Biostatistics 234: Lab 2

Fernando Mora

1. Summarize briefly the effects on all parameters of changing from prior A to B to C. (Briefly = one sentence total; two only if really necessary).
   Parameters for betas 0 through 3 tended to increase in absolute value as prior precision terms decreased, while the Roof coefficient actually decreased.

2. Give a table of inferences for the coefficient of roofs for the three priors. Briefly explain why it comes out as it does.
   In table 1 below, we see how the estimated coefficient mean for Roof decreases as our prior precision increases from models A to C. This is attributable to the strong co-linearity of Roof and Intercept, a weaker prior precision seems to underestimate the influence of the state of the Roof on the posterior distribution.

Table 1: Estimates and Standard Deviations of Roof Coefficients for 3 Prior Models

Model prior	A	B	C
mean	2.60	1.20	0.86
sd	2.24	4.97	10.59
2.5%	-1.80	-8.58	-20.04
97.5%	6.99	10.90	21.58
p	0.88	0.59	0.53

3. For one of the three priors:

a. Show summaries of the futurefit, futureobs, futuretail in a properly formatted table for the house in perfect condition.

Table 2: Summary of Paramters for Home in Perfect Condition

Parameter	Mean	SD	2.5	97.5	p>0
futurefit	6.23	2.93	0.53	12.00	0.98
futureobs	6.26	7.60	-8.52	21.12	0.79
futuretail	17.64	3.22	11.52	24.13	1.00

b. Which house is in the worst condition? Calculate the three futurefit, futureobs and futuretail variables for this house and provide a formatted table.

Table 3: Summary of Paramters for Home in Worst Condition

Parameter	Mean	SD	2.5	97.5	p>0
futurefit	18.37	2.73	13.07	23.68	1.00
futureobs	18.41	7.52	3.75	33.23	0.99
futuretail	17.64	3.22	11.52	24.13	1.00

4. For prior (C), what two coefficients (including the intercept) have the highest posterior correlation? Briefly explain why.
   The Intercept `beta0` and the `Roof` parameter using prior C had an extremely sharp correlation on a scatterplot, with a correlation coefficient of -0.98. I'd assume the weak prior precision term affects the weight given to beta4 such that the unchanging value of the beta4 vector begins to correlate strongly with the mean.

5. Briefly interpret the three variables futurefit, futureobs, futuretail in your own words.
   `Futurefit` would provide us with the hypothetical cost, in thousands of dollars, to update or repair a home that had some specified scores on the house rating instrument. `Futureobs` would be a predictive distribution of the mean predicted cost of futurefit, with a wider variance. `Futuretail` is the 95th percentage point of the cumulative distribution function, or the value at which we might lose confidence in our estimate for `futuretail.`

6. Suppose we pool the two data sets after the inflation correction. Also, the expert at the housing department told you he thought each unit increase in any rating scale ought to increase the cost by around $1000. You're not sure that all coefficients should be positive. Suggest priors (all regression coefficients and for sigma^2) to use now. Write one or two sentences justifying your priors.
   Possibe priors for such a model might look like:
   $y_i \sim No(\mu_i, \tau)$
   $\mu_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$
   $\beta_j \sim N(\mu = 1, \sigma^2 = 1) for\ j\ in\ 1:4$
   $\tau \sim Gamma(2, 0.5)$
   $\sigma = \frac{1}{\sqrt{\tau}}$
   Where I assume a continuous normal density for outcome `y`, specifying our regression model for μ_i and a precision term τ. The β coefficients according to our expert would equal one in our data, and I chose a wide standard deviation $\sigma = 1$, which would also equal one for σ^2. For tau's shape α and scale β, I relied on the gamma distribution information on wikipedia, and chose a properly conservative density where $\alpha = 2$ and $\beta = 2$. Thus, variance was $\frac{\alpha}{\beta^2} = \frac{2}{2^2} = 0.5$