Try again once you are ready.

Required to pass: 80% or higher

You can retake this quiz up to 3 times every 8 hours.

Back to Week 4

Retake

0 / 1
point

1.

In a Bayesian simple linear regression $y \sim N(\alpha + x\beta, \sigma^2)$

Suppose our priors on the parameters α, β, σ^2 are independent and that the prior on β is $N(0, 1)$.

Then the posterior mean of β will be closer to zero than the least squares estimate. True or False?

True

False

This should not be selected

By imposing a prior on β with mean 0 and variance 1 the posterior will be a mixture of this prior and likelihood, as such the posterior mean will be at least slightly pulled closer to the prior mean than the least squares estimate, which uses only the likelihood to estimate β.

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

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

---

1 / 1
point

2.

A linear model was estimated using Bayesian methods to predict the height of a male based on his age. All males used in the data are between the ages of 3 to 9 years old. Is it appropriate to use this model to predict the height of a 21 year old man?

- No, since extrapolating outside the range of age observed in the data set may result in a nonsensical prediction.

Correct

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.

- No, since heights may be skewed right, which violates the assumption of normality.
- Yes, as long as proper priors are given to the parameters to ensure that the posterior is proper.
- Yes, an advantage of Bayesian statistics is its ability to generate predictions and express uncertainty in terms of probabilities.

---

0 / 1
point

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?

- 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.

This should not be selected

As the sample size increases, the expected number of points that deviate by k standard deviations also increases. Hint - remember that residuals are normally distributed and hence we can use the command $(1 - 2 * pnorm(-k))^N$ to find the probability that all N points are within k standard deviations of their predicted value.

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.

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.

---

1 / 1
point

4.
Suppose we use Bayesian methods (with a prior distribution) to fit a linear model in order to predict the final sale price of a home based on quantifiable attributes of the home. If the 95% posterior predictive interval of a new home (not in the data set) is (312,096, 392,097), which of the following statements represents a correct interpretation of this interval?

The probability that the house will sell for between 312,096 and 392,097 is 0.95.

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

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

95% of houses with the same attributes as this house have will be sold for prices between 312,096 and 392,097.

95% of houses with the same attributes as this house have historically sold for prices between 312,096 and 392,097.

This house would be sold for between 312,096 and 392,097 95% of the time.

---

0 / 1
point

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

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

This should not be selected

The formula for the BIC is $-2L + \log(n)k$, where L corresponds to the maxmimum value of the log-likelihood, n is the sample size, and k is the number of parameters in the model.

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

- Use principled statistical methods to select a single parsimonious 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
- The maximum value of the log-likelihood under the current model and the number of parameters in the model

---

1 / 1
point

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

- 4
- 5
- 16

Correct

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

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

- 32

---

1 / 1
point

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?

- If a jump to Model C 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.

- If a jump to Model A is proposed, this jump is never accepted.

---

1 / 1
point

8.
Which of the following is **not** a useful method of checking a linear model after it is fit?

- Plotting the residuals to check for non-normally distributed residuals.
- Ensuring that R^2 is as close to 1 as possible.

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

- Deduce how wrong model assumptions affect model results.

- Comparing the distribution of fitted values to the distribution of observed data.
- Examining the influence of potential outliers on the parameters of the model.

---

0 / 1
point

9.
Which of the following is an advantage of using the Zellner-Siow-Cauchy prior in Bayesian model averaging?

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

This should not be selected
For a fixed value of g, the Bartlett-Lindley paradox means that as the sample size increases, the null model will receive higher posterior probability. The Zellner-Siow-Cauchy prior allows g to increase with n to avoid this problem while allowing for uncertainty about g, assigning n/g a Gamma prior distribution.

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

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

- It prevents BMA from disproportionately favoring the null model as a result of the Bartlett-Lindley paradox
- It allows for uncertainty in the prior variance parameter g
- Both b and c

---

1 / 1
point

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 "highest probability model"?

- Selecting the model that generates predictions most similar to those obtained from averaging over the model space.
- Selecting the model with the highest posterior model probability.

Correct
The median probability model includes only the coefficients with posterior model inclusion probabilities above 0.5.

Including only the coefficients with posterior model inclusion probability above 0.5