Andrew J. Spieker, PhD
BIOS 7345 - Advanced Regression Analysis I
Collection of problems for Fall 2024 (Version: 12/02/2025)

Instructions: Responses are due by Box by close-of-business on the due date (5:00p on Fridays). You should word-process your responses (e.g., with LaTeX) for the more data-analytic or simulation-based problems. It is acceptable to hand-write and scan your responses to problems that are more mathematically intensive. When you use software, you should **always** turn in your annotated code as an appendix.

---

1. Let $\mathbf{X}$ be the 2×2 matrix shown below. You may use software for this problem.

$$\mathbf{X} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

   (a) Use five distinct arguments/methods of your choosing to conclude that $\mathbf{X}$ is invertible.

   (b) Without actually computing them by hand, state anything you can about the eignvalues and eigenvectors of $\mathbf{X}$ that can be learned merely from examining $\mathbf{X}$.

   (c) Determine the singular value decomposition of $\mathbf{X}$, and with it, determine $\mathbf{X}^{1/2}$.

2. (a) Express the quadratic form $y = 2x_1^2 + 2x_1x_2 + x_2^2$ in the form $y = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ (whenever you are going through an exercise involving this sort of machinery, $\mathbf{A}$ should *always* be chosen to be symmetric). With the assistance of software (e.g., Wolfram Alpha), briefly comment on the geometry and "definiteness" of this quadratic form.

   (b) Express the quadratic form $y = x_1^2 + 4x_1x_2 + x_2^2$ in the form $y = \mathbf{x}^\top \mathbf{A}\mathbf{x}$. With the assistance of software, briefly comment on the geometry and "definiteness" of this quadratic form.

   (c) Express the quadratic form $y = x_1^2 + 2x_1x_2 + x_2^2$ in the form $y = \mathbf{x}^\top \mathbf{A}\mathbf{x}$. With the assistance of software, briefly comment on the geometry and "definiteness" of this quadratic form.

3. Let $\mathbf{A}$ be a 4×2 matrix defined as follows:

$$\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

   (a) Compute the matrix $\mathbf{P} = \mathbf{A}(\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top$.

   (b) Without actually computing it, argue that $\det(\mathbf{P}) = 0$.

   (c) State the two natural vectors $\mathbf{c}$ that should have the property $\mathbf{P}\mathbf{c} = \mathbf{c}$.

   (d) Without doing the computation, argue that $\mathbf{c} = (0, 1, 1, 2)^\top$ also has the property $\mathbf{P}\mathbf{c} = \mathbf{c}$. Argue that $\mathbf{c} = (0, 1, 1, 3)^\top$, on the other hand, does not have the property $\mathbf{P}\mathbf{c} = \mathbf{c}$.

   (e) Determine the eigenvalues and eigenvectors of $\mathbf{P}$. You can use software as an aid to solve the computationally obnoxious parts of this problem, although you should otherwise show your work and explain your reasoning along the way.

4. Let $\mathbf{c}(\boldsymbol{\beta}) = \mathbf{X}^\top \text{diag}((\mathbf{x}_i^\top \boldsymbol{\beta})^2)(\mathbf{y} - \text{vec}((\mathbf{x}_i^\top \boldsymbol{\beta})^{-1}))$, where $\mathbf{X}$ is an $N \times K$ matrix $(N > K)$, $\mathbf{x}_i$ is a column vector of its i^{th} row, $\mathbf{y}$ is a length-N vector, and $\boldsymbol{\beta}$ is a length-K vector. Use matrix calculus rules to determine $\partial \mathbf{c}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ (you may assume that $\mathbf{x}_i^\top \boldsymbol{\beta} > 0$ for $i = 1, \ldots, N$).

5. Let $\mathbf{X}$ be an $N \times K$ matrix with $K < N$ and $\text{rank}(\mathbf{X}) = K$ ($\mathbf{X}$ has "full rank" in that its rank is the highest possible given its dimensions). Further, let $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ denote its SVD.

   (a) Verify that $(\mathbf{V}\mathbf{D}^\top\mathbf{D}\mathbf{V}^\top)^{-1} = \mathbf{V}(\mathbf{D}^\top\mathbf{D})^{-1}\mathbf{V}^\top$.

   (b) Write an expression for $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ of the form $\mathbf{U}\mathbf{A}\mathbf{U}^\top$, where $\mathbf{A}$ is some numeric matrix that does not depend upon $\mathbf{X}$, $\mathbf{V}$, or $\mathbf{D}$.

   (c) What are the singular values of $\mathbf{P}$?

   (d) Write an expression for $\mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ of the form $\mathbf{U}\mathbf{B}\mathbf{U}^\top$, where $\mathbf{B}$ is some matrix that does not depend upon $\mathbf{X}$, $\mathbf{V}$, or $\mathbf{D}$.

   (e) What are the singular values of $\mathbf{I} - \mathbf{P}$?

6. Let $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with

$$\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & p \\ 0 & \rho & 1 \end{pmatrix}.$$

   Let $X_1 = Y_1 + Y_2 + Y_3$ and let $X_2 = Y_1 - Y_2 - Y_3$. Determine the value(s) of ρ for which X_1 and X_2 are independent.

7. Let $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Determine the values of c_1 and c_2 for which $c_1(X_2 - X_1)^2 + c_2(X_1 + X_2)^2 \sim \chi^2_2$.

8. Let $\mathbf{x} = (X_1, \cdots, X_N)^T$ denote a random vector of length N with $\mathbf{E}[X_i] = \mu_i$ and $\text{Cov}[\mathbf{x}] = \mathbf{I}$.

   (a) Suppose $Y = \left(\sum_{i=1}^N X_i\right)^2$. Show that $\mathbf{E}[Y] = N + \left(\sum_{i=1}^N \mu_i\right)^2$.

   (b) Suppose $N = 3$ and that $\mathbf{x}$ comprises three standard normal random variables. Let $Z = X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_2X_3 + X_1X_3$. Show that $Z \sim \text{Exponential}(\lambda = 1/3)$.

9. Suppose that $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ denotes a length-three multivariate normal random vector, and let $Z = 2(Y_1Y_2 - Y_2Y_3 - Y_1Y_3)$.

   (a) Write $Z = \mathbf{y}^\top\mathbf{A}\mathbf{y}$ (i.e., as a quadratic form) for some matrix $\mathbf{A}$ that you determine.

   (b) Show that MGF of Z is given by $M_z(t) = (1 - 4t)^{-1/2}(1 + 2t)^{-1}$. You may use software as an aid to compute eigenvalues (that's really also giving you a hint). As another hint, I recommend starting with the definition of the MGF.

   (c) Argue that $Z \stackrel{d}{=} 2U_1 - U_2 - U_3$, where U_1, U_2, and U_3 are independent χ^2_1 random variables.

10. Let $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with $\mathbf{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ and $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}$. let $\mathbf{X}$ denote an $N \times K$ fixed matrix of full rank (including a column of ones for the intercept), and let $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$.

   (a) Derive expressions for $\mathbf{E}[\widehat{\boldsymbol{\epsilon}}]$, $\text{Cov}[\widehat{\boldsymbol{\epsilon}}]$, and $\text{Cov}[\widehat{\boldsymbol{\epsilon}}, \mathbf{P}\mathbf{y}]$.

   (b) Prove that $\sum_{i=1}^N (y_i - \widehat{y}_i) = 0$. This will be quite a short proof if you stick to linear algebra notation.

11. Let $X_i = 1(i > n)/\sqrt{2} - 1(i \leq n)/\sqrt{2}$, and let $Y_i = X_i\beta + \epsilon_i$ for $i = 1, \ldots, 2n$, with i.i.d. errors of mean zero.

   (a) Determine an expression for $\widehat{\beta}_1^{\text{OLS}}$ from an OLS fit to the model $\mathbf{E}[Y|X = x] = \beta_0 + \beta_1 x$; argue that $\widehat{\beta}_1^{\text{OLS}}$ is the BLUE of β despite the fact that it arises from a model that unnecessarily estimates an intercept.

   (b) Argue that the sample mean of Y among observations for which $X = 1/\sqrt{2}$ is *not* the BLUE of $\mathbf{E}[Y|X = 1/\sqrt{2}]$ if you know that, in truth, $\beta_0 = 0$. Determine the BLUE of $\mathbf{E}[Y|X = 1/\sqrt{2}]$ as part of your response.

   (c) Let $\beta = 1$. Illustrate via simulation that the value of $\text{Var}[\sqrt{n}(\widehat{\beta}_1 - 1)]$ does not depend upon error normality. Specifically, consider sample sizes of $n = 10$, $n = 100$, and $n = 1000$; consider cases (1) $\epsilon_i \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$, and (2) $\epsilon_i \sim \text{Laplace}(\text{“mean”} = 0, \text{“variance”} = 1)$.

   (d) Illustrate that $(\mathbf{x}^\top \mathbf{x})^{1/2}(\widehat{\beta}_1 - 1) \xrightarrow{d} \mathcal{N}(0, 1)$ by simulation for each case in part (c).

   (e) Should any of your conclusions from parts (c) and (d) change if the values of X are randomly sampled (i.e., taking on values $X = -1/\sqrt{2}$ and $X = 1/\sqrt{2}$ with equal probability)? Verify your answer by re-running the simulations of (c) and (d) under this setup.

12. Suppose $Y_i = \beta x_i + \epsilon_i$, $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ for $i = 1, \ldots, N$. Assume that $1 \leq x_i < \infty$ $\forall i$ with the x_i's fixed and distinct. Consider the following two estimators of $\beta \in \mathbb{R}$:

$$\widehat{\beta} = \frac{\sum_{i=1}^N x_i Y_i}{\sum_{i=1}^N x_i^2} \qquad \text{and} \qquad \widetilde{\beta} = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N x_i}.$$

   (a) Argue that $\text{Var}[\widehat{\beta}] < \text{Var}[\widetilde{\beta}]$ two ways: (1) Compute $\text{Var}[\widehat{\beta}]$ and $\text{Var}[\widetilde{\beta}]$, and then compare; and (2) invoke a key theorem and carefully justify why it applies.

   (b) Suppose $\text{Var}[\epsilon_i] = \sigma^2$. Determine a sequence of x_i's for which $\widehat{\beta}$ is not consistent for β. *Hint*: You'll want to remove the condition $x_i \geq 1$, and you may need to look up (or ask about) the behavior and convergence properties of certain infinite series.

13. Consider an ANOVA-style model involving a comparison of three treatment categories, each group having a total of n observations. Specifically, for $i = 1, \ldots, n$ and $j = 0, 1, 2$, suppose $Y_{ij} = \alpha_j + \epsilon_{ij}$, with $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

   (a) Write down the design matrix for this problem. What is its rank?

   (b) Consider the hypothesis $H_0 : \alpha_0 = \alpha_1 = \alpha_2$. Write down H_0 in the form $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.

   (c) Write down the restricted model under the null, H_0.

   (d) Showing your work, determine the (unrestricted) OLS estimate of $\boldsymbol{\alpha}$.

   (e) Determine an expression for the F-statistic that could be used to test H_0.

   (f) Assuming H_0 is true, what is the distribution of the F-statistic you determined in (e)?

   (g) Assuming instead that $\alpha_2 - \alpha_1 = \alpha_0 - \alpha_1 = 1$, what is the distribution of the F-statistic you determined in part (e)?

   (h) Assuming the same condition on $\boldsymbol{\alpha}$ put forth in part (g), and also assuming $\sigma^2 = 5$, determine the power of a 0.05-level F-test (one-sided/right-tailed) under per-group sample sizes of $n = 5$, $n = 10$, and $n = 20$. Confirm your answers with a simulation study.

14. Suppose $Y_i \sim \mathcal{N}(\beta x_i, \sigma^2)$ for independent observations $i = 1, \ldots, N$. Assume $1 \leq x_i < \infty \ \forall i$ with the x_i's fixed and not all the same value. Consider the OLS estimator of $\beta \in \mathbb{R}$:

$$\widehat{\beta} = \frac{\sum_{i=1}^N x_i Y_i}{\sum_{i=1}^N x_i^2}.$$

   (a) Suppose that $N = 3n$, with $x_i = 1$ for $i = 1, \ldots, n$, $x_i = 2$ for $i = n+1, \ldots, 2n$, and $x_i = 3$ for $i = 2n+1, \ldots, 3n$. Determine a suitable test statistic, F, for the hypothesis $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$ based on $\widehat{\beta}$. Then (assuming a level of $\alpha = 0.05$):

      (I) Determine the power to detect $\beta = 0.3$ if $N = 60$ ($n = 20$) and $\sigma^2 = 5$.

      (II) Determine the minimum sample size needed to detect $\beta = 0.5$ with 80% power if $\sigma^2 = 5$.

      (III) Determine the minimum value of β necessary to achieve 90% power if $N = 30$ ($n = 10$) and $\sigma^2 = 3$.

   You may (i.e., *should*) use R functions such as `pf()` and `qf()` to obtain your answers—but note that some amount of "guess-and-check" may be needed for (II) and (III).

   (b) Conduct and present the results of a simulation that confirms your answers to part (a).

   (c) Consider Part (I) of (a). Suppose that, rather than being in the case of normally distributed errors, we had that $Y_i = \beta x_i + \epsilon_i$, with $\epsilon_i \sim \text{Exponential}(\lambda = 1/\sqrt{5}) - \sqrt{5}$ (these errors have a mean of zero and a variance of five). Conduct a simulation study to investigate how this alters the power of the F test based on normal theory.

15. Suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X}$ is a fixed, full-rank $N \times K$ design matrix, $\mathbf{E}[\boldsymbol{\epsilon}] = \mathbf{0}$, and $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2 \mathbf{V}$ for a diagonal $\mathbf{V}$ with $V_{ii} > 0$ and $\sum_{i=1}^N \mathbf{V}_{ii} = 1$. Let $\widehat{\boldsymbol{\beta}}^*$ denote the weighted least squares estimate based on weights $\mathbf{W} = \mathbf{V}^{-1}$. Prove each of the following properties:

   (a) $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{V}^{-1}$ is idempotent.

   (b) $\mathcal{C}(\mathbf{P}) = \mathcal{C}(\mathbf{X})$ (I recommend showing any vector in one subspace is also in the other).

   (c) $\mathbf{E}[\widehat{\boldsymbol{\beta}}^*] = \boldsymbol{\beta}^*$ and $\text{Cov}[\widehat{\boldsymbol{\beta}}^*] = \sigma^2 (\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}$.

   (d) $\text{Cov}[\widehat{\boldsymbol{\epsilon}}] = \sigma^2 (\mathbf{I} - \mathbf{P}) \mathbf{V}$.

   (e) $\mathbf{E}[\text{RSS}] = \sigma^2 (N - K)$, where $\text{RSS} = (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^*)^\top \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^*)$.

16. Load the data set `fev.csv` and read the corresponding documentation. Suppose we seek to evaluate height as a predictor of mean FEV. Specifically, let Y denote FEV (L) and let $X = (\text{Height}/12)^3$ denote cubed height in cubic feet, which is a transformation that greatly improves the fit of the linear model $\mathbf{E}[Y|X = x] = \beta_0 + \beta_1 x$. Generate a scatter plot of X and Y. Construct a point estimate and 95% confidence interval for β_1 the following four ways:

   (a) Ordinary least squares.

   (b) A weighted least squares estimate based on fixed weights given by $W_i = X_i^{-2}$.

   (c) Iteratively re-weighted least squares with variance model $\text{Var}[Y|X = x] = (\theta_1 + |\widehat{Y}|^{\theta_2})^2$ (you can use the `gls()` function in R).

   (d) Iteratively re-weighted least squares with variance model $\text{Var}[Y|X = x] \propto \beta_0 + \beta_1 x$ (you should hard-code this one).

17. This problem is fundamentally about the "one-step" estimator. Suppose that for $i = 1, \ldots, N$ (independent observations), we have $X_i \sim \text{Uniform}(1, 30)$, either fixed or random—it doesn't matter—and $Y_i = 20 + 15X_i + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2 = 500^2/(20 + 15X_i))$. Consider estimating β_1 the following five ways:

   (I) Ordinary least squares.
   (II) IRLS (iterated until convergence) with $\boldsymbol{\beta}^{(0)}$ as the OLS estimate from method (I).
   (III) IRLS (iterated until convergence), with $\boldsymbol{\beta}^{(0)} = (1, 0)^\top$.
   (IV) $\boldsymbol{\beta}^{(1)}$, the first step beyond the initializer, from the IRLS procedure of method (II).
   (V) $\boldsymbol{\beta}^{(1)}$, the first step beyond the initializer, from the IRLS procedure of method (III).

   Note that methods (II) and (III) involve a weighting scheme of $\mathbf{W} \propto \mathbf{V}^{-1}$. Conduct a simulation study (M = 10,000 iterations) under a sample size of $N = 500$ in which you determine the mean estimate and empirical standard error across the five approaches. Comment on and account for patterns of note (you do not need to mathematically prove your answers).

18. Let $N = 100$, with $X_i \sim \mathcal{N}(0, 1)$ and $Z_i \sim \mathcal{N}(0, 1)$; $Z \perp\!\!\!\perp X$, either fixed or random—it doesn't matter. Let $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i^2 + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, 1)$. Illustrate via simulation (with M = 10,000 iterations) that the OLS estimate of α_1 from the model linear regression $\mathbf{E}[Y|X = x, Z = z] = \alpha_0 + \alpha_1 x + \alpha_2 z$ is unbiased for β_1. Although you needn't prove it, it conveniently turns out that the more general case is true: you can misspecify the functional form of the "Z" component in the model and nevertheless achieve unbiasedness.

19. Suppose the true data generating mechanism is given by $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}$, where $\mathbf{E}[\boldsymbol{\epsilon}] = \mathbf{0}$ and $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{I}$. However, we fit the "larger" model $\mathbf{E}[\mathbf{y}|\mathbf{X}_1, \mathbf{X}_2] = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$. Note that the dimensions of $\mathbf{X}_1$ and $\mathbf{X}_2$ are $(N \times K_1)$ and $(N \times K_2)$, respectively.

   (a) Verify (the easy way) the following formula for the inverse of a 2×2 block matrix:
   $$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix}$$

   (b) Use your answer to part (a) to show that $\text{Cov}[\widehat{\boldsymbol{\beta}}]$ can be written in the following form:
   $$\text{Cov}[\widehat{\boldsymbol{\beta}}] = \sigma^2 \begin{bmatrix} (\mathbf{X}_1^\top\mathbf{X}_1)^{-1} + \mathbf{F}\mathbf{G}^{-1}\mathbf{F}^\top & -\mathbf{F}\mathbf{G}^{-1} \\ -\mathbf{G}^{-1}\mathbf{F}^\top & \mathbf{G}^{-1} \end{bmatrix},$$
   specifically stating the matrices $\mathbf{F}$ and $\mathbf{G}$ in terms of $\mathbf{X}_1$ and $\mathbf{X}_2$. Note that $\text{Cov}[\widehat{\boldsymbol{\beta}}_1]$ is marked by the upper left-hand block.

   (c) Assuming that $K_1 \leq K_2$, show that $\mathbf{F}\mathbf{G}^{-1}\mathbf{F}^\top$ is positive definite unless $\mathbf{X}_1^\top\mathbf{X}_2 = \mathbf{0}$.

   (d) Briefly state the implications of the result you obtained in part (c).

20. Consider a simple linear regression model $\mathbf{E}[Y|X = x] = \beta x$ in which X and Y each have been centered to have mean zero so that there is no intercept, and X has been scaled so that $\sum_i x_i^2 = 1$. Consider estimation by OLS. Show that the graph of the upper confidence interval for $\mathbf{E}[Y|X = x]$ is hyperbolic. You may need to refresh your memory on the analytic geometry of conic sections and second-degree equations; reach out if you need to be pointed in the right direction. Showing that the confidence interval is actually a function of x is harder and you do not need to do this (although we may discuss it).

21. Consider the projection matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$, where (in this problem) the $N \times K$ design matrix, $\mathbf{X}$, is of full rank and has a column of ones for the intercept. Now that we've learned its connection to leverage, we can derive even more of its useful properties! Each problem includes a gentle hint, but reach out if you need more direction.

    (a) Show that $\sum_{i=1}^{N} \mathbf{P}_{ii} = K$. *Hint*: Use a property of the trace.

    (b) Show that $\sum_{i=1}^{N} \mathbf{P}_{ij} = \sum_{j=1}^{N} \mathbf{P}_{ij} = 1$. *Hint*: Use the fact that $\mathbf{X}$ has a column of ones.

    (c) Show that $\mathbf{P}_{ii}^2 + \sum_{j \neq i} \mathbf{P}_{ij}^2 = \mathbf{P}_{ii}$. *Hint*: Use the idempotence of $\mathbf{P}$.

    (d) Show that if $\mathbf{P}_{ii} = 1$, then the OLS fit passes through the point $(\mathbf{x}_i, Y_i)$. *Hint*: Use the result of part (c).

    (e) Show that $(1 - \mathbf{P}_{ii})^2 + \sum_{j \neq i} \mathbf{P}_{ij}^2 = 1 - \mathbf{P}_{ii}$. *Hint*: Begin by using the idempotence of $(\mathbf{I} - \mathbf{P})$.

    (f) Show that $N^{-1} \leq \mathbf{P}_{ii} \leq R_i^{-1}$, where R_i denotes the number of times that the observation $\mathbf{x}_i$ appears in the data set. Assume the covariates to be centered to have mean zero. *Hint*: Write $\mathbf{X}^\top\mathbf{X}$ as a block-diagonal matrix with a real-valued entry in the upper-left block and a $(K-1) \times (K-1)$ matrix in the lower-right block.

22. This problem is about what can go wrong when you fail to center a predictor prior to regularization. Consider the setting in which you seek to estimate shrunken coefficients from the simple linear regression model $\mathbf{E}[Y|X = x] = \beta_0 + \beta_1 x$ via the ridge penalty. For simplicity, and without any serious loss to generality, consider X to be uniformly distributed between 0 and 1. Given a sample size of $N > 2$, define the leverage for an observation $\mathbf{x} = (1, x)$ as:

    $$P_\lambda(x) = \mathbf{x}^\top(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{x},$$

    where $\lambda \geq 0$ marks the penalty and $\mathbf{X}$ is the $N \times 2$ design matrix for the uncentered data.

    (a) Briefly argue that $P_\lambda(x) > 0$.

    (b) Determine the value x_λ, at which $P_\lambda(x)$ is minimized. Conclude that $x_\lambda < x_0$ for $\lambda > 0$.

    (c) Show that $P_0(x) > P_\lambda(x)$ for $\lambda > 0$. *Hint*: Use Lemma 18.3.

    (d) Argue graphically/heuristically that for $\lambda > 0$, $P_\lambda(x)$ is not a function of $P_0(x)$.

    (e) Characterize the behavior of $P_\lambda(x)$ as $\lambda \nearrow \infty$ (i.e., for a fixed $N > 2$).

    (f) Characterize the behavior of $P_\lambda(x)$ as $N \nearrow \infty$ (i.e., for a fixed $\lambda > 0$).

23. Consider independent observations $(\mathbf{x}_1, Y_1), \ldots, (\mathbf{x}_N, Y_N)$, with the values of $\mathbf{x}_i$ fixed and known in advance, and σ^2 the shared outcome variance. Recall:

    $$\text{df}(\widehat{\mathbf{y}}) = \frac{1}{\sigma^2}\sum_{i=1}^{N} \text{Cov}[Y_i, \widehat{Y}_i],$$

    Let $\widehat{\boldsymbol{\beta}}_\lambda$ denote a penalized least squares estimate based on a ridge penalty of λ.

    (a) Prove that $\text{df}_\lambda(\widehat{\mathbf{y}}) = \text{tr}(\mathbf{P}_\lambda) = \text{tr}(\mathbf{X}(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^\top)$.

    (b) Let $d_1, \ldots, d_K$ denote the singular values of $\mathbf{X}$. Prove that

    $$\text{df}_\lambda(\widehat{\mathbf{y}}) = \sum_{k=1}^{K} \frac{d_k^2}{d_k^2 + \lambda}.$$

24. Let $\mathbf{X}$ denote an $N \times K$ design matrix of covariates that are fixed in advance, and each standardized to have mean zero and variance one; further assume that $K \leq N$. Further, let $\mathbf{Y}$ denote an $N \times 1$ outcome vector (centered to have mean zero).

   (a) Characterize the set, Λ, of all possible eigenvalues of $\mathbf{X}^\top\mathbf{X}$; characterize the set $\Lambda^* \subseteq \Lambda$, of all possible eigenvalues of $\mathbf{X}^\top\mathbf{X}$ such that $\mathbf{X}^\top\mathbf{X}$ non-singular.

   (b) Characterize all values of $\lambda \in \mathbb{R}$ such that $\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}$ is non-singular. For what values of $\lambda \in \mathbb{R}$ is $\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}$ is positive definite?

   (c) Consider the following $N = 5$ independent observations based on three covariates (observations not yet centered/scaled):

ID	X_1	X_2	X_3	Y
1	−1	1	0	−1
2	−1	1	0	2
3	0	0	0	−5
4	0	1	1	5
5	1	0	1	−1

   Let $\widehat{\boldsymbol{\beta}}_\lambda$ denote a solution (if any) to the penalized normal equations, $\mathbf{X}^\top\mathbf{y} = (\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I})\boldsymbol{\beta}$, for a given λ (following centering/scaling). It turns out that a *negative* ridge penalty is capable of producing a solution that minimizes expected prediction error, though your answer to parts (a)-(b) may prompt concerns about this. Graph each component of $\widehat{\boldsymbol{\beta}}_\lambda$ and $\widehat{\mathbf{y}}_\lambda$ over the range `lambda=seq(-5*pi, 0, 0.2)`. Comment on and account for your findings. Use software as an aid in the computationally cumbersome calculations.

25. Consider the following random draw as a toy data set:

```
set.seed(7345)
x <- runif(10, 0, 5)
Y <- 1 + x + rnorm(10, 0, 5)
```

   Let $\boldsymbol{\beta}|\tau \sim \mathcal{N}(\boldsymbol{\phi}, (1/\tau)\mathbf{V})$ and $\tau \sim \text{Gamma}(\alpha, \delta)$ denote the prior of a Bayesian analysis. We seek to understand how the posterior is impacted by different aspects of the setup (you may use the results presented in class). Consider the following as the default setup:

   - $\boldsymbol{\phi} = (\phi_0, \phi_1)^\top$, where $\phi_0 = 0$ and $\phi_1 = 1$.
   - $\mathbf{V} = v\mathbf{I} + c(\mathbf{J} - \mathbf{I})$ with $v = 3^2$ and $c = 1^2$.
   - $\alpha = 2$ and $\delta = 0.04$.

   For each of parts (a)-(e), produce graphs of the posterior mean and variance for β_1 under the default setup except varying the stated parameter as noted.

   (a) Vary ϕ_0 from -1 to 1.

   (b) Vary ϕ_1 from 0 to 2.

   (c) Vary v from 2^2 to 10^2.

   (d) Vary c from $-(1^2)$ to 2^2.

   (e) Vary α from 2 to 5 and let $\delta = 0.04(\alpha - 1)$ for each α.

26. This problem serves primarily to provide you with experience implementing computational Bayesian procedures for simple linear regression. To that end, generate the following toy data set for this problem:

```
set.seed(7345)
n <- 25
x <- rnorm(n, 5, 2)
y <- 10 + 2 * x + rnorm(n, 0, 5)
```

Keeping in mind that this problem is fundamentally a skill-building exercise and that the particulars of the setup that follows may not reflect the choices a Bayesian might make in the real world, consider the following priors: $\beta_0 \sim \mathcal{N}(\mu = 0, \sigma^2 = 50^2)$, $\beta_1 \sim \text{Gamma}(\alpha = 1, \delta = 1/2)$, and $\sigma \sim \text{Weibull}(\lambda = 6, k = 1.5)$, all independent.

(a) What key "hit-you-in-the-face" belief about the data generating mechanism is being conveyed by this choice of a prior distribution? Provide an example of a research question in which this might be inappropriate, and another in which it might very well be reasonable.

(b) Explain the appeal of keeping the three proposal functions in the same family as the three prior densities specified above.

(c) Describe an advantage and a disadvantage of choosing parameters for the proposal functions that render them approximately symmetric.

(d) Given three proposal windows ω_1, ω_2, and ω_3 that you will need to later specify (one for each of β_0, β_1, and σ, respectively), outline a Metropolis-Hastings algorithm to numerically generate draws from the posterior distribution. Implement it in R (in the process, you will need to decide on values for the proposal windows, decide where the burn-in period ends, and assess autocorrleation. You do not need to show the false starts that lead to your final choices, but you should show the diagnostics for your final choices).

(e) Obtain the following from the final posterior draws:

- The posterior mean and median for β_1.
- A quantile-based and a highest-posterior-density 95% credible interval for β_1.
- $\pi(\beta_1 > 1)$ and $\pi(\beta_1 > 1|\mathbf{y};\mathbf{x})$.

(f) Compare the posterior mean and median to one another and to $\widehat{\beta}_1^{\text{OLS}}$.

(g) Compare the 95% credible intervals to one another and to a 95% CI for β_1 from OLS.

(h) Compare $\pi(\beta_1 > 1|\mathbf{y};\mathbf{x})$ to a p-value for a test of $H_0 : \beta_1 \leq 1$ vs. $H_1 : \beta_1 > 1$.

27. This is a continuation of problem 26. The posterior predictive distribution for a future observation, $\tilde{y}(x)$, often has an analytically intractable form, but can be expressed as:

$$p(\tilde{y}(x)|\mathbf{y};\mathbf{x}) = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} p(\tilde{y}(x)|\boldsymbol{\theta};\mathbf{x})\pi(\boldsymbol{\theta}|\mathbf{y};\mathbf{x})d\boldsymbol{\theta}.$$

Obtaining numerical samples from the posterior predictive distribution turns out to be fairly straightforward post-MCMC. For each accepted draw $\boldsymbol{\theta}^{(j)}$ (following a burn-in period, of course), generate a random draw $y^{(j)}(x)$ from the density $f_{\theta^{(j)}}(y;x)$ at the fixed value of x under consideration. The collection of random draws constitutes a numerical approximation to the desired posterior predictive distribution. Determine the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles of the posterior predictive distribution for $\widetilde{y}(1)$, $\widetilde{y}(3)$, $\widetilde{y}(5)$, $\widetilde{y}(7)$, and $\widetilde{y}(9)$. Compare these ranges to those of 95% prediction intervals as derived from OLS.

28. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are fixed and known in advance, and that Y_i is presumed to follow a Poisson distribution that depends upon the value of $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$.

   (a) Factor the probability mass function for $Y \sim \text{Poisson}(\lambda)$ into canonical form. From this, identify or determine the following:

      - The natural parameter, θ, and the nuisance parameter, ϕ.
      - The values of $\mathbf{E}[Y]$, and $\text{Var}[Y]$.
      - The canonical link function associated with a GLM of $\mathbf{y}$ on $\mathbf{X}$.

   (b) Based on a GLM of $\mathbf{y}$ on $\mathbf{X}$ using the canonical link, identify or determine the following:

      - The mean model, and the mean-variance relationship, $V(\mu)$.
      - The score equations to solve for $\boldsymbol{\beta}$.
      - The formulas for a single iteration of a Newton-Raphson and a Gauss-Newton step.
      - A (likelihood-based) estimator of $\text{Cov}[\widehat{\boldsymbol{\beta}}]$.

   (c) Repeat part (b) with the choice of the identity link function.

29. Repeat problem 28, this time with Y_i following a $\text{Gamma}(\alpha, \beta)$ distribution that depends upon the value of $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ (*Hint*: though it seems weird, let $\theta = -\beta/\alpha$ and $\phi = 1/\alpha$).

30. This is a continuation of problem 29. Generate data according to the following code:

```
set.seed(7345)
n <- 500
X <- matrix(cbind(1, runif(n,1,5)), ncol = 2)
y <- rgamma(n, shape=2, rate=-2*(-1 - X[,2]/5))
```

   Hard-code the appropriate GLM to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{\text{SE}}(\widehat{\boldsymbol{\beta}})$; compare your answer that produced by the `glm()` function in R; account for the single most obvious difference you see.

31. Load the data set `assay.csv` and read the (very brief) documentation.

   (a) Create a scatter plot of the immunofluorescence assay (X) and the thermal assay (Y). Recognizing that linearly is closely approximated, investigate the heteroscedasticity using generalized least squares: `gls(therm~immuno, weights=varPower(form=~fitted(.)))`

   (b) Hard-code the following Gaussian GLMs; overlay the fitted curves on the scatter plot.

      (I) Identity link: $Y \sim \mathcal{N}(\mu = \beta_0 + \beta_1 X, \sigma^2)$.

      (II) Log link: $Y \sim \mathcal{N}(\mu = \exp(\beta_0 + \beta_1 X), \sigma^2)$.

      (III) Inverse link: $Y \sim \mathcal{N}(\mu = 1/(\beta_0 + \beta_1 X), \sigma^2)$.

   (c) Repeat part (b) with three Poisson GLMs; hard-code the GLM that corresponds to the canonical link and use the `glm()` function for the other two.

   (d) Repeat part (b) with three Gamma GLMs; use the `glm()` function for all three.

   (e) For each of the three GLMs involving the identity link, obtain a point estimate and normal-based 95% CI for the mean thermal assay among observations with an immunofluorescence assay value of 0.20. Briefly describe the most important reasons not to trust their validity. Sorry to be such a downer! We'll soon learn better methods.

32. A study was conducted to evaluate the association between high systolic blood pressure (SBP) and coronary heart disease (CHD). The results of the study are shown in the table below, stratified by age group (younger: <55; older: ≥55).

	Younger			Older		
	SBP < 165	SBP ≥ 165	Total	SBP < 165	SBP ≥ 165	Total
Did not develop CHD	280	40	320	140	20	160
Developed CHD	70	20	90	50	20	70
Total	350	60	410	190	40	230

(a) Poisson regression may be used to fit data in tabular form; the mean of a cell count is presumed related to indicators marked by that cell:

$$x_A = \begin{cases} 1 & \text{older} \\ 0 & \text{younger} \end{cases}, \quad x_S = \begin{cases} 1 & \text{SBP} \geq 165 \\ 0 & \text{SBP} < 165 \end{cases}, \text{ and } \quad x_C = \begin{cases} 1 & \text{Developed CHD} \\ 0 & \text{Did not develop CHD} \end{cases}.$$

Letting λ denote mean cell count, use the `glm()` function to fit the following models:

$$\begin{aligned}
\text{Model 1}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C \\
\text{Model 2}: &\quad \log(\lambda) = \beta_0 + \beta_S x_S \\
\text{Model 3}: &\quad \log(\lambda) = \beta_0 + \beta_A x_A \\
\text{Model 4}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S \\
\text{Model 5}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_A x_A \\
\text{Model 6}: &\quad \log(\lambda) = \beta_0 + \beta_S x_S + \beta_A x_A \\
\text{Model 7}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A \\
\text{Model 8}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A + \beta_{CS} x_C x_S \\
\text{Model 9}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A + \beta_{CA} x_C x_A \\
\text{Model 10}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A + \beta_{AS} x_A x_S \\
\text{Model 11}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A + \beta_{CS} x_C x_S + \beta_{CA} x_C x_A \\
\text{Model 12}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A + \beta_{CS} x_C x_S + \beta_{AS} x_A x_S \\
\text{Model 13}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A + \beta_{CA} x_C x_A + \beta_{AS} x_A x_S \\
\text{Model 14}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A + \beta_{CS} x_C x_S + \beta_{CA} x_C x_A + \beta_{AS} x_A x_S \\
\text{Model 15}: &\quad \log(\lambda) = \beta_0 + \beta_C x_C + \beta_S x_S + \beta_A x_A + \beta_{CS} x_C x_S + \beta_{CA} x_C x_A + \beta_{AS} x_A x_S + \beta_{CSA} x_C x_S x_A
\end{aligned}$$

Report the coefficient estimates in a table, with the eight coefficients spanning across the columns and the fifteen models spanning down the rows (some cells will be empty).

(b) The data can also be thought of as binomial counts, where the number in each of the four age/SBP groups who develop CHD is the binomial outcome. To that end, let p denote the probability of developing CHD; use the `glm()` function to fit the following models.

$$\begin{aligned}
\text{Model 1}: &\quad \text{logit}(p) = \beta_0 + \beta_S x_S \\
\text{Model 2}: &\quad \text{logit}(p) = \beta_0 + \beta_A x_A \\
\text{Model 3}: &\quad \text{logit}(p) = \beta_0 + \beta_S x_S + \beta_A x_A \\
\text{Model 4}: &\quad \text{logit}(p) = \beta_0 + \beta_S x_S + \beta_A x_A + \beta_{SA} x_S x_A
\end{aligned}$$

Report the coefficient estimates in a table having the analogous style as in part (a).

(c) Compare the estimates obtained in parts (a) and (b); comment on and explain the significance of the similarities between numbers in the tables (we will have an in-depth discussion about this problem—and analysis of tabular data—in class).

33. Suppose $\mathbf{x}_1, \dots, \mathbf{x}_N$ are randomly sampled vectors from a distribution that satisfies sensible regularity conditions, and that (conditional on $\mathbf{x}_i$) Y_i follows a Gamma distribution (where α is some unknown constant and β depends upon the linear predictor, $\eta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$).

    (a) Under the negative-inverse link, derive a sandwich-based variance estimator. Under what conditions on the mean model and mean-variance relationship should this variance estimator be valid?

    (b) Under the log link, derive a sandwich-based variance estimator that relies on correct specification of the mean model but allows misspecification of the mean-variance relationship.

    (c) Under the log link, derive a sandwich-based variance estimator that allows misspecification of both the mean model and the mean-variance relationship.

    (d) Under the identity link, derive a quasi-likelihood variance estimator that assumes that $\text{Var}[Y|\mathbf{x}] = \varphi(\mathbf{E}[Y|\mathbf{x}])^2$. Under what conditions on the mean model and mean-variance relationship should this variance estimator be valid?

    (e) Comment on the extent to which the assumptions necessary for the variance estimators in parts (a)-(d) to be correct depend upon $\mathbf{x}_1, \dots, \mathbf{x}_N$ having been randomly sampled. That is, how would the assumptions change if instead $\mathbf{x}_1, \dots, \mathbf{x}_N$ were fixed and known in advance?

34. Suppose $(X_1, Y_1), \dots, (X_N, Y_N)$ are independent samples. The outcome, Y, is distributed as follows (conditional on X):

$$Y \sim \text{Beta}\left(\alpha = \frac{1}{\gamma_0 + \gamma_1 X},\ \beta = \varphi - \frac{1}{\gamma_0 + \gamma_1 X}\right),$$

    for some suitable values of γ_0, γ_1, and φ (all unknown) such that $\alpha > 0$ and $\beta > 0$.

    (a) Argue that the mean model based on the inverse link, $\mathbf{E}[Y|X = x] = 1/(\beta_0 + \beta_1 x)$, is correctly specified; specifically identify the values of β_0 and β_1 in terms of γ_0, γ_1, and φ.

    (b) Argue that quasi-binomial and Poisson GLMs (both with inverse link) provide consistent estimators of of $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$, with the quasi-binomial GLM being the more efficient of the two asymptotically.

    (c) Consider a quasi-binomial GLM with an inverse link in particular. Derive a closed-form expression for an estimator of $\text{Cov}[\widehat{\boldsymbol{\beta}}]$ that relies on correct specification of the mean model and the mean-variance relationship.

    (d) Consider a Poisson GLM with an inverse link in particular. Derive a closed-form expression for an estimator of $\text{Cov}[\widehat{\boldsymbol{\beta}}]$ that allows misspecification of the mean model and the mean-variance relationship.

    (e) Conduct a simulation study that illustrates the validity of the estimators you derived in parts (c) and (d) under the described data generating mechanism. Please conduct $M = 5000$ simulation replicates, generate $X \sim \text{Uniform}(0, 1)$, and use $\gamma_0 = 0.5$, $\gamma_1 = 3$, $\varphi = 2.5$, and $N = 1000$. Further, you may use the `glm()` function to fit the models, but please hard-code any variance estimators.

35. A study was conducted to evaluate cryptorchidism (undescended testicle) as a risk factor for testicular cancer. Investigators were interested in evaluating whether the association was attributable to something systemic in men born cryptorchid, or to something in the localized environment of the undescended testicle. To address this question, they sought to evaluate whether cryptorchidism was more strongly associated with risk of ipsilateral ("same-side") testicular cancer as compared to contralateral ("opposite-side") testicular cancer. A total of $n_C = 2000$ controls, $n_L = 500$ left-sided cases, and $n_R = 500$ right-sided cases were sampled:

	Tumor laterality		
Cryptorchidism	Control	Case (left)	Case (right)
None	1960	450	440
Left	15	30	10
Right	15	10	40
Bilateral	10	10	10

The concept of "same" or "opposite" side has no meaning for controls, complicating implementation of usual case-control methods. To address this, let X_L and X_R be the indicators of cryptorchidism on the left and right sides, respectively, and consider the following multinomial model defined by constrained simultaneous equations:

$$\log\left(\frac{\text{P(Left-sided case}|X_L = x_L, X_R = x_R)}{\text{P(Control}|X_L = x_L, X_R = x_R)}\right) = \alpha_{0L} + \alpha_I x_L + \alpha_C x_R,$$

$$\log\left(\frac{\text{P(Right-sided case}|X_L = x_L, X_R = x_R)}{\text{P(Control}|X_L = x_L, X_R = x_R)}\right) = \alpha_{0R} + \alpha_C x_L + \alpha_I x_R.$$

Note, in particular, how α_I and α_C appear in both equations (color-coded for clarity).

(a) Use the fact that testicular cancer is rare in the population to express the quantities $\exp(\alpha_I)$ and $\exp(\alpha_C)$ as (approximate) risk ratios.

Let Y_L and Y_R denote indicators of left- and right-tumors (respectively); consider the model:

$$\begin{aligned}\log(\mathbf{E}[Z|X_L = x_L, X_R = x_R, Y_L = y_L, Y_R = y_R]) &= \beta_0 + \beta_1 x_L + \beta_2 x_R + \beta_3 x_L x_R + \beta_4 y_L + \beta_5 y_R \\ &\quad + \beta_6(x_L y_L + x_R y_R) + \beta_7(x_L y_R + x_R y_L),\end{aligned}$$

where Z is a corresponding count outcome (represented by the 4×3 table above).

(b) What do the quantities "$x_L x_R$," "$x_L y_L + x_R y_R$," and "$x_L y_R + x_R y_L$" represent?

(c) Argue that the multinomial model's structure is embedded within the log-linear model.

(d) Use `glm(..., family = poisson(link = "log"))` to fit the log-linear model; use this fit to report point estimates of $\exp(\alpha_I)$ and $\exp(\alpha_C)$.

(e) Employ a bootstrap procedure treating n_C, n_L, and n_R as fixed. Use this procedure to:

(i) Form quantile-based 95% confidence intervals for $\exp(\alpha_I)$ and $\exp(\alpha_C)$.

(ii) Form pivot-based 95% confidence intervals for $\exp(\alpha_I)$ and $\exp(\alpha_C)$.

(iii) Conduct a (Wald-based) test of the hypothesis $H_0 : \alpha_C = \alpha_I$ vs. $H_1 : \alpha_C \neq \alpha_I$.

Hard-code the bootstrap procedures, but use `glm()` within each iteration.

36. Load the data set `chemo.csv`, which comes from a study to evaluate doxorubicin as a chemotherapy agent at fixed concentrations. Subset the data set to concentrations, X, of $\geq$0.05 μmol/L, and consider a Poisson model (log link) of mean colony count, Y; treat X as a factor variable. For this problem, hard-code any test statistics, though you need not hard-code the regression models.

    (a) Showing your work, determine the form of a Wald-based test of the hypothesis that the mean colony count among plates given a concentration of 0.1 μmol/L is different from 139. Your answer will depend upon some covariance matrix for $\widehat{\boldsymbol{\beta}}$ that this part of the problem doesn't tell you how to estimate; do not try to simplify the test statistic.

    (b) Conduct the test of part (a) using a conditional bootstrap procedure.

    (c) Conduct the test of part (a) with an unconditional bootstrap.

    (d) Conduct the test of part (a) with a sandwich variance.

    (e) Conduct a score test of the hypothesis that the mean colony count differs between plates given a concentration of 0.5 and 1.0 μmol/L. Compare this to the result obtained from the score (Rao) test performed by the `anova()` function as a way to check your work.

37. This is a continuation of problem 36. Conduct the following hypothesis tests. Again, hard-code any test statistics, though you need not hard-code the regression models.

    (a) Conduct a likelihood ratio test of $H_0 : \mathbf{E}[Y|X = 1.0] = 0.5 \times \mathbf{E}[Y|X = 0.5]$. Compare these results to those obtained from the likelihood ratio test performed by the `anova()` function. Please show your work for this problem.

    (b) Conduct a likelihood ratio test of $H_0 : \mathbf{E}[Y|X = 1.0] = 0.5722 \times \mathbf{E}[Y|X = 0.5]$ (this problem also serves as a way for you to check your work in part (a)).

38. Revisit problem 31; focus your attention on the identity- and log-link Poisson-based models of part (c), but utilize the quasi-likelihood framework to accommodate overdispersion.

    (a) For each of the two models under consideration, compute the deviance residuals for each observation. Recall that if you use the `residuals()` function, you must scale the deviance residuals suitably.

        - Plot the deviance residuals (y-axis) against the immunofluorescence (x-axis).
        - Create a quantile-quantile plot of the deviance residuals.

        Very briefly comment on your findings.

    (b) For each of the two models under consideration:

        - Determine the influence of each observation (do this in a computationally efficient way, and use the one-step approach if applicable).
        - Generate a histogram of influences for each observation.
        - Generate a scatter plot with immunofluorescence on the x-axis and influence on the y-axis.

        Very briefly comment on your findings.

39. Your collaborators are studying output signals for a biomedical device. Given an input value, $X \geq 0$, the device produces a response, Y, having a sinusoidal relationship with X, but with random noise that tends to be of greater magnitude for larger inputs. Specifically, the mean model is given by $\mathbf{E}[Y|X = x] = \sin(\beta x)$, $\beta > 0$, which you may assume in this problem to be correctly specified. Your collaborators sought to calibrate the device to have response periodicity such that $\beta = 1.2 \pm 0.03$, a condition that would be straightforward to evaluate if there were no noise. Load the data set `signal.csv`, which includes data from an experiment of $n = 100$ independent runs at fixed, evenly spaced values of X. Your task as the team's biostatistician is to construct a point estimate and confidence interval for β to offer clarity on the device's calibration. You may assume in this problem that the input values are not subject to measurement error.

(a) Provide the form of a single Gauss-Newton step, $\beta^{(j+1)}$, for solving the nonlinear least squares estimating equation.

(b) Suggest an initializer, $\beta^{(0)}$. It does not need to be a consistent estimator—it can be a crude approximation, but it should be data driven in some way (even if visual).

(c) Provide a closed-form expression for an estimate of $\text{Var}[\widehat{\beta}]$. Briefly defend your choice of an estimator.

(d) Hard-code the Gauss-Newton algorithm for application to the data. Initialize based on your response in part (b); iterate until apparent convergence. Report your point estimate, $\widehat{\beta}$. You may confirm your answer with the `nls()` function. Provide a 95% Wald-based confidence interval for β based on your choice of a variance estimator in part (c).

(e) Given the investigators' desired level of calibration, how might you advise them?

(f) Suppose you seek to form a (point-wise) 95% confidence band for $\mathbf{E}[Y|x = x]$ over the range $0 < x < 10$. Describe why an approach based on the (first-order) delta method cannot accomplish this. Describe and implement a quantile-based nonparametric bootstrap approach to form a 95% confidence band for $\mathbf{E}[Y|X = x]$ over the range $0 < x < 10$.