LR2_Example

Anthony Stachowitz

June 4, 2018

Point to local file repository:

Set the local directory for this project.

```
setwd("C:\\Users\\Anthony\\Desktop\\R studio projects\\LR2-exercise")
```

Load the National Health Interview Survey data:

load dataset NatHealth2011.rds into NH11 and set labs to NH11$label attributes.

```
NH11 <- readRDS("dataSets/NatHealth2011.rds")
labs <- attributes(NH11)$labels
```

predict the probability of being diagnosed with hypertension based on age, sex, sleep, and bmi:

check structure of hypev.

```
str(NH11$hypev)
```

```
##  Factor w/ 5 levels "1 Yes","2 No",..: 2 2 1 2 2 1 2 2 1 2 ...
```

check levels of hypev.

```
levels(NH11$hypev)
```

```
## [1] "1 Yes"             "2 No"              "7 Refused"
## [4] "8 Not ascertained" "9 Don't know"
```

collapse all missing values to NA.

```
NH11$hypev <- factor(NH11$hypev, levels=c("2 No", "1 Yes"))
```

run our regression model.

```
hyp.out <- glm(hypev~age_p+sex+sleep+bmi,
               data=NH11, family="binomial")
coef(summary(hyp.out))
```

```
##                 Estimate   Std. Error    z value     Pr(>|z|)
## (Intercept) -4.269466028 0.0564947294 -75.572820 0.000000e+00
## age_p        0.060699303 0.0008227207  73.778743 0.000000e+00
## sex2 Female -0.144025092 0.0267976605  -5.374540 7.677854e-08
## sleep       -0.007035776 0.0016397197  -4.290841 1.779981e-05
## bmi          0.018571704 0.0009510828  19.526906 6.485172e-85
```

transform the coefficients to make them easier to interpret.

```
hyp.out.tab <- coef(summary(hyp.out))
hyp.out.tab[, "Estimate"] <- exp(coef(hyp.out))
hyp.out.tab
```

```
##               Estimate   Std. Error    z value     Pr(>|z|)
## (Intercept) 0.01398925 0.0564947294 -75.572820 0.000000e+00
## age_p       1.06257935 0.0008227207  73.778743 0.000000e+00
## sex2 Female 0.86586602 0.0267976605  -5.374540 7.677854e-08
## sleep       0.99298892 0.0016397197  -4.290841 1.779981e-05
## bmi         1.01874523 0.0009510828  19.526906 6.485172e-85
```

How much more likely is a 63 year old female to have hypertension compared to a 33 year old female?

Create a dataset with predictors set at desired levels.

```
predDat <- with(NH11,
                expand.grid(age_p = c(33, 63),
                            sex = "2 Female",
                            bmi = mean(bmi, na.rm = TRUE),
                            sleep = mean(sleep, na.rm = TRUE)))
```

predict hypertension at those levels

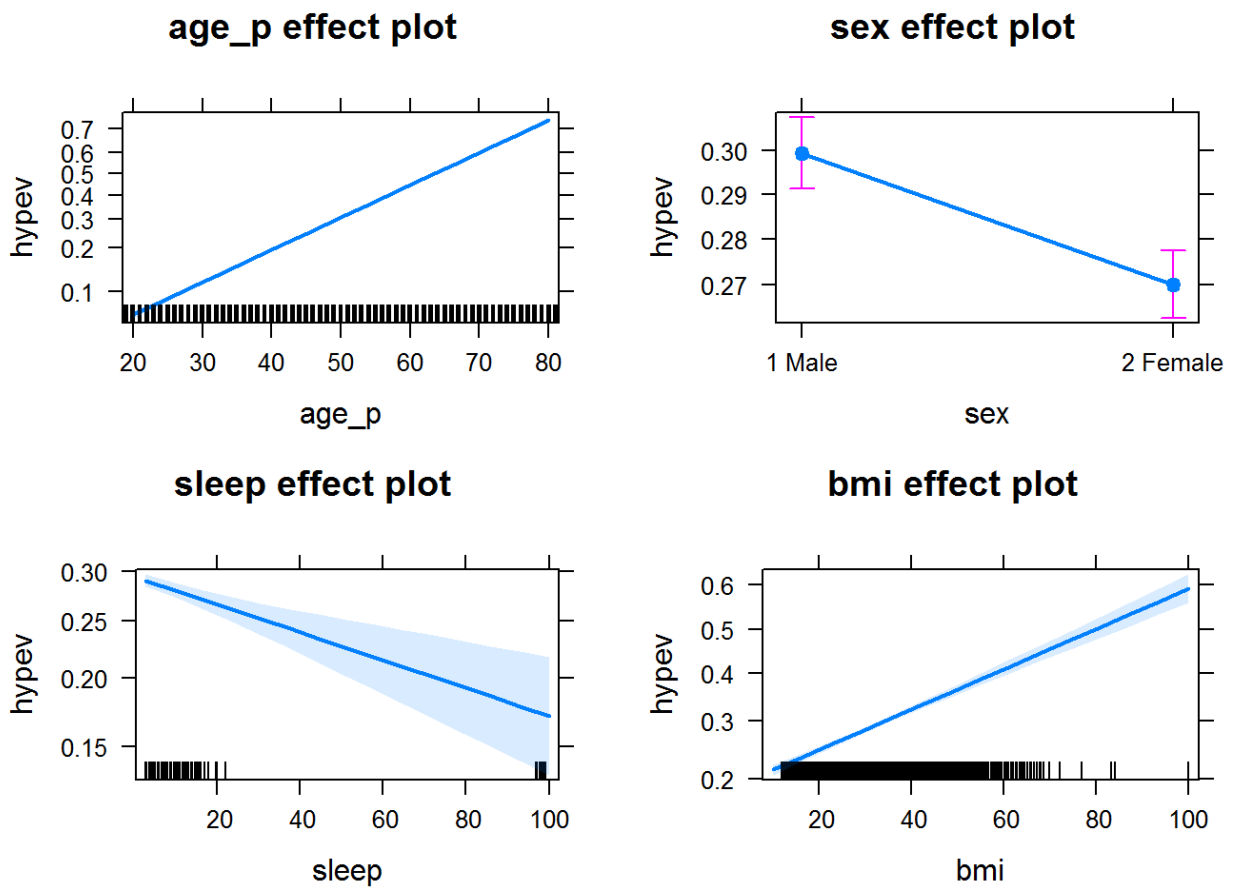
```
cbind(predDat, predict(hyp.out, type = "response",
                       se.fit = TRUE, interval="confidence",
                       newdata = predDat))
```

	age_p	sex	bmi	sleep	fit	se.fit	residual.scale
	<dbl>	<fctr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	33	2 Female	29.89565	7.86221	0.1289227	0.002849622	1
2	63	2 Female	29.89565	7.86221	0.4776303	0.004816059	1

2 rows

use the effects package to compute quantities of interest for us (cf. the Zelig package).

```
plot(allEffects(hyp.out))
```



Exercise: logistic regression

1. Use glm to conduct a logistic regression to predict ever worked (everwrk) using age (age_p) and marital status (r_maritl).

2. Predict the probability of working for each level of marital status.

Predict the probability of working for each level of marital status:

check structure of everwrk.

```
str(NH11$everwrk)
```

```
##  Factor w/ 5 levels "1 Yes","2 No",..: NA NA 1 NA NA NA NA NA 1 1 ...
```

check levels of everwrk.

```
levels(NH11$everwrk)
```

```
## [1] "1 Yes"             "2 No"              "7 Refused"
## [4] "8 Not ascertained" "9 Don't know"
```

collapse all missing values to NA.

```
NH11$everwrk <- factor(NH11$everwrk, levels=c("2 No", "1 Yes"))
```

Make marital status abbreviations for easier placement on the plot.

```
levels(NH11$r_maritl) <- c("U 14","Marr_IH","Marr_OH", "Marr_UK", "Wid", "Div", "Sep", "Nev Mar", "LW
P", "Unk")
```

run our regression model.

```
wrk.out <- glm(everwrk~age_p+r_maritl,
               data=NH11, family="binomial")
coef(summary(wrk.out))
```

```
##                    Estimate  Std. Error    z value     Pr(>|z|)
## (Intercept)      0.44024757 0.093537691  4.7066328 2.518419e-06
## age_p            0.02981220 0.001645433 18.1181481 2.291800e-73
## r_maritlMarr_OH -0.04967549 0.217309587 -0.2285932 8.191851e-01
## r_maritlWid     -0.68361771 0.084335382 -8.1059419 5.233844e-16
## r_maritlDiv      0.73011485 0.111680788  6.5375152 6.254929e-11
## r_maritlSep      0.12809081 0.151366140  0.8462316 3.974236e-01
## r_maritlNev Mar -0.34361068 0.069222260 -4.9638756 6.910023e-07
## r_maritlLWP      0.44358296 0.137769623  3.2197443 1.283050e-03
## r_maritlUnk     -0.39547953 0.492966577 -0.8022441 4.224118e-01
```

transform the coefficients to make them easier to interpret.

```
wrk.out.tab <- coef(summary(wrk.out))
wrk.out.tab[, "Estimate"] <- exp(coef(wrk.out))
wrk.out.tab
```

```
##                   Estimate  Std. Error    z value     Pr(>|z|)
## (Intercept)     1.5530917 0.093537691  4.7066328 2.518419e-06
## age_p           1.0302610 0.001645433 18.1181481 2.291800e-73
## r_maritlMarr_OH 0.9515382 0.217309587 -0.2285932 8.191851e-01
## r_maritlWid     0.5047875 0.084335382 -8.1059419 5.233844e-16
## r_maritlDiv     2.0753189 0.111680788  6.5375152 6.254929e-11
## r_maritlSep     1.1366562 0.151366140  0.8462316 3.974236e-01
## r_maritlNev Mar 0.7092050 0.069222260 -4.9638756 6.910023e-07
## r_maritlLWP     1.5582805 0.137769623  3.2197443 1.283050e-03
## r_maritlUnk     0.6733571 0.492966577 -0.8022441 4.224118e-01
```

What is the probability of work at each marital level??

Create a dataset with predictors set at desired levels.

```
predDat2 <- with(NH11,
                 expand.grid(r_maritl = c("Marr_IH","Marr_OH", "Wid", "Div", "Sep", "Nev Mar", "LWP",
"Unk"),
                             age_p = mean(age_p, na.rm = TRUE)))
```

predict if ever worked at those levels.

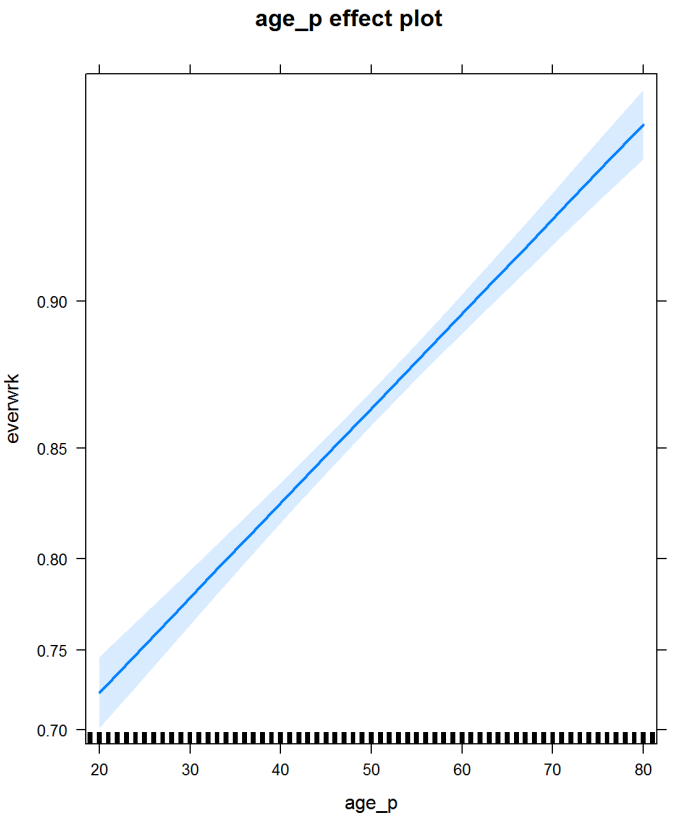
```
cbind(predDat2, predict(wrk.out, type = "response",
                        se.fit = TRUE, interval="confidence",
                        newdata = predDat2))
```

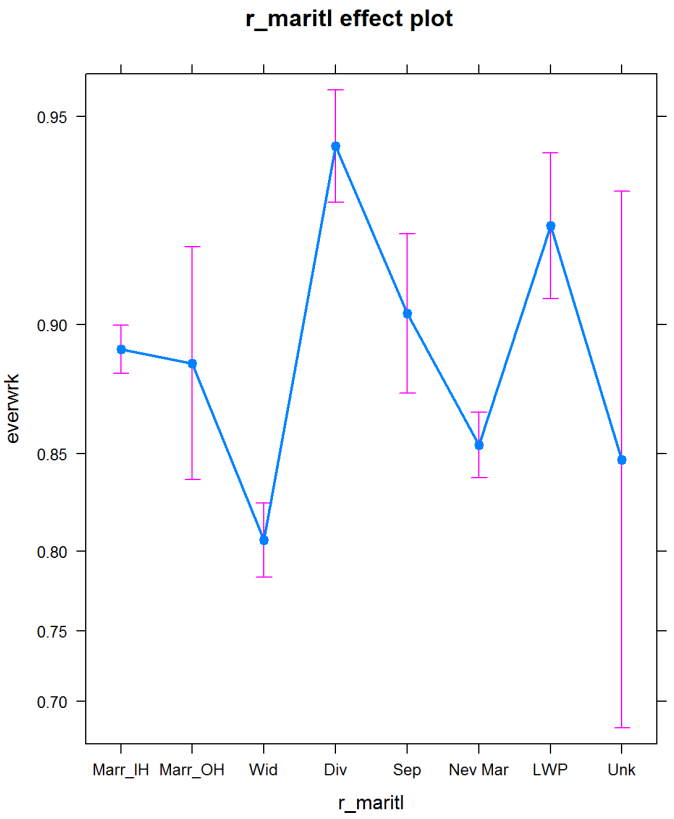
	r_maritl <fctr>	age_p <dbl>	fit <dbl>	se.fit <dbl>	residual.scale <dbl>
1	Marr_IH	48.10983	0.8669790	0.004999417	1
2	Marr_OH	48.10983	0.8611449	0.025466108	1
3	Wid	48.10983	0.7669001	0.013322250	1
4	Div	48.10983	0.9311585	0.006624603	1
5	Sep	48.10983	0.8810696	0.015188109	1
6	Nev Mar	48.10983	0.8221375	0.007606786	1
7	LWP	48.10983	0.9103642	0.010621449	1
8	Unk	48.10983	0.8144257	0.074230320	1

8 rows

use the effects package to compute quantities of interest for us (cf. the Zelig package).

```
plot(allEffects(wrk.out))
```


age_p effect plot
everwrk
0.90
0.85
0.80
0.75
0.70
20
30
40
50
60
70
80
age_p


r_maritl effect plot
everwrk
0.95
0.90
0.85
0.80
0.75
0.70
Marr_IH
Marr_OH
Wid
Div
Sep
Nev Mar
LWP
Unk
r_maritl