Homework 9

Due Thursday, 4/22/21 on Canvas.

1. This problem is an extension of Problem 4 from HW8.

   (a) Use REML + GLS to fit the model you proposed in part (c)(i) of Problem 4. Test the null hypothesis that there is no patient × treatment interaction. Make sure to report your null and alternative models.

   (b) Use the model you fitted above to estimate the parameters in the model you proposed part (c)(ii) of Problem 4. Which drug would you recommend an insomniac take if they slept an average of 2 hours a night without treatment? How about one that slept an average of 6 hours a night?

2. (REML is maximum quasi-likelihood on the residuals of $\boldsymbol{Y}$) Suppose $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ is full rank, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\mathbb{E}(\boldsymbol{\epsilon}) = \boldsymbol{0}_n$ and $\text{Var}(\boldsymbol{Y}) = \boldsymbol{V}(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \mathbb{R}^b$ and some function $\boldsymbol{V} : \mathbb{R}^b \to \mathbb{R}^{n\times n}$. Throughout this problem, assume that $\boldsymbol{V}(\boldsymbol{\theta})$ is positive definite and let $\boldsymbol{A}_X \in \mathbb{R}^{n\times(n-p)}$ be any matrix whose columns form a basis for $\ker\left(\boldsymbol{X}^T\right)$. Recall that $\boldsymbol{A}_X$ is not unique.

   E(Y~) = E(Y) - E(...e) - E(XB) = E(XB) - E(XB) = 0

   (a) Let $\boldsymbol{Q} = I_n - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T$. Show that $\boldsymbol{A}_X\left(\boldsymbol{A}_X^T\boldsymbol{A}_X\right)^{-1}\boldsymbol{A}_X^T = \boldsymbol{Q}$. multiply both sides with A

   Var(Y~) = Var(QY) = QVar(Y)Qt = QV(theta)Q (bc Q is symmetric)

   (b) Let $\tilde{\boldsymbol{Y}} = \boldsymbol{Q}\boldsymbol{Y} \in \mathbb{R}^n$ be the residuals after you regress out $\boldsymbol{X}$ from $\boldsymbol{Y}$. Find expressions for $\mathbb{E}\left(\tilde{\boldsymbol{Y}}\right)$ and $\text{Var}\left(\tilde{\boldsymbol{Y}}\right)$. Show that $\tilde{\boldsymbol{Y}}$ is a degenerate distribution (i.e. the support of $\tilde{\boldsymbol{Y}}$ is a strict subspace of $\mathbb{R}^n$).

   (c) Given that $\tilde{\boldsymbol{Y}}$ is degenerate, the usual normal likelihood is not an appropriate quasi-likelihood for $\tilde{\boldsymbol{Y}}$. To circumvent this, define

   $$\tilde{\ell}(\boldsymbol{\theta}) = -\frac{1}{2}\log\left[\det{}_+\left\{\boldsymbol{Q}\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{Q}\right\}\right] - \frac{1}{2}\tilde{\boldsymbol{Y}}^T\left\{\boldsymbol{Q}\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{Q}\right\}^\dagger\tilde{\boldsymbol{Y}}$$

   to be the likelihood of the degenerate normal distribution, where $\boldsymbol{B}^\dagger$ is the usual Moore-Penrose pseudoinverse of $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ and $\det_+(\boldsymbol{B})$, called the **pseudo determinant** of $\boldsymbol{B}$, is defined to be

   $$\det{}_+(\boldsymbol{B}) = \lim_{\epsilon\to 0}\frac{\det(\boldsymbol{B} + \epsilon I_n)}{\epsilon^{n-\text{rank}(\boldsymbol{B})}}.$$

   (i) Show that $\left\{\boldsymbol{Q}\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{Q}\right\}^\dagger = \boldsymbol{A}_X\left\{\boldsymbol{A}_X^T\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{A}_X\right\}^{-1}\boldsymbol{A}_X^T$.

   (ii) If $\boldsymbol{B} \in \mathbb{R}^n$ is symmetric positive semi-definite, show that $\det_+(\boldsymbol{B})$ is the product of the non-zero eigenvalues of $\boldsymbol{B}$. Use this to show that $\det_+\left\{\boldsymbol{Q}\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{Q}\right\} = \det\left(\boldsymbol{A}_X^T\boldsymbol{A}_X\right)^{-1}\det\left\{\boldsymbol{A}_X^T\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{A}_X\right\}$.

   Remark. *The function $\tilde{\ell}$ is the log-likelihood of $\tilde{\boldsymbol{Y}}$ if we assume $\tilde{\boldsymbol{Y}} \sim N(\boldsymbol{0}, \boldsymbol{Q}\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{Q})$. Note that implicit in $\tilde{\ell}$ is the requirement that $\tilde{\boldsymbol{Y}} \in ker(\boldsymbol{X}^T)$. The likelihood, by definition, is 0 for $\tilde{\boldsymbol{Y}} \notin ker(\boldsymbol{X}^T)$. For the probability purists, $\tilde{\ell}$ is the log of the density of $\tilde{\boldsymbol{Y}}$ (ignoring constants that do not depend on $\boldsymbol{\theta}$) with respect to the Lebesgue measure on $ker(\boldsymbol{X}^T)$.*

(d) Define the usual REML objective function to be

$$\ell_{REML}(\boldsymbol{\theta}) = -\frac{1}{2}\log\left[\det\left\{\boldsymbol{A}_X^T\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{A}_X\right\}\right] - \frac{1}{2}\left(\boldsymbol{A}_X^T\boldsymbol{Y}\right)^T\left\{\boldsymbol{A}_X^T\boldsymbol{V}(\boldsymbol{\theta})\boldsymbol{A}_X\right\}^{-1}\left(\boldsymbol{A}_X^T\boldsymbol{Y}\right).$$

Show that $\ell_{REML}(\boldsymbol{\theta}) = \tilde{\ell}(\boldsymbol{\theta}) + C$, where C is a constant that does not depend on $\boldsymbol{\theta}$. Use this to show that $\hat{\boldsymbol{\theta}}_{REML} = \arg\max_{\boldsymbol{\theta}\in\Theta}\ell_{REML}(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}\in\Theta}\tilde{\ell}(\boldsymbol{\theta})$. Why does this imply that $\hat{\boldsymbol{\theta}}_{REML}$ is invariant to the choice of $\boldsymbol{A}_X$?

(e) Lastly, suppose $\boldsymbol{\theta} = \sigma^2 > 0$ and that $\boldsymbol{V}(\boldsymbol{\theta}) = \sigma^2 I_n$ (i.e. the usual assumptions on the variance in standard linear regression). Using part (d) (i.e. that the REML estimator for σ^2 can be computed by maximizing $\tilde{\ell}$), show that $\hat{\sigma}^2_{REML} = MSE$, where $MSE = (n-p)^{-1}\tilde{\boldsymbol{Y}}^T\tilde{\boldsymbol{Y}}$ is the usual mean squared error.

3. The data Age1.txt and Age3.txt contain measurements of the concentration of bilirubin, a ubiquitous small molecule metabolite, in the blood plasma of children at age 1 and 3, respectively. The goal is to understand relationship between the concentration of bilirubin and recurrent wheeze, the latter of which is a diagnosis of ≥ 4 wheezing episodes in that year. The complete list of covariates is given below.

- `IndividualID`: A unique ID given to each child. There should be 33 children whose bilirubin concentration was measured at both age 1 and age 3. You may assume that data collected on different individuals are independent.
- `Bilirubin`: The log-concentration of bilirubin. Treat this as the dependent variable.
- `Wheeze`: A factor variable with 3 levels. level 0: no wheezing episodes in that year; level 1: 1-3 wheezing episodes in that year; level 2: ≥4 wheezing episodes in that year.
- `Diet`: A factor variable with 2 levels. level 0: exclusively breastfed for the first six months of life; level 1: not exclusively breastfed for the first six months of life.
- `Daycare`: A factor variable with 2 levels. level 0: did not attend daycare in the first year of life; level 1: attended daycare in the first year of life.
- `Sex`: A factor variable with 2 levels. level 0: male; level 1: female.

HW7, 3b

(a) By performing two separate linear regressions (one at age 1 and one at age 3) and treating all of the above-mentioned covariates in your model as additive fixed effects, estimate the expected difference in the log-concentration of bilirubin between recurrent wheezers (those with ≥ 4 wheezing episodes in that year) and healthy controls (those who did not wheeze in that year) at ages 1 and 3. Report 95% confidence interval for both expected differences.

(b) The inference you performed in part (a) presumably relied on approximating test statistics with a normal or t-distribution. Do you trust this approximation? Give an argument as to why you do or do not. Include plots if necessary.

(c) Let β_j be the expected difference in the log-concentration of bilirubin between recurrent wheezers and healthy controls at age j, and let $\hat{\beta}_j$ be its estimate you determined in part (a). Your collaborator has reason to believe that $\beta_1 = \beta_3 = \beta$. To estimate β, suppose you decide to meta-analyze the results at ages 1 and 3 by modelling $\hat{\beta}_1, \hat{\beta}_3$ as

Z test to compare?

$$\begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_3\end{pmatrix} \sim N\left(\mathbf{1}_2\beta, \operatorname{diag}\left(\hat{v}_1, \hat{v}_3\right)\right),$$

where $\hat{v}_j$ is the estimate for the variance of $\hat{\beta}_j$ determined in part (a). Assuming this model is correct, report a point estimate and 95% confidence interval for β.

(d) Do you think the model in part (c) is appropriate? Why or why not? Given your answer, do you suspect the confidence interval determined in part (c) is too narrow, too wide, or accurate? Justify your answer.

(e) Let y_{ji} be the log-bilirubin concentration at age j in individual $i = 1, \ldots, n_j$, and let $\boldsymbol{x}_{ji}$ be the covariates for individual i at age j used in part (a). Assume that $\mathbb{E}\left(y_{ji}\right) = \boldsymbol{x}_{ji}^T\boldsymbol{\gamma}_j$ and

$$\operatorname{Cov}\left(y_{ji}, y_{j'i'}\right) = \begin{cases}\sigma_j^2 & \text{if } (j,i) = (j',i') \\ \phi & \text{if } j \neq j' \text{ and } i = i' \\ 0 & \text{otherwise}\end{cases}.$$

diff person diff age

same person diff age => sign should be > 0

(i) What do you expect the sign of ϕ to be?

(ii) Based on the results from part (a), do you think it would be appropriate to let $\sigma_1^2 = \sigma_3^2$? Justify your answer.

(f) Using whatever method you deem most appropriate, estimate σ_j^2 and ϕ.

(g) Let $\boldsymbol{Y}_j = \left(y_{j1}, \ldots, y_{jn_j}\right)^T$ and define z_{ji} to be 1 if individual i at age j was diagnosed with recurrent wheeze, and 0 otherwise. Let $\boldsymbol{Z}_j = \left(z_{j1}, \ldots, z_{jn_j}\right)^T$. Suppose you parametrize the mean model for $\boldsymbol{Y}_j$ as

$$\mathbb{E}\left(\boldsymbol{Y}_j\right) = \boldsymbol{Z}_j\beta_j + \tilde{\boldsymbol{X}}_j\tilde{\boldsymbol{\gamma}}_j,$$

where β_j is defined in part (c) and $\tilde{\boldsymbol{X}}_j$ contains the other covariates at age j (that are not of interest). Show that for $\hat{\beta}_j$ the ordinary least squares estimate for β_j defined in part (c) and $\tilde{\boldsymbol{Q}}_j = I_{n_j} - \tilde{\boldsymbol{X}}_j\left(\tilde{\boldsymbol{X}}_j^T\tilde{\boldsymbol{X}}_j\right)^{-1}\tilde{\boldsymbol{X}}_j^T$,

$$\hat{\beta}_j = \left(\boldsymbol{Z}_j^T\tilde{\boldsymbol{Q}}_j\boldsymbol{Z}_j\right)^{-1}\boldsymbol{Z}_j^T\tilde{\boldsymbol{Q}}_j\boldsymbol{Y}_j.$$

(h) Using parts (f) and (g), estimate $\boldsymbol{V} = \operatorname{Var}\left\{\begin{pmatrix}\hat{\beta}_1 \\ \hat{\beta}_3\end{pmatrix}\right\}$. Comment on the off-diagonal elements of this matrix.

(i) Assume $\hat{\beta}_1, \hat{\beta}_3$ are jointly normal and that $\beta_1 = \beta_3 = \beta$. Use your estimate for $\boldsymbol{V}$ from part (h) to provide a point estimate and 95% confidence interval for β.