# Office Hours Questions

1. Have a question you would like answered in office hours? Add it to the list, along with your initials.
2. See a question you also have? Add your initials.
3. If you figure out your question before office hours, delete your initials.
4. If no initials remain, I'll assume no one is interested in the question anymore.

I can't promise to answer every question, but questions with more initials are more likely to be addressed. Sometimes I might include an answer or partial answer right in this document. That will depend on the question.

**New Questions:**

1. What kind of post-distribution histogram or density plot would be reasonable? Do they have to look like a normal-distribution? **A: Reasonable for what? There isn't a blanket answer to this. Certainly not all posterior distributions are normal. They could be skewed, or have heavier or lighter tails. Multimodal posterior distributions are possible, but most of the situations we have seen will produce unimodal distributions. Also, when you make a histogram, you are really looking at a marginal posterior for just one parameter. The posterior is a joint distribution over all of the parameters.**

**Answered Questions:**

1. (From 8.1e on PS8) Is there a way to isolate a certain state (in this case) to calculate the residual for it, or is it best to calculate all residuals and find the one you are looking for? smk **Answer: This depends a bit on how you are calculating residuals. If you use link(), for example, you could pass in new data just for your particular state, or a handful of states. If you are working from a posterior distribution of the parameters, you can calculate as many or few as you like.**
2. Conditional independence: is D⊥ ⊥M | A equivalent to M⊥ ⊥D | A? yh38 **Answer: Yes. Independence is symmetric.**
3. Difference of '.width = 0.95' and 'width = 0.95' in the apply() function? Yh38 **Answer: You use whatever names the function you are applying needs. That argument is called .width in mean_hdi() but could be called something else for a different function. Use R's help to see the names of the arguments for a function.**
4. How to unstandardized different parameters? Akd **Not really an Answer: This depends on the parameter and what sort of standardization was done. Note: We don't really unstandardize a parameter as much as compute a different parameter that would be related to unstandardized data. So you might have a (distribution for the) intercept of a regression line based on standardized data and you might want to**

**know what the intercept would be if we had not standardized. That's a different parameter. Example: If you have y_s = a0 + a1 x_s, you might like to know b0 and b1 so that y = b0 + b1 x. You just need to do the algebra and see how it works out. (Write the standardized forms and then rearrange in terms of the unstandardized forms of the data.)**

5. Could you please explain what are parameters in a model? Does it include sigma? **Answer: This is a little bit of a slippery question because the word parameter doesn't always get used the same way. But in the models we have seen, sigma has always been a parameter. If it has a prior (and a posterior) distribution, it is a parameter.**
6. Follow up question: how to unstandardized sigma? Can we use unstandardized() with sigma? Thank you! **Answer: This will depend on the model details. (One word answer: Algebra.) But let's suppose that you standardized your response and sigma has its usual role as in our linear models. Then that sigma is telling us about the variability of standardized responses. If we want to know the variability of responses on the natural scale, we just need to figure out what knowing the variability of (y - mean of y) / (sd of y) tells us about the variability of y. To get y from this, we need to multiply by sd of y. That multiplies the variability by sd of y. (If you multiply all values by 2, then they will be 2 times as spread out.) Then we need to shift by mean of y, but that doesn't affect the variability.**
7. How do you compute one posterior distribution from the posterior distribution of another model? **I don't understand this question.**
8. What defines a poor prediction in terms of residual?>1.1? **Answer: I think you are referring to one of the plots I made. I just wanted to highlight a few of states that were predicted least well in that data set. The cut-off of 1.1 (in absolute value) kept the plot from getting too busy. But there aren't really magic numbers here -- some things are easier to predict and you would expect smaller residuals. Some are harder to predict (ie, have more variability) and you would expect larger residuals. A physicist generally deals with much smaller residuals than a sociologist...**
9. Can you explain more about plot(coeftab(...))--I'm having trouble with how we are able to draw conclusions about each parameter (in relation to DAGs, which is spurious, which is a 'good' predictor)? **Answer: That plot shows a visual representation for the HDI of the marginal posterior for each parameter in a model. In many models, a parameter value of 0 makes a term drop out, in which case a predictor might have no impact on the predicted response. (This depends on the form of the model, of course.) In that case, we might check what the posterior tells us about what the model things about that parameter being 0. Is the model sure that that parameter is not 0? That tells us that that term can't be dropped. Are substantial portions of the posterior on each side of 0? Then the model isn't sure that the true value isn't 0.**
10. What is the preferred way of computing a 95% HDI for a parameter? Is it using link(), or is there a simpler method that just gives the numbers? Also, how do you make it on a natural scale versus not? I'm also unsure of how to find a HDI for a parameter vs. the

entire posterior. **Answer: link() and sim() are not about the distribution of a parameter, they are about the distribution of the average or individual response. To get an HDI for a parameter (or anything else, for that matter) you need its distribution. Once you have the distribution, a function like mean_hdi() will do the trick. Some functions -- like precis() -- take care of this for you. precis() reports HDIs for each of the parameters in its table. Answer 2: Finding the "natural scale for a parameter" depends on the model and on what exactly you mean. In a GLM It isn't always possible to put a parameter on the natural scale -- the parameters work together to come up with a response, and that total response can be translated to the natural scale using the inverse link function (if you have a GLM). But in some models, you can convert to the natural scale and see what that says about the parameters. If you are thinking about simple linear regression (possibly with multiple predictors, and you are worried about centering and standardizing, these transformations can often be "undone" to find out what the equivalent parameter would be on the natural scale. Answer 3: One word answer for all of this: ALGEBRA.**

11. In determining priors, are you allowed to create initial plots to help determine possible prior values, or is that cheating the process? **Answer: It is a minor cheat. When you are working with unfamiliar data, you need to have some context. Experts in that area probably don't learn anything important for selecting priors from these plots that they didn't already know. Remember, we want our priors to be "order of magnitude right" -- they don't need to be perfect. We just don't want our prior to allow things that are clearly unreasonable or to omit things that are at least reasonably plausible. Note: Standardizing is also learning from the data before you fit the model (the mean and standard deviation). Again, an expert in the data could probably tell you the right order of magnitude for the mean and standard deviation without looking at the data. Answer 2: Don't forget that priors also have other purposes, like regularization, or controlling how parameters "share information".**
12. Why does quap() need to search for the posterior mode? **Answer: That's how quap() works. It assumes the posterior is multivariate normal. That means it only needs to figure out the mode (peak) and the variances/covariances. It uses hill climbing to approximate the mode. Then it uses partial derivatives to approximate the variances and covariances. Then it uses a multivariate normal distribution that matches those values as its approximation to the posterior. This obviously works better if the posterior distribution really is (approximately) multivariate normal.**
13. When is using a lognormal distribution ideal, and what are common mu and sigma to use for them? **Answer: log-normal distributions are an example of a distribution that only produces non-negative values. Exponential, Cauchy, and half-normal are some other examples. I don't think there is an easy general purpose answer to the question about choosing mu and sigma, but you can experiment to find out what different values do to the prior for a parameter, and more importantly for the rest of the model.**

14. For 9.1e from PS09, how do we know if the results look reasonable. **Answer: The Metropolis (or HMC) algorithm is trying to (approximately) sample from the posterior, which combines information from the prior and the likelihood. So the question is whether the samples you are getting look like what the posterior should be like.**

**Answered Questions**

1. How do I ask a question? rjp
    - This question is an example of how to do it.
2. How do you find the log-normal distribution in 6.1?
3. I am struggling with Problem 7.2 on PS 7. First, I can understand conceptually how 'a' can be added to the weights instead of after multiplication, but I am unable to explain it in any meaningful way. Also, do you have any suggestions of how to alter the model without 'a'? ajr, bah
4. Can you go over link() and sim()? nwb
5. For Problem 7.3 of PS 7, I have made average and individual HDI plots for both models and graphed them side by side. I am still unsure of how to determine which is a better model. Could you explain in general and perhaps specifically for this problem how to determine effectiveness of models? smk, ca, ajr, akd
6. Could you please go through an example on how to do proper model comparison?