PHP 2514 HW 4

Blain Morin

April 9, 2018

Question 1

1a.)

We start by regressing follow-up acts on the treatment groups:

```
###Poisson regression with women only

attach(risky)

reg.1a = glm(fupacts ~ couples + women_alone, family = poisson)

detach(risky)

sjt.glm(reg.1a)
```

	fupacts		
	IRR	*CI*	*p*
(Intercept)	21.97	21.16 – 22.80	<.001
couples (1)	0.72	0.69 – 0.76	<.001
women_alone (1)	0.56	0.53 – 0.60	<.001
Observations		434	

The regression results above display the incident rate ratios for the treatment groups. Compared to the control group, couples that recieved conseling had .72 times the risk of having unprotected sex. Moreover, when women alone recieved counseling, couples had .56 the risk of having unprotected sex compared to the control group. Both groups were significant at the .05 level.

Next, we check the model fit.

First, we examine the residual plot:
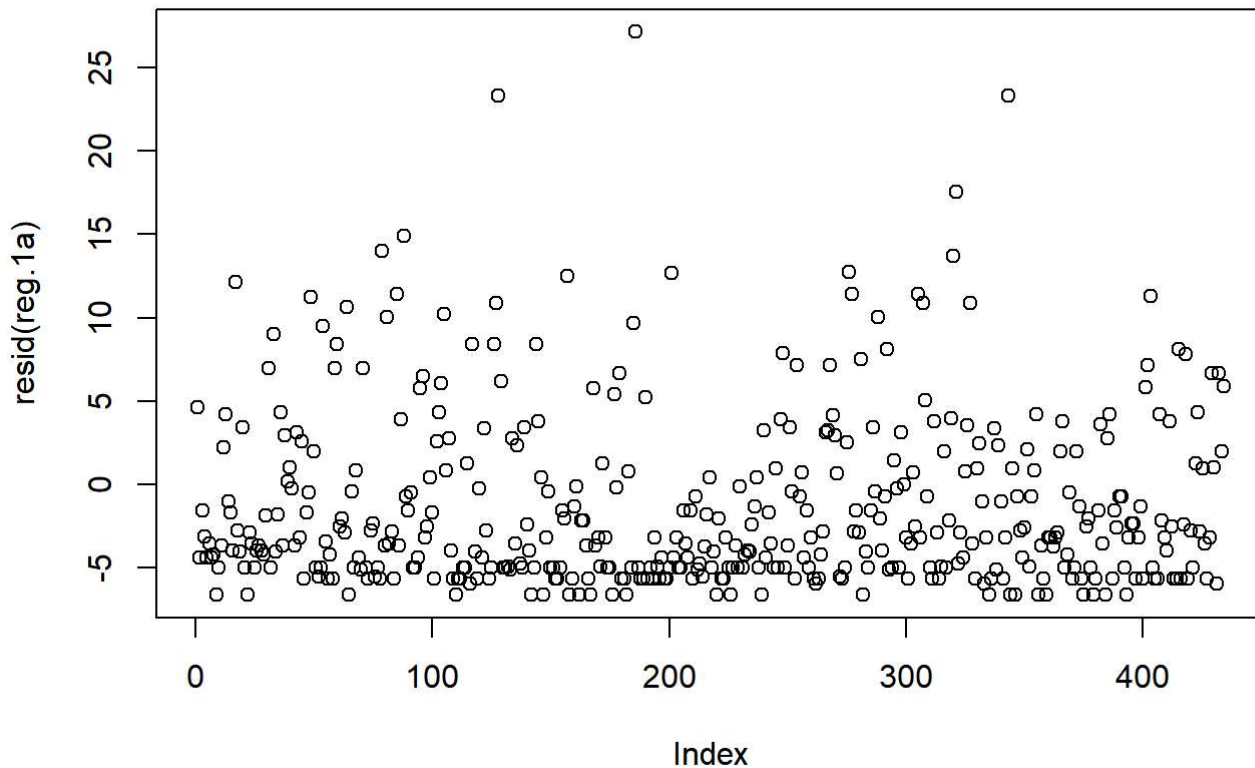
We see that the points are not randomly scattered about the zero line: there are many points above the positive 5 and very few points below negative 5. This is evidence that our model does not fit. We can further check the fit by checking the chi squared statistic of the deviance:

```
pchisq(reg.1a$deviance, df=reg.1a$df.residual, lower.tail=FALSE)
```

```
## [1] 0
```

Due to rounding, we get a zero probability. This means that there is evidence for overdispersion.

1b.)

Next, we run the model including sex, initial HIV status, and initial unprotected sex acts:

```
attach(risky)

reg.1b = glm(fupacts ~ couples + women_alone + sex + bs_hiv + bupacts, family = poisson)

detach(risky)

sjt.glm(reg.1b)
```

	fupacts		
	IRR	*CI*	*p*
(Intercept)	16.23	15.50 – 17.00	<.001
couples (1)	0.66	0.63 – 0.70	<.001
women_alone (1)	0.52	0.49 – 0.55	<.001
sex (woman)	1.11	1.06 – 1.17	<.001
bs_hiv (positive)	0.65	0.60 – 0.69	<.001
bupacts	1.01	1.01 – 1.01	<.001
Observations	434		

The above regression results show that all variables are significant. The interpretations for the factor variables are similar to part 1a. The interpretation for the continuous bupacts variable: for every 1 initial unprotected sex act, the relative risk of having unprotected sex increases by 1%.

We also check the fit of this regression.

First, we look at the residuals:
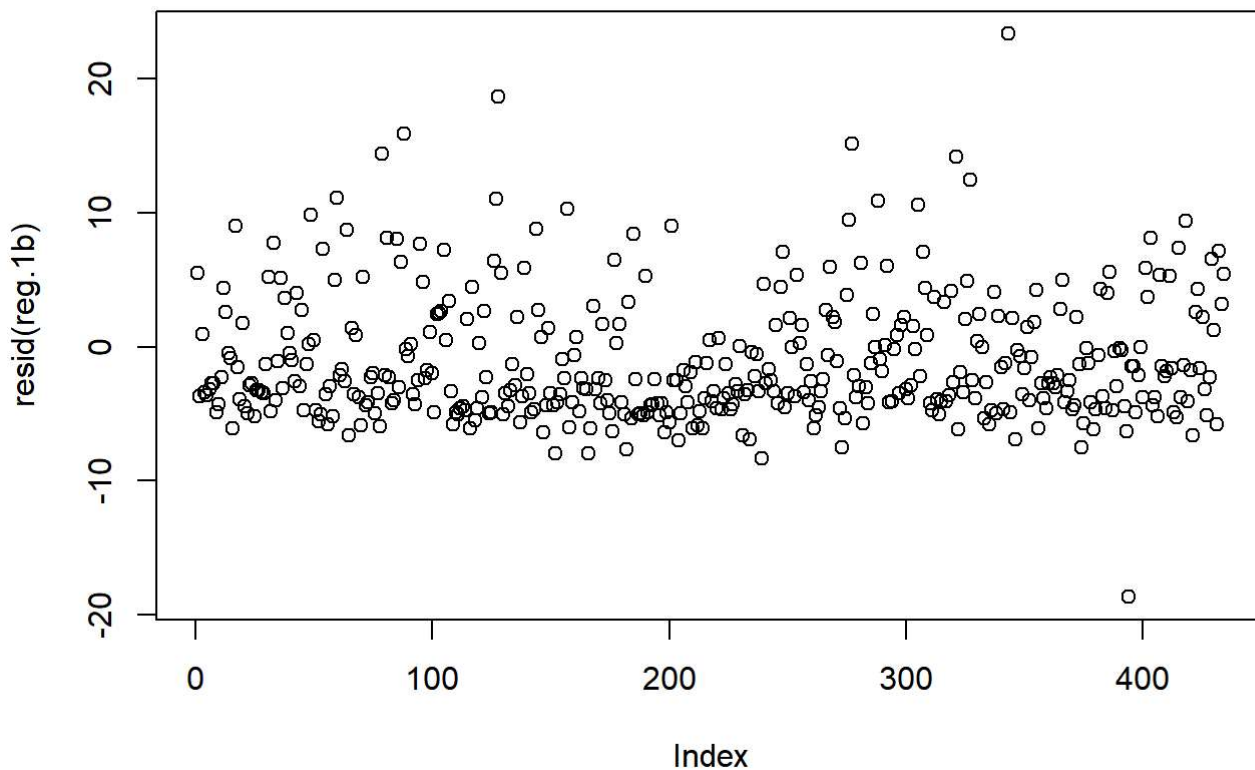
We see again that the residuals are not randomly scattered about the zero line. The spread of the positive residuals is higher than the spread of the negative ones.

We also check the fit using the chi squared statistic of the deviance:

```
pchisq(reg.1b$deviance, df=reg.1b$df.residual, lower.tail=FALSE)
```

```
## [1] 0
```

Again, this probability is zero due to rounding. This means that there is evidence for overdispersion.

1c.)

Since our first two models have evidence of overdispersion, we run the model using a quasi poisson:

```
attach(risky)

reg.1c = glm(fupacts ~ couples + women_alone + sex + bs_hiv + bupacts, family = quasipois
son)

detach(risky)

sjt.glm(reg.1c)
```

	fupacts		
	IRR	*CI*	*p*
(Intercept)	16.23	12.52 – 20.77	<.001
couples (1)	0.66	0.49 – 0.90	.008
women_alone (1)	0.52	0.37 – 0.72	<.001
sex (woman)	1.11	0.86 – 1.44	.404
bs_hiv (positive)	0.65	0.43 – 0.93	.024
bupacts	1.01	1.01 – 1.01	<.001
Observations		434	

We see that the confidence intervals are wider in the quasipoisson model. However, the betas for the treatment groups (couples and women alone) maintain significance. Therefore, there is evidence to suggest that the treatment decreases the risk of unprotected sex acts.

d.)

Since the data includes responses from both members of the couple, the observations are thus not independent. This violates our model assumptions.

Question 2

Poisson Distribution:

pmf: $\dfrac{\lambda^x \, e^{-\lambda}}{x!}$ $\qquad$ $\lambda > 0$, $x = 0, 1, 2 \dots$

Exponential Family:

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

$$\log\big(f(x)\big) = x \log \lambda - \lambda - \log(x!)$$

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

$$f(x) = \underbrace{\exp\{ -\log(x!) \}}_{\substack{= \\ \frac{1}{x!}}} \cdot \exp\left\{ \frac{x \log \lambda - \lambda}{1} \right\}$$

Thus for Poisson: $h(y, \phi) = \dfrac{1}{x!}$ $\qquad$ $\theta = \log \lambda$, $\lambda = e^{\theta}$

$$T(y) = x$$

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

$$b(\theta) = \lambda = e^{\theta}$$

$$a(\phi) = 1$$

$$E(X) = \frac{d}{d\theta} b(\theta) = \frac{d}{d\theta} e^{\theta} = e^{\theta} = \lambda$$

$$Var(X) = a(\phi) \frac{d^2}{d\theta^2} b(\theta) = 1 \frac{d}{d\theta} e^{\theta} = e^{\theta} = \lambda$$

Canonical Link

$$g(\mu) = \theta$$

$$\lambda = e^{\theta}$$

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

Thus, log is the canonical link

Question 3

```r
library(MASS)
library(dplyr)


### Function to calculate beta
itls = function(x, y, tolerance = .0000000000001) {

  ### Initiate beta vector
  beta = rep(1, ncol(x))

  ### Initiate delta
  delta = 100

  ### Iterate until beta doesn't change

  while (abs(delta) > tolerance) {

    eta = x %*% beta
    mu =  exp(eta)
    theta = log(mu)
    v = exp(theta)
    w = ginv(v *((1/mu)^2))
    z = eta + (y-mu) * (1/mu)
    z = z[1]
    new.beta = ginv(t(x) %*% diag(as.vector(w)) %*% x) %*% t(x) %*% diag(as.vector(w)) %*% as.matrix(z)


    delta = sum(new.beta - beta)
    beta = new.beta

  }

  beta

}


### Convert our data to a model matrix
### Predictor Matrix
x.risky = model.matrix(reg.1a)


### Outcome Matrix
y.risky = risky [,7]

### Put our data into the function
result = as.data.frame(cbind(our_function = itls(x.risky, y.risky), coef_from_part1 = coe
```

```
f(reg.1a)))
names(result)[1] = "coef_from_our_function"
```

Create result table
```
kable(result)
```

	coef_from_our_function	coef_from_part1
(Intercept)	3.0895984	3.0895984
couples1	-0.3224256	-0.3224256
women_alone1	-0.5721229	-0.5721229

Our function returns the same results as the glm call in part 1.