In-Class Problems

Name (print clearly): _____

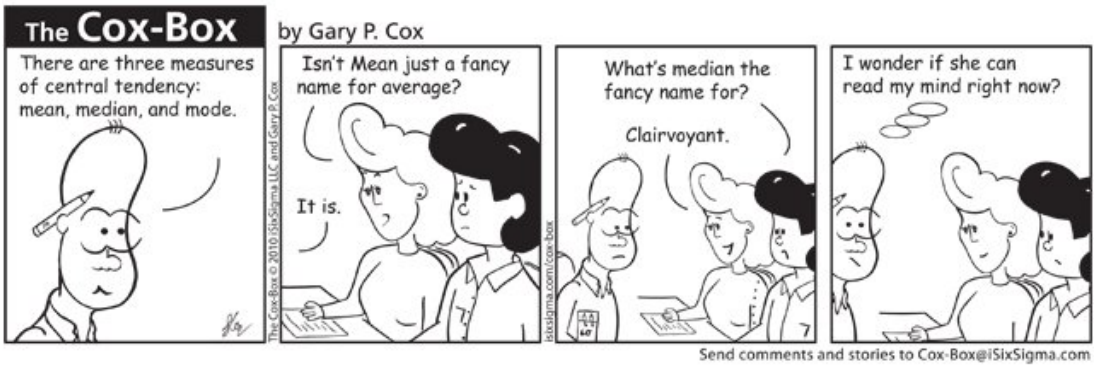
Do your work on these sheets of paper. Your paper will be scanned into gradescope, so avoid writing at the very edge of the paper.

1 Bayesian modeling centers on three important components: the prior, the likelihood, and the posterior.

a) *Briefly* (approximately 1 – 3 sentences for each) describe what each of these is and its role in the Bayesian modeling process. I'm looking for your answer to convince me that you understand what these are and can express that understanding accurately in words.

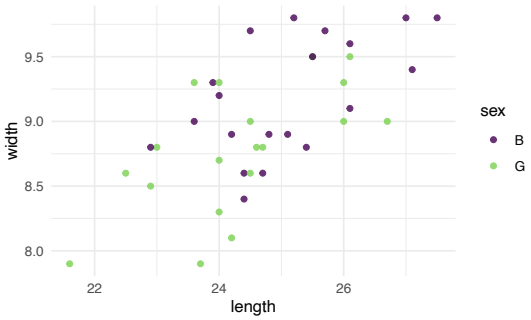
b) What important "word-equation" involves these three words?

c) Express this word equation in mathematical notation.



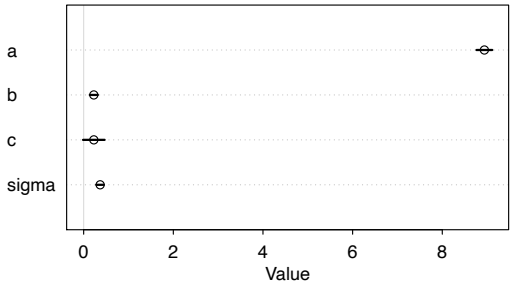
2 Below is the R code that fits a model using `quap()` using data on the length and width of 39 kids' feet (measured in cm) and whether they are boys or girls. See the next page for questions about this output.

```
gf_point(width ~ length, color = ~sex,
         data = KidsFeet, alpha = 0.8) %>%
  gf_refine(scale_color_viridis_d(end = 0.8))

KidsFeet <- KidsFeet %>%
  mutate(
    length_s = length - 25,
    male = as.numeric(sex == "B"))
```




```
# Two example kids
KidsFeet %>% filter(name %in% c("Laura", "Mark"))

   name birthmonth birthyear length width sex biggerfoot domhand length_s male
1  Mark          9        87   27.5   9.8   B          R       R      2.5    1
2 Laura          9        88   24.0   8.3   G          R       L     -1.0    0

Kids.model <- quap(
  alist(
    width ~ dnorm(mu, sigma),
    mu <- a + b * length_s + c * male,
    a ~ dnorm(10, 3),
    b ~ dlnorm(0, 1),
    c ~ dnorm(0, 2),
    sigma ~ dexp(1)),
  data = KidsFeet)
```

```
precis(Kids.model, prob = 0.95)

      mean   sd  2.5% 97.5%
a     8.94 0.09  8.76  9.12
b     0.23 0.05  0.14  0.32
c     0.23 0.12 -0.01  0.47
sigma 0.37 0.04  0.29  0.45

plot(precis(Kids.model, prob = 0.95))
```



Answer the following questions about this model based on the output above. (If you were doing this yourself, you might like to do additional things, but do what you can with what you have.)

a) Interpret row `a` of the precis output. What does this tell us about the model?

b) Interpret row `b` of the precis output. What does this tell us about the model?

c) Interpret row `c` of the precis output. What does this tell us about the model?

d) Interpret row `sigma` of the precis output. What does this tell us about the model?

e) Sketch what you think a desnity plot of the posterior distribution for `c` would look like. Be sure to label the x-axis of your plot. (No need to label the y-axis, I'm just interested in the overall shape.)

f) Compute the MAP predictions for Laura and Mark. (MAP = maximum a posteriori = mode of the posterior, which in this case is the same as mean of the posterior.)

g) Compute residuals for Laura and Mark.

h) Another modeler proposes replacing `b ~ dlnorm(0, 1)` with `b ~ dnorm(0, 1)`. What is the main difference between these two options? How do you expect making this change would affect the output of `precis()`?

3 On a table are three unmarked boxes. Each box contains black and white marbles, but in different proportions.

- Box A: 30% black, 70% white
- Box B: 40% black, 60% white
- Box C: 50% black, 50% white

Now consider the following (admittedly silly) situation. A fair six-sided die is rolled. If the result is 1, 2 or 3, Box A is selected. If the result is a 4 or a 5, Box B is selected. If the result is a 6, Box C is selected. From the selected box, two times a marble is removed, its color noted, and it is returned to the box and thoroughly mixed with the other marbles in that box.

You are told that two black marbles were drawn.

a) Given this result, what is the probability that they were drawn from Box A?

b) Given this result, which box is most likely to be the box these marbles were drawn from?

If you put work on this page, be sure to note that by the problem and to label things clearly here.