

# HW05

Zach White

November 6, 2017

## 1 Marginal and mutual independence.

Two balanced coins are flipped independently. Let  $X$  = whether the first flip resulted in a head (yes, no),  $Y$  = whether the second flip resulted in a head, and  $Z$  = whether both flips had the same result. Using the example, show that (marginal) independence of each pair of three variables does not imply that variables are mutually independent.

### 1.1 Solution

To show that the marginal pairwise independence of each of the random variables does not imply mutual independence, we note that the sample space  $S = \{HH, HT, TH, TT\}$ . We will show that  $Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y)$ ,  $Pr(X = x, Z = z) = Pr(X = x)Pr(Z = z)$ , and  $Pr(Y = y, Z = z) = Pr(Y = y)Pr(Z = z)$ , but we will show that  $Pr(X = x, Y = y, Z = z) \neq Pr(X = x)Pr(Y = y)Pr(Z = z)$  necessarily.

First, it is clear that  $\forall x \in X, y \in Y, Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y) = 0.25$ .

$$\begin{aligned} Pr(X = x) &= Pr(Y = y) = 0.5 \\ Pr(X = \text{head}, Y = \text{head}) &= Pr(HH) = 0.25 = Pr(X = \text{head})Pr(Y = \text{head}) \\ Pr(X = \text{head}, Y = \text{tail}) &= Pr(HT) = 0.25 = Pr(X = \text{head})Pr(Y = \text{tail}) \\ Pr(X = \text{tail}, Y = \text{head}) &= Pr(TH) = 0.25 = Pr(X = \text{tail})Pr(Y = \text{head}) \\ Pr(X = \text{tail}, Y = \text{tail}) &= Pr(TT) = 0.25 = Pr(X = \text{tail})Pr(Y = \text{tail}) \end{aligned}$$

Second, we show that  $\forall x \in X, z \in Z, Pr(X = x, Z = z) = Pr(X = x)Pr(Z = z)$ . To begin, note that  $Pr(X = x) = 0.5$  and  $Pr(Z = z) = 0.5$

$$\begin{aligned} Pr(X = \text{head}, Z = \text{TRUE}) &= Pr(HH) = 0.25 = Pr(X = \text{head})Pr(Z = \text{TRUE}) \\ Pr(X = \text{head}, Z = \text{FALSE}) &= Pr(HT) = 0.25 = Pr(X = \text{head})Pr(Z = \text{FALSE}) \\ Pr(X = \text{tail}, Z = \text{FALSE}) &= Pr(TH) = 0.25 = Pr(X = \text{tail})Pr(Z = \text{FALSE}) \\ Pr(X = \text{tail}, Z = \text{TRUE}) &= Pr(TT) = 0.25 = Pr(X = \text{tail})Pr(Z = \text{TRUE}) \end{aligned}$$

Third, we show that  $\forall y \in Y, z \in Z, Pr(Y = y, Z = z) = Pr(Y = y)Pr(Z = z)$ . Note that  $Pr(Y = y) = 0.5$  and  $Pr(Z = z) = 0.5$

$$\begin{aligned} Pr(Y = \text{head}, Z = \text{TRUE}) &= Pr(HH) = 0.25 = Pr(Y = \text{head})Pr(Z = \text{TRUE}) \\ Pr(Y = \text{tail}, Z = \text{FALSE}) &= Pr(HT) = 0.25 = Pr(Y = \text{head})Pr(Z = \text{FALSE}) \\ Pr(Y = \text{head}, Z = \text{FALSE}) &= Pr(TH) = 0.25 = Pr(Y = \text{tail})Pr(Z = \text{FALSE}) \\ Pr(Y = \text{tail}, Z = \text{TRUE}) &= Pr(TT) = 0.25 = Pr(Y = \text{tail})Pr(Z = \text{TRUE}) \end{aligned}$$

Thus,  $X$ ,  $Y$ , and  $Z$  are all marginally pairwise independent. However, It is clear to see that  $Pr(X = \text{head}, Y = \text{tails}, Z = \text{FALSE}) = 0 \neq (\frac{1}{2})(\frac{1}{2})(\frac{1}{2}) = \frac{1}{8}$ . Thus these are not mutually independent.

## 2 Interpreting interactions

In a study of low birth weight (defined as an infant weighing less than 2500mg), researchers are interested in the roles of maternal age (in years) and adequacy of prenatal care in predicting the outcome. They consider three variables: the outcome of low birth weight ( $LBW=1$  if yes and  $0=no$ ), age (centered at the sample mean of 23 years so that 0 represents 23 years, -1 represents 22 years, 5 represents 28 years, etc.), and care ( $1=adequate$  prenatal care,  $0=inadequate$  care). The investigators fit the following model to the data:

$$\text{logit}(Pr(LBW_i = 1)) = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{care}_i + \beta_3 \text{age}_i \text{care}_i$$

and obtained the estimates

$$\hat{\beta}' = (-0.52, 0.04, -0.47, -0.18)'$$

and

$$\hat{\text{Cov}}(\hat{\beta}) = \begin{bmatrix} 0.046 & 0.002 & -0.046 & -0.002 \\ 0.002 & 0.002 & -0.002 & -0.002 \\ -0.046 & -0.002 & 0.111 & 0.004 \\ -0.002 & -0.002 & 0.004 & 0.005 \end{bmatrix}$$

Prepare an appropriate graphical display to illustrate the associations (point and interval estimates of odds ratios) between adequacy of care, age, and low birth weight at each combination of levels of adequacy of care and maternal age (assume age takes integer values from 17-36). In addition, provide specific interpretations of each parameter estimate (including the intercept) in this model. Include reproducible code for your graphical display.

### 2.1 Solution

I show three different plots. The first two plots show the  $\text{logit}(Pr(LBW\_i = 1))$  and  $Pr(LBW\_i = 1)$  by age and care. The different colors represent whether or not an individual received adequate care or not. Both of these plots are also accompanied by a 95% confidence interval that we found using the  $\hat{\text{Cov}}(\hat{\beta})$  matrix, specifically, we used the variances of the  $\beta$  coefficients and the covariances between  $\beta$  coefficients to then find the standard error.

```
b.vals = c(-.52, .04, -.047, -.18)
age.vals = 17:36 - 23
X.mat = cbind(rep(1, 2*length(age.vals)),
               age.vals,
               rep(c(1, 0), c(length(age.vals), length(age.vals))),
               age.vals * rep(c(1, 0), c(length(age.vals), length(age.vals))))
colnames(X.mat) = c("Intercept", "Age", "Care", "Age_Care")
res.mat = as_data_frame(X.mat)
res.mat$logit.odds = c(X.mat %*% b.vals)
res.mat$odds = exp(res.mat$logit.odds)
res.mat$probs = res.mat$odds / (1 + res.mat$odds)

cov.mat = matrix(c(0.046, 0.002, -0.046, -0.002,
                   0.002, 0.002, -0.002, -0.002,
                   -0.046, -0.002, 0.111, 0.004,
                   -0.002, -0.002, 0.004, 0.005), ncol = 4, byrow = TRUE)
comp.vec = c(1, 1, 1, 1)
age.vec = c(1, 1, 0, 0)
```

```

inv.logit = function(x){exp(x) / (1+ exp(x))}
res.mat = res.mat %>% mutate(var = cov.mat[1,1] +
                             Age^2 * cov.mat[2,2] +
                             Care^2 * cov.mat[3,3] +
                             Age_Care^2 * cov.mat[4,4] +
                             2*Age*cov.mat[1,2] +
                             2*Care*cov.mat[1,3] +
                             2 * Age_Care*cov.mat[1,4] +
                             2 * Age*Care * cov.mat[2,3] + 2 * Age * Age_Care * cov.mat[2,4] +
                             2 * Care * Age_Care * cov.mat[3,4],
                             se = sqrt(var),
                             lower.logit = logit.odds - 1.96 * se,
                             upper.logit = logit.odds + 1.96 * se,
                             lower.odds = exp(lower.logit),
                             upper.odds = exp(upper.logit),
                             low.prob = inv.logit(lower.logit),
                             upper.prob = inv.logit(upper.logit)
                             )

res.mat$Care = ifelse(res.mat$Care == 1, "Care","No_care")

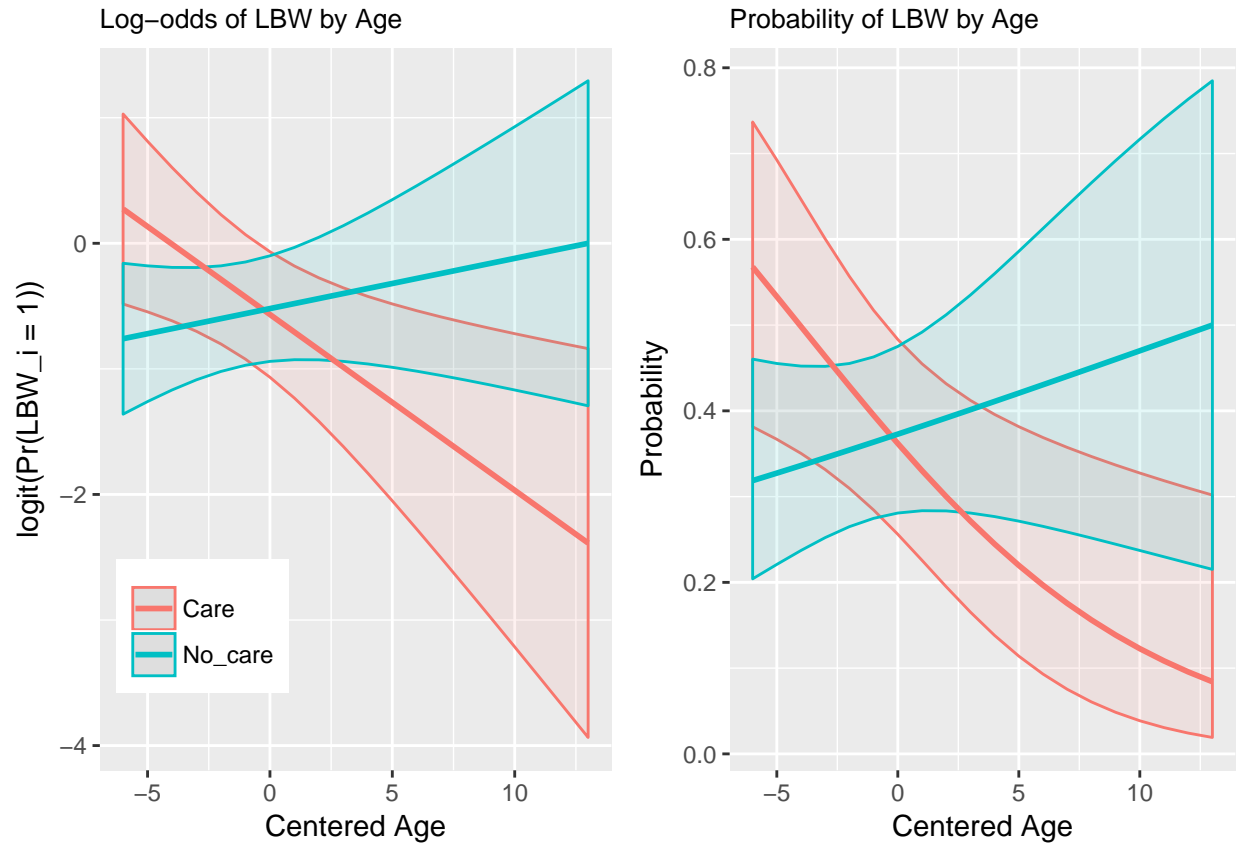
logit.plot = ggplot(res.mat, aes(x = Age,y = logit.odds, group = factor(Care),color = factor(Care))) +
  geom_ribbon(aes(ymin=lower.logit, ymax=upper.logit, fill = Care), alpha=0.1)+
  geom_line(size = 1) +
  theme(legend.position = c(.2, .2)) +
  guides(fill = FALSE) +
  labs(subtitle = "Log-odds of LBW by Age",
       x = "Centered Age",
       y = "logit(Pr(LBW_i = 1))",
       color = NULL)

odds.plot = ggplot(res.mat, aes(x = Age, y = odds, group = Care, color = Care)) +
  geom_ribbon(aes(ymin = lower.odds, ymax = upper.odds, fill = Care), alpha = .1) +
  geom_line(size = 1) +
  theme_bw() +
  guides(fill = FALSE) +
  labs(subtitle = "Odds- of LBW by Age and Care",
       x = "Centered Age",
       y = "Odds of Low-Birthweight")

prob.plot = ggplot(res.mat, aes(x = Age,y = probs, group = factor(Care),color = factor(Care))) +
  geom_ribbon(aes(ymin = low.prob,ymax = upper.prob,fill = Care),alpha = .1) +
  geom_line(size = 1) +
  guides(fill = FALSE, color = FALSE, group = FALSE) +
  labs(subtitle = "Probability of LBW by Age",
       x = "Centered Age",
       y = "Probability",
       color = "Pre-natal Care")

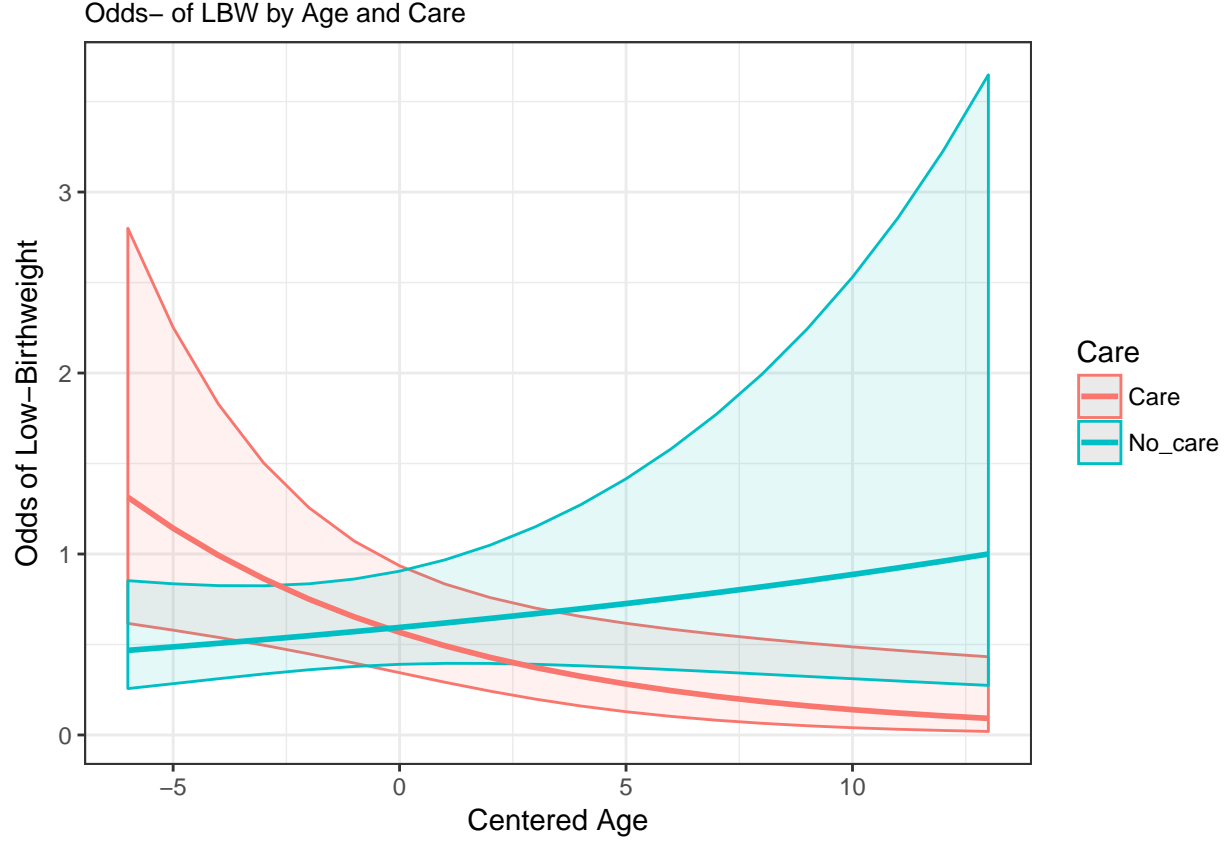
grid.arrange(logit.plot,prob.plot,ncol = 2)

```



Analyzing the above plot, it is clear that the interaction effect is quite strong. These plots are of the log-odds and probability of low-birthweight. In the first plot, we see that these lines cross at around the intercept, which would suggest that at age 23, the odds of having a child with low-birthweight is about the same with or without adequate care. However, as age increases, we see a large discrepancy appear between women with and without appropriate care. For individuals with appropriate care, the probability of low birthweight decreases while it increases in those without appropriate care. Another thing of note is that, the confidence intervals overlap consistently in both of these plots, which mean that this difference may not be significant, but we would need to perform more tests to find that.

```
odds.plot
```



The plot above describes the same store, but on the scale of odds of low-birthweight compared to the baseline. All the plots show evidence of at least some interaction effect because the lines are clearly not parallel, and they cross at our intercept of 23 years old. These plots confirm the significance of the interaction in this study. According to these plots, adequate care seems to reverse the effects of age on low-birthweight. Intuitively, this result should be limited because there is a time when women can no longer have children, and thus these effects are probably not as simple as these plots suggest. However, this result also makes sense because generally, we are told that higher aged pregnancies are associated with higher risk pregnancies, and so the importance of adequate care for higher-aged pregnancies is especially important. These results support that.

The interpretation of the coefficients is as follows:

$e^{\beta_0} = .5945$  (0.3904821, 0.9051751) represents the odds of Low-birthweight for an 23 year-old individual without adequate care.

$e^{\beta_1} = 1.0408$  (0.9534638, 1.1361597) represents the odds ratio of low-birthweight for a change of one year of age for someone with inadequate care.

$e^{\beta_2} = 0.9541$  (0.496582, 1.833096) represents the odds ratio for low-birthweight between a 23 year-old individual who receives adequate care and one who does not.

$e^{\beta_3} = 0.8353$  (0.7271715, 0.9594385) represents an change in the odds ratio associated with a change of one year of age for an individual with adequate care compared to inadequate care.

### 3 Interrater Agreement for Diagnosis of Epileptic Seizures.

Consider data from a study of interrater agreement in the diagnosis of type of epileptic seizures. Two neurologists assessed the medical records of 100 patients and were asked to determine for each patient whether

the patient had experienced generalized epileptic seizures, partial epileptic seizures, or no epileptic seizures. The diagnoses of both neurologists are shown below.

Neurologist 1	No Seizures	Partial	Generalized
No Seizures	5	4	1
Partial	7	39	9
Generalized	3	15	17

Fit the interrater agreement models discussed in class to determine the type of agreement between the two neurologists. Describe your findings, including a discussion of whether the neurologists differ in their tendency to diagnose each type of seizures overall, the type of agreement (if any) between the neurologists in their diagnoses, and any particular diagnoses that may be more challenging for the neurologists to make. Be sure to include fully reproducible code with your answer.

### 3.1 Solution

We begin by exploring Cohen's kappa statistics, which can be used as a summary of agreement between two raters. Although the simplicity of this statistics is nice, it can also hide valuable information about the nature of matching and is limited in nature. It is important to note that kappa is between 0.3 and 0.38 for both weighted and unweighted of both weighting systems: equal and Fleiss-Cohen. According to the Landis and Koch scale, this is associated with fair agreement. Under both these tests, we see that the weighted value is higher than the unweighted. We will not dwell on these results, and we will now fit actual agreement models.

```
seizure = matrix(c(5,4,1,7,39,9,3,15,17), ncol = 3, byrow = TRUE)
dimnames(seizure) = list(rater1 = c("no_seizures", "partial", "generalized"), rater2 = c("no_seizures", "partial", "generalized"))
Kappa(seizure)

##           value      ASE      z  Pr(>|z|)
## Unweighted 0.3176 0.08231 3.859 1.141e-04
## Weighted   0.3395 0.08237 4.121 3.766e-05

Kappa(seizure, weights = "Fleiss-Cohen")

##           value      ASE      z  Pr(>|z|)
## Unweighted 0.3176 0.08231 3.859 0.0001141
## Weighted   0.3704 0.09746 3.800 0.0001445

seiz.tab = as.data.frame(as.table(seizure))
# Uniform Exactness
seiz.tab$unifexact = as.numeric(seiz.tab$rater1 == seiz.tab$rater2)
# Pattern of Exact Agreement
seiz.tab$exact1 = as.numeric((seiz.tab$rater1 == "no_seizures") & (seiz.tab$rater2 == "no_seizures"))
seiz.tab$exact2 = as.numeric((seiz.tab$rater1 == "partial") & (seiz.tab$rater2 == "partial"))
seiz.tab$exact3 = as.numeric((seiz.tab$rater1 == "generalized") & (seiz.tab$rater2 == "generalized"))
```

For these models, we will use log-linear modeling approaches. We will test three different models under two different paradigms. The first paradigm is that the result is not ordered, or for our case, generalized epileptic seizures, partial epileptic seizures, and no epileptic seizures are not ordered. The second paradigm is exploring linear by linear association model, which assigns ordinal scores to the possible results and assumes that there is a natural ordering in the seizure results. Within each of these paradigms, we will test three different models. We test the independence model, where we just test the raters. We then test the quasi-independence model, where we add a parameter to accommodate exact matching. This represents the matching in the diagonal of the table above. Finally, we test the quasi-independence model for patterns of exact matching, where we can explore where the matching occurs in the table. In this case, we break down the diagonal even more

and add a parameter for each of the diagonal elements. Once we have fit these six different models, we will informally compare their deviances and degrees of freedom, we do this informally because we can't use analysis of deviance for unnested models, which characterizes the unordered and ordered models.

```
# Independence Mode
m.ind = glm(Freq ~ rater1 + rater2, data = seiz.tab,
            family = poisson(log))
# Uniform Exact Association Model
m.ue = glm(Freq ~ rater1 + rater2 + unifexact, data = seiz.tab,
            family = poisson(log))
# Pattern of Exact Agreement
m.pea = glm(Freq ~ rater1 + rater2 + exact1 + exact2 + exact3, data = seiz.tab,
            family = poisson(log))

# Ordinal Scores
seiz.tab$r1score = rep(c(0,1,2),3)
seiz.tab$r2score = rep(c(0,1,2),c(3,3,3))

# Linear by Linear ASsociation Model
m.lbl = glm(Freq ~ rater1 + rater2 + r1score:r2score, data = seiz.tab,
            family = poisson(log))
m.lbl.ue = glm(Freq ~ rater1 + rater2 + r1score:r2score + unifexact, data = seiz.tab,
               family = poisson(log))
m.lbl.pea = glm(Freq ~ rater1 + rater2 + r1score:r2score + exact1 + exact2 + exact3, data = seiz.tab,
                family = poisson(log))
#m.lbl.ue[["deviance"]]
#m.lbl.ue[["df.residual"]]

deviance_frame = data_frame(
  model_name = c("Independence","Uniform Exact","Pattern of Exact Agreement","Linear-by-Linear","LxL UE",
  deviance = c(m.ind[["deviance"]],m.ue[["deviance"]],m.pea[["deviance"]],m.lbl[["deviance"]],m.lbl.ue[
  dof = c(m.ind[["df.residual"]],m.ue[["df.residual"]],m.pea[["df.residual"]],m.lbl[["df.residual"]],m.lbl.ue[
)
options(digits = 9)
deviance_frame[,2] = round(deviance_frame[,2],4)
kable(deviance_frame)
```

model_name	deviance	dof
Independence	18.9710	4
Uniform Exact	2.4066	3
Pattern of Exact Agreement	0.0004	1
Linear-by-Linear	3.6322	3
LxL UE	0.1491	2

The linear by linear for pattern of exact agreement is overfit and so we don't include it in the table above.

An important consideration in this analysis is the overall lack of degrees of freedom. We only have 9 total degrees of freedom, and it is important to note that in all of these log-linear models, we have at least five terms:  $\lambda$ ,  $\lambda_1^X$ ,  $\lambda_2^X$ ,  $\lambda_1^Y$ ,  $\lambda_2^Y$ . This is a difficult problem because we aren't necessarily interested in these parameters. We are mostly interested in the agreement parameters. And when we add other facets to the model, like uniform exact agreement or pattern of agreement, the available degrees of freedom are used up quickly.

It is clear that when we do analysis of deviance within the unordered model, adding the uniform exactness indicator into the model provides significant improvement o the independence model. However, the pattern of exact agreement does not provide significant improvement. No models in the ordered case provide significant improvement to the independent linear by linear model. The best model seems to be either the unordered model with uniform exactness or the linear by linear association model.

Also, the uniform exactnes parameter is clearly significant in the unordered case. It seems like there is a significant degree of agreement generally, but it is not as clear when we break it into the overall pattern of exact agreement.

```
kable(summary(m.pea)$coefficients)
```

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.232810727	0.653794114	-0.356091807	0.721771813
rater1partial	2.180839177	0.678907806	3.212275890	0.001316878
rater1generalized	1.326462795	0.514036656	2.580482889	0.009866225
rater2partial	1.615387232	0.562440205	2.872104832	0.004077477
rater2generalized	0.247545453	0.468813326	0.528025634	0.597481545
exact1	1.842248640	0.792115358	2.325732763	0.020032812
exact2	0.100145965	0.642858901	0.155782187	0.876204712
exact3	1.492015823	0.681451610	2.189467014	0.028562914

We can't asnwer the last question about where there is disagreement using our preferred model of uniform exact agreement. As a result, we use a more saturated model in the patterns of exact matching to understand where there is agreement or disagreement. To assess which type of seizure is most difficult to diagnose, we look at the coefficients for the different matching category. For both no seizures and generalized seizures, there is matching beyond independent raters, but the same does not hold true for the partialized seizures. This could indicate that this is the most difcult to diagnose.