Take Home Quiz 3

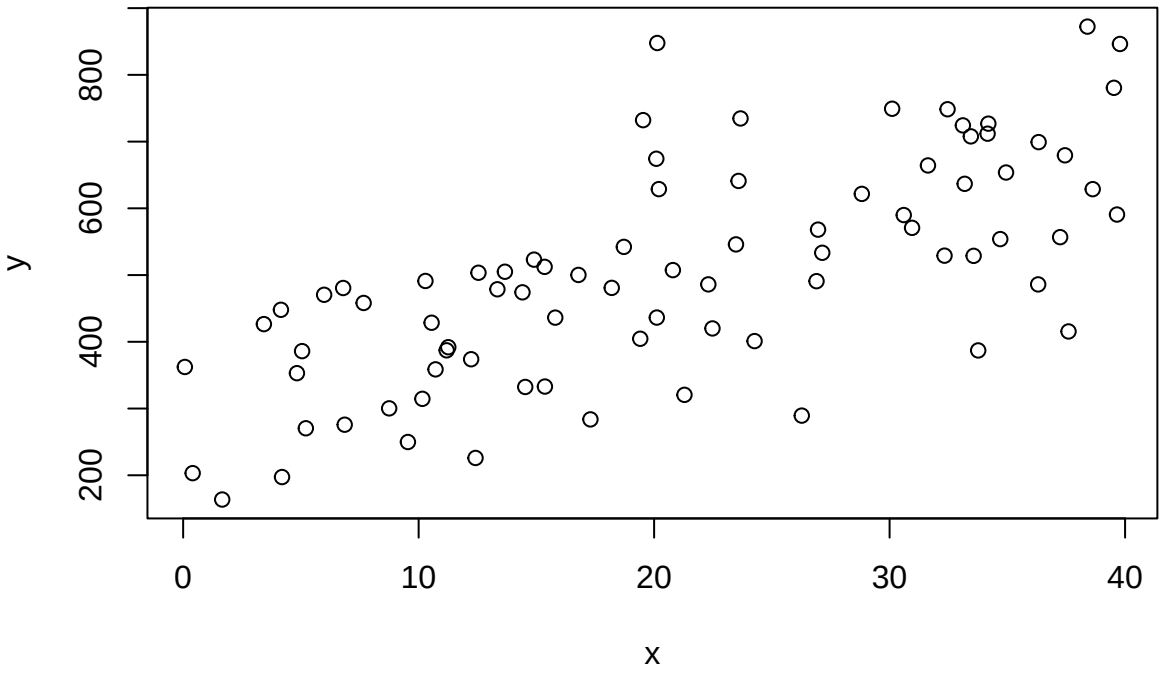
Solutions

Due Tuesday Nov 26 at 11:00am

There is no time limit, but this quiz should take you about 30-40 minutes. Place your answers into this markdown document, knit it, and hand in the result as a PDF or Word document. You may use R, any reference material, and information already available on the internet. Do not work together, do not use AI, and do not get help from anyone but Dr. Clair.

Problem 1 (10 points)

Consider the normal regression model $Y_i \sim N(\mu_i, \sigma)$ with $\mu_i = \beta_0 + \beta_1 X_i$ From this plot, give your best estimate (by eye) of β_0, β_{0c} (centered), β_1 and σ.



Solution β_0 looks to be between 100 and 400. β_{0c} is between 300 and 600. β_1 is between 10 and 20. σ is around 50 to 150.

Problem 2 (10 points)

In the regression model from Problem 1, suppose we choose priors $\beta_0 \sim N(200, 100)$ and $\beta_1 \sim N(0, 40)$.

a. Give an example of any prior plausible line by writing its equation.

b. Give an example of any line that would not be prior plausible, by writing its equation.

Solution Many answers. $y = 200 + 0 \cdot x$ is certainly plausible (it's the most plausible, in fact). $y = 10000000 + 10000000x$ would not be plausible.

Problem 3 (10 points)

This problem and the last two use data on blood alcohol content, available on our web page at https://turtlegraphics.org/bayes/data/bac.csv

16 students from The Ohio State University were randomly assigned to drink a number of cans of beer. Then their blood alcohol content (BAC) was measured.

Use rstanarm to build a normal regression model of `bac` on `beers` using the default weakly informative priors, 4 chains, and 10000 iterations per chain.

Report the mean posterior beers coefficient to at least three decimal places. What does this number tell you about beer drinking?

Solution The mean posterior beers coefficient is 0.018, which says that on average consuming one beer will raise BAC by 0.018.

```
suppressMessages(library(rstanarm))
bac <- read.csv("https://turtlegraphics.org/bayes/data/bac.csv")
bac_mod <- stan_glm(bac ~ beers, data=bac, iter=10000)
```

```
bac_mod$coefficients
```

```
## (Intercept)       beers
## -0.01226222  0.01787314
```

Problem 4 (10 points)

What prior did rstanarm choose for the σ parameter? What is the mean of the σ prior distribution?

Solution The prior is $\sigma \sim \text{Exp}(23)$ which has mean $1/23 = 0.043$.

```
prior_summary(bac_mod)
```

```
## Priors for model 'bac_mod'
## ------
## Intercept (after predictors centered)
##   Specified prior:
##     ~ normal(location = 0.074, scale = 2.5)
##   Adjusted prior:
##     ~ normal(location = 0.074, scale = 0.11)
##
## Coefficients
##   Specified prior:
##     ~ normal(location = 0, scale = 2.5)
##   Adjusted prior:
##     ~ normal(location = 0, scale = 0.05)
##
## Auxiliary (sigma)
##   Specified prior:
##     ~ exponential(rate = 1)
##   Adjusted prior:
##     ~ exponential(rate = 23)
## ------
## See help('prior_summary.stanreg') for more details
```

Problem 5 (10 points)

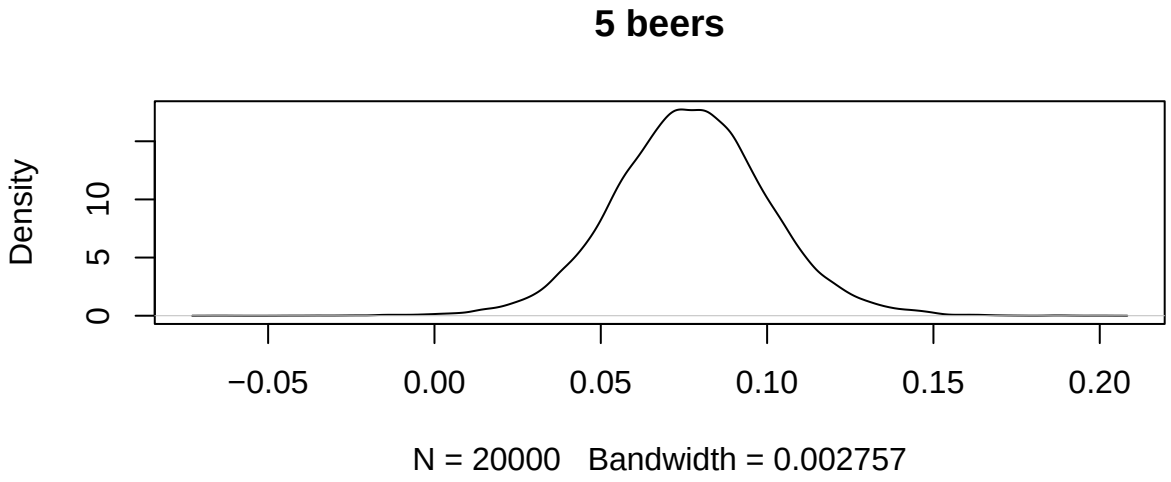
Make a density plot of the posterior predictive distribution for a student that drinks 5 beers.

Make a density plot of the posterior predictive distribution for a student that drinks 9 beers.

Both distributions use the same σ values, but one is noticeably wider. Why?

Solution The 9 beer predictions are wider (less accurate) because the variation in the regression line is greater away from the center of the data.

```
pp <- posterior_predict(bac_mod, newdata = data.frame(beers = c(5,9)))
plot(density(pp[,1]), main="5 beers")
```




```
plot(density(pp[,2]), main="9 beers")
```