISPERSION IS ACCOUNTED FOR IN SVYGLM, PER LUMLEY)

# ZERO INFLATED NEGATIVE BINOMIAL NOT AVAILABLE IN R SURVEY PACKAGE
```