

Advanced Regression Methods for Independent Data

STAT/BIOST 570, 2020

Practical Aspects of Generalized Linear Models

Mauricio Sadinle
Department of Biostatistics
University of Washington
msadinle@uw.edu

Generalized Linear Models in R

We'll present some examples to illustrate how to fit and check GLMs using R.

In R, GLMs are implemented in the `glm` function

```
glm(formula, family = gaussian, data, weights, subset,  
    na.action, start = NULL, etastart, mustart, offset,  
    control = list(...), model = TRUE, method = "glm.fit",  
    x = FALSE, y = TRUE, singular.ok = TRUE, contrasts = NULL, ...)
```

where the `family` argument is specified as one of

```
binomial(link = "logit") # accepts the links logit, probit, cauchit, log, cloglog  
gaussian(link = "identity") # accepts identity, log, inverse  
Gamma(link = "inverse") # accepts inverse, identity, log  
inverse.gaussian(link = "1/mu^2") # accepts 1/mu^2, inverse, identity, log  
poisson(link = "log") # accepts log, identity, sqrt
```

Pseudo-examples:

```
glm(y~x1+x2, family = Gamma) # Gamma GLM with inverse link (default)  
glm(y~x1+x2, family = Gamma(link="log")) # Gamma GLM with log link  
glm(y~x1+x2, family = poisson) # Poisson GLM with log link (default)
```

Back to the FEV Data

- FEV: forced expiratory volume. FEV1: amount of air you can force from your lungs in one second.
- Data from 654 children and youths ages 3–19 in East Boston, 1980. (Childhood Respiratory Disease Study).
- For more information visit: <http://www.statsci.org/data/general/fev.html>

```
> url <- "http://www.statsci.org/data/general/fev.txt"
> data <- read.table(file = url, header = T, sep="\t", stringsAsFactors = F)
> data$Sex <- factor(data$Sex, levels=c("Male","Female"), labels=c(0,1))
> data$Smoker <- factor(data$Smoker, levels=c("Non","Current"), labels=c(0,1))
> data$HeightC <- data$Height - 60
> data$AgeC <- data$Age - 10
> data$HeightCSq <- data$HeightC^2
```

Back to the FEV Data

Based on an exploratory analysis, we last looked at a normal linear model with regression function:

$$\begin{aligned} E(\text{FEV1} \mid \mathbf{X}) = & \beta_0 + \beta_1 \text{Height} + \beta_2 \text{Height}^2 + \beta_3 \text{Age} + \\ & (\beta_4 + \beta_5 \text{Height} + \beta_6 \text{Height}^2 + \beta_7 \text{Age}) I(\text{Sex}=\text{Female}) + \\ & (\beta_8 + \beta_9 \text{Height} + \beta_{10} \text{Height}^2 + \beta_{11} \text{Age}) I(\text{Smoker}=\text{Yes}) \end{aligned}$$

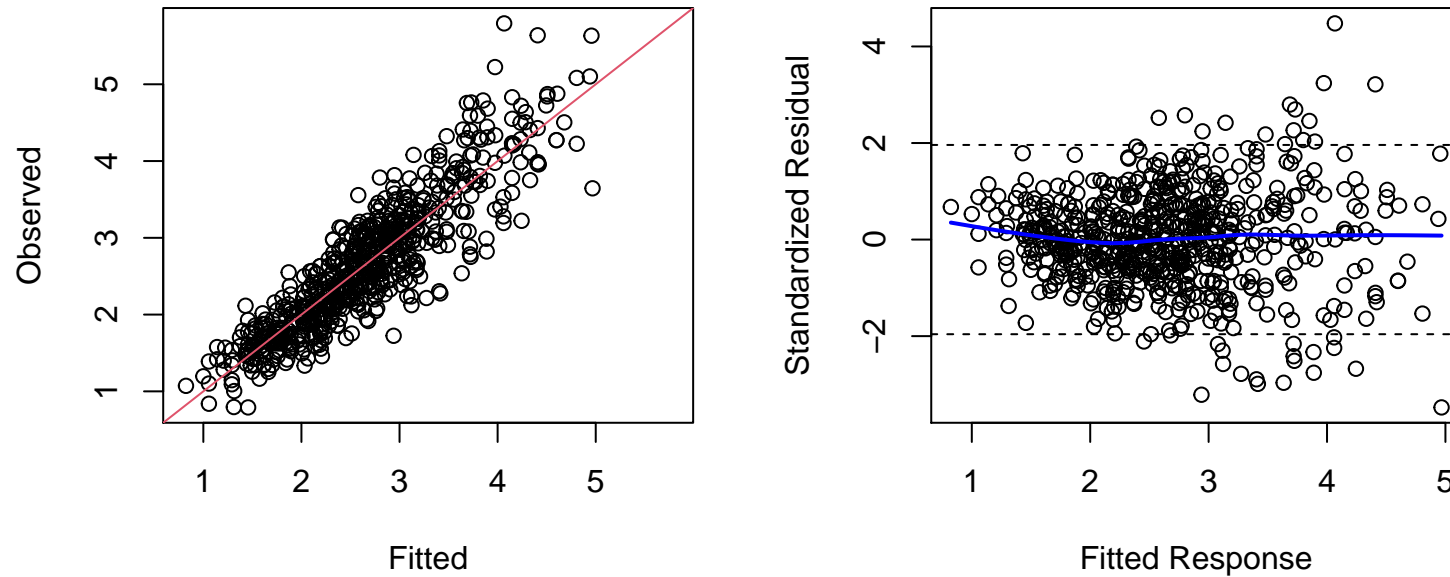
```
> model3 <- lm(  
+       FEV ~ (HeightC+HeightCSq+AgeC)*(Sex+Smoker),  
+       data=data)
```

Back to the FEV Data

We were mainly concerned about the assumption of homoskedasticity

```
> limits3 <- range(data$FEV, fitted(model3))
> par(mfrow = c(1,2))
> plot(fitted(model3), data$FEV, xlab="Fitted", ylab="Observed", xlim=limits3, ylim=limits3)
> abline(a=0, b=1, col=2)
> library(MASS)
> resid_mod3 <- stdres(model3)
> plot(fitted(model3), stdres(model3),
+   ylab = "Standardized Residual", xlab = "Fitted Response")
> lines(lowess(fitted(model3), resid_mod3), lwd = 2, col = "blue")
> abline(h = 1.96, lty = 2)
> abline(h = -1.96, lty = 2)
```

Back to the FEV Data

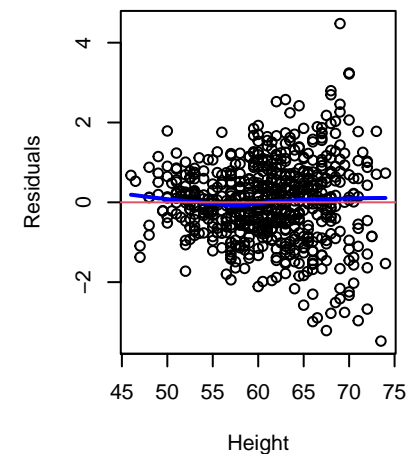
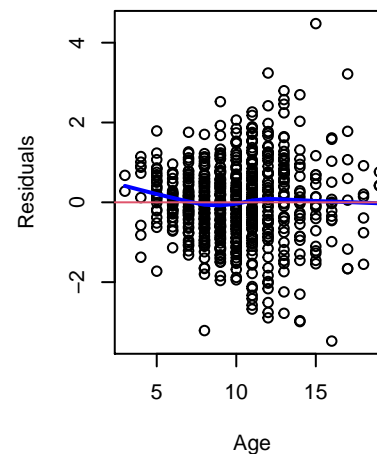
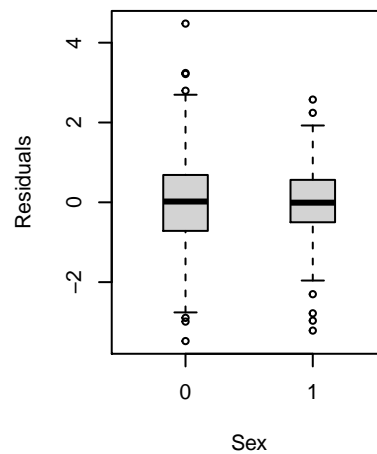
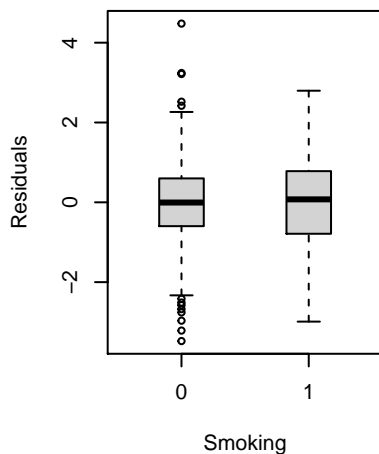


Lack of homoscedasticity seems to be the main issue here, so the normal linear model doesn't hold

Back to the FEV Data

Residuals vs covariates: variance of response is not constant

```
> par(mfrow = c(1,4))
> plot(resid_mod3 ~ data$Smoker, xlab = "Smoking", ylab = "Residuals",
+   col = 'lightgray', boxwex = .35)
> plot(resid_mod3 ~ data$Sex, xlab = "Sex", ylab = "Residuals",
+   col = 'lightgray', boxwex = .35)
> plot(data$Age, resid_mod3, xlab = "Age", ylab = "Residuals",)
> lines(lowess(data$Age, resid_mod3), lwd = 2, col = "blue")
> abline(h=0, col=2)
> plot(data$Height, resid_mod3, xlab = "Height", ylab = "Residuals",)
> lines(lowess(data$Height, resid_mod3), lwd = 2, col = "blue")
> abline(h=0, col=2)
```



Back to the FEV Data

Remember that under the gamma GLM

$$\text{var}(Y_i | \mathbf{x}_i) = \alpha E(Y_i | \mathbf{x}_i)^2,$$

which might be appropriate for our data

It is OK to use the identity link here since the outcome is away from zero

Let's consider a gamma GLM with identity link:

```
> FEV_Gamma <- glm(  
+   FEV ~ (HeightC+HeightCSq+AgeC)*(Sex+Smoker),  
+   family = Gamma(link="identity"),  
+   data=data)
```

In this case, $\hat{\alpha}$ is

```
> summary(FEV_Gamma)$dispersion
```

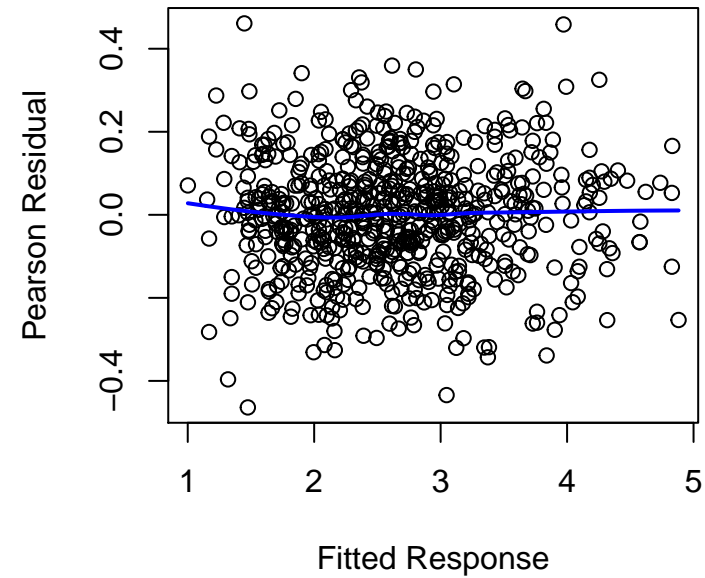
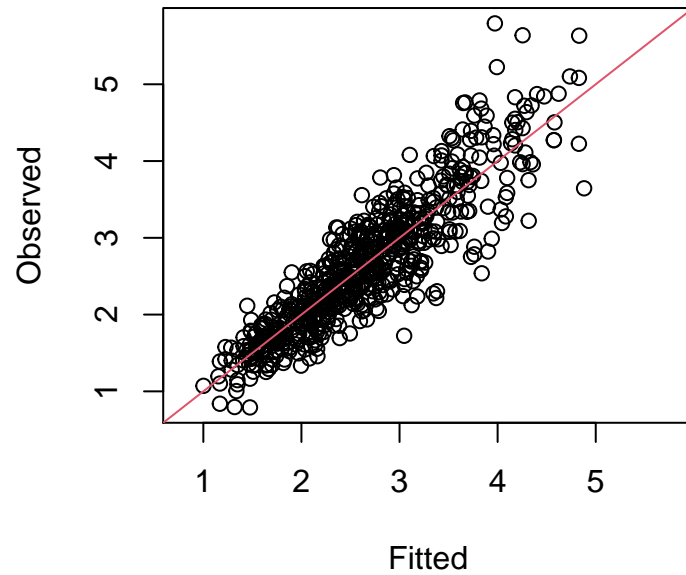
```
[1] 0.02015759
```


Back to the FEV Data

For the gamma GLM to be appropriate, we need to *not* find patterns in the Pearson residuals

```
> resid_Gamma <- residuals(FEV_Gamma, type = "pearson")
> fitted_Gamma <- fitted(FEV_Gamma)
> limits <- range(data$FEV, fitted_Gamma)
> par(mfrow = c(1,2))
> plot(fitted_Gamma, data$FEV, xlab="Fitted", ylab="Observed", xlim=limits, ylim=limits)
> abline(a=0, b=1, col=2)
> library(MASS)
> plot(fitted_Gamma, resid_Gamma,
+      ylab = "Pearson Residual", xlab = "Fitted Response")
> lines(lowess(fitted_Gamma, resid_Gamma), lwd = 2, col = "blue")
> abline(h = 1.96, lty = 2)
> abline(h = -1.96, lty = 2)
```

Back to the FEV Data

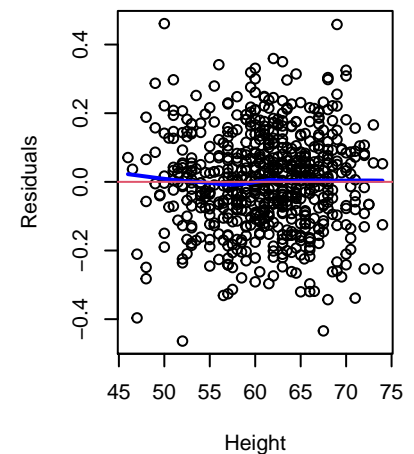
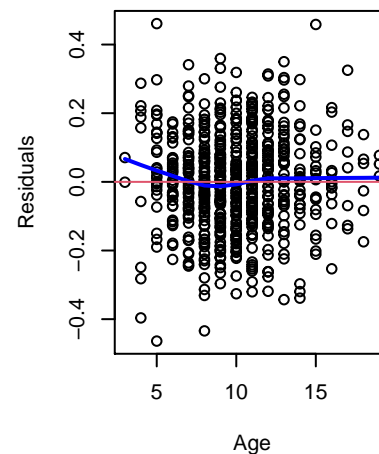
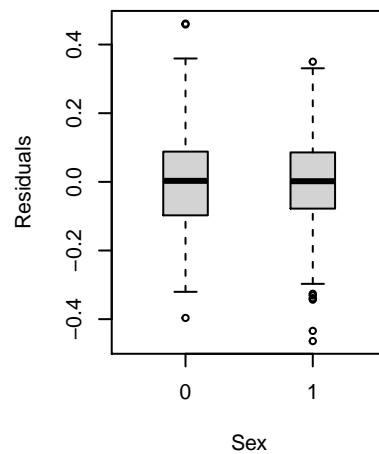
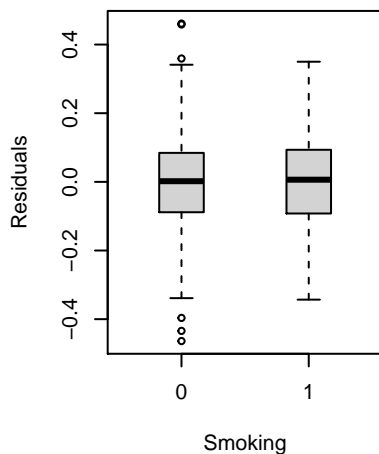


Looks pretty good!

Back to the FEV Data

Residuals vs covariates look good too:

```
> par(mfrow = c(1,4))
> plot(resid_Gamma ~ data$Smoker, xlab = "Smoking", ylab = "Residuals",
+ col = 'lightgray', boxwex = .35)
> plot(resid_Gamma ~ data$Sex, xlab = "Sex", ylab = "Residuals",
+ col = 'lightgray', boxwex = .35)
> plot(data$Age, resid_Gamma, xlab = "Age", ylab = "Residuals",)
> lines(lowess(data$Age, resid_Gamma), lwd = 2, col = "blue")
> abline(h=0, col=2)
> plot(data$Height, resid_Gamma, xlab = "Height", ylab = "Residuals",)
> lines(lowess(data$Height, resid_Gamma), lwd = 2, col = "blue")
> abline(h=0, col=2)
```



Back to the FEV Data

We had the following question: is there any association between FEV and smoking after controlling for other variables in this model?

We can test the null hypothesis

$$H_0 : \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$$

We can do this by comparing FEV_Gamma with a model that does not include the Smoker variable, say FEV_Gamma_red:

```
> FEV_Gamma_red <- glm(  
+   FEV ~ (HeightC+HeightCSq+AgeC)*Sex,  
+   family = Gamma(link="identity"),  
+   data=data)
```

Back to the FEV Data

```
> library(lmtest)
> waldtest(FEV_Gamma_red, FEV_Gamma, test = "Chisq") # Wald test
```

Wald test

```
Model 1: FEV ~ (HeightC + HeightCSq + AgeC) * Sex
Model 2: FEV ~ (HeightC + HeightCSq + AgeC) * (Sex + Smoker)
  Res.Df Df    Chisq Pr(>Chisq)
1     646
2     642  4 3.0416    0.5509
```

```
> anova(FEV_Gamma_red, FEV_Gamma, test = "Rao") # Score test
```

Analysis of Deviance Table

```
Model 1: FEV ~ (HeightC + HeightCSq + AgeC) * Sex
Model 2: FEV ~ (HeightC + HeightCSq + AgeC) * (Sex + Smoker)
  Resid. Df Resid. Dev Df Deviance      Rao Pr(>Chi)
1       646      13.462
2       642      13.404  4 0.058237 0.056386    0.5923
```

Back to the FEV Data

```
> anova(FEV_Gamma_red, FEV_Gamma, test = "LR") # LR test
```

Analysis of Deviance Table

```
Model 1: FEV ~ (HeightC + HeightCSq + AgeC) * Sex
Model 2: FEV ~ (HeightC + HeightCSq + AgeC) * (Sex + Smoker)
```

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	646	13.462			
2	642	13.404	4	0.058237	0.5766

```
> # LR test 'by hand', from the individual model deviances
> (dev_M0vsM1 <- FEV_Gamma_red$deviance - FEV_Gamma$deviance)
```

```
[1] 0.05823693
```

```
> (scaled_dev_M0vsM1 <- dev_M0vsM1/summary(FEV_Gamma)$dispersion)
```

```
[1] 2.889083
```

```
> 1-pchisq(scaled_dev_M0vsM1, 4)
```

```
[1] 0.5765554
```

Therefore the larger model is *not* a better fit, and so we lack evidence that FEV is associated with smoking status after controlling for other variables in this model.

A Count Data Example

We will use the data from Deb and Trivedi (1997), *Demand for Medical Care by the Elderly: A Finite Mixture Approach*, Journal of Applied Econometrics.

Data on 4406 individuals, aged 66 and over, who are covered by Medicare (a public insurance program in the US for people over 65). Taken from the National Medical Expenditure Survey (NMES, 1987 and 1988)

These data were also used by Zeileis (2006) *Object-oriented Computation of Sandwich Estimators*, Journal of Statistical Software, which describes some functionality of the sandwich R package

You can download the data from
<https://www.jstatsoft.org/article/view/v016i09/0>

A Count Data Example

From Deb and Trivedi (1997):

OFP	# of physician office visits
OFNP	# of non-physician office visits
OPP	# of physician hospital outpatient visits
OPNP	# of non-physician hospital outpatient visits
EMR	# of emergency room visits
HOSP	# of hospital stays
EXCLHLTH	= 1 if self-perceived health is excellent
POORHLTH	= 1 if self-perceived health is poor
NUMCHRON	# of chronic conditions (cancer, heart attack, gall bladder problems, emphysema, arthritis, diabetes, other heart disease)
ADLDIFF	= 1 if the person has a condition that limits activities of daily living
NOREAST	= 1 if the person lives in northeastern US
MIDWEST	= 1 if the person lives in the midwestern US
WEST	= 1 if the person lives in the western US
AGE	age in years (divided by 10)
BLACK	= 1 if the person is African American
MALE	= 1 if the person is male
MARRIED	= 1 if the person is married
SCHOOL	# of years of education
FAMINC	family income in \$10000
EMPLOYED	= 1 if the person is employed
PRIVINS	= 1 if the person is covered by private health insurance
MEDICAID	= 1 if the person is covered by Medicaid

Initial Exploratory Analysis

```
> load("DebTrivedi.rda")
> str(DebTrivedi)
```

```
'data.frame':      4406 obs. of  19 variables:
 $ ofp      : int  5 1 13 16 3 17 9 3 1 0 ...
 $ ofnp     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ opp      : int  0 2 0 5 0 0 0 0 0 0 ...
 $ opnp     : int  0 0 0 0 0 0 0 0 0 0 ...
 $ emer     : int  0 2 3 1 0 0 0 0 0 0 ...
 $ hosp     : int  1 0 3 1 0 0 0 0 0 0 ...
 $ health   : Factor w/ 3 levels "average","excellent",...: 1 1 3 3 1 3 1 1 1 1 ...
 $ numchron: int  2 2 4 2 2 5 0 0 0 0 ...
 $ adldiff  : Factor w/ 2 levels "no","yes": 1 1 2 2 2 2 1 1 1 1 ...
 $ region   : Factor w/ 4 levels "midwest","noreast",...: 3 3 3 3 3 3 1 1 1 1 ...
 $ age      : num  6.9 7.4 6.6 7.6 7.9 6.6 7.5 8.7 7.3 7.8 ...
 $ black    : Factor w/ 2 levels "no","yes": 2 1 2 1 1 1 1 1 1 1 ...
 $ gender   : Factor w/ 2 levels "female","male": 2 1 1 2 1 1 1 1 1 1 ...
 $ married  : Factor w/ 2 levels "no","yes": 2 2 1 2 2 1 1 1 1 1 ...
 $ school   : int  6 10 10 3 6 7 8 8 8 8 ...
 $ faminc   : num  2.881 2.748 0.653 0.659 0.659 ...
 $ employed: Factor w/ 2 levels "no","yes": 2 1 1 1 1 1 1 1 1 1 ...
 $ privins  : Factor w/ 2 levels "no","yes": 2 2 1 2 2 1 2 2 2 2 ...
 $ medicaid: Factor w/ 2 levels "no","yes": 1 1 2 1 1 2 1 1 1 1 ...
```

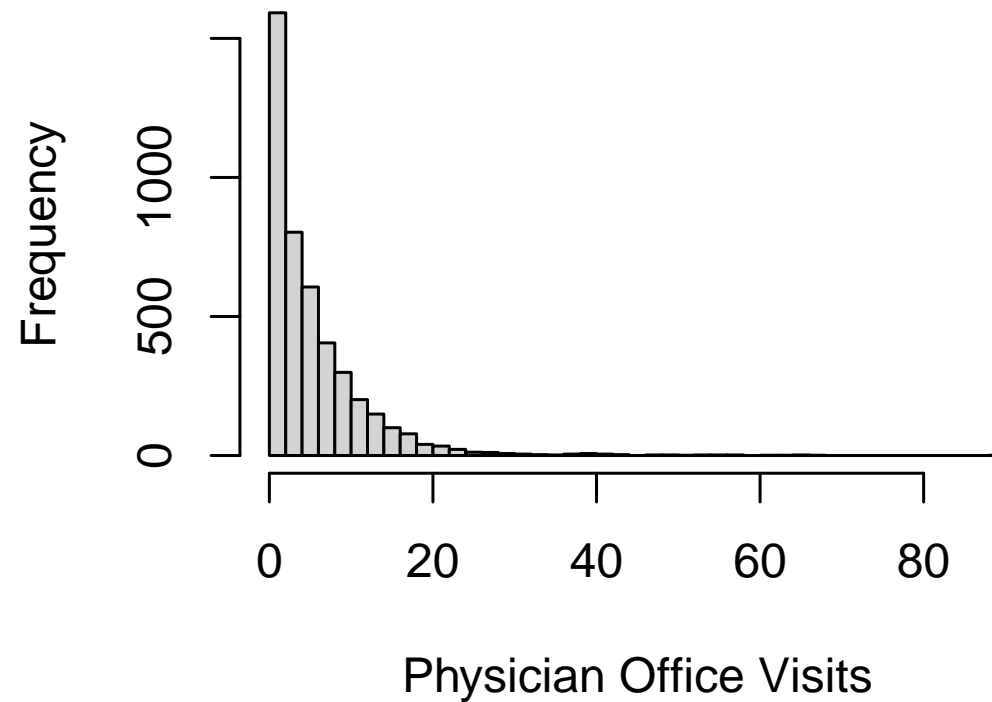
Zeileis (2006) used the model

```
ofp ~ health + age + gender + married + faminc + privins
```

and so will focus on these variables for illustrative purposes

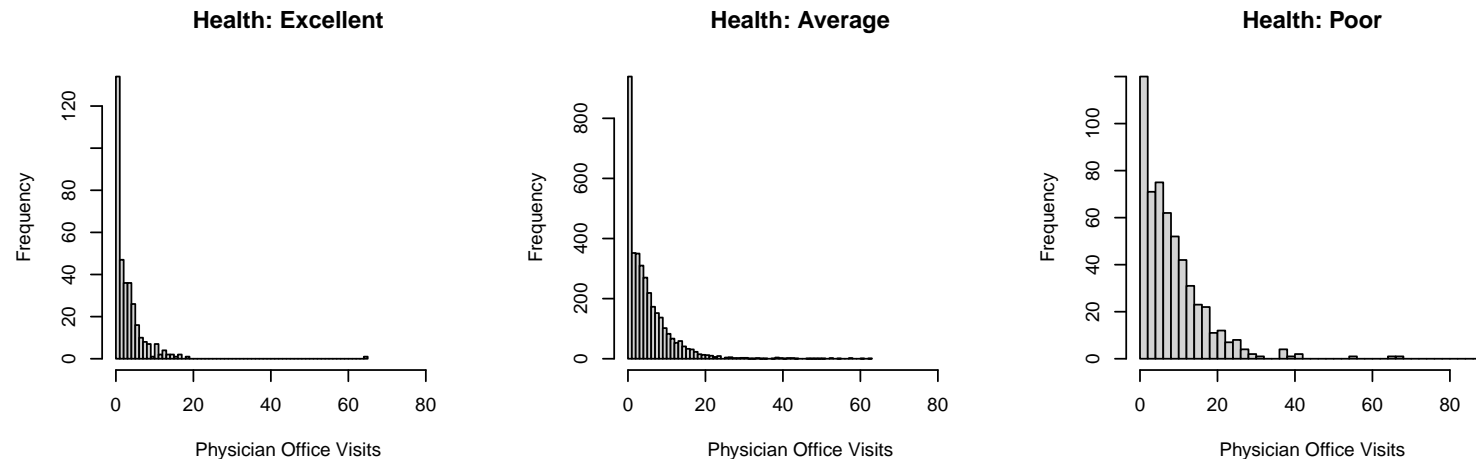
Initial Exploratory Analysis

```
> hist(DebTrivedi$ofp, 50, xlab="Physician Office Visits", main="")
```



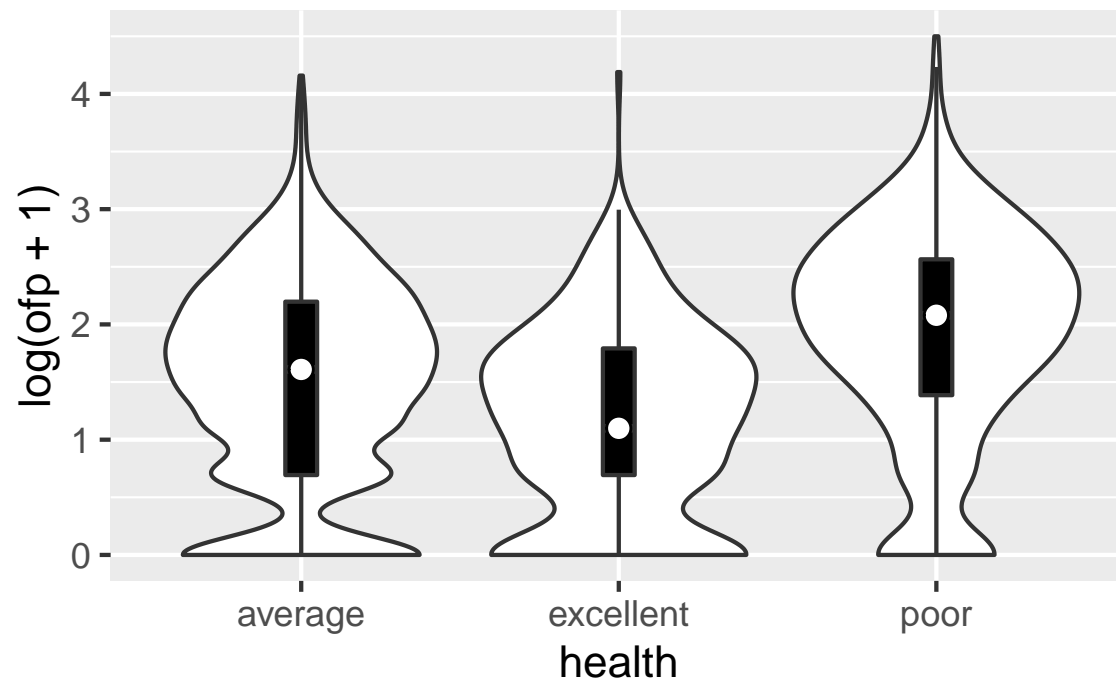
Initial Exploratory Analysis

```
> par(mfrow=c(1,3))
> ofp_range <- range(DebTrivedi$ofp); xlab <- "Physician Office Visits"
> with(DebTrivedi,
+   {hist(ofp[health=="excellent"], 50, xlab=xlab, xlim=ofp_range, main="Health: Excellent")
+     hist(ofp[health=="average"], 50, xlab=xlab, xlim=ofp_range, main="Health: Average")
+     hist(ofp[health=="poor"], 50, xlab=xlab, xlim=ofp_range, main="Health: Poor")
+   })
```



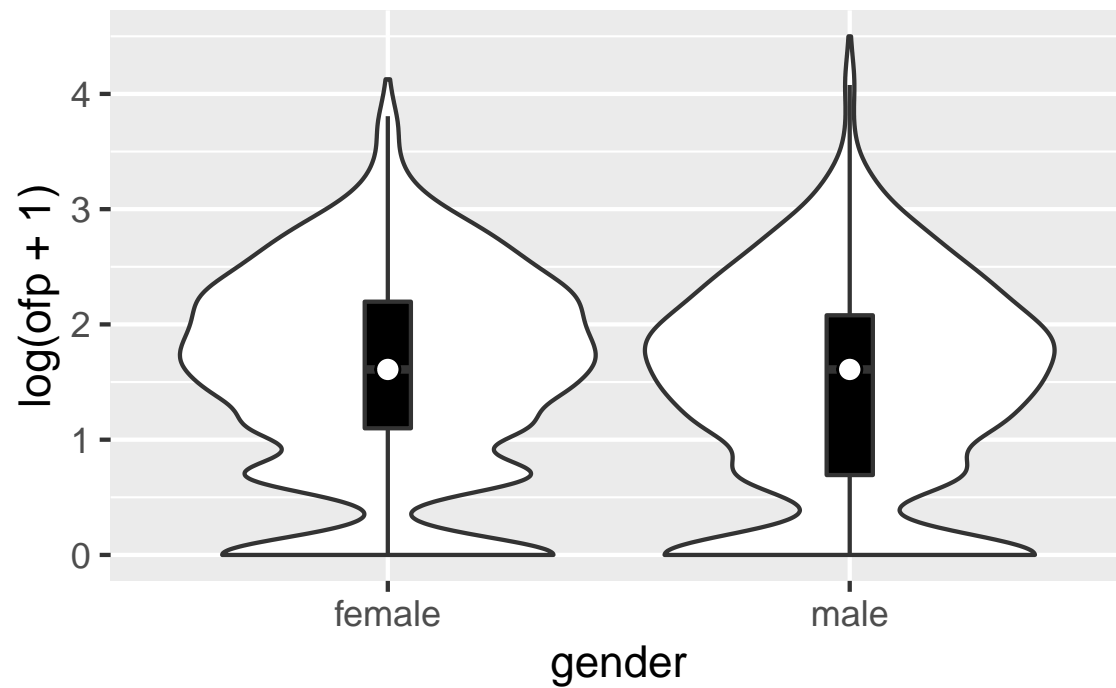
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = health, y = log(ofp+1))) +  
+   geom_violin() +  
+   geom_boxplot(width = .1, fill = "black", outlier.colour = NA) +  
+   stat_summary(fun.y = median, geom = "point", fill = "white", shape = 21, size = 2.5)
```



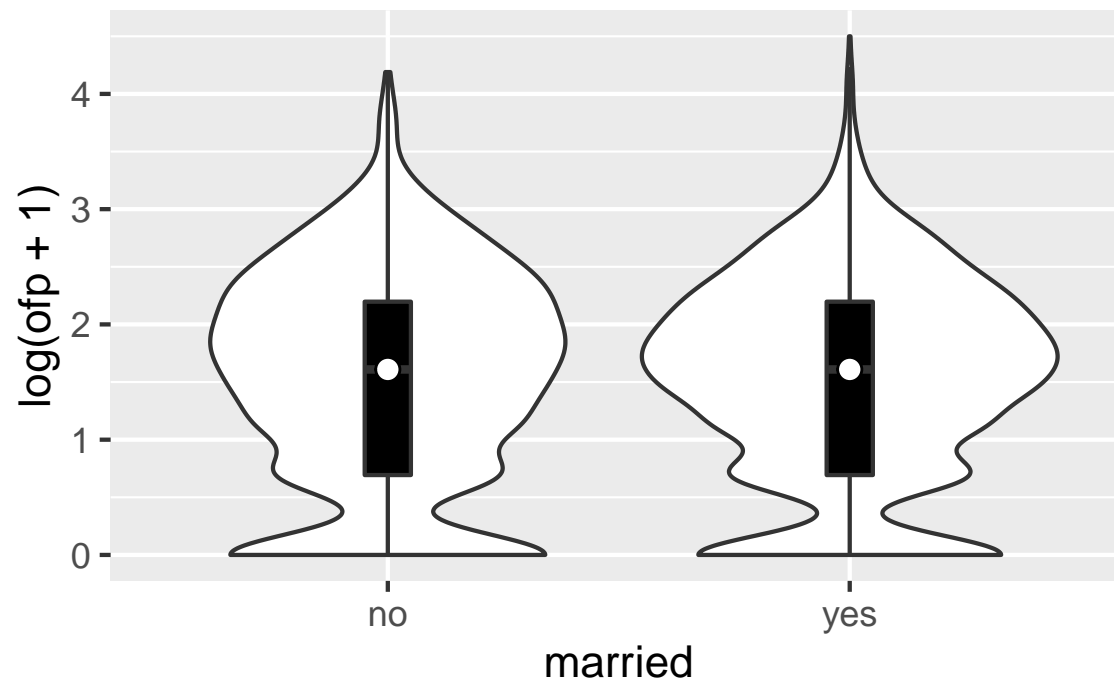
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = gender, y = log(ofp+1))) +  
+   geom_violin() +  
+   geom_boxplot(width = .1, fill = "black", outlier.colour = NA) +  
+   stat_summary(fun.y = median, geom = "point", fill = "white", shape = 21, size = 2.5)
```



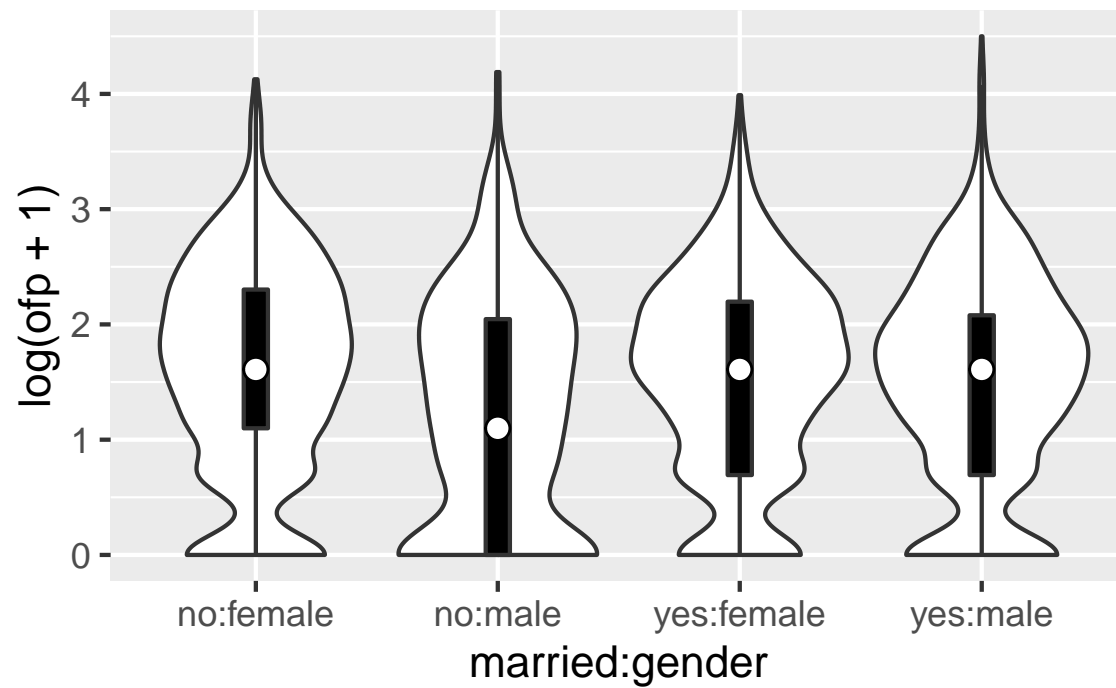
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = married, y = log(ofp+1))) +  
+   geom_violin() +  
+   geom_boxplot(width = .1, fill = "black", outlier.colour = NA) +  
+   stat_summary(fun.y = median, geom = "point", fill = "white", shape = 21, size = 2.5)
```



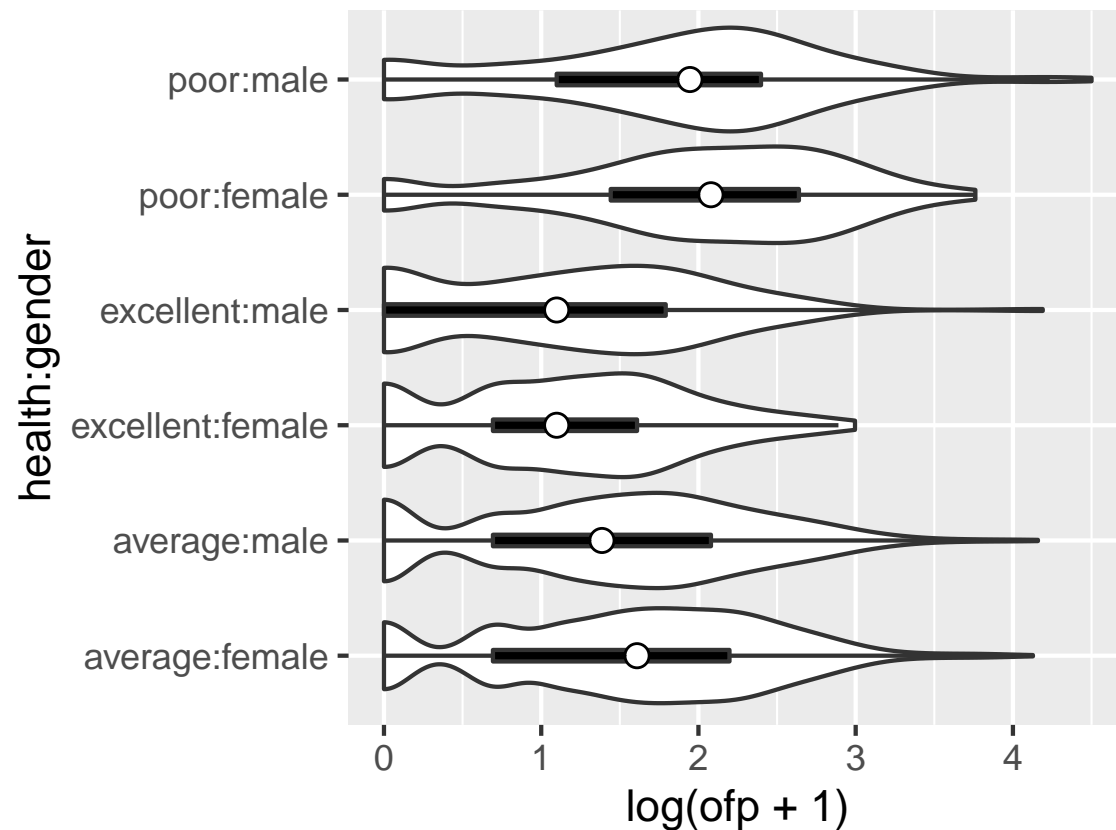
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = married:gender, y = log(ofp+1))) +  
+   geom_violin() +  
+   geom_boxplot(width = .1, fill = "black", outlier.colour = NA) +  
+   stat_summary(fun.y = median, geom = "point", fill = "white", shape = 21, size = 2.5)
```



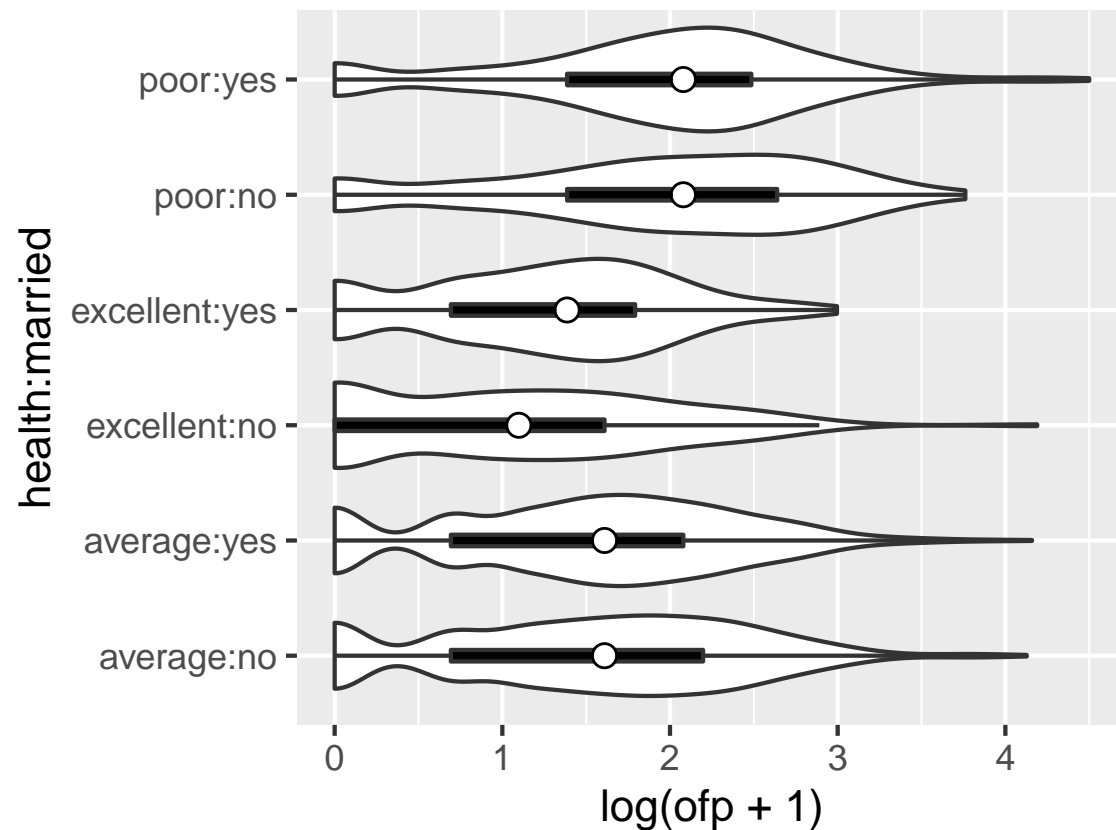
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = health:gender, y = log(ofp+1))) + geom_violin() +  
+   geom_boxplot(width = .1, fill = "black", outlier.colour = NA) +  
+   stat_summary(fun.y = median, geom = "point", fill = "white", shape = 21, size = 2.5) +  
+   coord_flip()
```



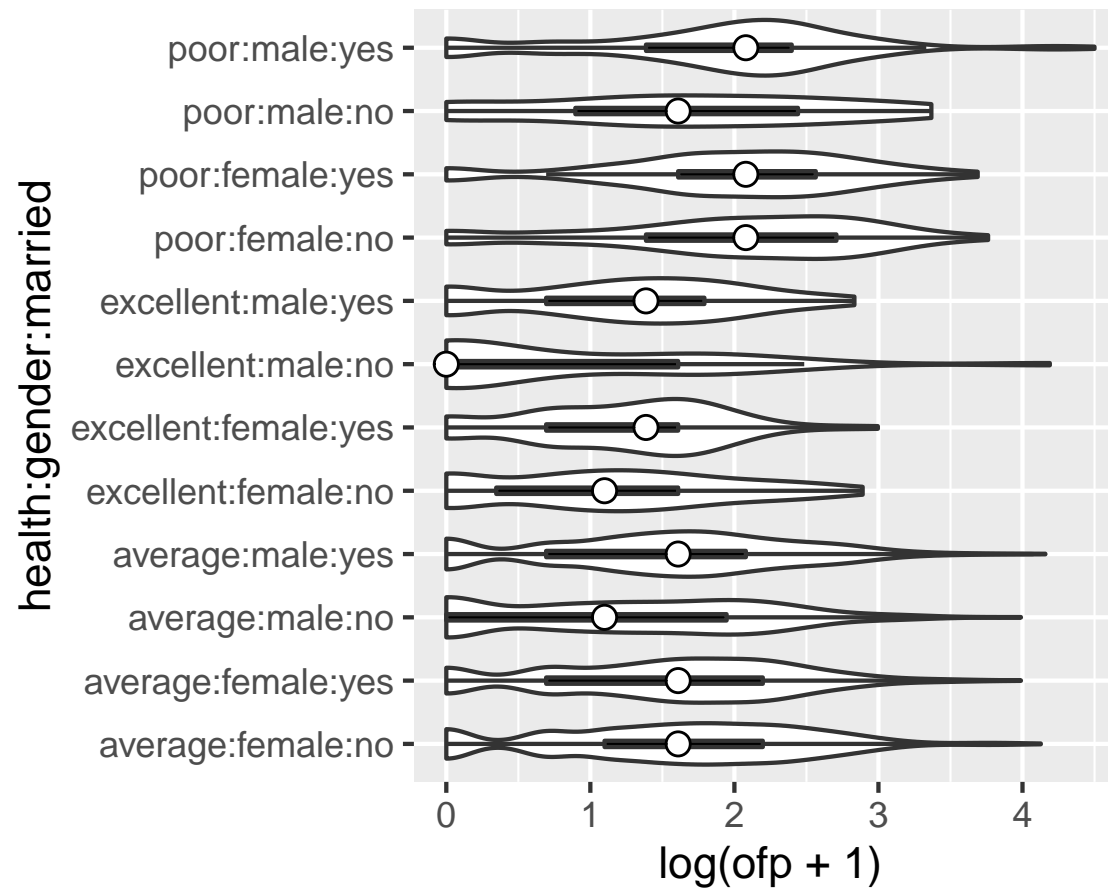
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = health:married, y = log(ofp+1))) + geom_violin() +  
+   geom_boxplot(width = .1, fill = "black", outlier.colour = NA) +  
+   stat_summary(fun.y = median, geom = "point", fill = "white", shape = 21, size = 2.5) +  
+   coord_flip()
```



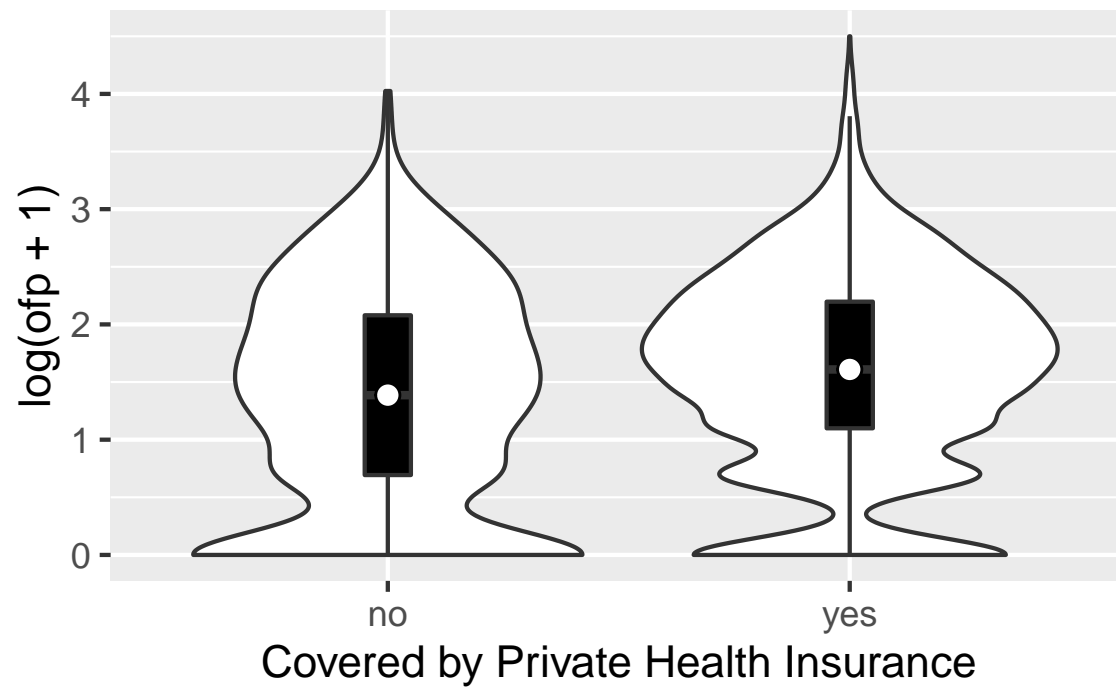
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = health:gender:married, y = log(ofp+1))) + geom_violin() +  
+ geom_boxplot(width = .1, fill = "black", outlier.colour = NA) +  
+ stat_summary(fun.y = median, geom = "point", fill = "white", shape = 21, size = 2.5) +  
+ coord_flip()
```



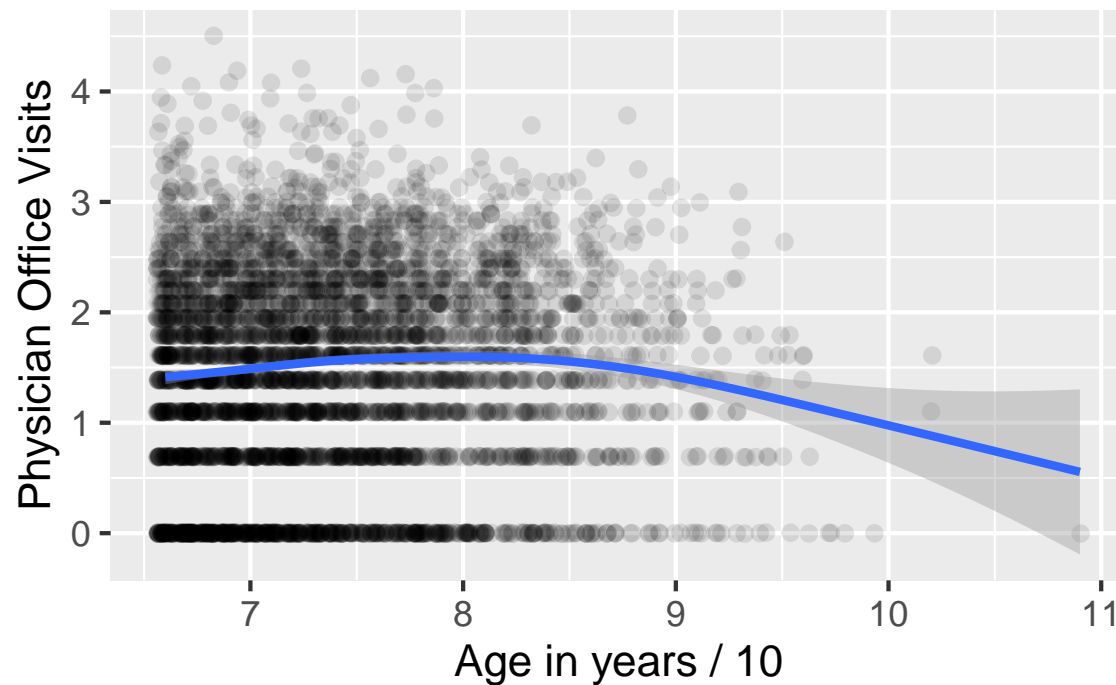
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = privins, y = log(ofp+1))) +  
+   geom_violin() + geom_boxplot(width = .1, fill = "black", outlier.colour = NA) +  
+   stat_summary(fun.y = median, geom = "point", fill = "white", shape = 21, size = 2.5) +  
+   xlab("Covered by Private Health Insurance")
```



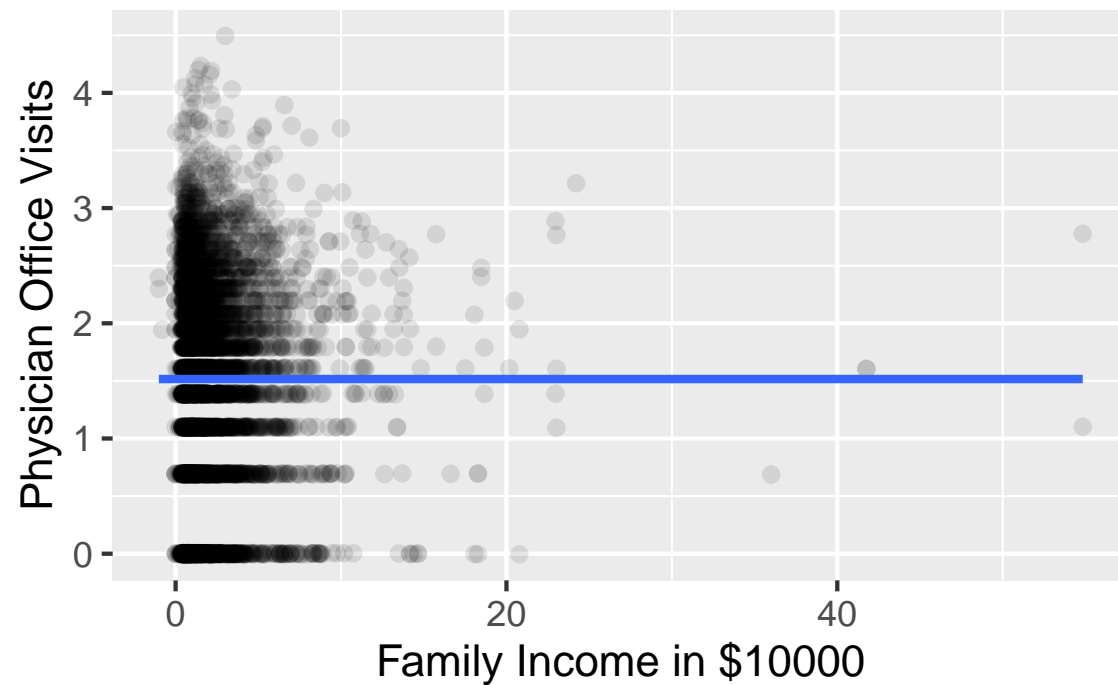
Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = age, y = log(ofp+1))) +  
+   geom_point(alpha = .1, position = "jitter") +  
+   stat_smooth() + ylab("Physician Office Visits") + xlab("Age in years / 10")
```



Initial Exploratory Analysis

```
> ggplot(DebTrivedi, aes(x = faminc, y = log(ofp+1))) +  
+   geom_point(alpha = .1, position = "jitter") +  
+   stat_smooth() + ylab("Physician Office Visits") + xlab("Family Income in $10000")
```



Fitting Generalized Linear Models

We will start with a Poisson GLM.

The following mean model seems reasonable given our previous data exploration

```
> DebTrivedi <- within(DebTrivedi, married <- relevel(married, ref = 2))
> DebTrivedi$age2 <- DebTrivedi$age^2
> glm_pois <- glm(ofp ~ faminc + age + age2 + privins + health*gender*married,
+               data = DebTrivedi, family = poisson(link = "log"))
```

Fitting Generalized Linear Models

```
> library(lmtest)
> coeftest(glm_pois)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-5.56485456	0.77334663	-7.1958	6.209e-13	***
faminc	0.00495037	0.00217587	2.2751	0.022899	*
age	1.87532273	0.20309799	9.2336	< 2.2e-16	***
age2	-0.12526063	0.01326983	-9.4395	< 2.2e-16	***
privinsyes	0.28442945	0.01653981	17.1967	< 2.2e-16	***
healthexcellent	-0.54992640	0.06569215	-8.3713	< 2.2e-16	***
healthpoor	0.52466828	0.03482476	15.0660	< 2.2e-16	***
gendermale	0.01735944	0.01975549	0.8787	0.379556	
marriedno	0.16834455	0.01965372	8.5655	< 2.2e-16	***
healthexcellent:gendermale	0.16488771	0.08306020	1.9852	0.047127	*
healthpoor:gendermale	0.00710958	0.04593409	0.1548	0.876996	
healthexcellent:marriedno	-0.07182619	0.08418571	-0.8532	0.393555	
healthpoor:marriedno	0.00014626	0.04263093	0.0034	0.997263	
gendermale:marriedno	-0.38182992	0.03675448	-10.3887	< 2.2e-16	***
healthexcellent:gendermale:marriedno	0.34843884	0.13282644	2.6233	0.008709	**
healthpoor:gendermale:marriedno	0.07532543	0.07678206	0.9810	0.326578	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

As expected, one of the three-way interactions turned out to be significant

Comparing Models

We now present likelihood-based tests to test a reduced model.

Say we want to test whether marital status modifies the association of health/gender and the number of visits. In this reduced model the effect of marital status is additive:

```
> glm_pois_red <- glm(ofp ~ faminc + age + age2 + privins + married + health*gender,  
+ data = DebTrivedi, family = poisson)  
> library(lmtest)  
> coeftest(glm_pois_red)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-5.6176099	0.7719123	-7.2775	3.400e-13	***
faminc	0.0046488	0.0021769	2.1355	0.03272	*
age	1.8981875	0.2026714	9.3658	< 2.2e-16	***
age2	-0.1261455	0.0132412	-9.5267	< 2.2e-16	***
privinsyes	0.2839862	0.0165118	17.1990	< 2.2e-16	***
marriedno	0.0604635	0.0144932	4.1719	3.021e-05	***
healthexcellent	-0.5937832	0.0410910	-14.4505	< 2.2e-16	***
healthpoor	0.5280521	0.0202545	26.0709	< 2.2e-16	***
gendermale	-0.0992863	0.0158145	-6.2782	3.426e-10	***
healthexcellent:gendermale	0.2613004	0.0603016	4.3332	1.469e-05	***
healthpoor:gendermale	0.0015054	0.0331895	0.0454	0.96382	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Comparing Models

```
> waldtest(glm_pois_red, glm_pois, test = "Chisq") # Wald test
```

Wald test

```
Model 1: ofp ~ faminc + age + age2 + privins + married + health * gender
Model 2: ofp ~ faminc + age + age2 + privins + health * gender * married
```

```
Res.Df Df  Chisq Pr(>Chisq)
1    4395
2    4390  5 128.41  < 2.2e-16 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> anova(glm_pois_red, glm_pois, test = "LR") # LR test
```

Analysis of Deviance Table

```
Model 1: ofp ~ faminc + age + age2 + privins + married + health * gender
Model 2: ofp ~ faminc + age + age2 + privins + health * gender * married
```

```
Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      4395      25200
2      4390      25066  5   134.13 < 2.2e-16 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Comparing Models

```
> anova(glm_pois_red, glm_pois, test = "Rao") # Score test
```

Analysis of Deviance Table

Model 1: ofp ~ faminc + age + age2 + privins + married + health * gender

Model 2: ofp ~ faminc + age + age2 + privins + health * gender * married

	Resid. Df	Resid. Dev	Df	Deviance	Rao	Pr(>Chi)
1	4395	25200				
2	4390	25066	5	134.13	129.1	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

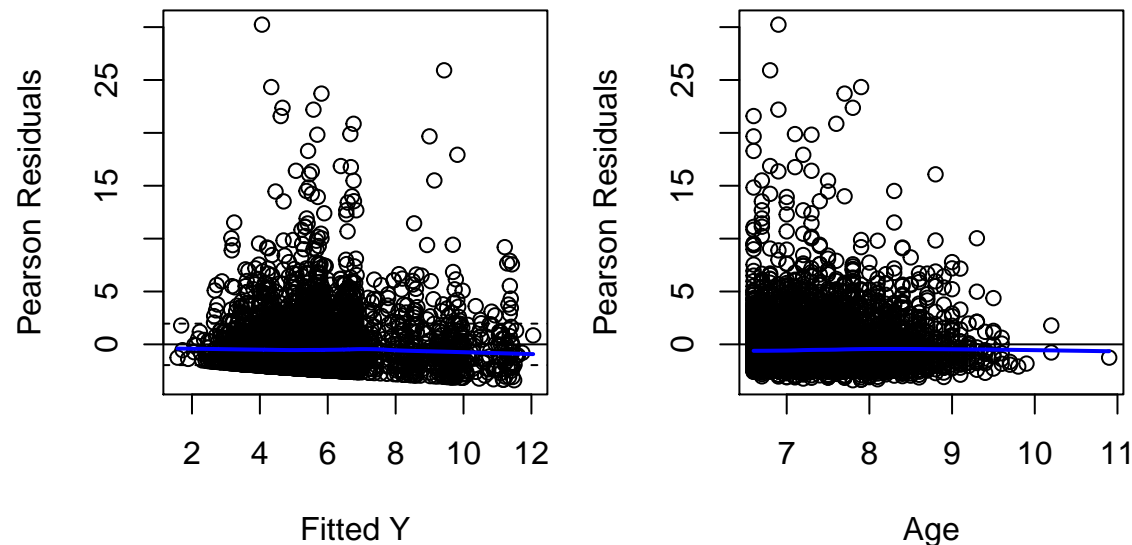
Therefore the larger model is a better fit, which if the model was correct would provide evidence that marital status modifies the association of the some combinations of health/gender and the number of visits.

Note that these tests rely on the Poisson model being correctly specified, in particular its mean/variance relationship

Diagnostics

Something we will want to check are diagnostics, in particular residuals

```
> pearson <- residuals(glm_pois, type = "pearson")
> yhat <- fitted(glm_pois)
> par(mar=c(4,4,1,1))
> par(mfrow = c(1,2))
> plot(yhat, pearson, ylab = "Pearson Residuals", xlab = "Fitted Y")
> lines(lowess(yhat, pearson), lwd = 2, col = "blue")
> abline(h=0); abline(h = 1.96, lty = 2); abline(h = -1.96, lty = 2)
> plot(DebTrivedi$age, pearson, ylab = "Pearson Residuals", xlab = "Age")
> abline(h=0); lines(lowess(DebTrivedi$age, pearson), lwd = 2, col = "blue")
```



If the Poisson model was correct, how should the residuals look like?

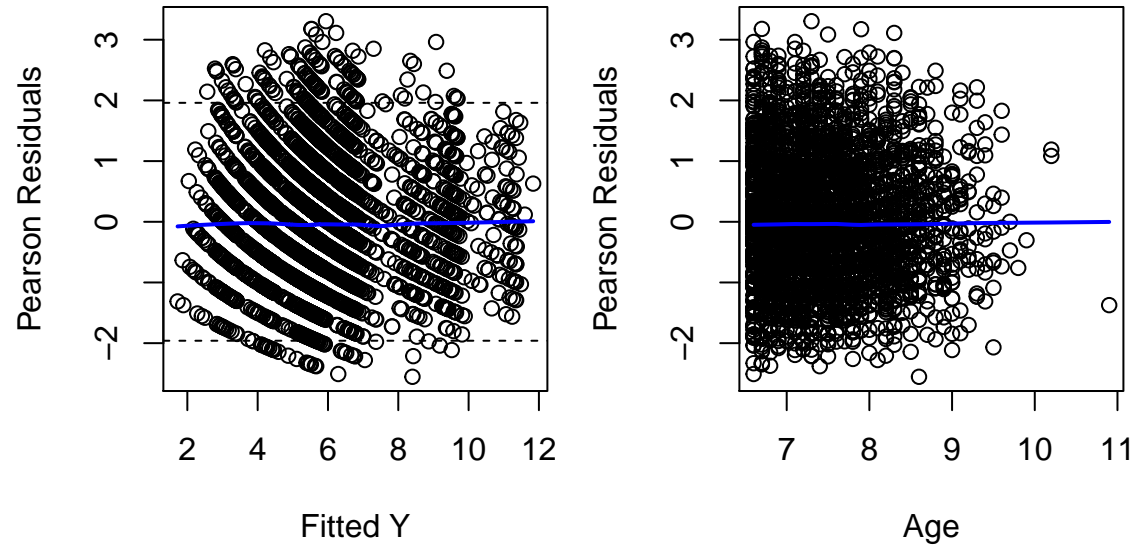
Diagnostics

One way to find out: simulate from a Poisson model as if your estimated model was the true one, fit model with new response and check residuals

```
> set.seed(1221)
> check_resp1 <- rpois(nrow(DebTrivedi),fitted(glm_pois))
> check_glm1 <- glm(check_resp1 ~ faminc + age + age2 + privins + health*gender*married,
+                   data = DebTrivedi, family = poisson)

> check_pearson <- residuals(check_glm1, type = "pearson")
> check_yhat <- fitted(check_glm1)
> par(mar=c(4,4,1,1))
> par(mfrow = c(1,2))
> plot(check_yhat, check_pearson, ylab = "Pearson Residuals", xlab = "Fitted Y")
> lines(lowess(check_yhat, check_pearson), lwd = 2, col = "blue")
> abline(h = 1.96, lty = 2); abline(h = -1.96, lty = 2)
> plot(DebTrivedi$age, check_pearson, ylab = "Pearson Residuals", xlab = "Age")
> lines(lowess(DebTrivedi$age, check_pearson), lwd = 2, col = "blue")
```

Diagnostics



Don't look anything like what we obtained with the actual response variable, so Poisson model doesn't seem OK here