1. True or False: The mean and standard deviation of the posterior distribution of a slope or intercept parameter in Bayesian linear regression is equal to the least squares estimate and corresponding standard error if the reference prior is used and normally distributed errors are assumed. **1 / 1 point**

- (●) True
- ( ) False

Correct

When the reference prior $p(\beta, \sigma^2) \propto 1/\sigma^2$ is used, the marginal posterior distribution of β follows a t-distribution, with mean and standard deviation equal to the OLS estimate and standard error respectively.

This question refers to the following learning objective(s):

- Understand the basics of Bayesian linear regression and how it relates to Frequentist regression.

2. A simple linear model (either Bayesian or frequentist) that tries to predict an individual's height from his/her age is unlikely to perform well, since human growth rates are non-linear with regard to age. Specifically, humans tend to grow quickly early in life, stop growing at through most of adulthood, and sometimes shrink somewhat when they get old. Which of the following modifications to a simple linear regression model should you prefer? **1 / 1 point**

- ( ) Imposing strong prior distributions on the parameters in a Bayesian analysis.
- ( ) Including other relevant covariates such as weight or income.
- ( ) Log-transforming the dependent variable (height) to account for skewness.
- (●) Including terms of age^2 and or $\log(age)$ as covariates in the model.

Correct

Including transformations of the independent variable such as $\log(age)$ and age^2 as covariates in a model is often a great way to capture non-linear relationships within the context of a linear model, which is easy to work with in both Bayesian and Frequentist settings.

This question refers to the following learning objective(s):

- Identify the assumptions of linear regression and assess when a model may need to be improved.

3. You fit a linear model on 1000 data points and identify a point that lies 3 standard deviations above its predicted value. Should you worry about this potential outlier? Why or why not?

1 / 1 point

- ○ Yes, because outliers can have high leverage and result in a poorly fit model.
- ○ No, because the probability that all 1000 points will be within 3 standard deviations of their predicted values is 0.74, so it is not implausible to observe a point 3 standard deviations away from its predicted value.
- ○ Yes, since the probability of a point deviating from its predicted value by at least 3 standard deviations is roughly 0.003, which suggests that the point is an outlier.
- ◉ No, because the probability that all 1000 points will be within 3 standard deviations of their predicted values is 0.07, so it is unsurprising to observe a point 3 standard deviations away from its predicted value.

✓ **Correct**

This question refers to the following learning objective(s):

- Check the assumptions of a linear model
- Identify outliers and high leverage points in a linear model.

4. Suppose a researcher is using Bayesian multiple regression to quantify the effect of vitamin C on cancer patient mortality. The central 95% posterior credible interval of the coefficient of vitamin C dosage is (-0.19, -0.07). Assuming the model assumptions are valid, what can we say about the effect of vitamin C on cancer patient mortality?

1 / 1 point

- ○ We reject the null hypothesis of no difference, since the 95% credible interval does not include zero.
- ◉ The posterior probability that the coefficient of vitamin C is greater than zero is low, so there is a high posterior probability of a negative association between vitamin C and cancer patient mortality.
- ○ There is not enough information to quantify the effect of vitamin C on cancer patient mortality.

✓ **Correct**

This question refers to the following learning objective(s):

- Interpret Bayesian credible and predictive intervals in the context of multiple linear regression.

5. Which of the following goes into the calculation of the Bayesian Information Criterion (BIC)? **1 / 1 point**

- ○ The maximum value of the log-likelihood under the current model, a constant penalty, and the number of parameters in the model
- ○ The maximum value of the log-likelihood under the current model and the number of parameters in the model
- ◉ The maximum value of the log-likelihood under the current model, the sample size, and the number of parameters in the model
- ○ The maximum value of the log-likelihood under the current model

✓ **Correct**

This question refers to the following learning objective(s):

- Use principled statistical methods to select a single parsimonious model.

6. In a linear model with an intercept term (that is always included) and 4 potential predictors, how many possible models are there? **1 / 1 point**

- ○ 4
- ○ 5
- ◉ 16
- ○ 32

✓ **Correct**

This question refers to the following learning objective(s):

- Implement Bayesian model averaging for both prediction and variable selection.

7. Suppose that a MCMC sampler is currently visiting model B. Model A has a higher posterior probability than model B and Model C has a lower posterior probability than model B. Which of the following statements is true in the MCMC algorithm?

1 / 1 point

- ○ If a jump to Model C is proposed, this jump is never accepted.
- ○ If a jump to Model A is proposed, this jump is never accepted.
- ○ If a jump to Model C is proposed, this jump is always accepted.
- ◉ If a jump to Model A is proposed, this jump is always accepted.

✓ **Correct**

This question refers to the following learning objective(s):

- Understand the importance and use of MCMC within Bayesian model averaging.

8. Which of the following is not an assumption made in Bayesian multiple regression?

1 / 1 point

- ○ The errors have zero autocorrelation.
- ○ The errors have constant variance.
- ○ The errors are independent.
- ◉ The errors follow a t-distribution.

✓ **Correct**

In Bayesian multiple regression, we assume the errors to be **normally** distributed. It is possible to run a Bayesian model with a different distribution of errors, but this is outside the scope of this course.

This question refers to the following learning objective(s):

- Deduce how wrong model assumptions affect model results.

9. Why is the Zellner g-prior useful in Bayesian model averaging?

1 / 1 point

- ◉ It simplifies prior elicitation down to two components, the prior mean and g

- [ ] It helps shrink the coefficients towards 0, which is important if the variables are highly correlated

- [ ] It prevents BMA from disproportionately favoring the null model as a result of the Bartlett-Lindley paradox

✓ **Correct**

This question refers to the following learning objective(s):

- Understand the purpose of prior distributions within Bayesian model averaging.

10. When selecting a single model from an ensemble of models in the case of Bayesian model averaging, which of the following selection procedures corresponds to choosing the "median probability model"? **1 / 1 point**

- [ ] Selecting the model with the highest posterior model probability.

- [ ] Selecting the model that generates predictions most similar to those obtained from averaging over the model space.

- [x] Including only the coefficients with posterior model inclusion probability above 0.5.

✓ **Correct**

The highest probability model simply selects the model with the highest posterior probability. This can be problematic if many diverse models have similar posterior probabilities. Refer to lecture, "Decisions under Model Uncertainty."