

Tutorial 8

Statistical Computation and Analysis
Spring 2025

Tutorial Outline

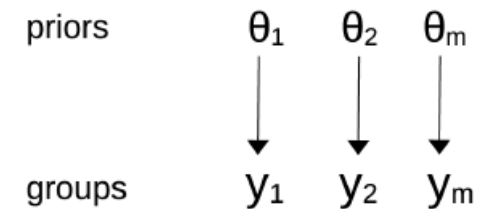
- Hierarchical models
 - Shrinkage
- Multiple Regression

Hierarchical models

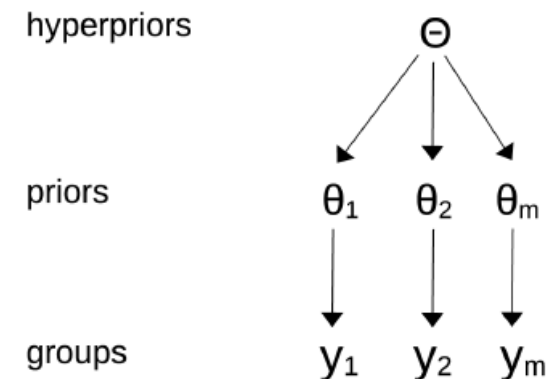
- Until now, we either had one group with a bunch of data (coin flips, number of email in the class, etc.)
- Or, we had separate groups (tips / bills on separate days).
- Sometimes, we would like to have separate groups that share information.
 - Different classes that all belong to the same school.
 - Different patients that are all in the same hospital.

Hierarchical models

- Does not allow sharing information.



- Allows for sharing information between the groups, while also allowing the separate groups.



Hierarchical models

- In hierarchical models, the parameters of the prior distributions are themselves given a prior distribution.
- These higher-level priors are often called hyperpriors.
 - The parameters of the prior distributions belong to a common population of parameters.

Hierarchical models

- We want to analyze the water quality in a city comprised of neighborhoods.
- Option 1: Analyze the water quality in each neighborhood separately.
 - We obtain a more detailed view of the problem and averaging can cause us to lose information.
- Option 2: Combine the data from all the neighborhoods.
 - We obtain a bigger sample size and hence a more accurate estimation.

Hierarchical models

- Option 3: Hierarchical model.
- Water samples from three neighborhoods.
- If the lead content is above a threshold – the water is contaminated.
- **For each neighborhood, we have 30 samples, where the value of the sample is 0 if it's contaminated and 1 if it is good.**

Hierarchical models

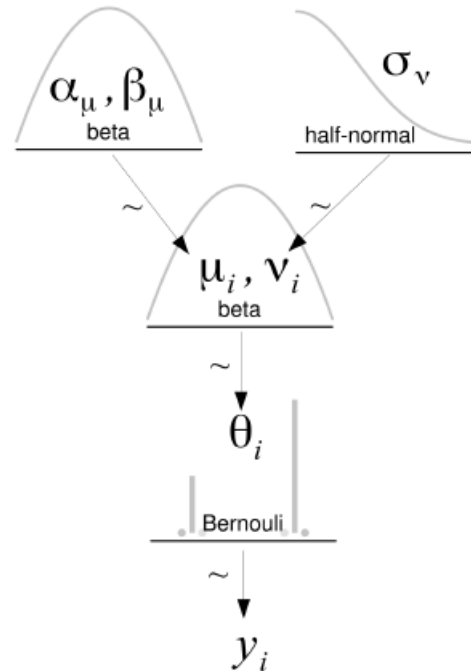
- We are measuring the chance for a contaminated sample, or not -> binomial distribution.
- We will use a hierarchical model to share information between the neighborhoods.

$$\mu \sim \text{Beta}(\alpha_\mu, \beta_\mu)$$

$$\nu \sim \text{HN}(\sigma_\nu)$$

$$\theta_i \sim \text{Beta}(\mu, \nu)$$

$$y_i \sim \text{Bernoulli}(\theta_i)$$



Note: we are using the mean and precision (inverse of the standard deviation) in the beta prior.

Hierarchical models

```
coords = {"neighborhood": np.arange(len(N_samples)), "neighborhood_data": group_idx}

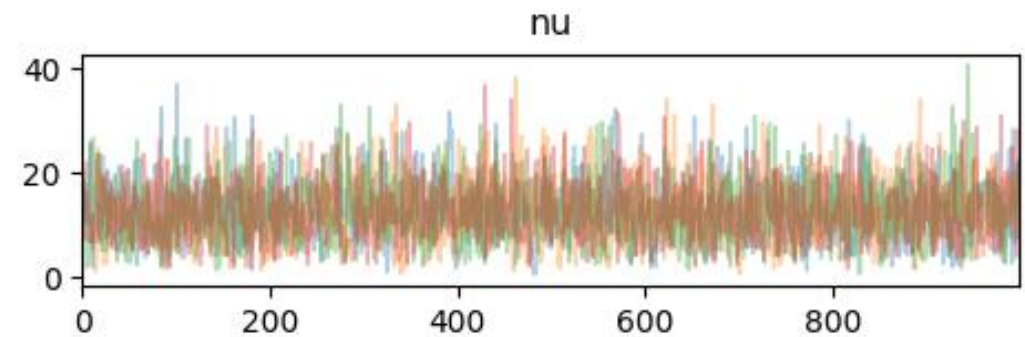
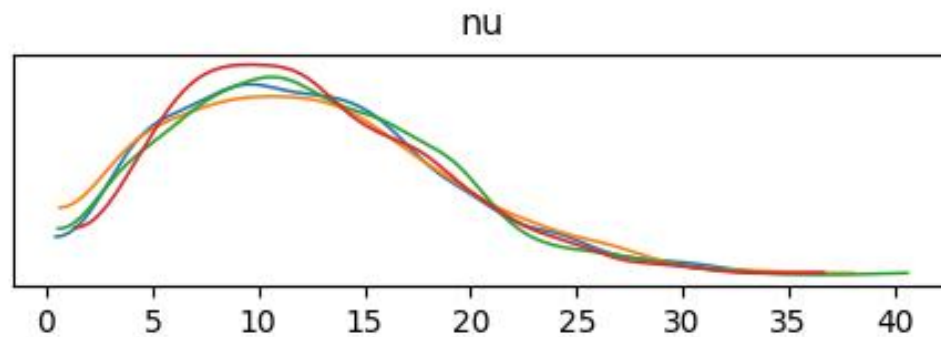
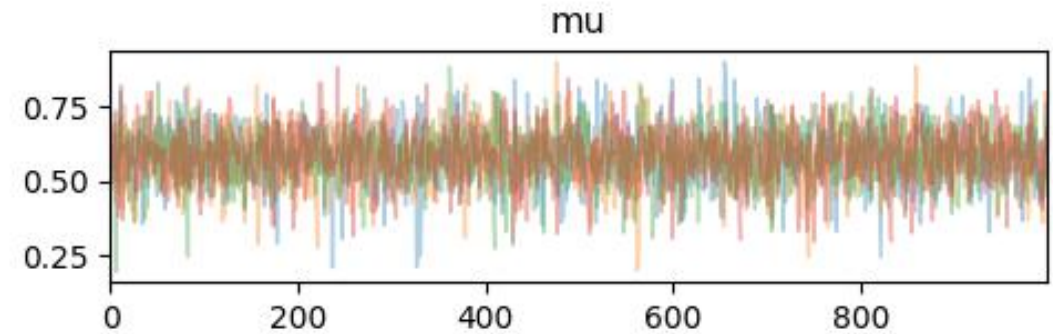
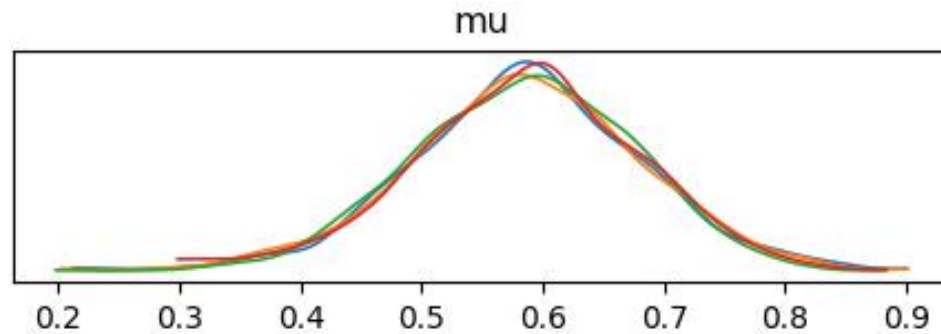
with pm.Model(coords = coords) as model_h1:
    # hyperpriors
    mu = pm.Beta('mu', 1, 1)
    nu = pm.HalfNormal('nu', 10)
    # prior
    theta = pm.Beta('theta', mu=mu, nu=nu, dims = "neighborhood")
    # likelihood
    y = pm.Bernoulli('y', p=theta[group_idx], observed = data, dims = "neighborhood_data")
```

Hierarchical models

- We are going to run the model on three versions of the data:
 1. There are 18/30 contaminated water samples in all three neighborhoods.
 2. There are 3/30 contaminated water samples in all three neighborhoods.
 3. There are 3/30 contaminated water samples in two neighborhoods and 18/30 in one.

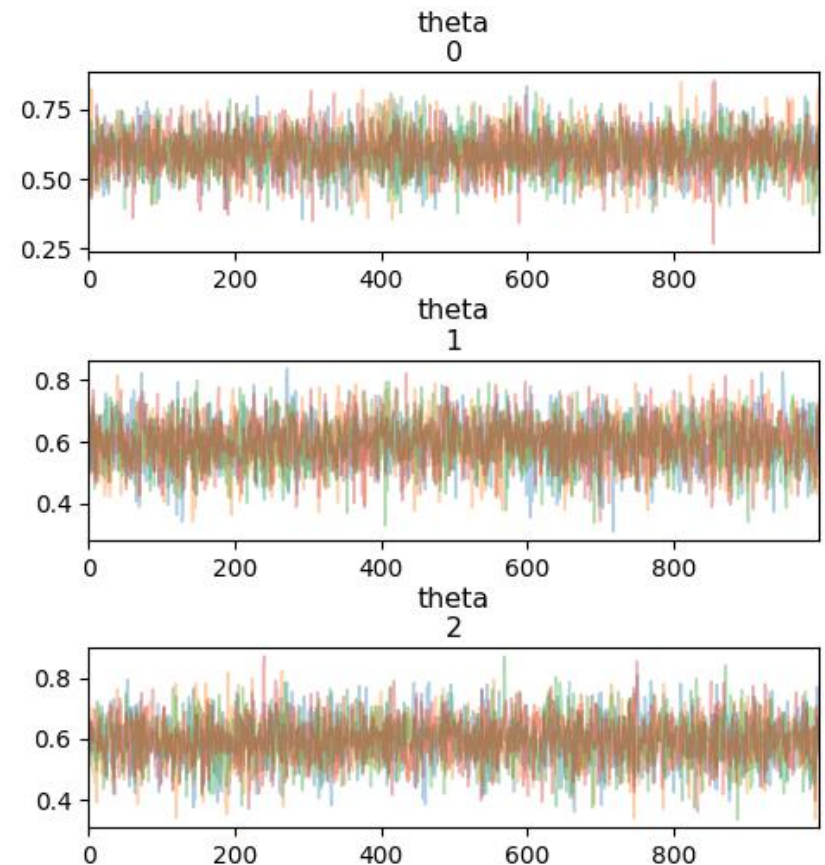
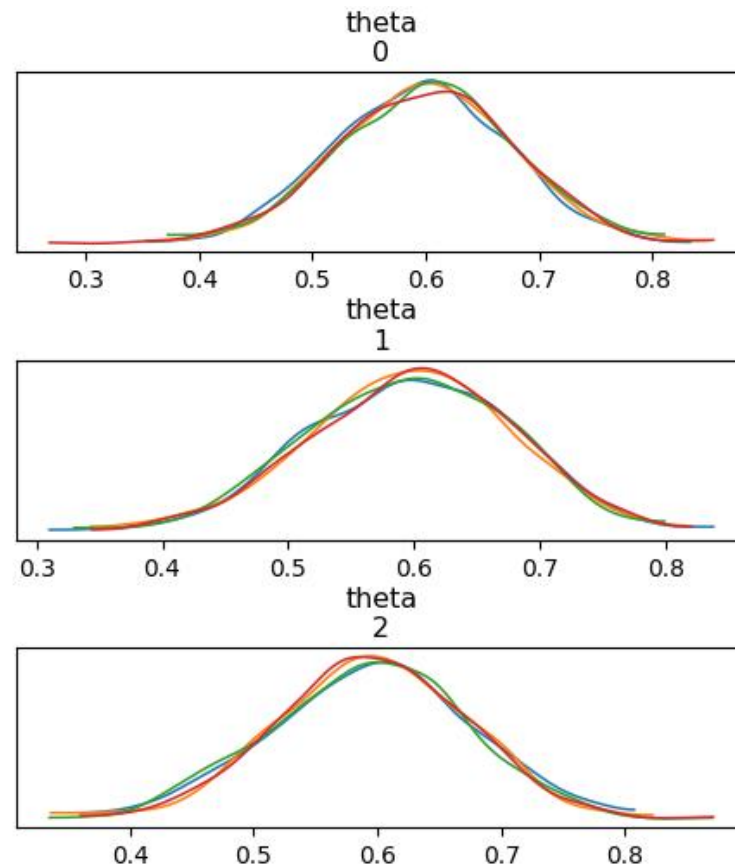
Hierarchical models

- Examine the posterior distributions for the prior parameters (version 1).



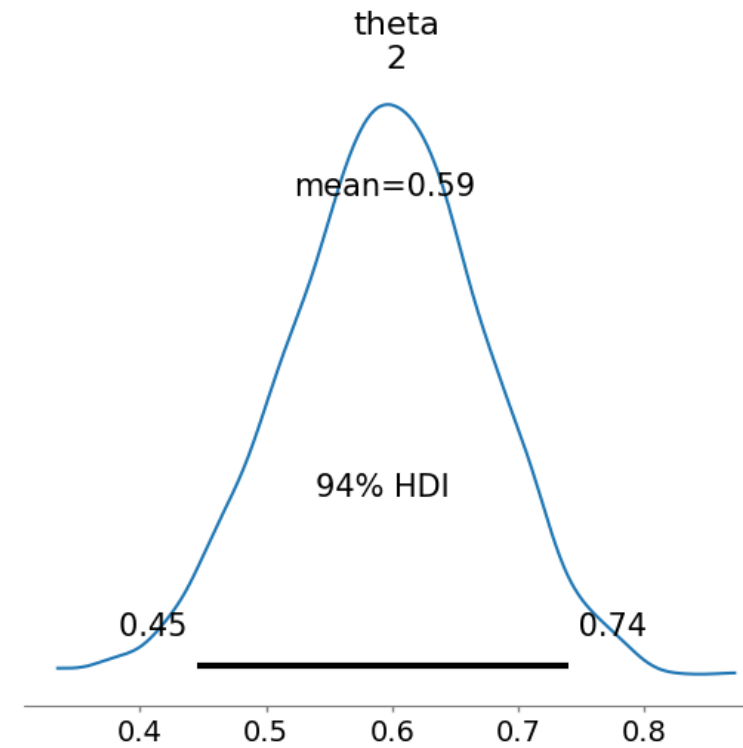
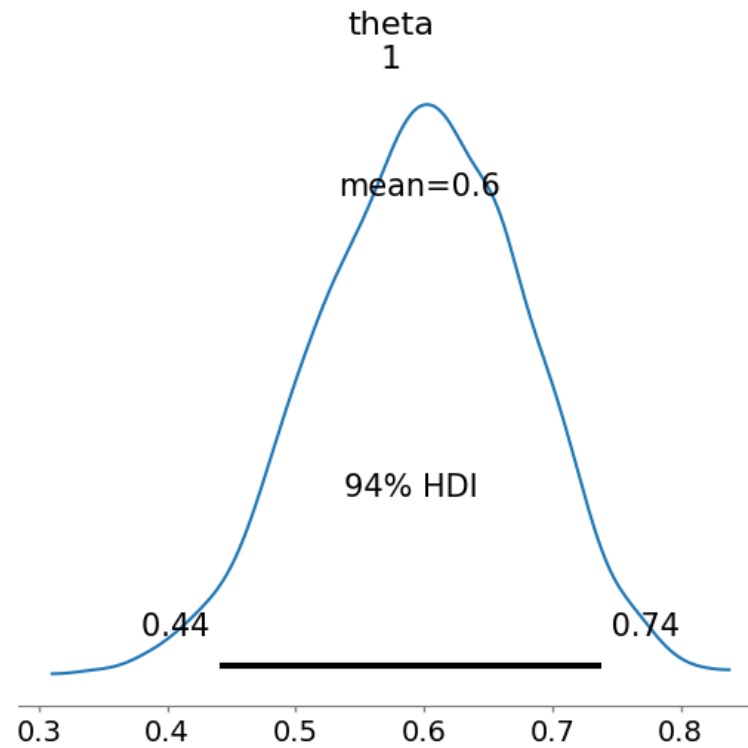
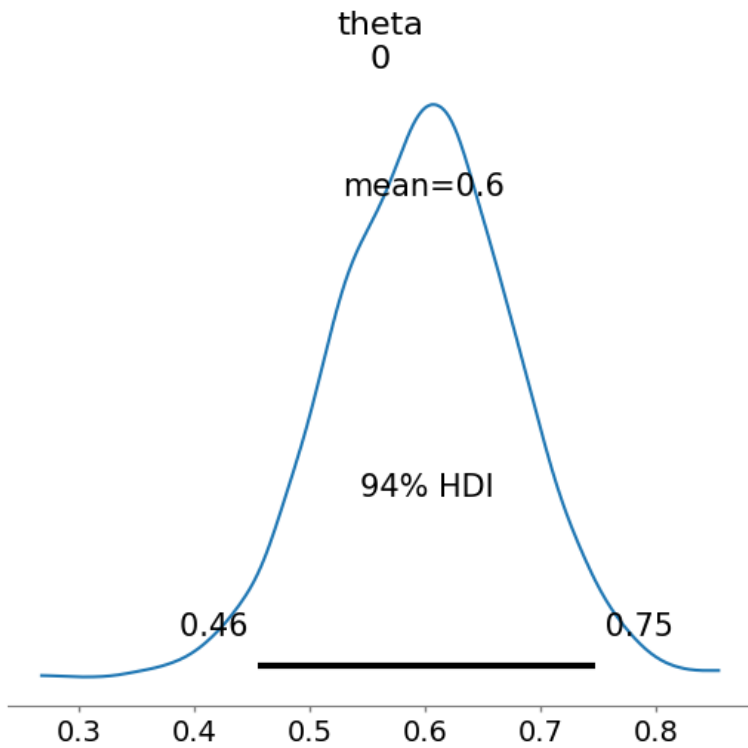
Hierarchical models

- For the theta
 - They all look very similar, which is expected as the data for the three neighborhoods was the same.

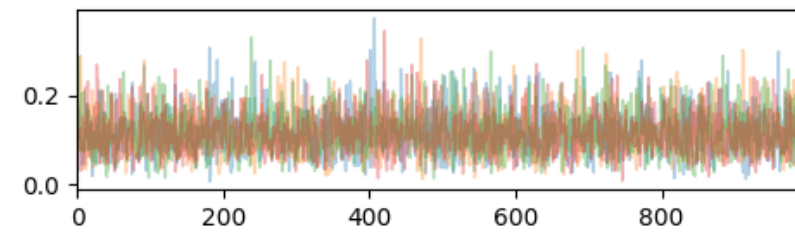
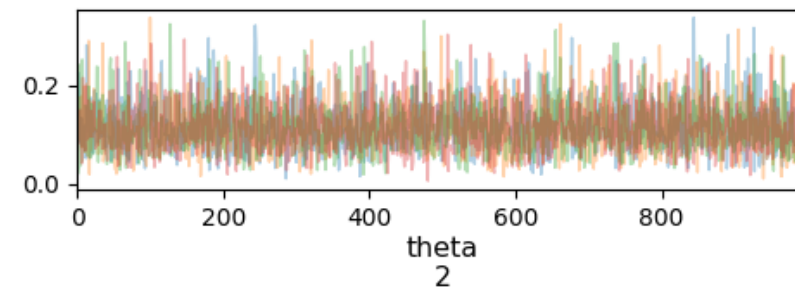
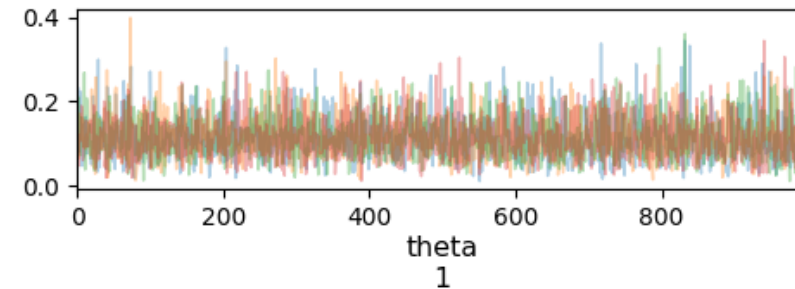
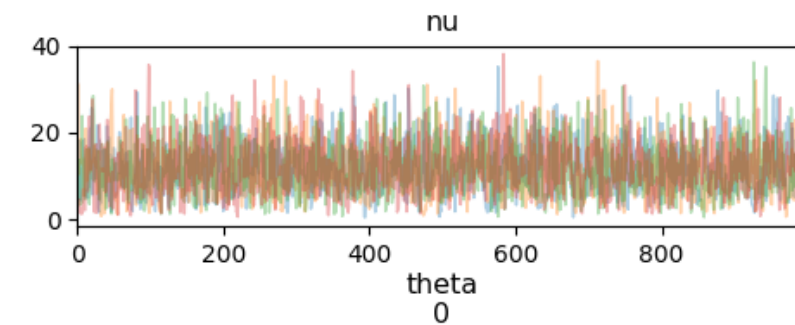
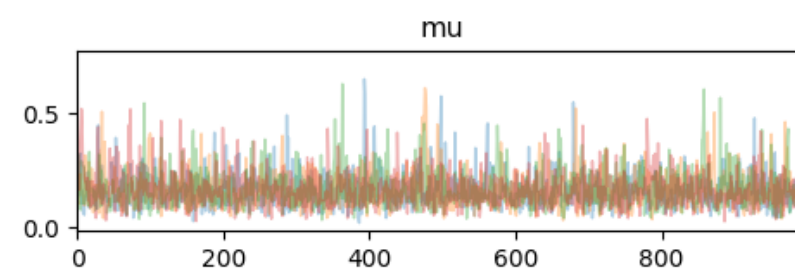
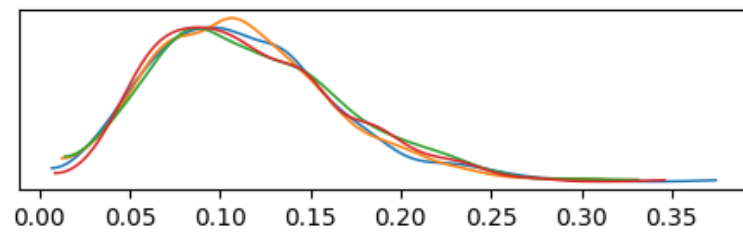
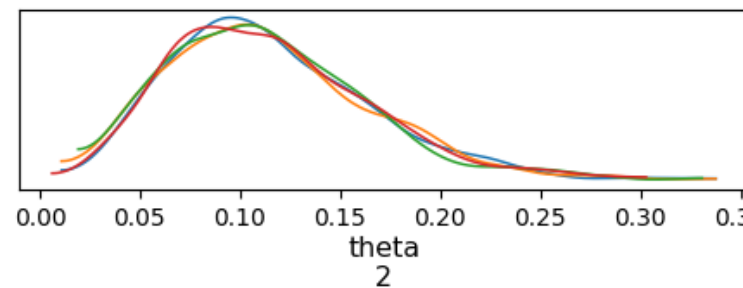
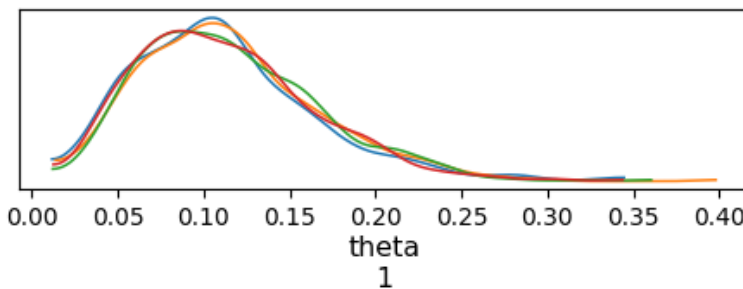
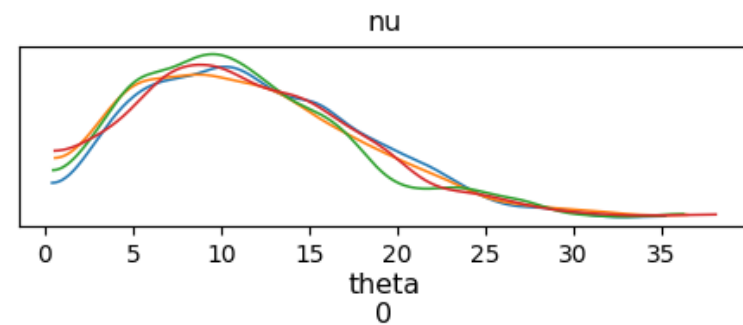
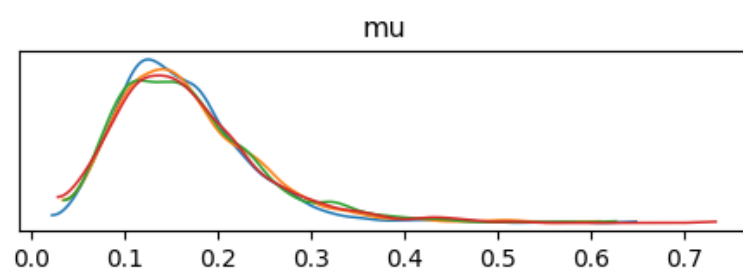


Hierarchical models

- Mean and HDI

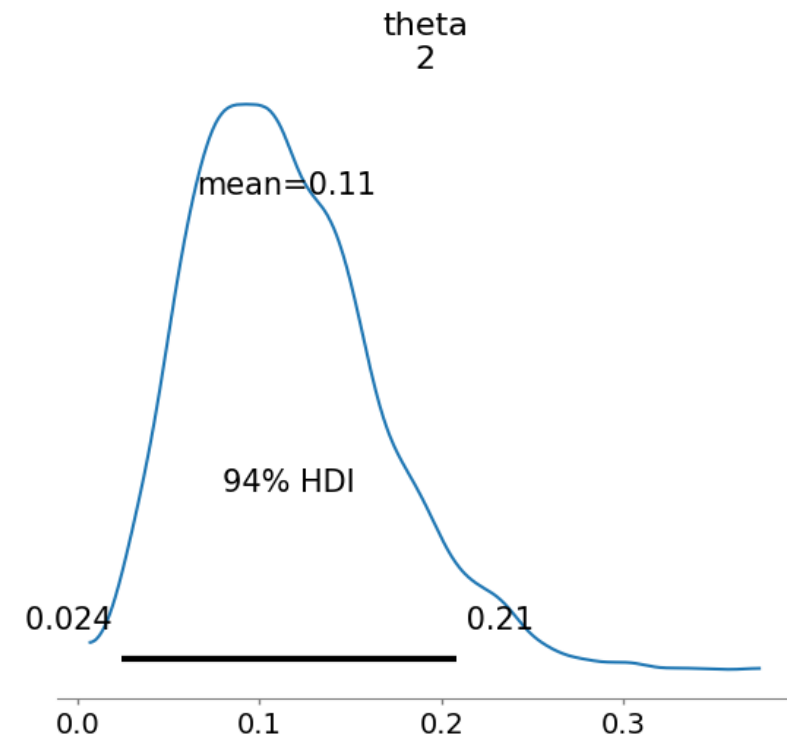
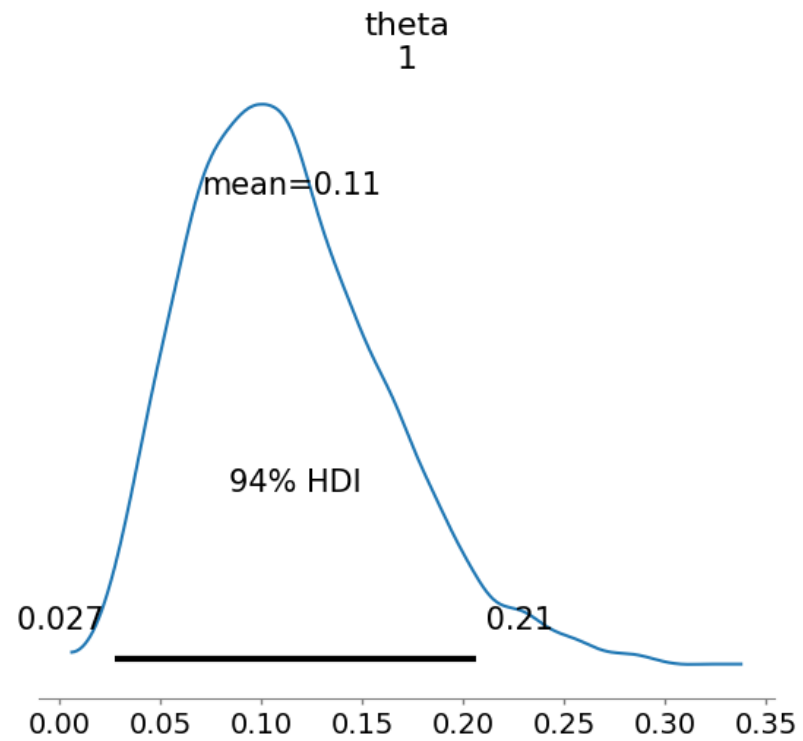
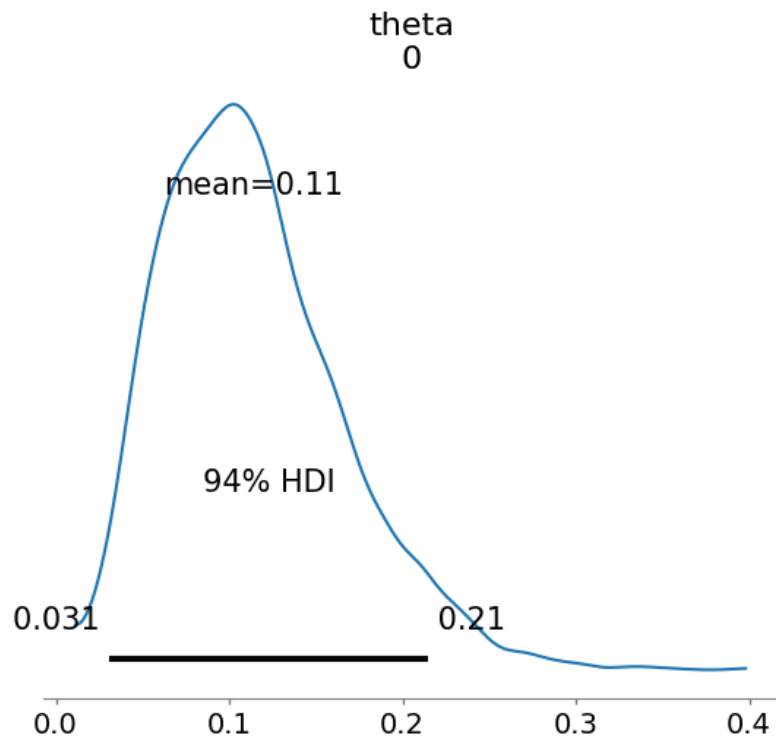


■ Version 2:

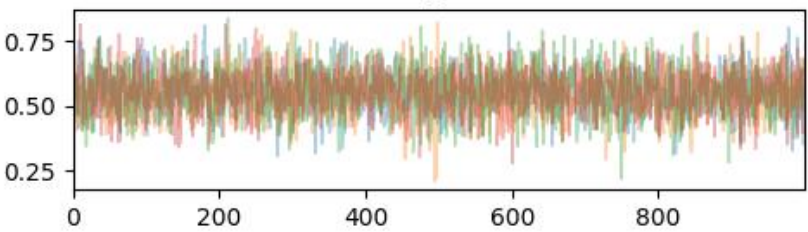
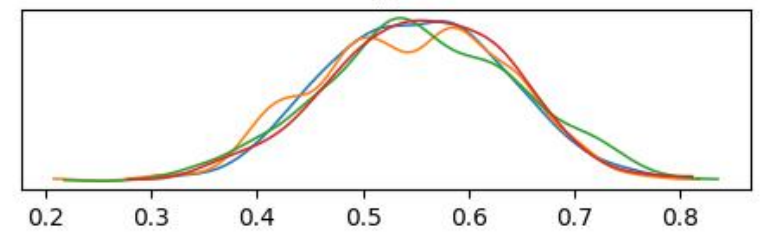
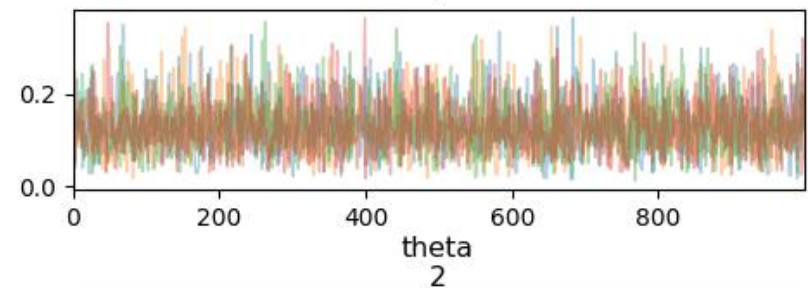
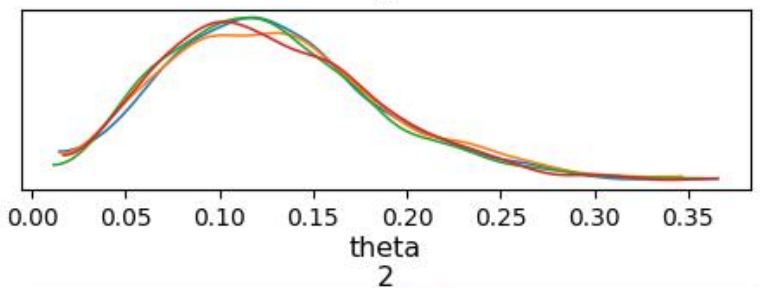
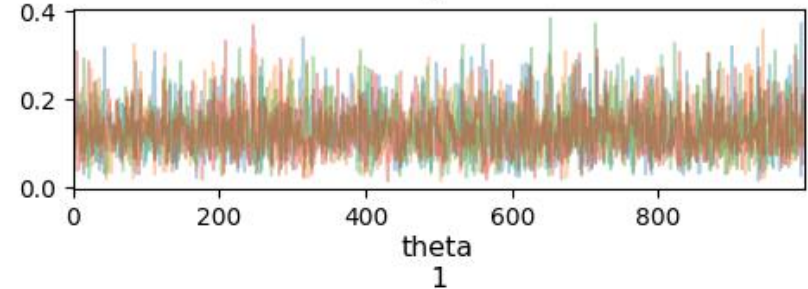
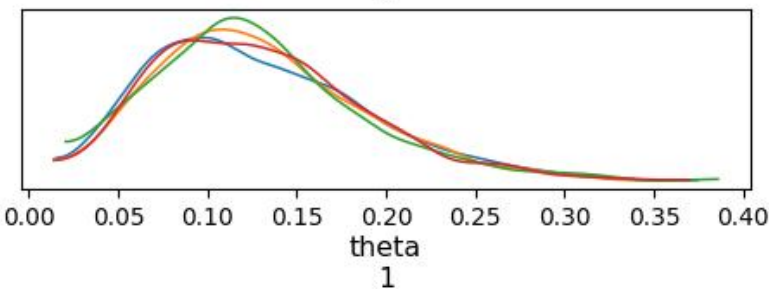
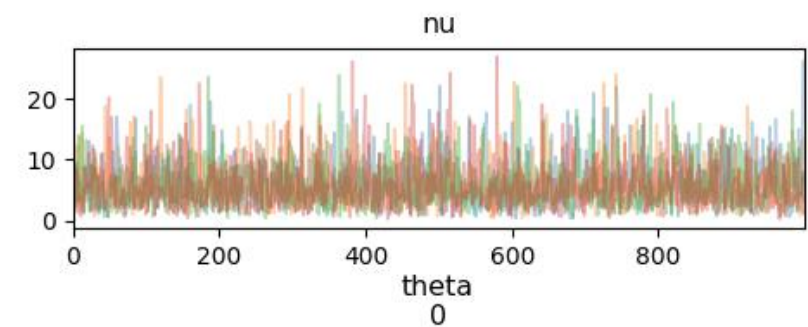
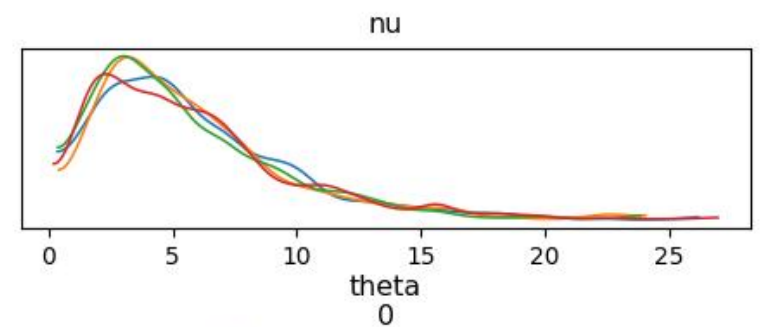
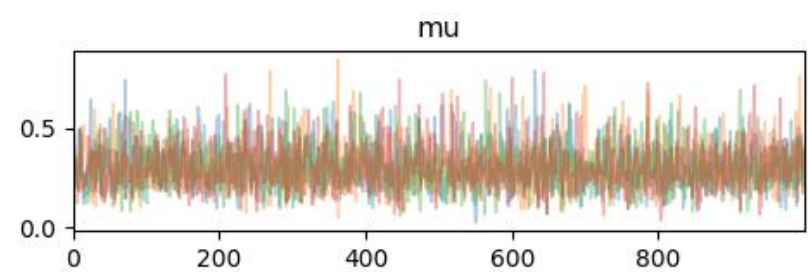
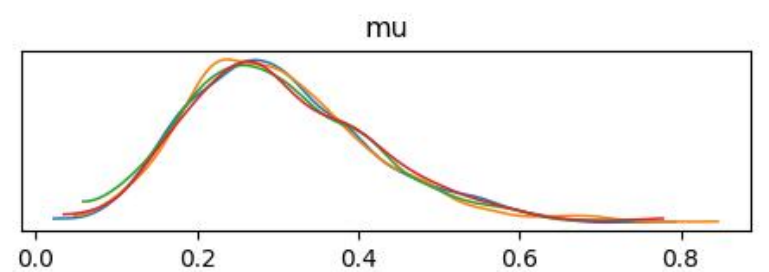


Hierarchical models

- Mean and HDI

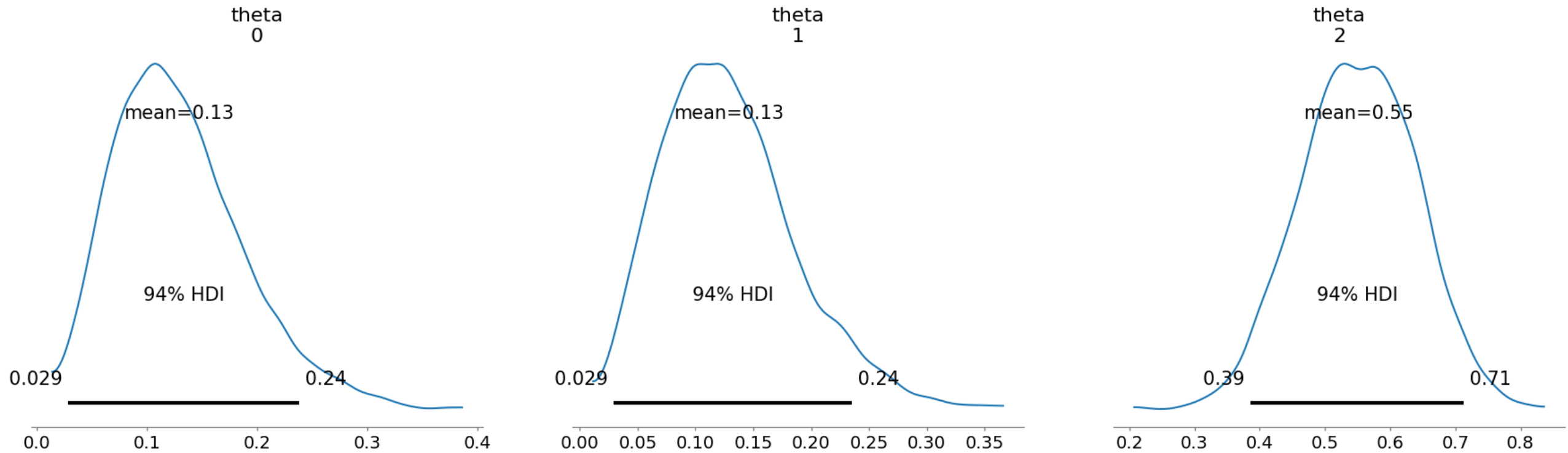


■ Version 3:



Hierarchical models

- Mean and HDI



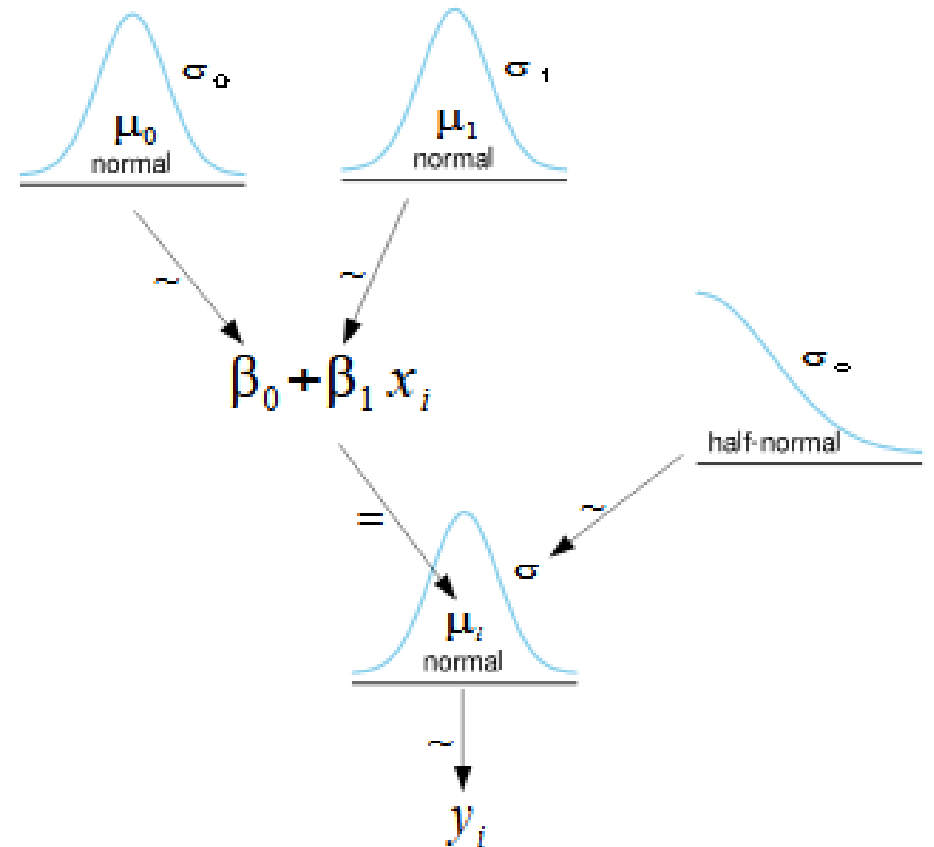
Shrinkage

- In version 3, we would expect to see mean values of **0.11** for the neighborhoods with **three** contaminated samples and **0.6** for the neighborhood with **18** contaminated samples.
- Instead, we got mean values of **0.13** for the neighborhoods with **three** contaminated samples and **0.55** for the neighborhood with **18** contaminated samples.
- This is called **shrinkage**.
- The estimates have shrunk toward the common mean due to each group informing the rest, and each group being informed by the estimation of the others.

Simple Linear Regression

- Review: We began to look at a case in which the mean depends on another variable.
 - Average height as a function of age.

$$y_i \sim N(\mu_i, \sigma)$$
$$\mu_i = \beta_0 + \beta_1 x_i$$
$$\beta_0 \sim \text{Prior0}(\theta_0)$$
$$\beta_1 \sim \text{Prior1}(\theta_1)$$
$$\sigma \sim \text{Prior2}(\theta_\sigma)$$



Multiple Regression

- We saw the case in which there is one dependent variable and one independent variable.
- There can also be more than one independent variable.
 - For example, height as a function of both weight and age.
 - In this case: $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$

Multiple Regression

- Radon is a radioactive gas that can cause lung cancer and thus it is something that would be undesirable in a domicile.
 - The presence of a basement may increase the radon levels in a household as radon may enter the household more easily through the ground.
 - We have a dataset of the radon levels at homes in Minnesota as well as the county of the home, and the presence of a basement

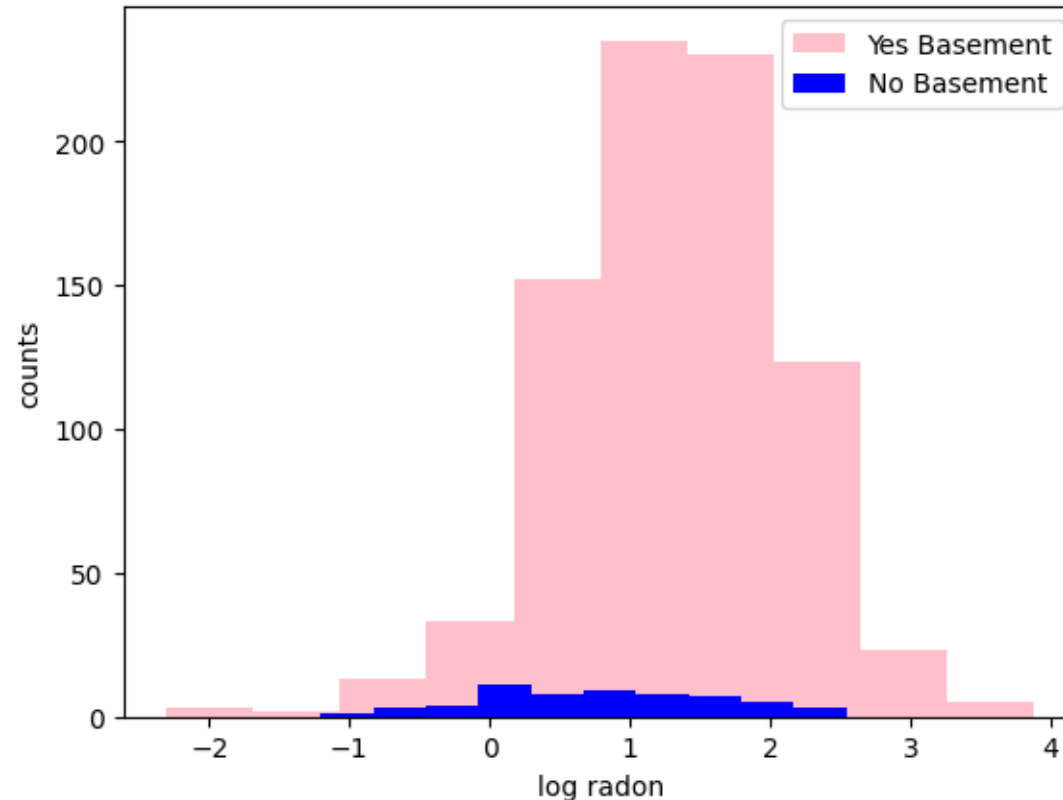
Multiple Regression

■ Load data

	Unnamed: 0	idnum	state	state2	stfips	zip	region	typebldg	floor	room	...	pcterr	adjwt	dupflag	zipflag	cntyfips	county	fips	Uppm	county_code
0	0	5081.0	MN	MN	27.0	55735	5.0	1.0	1.0	3.0	...	9.7	1146.499190	1.0	0.0	1.0	AITKIN	27001.0	0.502054	0
1	1	5082.0	MN	MN	27.0	55748	5.0	1.0	0.0	4.0	...	14.5	471.366223	0.0	0.0	1.0	AITKIN	27001.0	0.502054	0
2	2	5083.0	MN	MN	27.0	55748	5.0	1.0	0.0	4.0	...	9.6	433.316718	0.0	0.0	1.0	AITKIN	27001.0	0.502054	0
3	3	5084.0	MN	MN	27.0	56469	5.0	1.0	0.0	4.0	...	24.3	461.623670	0.0	0.0	1.0	AITKIN	27001.0	0.502054	0
4	4	5085.0	MN	MN	27.0	55011	3.0	1.0	0.0	4.0	...	13.8	433.316718	0.0	0.0	3.0	ANOKA	27003.0	0.428565	1
...
914	914	5995.0	MN	MN	27.0	55363	5.0	1.0	0.0	4.0	...	4.5	1146.499190	0.0	0.0	171.0	WRIGHT	27171.0	0.913909	83
915	915	5996.0	MN	MN	27.0	55376	5.0	1.0	0.0	7.0	...	8.3	1105.956867	0.0	0.0	171.0	WRIGHT	27171.0	0.913909	83
916	916	5997.0	MN	MN	27.0	55376	5.0	1.0	0.0	4.0	...	5.2	1214.922779	0.0	0.0	171.0	WRIGHT	27171.0	0.913909	83
917	917	5998.0	MN	MN	27.0	56297	5.0	1.0	0.0	4.0	...	9.6	1177.377355	0.0	0.0	173.0	YELLOW MEDICINE	27173.0	1.426590	84
918	918	5999.0	MN	MN	27.0	56297	5.0	1.0	0.0	4.0	...	8.0	1214.922779	0.0	0.0	173.0	YELLOW MEDICINE	27173.0	1.426590	84

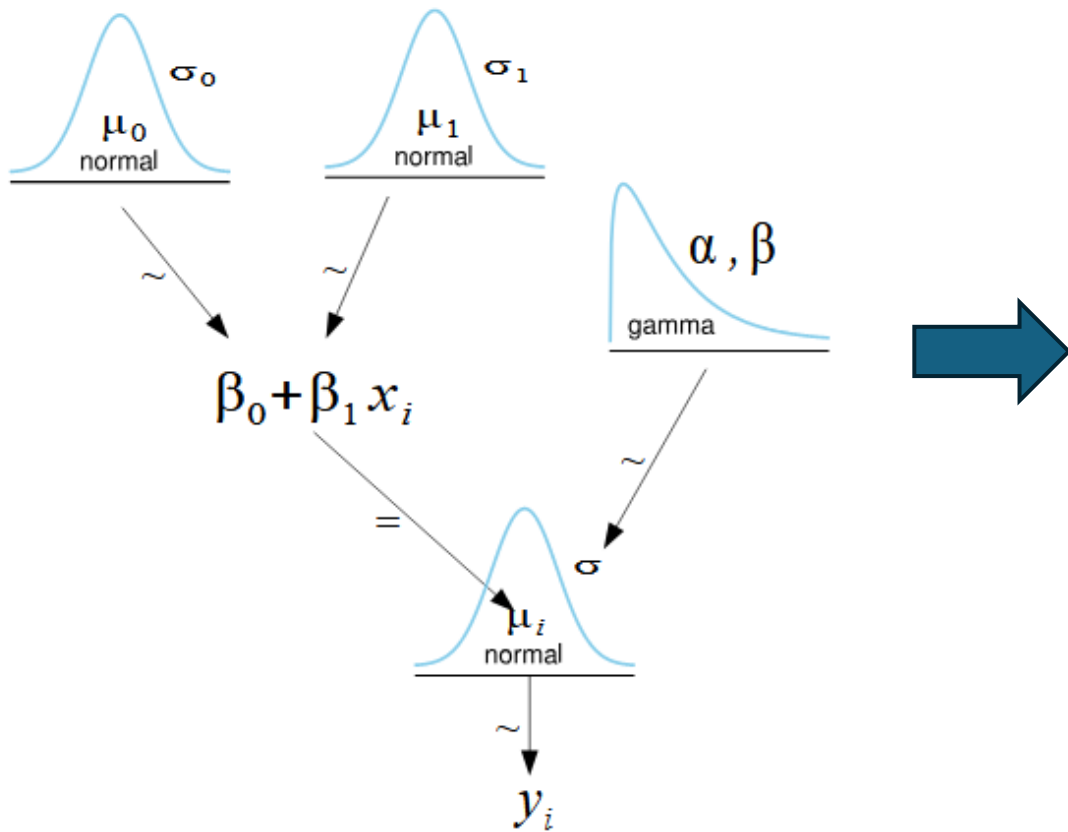
Simple Regression

- Start with assessing the effect only of the presence of a basement on the radon levels.



Simple Regression

- Define our model



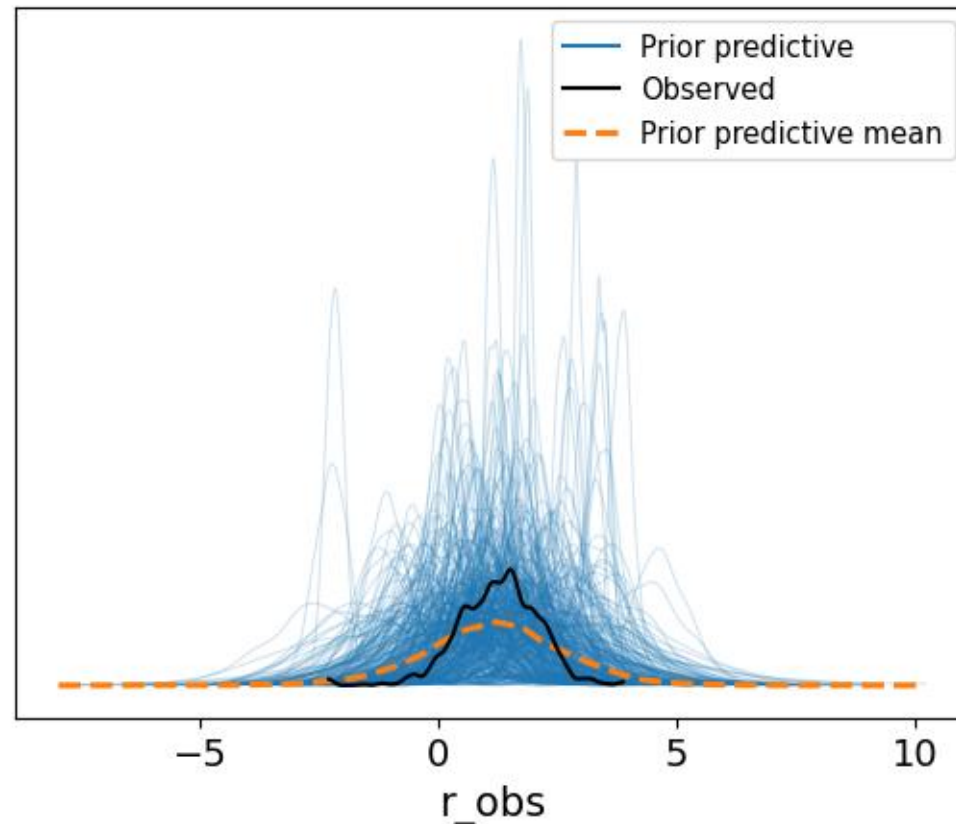
```
c_basement = {"basement": b_codes, "data": np.arange(len(r))}

with pm.Model(coords=c_basement) as m_basement:
    b_i = pm.Data('b_i', b_idx, dims="data")
    beta_0 = pm.Normal('beta_0', mu=mu_0, sigma=sigma_0)
    beta_b = pm.Normal('beta_b', mu=0, sigma=sigma_b, dims="basement")

    mu_r = pm.Deterministic('mu_r', beta_0 + beta_b[b_i], dims="data")
    sigma_r = pm.Gamma('sigma_r', mu=mu_sigma_r, sigma=sigma_sigma_r)
    r_obs = pm.Normal('r_obs', mu=mu_r, sigma=sigma_r, observed=r, dims="data")
```


Simple Regression

- Prior predictive check
 - Our priors look reasonable given the sampled datasets
 - Now we can sample



Simple Regression

■ Our inference data object

id_basement

arviz.InferenceData









▼ posterior

xarray.Dataset

► Dimensions: (chain: 4, draw: 1000, basement: 2, data: 878)

► Coordinates: (4)

▼ Data variables:

β_0	(chain, draw)	float64	0.4527 1.27 1.028 ... 0.8074 0.8166		
β_b	(chain, draw, basement)	float64	0.522 0.8353 ... 0.05146 0.501		
μ_r	(chain, draw, data)	float64	0.9748 1.288 1.288 ... 1.318 1.318		
σ_r	(chain, draw)	float64	0.8306 0.8318 ... 0.7977 0.796		

► Indexes: (4)

► Attributes: (6)

► sample_stats

► prior

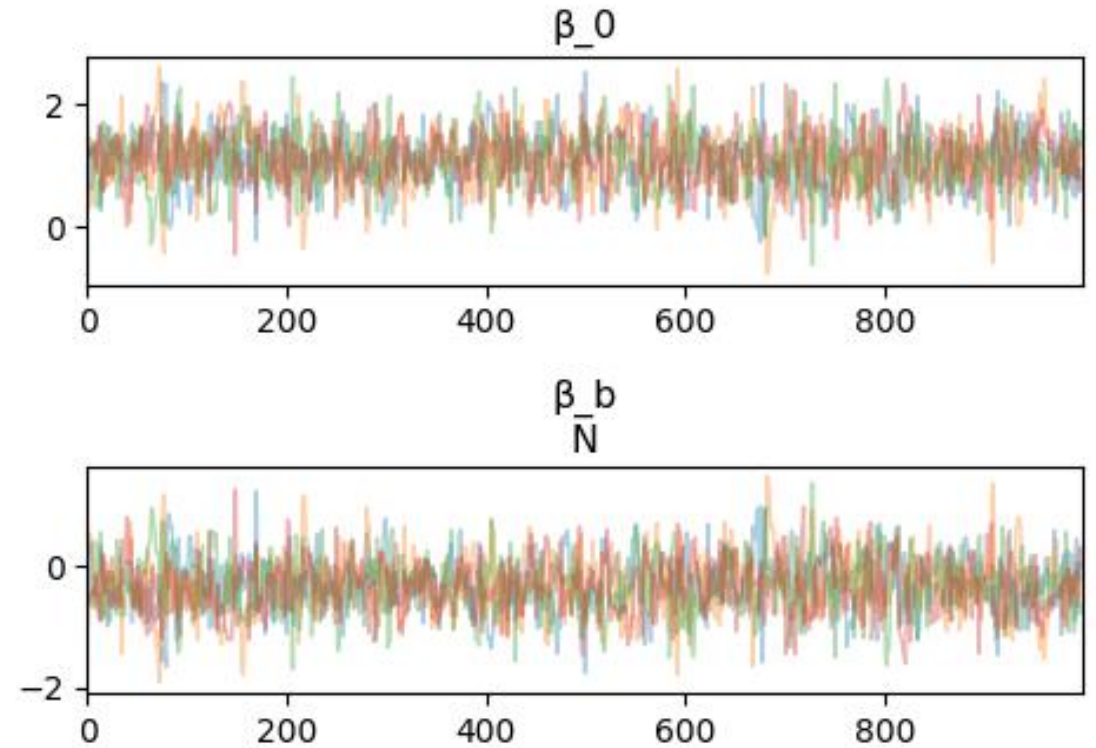
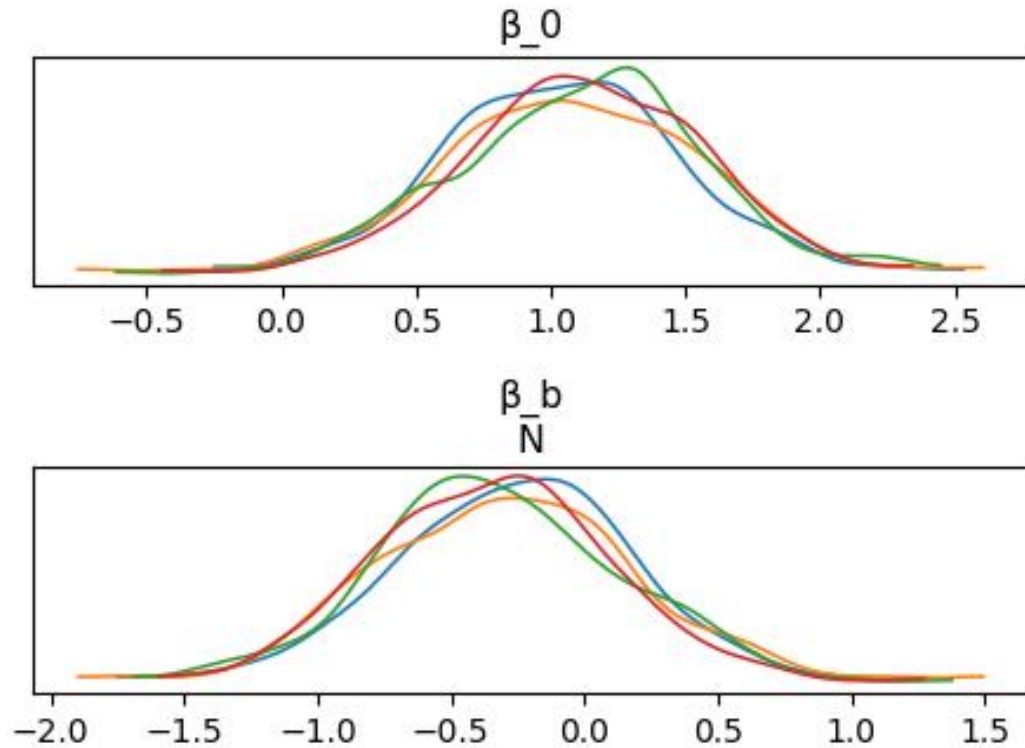
► prior_predictive

► observed_data

► constant_data

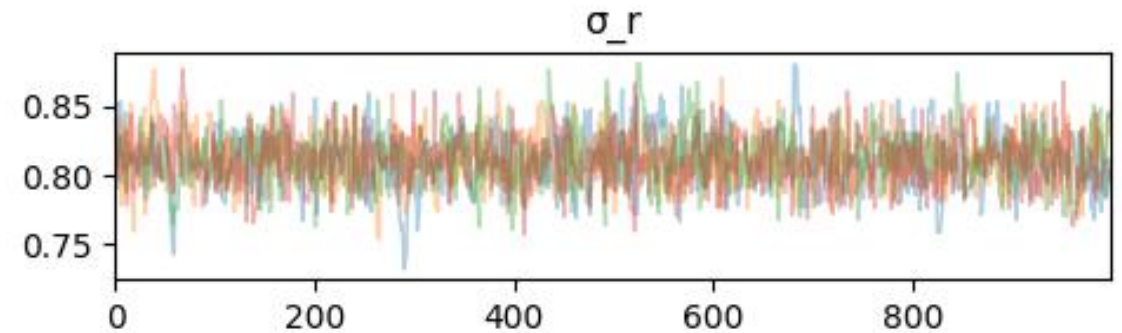
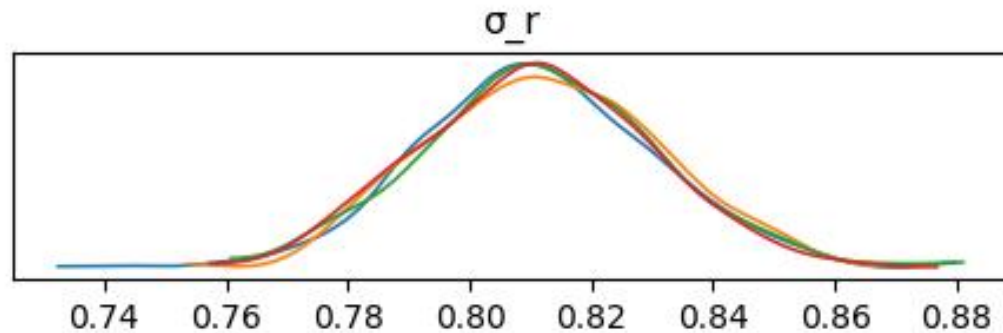
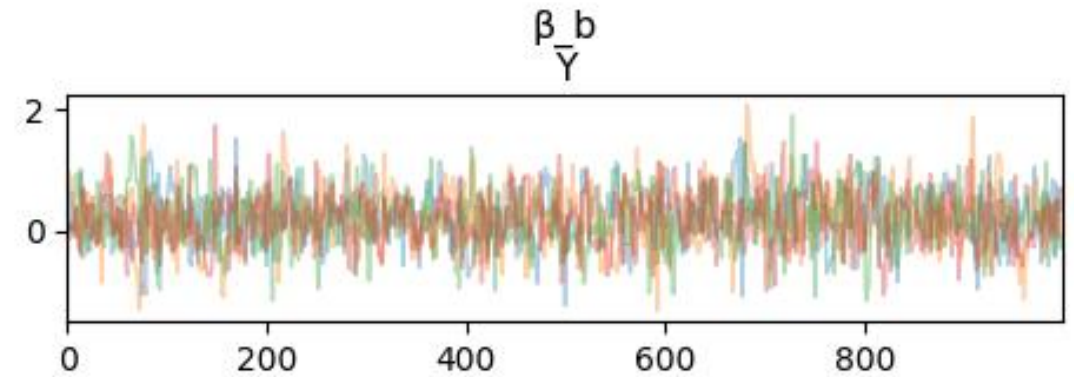
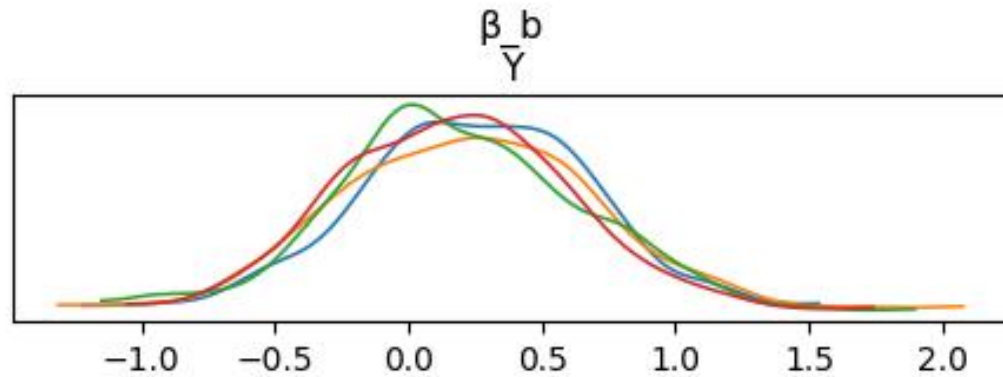
Simple Regression

- Examine our sampling and posterior



Simple Regression

- Examine our sampling and posterior



- We can see that the four chains overlap.

Simple Regression

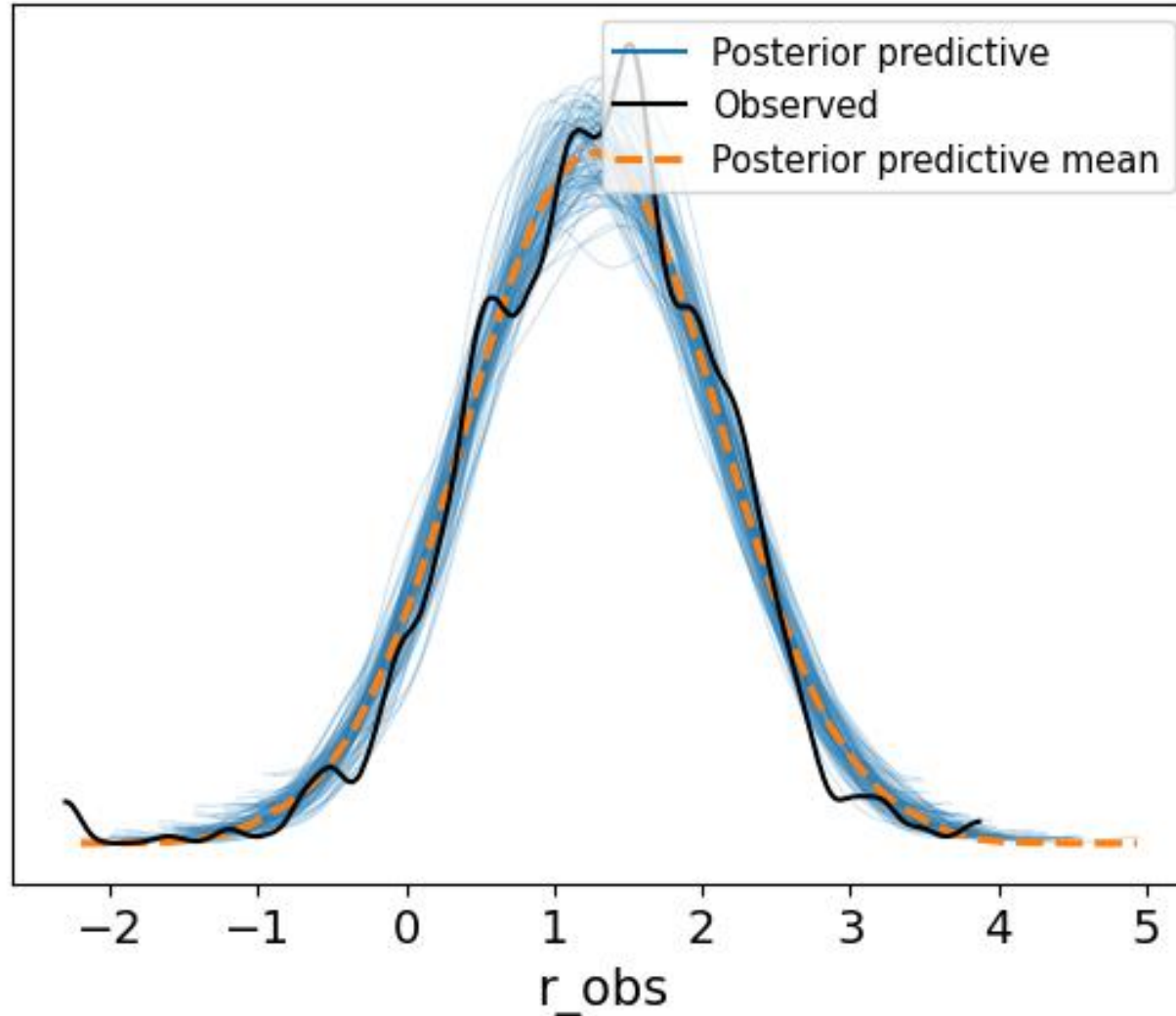
- And diagnostics
 - They fulfil the criteria we learned

	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
β_0	0.014	0.011	1126.0	1365.0	1.0
$\beta_b[N]$	0.014	0.010	1149.0	1415.0	1.0
$\beta_b[Y]$	0.014	0.011	1121.0	1385.0	1.0
σ_r	0.001	0.000	1392.0	1371.0	1.0



Simple Regression

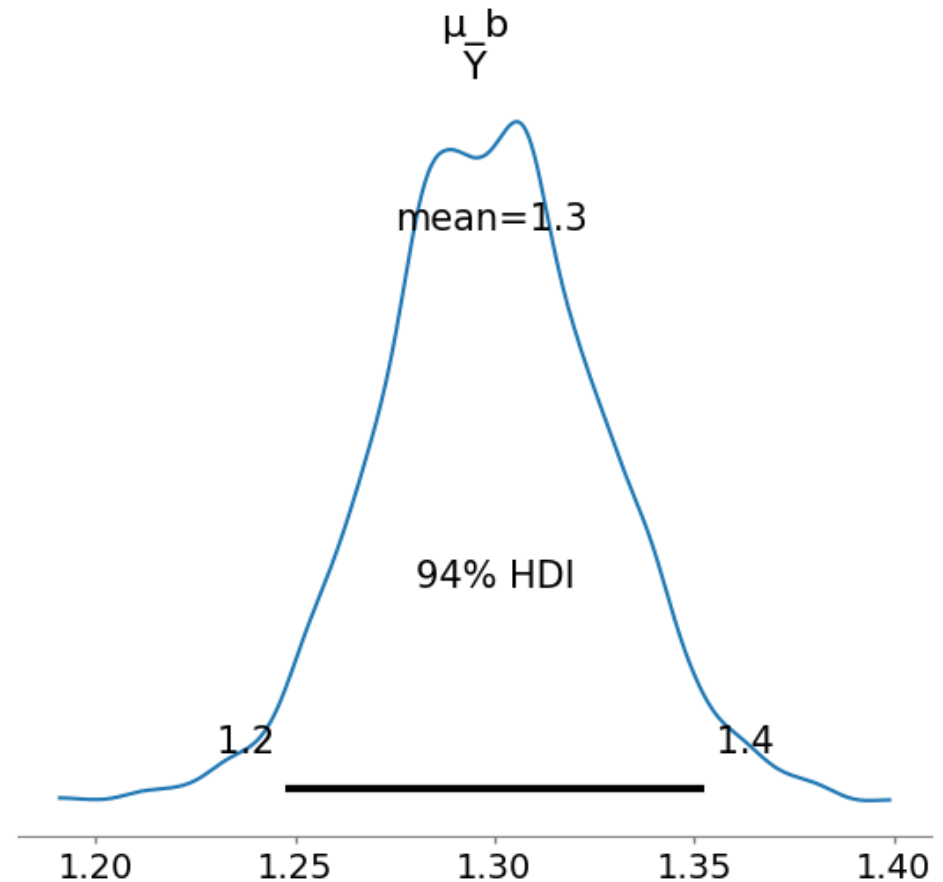
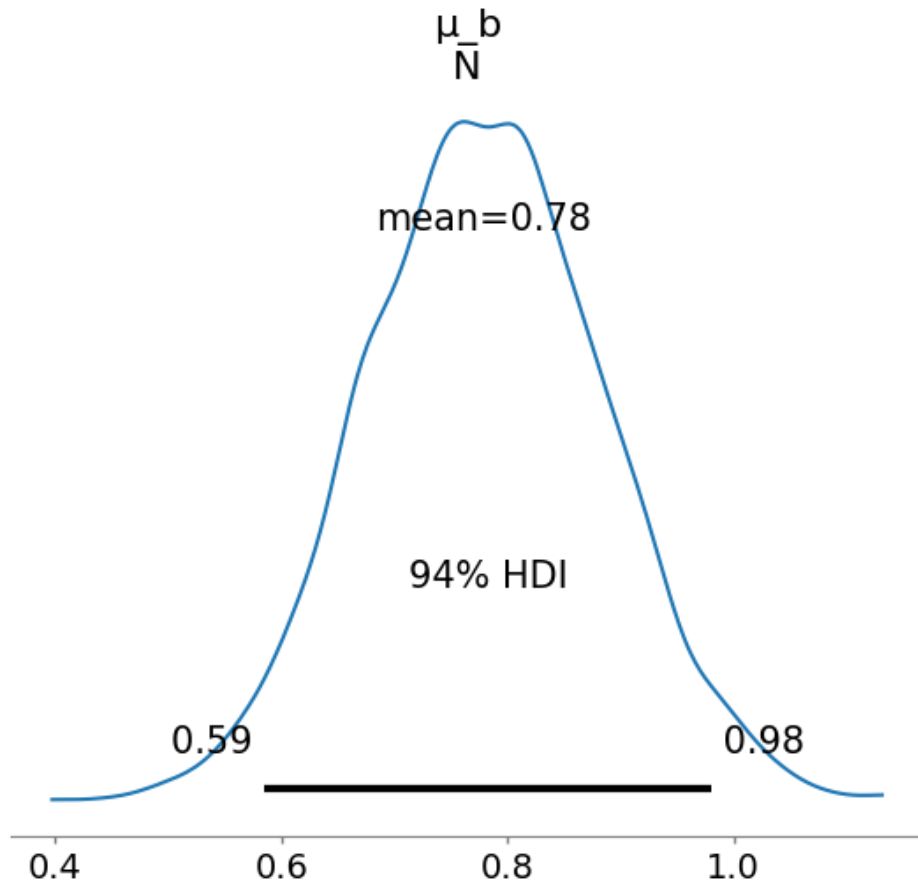
- Posterior Predictive Checks



Simple Regression

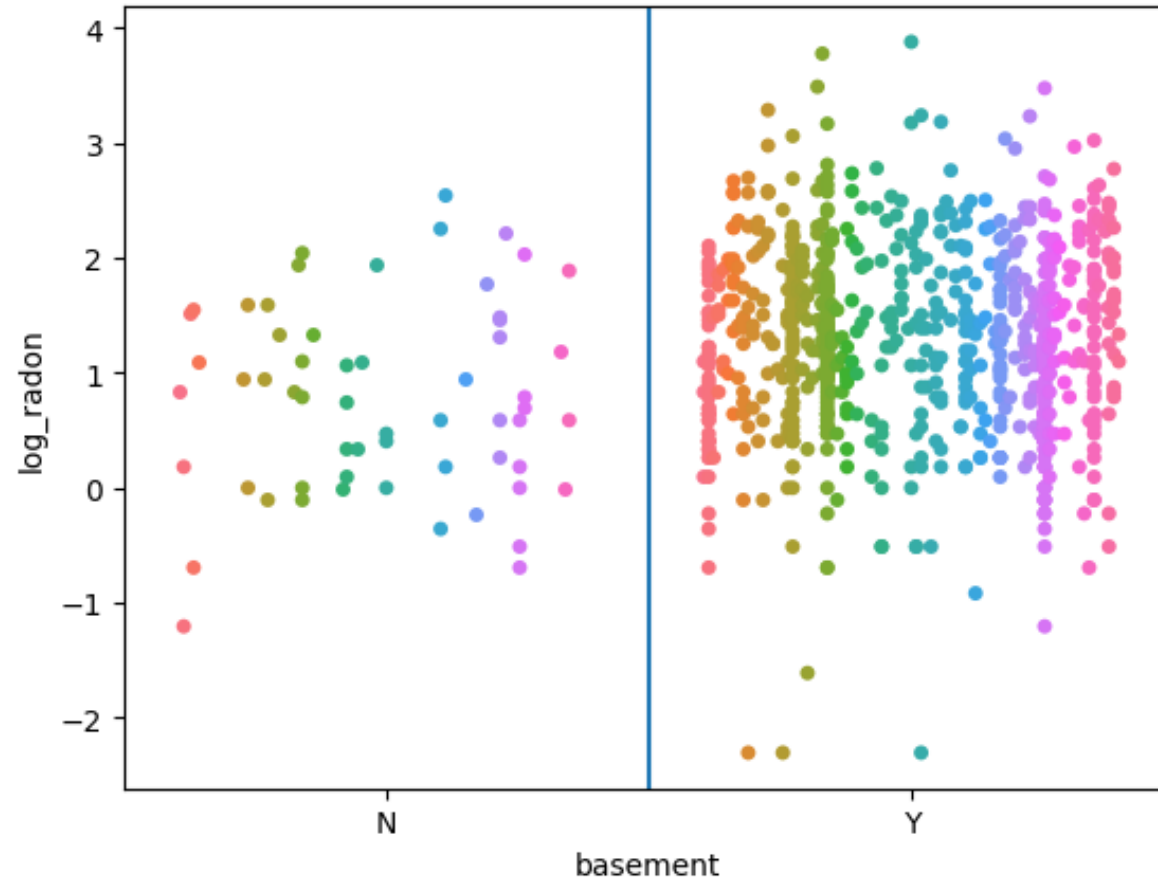
- Posterior Inference

- The appears to be less radon in houses without basements.



Multiple Regression

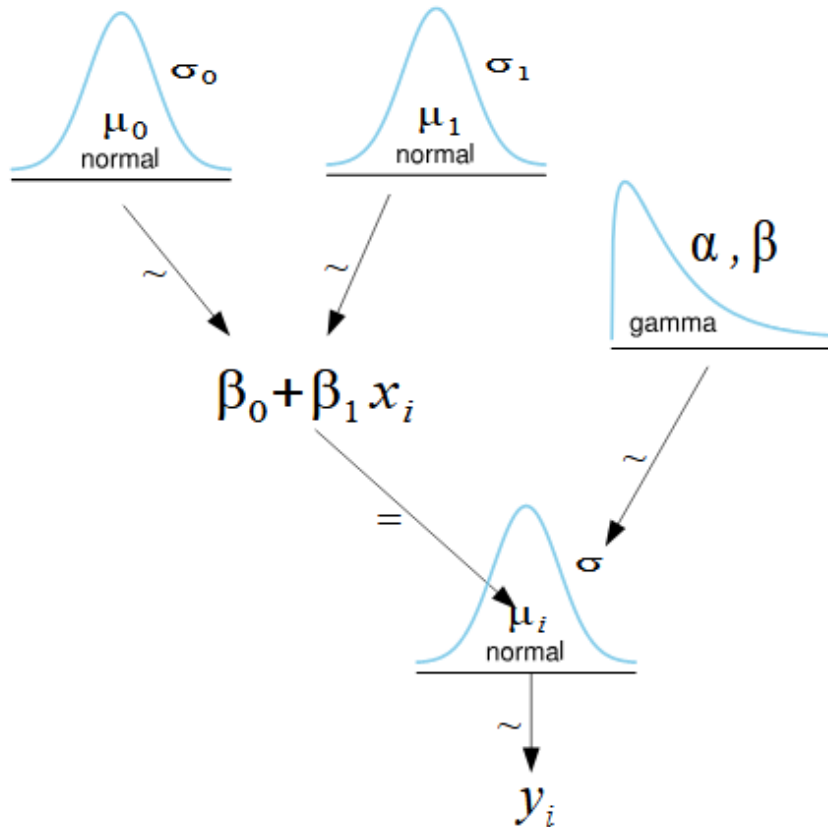
- Now let's add the effect of the county.
- Examine data (colors show the different counties)



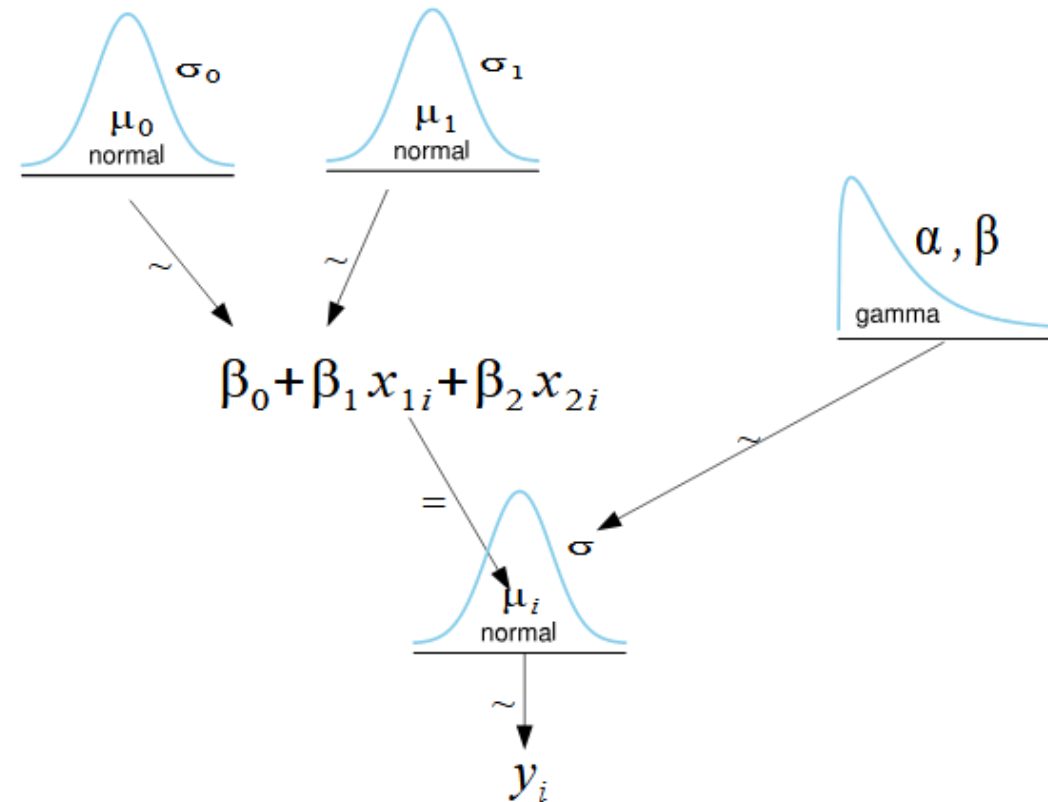
Multiple Regression

- Create model for the multiple regression.

What we had for only the basement

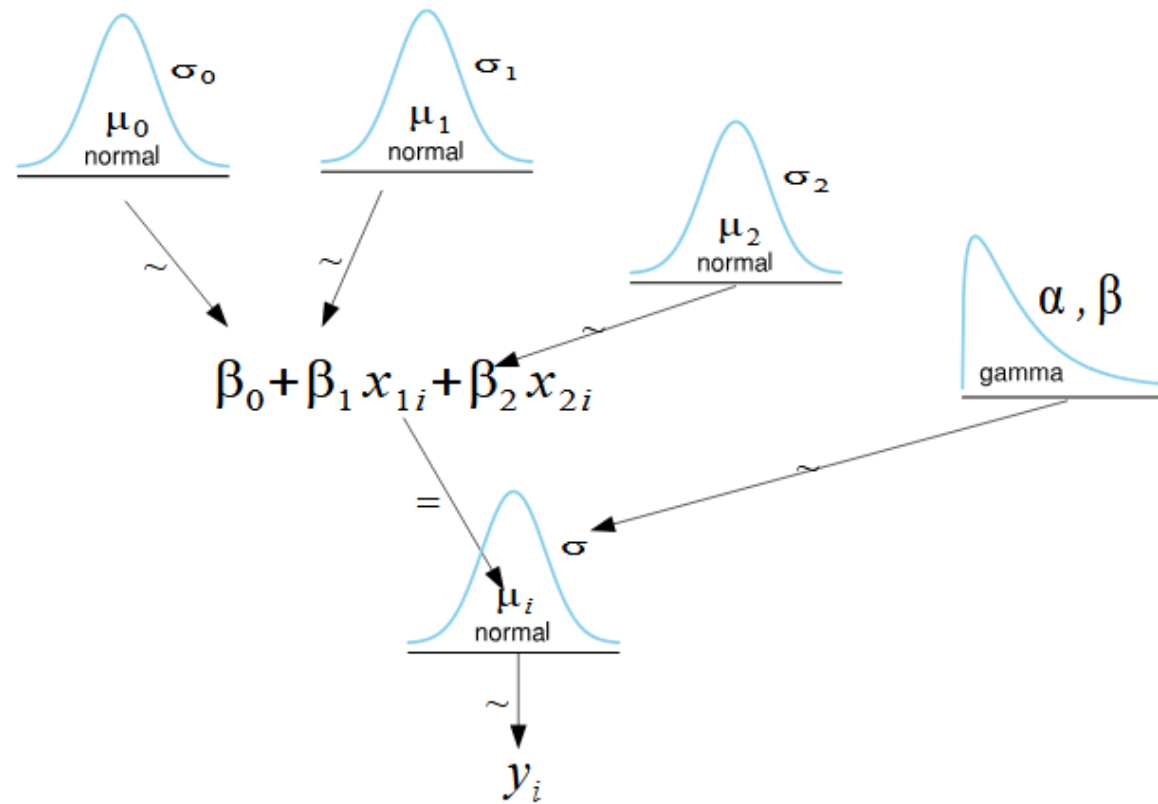


Adding in county (still need prior)



Multiple Regression

- Create model for the multiple regression.



Multiple Regression

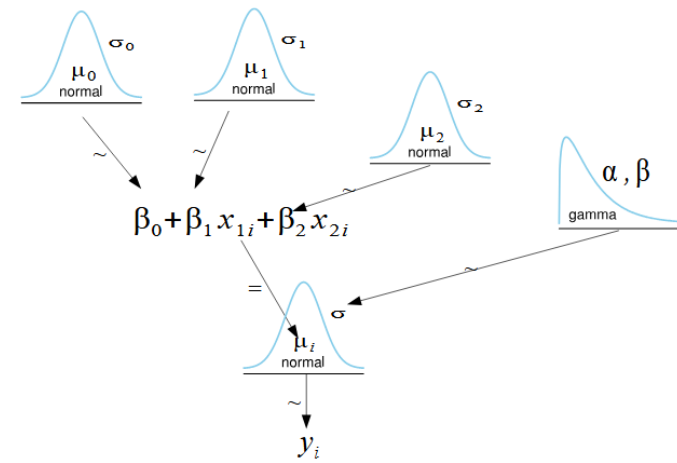
- And the code:

```
c_county = {"basement": b_codes, "county": c_codes, "data": np.arange(len(r))}

with pm.Model(coords=c_county) as m_county1:
    b_i = pm.Data('b_i', b_idx, dims="data")
    c_i = pm.Data('c_i', c_idx, dims="data")

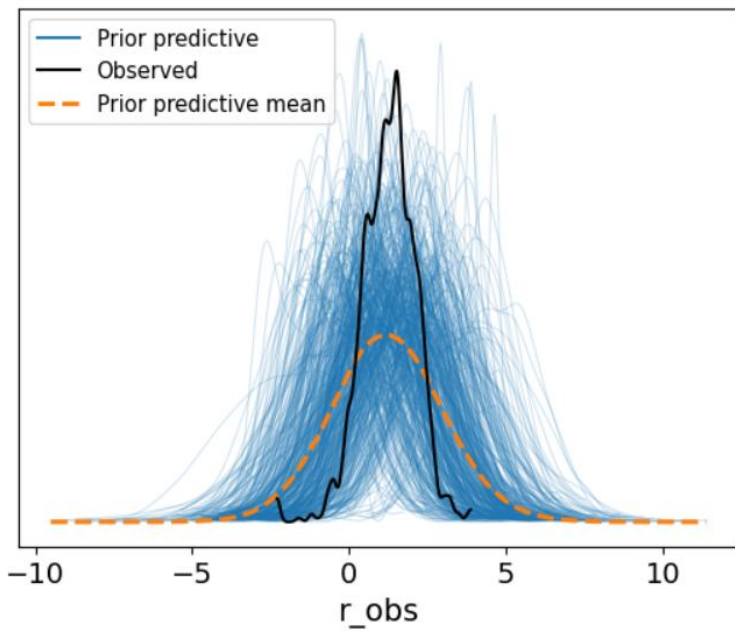
    beta_0 = pm.Normal('beta_0', mu=mu_0, sigma=sigma_0)
    beta_b = pm.Normal('beta_b', mu=0, sigma=sigma_b, dims="basement")
    beta_c = pm.Normal('beta_c', mu=0, sigma=sigma_c, dims="county")

    mu_r = pm.Deterministic('mu_r', beta_0 + beta_c[c_i] + beta_b[b_i], dims="data")
    sigma_r = pm.Gamma('sigma_r', mu=mu_sigma_r, sigma=sigma_sigma_r)
    r_obs = pm.Normal('r_obs', mu=mu_r, sigma=sigma_r, observed=r, dims="data")
```

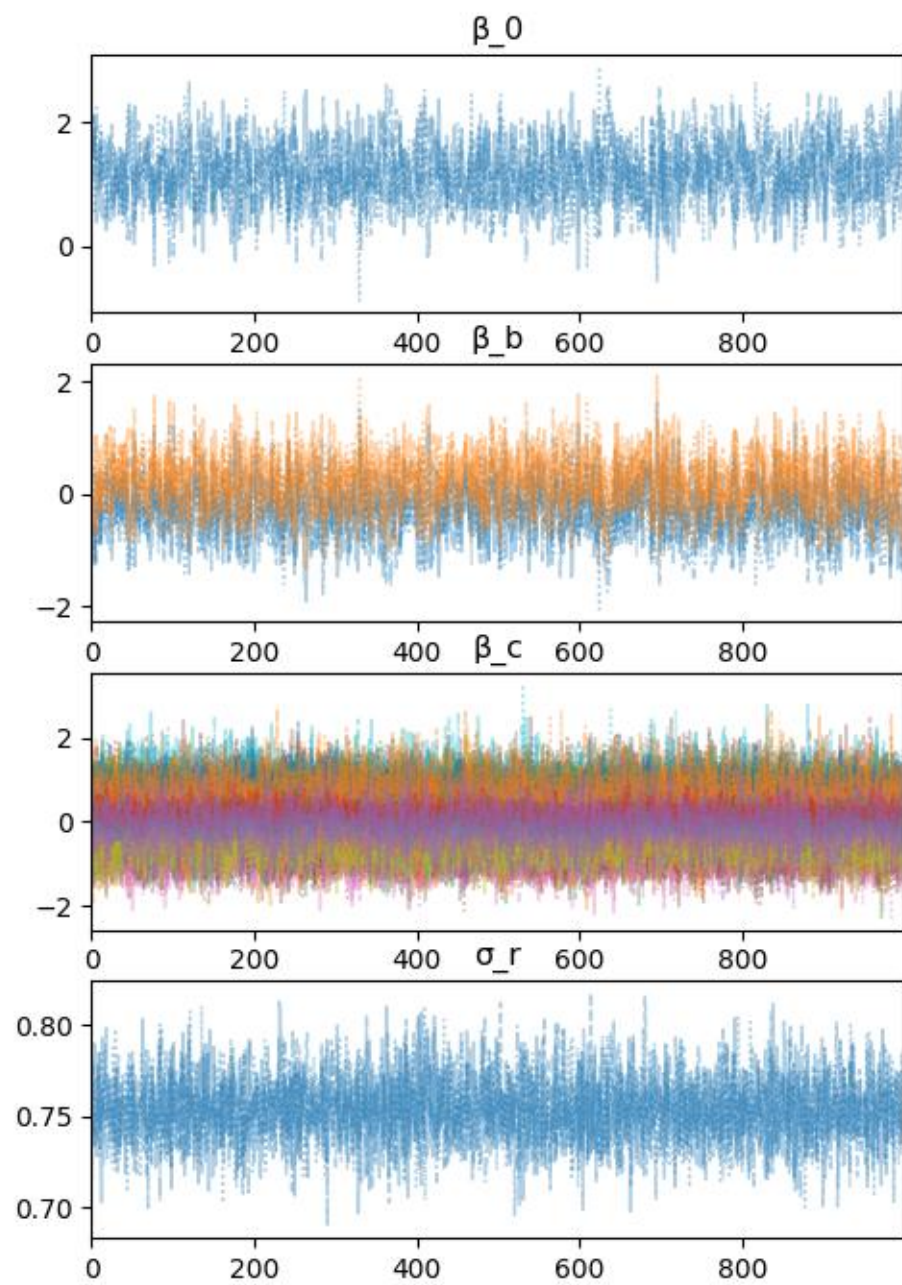
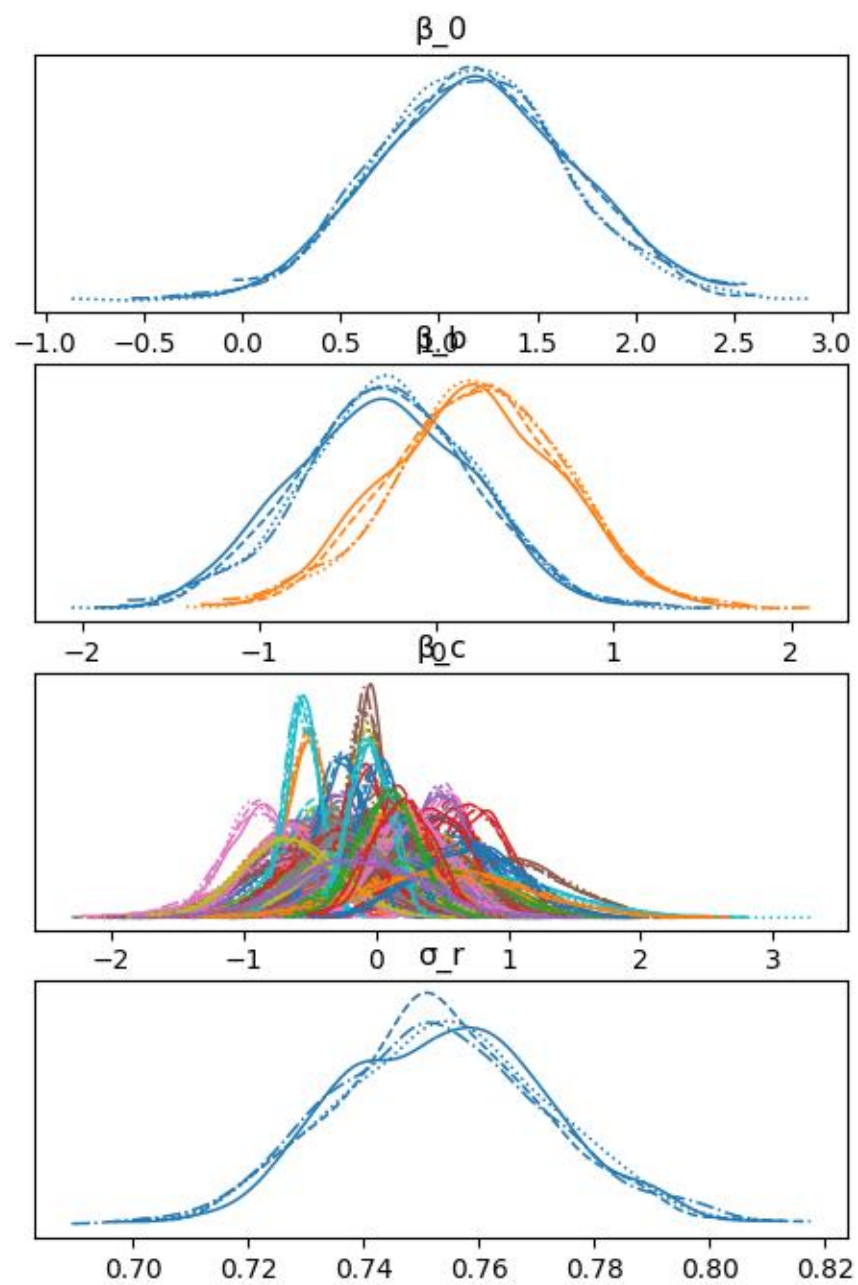


Multiple Regression

- Now we do:
 - Prior predictive checks
 - Sample
 - Diagnostics

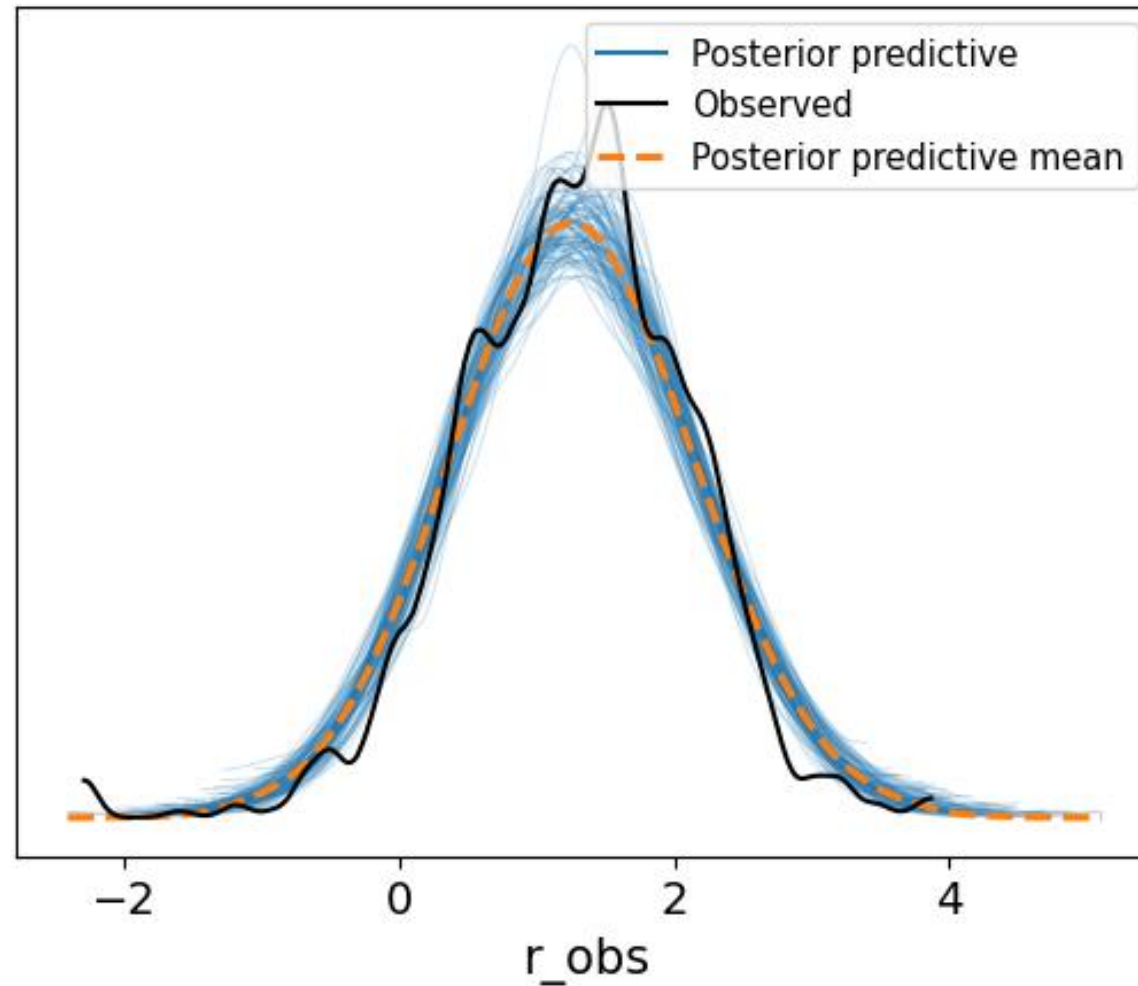


	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
β_0	0.015	0.008	1136.0	2080.0	1.00
$\beta_b[N]$	0.014	0.008	1260.0	2204.0	1.00
$\beta_b[Y]$	0.014	0.007	1208.0	2186.0	1.00
$\beta_c[AITKIN]$	0.007	0.006	2658.0	2461.0	1.00
$\beta_c[ANOKA]$	0.006	0.003	615.0	1326.0	1.01
...
$\beta_c[WILKIN]$	0.009	0.011	4971.0	2470.0	1.00
$\beta_c[WINONA]$	0.007	0.004	1201.0	2273.0	1.00
$\beta_c[WRIGHT]$	0.007	0.004	1184.0	2407.0	1.00
$\beta_c[YELLOW\ MEDICINE]$	0.007	0.008	4190.0	3117.0	1.00
σ_r	0.000	0.000	4784.0	3175.0	1.00



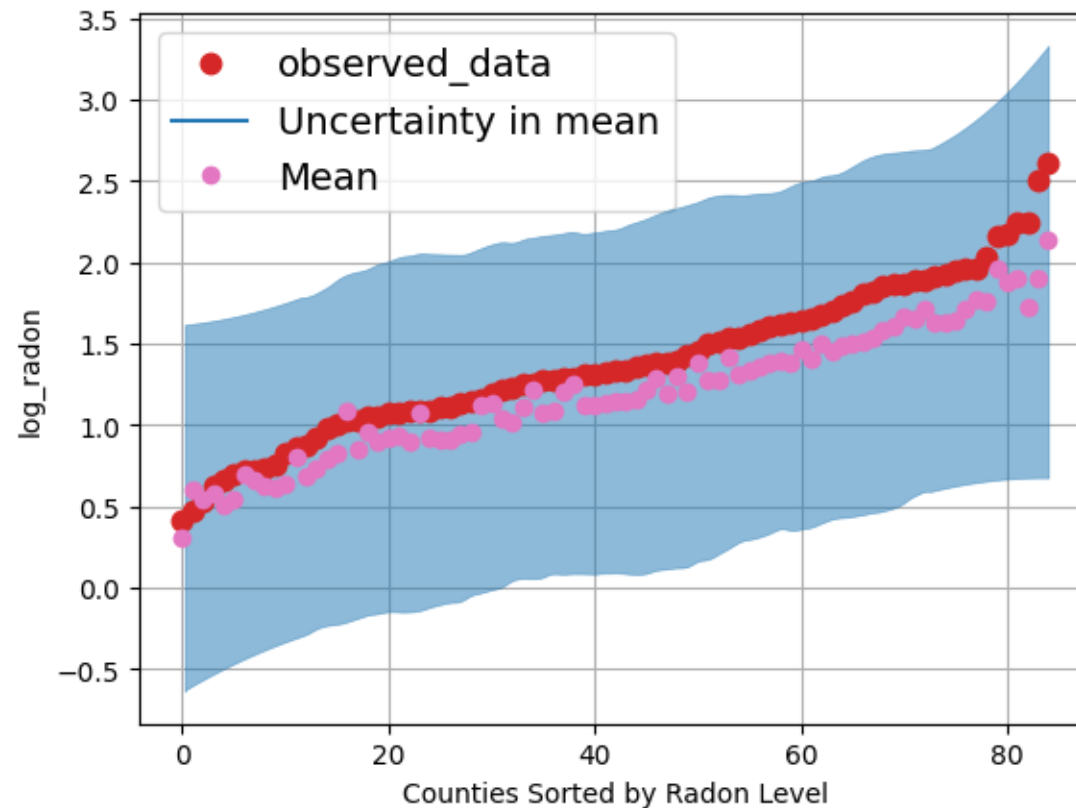
Multiple Regression

- Posterior Predictive Check



Multiple Regression

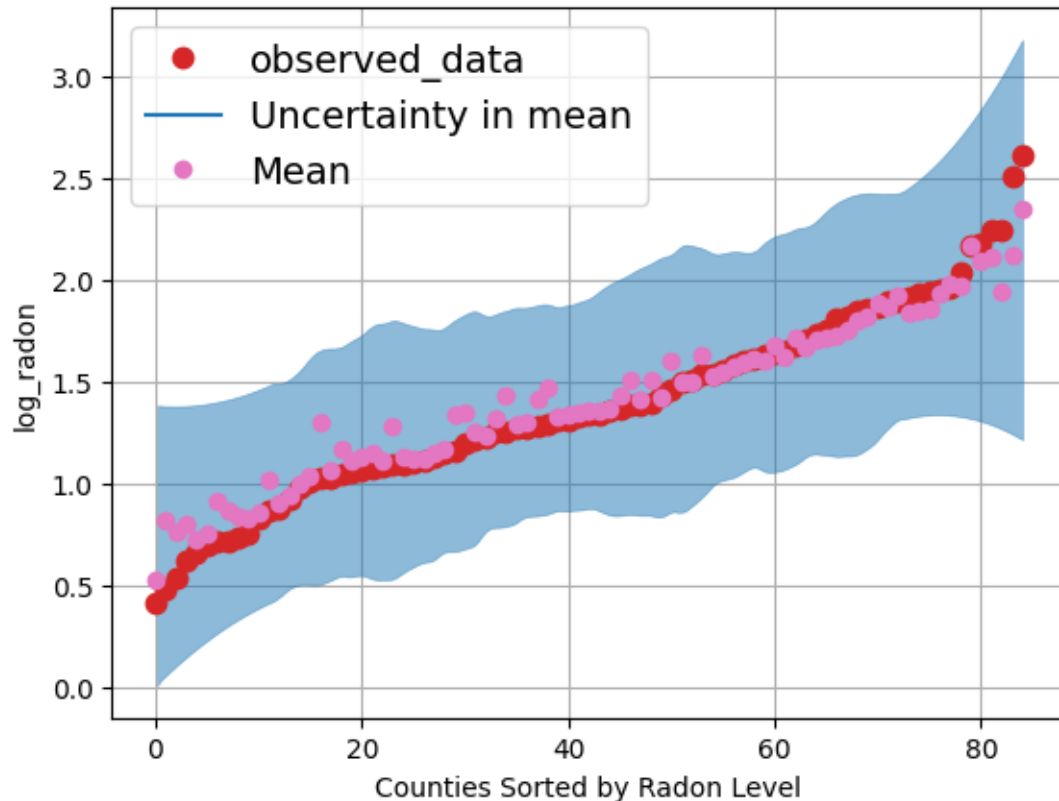
- Examine the average levels of radon for the different counties and those predicted by the model.



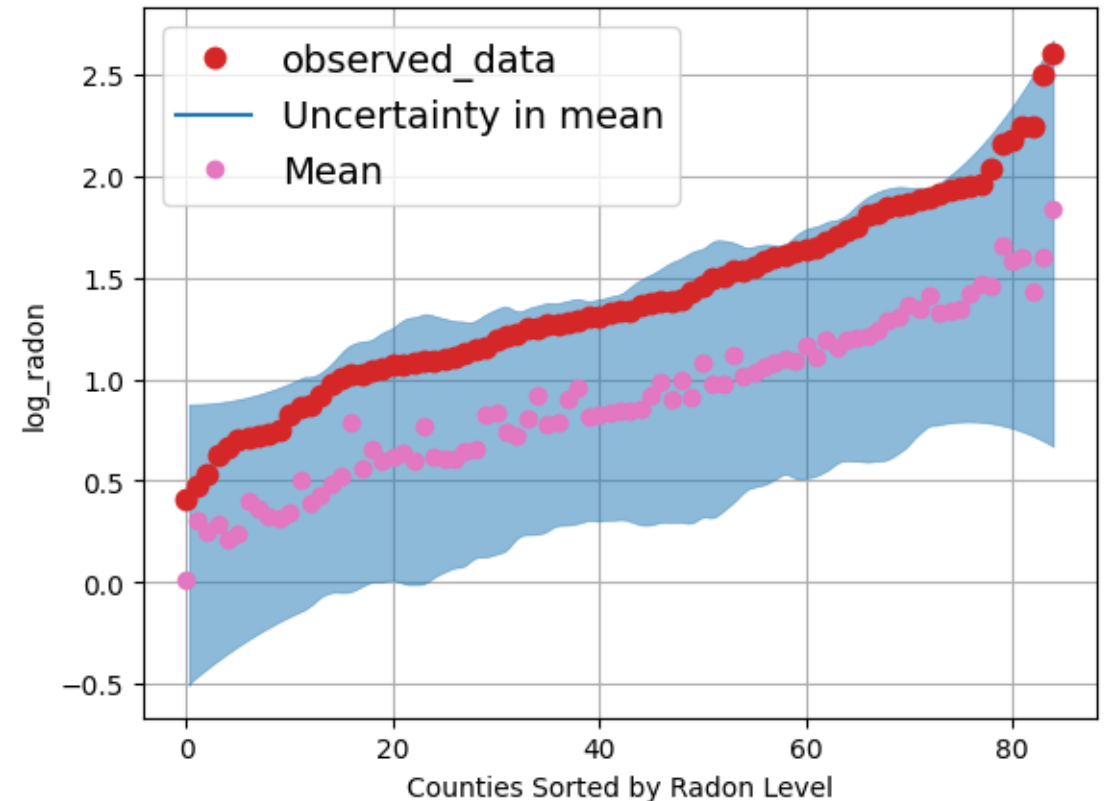
Multiple Regression

- We can also separate for the estimates for houses with and without basements separately:
- The mean of all the data is closer to that of only the basements (the majority of the data)

With Basement

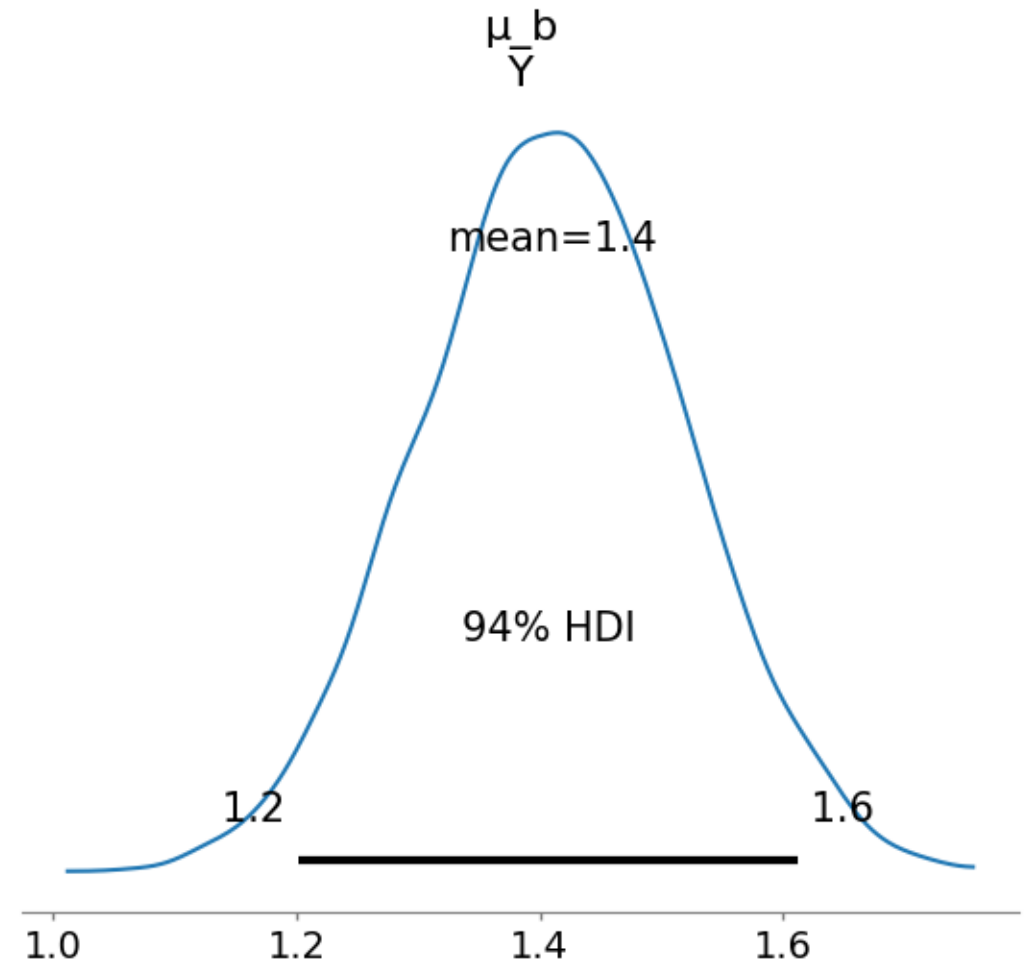
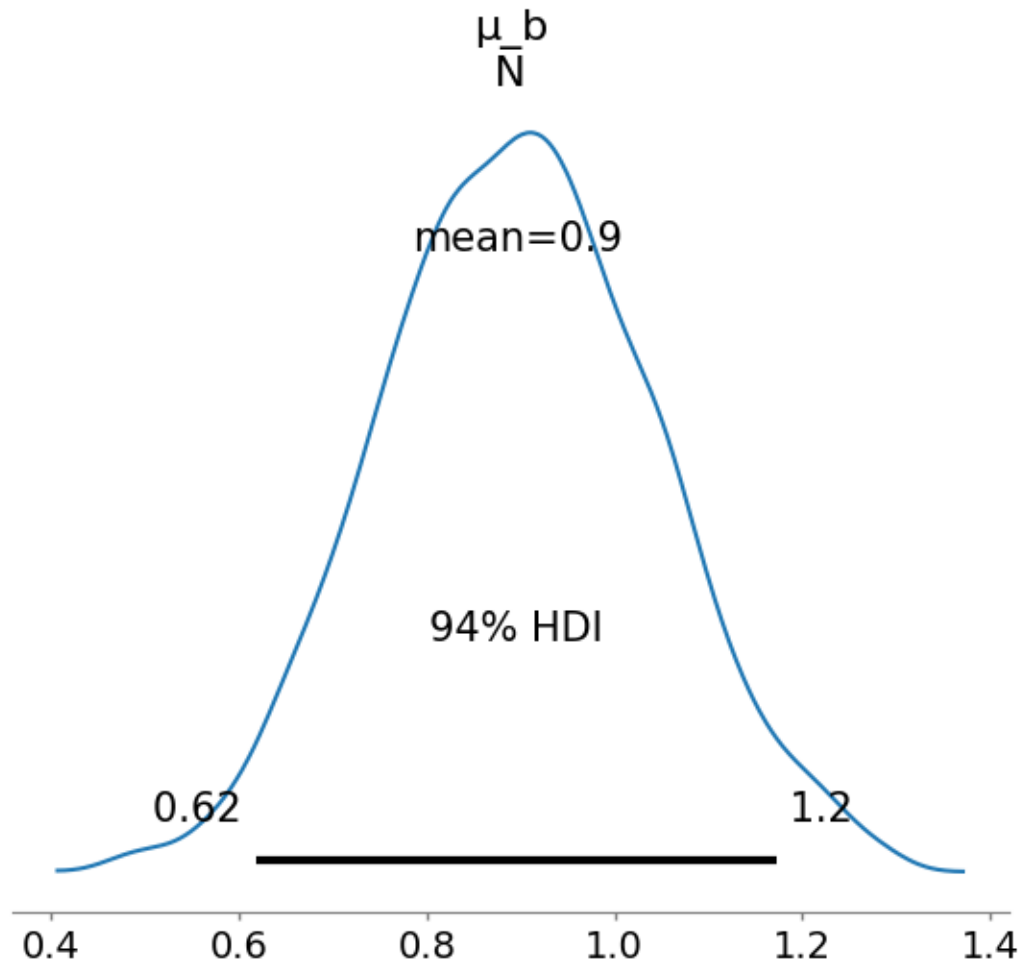


Without Basement



Multiple Regression

- Effect of basement (increases radon levels)

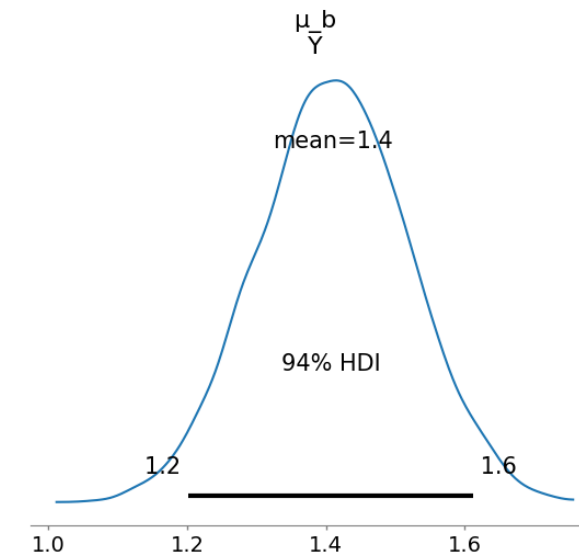
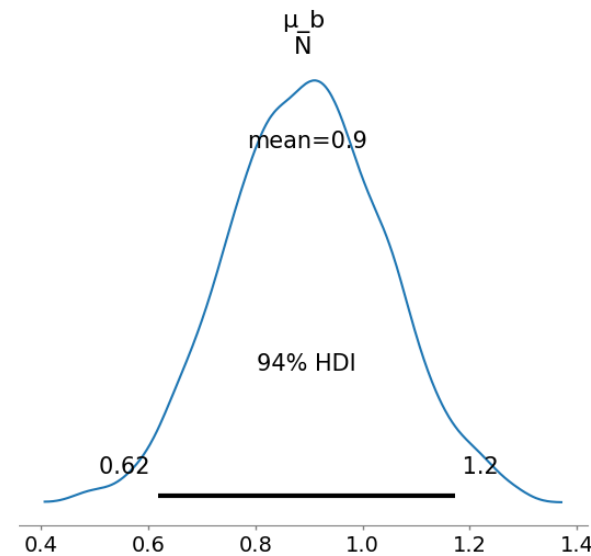
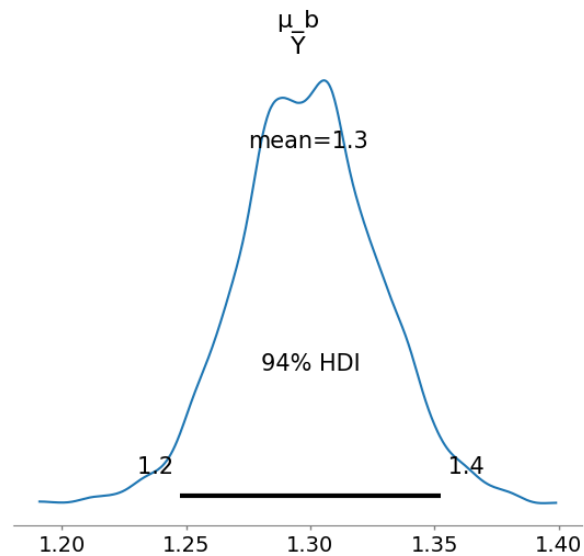
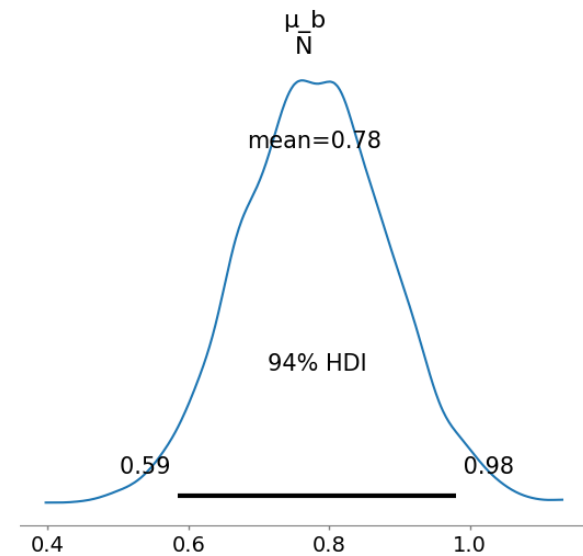


Multiple Regression

- We can compare between the results when including and not including the county data.

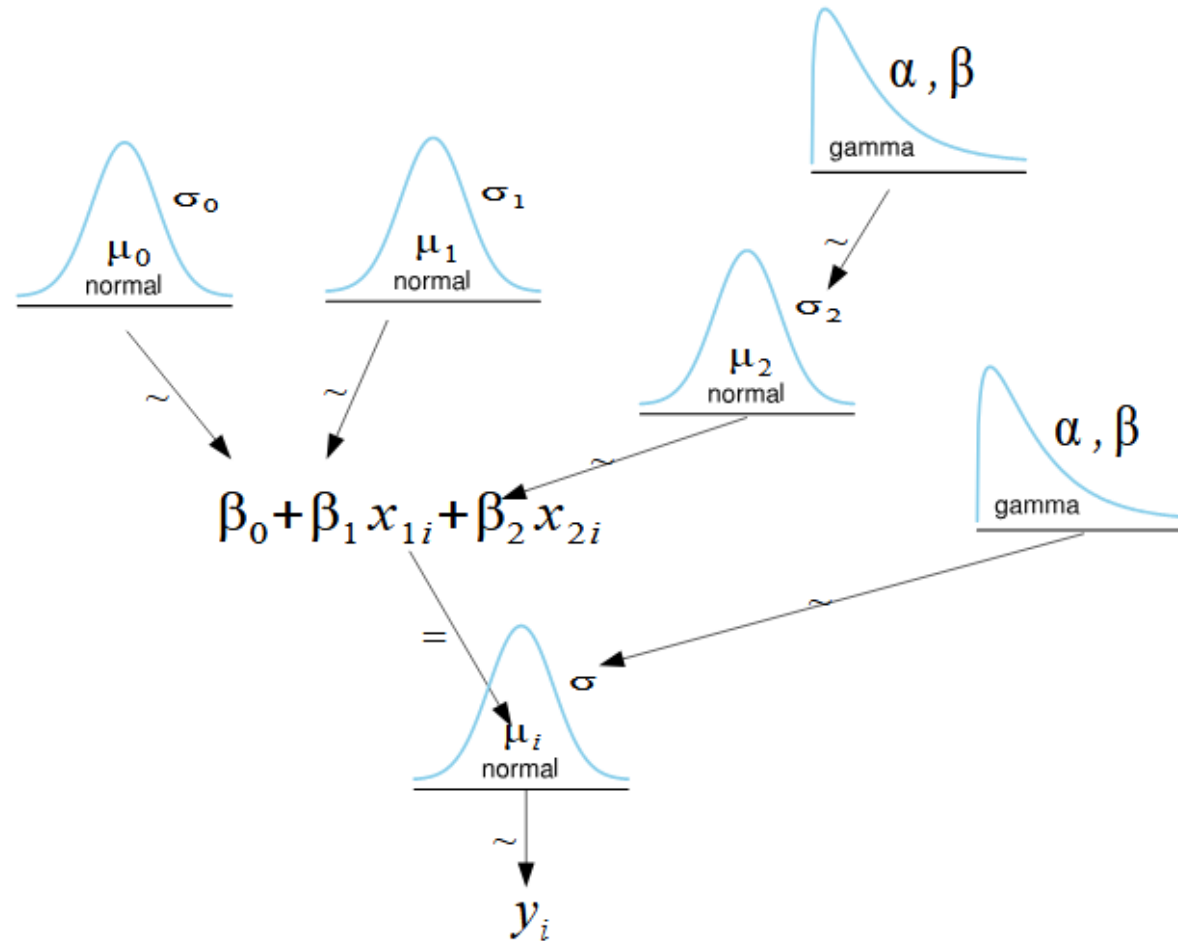
Not including

Including



Multiple Hierarchical Regression

- Create hierarchical model for the multiple regression.



Multiple Hierarchical Regression

- And the code:

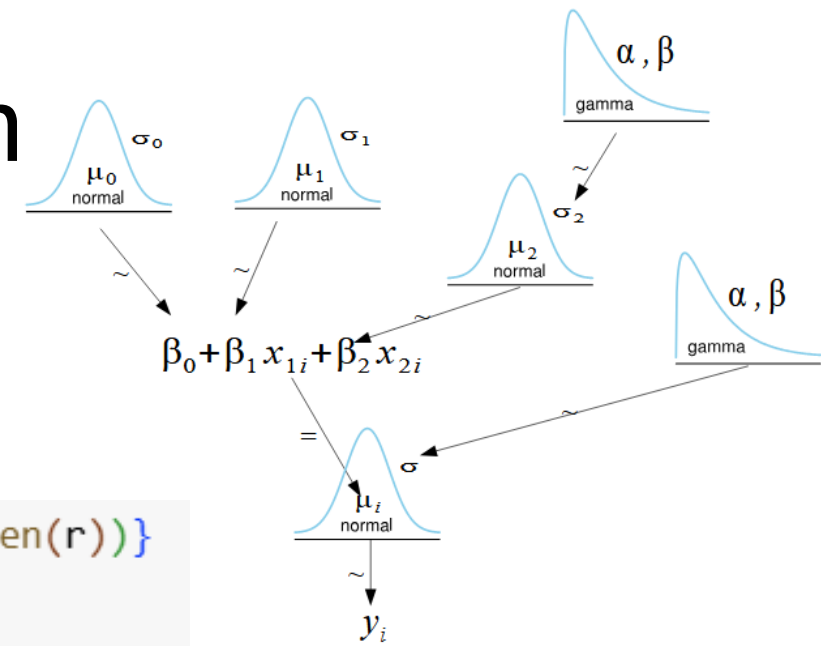
```
c_county = {"basement": b_codes, "county": c_codes, "data": np.arange(len(r))}

with pm.Model(coords=c_county) as m_county:
    b_i = pm.Data('b_i', b_idx, dims="data")
    c_i = pm.Data('c_i', c_idx, dims="data")

    sigma_c = pm.Gamma('sigma_c', mu=mu_sigma_c, sigma=sigma_sigma_c)

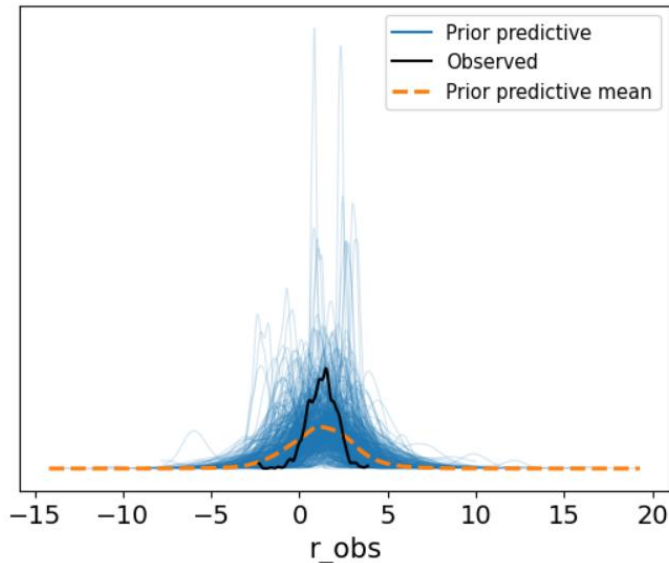
    beta_0 = pm.Normal('beta_0', mu=mu_0, sigma=sigma_0)
    beta_b = pm.Normal('beta_b', mu=0, sigma=sigma_b, dims="basement")
    beta_c = pm.Normal('beta_c', mu=0, sigma=sigma_c, dims="county")

    mu_r = pm.Deterministic('mu_r', beta_0 + beta_c[c_i] + beta_b[b_i], dims="data")
    sigma_r = pm.Gamma('sigma_r', mu=mu_sigma_r, sigma=sigma_sigma_r)
    r_obs = pm.Normal('r_obs', mu=mu_r, sigma=sigma_r, observed=r, dims="data")
```



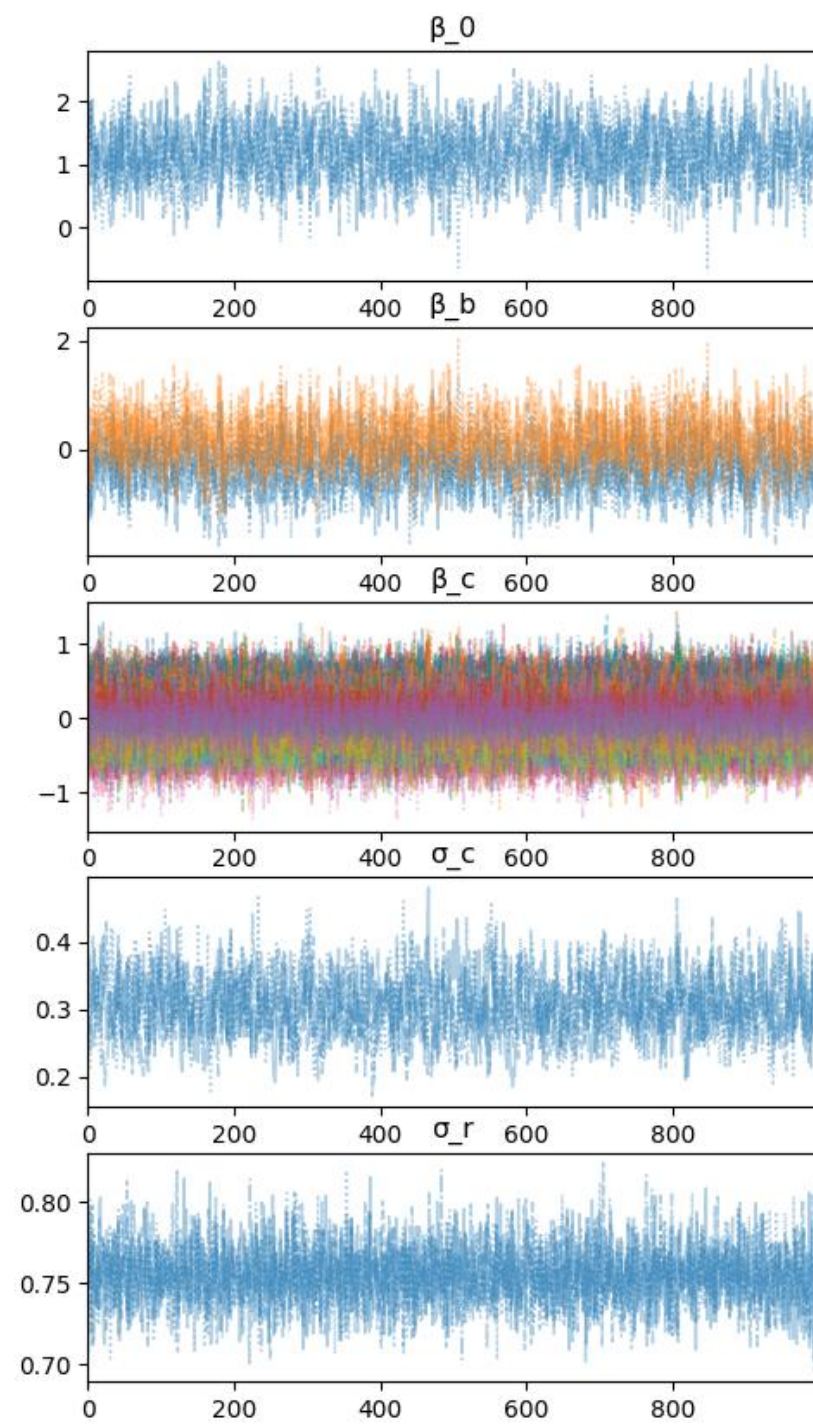
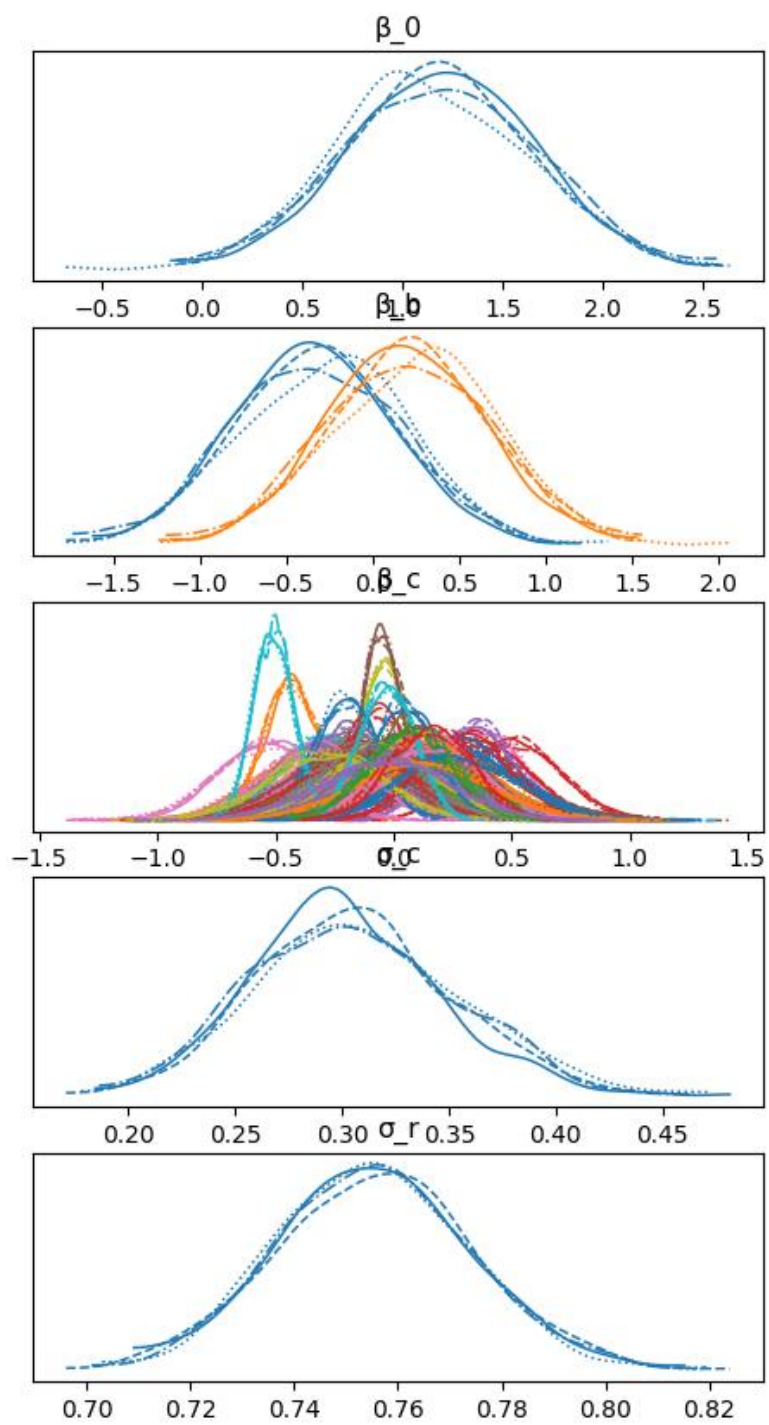
Multiple Hierarchical Regression

- Now we do:
 - Prior predictive checks
 - Sample
 - Diagnostics



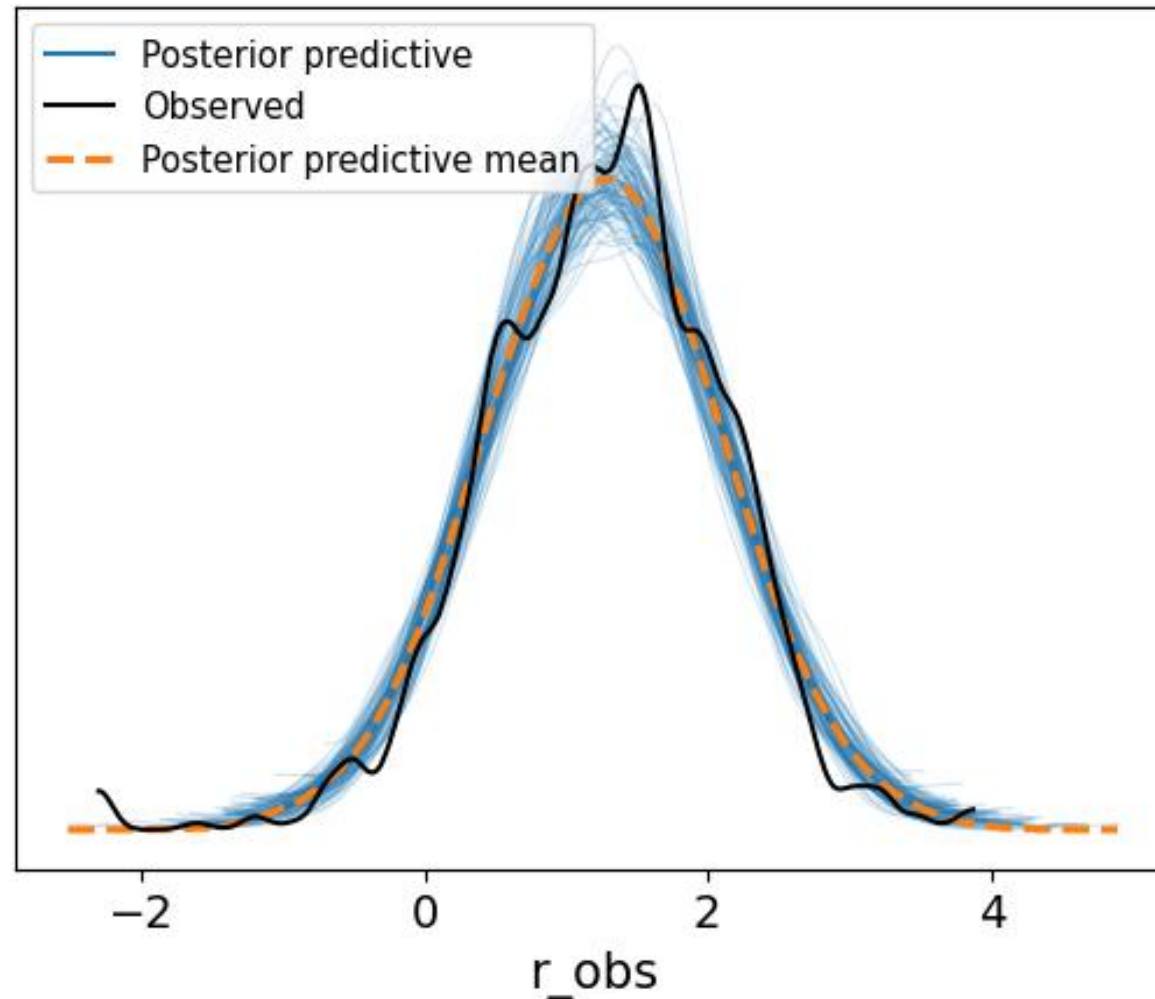
	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
β_0	0.013	0.007	1381.0	2478.0	1.0
$\beta_b[N]$	0.013	0.007	1430.0	2277.0	1.0
$\beta_b[Y]$	0.013	0.007	1406.0	2420.0	1.0
$\beta_c[AITKIN]$	0.004	0.004	4473.0	3031.0	1.0
$\beta_c[ANOKA]$	0.002	0.002	3329.0	2528.0	1.0
...
$\beta_c[WINONA]$	0.002	0.003	5350.0	2819.0	1.0
$\beta_c[WRIGHT]$	0.003	0.003	3667.0	3194.0	1.0
$\beta_c[YELLOW\ MEDICINE]$	0.004	0.004	4760.0	2851.0	1.0
σ_c	0.002	0.001	909.0	1402.0	1.0
σ_r	0.000	0.000	3939.0	3220.0	1.0

90 rows x 5 columns



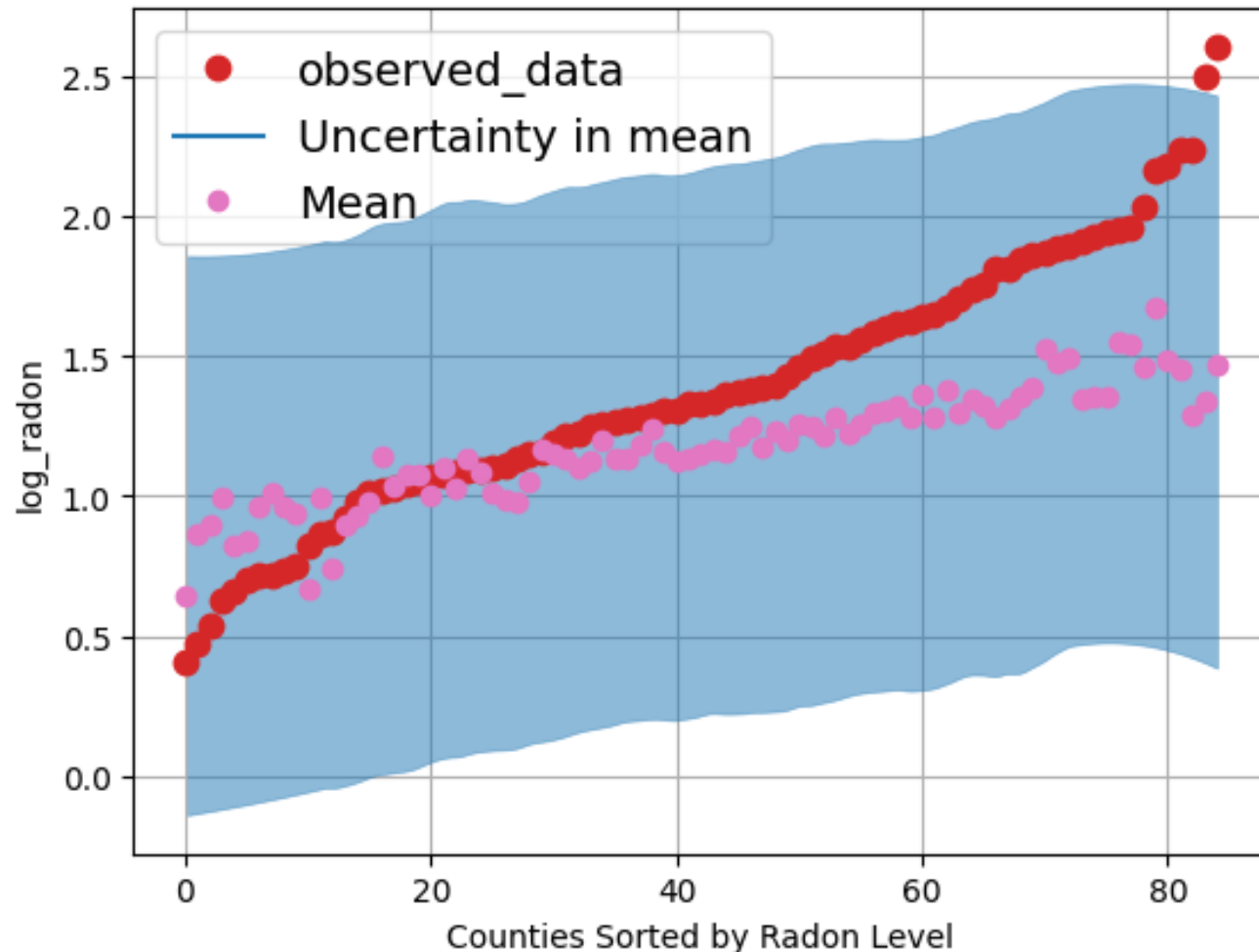
Multiple Hierarchical Regression

- Posterior Predictive Check



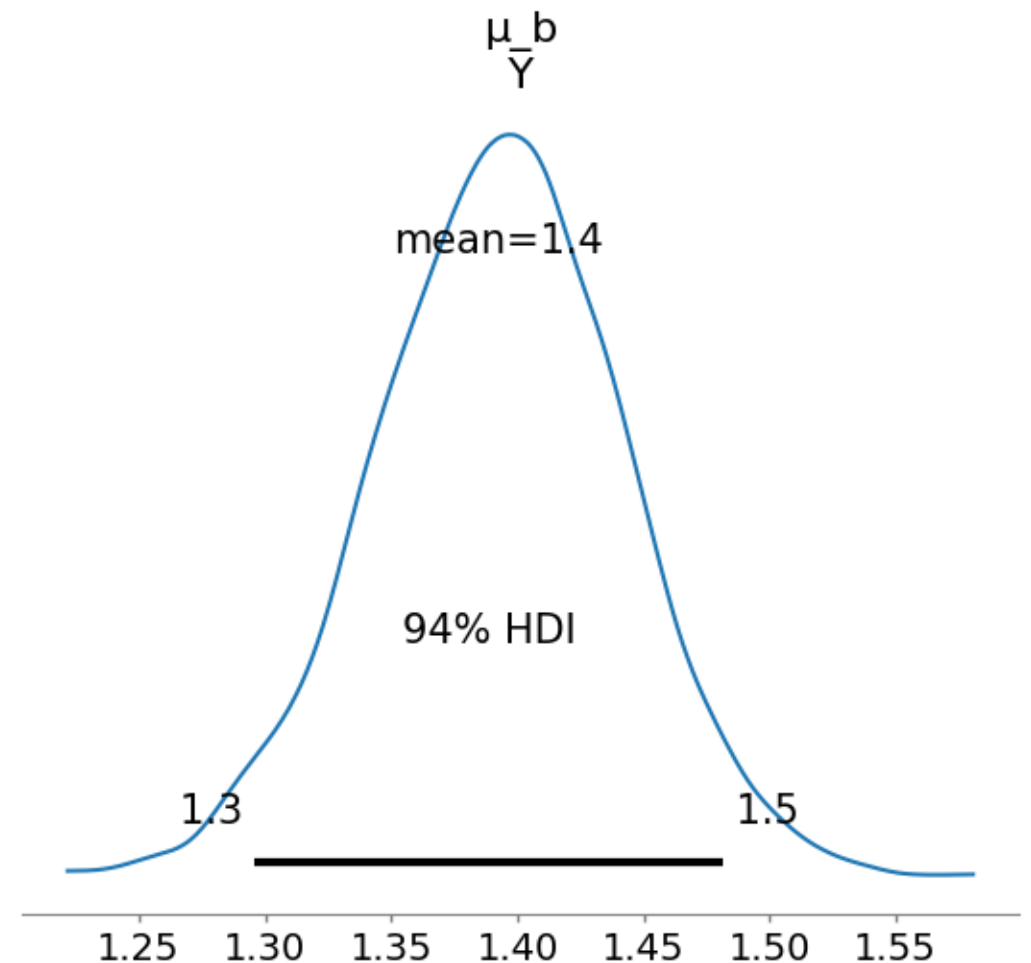
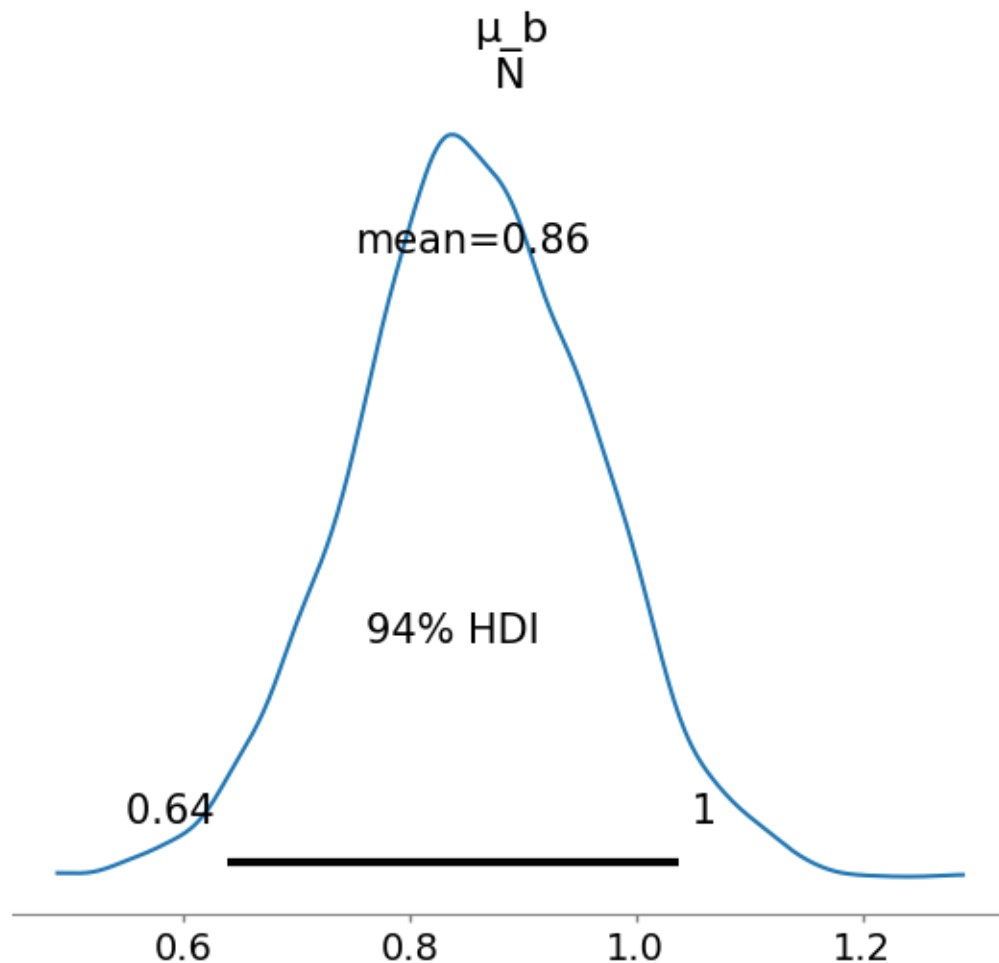
Multiple Hierarchical Regression

- Average levels of radon for the different counties and those predicted by the model.



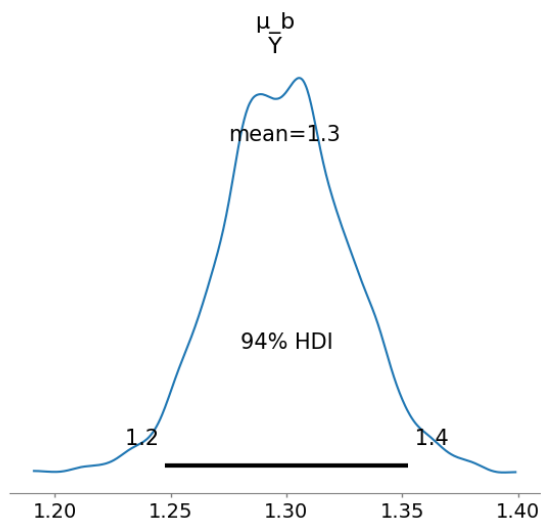
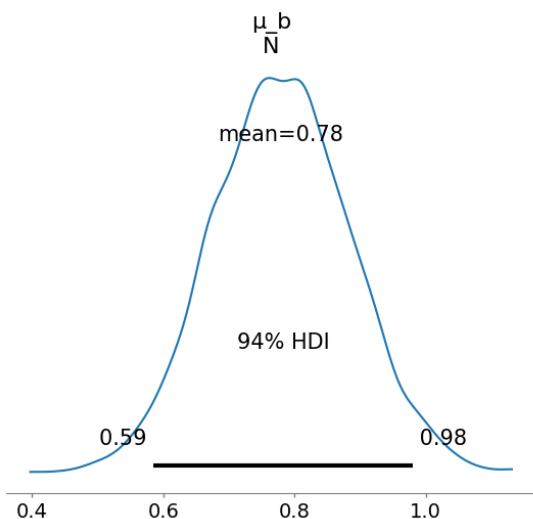
Multiple Hierarchical Regression

- Effect of basement for this model:

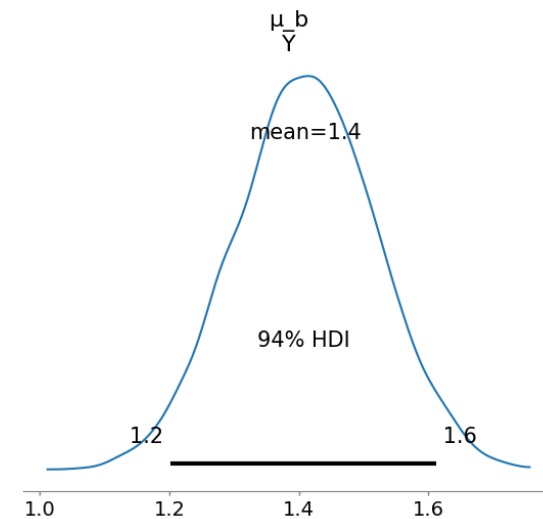
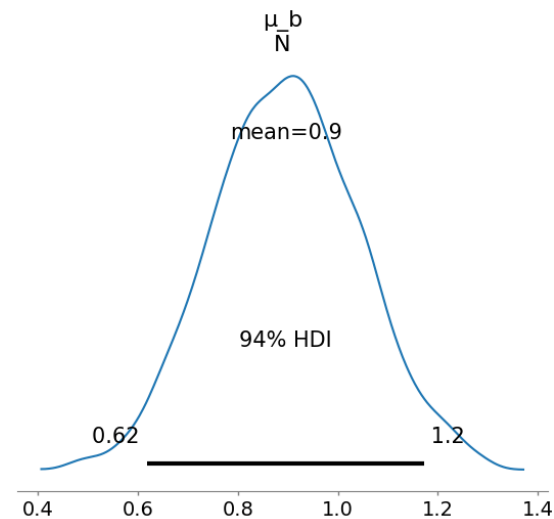


Compare the Three Models

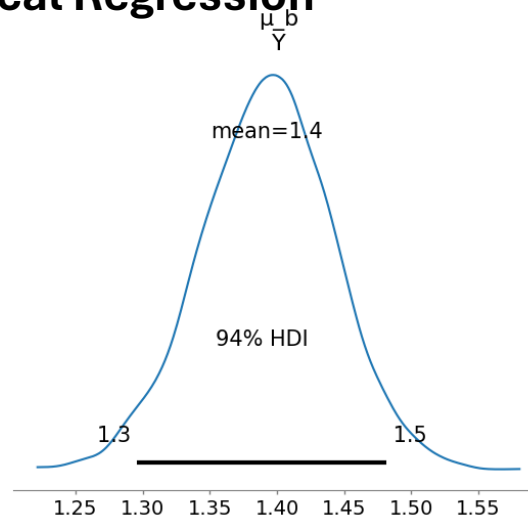
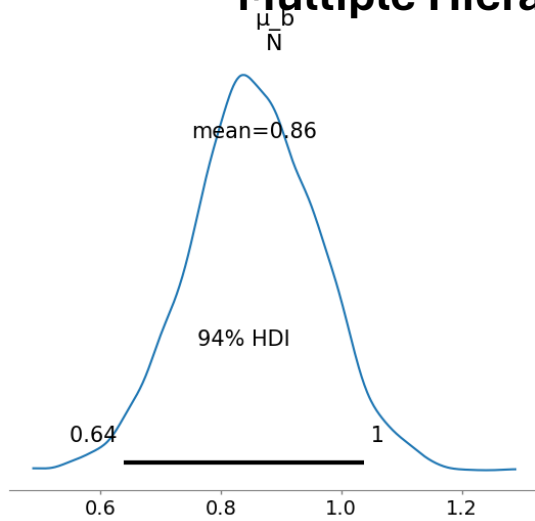
Only the basement



Multiple Regression



Multiple Hierarchical Regression



Compare the Models

- The hierarchical linear regression led to:
 - A larger difference between the estimate of radon levels in houses with and without basements.
 - More certainty in the estimates (narrower HDIs).
 - This is true for both the comparison with the multiple regression (not hierarchical) model and the simple model including only basement information.