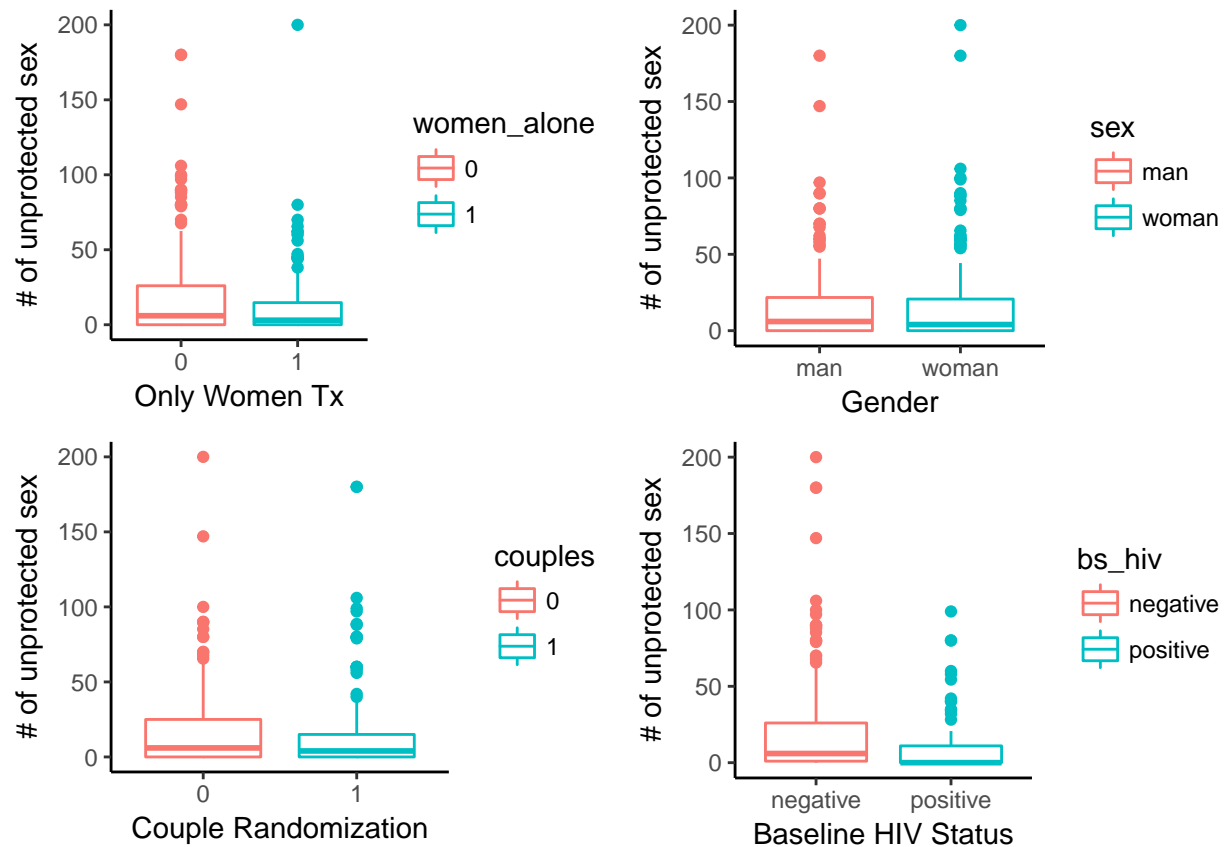


# Homework 4

*Allan Kimaina*

**Question 1: Do Problem 1 in chapter 6 of Gelman and Hill. In this problem, fupacts is the outcome; sex, baseline HIV status and baseline number of sex acts are the pre-treatment variables for question (b). The treatment groups require a bit of manipulation of the variables women\_alone and couples.**

## Data Exploration



A few initial notes to make before we do analysis:

- **Gender Effect (sex)** - In general the number individuals who engaged in unprotected sex acts after counseling sessions is higher in men compared to women. However this could be confounded by several other factors
- **Only Women Effect (women\_alone)** - From the plot we see that the number of individuals who engaged in unprotected sex acts after counseling sessions is higher in couple treatment (man + woman) compared with women only treatment.
- **Couple Effect (couple)** - The plot depicts that the number of individuals who engaged in unprotected sex acts after counseling sessions is higher in group randomized to couples compared with group randomized to non couples.
- **Baseline HIV Status (bs\_hiv)** - The number individuals who engaged in unprotected sex acts after counseling sessions is higher in group with -ve baseline HIV status compared with group with positive baseline HIV status.

## Part A

Model this outcome as a function of treatment assignment using a Poisson regression. Does the model fit well? Is there evidence of overdispersion?

Does the model fit well?

| Model 1        |                 |
|----------------|-----------------|
| (Intercept)    | 2.92 (0.01)***  |
| women_alone1   | -0.41 (0.03)*** |
| AIC            |                 |
| BIC            |                 |
| Log Likelihood |                 |
| Deviance       | 13070.07        |
| Num. obs.      | 434             |

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

Table 1: Statistical models

From this model, the predictor (women\_alone1) is statistically significant; However, the model does not fit well because the the ratio of residual deviance and degree of freedom (30) is much greater than 1. In fact a chi-square test based on the residual deviance and degrees of freedom indicates that we have strong evidence to reject that the null hypothesis that our model correctly fits the data (with a p-value of 0). Perhaps we might be having missing data or missing covariates or higher level of overdispersion.

Is there evidence of overdispersion?

## Overdispersion test

```
## data: riskyBehaviours.fit1
## z = 4.9312, p-value = 4.086e-07
## alternative hypothesis: true alpha is greater than 0
## sample estimates:
##      alpha
## 41.99409
```

Using Overdispersion test by Cameron & Trivedi (1990), We have a strong & significant evidence of overdispersion 42 (p-value = 4.086e-07). However, this problem of overdispersion might be confounded with the problem of omitted covariates or missingness. Given that we have only one predictor in our model, perhaps this is the problem, lets evaluate this notion in part (b) of this question.

## Part B

Extend the model to include pre-treatment measures of the outcome and the additional pre-treatment variables included in the dataset. Does the model fit well? Is there evidence of overdispersion??

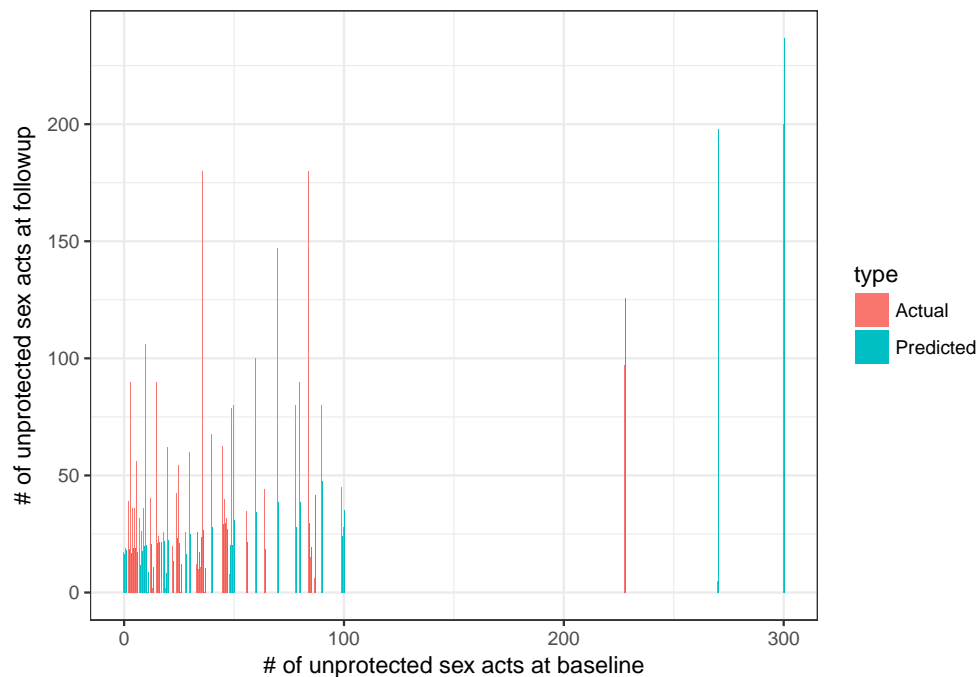
| Model 2: Poisson |                 |
|------------------|-----------------|
| (Intercept)      | 3.07 (0.02)***  |
| sexwoman         | 0.11 (0.02)***  |
| couples1         | -0.41 (0.03)*** |
| women_alone1     | -0.66 (0.03)*** |
| cBupacts         | 0.69 (0.01)***  |
| bs_hivpositive   | -0.44 (0.04)*** |
| AIC              |                 |
| BIC              |                 |
| Log Likelihood   |                 |
| Deviance         | 10204.40        |
| Num. obs.        | 434             |

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

Table 2: Statistical models

### Does the model fit well?

Still the model does not fit well, however the model is much better than the previous one. The rule of thumb dictates that the ratio of residual deviance to degree of freedom should be close to 1, however we get 23 which is still way far from 1. Furthermore, the chi-square test based on the residual deviance and degrees of freedom has a p-value of 0.



From the above plot we see that this model is performing poorly in prediction, i.e we have huge residue especially when predicted counts increases. Perhaps we still have omitted predictors in our model or missing value that would have accounted for the extra variance.

Is there evidence of overdispersion??

```
## Overdispersion test
##
## data:  riskyBehaviours.fit2
## z = 5.5727, p-value = 1.254e-08
## alternative hypothesis: true alpha is greater than 0
## sample estimates:
##    alpha
## 28.65966
```

We still have evidence of overdispersion, However not as pronounced as in the previous model. Using Overdispersion test by Cameron & Trivedi (1990), we get an overdispersion value of 28.66 which is much less compared to the previous model. It is important to note that this value is still huge. Fitting overdispersed Poisson model may remedy the problems of overdispersion, lets check this notion in the next question

## Part C

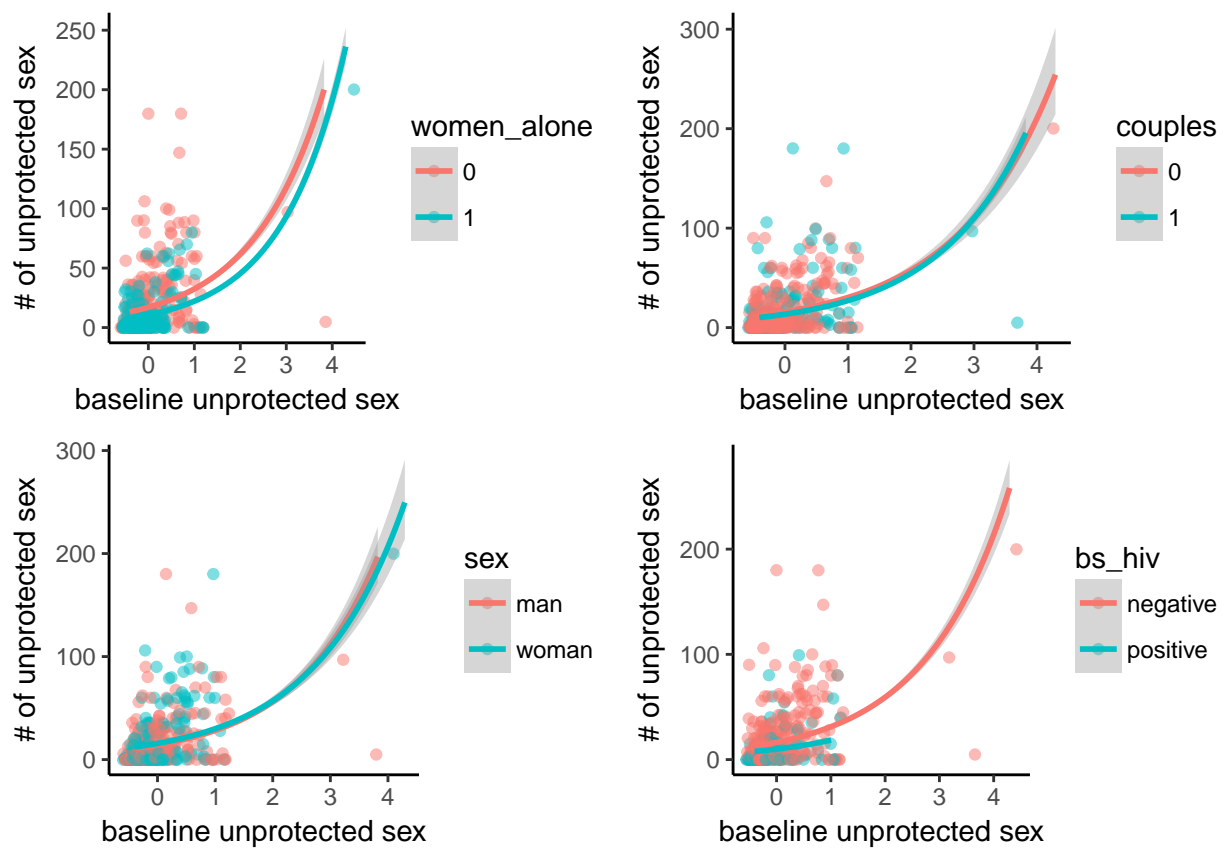
Fit an overdispersed Poisson model. What do you conclude regarding effectiveness of the intervention?

| Model 3: QuasiPoisson |                 |
|-----------------------|-----------------|
| (Intercept)           | 3.07 (0.12)***  |
| women_alone1          | -0.66 (0.17)*** |
| sexwoman              | 0.11 (0.13)     |
| cBupacts              | 0.69 (0.06)***  |
| couples1              | -0.41 (0.15)**  |
| bs_hivpositive        | -0.44 (0.19)*   |
| AIC                   |                 |
| BIC                   |                 |
| Log Likelihood        |                 |
| Deviance              | 10204.40        |
| Num. obs.             | 434             |

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

Table 3: Statistical models

What do you conclude regarding effectiveness of the intervention?



In general, the intervention was effective in the sense that it decreased the number of unprotected sex acts at followup. Evaluating each significant predictor, we see that:

- **Only Women Intervention Effects (women\_only1)** - Holding the other predictors constant at the same level, the number of individuals who engaged in unprotected sex acts after counseling in intervention group where only the woman participated was reduced by  $(1 - \exp(-0.66)) \times 100 = 48\%$ . Notice from the plot that there is a wide distance between the 2 regression curves of women\_alone(1) compared to women\_alone(0).
- **Couple Effects (couples1)** - Holding the other predictors constant at the same level, the number of individuals who engaged in unprotected sex acts after counseling in intervention group where both partners participated was reduced by  $(1 - \exp(-0.41)) \times 100 = 34\%$ .
- **Baseline HIV Status (bs\_hivpositive)** - Individuals having a positive baseline HIV status realized a 36% reduction effect in number of unprotected sex acts after counseling intervention (holding the other predictors constant).
- **Baseline number of unprotected** - As expected, the number of unprotected sex at baseline was directly proportional to number of unprotected sex acts after counseling intervention (holding the other predictors constant).
- **Sex - Gender** was statistically insignificant. Additionally, from the plot we see that the 2 regression curves completely overlap.

In summary the intervention realized greater reduction effects in number of unprotected sex when “only the woman” participated in the counselling session. However, we see that this reduction effect was less when both partners participated in counselling session. This is quite odd, perhaps because of the differences in the baseline as explained in the data explanatory section of this analysis.

**Q1.D: These data include responses from both men and women from the participating couples. Does this give you any concern with regard to our modeling assumptions?**

Yes, the responses of participants who are part of the same couple wouldn't be independent because we expect their response to be highly correlated in the sense that partners tend to imitate each other's responses as opposed to when interviewed individually. This clearly violates the independence assumption of observation i.e. one observation should not affect or provide any information on another observation.

**Question 2: Show that the Poisson model is a member of the exponential family and derive its canonical parameter.**

Let  $y \sim \text{Poisson}(\lambda)$

$\Rightarrow$  By definition the PDF of Poisson:

$$f(y) = e^{-\lambda} \cdot \frac{\lambda^y}{y!}$$

$\Rightarrow$  We know that the exponential family:

$$f(y; \theta) = h(y, \phi) \exp \left\{ \frac{T(y) c(\theta) - b(\theta)}{a(\phi)} \right\} \dots (i)$$

$\Rightarrow y \sim \text{poisson}(\lambda)$  can be expressed as: (i)

$$f(y) = \exp \{ y \log(\lambda) - \lambda - \log(y!) \}$$

$$f(y) = \frac{1}{y!} \exp \left\{ \frac{y \log(\lambda) - \lambda}{1} \right\} \dots (ii)$$

$\Rightarrow$  Observe that:

$$f(y) = \underbrace{\frac{1}{y!}}_{h(y; \phi)} \exp \left\{ \underbrace{\frac{T(y) c(\theta) - b(\theta)}{a(\phi)}}_{\frac{y \log(\lambda) - \lambda}{1}} \right\} \Rightarrow \frac{\phi}{y!} \exp \left\{ \frac{y \theta - e^\theta}{\phi} \right\}$$

Where  $\theta = \log(\lambda)$   
 $\phi = 1$

$\Rightarrow$  Therefore:

$$h(y; \phi) = \frac{1}{y!} \Rightarrow \frac{\phi}{y!}$$

$$T(y) = y \Rightarrow y$$

$$c(\theta) = \log(\lambda) \Rightarrow \theta$$

$$a(\phi) = 1 \Rightarrow \phi$$

$$b(\theta) = \lambda \Rightarrow e^\theta$$

From above we find that canonical parameter:  $\theta = \log \lambda$

Figure 1:



**Question 3: Write an iteratively reweighted least squares algorithm (see slides 8 and 9 of the generalized linear models lecture) to fit Poisson regression. Test it out on problem 1. How close do you get to the right answer?**

Collaborated with: Julia and Isaac

```
poissonIrls <- function(response,predictors,df){
  # Convert factors into binary (0 or 1)
  for (i in 1:ncol(df)){
    if (is.factor(df[,i])){
      df[,i] = as.numeric(df[,i]) - 1
    }
  }

  betas = rep(1,length(predictors)+1) # init coefficients
  y = df[,response] # extract response
  x = matrix(NA,nrow(df),length(predictors)) # extract predictors

  for (i in 1:ncol(x)){
    x[,i] = df[,predictors[i]]
  }
  X = cbind(1,x)

  # IRLS =  $B(t+1) = B(t) + J(-1)B(t)u(B(t))$ 
  b_old = rep(0,5)
  diff = 1
  # Iterate until  $|B(t+1) - B(t)| < e$ , where  $e = 0.05$ 
  while (diff>0.05){
    yhat = betas[1] #  $Y = b_0 + b_1x_1 + b_2x_2 + \dots$ 
    for (j in 1:ncol(x)){
      yhat = yhat + betas[j+1]*x[,j]
    }

    # Creates  $B(t+1)$ 
    b_new = as.matrix(yhat + ((y - exp(yhat)) / exp(yhat)))
    W = diag(exp(yhat))

    # solves beta coefficients
    betas = solve(t(X)%*%
                  W%*%
                  X)%*%
            t(X)%*%
            W%*%
            b_new

    # Computes  $|B(t+1) - B(t)|$ 
    diff = abs(sum(b_old)-sum(b_new))
    b_old = b_new
  }
  betas = t(betas)
```

```

colnames(betas) = c("Intercept",predictors) # add col names
return(betas) # return coefficients
}

```

Test and Compare

```

# inbuilt GLM
glm(fupacts ~ sex + couples + women_alone + cBupacts , data=riskyBehaviours.df, family = poisson(link="log"))

##      (Intercept)      sexwoman      couples1 women_alone1      cBupacts
##    3.02291411    0.09728765   -0.47384461   -0.69358123    0.71275149

# Poisson IRLS
poissonIrls(response="fupacts", predictors = c("sex","couples","women_alone","cBupacts"), df=riskyBehaviours.df)

##      Intercept      sex      couples women_alone cBupacts
## [1,]  3.022914  0.09728765 -0.4738446  -0.6935812  0.7127515

```

How close do you get to the right answer?

Really close

## Source Code

```
#' ---  
# title: "Homework 4"  
# author: "Allan Kimaina"  
# header-includes:  
# - \usepackage{pdflscape}  
# - \newcommand{\blandscape}{\begin{landscape}}  
# - \newcommand{\elandscape}{\end{landscape}}  
# output:  
# pdf_document: default  
# html_document: default  
# ---  
#  
#  
#knitr::opts_chunk$set(echo = F)  
  
# loadl lm package  
library(dplyr)  
library(car)  
library(sjPlot)  
library(sjmisc)  
library(sjlabelled)  
library(ggpubr)  
library(ggpmisc)  
library(gridExtra)  
library(stargazer)  
library(e1071)  
library(jtools)  
library(effects)  
library(multcompView)  
library(ggplot2)  
library(ggrepel)  
library(MASS)  
library(broom)  
library(ggcorrplot)  
library(leaps)  
library(relaimpo)  
library(olsrr)  
  
# load GLM packages  
library(ROCR)  
library(arm)  
library(foreign)  
library(nnet)  
library(VGAM)  
library(ordinal)  
library(ModelGood)  
library(InformationValue)  
library(rms)  
library(texreg) # for latex table  
library(AER) # for overdispersion
```

```

# load required data
riskyBehaviours.df=read.csv("data/risky_behaviors.csv")
# Data wrangling
riskyBehaviours.df$couples <- factor(riskyBehaviours.df$couples)
riskyBehaviours.df$women_alone <- factor(riskyBehaviours.df$women_alone)
riskyBehaviours.df$cBupacts <- (riskyBehaviours.df$bupacts - mean(riskyBehaviours.df$bupacts)) / (2 * s
#hist(riskyBehaviours.df$fupacts)

#'
#' \onecolumn
#'
#' # Question 1: Do Problem 1 in chapter 6 of Gelman and Hill. In this problem, fupacts is the outcome;
#'
#' ## Data Exploration
#'

de1<- ggplot(riskyBehaviours.df) +
  geom_boxplot(aes(women_alone, fupacts, col = women_alone)) +
  theme_classic()+ guides(fill=FALSE)+
  labs(y = "# of unprotected sex", x="Only Women Tx")

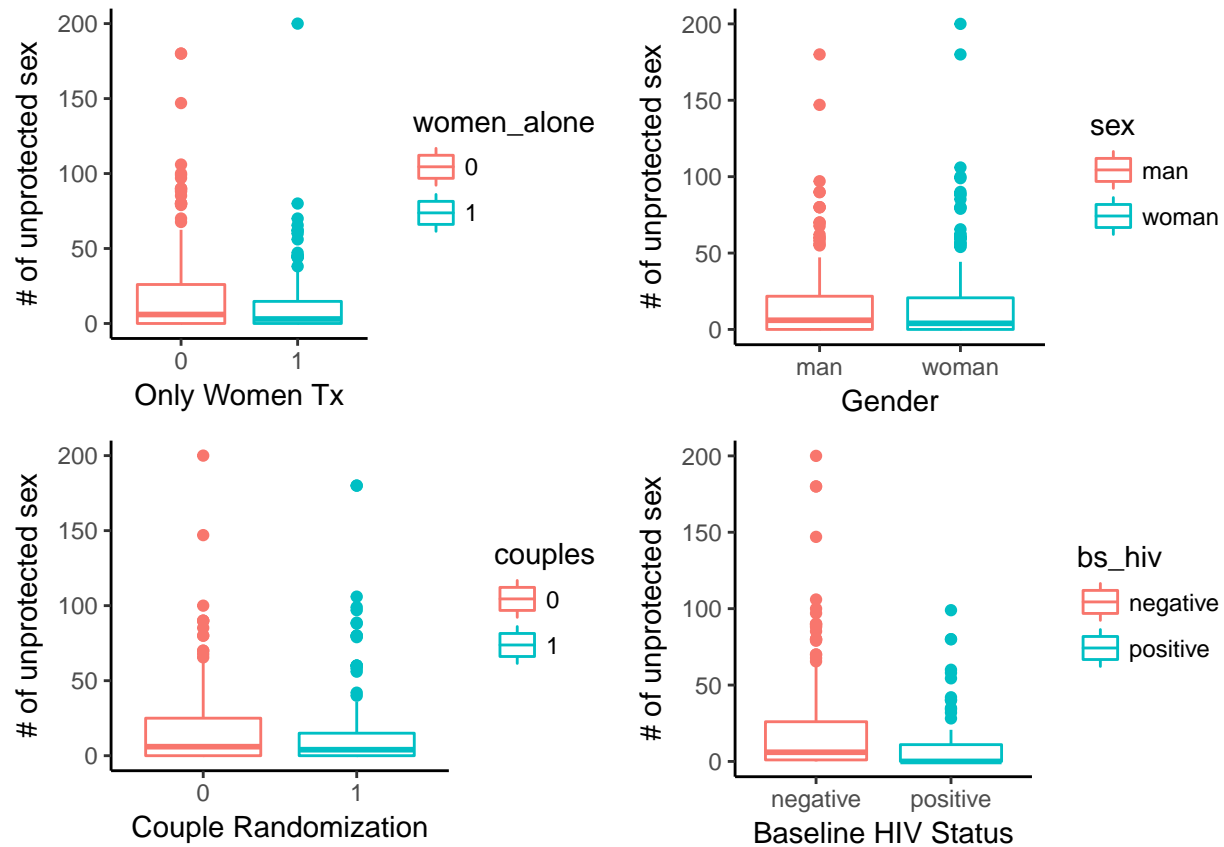
de2<- ggplot(riskyBehaviours.df) +
  geom_boxplot(aes(sex, fupacts, col = sex)) +
  theme_classic()+ guides(fill=FALSE)+
  labs(y = "# of unprotected sex", x="Gender")

de3<- ggplot(riskyBehaviours.df) +
  geom_boxplot(aes(couples, fupacts, col = couples)) +
  theme_classic()+ guides(fill=FALSE)+
  labs(y = "# of unprotected sex", x="Couple Randomization")

de4<- ggplot(riskyBehaviours.df) +
  geom_boxplot(aes(bs_hiv, fupacts, col = bs_hiv)) +
  theme_classic()+ guides(fill=FALSE)+
  labs(y = "# of unprotected sex", x="Baseline HIV Status")

grid.arrange(de1, de2, de3, de4,
  nrow = 2.)

```



```

#
# A few initial notes to make before we do analysis:
#
# * `Gender Effect (sex)` - In general the number individuals who engaged in unprotected sex acts af
# * `Only Women Effect (women_alone)` - From the plot we see that the number of individuals who enga
# * `Couple Effect (couple)` - The plot depicts that the number of individuals who engaged in unprot
# * `Baseline HIV Status (bs_hiv)` - The number individuals who engaged in unprotected sex acts after
#
# \onecolumn
#
## Part A
# Model this outcome as a function of treatment assignment using a Poisson regression. Does the model
#
#
# ### `Does the model fit well?`
#
riskyBehaviours.fit1 <- glm(fupacts ~ women_alone, family=poisson, data=riskyBehaviours.df)

#summary(riskyBehaviours.fit1)
texreg(list(riskyBehaviours.fit1), single.row=TRUE, float.pos = "h")

##
## \begin{table}[h]

```

```

## \begin{center}
## \begin{tabular}{l c }
## \hline
## & Model 1 \\
## \hline
## (Intercept) & $2.92 \; ; \; (0.01)^{***}$ \\
## women\_alone1 & $-0.41 \; ; \; (0.03)^{***}$ \\
## \hline
## AIC & \\
## BIC & \\
## Log Likelihood & \\
## Deviance & 13070.07 \\
## Num. obs. & 434 \\
## \hline
## \multicolumn{2}{l}{\scriptsize{$^{***}$p<0.001$, $^{**}$p<0.01$, $^*$p<0.05$}}
## \end{tabular}
## \caption{Statistical models}
## \label{table:coefficients}
## \end{center}
## \end{table}

## calculate and store predicted values

#'
#'
#'
riskyBehaviours.fit1$deviance/riskyBehaviours.fit1$df.residual

## [1] 30.25479

pchisq(riskyBehaviours.fit1$deviance, df=riskyBehaviours.fit1$df.residual, lower.tail=FALSE)

## [1] 0

#'
#'
#' From this model, the predictor (women_alone1) is statistically significant; However, the model does not
#'
#'
#'
#' ### `Is there evidence of overdispersion?`
#'
#'
#'
# Using Overdispersion test by Cameron & Trivedi (1990)
# http://www.sciencedirect.com/science/article/pii/S030440769090014K
# Poisson distribution assume variance is equal to the mean. H0 : c=0 and H1 c !=0
disp <- dispersiontest(riskyBehaviours.fit1,trafo=1)
#'
#'
#' ...
#' ## Overdispersion test
#'
#' ## data: riskyBehaviours.fit1

```

```

#' ## z = 4.9312, p-value = 4.086e-07
#' ## alternative hypothesis: true alpha is greater than 0
#' ## sample estimates:
#' ##      alpha
#' ## 41.99409
#' ```
#'
#' Using Overdispersion test by Cameron & Trivedi (1990), We have a strong & significant evidence of overdispersion
#'
#' \onecolumn
#'
#' ## Part B
#' Extend the model to include pre-treatment measures of the outcome and the additional pre-treatment variables
#'
#'
#'
riskyBehaviours.fit2 <- glm(fupacts ~ sex + couples + women_alone + cBupacts + bs_hiv, family=poisson, data=riskyBehaviours)

#summary(riskyBehaviours.fit2)
texreg(list(riskyBehaviours.fit2),
  custom.model.names = c("Model 2: Poisson"),
  single.row=TRUE, float.pos = "h")

##
## \begin{table}[h]
## \begin{center}
## \begin{tabular}{l c }
## \hline
## & Model 2: Poisson \\
## \hline
## (Intercept) & $3.07 \backslash; (0.02)^{\{***\}}$ \\
## sexwoman & $0.11 \backslash; (0.02)^{\{***\}}$ \\
## couples1 & $-0.41 \backslash; (0.03)^{\{***\}}$ \\
## women\_alone1 & $-0.66 \backslash; (0.03)^{\{***\}}$ \\
## cBupacts & $0.69 \backslash; (0.01)^{\{***\}}$ \\
## bs\_hivpositive & $-0.44 \backslash; (0.04)^{\{***\}}$ \\
## \hline
## AIC & \\
## BIC & \\
## Log Likelihood & \\
## Deviance & 10204.40 \\
## Num. obs. & 434 \\
## \hline
## \multicolumn{2}{l}{\scriptsize{$^{\{***\}}p<0.001$, $^{\{**\}}p<0.01$, $^{\{*\}}p<0.05$}}
## \end{tabular}
## \caption{Statistical models}
## \label{table:coefficients}
## \end{center}
## \end{table}

#'
#'
```

```

#' ### `Does the model fit well?`
#'
#'

riskyBehaviours.df$fit <- predict(riskyBehaviours.fit2, type="response")
df <- riskyBehaviours.df[with(riskyBehaviours.df, order(women_alone, bupacts)), ]
plot.df <- ggplot(df, aes(x = bupacts, y = fit, colour = women_alone)) +
  geom_point(aes(y = fupacts), alpha=.5, position=position_jitter(h=.2)) +
  geom_smooth(method = "glm",
              method.args = list(family = "poisson"))+
  labs(y = "# of unprotected sex acts at followup", x="# of unprotected sex acts at baseline")+
  theme_classic()+ guides(fill=FALSE)

#'
#'
#'
riskyBehaviours.fit2$deviance/riskyBehaviours.fit2$df.residual

## [1] 23.84207

pchisq(riskyBehaviours.fit2$deviance, df=riskyBehaviours.fit2$df.residual, lower.tail=FALSE)

## [1] 0

#'
#' Still the model does not fit well, however the model is much better than the previous one. The rule
#'
#'
#'
df3 <- riskyBehaviours.df
df3$type = factor("Actual")

df4 <- riskyBehaviours.df
df4$type = factor("Predicted")
df4$fupacts <- riskyBehaviours.df$fit

df5 <- rbind(df3, df4)

ggplot(df5, aes(x=bupacts, y=fupacts, fill=type)) +
  geom_bar(stat="identity", position=position_dodge())+
  xlab("# of unprotected sex acts at baseline") + ylab("# of unprotected sex acts at followup") +
  theme_bw()

```





```

#' ## alternative hypothesis: true alpha is greater than 0
#' ## sample estimates:
#' ##      alpha
#' ## 28.65966
#' ``
#'
#' We still have evidence of Overdispersion, However not as pronounced as in the previous model. Using
#'
#' \onecolumn
#'
#' ## Part C
#' Fit an overdispersed Poisson model. What do you conclude regarding effectiveness of the intervention
#'
#'
#'
'
```

```

riskyBehaviours.fit3 <- glm(fupacts ~ women_alone + sex + cBupacts + couples + bs_hiv, family=quasipois)

texreg(list(riskyBehaviours.fit3),
  custom.model.names = c("Model 3: QuasiPoisson"),
  single.row=TRUE, float.pos = "h")

```

```

##
## \begin{table}[h]
## \begin{center}
## \begin{tabular}{l c }
## \hline
## & Model 3: QuasiPoisson \\
## \hline
## (Intercept)      & $3.07 \; \backslash; (0.12)^{\{***\}}$ \\\
## women\_alone1    & $-0.66 \; \backslash; (0.17)^{\{***\}}$ \\\
## sexwoman         & $0.11 \; \backslash; (0.13)$ \\\
## cBupacts         & $0.69 \; \backslash; (0.06)^{\{***\}}$ \\\
## couples1         & $-0.41 \; \backslash; (0.15)^{\{**\}}$ \\\
## bs\_hivpositive  & $-0.44 \; \backslash; (0.19)^{\{*\}}$ \\\
## \hline
## AIC              & \\\
## BIC              & \\\
## Log Likelihood   & \\\
## Deviance         & 10204.40 \\\
## Num. obs.        & 434 \\\
## \hline
## \multicolumn{2}{l}{\scriptsize{$^{\{***\}}p<0.001$, $^{\{**\}}p<0.01$, $^{\{*\}}p<0.05$}}
## \end{tabular}
## \caption{Statistical models}
## \label{table:coefficients}
## \end{center}
## \end{table}

```

```

#display(riskyBehaviours.fit2)
#'
'
```

```

#'
#'
#'
#'
#' ### `What do you conclude regarding effectiveness of the intervention?`
#'
#'

riskyBehaviours.df$fit <- predict(riskyBehaviours.fit3, type="response")
df <- riskyBehaviours.df[with(riskyBehaviours.df, order(women_alone, bupacts)), ]

plot1 <- ggplot(df, aes(x = cBupacts, y = fit, colour =women_alone )) +
  geom_point(aes(y = fupacts), alpha=.5, position=position_jitter(h=.2,w=.2)) +
  geom_smooth(method = "glm",
              method.args = list(family = "quasipoisson"))+
  labs(y = "# of unprotected sex", x="baseline unprotected sex")+
  theme_classic()+ guides(fill=FALSE)

plot2 <- ggplot(df, aes(x = cBupacts, y = fit, colour =couples )) +
  geom_point(aes(y = fupacts), alpha=.5, position=position_jitter(h=.2,w=.2)) +
  geom_smooth(method = "glm",
              method.args = list(family = "quasipoisson"))+
  labs(y = "# of unprotected sex", x="baseline unprotected sex")+
  theme_classic()+ guides(fill=FALSE)

plot3 <- ggplot(df, aes(x = cBupacts, y = fit, colour =sex )) +
  geom_point(aes(y = fupacts), alpha=.5, position=position_jitter(h=.2,w=.2)) +
  geom_smooth(method = "glm",
              method.args = list(family = "quasipoisson"))+
  labs(y = "# of unprotected sex", x="baseline unprotected sex")+
  theme_classic()+ guides(fill=FALSE)

plot4 <- ggplot(df, aes(x = cBupacts, y = fit, colour =bs_hiv )) +
  geom_point(aes(y = fupacts), alpha=.5, position=position_jitter(h=.2,w=.2)) +
  geom_smooth(method = "glm",
              method.args = list(family = "quasipoisson"))+
  labs(y = "# of unprotected sex", x="baseline unprotected sex")+
  theme_classic()+ guides(fill=FALSE)

grid.arrange(plot1, plot2, plot3, plot4,
              nrow = 2.)

```



```

#' \onecolumn
#'
#' 
#'
#' \onecolumn
#'
#' # Question 3: Write an iteratively reweighted least squares algorithm (see slides 8 and 9 of the gen
#'
#'
#' Collaborated with: Julia and Isaac
#'
#'

poissonIrls <- function(response,predictors,df){
  # Convert factors into binary (0 or 1)
  for (i in 1:ncol(df)){
    if (is.factor(df[,i])){
      df[,i] = as.numeric(df[,i]) - 1
    }
  }

  betas = rep(1,length(predictors)+1) # init coefficients
  y = df[,response] # extract response
  x = matrix(NA,nrow(df),length(predictors)) # extract predictors

  for (i in 1:ncol(x)){
    x[,i] = df[,predictors[i]]
  }
  X = cbind(1,x)

  # IRLS =  $B(t+1) = B(t) + J(-1)B(t)u(B(t))$ 
  b_old = rep(0,5)
  diff = 1
  # Iterate until  $|B(t+1) - B(t)| < e$ , where  $e = 0.05$ 
  while (diff>0.05){
    yhat = betas[1] #  $Y = b_0 + b_1x_1 + b_2x_2 + \dots$ 
    for (j in 1:ncol(x)){
      yhat = yhat + betas[j+1]*x[,j]
    }

    # Creates  $B(t+1)$ 
    b_new = as.matrix(yhat + ((y - exp(yhat)) / exp(yhat)))
    W = diag(exp(yhat))

    # solves beta coefficients
    betas = solve(t(X)%*%
                  W)%*%
                  X)%*%
                  t(X)%*%
                  W)%*%
                  b_new
  }
}

```

```

    # Computes | B(t+1) - B(t) |
    diff = abs(sum(b_old)-sum(b_new))
    b_old = b_new
  }
  betas = t(betas)
  colnames(betas) = c("Intercept",predictors) # add col names
  return(betas) # return coefficients
}

#'
#'
#' ### `Test and Compare`
#'
#'
#'
# inbuilt GLM
glm(fupacts ~ sex + couples + women_alone + cBupacts , data=riskyBehaviours.df, family = poisson(link="log"))

##      (Intercept)      sexwoman      couples1 women_alone1      cBupacts
##      3.02291411      0.09728765     -0.47384461     -0.69358123      0.71275149

# Poisson IRLS
poissonIrls(response="fupacts", predictors = c("sex","couples","women_alone","cBupacts"), df=riskyBehaviours.df)

##      Intercept      sex      couples women_alone cBupacts
## [1,]  3.022914  0.09728765 -0.4738446  -0.6935812  0.7127515

#'
#'
#'
#' ### `How close do you get to the right answer?`
#'
#' Really close
#'
#' \onecolumn
#'
#' # Source Code
#'
#'
#'
```