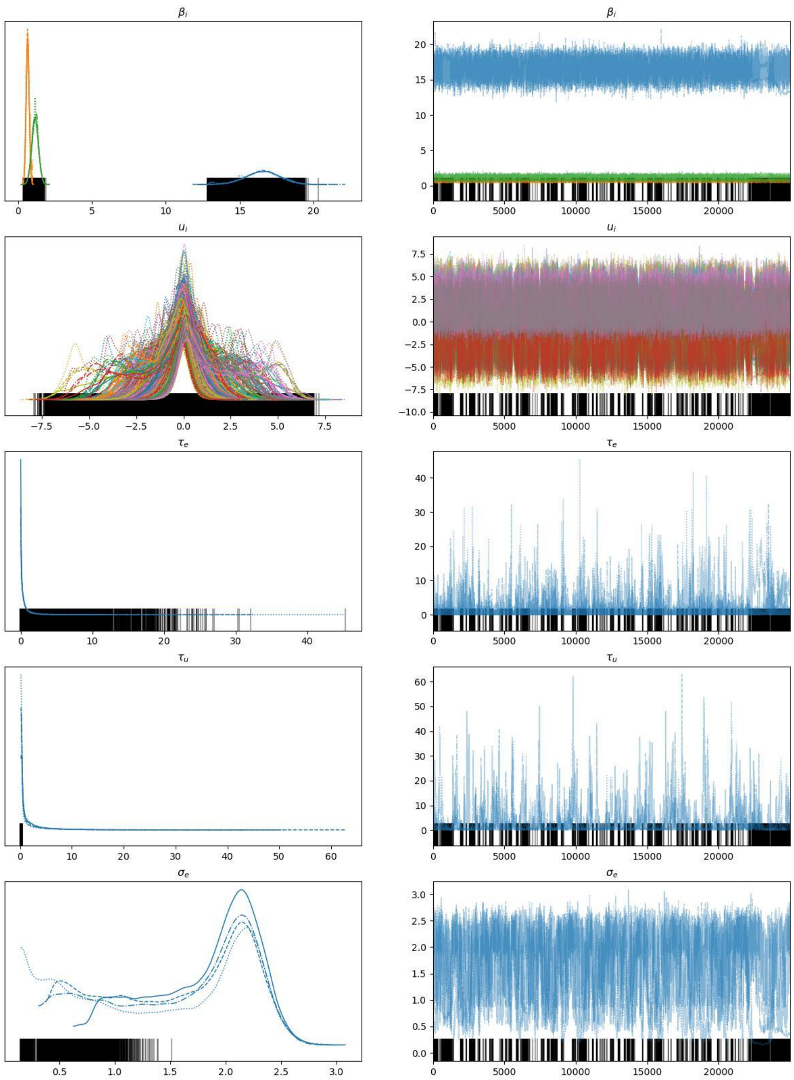
Questions

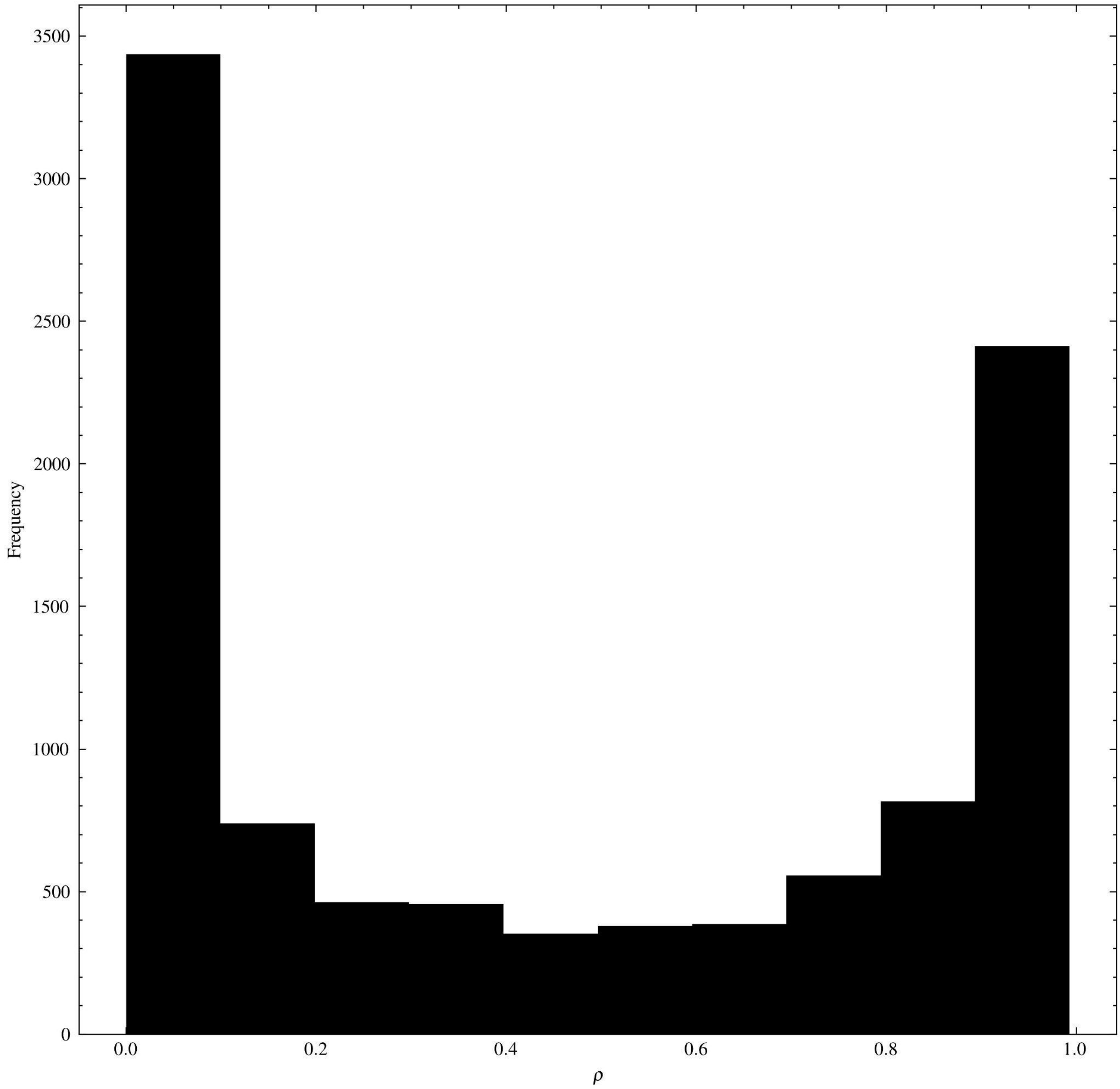
1. Orthodontic Distance. A longitudinal study was conducted to understand the effect of age and sex on the orthodontic distance (y). Measurements on 27 children are given in the file ortho.csv. There are a total of 16 boys and 11 girls, which are identified in the dataset using the column Subject. Consider the following random effects model:

$$y_{ij} \mid \beta_0, \beta_1, \beta_2, u_i, \sigma_e^2 \sim\sim^{ind.} \quad N\big(\beta_0 + \beta_1 age_{ij} + \beta_2 \, \mathrm{sex} \, x_i + u_i, \sigma_e^2\big),$$
$$u_i \mid \sigma_u^2 \sim \text{ iid } N\big(0, \sigma_u^2\big),$$

a) Posterior distributions for the requested parameters were calculated using PyMC3. Results are below:



b) Intra-class correlation coefficient histogram:



It appears to be significantly different from 0

c) The previous model was fitted without the random effects and tables presenting the results are presented below.
With Random Effects:

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
$\beta_i[0]$	16.521	1.161	14.399	18.739	0.043	0.03	822.0	346.0	1.01
$\beta_i[1]$	0.663	0.104	0.472	0.859	0.004	0.003	754.0	308.0	1.01
$\beta_i[2]$	1.163	0.222	0.738	1.577	0.002	0.001	11661.0	22719.0	1.0
τ_e	0.931	2.194	0.0	3.67	0.166	0.117	462.0	202.0	1.01
τ_u	2.263	4.316	0.119	9.326	0.181	0.128	298.0	426.0	1.02
σ_e	1.594	0.658	0.376	2.48	0.042	0.03	242.0	66.0	1.04

Without Random effects:

	mean	sd	hdi_3%	hdi_97%	mcse_mean	mcse_sd	ess_bulk	ess_tail	r_hat
$\beta_i[0]$	16.543	1.112	14.455	18.658	0.006	0.004	40327.0	49289.0	1.0
$\beta_i[1]$	0.66	0.099	0.474	0.848	0.0	0.0	40280.0	48362.0	1.0
$\beta_i[2]$	1.161	0.224	0.743	1.583	0.001	0.001	76292.0	61983.0	1.0
τ_e	0.225	0.293	0.0	0.756	0.001	0.001	40744.0	27520.0	1.0
σ_e	2.286	0.16	1.992	2.592	0.001	0.0	66606.0	62648.0	1.0

Confidence intervals for the regression parameters haven't been affected by the presence of random effects.

2. **Nanowire density**. Consider the problem of predicting the density of nanowires y with respect to the thickness of polymer films x in a solution-based growth process (see Figure 1). Eight experiments were conducted with two replicates (except for one run). The data are in the file nanowire.csv. The density of nanowires is assumed to follow a Poisson distribution with mean:

$$\mu(x) = \theta_1 \exp\left(-\theta_2 x^2\right) + \theta_3 \left\{1 - \exp\left(-\theta_2 x^2\right)\right\} \Phi(-x/\theta_4)$$

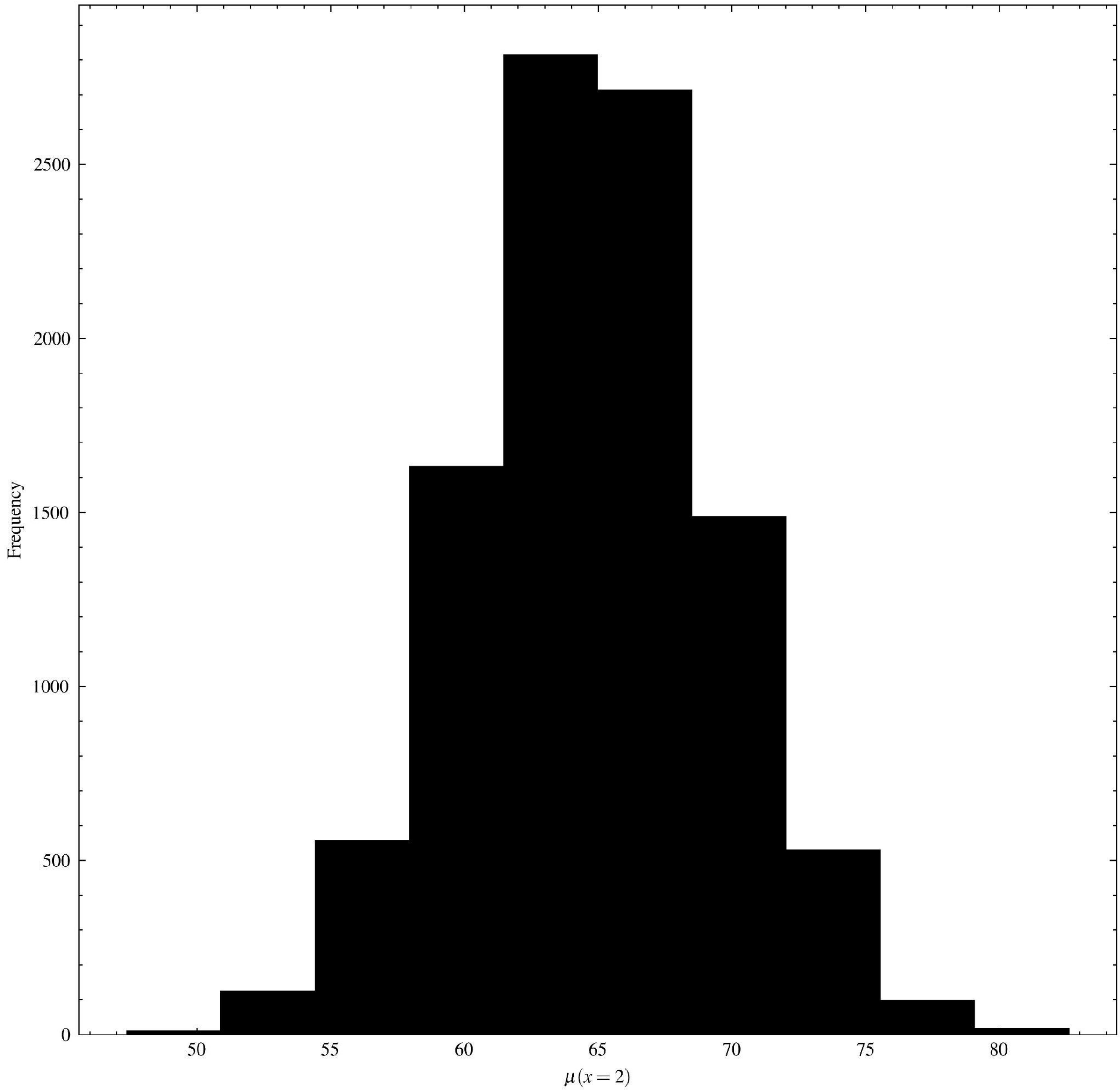
where $\Phi(\cdot)$ is the standard normal CDF - note that there is a phi () function in BUGS for this, and in pyme you may use the invprobit() function. Assume the following prior distribution for the parameters:

$$\begin{aligned} \log\theta_1, \log\theta_3, \log\theta_4 \quad &\sim^{iid} \quad N(0, \sigma^2 = 10) \\ \theta_2 &\sim U(0,1). \end{aligned}$$

a) The model was implemented in PyMC3 and the resulting summaries with the according intervals are shown below:

	mean	sd	hdi 3%	hdi 97%	mcse mean	mcse sd	ess bulk	ess tail	r hat
θ_1	123.019	7.853	108.59	138.029	0.084	0.059	8702.0	8259.0	1.0
θ_3	27.608	7.848	14.02	42.228	0.107	0.076	5476.0	5767.0	1.0
θ_4	11.473	2.881	7.318	15.875	0.038	0.027	5774.0	6080.0	1.0
θ_2	0.186	0.027	0.135	0.235	0.0	0.0	7396.0	7356.0	1.0
$\mu[0]$	123.019	7.853	108.59	138.029	0.084	0.059	8702.0	8259.0	1.0
$\mu[1]$	123.019	7.853	108.59	138.029	0.084	0.059	8702.0	8259.0	1.0
$\mu[2]$	46.518	4.615	37.614	54.968	0.043	0.031	11262.0	10155.0	1.0
$\mu[3]$	46.518	4.615	37.614	54.968	0.043	0.031	11262.0	10155.0	1.0
$\mu[4]$	10.091	1.599	7.141	13.083	0.019	0.013	7437.0	7660.0	1.0
$\mu[5]$	10.091	1.599	7.141	13.083	0.019	0.013	7437.0	7660.0	1.0
$\mu[6]$	6.619	1.099	4.621	8.719	0.012	0.009	7881.0	8930.0	1.0
$\mu[7]$	6.619	1.099	4.621	8.719	0.012	0.009	7881.0	8930.0	1.0
$\mu[8]$	4.808	0.753	3.458	6.261	0.007	0.005	13260.0	11340.0	1.0
$\mu[9]$	4.808	0.753	3.458	6.261	0.007	0.005	13260.0	11340.0	1.0
$\mu[10]$	3.382	0.67	2.089	4.584	0.006	0.004	12352.0	10053.0	1.0
$\mu[11]$	3.382	0.67	2.089	4.584	0.006	0.004	12352.0	10053.0	1.0
$\mu[12]$	1.001	0.535	0.151	1.972	0.006	0.005	6903.0	7419.0	1.0
$\mu[13]$	0.406	0.36	0.005	1.055	0.005	0.003	6355.0	7024.0	1.0
$\mu[14]$	0.406	0.36	0.005	1.055	0.005	0.003	6355.0	7024.0	1.0

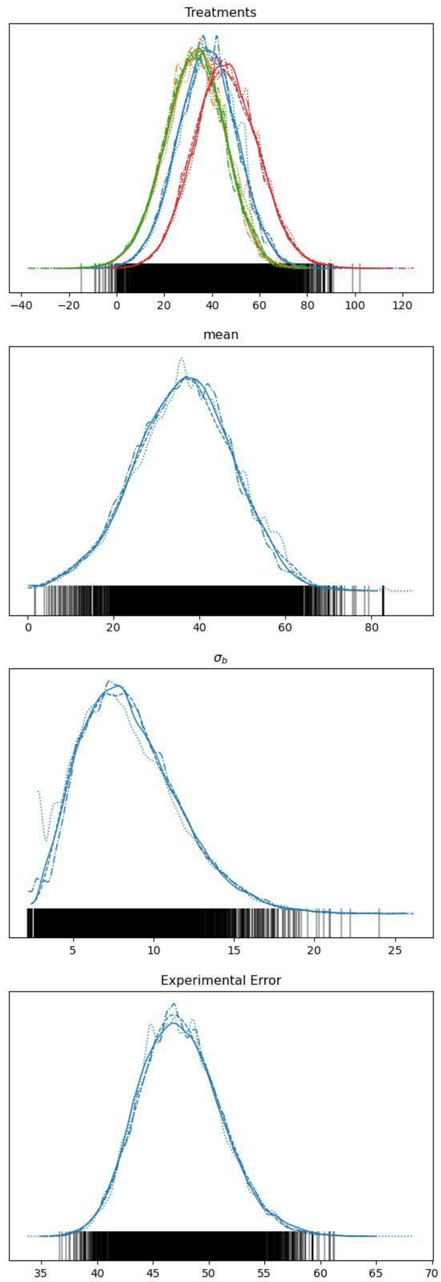
b) The posterior distribution when x=2 was obtained using the Deterministic wrapper:



3) **Color Attraction for Oulema melanopus.** Some colors are more attractive to insects than others. Wilson and Shade $(1967)^{1}$ conducted an experiment aimed at determining the best color for attracting cereal leaf beetles (Oulema melanopus). Six boards in each of four selected colors (lemon yellow, white, green, and blue) were placed in a field of oats during summer time. The following table (modified from Wilson and Shade, 1967) gives data on the number of cereal leaf beetles trapped:

Board color	Insects trapped					
Lemon yellow	45	59	48	46	38	47
White	21	12	14	17	13	17
Green	16	11	20	21	14	7
Blue	37	32	15	25	39	41

a) Bayesian ANOVA analysis was done using PyMC3. The STZ constraint was done by imposing a zero mean to the treatments as advised by Krutchske. Results are presented below:



b) From this Anova analysis we can conclude there is no clear evidence of statistical significance for the bugs actually preferring a color. This is due mainly to the fact that the treatments have meaningful overlap and the experimental error is on the order of the measured values.