Lesson 10

Quiz, 8 questions

8/8 points (100%)

Congratulations! You passed!

Next Item

1 / 1 points

1.
With Poisson regression, we use the log link function so that $\log(\mathrm{E}(y)) = \log(\lambda) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$.

Suppose we have two covariate values $x_{1,i} = 0.8$ and $x_{2,i} = 1.2$ and we know the values of the coefficients: $\beta_0 = 1.5$, $\beta_1 = -0.3$, and $\beta_2 = 1.0$.

Calculate $\mathrm{E}(y_i)$ in this case. Round your answer to one decimal place.

11.7

Correct Response
This is just $\lambda_i = \exp(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i})$.

---

1 / 1 points

2.
Re-run the JAGS model for the Poisson regression on doctor visits from the lesson. Calculate the DIC for the original model. Now remove the interaction term from the model and fit the simpler additive model. Again compute the DIC. If we use predictive performance as a criterion for selecting models, what do we conclude?

- The original model with interaction has a **lower** value of DIC than the simpler model, so we retain the **simpler** model.
- The original model with interaction has a **higher** value of DIC than the simpler model, so we retain the **simpler** model.
- The original model with interaction has a **higher** value of DIC than the simpler model, so we retain the **original** model.
- The original model with interaction has a **lower** value of DIC than the simpler model, so we retain the **original** model.

Correct

Using the interaction term improves the model predictive performance according to the DIC.

Lesson 10

Quiz, 8 questions

8/8 points (100%)

1 / 1 points

3.

In the original model, the posterior mean estimate of the coefficient for $\mathbf{badh}$ (bad health) was 1.56. What is the correct interpretation of this coefficient if this were the true value?

- Being in bad health is associated with a 1.56 decrease in the log of the number of doctor visits.
- Being in bad health is associated with a 1.56 increase in the expected number of doctor visits.
- Being in bad health is associated with a 1.56 decrease in the log of the expected number of doctor visits.
- Being in bad health is associated with a 1.56 increase in the log of the expected number of doctor visits.

Correct

Since $\mathbf{badh}$ was an indicator (dummy) variable, the interpretation of the coefficient is simply the effect of this variable being "on."

1 / 1 points

4.

In the previous course, we briefly discussed Poisson processes. The mean of a Poisson distribution can be thought of as a rate at which the events we count are occurring. Hence, it is natural to imagine that if we are observing for twice as long, we would expect to count about twice as many events (assuming the rate is steady). If t is the amount of time that we observe, and λ is the rate of events per unit of time, then the expected number of events is $t\lambda$ and the distribution of the number of events in this time interval is $\mathrm{Poisson}(t\lambda)$.

Suppose that a retail store receives an average of 15 customer calls per hour, and that the calls approximately follow a Poisson process. If we monitor calls for two hours, what is the probability that there will be fewer than 22 calls in this time period? Round your answer to two decimal places.

0.05

Correct Response

Let X be the number of calls in this interval. Then $X \sim \mathrm{Pois}(30)$ and the probability of fewer than 22 calls can be calculated in R.

```
ppois(q=21, lambda=30)
```

Lesson 10

Quiz, 8 questions

8/8 points (100%)

1 / 1 points

5.

On average, this retailer receives 0.01 calls per customer per day. They notice, however, that one particular group of customers tends to call more frequently.

To test this, they select 90 days to monitor 224 customers, 24 of which belong to this group (call it group 2). Not all customers had accounts for the full 90 day period, but we do know how many of the 90 days each was active. We also have the age of the customer, the group to which the customer belongs, and how many calls the customer placed during the period they were active. The data are attached as `callers.csv`.

callers.csv

Download and read the data into R using the following code.

```
dat = read.csv(file="callers.csv", header=TRUE)
  ## set R's working directory to the same directory
  ## as this file, or use the full path to the file.
```

Try plotting some of the variables to understand some of the relationships. If one of the variables is categorical, a box plot is a good choice.

Which of the following plots would be most useful to the retailer to informally explore their hypothesis that customers from group 2 call at a higher rate than the other customers?

- `age` vs. `isgroup2`
- `calls` / `days_active` vs. `isgroup2`

Correct

This is the best choice. The first variable is the observed call rate for the customer, which is better than using the total number of calls because that does not account for possible differences in how long the account was active.

The `age` vs. `isgroup2` plot is also very important. If the customers in group 2 tend to be different in age than the other customers, then we will not be able to tell whether group membership or age (or another variable related to both) is driving the difference in call rates.

- `calls` / `days_active` vs. `age`
- `calls` vs. `isgroup2`

1 / 1

points

Since we know how many days each customer was active and the data are counts, it seems reasonable to use a Poisson process model. We will assume that the customers' calls are independent, and that the calling rate per day active for person i, λ_i is unique to each customer and does not change throughout the monitoring period.

It makes sense to model λ_i using our two covariates, `age` and `isgroup2`. How would the likelihood specification for this model look in JAGS?

```
for (i in 1:length(calls)) {
  calls[i] ~ dpois( lam[i] )
  log(lam[i]) = b0 + b[1]*age[i] + b[2]*isgroup2[i]
}
```

```
for (i in 1:length(calls)) {
  calls[i] ~ dpois( days_active[i] * lam[i] )
  log(lam[i]) = b0 + b[1]*age[i] + b[2]*isgroup2[i]
}
```

Correct

The Poisson process part comes in when we account for `days_active`, and the regression part comes in our model for λ_i.

```
for (i in 1:length(calls)) {
  calls[i] ~ dpois( days_active[i] * lam[i] )
  lam[i] = b0 + b[1]*age[i] + b[2]*isgroup2[i]
}
```

```
for (i in 1:length(calls)) {
  calls[i] ~ dpois( lam[i] )
  lam[i] = b0 + b[1]*age[i] + b[2]*isgroup2[i]
}
```

1 / 1

points

7.

Complete fit the model in JAGS using $\mathrm{N}(0, 10^2)$ priors for the intercept and both coefficients. Be sure to check for MCMC convergence and look at the residuals (don't forget to multiply `lam_hat` by `days_active` to obtain the model's predicted mean number of calls).

What is the posterior probability that β_2, the coefficient for the indicator `isgroup2` is greater than 0? Round your answer to two decimal places.

1.00

Correct Response

Even the posterior probability that $\beta_2 > 1.0$ is greater than 0.99. There is strong evidence that the coefficient is positive.

The striped pattern in the residuals is due to the fact that the data are discrete counts.

Lesson 10

Quiz, 8 questions

8/8 points (100%)

8. 1 / 1 points

What formal conclusions should be made about the retailer's hypothesis?

- The data contain no evidence of association between call rates and group membership.
- While accounting for time active and age, customer membership in group 2 is associated with higher call rates than for customers not in group 2.

Correct
There is strong evidence that membership in group 2 is associated with higher call rates.

- While accounting for time active and age, customer membership in group 2 is associated with lower call rates than for customers not in group 2.
- We are unable to conclude whether the calling rate discrepancy is due to group membership or age because members of group 2 are generally of different age than customers not in group 2.