

1 Potato Leafhopper (10 points)

By running the OpenBUGS code, we obtain the following results:

	mean	sd	MC_error	val5.0pc	val95.0pc	start	sample
beta0	116.6	10.86	0.3888	98.19	134.2	1001	100000
beta1	-1.282	0.1513	0.005419	-1.527	-1.026	1001	100000
develop[3]	29.92	5.985	0.02961	20.14	39.69	1001	100000
develop[17]	19.16	5.982	0.03058	9.413	29.01	1001	100000
develop[24]	1.578	6.554	0.09873	-8.986	12.39	1001	100000
develop[29]	20.52	5.979	0.02474	10.81	30.33	1001	100000
temp[2]	71.48	4.299	0.0185	64.57	78.63	1001	100000
temp[30]	70.98	4.285	0.01687	64.13	78.16	1001	100000

Equivalent Python solution:

	mean	sd	hdi_5%	...	ess_bulk	ess_tail	r_hat
beta0	116.757	11.020	98.510	...	6173.0	6768.0	1.0
beta1	-1.285	0.153	-1.552	...	6086.0	6653.0	1.0
x_imputed_missing[0]	71.511	4.346	64.554	...	10012.0	6785.0	1.0
x_imputed_missing[1]	71.062	4.390	63.782	...	10262.0	8119.0	1.0
lik_missing[0]	29.889	6.063	19.888	...	10934.0	7725.0	1.0
lik_missing[1]	19.222	6.131	9.228	...	11280.0	7587.0	1.0
lik_missing[2]	1.487	6.774	-9.582	...	8656.0	8455.0	1.0
lik_missing[3]	20.508	6.144	10.462	...	12310.0	7517.0	1.0
tau	5.854	0.946	4.340	...	6204.0	7205.0	1.0

- (a) (**1 point**) The estimated parameter β_0 and β_1 are 116.6 and -1.282 , respectively. The fitted linear regression model is

$$Develop = 116.6 - 1.282 \cdot Temp.$$

- (b) (**6 points**, 0.5 for each mean and credible set of the missing values) The results for the imputed missing values are as follows:

develop[3] estimate is 29.92, and the 90% credible set is [20.14, 39.69]

develop[17] estimate is 19.16, and the 90% credible set is [9.413, 29.01]

develop[24] estimate is 1.578, and the 90% credible set is $[-8.986, 12.39]$

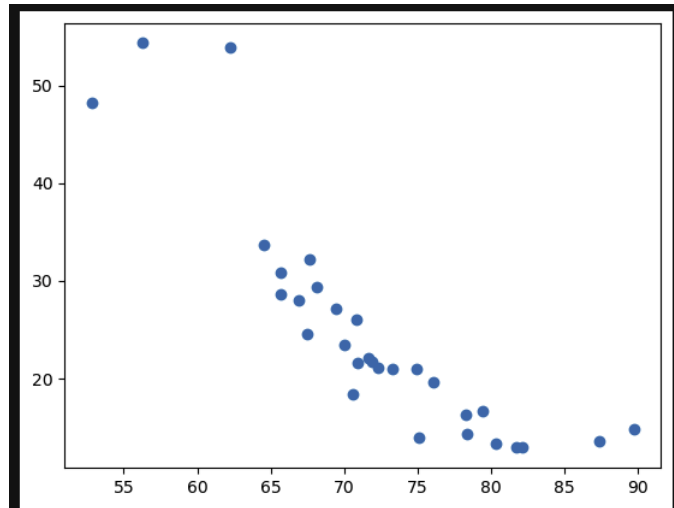
develop[29] estimate is 20.52, and the 90% credible set is $[-8.986, 12.39]$

temp[2] estimate is 71.48, and the 90% credible set is [64.57, 78.63]

temp[30] estimate is 70.98, and the 90% credible set is [64.13, 78.16]

- (c) (**1 point** for identifying what was problematic and **2 points** for showing how a different model remedies this)

We plot temp vs develop to get a sense of the data:



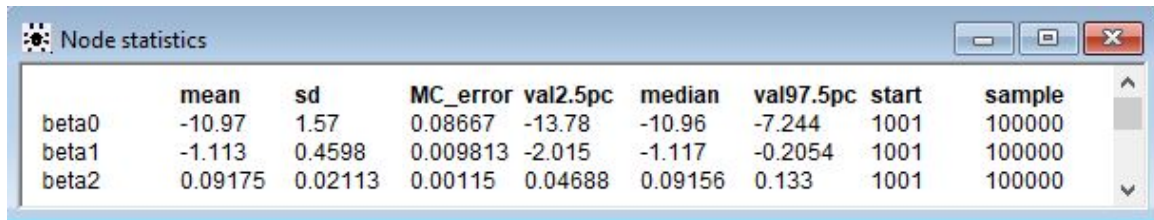
A linear model seems appropriate for most of this domain, but large values of `temp` seem problematic as there cannot be a negative development time. For example, the imputed Develop time at $Temp = 89.7$ is only 1.578, and for higher temperatures this value could be negative. As such, a truncated likelihood or nonlinear models may allow for better predictions when the value of `temp` is large. All such discussions and results should receive credit. One quick formal solution is to use a lognormal likelihood instead of a normal likelihood. The updated predictions are more reasonable, especially `develop[24]` which corresponds to $Temp = 89.7$:

Node statistics								
	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta0	6.643	0.3002	0.01069	6.031	6.645	7.224	1001	100000
beta1	-0.04934	0.004182	1.489E-4	-0.05739	-0.04937	-0.0408	1001	100000
develop[3]	27.72	4.671	0.02282	19.69	27.32	38.03	1001	100000
develop[17]	18.32	3.086	0.01566	13.0	18.05	25.17	1001	100000
develop[24]	9.339	1.728	0.02557	6.42	9.176	13.19	1001	100000
develop[29]	19.31	3.255	0.01347	13.72	19.04	26.53	1001	100000
temp[2]	68.88	3.234	0.01162	62.61	68.86	75.41	1001	100000
temp[30]	68.36	3.227	0.01119	62.11	68.31	74.91	1001	100000

	mean	sd	hdi_5%	...	ess_bulk	ess_tail	r_hat
beta0	6.661	0.310	6.172	...	6865.0	6712.0	1.0
beta1	-0.050	0.004	-0.057	...	6769.0	6946.0	1.0
x_imputed_missing[0]	68.969	3.342	63.626	...	11425.0	7366.0	1.0
x_imputed_missing[1]	68.412	3.273	63.256	...	10615.0	7753.0	1.0
tau	0.164	0.027	0.122	...	6878.0	7562.0	1.0
lik_missing[0]	27.727	4.759	19.890	...	11416.0	7624.0	1.0
lik_missing[1]	18.309	3.133	13.109	...	10239.0	7370.0	1.0
lik_missing[2]	9.311	1.742	6.547	...	10886.0	8376.0	1.0
lik_missing[3]	19.314	3.331	14.064	...	10462.0	6935.0	1.0

2 Dukes' C Colorectal Cancer and Diet Treatment (10 points)

We run the OpenBUGS code and obtain the following results:



	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta0	-10.97	1.57	0.08667	-13.78	-10.96	-7.244	1001	100000
beta1	-1.113	0.4598	0.009813	-2.015	-1.117	-0.2054	1001	100000
beta2	0.09175	0.02113	0.00115	0.04688	0.09156	0.133	1001	100000

(5 points) Based on the results, we see that the estimated values of the coefficients are -10.97 , -1.113 and 0.09175 for β_0 , β_1 and β_2 respectively. Small numerical differences are acceptable for full credit. Equivalent python output:

	mean	sd	hdi_2.5%	hdi_97.5%	...	mcse_sd	ess_bulk	ess_tail	r_hat
beta0	-10.196	2.267	-14.612	-5.819	...	0.036	2019.0	2721.0	1.0
beta1	-1.090	0.458	-1.981	-0.173	...	0.004	7673.0	7133.0	1.0
beta2	0.084	0.029	0.024	0.139	...	0.000	2105.0	2908.0	1.0
alpha	1.663	0.277	1.126	2.208	...	0.003	3300.0	4223.0	1.0

(5 points) As the 95% credible set of coefficient β_1 is $[-2.015, -0.2054]$, we conclude that **linoleic acid treatment has a significant effect on survival time**. Specifically since the mean and credible set are both negative, this means that the treatment **is beneficial** for survival time. Concluding that the treatment is significant is enough for full credit.

Python output for the median survival time of the 2 different groups (optional):

```
Median survival time (months) without treatment:
mean      17.854551
sd         5.226143
hdi_2.5%   10.023245
hdi_97.5%  28.130347
dtype: float64

Median survival time (months) with treatment:
mean      34.780082
sd        11.953286
hdi_2.5%   17.539367
hdi_97.5%  56.829571
dtype: float64
```

Python code snippet for median survival time of the 2 different groups (optional):

```
median0 = pm.Deterministic(
    "median0",
    (np.log(2) * (at.exp(beta0 + beta1 * 0 + beta2 * age)) ** (-1 / alpha)),
)
median1 = pm.Deterministic(
    "median1",
    (np.log(2) * (at.exp(beta0 + beta1 * 1 + beta2 * age)) ** (-1 / alpha)),
)
```