STAT611 **Homework 7** **Ungraded**

1. Consider the model

$$y_i = \beta + bx_i + e_i \text{ for } i = 1, \ldots, n \tag{1}$$

where $x_1, \ldots, x_n$ are known fixed values, β is an unknown parameter, $b \sim N(0, \sigma_b^2)$ independently of $e_1, \ldots, e_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$. This is a simple linear regression model with a random slope coefficient. The parameter space for the model is $\{(\beta, \sigma_b^2, \sigma^2) : \beta \in \mathbb{R}, \sigma_b^2 > 0, \sigma^2 > 0\}$. Let $\underline{y} = (y_1, \ldots, y_n)'$, $\underline{x} = (x_1, \ldots, x_n)'$, and $X = [\underline{1}, \underline{x}]$. Assume that $\text{rank}(X) = 2$.

   (a) Determine $\text{Var}(\underline{y})$.

   (b) Consider the quadratic forms $\underline{y}'(P_X - P_{\underline{x}})\underline{y}/\sigma^2$ and $\underline{y}'(I - P_X)\underline{y}/\sigma^2$.

      i. Prove that $\underline{y}'(P_X - P_{\underline{x}})\underline{y}/\sigma^2$ has a χ^2 distribution under model (1) and provide fully simplified, non-matrix expressions for its degrees of freedom and non-centrality parameter.

      ii. Prove that $\underline{y}'(I - P_X)\underline{y}/\sigma^2$ has a χ^2 distribution under model (1) and provide fully simplified, non-matrix expressions for its degrees of freedom and non-centrality parameter.

      iii. Are the quadratic forms independent of one another?

   (c) Explain how to use the quadratic forms in part (b) to test $H_0 : \beta = 0$ vs. $H_A : \beta \neq 0$. Specifically,

      i. provide a test statistic,

      ii. state its distribution under H_0,

      iii. state its distribution under H_A, and

      iv. explain how you would use the test statistic to obtain a p-value for the test of H_0 vs. H_A.

Note that the next two problems and their solutions use n instead of T and p instead of K to denote the sample size and the number of columns of X, respectively. Likewise, p^* is used instead of K^* to denote the rank of the design matrix.

2. Suppose $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ where $\underline{\varepsilon} \sim N(0, \sigma^2 I)$. Consider partitioning $\underline{y}$ into its first element and the rest as $\underline{y} = [y_1, \underline{y}_2']'$. Similarly partition X as $X = [\underline{x}_1, X_2']'$ so that $\underline{x}_1'$ represents the first row of the X matrix. Suppose that X and X_2 are both of full column rank. The *deleted residual* for the first observation is defined as $d_1 = y_1 - \hat{y}_{1(1)}$, where $\hat{y}_{1(1)} = \underline{x}_1'\hat{\underline{\beta}}_{(1)} = \underline{x}_1'(X_2'X_2)^{-1}X_2'\underline{y}_2$.

(a) Show that $\underline{x}_1'(X_2'X_2)^{-1}X_2'\underline{y}_2$ is the BLUP of y_1 if we were to delete y_1 from the data set.

(b) Prove that $d_1 = \frac{y_1 - \hat{y}_1}{1 - \underline{x}_1'(X'X)^{-1}\underline{x}_1}$, where $\hat{y}_1 = \underline{x}_1'\hat{\underline{\beta}} = \underline{x}_1'(X'X)^{-1}X'\underline{y}$. This equation shows that the deleted residual is the ordinary residual divided by one minus the observation's *leverage* (recall the diagonal elements of $X(X'X)^{-1}X'$ are known as the leverage values). Note that this expression for the deleted residual allows the computation of all deleted residuals from the fit of a single multiple regression, a fact that you most likely learned about in STAT 500.

(c) Determine the distribution of d_1.

(d) Let $\hat{\sigma}^2_{(1)} = (\underline{y}_2 - X_2\hat{\underline{\beta}}_{(1)})'(\underline{y}_2 - X_2\hat{\underline{\beta}}_{(1)})/(n-p-1)$. Determine the distribution of

$$\frac{d_1}{\hat{\sigma}_{(1)}\sqrt{1 + \underline{x}_1'(X_2'X_2)^{-1}\underline{x}_1}}.$$

(e) Now suppose $\text{Var}(\varepsilon_1) = \sigma^2 + \sigma^2_\delta$, where $\sigma^2_\delta \geq 0$. Explain how the statistic in part (d) could be used to test the hypotheses

$$H_0 : \sigma^2_\delta = 0 \text{ vs. } H_A : \sigma^2_\delta > 0.$$

3. Suppose $y_1, \ldots, y_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$.

(a) Provide a vector of LIN error contrasts that could be used to obtain a REML estimate of σ^2.

(b) Specify the joint distribution of the error contrasts provided in part (a).

(c) Write down the likelihood function for the error contrasts provided in part (a).

(d) Find a simplified expression for the REML estimate of σ^2 from the error contrasts provided in part (a) by finding the value of σ^2 that maximizes the likelihood provided in part (c).

(e) How does your answer to part (c) compare to the usual unbiased estimate of σ^2?

(f) Would your answer to part (d) change if you were to choose a different set of LIN error contrasts in part (a)?