Exam 2

March 30, 2026

For all problems, you may assume that we have loaded the `alr4` package with `library(alr4)`.

1. (a) In multiple regression, the residual plots must follow the number of predictors given (for example, if we have 3 predictors, then we must have 3 residual plots for this model).
   i. True
   ii. False

   (b) In multiple regression, the degrees of freedom used for the p-values from the t-tests of the coefficients
   i. Is 1
   ii. Is p
   iii. Are 1 for the intercept, p for each input variable
   iv. None of the above.

   (c) Consider a dataset with response Y and predictors A, B and C. Let $D = C - B - A$ be a new predictor. If we fit regression models `Y~A+B+C` and `Y~A+B+D`,
   i. The estimates for C and D must be the same.
   ii. The estimates for C and D must be different.
   iii. The estimates for C and D can be the same or different depending on data.
   iv. The model `Y~A+B+D` will not run.

   (d) Suppose that we have added variable plots of a regression model with multiple predictors (say, X1, X2, X3). If we remove one of the variables (say, X3) in the original model, the slopes of X1 and X2 in the added variable plot of the new model.
   i. Must be the same as the slopes of the new multiple regression model.
   ii. Must have the same sign as the slopes of X1 and X2 of the original model.
   iii. Both of the above
   iv. None of the above

   (e) The slope lines in the effects plots for a multiple regression model must pass through
   i. The origin
   ii. The mean of all X values on the x-axis
   iii. The mean of all X values on the y-axis
   iv. None of the above

2. Consider the dataset `MinnWater`

   (a) If we want to consider a model with `log(muniUse)` as the response, and `year` and `muniPrecip` as predictors, write R codes that will compute the 98% prediction interval for the `year` 2015 and `muniPrecip` value of 20.

   (b) Repeat part (a) with matrix computation in R (assume here that values of n, $\mathbf{X}$, $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ are already computed).

3. Suppose that we are given a regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$$

   for $i = 1, 2, 3$. Let Y have the values 5,3,4, let X_1 have the values -1,0,1, and let X_2 have the values -1,2,-1.

   (a) Write down the R codes (with matrix computation) to compute $\hat{\beta}_2$.

   (b) Suppose that $\hat{\sigma}^2$ is known and given by $\hat{\sigma}^2 = 3$. Write down the R codes (with matrix computation) to compute $\text{se}(\hat{\beta}_2)$; you may assume and use the codes from part (a) above.

   (c) (BONUS) Suppose that $\hat{\sigma}^2$ is known and given by $\hat{\sigma}^2 = 3$. Calculate $\text{se}(\hat{\beta}_2)$ without R.

4. Suppose that we are given a regression model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, with $E(\mathbf{e}) = \mathbf{0}$ and $\text{Var}(\mathbf{e}) = \sigma^2\mathbf{I}$. If $\mathbf{1}_n$ is an $n \times 1$ vector of 1's, show that $\text{Var}(\mathbf{1}_n'\widehat{\mathbf{Y}}) = \sigma^2 \cdot c$, and find c.