Andrew J. Spieker, PhD
BIOS 7345 - Advanced Regression Analysis I
Collection of problems for Fall 2023 (Version: 11/09/2023)

Instructions: Responses are due by Box prior to the start of class (10:30a) on the due date. 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) Determine the singular value decomposition of $\mathbf{X}$.

   (c) Determine $\mathbf{X}^{1/2}$.

   (d) Compute the Moore-Penrose pseudoinverse, $\mathbf{X}^{-}$, and confirm that $\mathbf{X}^{-} = \mathbf{X}^{-1}$.

2. Let $\mathbf{X}$ be a 2×2 symmetric matrix (i.e., having the form shown below):

$$\mathbf{X} = \begin{bmatrix} X_1 & X_2 \\ X_2 & X_3 \end{bmatrix}.$$

   (a) State what the symmetry property implies about the eigenvalues and eigenvectors of $\mathbf{X}$.

   (b) For the remainder of the problem, assume that $\mathbf{X}$ is a projection matrix. Determine the constraints (in terms of X_1, X_2, and X_3) that are implied by the idempotence of $\mathbf{X}$. These equations have multiple solutions, but do not bother attempting to identify them.

   (c) Relying only on your general knowledge of projection matrices, state a specific numeric combination of X_1, X_2, and X_3 (not all zero) such that $\mathbf{X}$ is a projection matrix. Verify that your selection satisfies the constraints you determined in part (b).

   (d) State what the projection property implies about the determinant of $\mathbf{X}$.

   (e) State what the projection property implies about trace($\mathbf{X}$).

3. Suppose that $\mathbf{A}$ is a 3×2 matrix given by:

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

   (a) Showing your work, derive the singular value decomposition of $\mathbf{A}$.

   (b) Using the $\mathbf{U}_{\text{mini}}$ and $\mathbf{D}_{\text{mini}}$ method, derive $\mathbf{A}^{-}$, the Moore-Penrose pseudoinverse of $\mathbf{A}$.

   (c) Show that $\mathbf{A}^{-}$ also happens to be a left-inverse of $\mathbf{A}$ (i.e., $\mathbf{A}^{-}\mathbf{A} = \mathbf{I}$).

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

$$\begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 1 \end{bmatrix}$$

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

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

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

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

   (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.

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 $\mathbf{X} = \mathbf{UDV}^T$ denote its SVD.

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

   (b) Write an expression for $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ of the form $\mathbf{UAU}^T$, 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 $\mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ of the form $\mathbf{UBU}^T$, 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 $\mathbf{G}$ denote a generalized inverse of $\mathbf{X}^T\mathbf{X}$. Prove the following facts. You will find the result of Lemma 1.2 extremely useful in proving some of these.

   (a) $\mathbf{G}^T$ is also a generalized inverse of $\mathbf{X}^T\mathbf{X}$.

   (b) $\mathbf{GX}^T\mathbf{XG}^T$ is a symmetric reflexive generalized inverse of $\mathbf{X}^T\mathbf{X}$.

   (c) $\mathbf{GX}^T$ is a generalized inverse of $\mathbf{X}$ and $\mathbf{XG}$ is a generalized inverse of $\mathbf{X}^T$.

   (d) If $\widetilde{\mathbf{G}}$ is also a generalized inverse of $\mathbf{X}^T\mathbf{X}$, then $\mathbf{XGX}^T = \mathbf{X}\widetilde{\mathbf{G}}\mathbf{X}^T$.

   (e) $\mathbf{XGX}^T$ is symmetric.

7. Let $\mathbf{c}$ denote a length-N vector of constants and $\mathbf{x}$ a random length-N vector.

   (a) Show that:

$$\mathbf{E}[(\mathbf{x} - \mathbf{c})(\mathbf{x} - \mathbf{c})^T] = \text{Cov}[\mathbf{x}] + (\mathbf{E}[\mathbf{x}] - \mathbf{c})(\mathbf{E}[\mathbf{x}] - \mathbf{c})^T.$$

   (b) Let $\mathbf{\Sigma} = \text{Var}[\mathbf{x}]$, with σ_{ij} denoting the (i, j) entry of $\mathbf{\Sigma}$. Show that:

$$\mathbf{E}[||\mathbf{x} - \mathbf{c}||^2] = \sum_i \sigma_{ii} + ||\mathbf{E}[\mathbf{x}] - \mathbf{c}||^2.$$

8. Suppose $\mathbf{x}$ is a random vector having covariance matrix $\boldsymbol{\Sigma}$, and let $\mathbf{a}$ and $\mathbf{b}$ denote constant vectors (having the same length as $\mathbf{x}$). Show that $\text{Cov}[\mathbf{a}^T\mathbf{x}, \mathbf{b}^T\mathbf{x}] = \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{b}$. For this problem, do not merely rely on the more general formula for $\text{Cov}[\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{x}]$ when $\mathbf{A}$ and $\mathbf{B}$ are constant matrices, but show this by brute force.

9. Suppose $\mathbf{x}$ is a multivariate normal random length-N vector having mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and let $\mathbf{A} \succcurlyeq 0$ denote a constant $N \times N$ matrix. Derive a formula for $\text{Cov}[\mathbf{x}, \mathbf{x}^T\mathbf{A}\mathbf{x}]$ that depends only on $\boldsymbol{\Sigma}$, $\mathbf{A}$, and $\boldsymbol{\mu}$.

10. Suppose that $X_1, \ldots, X_N$ are real-valued random variables satisfying $X_{i+1} = \rho X_i$ for a known constant ρ and $\text{Var}[X_1] = \sigma^2$. Determine the variance-covariance matrix of $(X_1, \ldots, X_N)$.

11. Let X_1, X_2, and X_3 are i.i.d. standard normal random variables.

$$\begin{aligned} Y_1 &= \frac{1}{\sqrt{3}}(X_1 + X_2 + X_3), \\ Y_2 &= \frac{1}{\sqrt{2}}(X_1 - X_2), \\ Y_3 &= \frac{1}{\sqrt{6}}(X_1 + X_2 - 2X_3). \end{aligned}$$

    Show that Y_1, Y_2, and Y_3 are i.i.d. distributed standard normal random variables.

12. 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.

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

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

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

    (b) Determine the MGF of Z. You may use software as an aid along the way—though I recommend starting with the definition of the MGF.

    (c) Show that $Z =_d 2U_1 - U_2 - U_3$, where U_1, U_2, and U_3 are independent χ^2_1 random variables.

15. Suppose $X \sim \chi^2_N(\lambda)$. Use the MGF for a non-central chi-squared distribution to show that $\mathbf{E}[X] = N + 2\lambda$ and $\text{Var}[X] = 2N + 8\lambda$. *Hint*: For this problem, I recommend making use of the following two formulas based on the cumulant-generating function:

$$\mathbf{E}[X] = \frac{d}{dt}\log(M_X(t))\bigg|_{t=0} \text{ and } \text{Var}[X] = \frac{d^2}{dt^2}\log(M_X(t))\bigg|_{t=0}.$$

16. 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, let $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Derive expressions for $\mathbf{E}[\widehat{\boldsymbol{\epsilon}}]$, $\text{Cov}[\widehat{\boldsymbol{\epsilon}}]$, and $\text{Cov}[\widehat{\boldsymbol{\epsilon}}, \mathbf{Py}]$.

17. 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 having mean zero and variance one.

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

    (b) Argue that the sample mean of Y among observations for which $X = 1/\sqrt{2}$ is unbiased for $\mathbf{E}[Y|X = 1/\sqrt{2}]$ but is *not* the BLUE if you have knowledge of the absence of an intercept. Please state the BLUE for $\mathbf{E}[Y|X = 1/\sqrt{2}]$ as part of your response.

    (c) Letting $\beta = 1$, illustrate via simulation that the value of $\text{Var}[\sqrt{n}(\widehat{\beta}_1 - 1)]$ does not depend upon normality of errors. Specifically, consider the following specifications for the distribution of ϵ in the case where $n = 10$, $n = 100$, and $n = 1000$: (1) $\epsilon_i \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)$, (2) $\epsilon_i \sim \text{Laplace}(\mu = 0, \sigma^2 = 1)$, and (3) $\epsilon_i \sim \text{Gamma}(\alpha = 2, \beta = \sqrt{2}) - 2/\sqrt{2}$.

    (d) Illustrate that $(\mathbf{x}^T\mathbf{x})^{1/2}(\widehat{\beta}_1 - 1) \longrightarrow_d \mathcal{N}(0, 1)$ by simulation for each case. You may find quantile-quantile plots helpful.

    (e) Should any of your conclusions from parts (c) and (d) change if the values of X are randomly sampled (i.e., taking on the values $X = -0.5$ and $X = 0.5$ with equal probability)? Verify your answer by re-running the simulations of (c) and (d) under this setup.

18. Consider a linear regression model based on N observations that includes an intercept (assume $\mathbf{X}$ is full-rank). Prove that $\sum_{i=1}^{N}(y_i - \widehat{y}_i) = 0$ (a two-line proof if you use linear algebra).

19. Consider the linear regression model $\mathbf{E}[Y|X = x] = \beta_0 + \beta_1 x + \beta_2(3x^2 - 2)$ for $i = 1, 2, 3$, with $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$. Determine (in terms of y_1, y_2, and y_3) the least squares estimates of β_0, β_1, and β_2 and show that the least squares estimates of β_0 and β_1 remain unchanged if the "$\beta_2(3x^2 - 2)$" term is dropped from the model.

20. Consider an ANOVA-style model involving a 2×2 factorial design of two binary treatment groups that are presumed not to interact. Assume each group has a total of N observations. Specifically, for $i = 1, \ldots, N$, $j = 0, 1$, and $k = 0, 1$:

$$Y_{ijk} \quad = \quad \mu + \alpha_j + \gamma_k + \epsilon_{ijk},$$

    where $\mathbf{E}[\epsilon_{ijk}] = 0$ and $\text{Var}[\epsilon_{ijk}] = \sigma^2$. For this problem, please show your work as appropriate—though you may use software for the parts of this problem that would be unwieldy by hand (e.g., computing the g-inverse of any matrix that is larger than a 2×2).

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

    (b) Is μ estimable? Justify your answer.

    Hint for (c) onward: Let $N = 1$ and simply generalize the answers on the back end.

    (c) Use the g-inverse method to find a solution to the normal equations.

    (d) Use the re-parameterization method to find a solution to the normal equations.

(e) Use the method of imposing identifiability constraints (also called side conditions) to find a solution to the normal equations.

(f) Verify that the same value of $\widehat{\mathbf{y}}$ is produced for the three methods you implemented in parts (c), (d), and (e).

(g) For this model, how many linearly independent estimable functions should there be?

(h) Is $\mu + \alpha_0$ estimable? Justify your answer.

(i) Is $\mu + \alpha_1 + \gamma_0$ estimable? Justify your answer.

(j) Determine the unique BLUE for the difference in mean Y that compares those receiving $k = 1$ and $j = 0$ to those receiving $j = 1$ and $k = 0$. Justify your answer by citing any theorems you invoke and determining whether the conditions of those theorems are met.

21. In this problem, you will illustrate the Gauss-Markov theorem by simulation. Let $N = 100$, $X_i = 1 + i/50$, and $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, where $\epsilon_i \sim \mathcal{N}(0, \sigma^2 = g(X_i))$, and let $\beta_0 = 0$ and $\beta_1 = 1$. Consider three ways of generating data: (1) $g(X) = 1$, (2) $g(X) = X$, and (3) $g(X) = 1/X$. For each case, consider estimating β_1 by ordinary least squares. Further, consider three ways to estimate β_1: (1) ordinary least squares, (2) weighted least squares with weights given by $w(X) = X^{-1}$, and (3) weighted least squares with weights given by $w(X) = X$. Based on 10,000 simulation replicates, fill in the rightmost two columns of the table below. Comment on and account for your findings.

True variance	Weights	$\mathbf{E}[\widehat{\beta}_1]$	$\text{SD}[\widehat{\beta}_1]$
$g(X) = 1$	$w(X) = 1$	–	–
$g(X) = 1$	$w(X) = 1/X$	–	–
$g(X) = 1$	$w(X) = X$	–	–
$g(X) = X$	$w(X) = 1$	–	–
$g(X) = X$	$w(X) = 1/X$	–	–
$g(X) = X$	$w(X) = X$	–	–
$g(X) = 1/X$	$w(X) = 1$	–	–
$g(X) = 1/X$	$w(X) = 1/X$	–	–
$g(X) = 1/X$	$w(X) = X$	–	–

Note: $\mathbf{E}[\widehat{\beta}_1]$ refers to the mean estimate and $\text{SD}[\widehat{\beta}_1]$ refers to the empirical standard error (the standard deviation of estimates) across simulations.

22. 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 five ways:

(a) Ordinary least squares.

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

(c) A weighted least squares estimate based on fixed weights given by $W = X$ (comment on why this is clearly *not* a good idea from an efficiency standpoint).

(d) 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).

(e) Iteratively re-weighted least squares with variance model $\text{Var}[Y|X = x] = \alpha_0 + \alpha_1 X$ (you should hard-code this one).

23. Suppose $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon$, where $\mathbf{X}$ is a fixed $N \times K$ design matrix of full rank, and $\text{Cov}[\boldsymbol{\epsilon}] = \sigma^2\mathbf{V}$ for a diagonal matrix $\mathbf{V}$ 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) $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}$ 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) $\widehat{\boldsymbol{\beta}}^*$ is unbiased.

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

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

24. 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 $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. Justify how you know this is a testable hypothesis.

    (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.

25. Consider a 2×2 factorial design in which a total of $4n$ patients are randomly by evenly allocated into each of four groups, defined by combinations of $X = 0, 1$ and $Z = 0, 1$. Specifically, for $i = 1, \ldots, 4n$, suppose $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + \epsilon_i$, with ϵ_i having mean zero and common variance σ^2 (homoscedasticity). The design matrix is of full rank in this case.

    (a) Consider the hypothesis in which you seek to test whether receipt of both X and Z together is different from receipt of X alone (in terms of effects on mean Y). Write down H_0 in the form $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$. Justify how you know this is a testable hypothesis.

    (b) Determine an expression for the F-statistic that could be used to test H_0. Please express your answer in terms of the elements of $\widehat{\boldsymbol{\beta}}$, n, and $\widehat{\sigma}^2$ (i.e., from the unrestricted model).

    (c) Assuming that H_0 is true, what is the distribution of the F-statistic you determined in part (b) if $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$?

    (d) Suppose that $\epsilon_i \sim \text{Gamma}(\alpha, \beta) - \alpha/\beta$. Illustrate by simulation that a 0.05-level F-test still has the correct type 1 error rate asymptotically even for a selection of α that corresponds to a distribution having far greater skewness and kurtosis than a normal distribution (β does not affect the skewness or the kurtosis; I recommend choosing $\alpha = 0.1$,). You should illustrate this over a reasonable range of n.

26. Consider an ANOVA-style model involving a 2×2 factorial design of two binary treatment groups that are presumed not to interact. Assume each group has a total of N observations. Specifically, for $i = 1, \ldots, N$, $j = 0, 1$, and $k = 0, 1$: $Y_{ijk} = \mu + \alpha_j + \gamma_k + \epsilon_{ijk}$, with $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$. Note that this is the model that was considered in Problem 20.

    (a) Argue that $H_0 : \gamma_0 = \gamma_1$ is a testable hypothesis; determine a corresponding F-statistic.

    (b) Argue that $H_0 : \mu + \alpha_0 + \gamma_0 = 0$ and $H_0 : \gamma_1 - \gamma_0 = \alpha_1 - \alpha_0 = 0$ are testable hypotheses. Provide plain-language interpretations of these hypotheses.

    (c) Argue that $H_0 : \alpha_1 = \gamma_0$ is not a testable hypothesis.

27. If $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is a testable hypothesis and $\text{rank}(\mathbf{C}) = Q$, then the least squares estimator under the restricted model takes the following form:

$$\widehat{\boldsymbol{\beta}}_H \;=\; \widehat{\boldsymbol{\beta}} - (\mathbf{X}^T\mathbf{X})^{-}\mathbf{C}^T(\mathbf{C}(\mathbf{X}^T\mathbf{X})^{-}\mathbf{C}^T)^{-1}\mathbf{C}\widehat{\boldsymbol{\beta}},$$

    where $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-}\mathbf{X}^T\mathbf{y}$ denotes a solution to the normal equations under the unrestricted model. Prove this directly using Lagrange multipliers to solve the optimization problem:

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \; \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \text{ subject to the constraint } \mathbf{C}\boldsymbol{\beta} = \mathbf{0}.$$

28. Prove the (one-way) ANOVA decomposition:

$$\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ji} - \overline{y}_{\cdot\cdot})^2 \;=\; \sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ji} - \overline{y}_{j\cdot})^2 + \sum_{j=1}^{J} n_j(\overline{y}_{j\cdot} - \overline{y}_{\cdot\cdot})^2.$$

29. A series of $N + 1$ observations, call them $Y_1, \ldots, Y_{N+1}$ are known to be uncorrelated with a shared variance σ^2. Following the collection of the first N observations, it is suspected that there is a sudden change in the mean of the distribution. Determine a test statistic for the hypothesis $H_0 : \mu_N = \mu_{N+1}$, where μ_N denotes the mean of the first N observations and μ_{N+1} denotes the mean of the final observation.

30. Consider a balanced two-way ANOVA setting. Obtain an F-statistic for the test $H_0 : \mu_{jk} = \mu$ vs. $H_1 : \mu_{jk} = \mu + \alpha_{jk}$ for all $j = 1, \ldots, J$ and $k = 1, \ldots, K$. What is the distribution of the F-statistic if the errors are normally distributed and homoscedastic? Please use the ANOVA decomposition rather than brute-forcing matrix algebra.

31. This is a continuation of Problem 21, in which we will now *estimate* the standard errors using the weighted least squares variance formula. For simplicity, remove the case in which $g(X) = 1/X$ from consideration (it does not satisfy all the necessary regularity conditions). Increase the number of simulation replicates to 20,000. Let $\widehat{\text{SE}}[\widehat{\beta}_1]$ denote the average estimated standard error across simulations (i.e., based on the correct weight matrix). Conduct the simulation again, filling in the additional column to the right. Comment on and account for your findings.

True variance	Weights	$\mathbf{E}[\widehat{\beta}_1]$	$\text{SD}[\widehat{\beta}_1]$	$\widehat{\text{SE}}[\widehat{\beta}_1]$
$g(X) = 1$	$w(X) = 1$	–	–	–
$g(X) = 1$	$w(X) = 1/X$	–	–	–
$g(X) = 1$	$w(X) = X$	–	–	–
$g(X) = X$	$w(X) = 1$	–	–	–
$g(X) = X$	$w(X) = 1/X$	–	–	–
$g(X) = X$	$w(X) = X$	–	–	–

32. 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 $\mathbf{E}[\mathbf{y}|\mathbf{X}_1, \mathbf{X}_2] = \mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2$.

(a) First, verify the formula for the inverse of a 2×2 block matrix. Specifically:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix}$$

(b) Use your answer to part (a) to show that $\text{Cov}[\widehat{\boldsymbol{\beta}}]$ can be written in the following form:

$$\text{Cov}[\widehat{\boldsymbol{\beta}}] = \sigma^2 \begin{bmatrix} (\mathbf{X}_1^T\mathbf{X}_1)^{-1} + \mathbf{F}\mathbf{G}^{-1}\mathbf{F}^T & -\mathbf{F}\mathbf{G}^{-1} \\ -\mathbf{G}^{-1}\mathbf{F}^T & \mathbf{G}^{-1} \end{bmatrix},$$

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) Show that $\mathbf{F}\mathbf{G}^{-1}\mathbf{F}^T$ is positive definite unless $\mathbf{X}_1^T\mathbf{X}_2 = \mathbf{0}$.

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

33. Consider a simple linear regression model $\mathbf{E}[Y|X = x] = \beta x$ in which X and Y each have mean zero so that there is no intercept; further, you may assume X to be scaled with $\sum_i x_i^2 = 1$. for convenience. Prove that the graph of the upper confidence interval for $\mathbf{E}[Y|X = x]$ is hyperbolic as a function of x. You may need to refresh your memory on the analytic geometry of conic sections and second-degree equations.

34. Consider the projection matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$, 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!

(a) Show that $\sum_{i=1}^N \mathbf{P}_{ii} = K$.

(b) Show that $\sum_{i=1}^N \mathbf{P}_{ij} = \sum_{j=1}^N \mathbf{P}_{ij} = 1$.

(c) Show that $\mathbf{P}_{ii}^2 + \sum_{j \neq i} \mathbf{P}_{ij}^2 = \mathbf{P}_{ii}$.

(d) Show that if $\mathbf{P}_{ii} = 1$, then the OLS fit passes through the point $(\mathbf{x}_i, Y_i)$.

(e) Show that $(1 - \mathbf{P}_{ii})^2 + \sum_{j \neq i} \mathbf{P}_{ij}^2 = 1 - \mathbf{P}_{ii}$.

(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.

35. Let $\widehat{\boldsymbol{\beta}}$ denote the OLS estimate of $\boldsymbol{\beta}$ based on the fixed, full-rank $N \times K$ design matrix, $\mathbf{X}$, and outcome vector $\mathbf{y}$. Derive a formula for the influence of the i^{th} observation. Specifically, let $\widehat{\boldsymbol{\beta}}_{-i}$ denote the least squares estimate without the i^{th} observation included; show that:

$$\widehat{\boldsymbol{\beta}}_{-i} = \widehat{\boldsymbol{\beta}} - \frac{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i\widehat{\epsilon}_i}{1 - \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i}.$$

Here is a hint for this problem. Let $\mathbf{X}_{-i}$ denote the $(N-1) \times K$ design matrix with the i^{th} observation excluded. It happens that $\mathbf{X}^T\mathbf{X} - \mathbf{x}_i\mathbf{x}_i^T = \mathbf{X}_{-i}^T\mathbf{X}_{-i}$. The Sherman-Morrison-Woodbury formula is given by $(\mathbf{A} + \mathbf{B}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}$, which you can use with the neat fact described previously to obtain a nice expression for $(\mathbf{X}_{-i}^T\mathbf{X}_{-i})^{-1}$.

36. Consider a total of N 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. We've noted that degrees of freedom can be conceptualized in the following special way:

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

where $\widehat{Y}$ denotes the fitted value. Let $\widehat{\boldsymbol{\beta}}_\lambda$ denote a penalized least squares estimate based on a ridge penalty with tuning parameter λ.

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

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

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

37. 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 combinations of eigenvalues of $\mathbf{X}^T\mathbf{X}$.

(b) Characterize the set $\Lambda^* \subseteq \Lambda$, of all possible combinations of eigenvalues of $\mathbf{X}^T\mathbf{X}$ such that $\mathbf{X}^T\mathbf{X}$ non-singular.

(c) Characterize all values of $\lambda \in \mathbb{R}$ such that $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}$ is non-singular.

(d) Characterize all values of $\lambda \in \mathbb{R}$ such that $\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I}$ is positive definite.

(e) Argue that for $\lambda \gg n$, the coefficient path for a single coefficient—namely, $\{(\lambda, \widehat{\beta}_{\lambda;j})\}$, resembles the graph of the hyperbolic function $\widehat{\beta}_{\lambda;j} = k/\lambda$.

38. This problem continues from the setup of Problem 37. 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

Please use statistical software as an aid in the computationally cumbersome parts of the problems that follow, though you should justify your steps.

(a) Let $\widehat{\boldsymbol{\beta}}_\lambda$ denote a solution (if any) to the penalized normal equations, $\mathbf{X}^T\mathbf{y} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})\boldsymbol{\beta}$, for a given λ (following centering/scaling). Argue that $\widehat{\boldsymbol{\beta}}_{\lambda=0}$ exists but is not unique.

(b) It has been shown that a *negative* ridge penalty is capable of producing a solution that minimizes expected prediction error; however, your answer to Problem 37 may prompt concerns. Graph each component of $\widehat{\boldsymbol{\beta}}_\lambda$ and $\widehat{\mathbf{y}}_\lambda$ as a function of λ over `lambda=seq(-2*pi, 0, 0.2)`. Comment on and account for your findings.

39. 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 $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$.

    (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 dispersion parameter, ϕ.
    - $\mathbf{E}[Y]$, 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.
    - $\mathbf{V}$, the matrix that marks the mean-variance relationship.
    - The score equations to solve for $\boldsymbol{\beta}$.
    - The formula for one iteration of a Newton-Raphson step.
    - The formula for one iteration of a Gauss-Newton step.
    - A (likelihood-based) estimator of $\text{Var}[\widehat{\boldsymbol{\beta}}]$.

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

40. Repeat Problem 39, this time with Y_i following a $\text{Gamma}(\alpha, \beta)$ distribution that depends upon the value of $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ (*Hint*: though it seems weird, let $\theta = -\beta/\alpha$ and $\phi = 1/\alpha$). 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.

41. 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) Fit the following GLMs (likelihood); overlay the three fitted curves on the scatter plot.

    - $Y \sim \mathcal{N}(\mu = \beta_0 + \beta_1 X, \sigma^2)$; hard-code this.
    - $Y \sim \mathcal{N}(\mu = \exp(\beta_0 + \beta_1 X), \sigma^2)$; hard-code this.
    - $Y \sim \mathcal{N}(\mu = 1/(\beta_0 + \beta_1 X), \sigma^2)$; hard-code this.

    (c) Repeat part (b) with three Poisson GLMs, but use the `glm()` function.

    (d) Repeat part (b) with three Gamma GLMs, but use the `glm()` function.

    (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.

42. 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:

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

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.

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

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.

43. Suppose $\mathbf{x}_1, \ldots, \mathbf{x}_N$ are randomly sampled vectors from a distribution that satisfies sensible regularity conditions, and that Y_i follows a Gamma distribution that depends upon the value of $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ (you should use the (α, β) parameterization of the Gamma distribution).

    (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 $\mathbf{x}_1, \ldots, \mathbf{x}_N$ having been randomly sampled. That is, how would the assumptions change if instead $\mathbf{x}_1, \ldots, \mathbf{x}_N$ were fixed and known in advance?

44. Revisit problem 41, focusing your attention on the Poisson model with the identity link. Obtain (hard-coded) estimates of $\text{SE}[\widehat{\beta}_1]$ based on the following approaches.

    (a) A sandwich estimator that relies on a correct mean model but allows a misspecified mean-variance relationship (show work, and then compare to `sandwich()`).

    (b) A quasi-Poisson approach that assumes $\text{Var}[Y|X = x] = \varphi\mathbf{E}[Y|X = x]$ (show work).

    (c) A bootstrap estimator that treats X as random (reflects the study).

45. Revisit problem 41, and consider the mean model $\log(\mathbf{E}[Y|X = x]) = \beta_0 + \beta_1 x$. Construct point estimates and 95% CIs for $\mathbf{E}[Y|X = 0.8]$ using the methods of (a)-(f). You need not hard-code the models or the sandwich variance of part (d), but do hard-code the bootstrap procedures of (e)-(f).

    (a) A Poisson-based approach assuming the mean and variance models are correct.

    (b) A quasi-Poisson approach that assumes $\text{Var}[Y|X = x] = \varphi\mathbf{E}[Y|X = x]$.

    (c) Likelihood-based negative binomial regression.

    (d) A Poisson-based approach with a sandwich variance.

    (e) A quantile bootstrap based on the Poisson model.

    (f) A pivot-based bootstrap based on the Poisson model.

46. 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 ≥ 0.05 μmol/L, and consider a Poisson model (log link) of mean colony count, Y; treat X as a factor variable. Conduct the following hypothesis tests. Hard-code any test statistics, though you need not hard-code the regression models.

(a) 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; use a conditional bootstrap procedure. Is this a valid approach?

(b) Repeat part (a) with an unconditional bootstrap. Is this a valid approach?

(c) Repeat part (a) with a sandwich variance. Is this a valid approach?

(d) 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 double-check your work.

(e) 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.

(f) 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 helps you check your work in part (e)).

47. Revisit (and re-do) problem 46, parts (a)-(c), but this time employ a Poisson model, and treat the (log) mean count as a linear function of doxorubicin concentration (that is, the mean model is given by $\log \mathbf{E}[Y|X = x] = \beta_0 + \beta_1 x$ and the mean-variance relationship model is given by $V(\mu) = \mu$).