Quiz 2

Sonya Eason

Exercise 1

Poisson regression assumes

1. independence of events observed
2. $log $ is a linear function of X
3. mean is equal to variance for each level of predictors
4. response is a count of occurrences within a space or time

Exercise 2

The response is the patients' number of relapses within five years of initial treatment.

$$

$$

Exercise 3

Possible values are 0, 1, 2, all the way up to infinity but contextually, this means the max number of relapses.

Exercise 4

Lambda represents the mean rate of a patient's relapses per year.

Exercise 5

A zero-inflated model could be suitable here because there are likely to be many observations of zero relapses, more than we would expect in a typical Poisson. Also, ZIP is good since there are two types of zeroes: the true zeroes i.e., the patients without the disease who have been cured from initial treatment and the false zeroes i.e, the patients who have not relapsed yet but are not totally cured.

Exercise 6

The assumption that mean and variance are the sum are equal is often unrealistic for observed data. Oftentimes, there's more variability than the theoretical Poisson's mean would convey. This phenomenon is overdispersion.

Exercise 7

To correct for overdispersion, either of these methods could work: Quasi-Poisson where we inflate standard errors or using negative-binomial regression model.

Exercise 8

The predictor is income, in ten thousands of dollars.

```
exp(2.1)
```

```
[1] 8.16617
```

Interpretation: For each 10,000 dollars that income increases, the mean number of credit cards used is expected to multiply by a factor of 8.17.

Exercise 9

We can assess linearity by plotting log of empirical mean of credit card use against the incomes. This will give an initial idea about whether there's a linear relationship. To actually implement this, we'd have to consider the different levels of income and compute the mean credit card use for each income level. From here, we'd take the log of each of these means, so that we could plot income level vs log mean.

Exercise 10

To assess the equal mean and variance assumption, we could look at both the overall mean and variance in credit card use and the means and variances in credit card use by levels of income. For example, it may be useful to look at each 10k group of income and get the mean and variance in credit card use. From the corresponding calculating means and variances calculating, we can determine whether the means and variances are close enough to be considered equal and see whether this mean = variance condition is satisfied.

Exercise 11

The response is number of dates arranged online. The predictor was age. The offset was time online. This offset does make sense because we'd expect number of dates arranged online to vary by time spent online, so adjusting for this will allow us focus on how the predictor of age relates to number of dates arranged online.

Exercise 12

Given λ is the mean number of dates respondents arranged online.

$$log(\frac{\lambda}{time}) = \beta_o + \beta_1 age$$

$$log(\lambda) = \beta_0 + \beta_1 * age + log(time)$$

Exercise 13

For a ZIP model, the true zeroes would be the individuals who do not do online dating. The other zeroes would be those who try online dating but just haven't arranged an online date in the past three months.