Lesson 7 Part A

Quiz, 8 questions

8/8 points (100%)

Congratulations! You passed!

Next Item

1/1 points

1. In a normal linear regression model with $\mathrm{E}(y_i) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i}$, which of the following gives the correct interpretation of β_2?

- When $x_{2,i} = 0$, a one unit change in $x_{1,i}$ and $x_{3,i}$ results in a β_2 in y_i.
- While holding $x_{1,i}$ and $x_{3,i}$ constant, a one unit change in $x_{2,i}$ results in a β_2 change in the expectation of y_i.

Correct
This is how much the expected response increases when we increase $x_{2,i}$ by 1.0 while controlling for the other predictors.

- While holding $x_{2,i}$ constant, the expectation of y_i is β_2.
- While holding $x_{1,i}$ and $x_{3,i}$ constant, a one unit change in $x_{2,i}$ results in a β_2 change in y_i.

1/1 points

2. Which of the following model specifications for $\mathrm{E}(y_i)$ is not a valid linear model?

- $\beta_0 + \beta_1 \sin(2\pi x_{1,i}) + \beta_2 x_{2,i}$
- $\beta_0 + \beta_1 \log(x_{1,i}) + \beta_2 x_{2,i}^2$
- $\beta_0 + \exp(\beta_1 x_{1,i}) + \beta_2 x_{2,i}^2$

Correct
This model is not linear in the coefficients. We are free to transform the predictors and the response, but the model itself must be linear.

- $\beta_0 + \beta_1 x_{1,i} + \beta_2 (x_{1,i}/x_{2,i})$

1/1 points

3. Consider the Anscombe data set in R which can be accessed with the following code:

```
library("car")  # load the 'car' package
data("Anscombe")  # load the data set
?Anscombe  # read a description of the data
head(Anscombe)  # look at the first few lines of the data
pairs(Anscombe)  # scatter plots for each pair of variables
```

Suppose we are interested in relating per-capita education expenditures to the other three variables. Which variable appears to have the strongest linear relationship with per-capita education expenditures?

- Proportion of population under age 18
- Proportion of population that is urban
- None of these variables appears to have a linear relationship with education expenditures.
- Per-capita income

Correct
It appears that increases in income generally co-occur with increases in education expenditures.

1/1 points

4. Fit a reference (noninformative) Bayesian linear model to the Anscombe data with education expenditures as the response variable and include all three other variables as predictors. Use the `lm` function in R.

What is the posterior mean estimate of the intercept in this model? Round your answer to one decimal place.

-286.8

Correct Response

1 / 1 points

5. In our reference analysis of the Anscombe data, the intercept is estimated to be negative. Does this parameter have a meaningful interpretation?

- No, this model should not have an intercept term at all.
- No, there must be something wrong with the model because expenditures can never be negative.
- Yes, it represents expected expenditures in a state with average income, average percent youth, and average percent urban.
- (selected) No, it represents expected expenditures in a state with 0 average income, 0 percent youth, and 0 percent urban which doesn't exist.

Correct
Although this parameter is not very interpretable in this particular example, it is necessary to the model.

One strategy for making the intercept more interpretable would be to subtract the averages of the predictors from their values (i.e., $x^*_{1,i} = (x_{1,i} - \bar{x}_1)$). Then the intercept would represent expected expenditures in a state with average income, average percent youth, and average percent urban.

---

1 / 1 points

6. Use the code below to fit a linear regression model to the Anscombe data in JAGS. You will need to finish setting up and running the model.

```
library("rjags")

mod_string = " model {
    for (i in 1:length(education)) {
        education[i] ~ dnorm(mu[i], prec)
        mu[i] = b0 + b[1]*income[i] + b[2]*young[i] + b[3]*urban[i]
    }

    b0 ~ dnorm(0.0, 1.0/1.0e6)
    for (i in 1:3) {
        b[i] ~ dnorm(0.0, 1.0/1.0e6)
    }

    prec ~ dgamma(1.0/2.0, 1.0*1500.0/2.0)
      ## Initial guess of variance based on overall
      ## variance of education variable. Uses low prior
      ## effective sample size. Technically, this is not
      ## a true 'prior', but it is not very informative.
    sig2 = 1.0 / prec
    sig = sqrt(sig2)
} "

data_jags = as.list(Anscombe)

```

Before proceeding to inference, we should check our model. The first step is to check our MCMC chains. Do there appear to be any problems with the chains?

- No, a few thousand iterations will be sufficient for these chains.
- Yes, scale reduction factors are well above 1.0. The chains are not exploring the same distribution.
- (selected) Yes, there is very high autocorrelation among the coefficients. It would be good to run the chain for 100,000+ iterations to get reliable estimates.

Correct
This is a common issue in Gibbs samplers, where parameters are updated one-at-a-time.

- Yes, there is very high autocorrelation for `sig`. We should help the chain for `sig` by fixing the initial value.

---

1 / 1 points

7. Which of the following is **not** a condition we can check using a residual plot with predicted values on the x-axis and residuals on the y-axis?

- (selected) Independence of the observations

Correct
One way to check this assumption is by plotting predicted values against the data index. In the Anscombe data, we could check to see if residuals are more similar for states that are geographically close than for states that are not geographically close. If that is true, there may be spatial correlation in the data.

- Linearity of the relationship between predictors and the response
- Constant error variance
- Presence of outliers

---

1 / 1 points

8. Check the residual plot described in Question 7 for the Anscombe data. Since the estimates of the coefficients in the reference model are very close to those in the JAGS model, we will just look at the residuals of the reference model. This plot is the first that appears when you run the following code:

```
plot(mod_lm)
# here mod_lm is the object saved when you run lm()
```

Lesson 7 Part A

Quiz, 8 questions

8/8 points (100%)

Do there appear to be any issues with this fit?

- (selected) Yes, the error variability appears to increase as predicted values increase.

Correct
There are alternative versions of linear models that address this issue, but we will not pursue them in this course.

- Yes, there is a curved pattern or shape to the residuals, indicating a nonlinear relationship between the variables.
- Yes, the observations appear not to be independent.
- Yes, there are a few extreme outliers.
- No, this plot raises no concerns.