

# Homework3Answers.R

*dbarron*

*Mon Feb 19 12:35:05 2018*

```
#####  
## Week 3 Homework answers  
#####  
  
library(AER)  
  
## Loading required package: car  
## Loading required package: lmtest  
## Loading required package: zoo  
  
##  
## Attaching package: 'zoo'  
  
## The following objects are masked from 'package:base':  
##  
##   as.Date, as.Date.numeric  
  
## Loading required package: sandwich  
## Loading required package: survival  
  
data(ResumeNames)  
help('ResumeNames')  
  
## starting httpd help server ...  
  
## done  
  
str(ResumeNames)  
  
## 'data.frame':   4870 obs. of  27 variables:  
## $ name       : Factor w/ 36 levels "Allison","Anne",...: 1 7 23 24 3 15 5 22 24 36 ...  
## $ gender     : Factor w/ 2 levels "male","female": 2 2 2 2 2 1 2 2 2 1 ...  
## $ ethnicity  : Factor w/ 2 levels "cauc","afam": 1 1 2 2 1 1 1 2 2 2 ...  
## $ quality    : Factor w/ 2 levels "low","high": 1 2 1 2 2 1 2 2 1 2 ...  
## $ call       : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...  
## $ city       : Factor w/ 2 levels "boston","chicago": 2 2 2 2 2 2 2 2 2 2 ...  
## $ jobs       : int   2 3 1 4 3 2 2 4 3 2 ...  
## $ experience : int   6 6 6 6 22 6 5 21 3 6 ...  
## $ honors     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 2 1 1 1 1 ...  
## $ volunteer  : Factor w/ 2 levels "no","yes": 1 2 1 2 1 1 2 2 1 2 ...  
## $ military   : Factor w/ 2 levels "no","yes": 1 2 1 1 1 1 1 1 1 1 ...  
## $ holes      : Factor w/ 2 levels "no","yes": 2 1 1 2 1 1 1 2 1 1 ...  
## $ school     : Factor w/ 2 levels "no","yes": 1 2 2 1 2 1 2 1 1 2 ...  
## $ email      : Factor w/ 2 levels "no","yes": 1 2 1 2 2 1 2 2 1 2 ...  
## $ computer   : Factor w/ 2 levels "no","yes": 2 2 2 2 2 1 2 2 2 1 ...  
## $ special    : Factor w/ 2 levels "no","yes": 1 1 1 2 1 2 2 2 2 2 ...  
## $ college    : Factor w/ 2 levels "no","yes": 2 1 2 1 1 2 2 1 2 2 ...  
## $ minimum    : Factor w/ 14 levels "none","0","0.5",...: 8 8 8 8 14 1 1 14 1 1 ...  
## $ equal      : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
```

```

## $ wanted      : Factor w/ 6 levels "manager","supervisor",...: 2 2 2 2 3 6 6 3 6 6 ...
## $ requirements: Factor w/ 2 levels "no","yes": 2 2 2 2 2 1 1 2 1 1 ...
## $ reqexp      : Factor w/ 2 levels "no","yes": 2 2 2 2 2 1 1 2 1 1 ...
## $ reqcomm     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ reqeduc     : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
## $ reqcomp     : Factor w/ 2 levels "no","yes": 2 2 2 2 2 1 1 2 1 1 ...
## $ reqorg      : Factor w/ 2 levels "no","yes": 1 1 1 1 2 1 1 2 1 1 ...
## $ industry    : Factor w/ 7 levels "manufacturing",...: 1 1 1 1 6 4 4 6 4 4 ...

```

```
xtabs(~ call, data = ResumeNames)
```

```

## call
## no yes
## 4478 392

```

```
summary(ResumeNames)
```

```

##      name      gender  ethnicity  quality      call
## Tamika : 256  male  :1124  cauc:2435  low :2424  no :4478
## Anne   : 242  female:3746  afam:2435  high:2446  yes: 392
## Allison: 232
## Latonya: 230
## Emily  : 227
## Latoya : 226
## (Other):3457
##      city      jobs      experience  honors      volunteer
## boston :2166  Min.   :1.000  Min.   : 1.000  no :4613  no :2866
## chicago:2704  1st Qu.:3.000  1st Qu.: 5.000  yes: 257  yes:2004
##      Median :4.000  Median : 6.000
##      Mean   :3.661  Mean   : 7.843
##      3rd Qu.:4.000  3rd Qu.: 9.000
##      Max.   :7.000  Max.   :44.000
##
## military  holes      school  email      computer  special
## no :4397  no :2688  no :2145  no :2536  no : 874  no :3269
## yes: 473  yes:2182  yes:2725  yes:2334  yes:3996  yes:1601
##
##
##
##
## college      minimum  equal      wanted      requirements
## no :1366  none   :2746  no :3452  manager      : 741  no :1036
## yes:3504  some   :1064  yes:1418  supervisor   : 376  yes:3834
##      2      : 356      secretary   :1621
##      3      : 331      office support: 578
##      5      : 163      retail sales : 818
##      1      : 142      other       : 736
##      (Other): 68
## reqexp  reqcomm  reqeduc  reqcomp  reqorg
## no :2750  no :4262  no :4350  no :2741  no :4516
## yes:2120  yes: 608  yes: 520  yes:2129  yes: 354
##
##
##

```

```
##
##
##                industry
## manufacturing      : 404
## transport/communication : 148
## finance/insurance/real estate : 414
## trade              :1042
## business/personal services :1304
## health/education/social services: 754
## unknown           : 804

mod1 <- glm(call ~ ethnicity + gender + quality, data = ResumeNames, family = binomial)
summary(mod1)
```

```
##
## Call:
## glm(formula = call ~ ethnicity + gender + quality, family = binomial,
##      data = ResumeNames)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4770  -0.4356  -0.3867  -0.3525   2.4207
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.4350     0.1348 -18.058 < 2e-16 ***
## ethnicityafam -0.4399     0.1074  -4.096 4.2e-05 ***
## genderfemale   0.1276     0.1289   0.990 0.3222
## qualityhigh    0.1914     0.1059   1.806 0.0709 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2726.9  on 4869  degrees of freedom
## Residual deviance: 2705.7  on 4866  degrees of freedom
## AIC: 2713.7
##
## Number of Fisher Scoring iterations: 5
```

```
mod2 <- update(mod1, . ~ . + military)
summary(mod2)
```

```
##
## Call:
## glm(formula = call ~ ethnicity + gender + quality + military,
##      family = binomial, data = ResumeNames)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.4896  -0.4342  -0.3977  -0.3520   2.4612
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.41253     0.13496 -17.876 < 2e-16 ***
```

```

## ethnicityafam -0.43636    0.10743   -4.062 4.87e-05 ***
## genderfemale  0.09844    0.12966    0.759  0.4477
## qualityhigh   0.25295    0.10982    2.303  0.0213 *
## militaryyes  -0.38319    0.20498   -1.869  0.0616 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2726.9 on 4869 degrees of freedom
## Residual deviance: 2701.9 on 4865 degrees of freedom
## AIC: 2711.9
##
## Number of Fisher Scoring iterations: 5
mod3 <- update(mod2, . ~ . + industry)
summary(mod3)

##
## Call:
## glm(formula = call ~ ethnicity + gender + quality + military +
##      industry, family = binomial, data = ResumeNames)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6364  -0.4433  -0.3950  -0.3402   2.5694
##
## Coefficients:
##              Estimate Std. Error z value
## (Intercept)    -2.77328    0.25013  -11.087
## ethnicityafam  -0.43767    0.10761   -4.067
## genderfemale    0.04136    0.13632    0.303
## qualityhigh     0.24484    0.11003    2.225
## militaryyes    -0.33869    0.20590   -1.645
## industrytransport/communication  0.99309    0.32668    3.040
## industryfinance/insurance/real estate  0.27647    0.29183    0.947
## industrytrade    0.25784    0.25203    1.023
## industrybusiness/personal services  0.44651    0.24205    1.845
## industryhealth/education/social services  0.65809    0.25239    2.607
## industryunknown  0.34928    0.25718    1.358
##              Pr(>|z|)
## (Intercept)    < 2e-16 ***
## ethnicityafam  4.76e-05 ***
## genderfemale    0.76160
## qualityhigh     0.02607 *
## militaryyes     0.09999 .
## industrytransport/communication  0.00237 **
## industryfinance/insurance/real estate  0.34345
## industrytrade    0.30627
## industrybusiness/personal services  0.06508 .
## industryhealth/education/social services  0.00912 **
## industryunknown  0.17443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```

## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2726.9 on 4869 degrees of freedom
## Residual deviance: 2686.7 on 4859 degrees of freedom
## AIC: 2708.7
##
## Number of Fisher Scoring iterations: 5
anova(mod2, mod3, test = 'LRT')

## Analysis of Deviance Table
##
## Model 1: call ~ ethnicity + gender + quality + military
## Model 2: call ~ ethnicity + gender + quality + military + industry
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1      4865      2701.9
## 2      4859      2686.7  6    15.22  0.01861 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

anova(mod2, mod3)

## Analysis of Deviance Table
##
## Model 1: call ~ ethnicity + gender + quality + military
## Model 2: call ~ ethnicity + gender + quality + military + industry
##   Resid. Df Resid. Dev Df Deviance
## 1      4865      2701.9
## 2      4859      2686.7  6    15.22

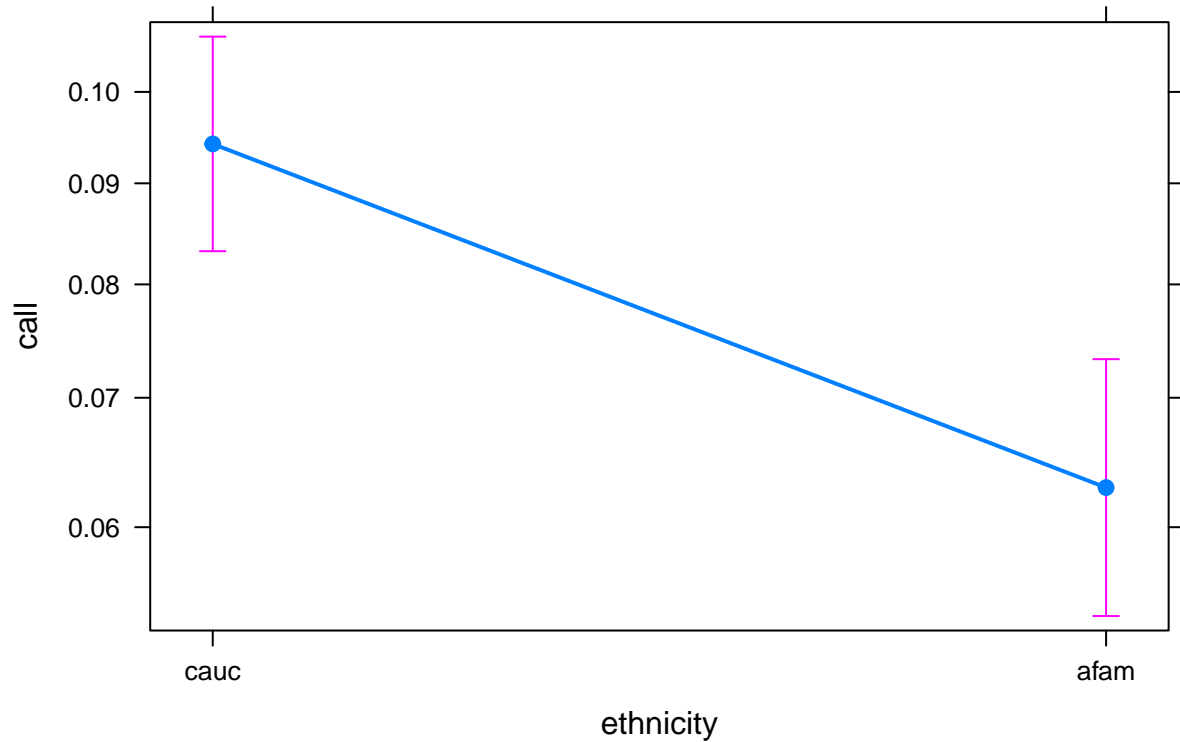
library(effects)

## Loading required package: carData
##
## Attaching package: 'carData'
##
## The following objects are masked from 'package:car':
##
##   Guyer, UN, Vocab
##
## lattice theme set by effectsTheme()
## See ?effectsTheme for details.

plot(Effect('ethnicity', mod3))

```

## ethnicity effect plot



```
stp <- step(mod1, scope = ~ ethnicity + gender + quality + military + industry + jobs +  
  experience + honors)
```

```
## Start: AIC=2713.67  
## call ~ ethnicity + gender + quality  
##  
##           Df Deviance   AIC  
## + honors   1  2687.2 2697.2  
## + experience 1  2690.4 2700.4  
## + industry  6  2689.6 2709.6  
## + military  1  2701.9 2711.9  
## - gender   1  2706.7 2712.7  
## <none>     2705.7 2713.7  
## - quality  1  2708.9 2714.9  
## + jobs    1  2705.6 2715.6  
## - ethnicity 1  2722.8 2728.8  
##  
## Step: AIC=2697.15  
## call ~ ethnicity + gender + quality + honors  
##  
##           Df Deviance   AIC  
## + experience 1  2676.1 2688.1  
## + industry  6  2670.1 2692.1  
## + military  1  2683.5 2695.5  
## - gender   1  2687.9 2695.9  
## - quality  1  2689.2 2697.2
```

```

## <none>          2687.2 2697.2
## + jobs          1  2686.9 2698.9
## - ethnicity     1  2704.1 2712.1
## - honors        1  2705.7 2713.7
##
## Step: AIC=2688.06
## call ~ ethnicity + gender + quality + honors + experience
##
##           Df Deviance   AIC
## + industry  6  2660.2 2684.2
## - gender    1  2676.7 2686.7
## - quality   1  2677.3 2687.3
## <none>      1  2676.1 2688.1
## + military  1  2674.9 2688.9
## + jobs      1  2675.6 2689.6
## - experience 1  2687.2 2697.2
## - honors    1  2690.4 2700.4
## - ethnicity 1  2693.0 2703.0
##
## Step: AIC=2684.16
## call ~ ethnicity + gender + quality + honors + experience + industry
##
##           Df Deviance   AIC
## - gender    1  2660.2 2682.2
## - quality   1  2661.4 2683.4
## <none>      1  2660.2 2684.2
## + military  1  2659.3 2685.3
## + jobs      1  2659.6 2685.6
## - industry  6  2676.1 2688.1
## - experience 1  2670.1 2692.1
## - honors    1  2675.3 2697.3
## - ethnicity 1  2677.1 2699.1
##
## Step: AIC=2682.25
## call ~ ethnicity + quality + honors + experience + industry
##
##           Df Deviance   AIC
## - quality   1  2661.5 2681.5
## <none>      1  2660.2 2682.2
## + military  1  2659.3 2683.3
## + jobs      1  2659.7 2683.7
## + gender    1  2660.2 2684.2
## - industry  6  2676.7 2686.7
## - experience 1  2670.2 2690.2
## - honors    1  2675.5 2695.5
## - ethnicity 1  2677.1 2697.1
##
## Step: AIC=2681.46
## call ~ ethnicity + honors + experience + industry
##
##           Df Deviance   AIC
## <none>      1  2661.5 2681.5
## + quality   1  2660.2 2682.2
## + military  1  2661.1 2683.1

```

```
## + jobs      1  2661.2 2683.2
## + gender    1  2661.4 2683.4
## - industry  6  2677.9 2685.9
## - experience 1  2672.2 2690.2
## - honors    1  2677.6 2695.6
## - ethnicity 1  2678.3 2696.3
```

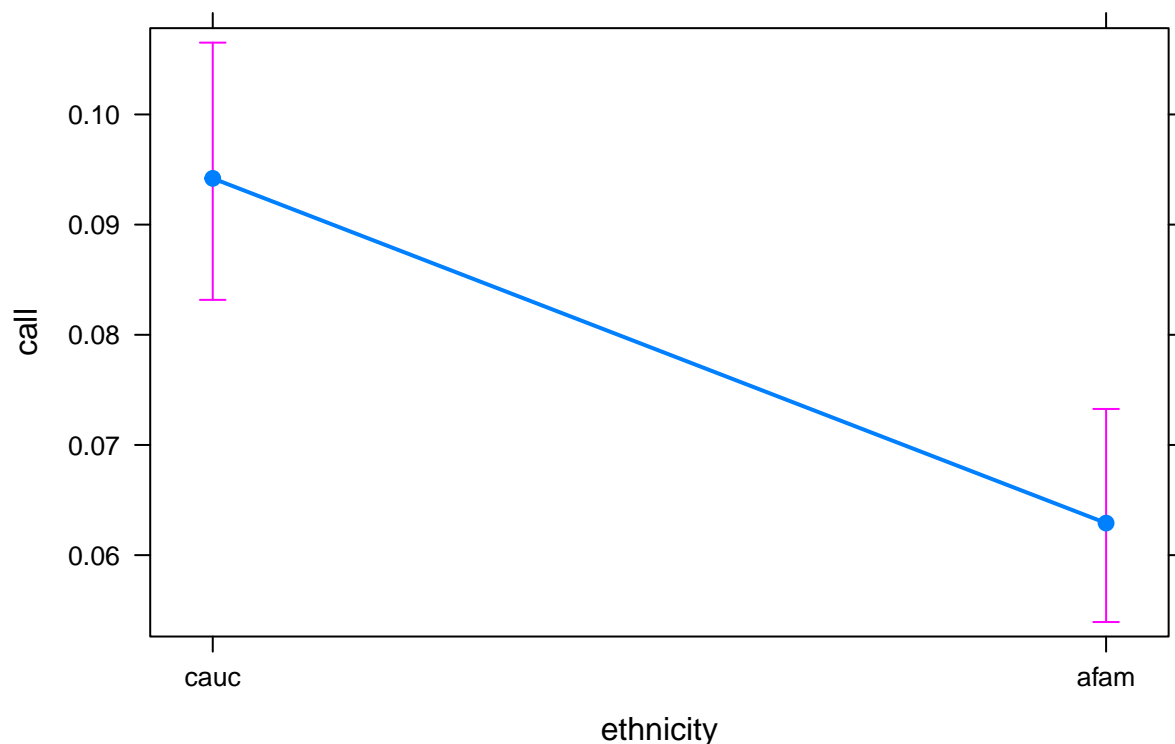
```
summary(stp)
```

```
##
## Call:
## glm(formula = call ~ ethnicity + honors + experience + industry,
##      family = binomial, data = ResumeNames)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9050  -0.4325  -0.3831  -0.3305   2.5866
##
## Coefficients:
##              Estimate Std. Error z value
## (Intercept)   -2.965851   0.239381 -12.390
## ethnicityafam -0.439000   0.107954  -4.067
## honorsyes      0.786467   0.183086   4.296
## experience     0.031829   0.009449   3.368
## industrytransport/communication 1.006558   0.327223   3.076
## industryfinance/insurance/real estate 0.251802   0.291493   0.864
## industrytrade  0.242296   0.252318   0.960
## industrybusiness/personal services 0.452455   0.241999   1.870
## industryhealth/education/social services 0.670778   0.250861   2.674
## industryunknown 0.365691   0.257528   1.420
##              Pr(>|z|)
## (Intercept)   < 2e-16 ***
## ethnicityafam 4.77e-05 ***
## honorsyes     1.74e-05 ***
## experience    0.000756 ***
## industrytransport/communication 0.002098 **
## industryfinance/insurance/real estate 0.387678
## industrytrade 0.336913
## industrybusiness/personal services 0.061531 .
## industryhealth/education/social services 0.007497 **
## industryunknown 0.155606
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 2726.9  on 4869  degrees of freedom
## Residual deviance: 2661.5  on 4860  degrees of freedom
## AIC: 2681.5
##
## Number of Fisher Scoring iterations: 5
```

```
plot(Effect('ethnicity', mod3), type = 'response')
```



## ethnicity effect plot



```
stp2 <- step(mod1, scope = ~ gender + ethnicity + quality + city + jobs + experience + honors + volunteer + military + holes + school + email + computer + special + college + minimum + equal + wanted + requirements + reqexp + reqcomm + reqeduc + reqcomp + reqorg + industry, direction = 'forward', trace = 0)
```

```
summary(stp2)
```

```
##  
## Call:  
## glm(formula = call ~ ethnicity + gender + quality + special +  
##   city + holes + honors + experience + reqeduc + reqcomp +  
##   equal + reqorg + industry + jobs + college, family = binomial,  
##   data = ResumeNames)  
##  
## Deviance Residuals:  
##   Min       1Q   Median       3Q      Max  
## -1.0759 -0.4442 -0.3485 -0.2797  2.9176  
##  
## Coefficients:  
##               Estimate Std. Error z value  
## (Intercept)    -2.99491    0.34724  -8.625  
## ethnicityafam  -0.45545    0.10926  -4.168  
## genderfemale    0.09406    0.14594   0.644  
## qualityhigh     0.21135    0.11619   1.819  
## specialyes      0.73879    0.11645   6.344  
## citychicago   -0.36599    0.12742  -2.872
```

```

## holesyes                0.39575    0.12156    3.256
## honorsyes              0.46299    0.19835    2.334
## experience             0.03225    0.01063    3.035
## reqeducyes            -0.53778    0.21225   -2.534
## reqcompyes            -0.24148    0.12547   -1.925
## equalyes              0.24716    0.12450    1.985
## reqorgyes            -0.49544    0.26505   -1.869
## industrytransport/communication 0.87204    0.33267    2.621
## industryfinance/insurance/real estate 0.19762    0.29756    0.664
## industrytrade         0.04966    0.25766    0.193
## industrybusiness/personal services 0.30898    0.24702    1.251
## industryhealth/education/social services 0.51129    0.25737    1.987
## industryunknown       0.33335    0.26304    1.267
## jobs                  -0.10141    0.05089   -1.993
## collegeyes           0.22447    0.13000    1.727
##
## Pr(>|z|)
## (Intercept)          < 2e-16 ***
## ethnicityafam       3.07e-05 ***
## genderfemale        0.51926
## qualityhigh        0.06891 .
## specialyes         2.23e-10 ***
## citychicago       0.00408 **
## holesyes           0.00113 **
## honorsyes          0.01958 *
## experience          0.00241 **
## reqeducyes         0.01129 *
## reqcompyes         0.05428 .
## equalyes           0.04711 *
## reqorgyes          0.06159 .
## industrytransport/communication 0.00876 **
## industryfinance/insurance/real estate 0.50662
## industrytrade      0.84717
## industrybusiness/personal services 0.21100
## industryhealth/education/social services 0.04697 *
## industryunknown    0.20506
## jobs               0.04628 *
## collegeyes        0.08423 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 2726.9 on 4869 degrees of freedom
## Residual deviance: 2568.2 on 4849 degrees of freedom
## AIC: 2610.2
##
## Number of Fisher Scoring iterations: 5
plot(Effect('ethnicity', stp2), type = 'response')

```

ethnicity effect plot

