Análisis y Tratamiento de Datos con R: Departamento de Matemática

*

24 de noviembre de 2017

Modelos lineales generalizados

Ejemplo 1

```
library(effects)
```

```
## Loading required package: carData
```

```
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.
```

```
data(Arrests)
dim(Arrests)
```

```
## [1] 5226    8
```

```
Arrests$year <- as.factor(Arrests$year)
arrests.mod <- glm(released ~ employed + citizen + checks + colour*year + colour*age,
                   family=binomial, data=Arrests)
summary(arrests.mod)
```

```
##
## Call:
## glm(formula = released ~ employed + citizen + checks + colour *
##     year + colour * age, family = binomial, data = Arrests)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4787   0.3241   0.4485   0.6262   1.7132
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.344433   0.310075   1.111 0.266651
## employedYes          0.735064   0.084770   8.671  < 2e-16 ***
## citizenYes           0.585984   0.113772   5.151  2.6e-07 ***
## checks              -0.366642   0.026032 -14.084  < 2e-16 ***
## colourWhite          1.212517   0.349775   3.467 0.000527 ***
```

*,

```
## year1998             -0.431179   0.260359  -1.656 0.097702 .
## year1999             -0.094434   0.261545  -0.361 0.718052
## year2000             -0.010898   0.259207  -0.042 0.966465
## year2001              0.243063   0.263015   0.924 0.355413
## year2002              0.212955   0.353279   0.603 0.546644
## age                   0.028728   0.008619   3.333 0.000859 ***
## colourWhite:year1998  0.651956   0.313490   2.080 0.037555 *
## colourWhite:year1999  0.155950   0.307043   0.508 0.611516
## colourWhite:year2000  0.295754   0.306203   0.966 0.334108
## colourWhite:year2001 -0.380541   0.304054  -1.252 0.210731
## colourWhite:year2002 -0.617318   0.419255  -1.472 0.140909
## colourWhite:age      -0.037373   0.010200  -3.664 0.000248 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4776.3  on 5225  degrees of freedom
## Residual deviance: 4257.1  on 5209  degrees of freedom
## AIC: 4291.1
##
## Number of Fisher Scoring iterations: 5
```

```
library(car) # for the Anova function
```

```
##
## Attaching package: 'car'

## The following objects are masked from 'package:carData':
##
##     Guyer, UN, Vocab
```

```
Anova(arrests.mod)
```

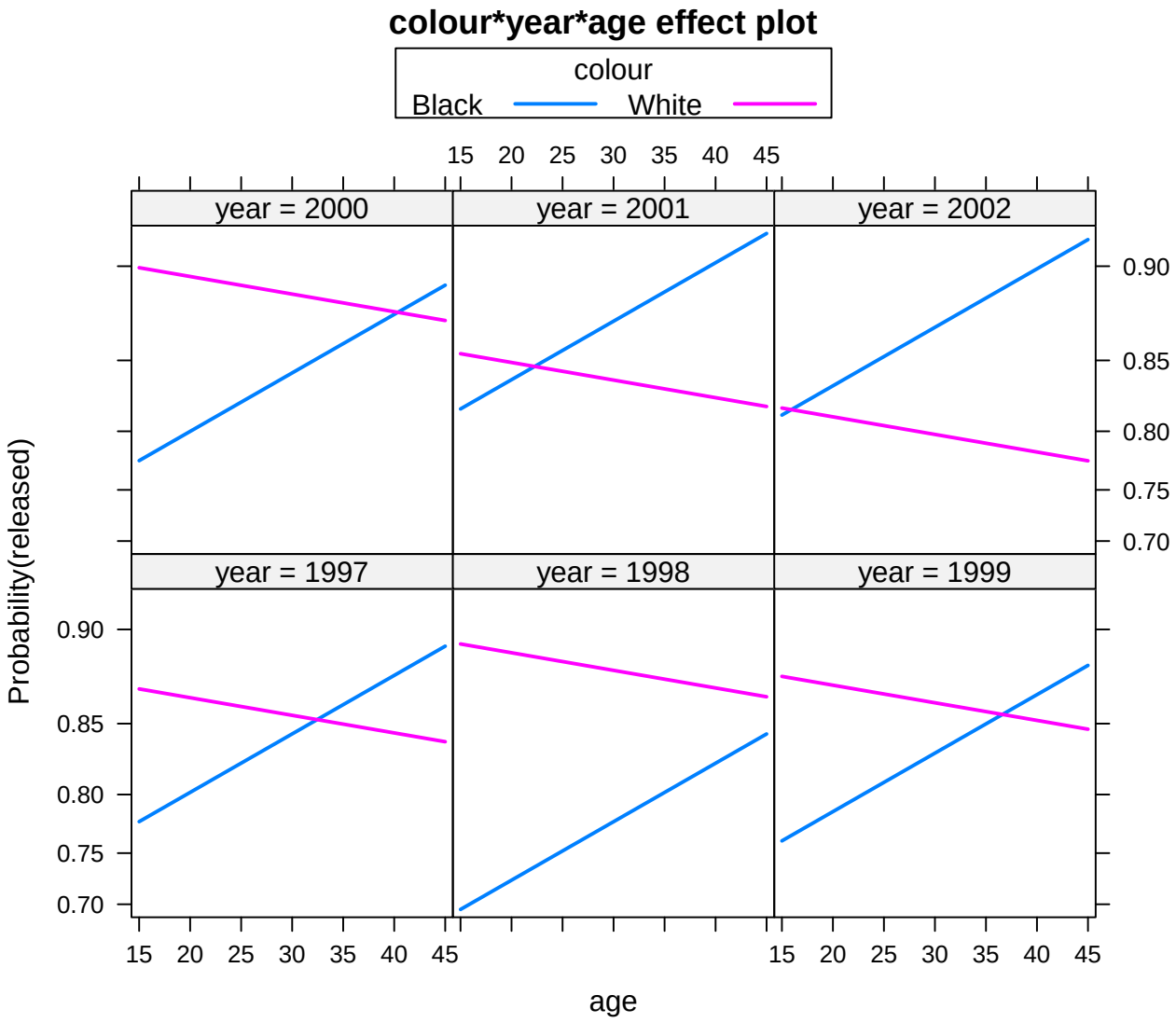
```
## Analysis of Deviance Table (Type II tests)
##
## Response: released
##             LR Chisq Df Pr(>Chisq)
## employed      72.673  1  < 2.2e-16 ***
## citizen       25.783  1  3.820e-07 ***
## checks       205.211  1  < 2.2e-16 ***
## colour        19.572  1  9.687e-06 ***
## year           6.087  5  0.2978477
## age            0.459  1  0.4982736
## colour:year   21.720  5  0.0005917 ***
## colour:age    13.886  1  0.0001942 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# model.matrix(arrests.mod)
```

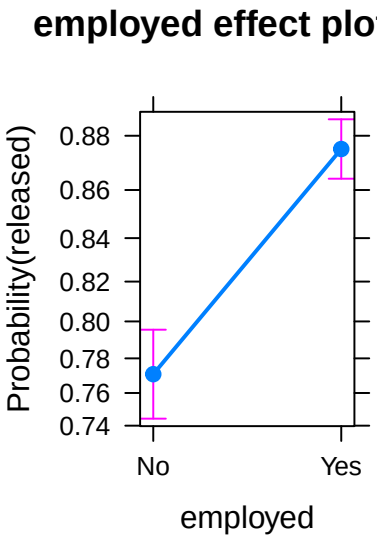
Effects plot

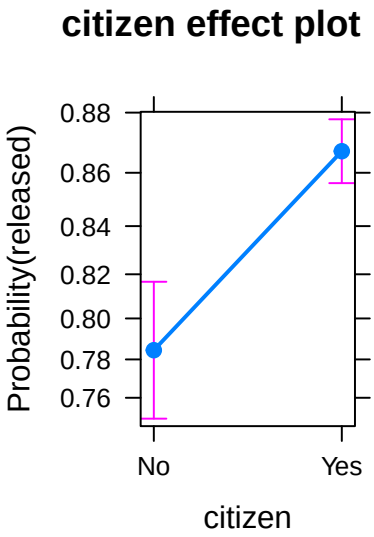
```
plot(effect("colour:year:age", arrests.mod, xlevels=list(age=15:45)), multiline=TRUE, ylab="Probability
```

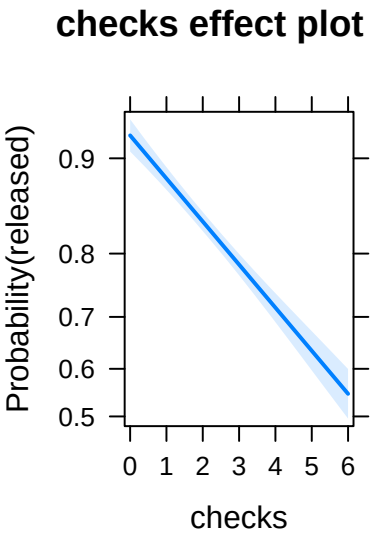
```
## NOTE: colour:year:age does not appear in the model
```

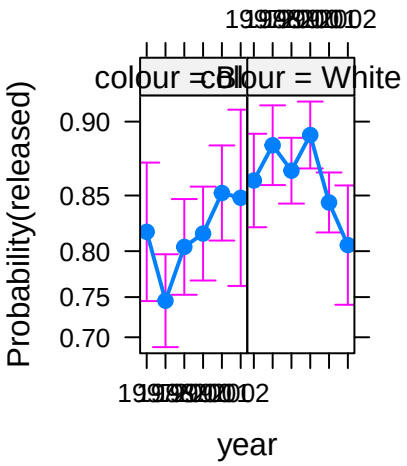


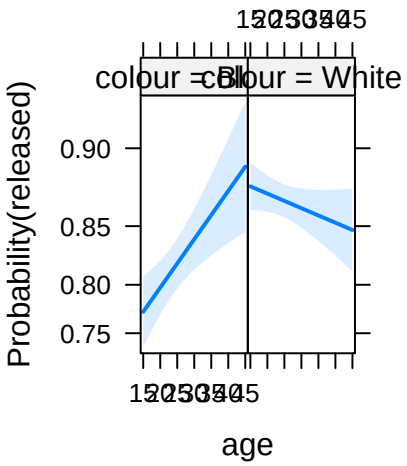

```
arrests.effects <- allEffects(arrests.mod, xlevels=list(age=15:45))
plot(arrests.effects, ylab="Probability(released)", rug=FALSE)
```





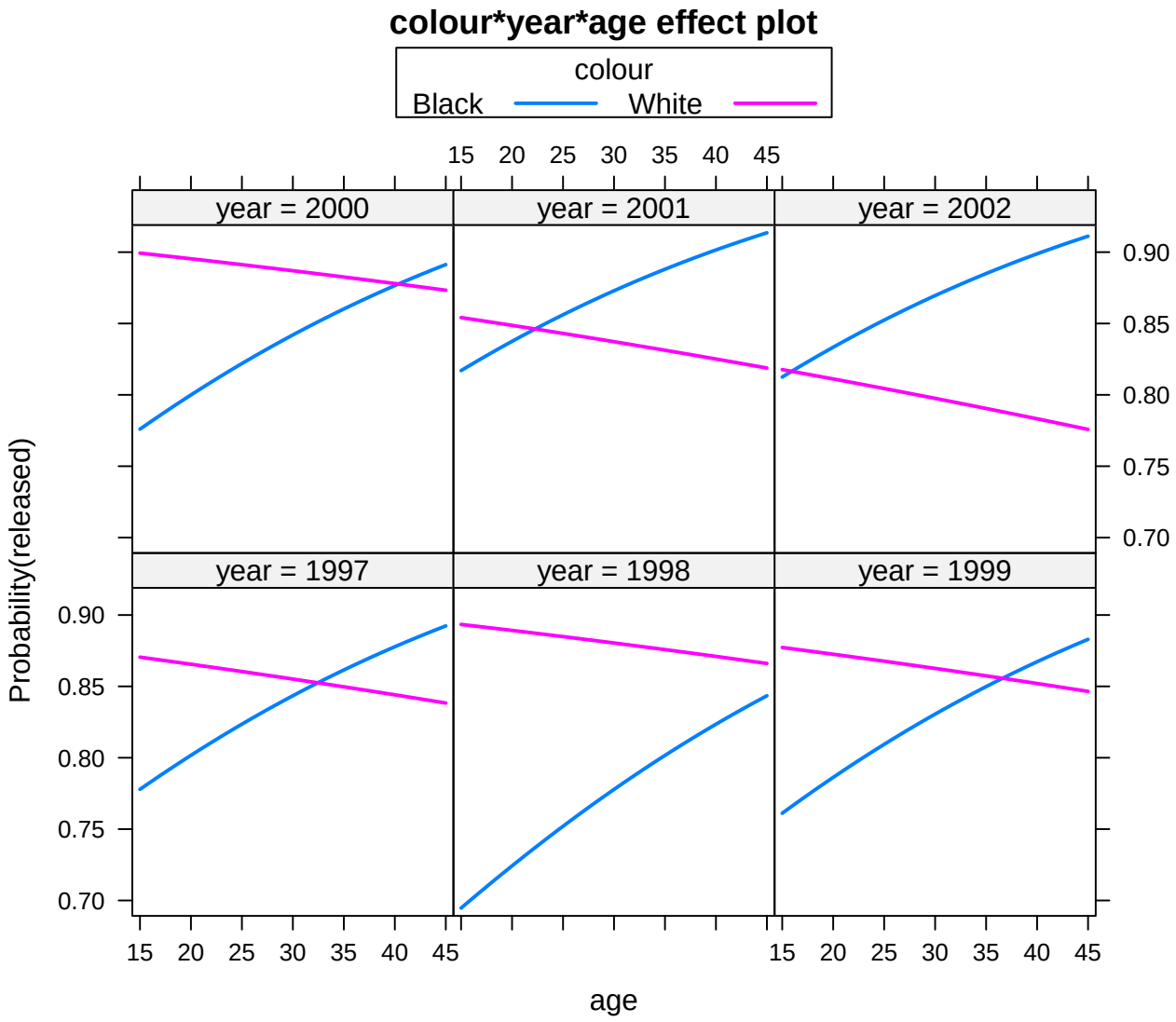







```
plot(effect("colour:year:age", arrests.mod, xlevels=list(age=15:45)),rescale.axis=FALSE, multiline=TRUE
```

```
## NOTE: colour:year:age does not appear in the model
```




```
colour.year <- effect("colour*year", arrests.mod)
colour.year
```

```
##
##  colour*year effect
##         year
## colour       1997      1998      1999      2000      2001      2002
##   Black 0.8186570 0.7457544 0.8042136 0.8170336 0.8519954 0.8481584
##   White 0.8615923 0.8858825 0.8687663 0.8922021 0.8443688 0.8059984
```

```
summary(colour.year)
```

```
##
##  colour*year effect
##         year
## colour       1997      1998      1999      2000      2001      2002
##   Black 0.8186570 0.7457544 0.8042136 0.8170336 0.8519954 0.8481584
##   White 0.8615923 0.8858825 0.8687663 0.8922021 0.8443688 0.8059984
##
```

```
##  Lower 95 Percent Confidence Limits
##        year
## colour       1997      1998      1999      2000      2001      2002
##   Black 0.7453153 0.6868619 0.7527357 0.7689626 0.8103962 0.7628886
##   White 0.8228754 0.8581311 0.8434593 0.8703397 0.8180851 0.7409054
##
##  Upper 95 Percent Confidence Limits
##        year
## colour       1997      1998      1999      2000      2001     2002
##   Black 0.8744366 0.7968468 0.8471505 0.8569631 0.8857544 0.906521
##   White 0.8929472 0.9087823 0.8905131 0.9107562 0.8674701 0.857875
```

En un modelo de regresión más general:

```
data(Prestige)
library(splines) # for bs
prestige.mod <- lm(prestige ~ log(income) + bs(education, df=3) + poly(women, 2),
data=Prestige)
summary(prestige.mod)
```

```
##
## Call:
## lm(formula = prestige ~ log(income) + bs(education, df = 3) +
##     poly(women, 2), data = Prestige)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.960  -4.983   0.372   3.787  17.092
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -72.916     15.494  -4.706 8.58e-06 ***
## log(income)               12.672      1.836   6.901 5.74e-10 ***
## bs(education, df = 3)1    -8.197      7.822  -1.048  0.29735
## bs(education, df = 3)2    25.660      5.497   4.668 9.97e-06 ***
## bs(education, df = 3)3    30.418      4.585   6.634 2.00e-09 ***
## poly(women, 2)1           11.978      9.384   1.276  0.20489
## poly(women, 2)2           18.465      6.828   2.704  0.00811 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.721 on 95 degrees of freedom
## Multiple R-squared:  0.8564, Adjusted R-squared:  0.8474
## F-statistic: 94.46 on 6 and 95 DF,  p-value: < 2.2e-16
```

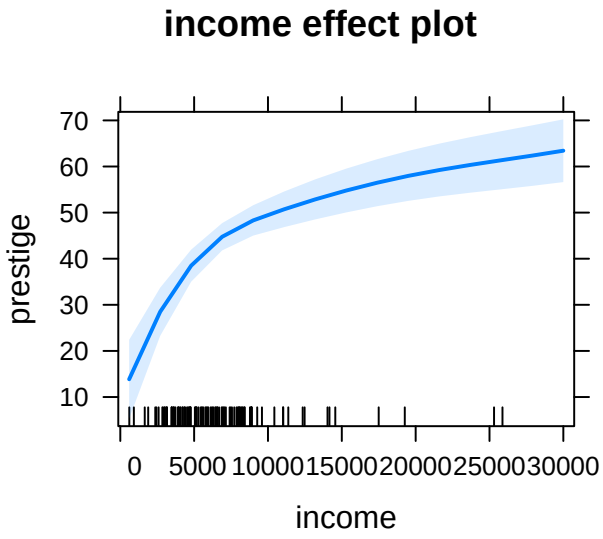
```
Anova(prestige.mod)
```

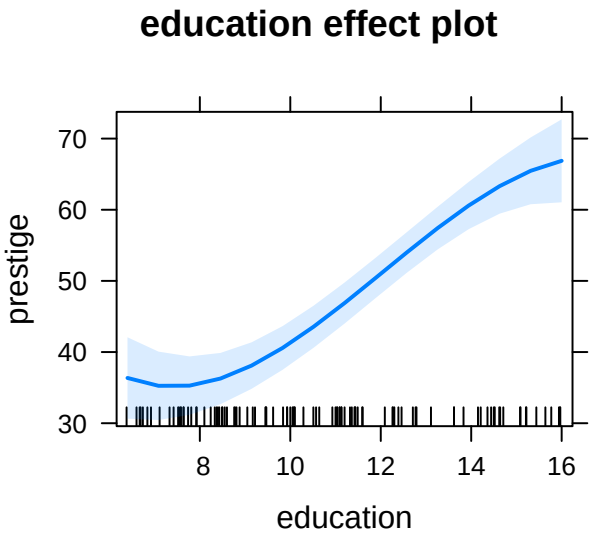
```
## Anova Table (Type II tests)
##
## Response: prestige
##                       Sum Sq Df F value    Pr(>F)
## log(income)           2151.5  1 47.6249 5.744e-10 ***
## bs(education, df = 3) 6067.0  3 44.7653 < 2.2e-16 ***
## poly(women, 2)         400.2  2  4.4295   0.01448 *
```

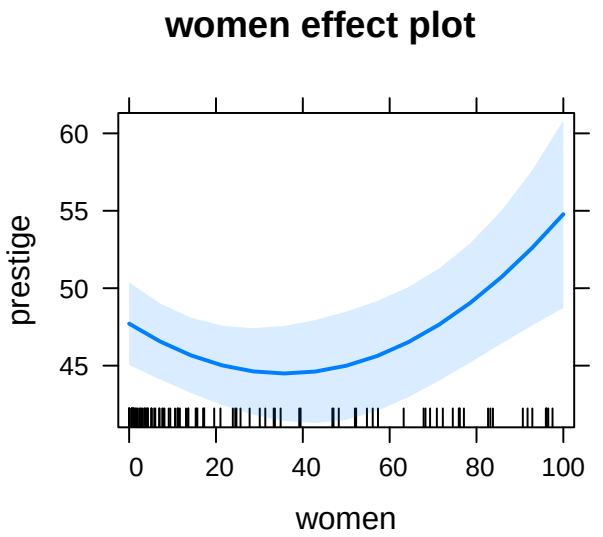
```
## Residuals              4291.8 95
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(allEffects(prestige.mod, default.levels=50))
```

```
## Warning in bs(education, degree = 3L, knots = numeric(0), Boundary.knots =
## c(6.38, : some 'x' values beyond boundary knots may cause ill-conditioned
## bases
```







Otro modelo con interacciones en las covariables:

```
data(Cowles)
cowles.mod <- glm(volunteer ~ sex + neuroticism*extraversion, data=Cowles, family=binomial)
summary(cowles.mod)
```

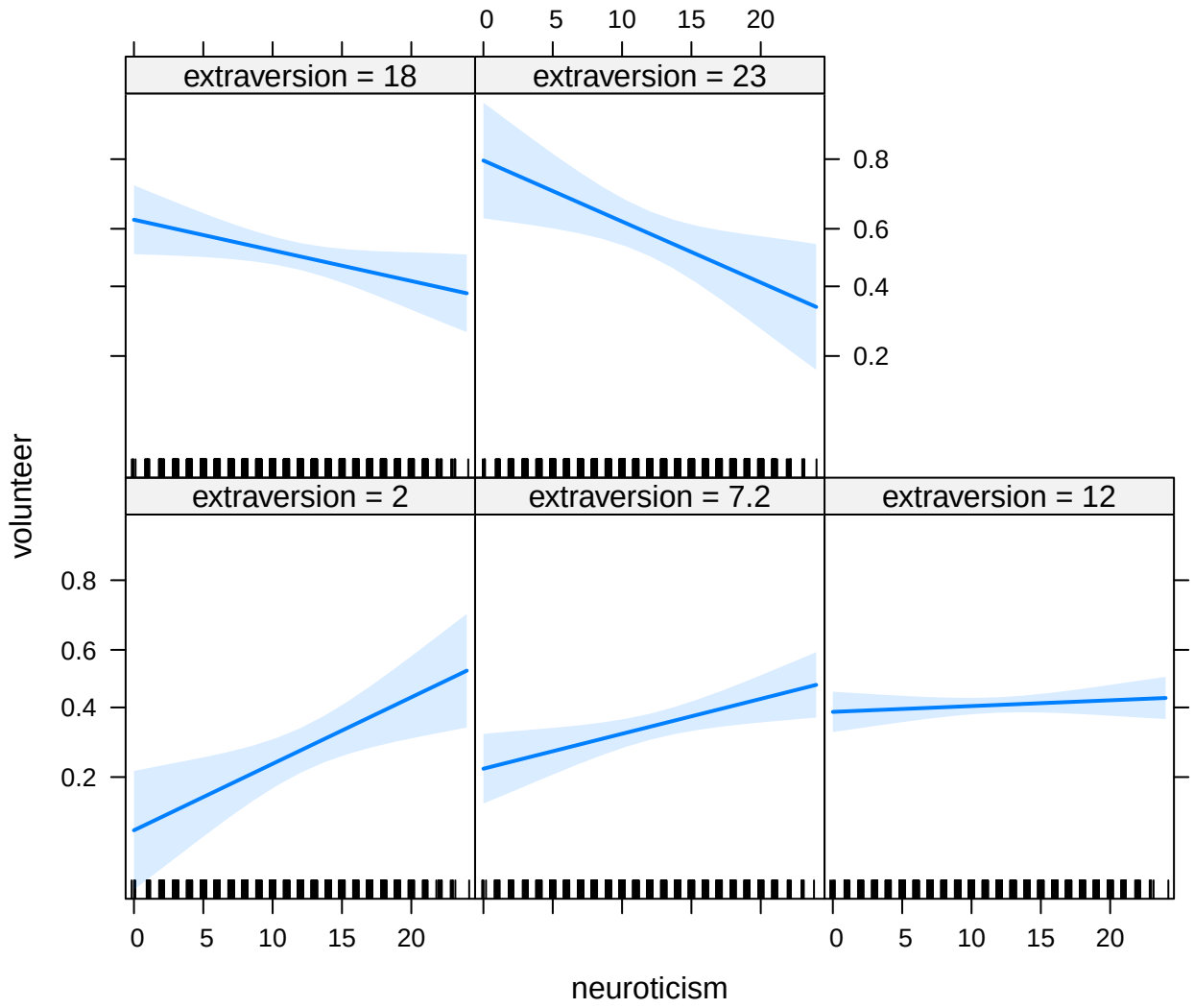
```
##
## Call:
## glm(formula = volunteer ~ sex + neuroticism * extraversion, family = binomial,
##     data = Cowles)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4749  -1.0602  -0.8934   1.2609   1.9978
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -2.358207   0.501320  -4.704 2.55e-06 ***
## sexmale                   -0.247152   0.111631  -2.214  0.02683 *
## neuroticism                0.110777   0.037648   2.942  0.00326 **
## extraversion               0.166816   0.037719   4.423 9.75e-06 ***
## neuroticism:extraversion -0.008552   0.002934  -2.915  0.00355 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1933.5  on 1420  degrees of freedom
## Residual deviance: 1897.4  on 1416  degrees of freedom
## AIC: 1907.4
##
## Number of Fisher Scoring iterations: 4
```

```
Anova(cowles.mod)
```

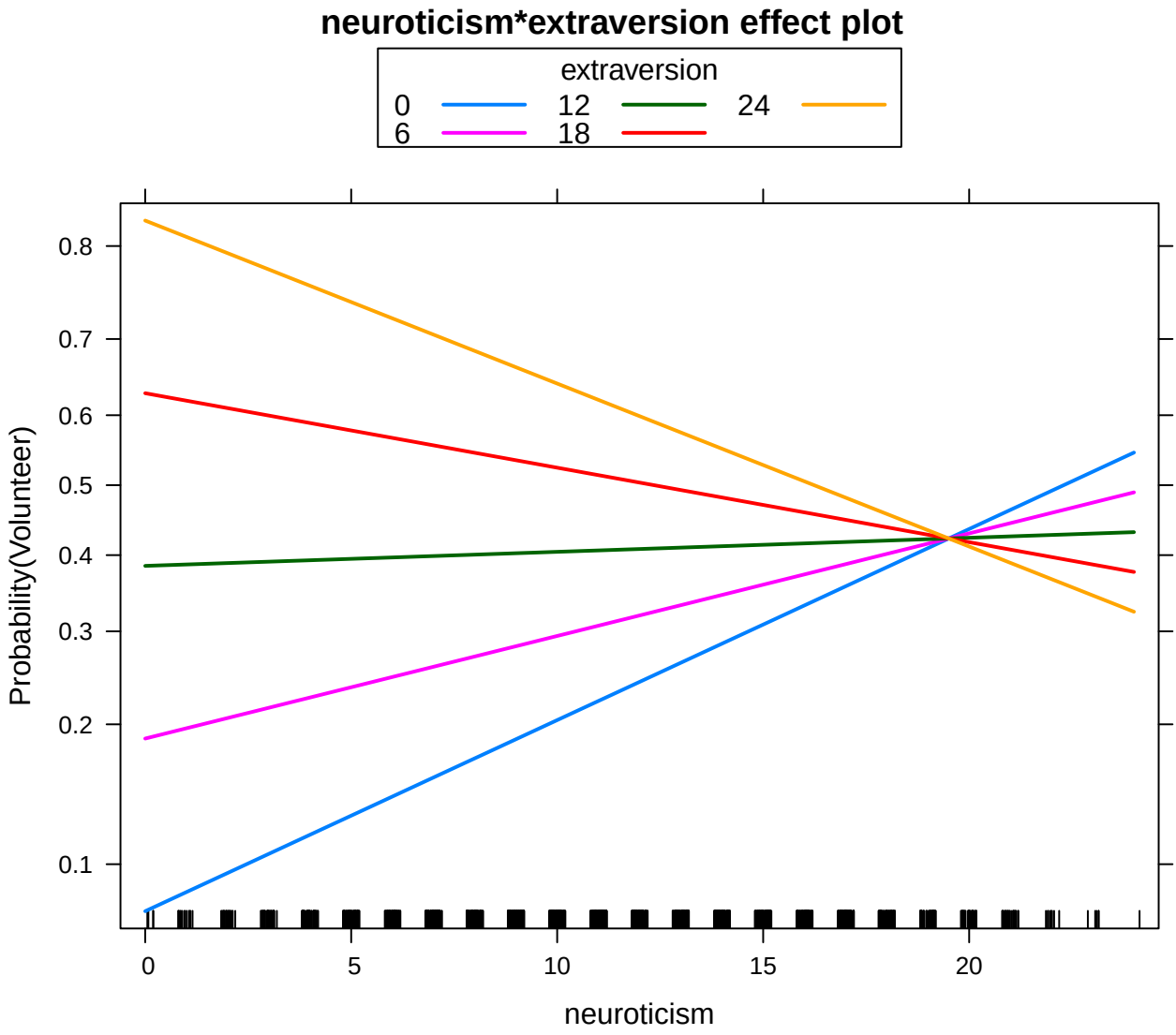
```
## Analysis of Deviance Table (Type II tests)
##
## Response: volunteer
##                           LR Chisq Df Pr(>Chisq)
## sex                         4.9184  1   0.026572 *
## neuroticism                 0.3139  1   0.575316
## extraversion               22.1372  1  2.538e-06 ***
## neuroticism:extraversion    8.6213  1   0.003323 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
plot(effect("neuroticism*extraversion", cowles.mod))
```



Alternativamente:

```
plot(effect("neuroticism*extraversion", cowles.mod,xlevels=list(neuroticism=0:24, extraversion=seq(0, 2
```




```
ne.effect <- effect("neuroticism*extraversion", cowles.mod, xlevels=list(neuroticism=seq(0, 24, 6), ext
ne.sumry <- summary(ne.effect, type="link")
fit <- ne.sumry$effect # fitted values for effect
lower <- ne.sumry$lower
upper <- ne.sumry$upper
par(mar=c(5, 4, 4, 4)+.1) # leave some extra room
plot(c(0, 30), range(c(lower, upper)), type="n", xaxt="n",
xlab="Extraversion", ylab="Logit of Volunteering")
axis(1, at=seq(0, 24, 6))
text(25, 1.8, "Neuroticism", adj=0)
probabilityAxis() # right-side axis, from car package
neuroticism <- seq(0, 24, 6)
for (neuro in 1:5){ # plot effects
   lines(0:24, fit[neuro,], lwd=2)
   text(25, fit[neuro, 25], paste("N = ", neuroticism[neuro]), adj=0)
}
extraversion <- 0:24
extra <- seq(1, 25, by=6)
for (neuro in c(1, 3, 5)){ # plot confidence bars
```

```
  arrows(extraversion[extra], lower[neuro, extra],
  extraversion[extra], upper[neuro, extra],
  angle=90, code=3, lty=2, length=0.05, col="red")
}
```