STATISTICAL MODELLING Part IIC / Michaelmas 2024
Example Sheet 1 (of 4)

In all the questions that follow, X is an n by p design matrix with full column rank and P is the orthogonal projection on to the column space of X. Also, let X_0 be the matrix formed from the first $p_0 < p$ columns of X and let P_0 be the orthogonal projection on to the column space of X_0. The vector $Y \in \mathbb{R}^n$ will be a vector of responses and we will define $\hat{\beta} := (X^T X)^{-1} X^T Y$, $\hat{\beta}_0 := (X_0^T X_0)^{-1} X_0^T Y$ and $\hat{\sigma}^2 := \|(I-P)Y\|^2/(n-p)$. By normal linear model, we mean the model $Y = X\beta + \varepsilon$, $\varepsilon \mid X \sim N_n(0, \sigma^2 I)$.

1. Show that
$$\|(P-P_0)Y\|^2 = \|(I-P_0)Y\|^2 - \|(I-P)Y\|^2 = \|PY\|^2 - \|P_0 Y\|^2.$$

2. Consider a random variable T that follows the F-distribution with d_1 and d_2 degrees of freedom.
   (a) When $d_1 = 1$, show that the distribution of T is the same as the square of a random variable that follows the t-distribution of d_2 degrees of freedom.
   (b) What can you say about the distribution of T when $d_2 \to \infty$? Use `R` to verify your conclusion. You may find some of these functions useful: `df`/`pf`, `dchisq`/`pchisq`, `hist`/`ecdf`/`qqplot`.

3. Show that the maximum likelihood estimator of σ^2 in the normal linear model is $\hat{\sigma}^2_{\text{MLE}} = \|(I - P)Y\|^2/n$. Find the distribution of $\hat{\sigma}^2_{\text{MLE}}$ and conclude that $\hat{\sigma}^2_{\text{MLE}}$ is a biased estimator of σ^2 but $\hat{\sigma}^2$ is unbiased. Is $\hat{\sigma}^2$ still unbiased unbiased when ε is not normally distributed (but has the same mean and variance given X)? Finish this question by constructing a confidence interval of σ^2 with level $1 - \alpha$ in the normal model.

4. In the normal linear model, show that the likelihood ratio statistic for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ can be written as a monotonically increasing function of $\|(I - P_0)Y\|^2 / \|(I - P)Y\|^2$.

5. Let the cuboid C be defined $C := \prod_{j=1}^{p} C_j(\alpha/p)$, where
$$C_j(\alpha) = \left[\hat{\beta}_j - \sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}}\, t_{n-p}(\alpha/2),\ \hat{\beta}_j + \sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}}\, t_{n-p}(\alpha/2)\right].$$
   Assuming the normal linear model, show that $\mathbb{P}(\beta \in C) \geq 1 - \alpha$.

6. Data are available on weights of two groups of three rats at the beginning of a fortnight and at its end. During the fortnight, one group was fed normally, and the other was given a growth inhibitor. The weights of the k^{th} rat in the j^{th} group before and after the fortnight are X_{jk} and Y_{jk} respectively. It is assumed that $Y_{jk} = \alpha_j + \beta_j X_{jk} + \varepsilon_{jk}$ where ε_{jk}, $j = 1, 2$, $k = 1, 2, 3$ are independent normally distributed noise with mean 0 and unknown variance.
   (a) Let W be the vector of responses, so $W = (Y_{11}, Y_{12}, Y_{13}, Y_{21}, Y_{22}, Y_{23})^T$, and similarly let δ be the vector of random errors. Write down the model above in the form $W = A\theta + \delta$, giving the design matrix A explicitly and expressing the vector of parameters θ in terms of the α_j and β_j.
   (b) The model is to be reparametrised in such a way that it can be specialised to (i) two parallel lines for the two groups, (ii) two lines with the same intercept, (iii) one common line for both groups, just by setting parameters to zero. Give one design matrix that can be made to correspond to (i), (ii) and (iii), just by dropping columns, specifying which columns are to be dropped for which cases.

7. Let Σ be a known $n \times n$ positive definite matrix. Consider the following linear model
$$Y = X\beta + \varepsilon, \qquad \varepsilon \mid X \sim N(0, \sigma^2 \Sigma),$$
   where β and σ^2 are unknown. Let $\hat{\beta}_\Sigma$ denote the maximum likelihood estimator of β.
   (a) Write down the optimisation problem solved by $\hat{\beta}_\Sigma$ and an expression for $\hat{\beta}_\Sigma$, assuming $X^T \Sigma^{-1} X$ is invertible.

(b) Show that $\hat{\beta}_\Sigma$ is the best linear unbiased estimator (BLUE) in this model.

(c) How does the optimisation problem and solution look like when Σ is diagonal? Read the `R` documentation for `lm` and find out how you can solve this problem numerically.

8. Consider the model $Y = \mu + \varepsilon$ where $\mathbb{E}(\varepsilon) = 0$ and $\text{Var}(\varepsilon) = \sigma^2 I$ and $\mu \in \mathbb{R}^n$ is a non-random vector. Suppose we have have performed least squares of Y on a fixed design matrix $X \in \mathbb{R}^{n\times p}$, so the fitted values are $\hat{\mu} = PY$. Show that if $Y^* = \mu + \varepsilon^*$ where ε^* is an independent and identically distributed copy of ε, then

$$\frac{1}{n}\mathbb{E}(\|\hat{\mu} - Y^*\|^2) = \sigma^2 + \frac{1}{n}\|(I-P)\mu\|^2 + \frac{\sigma^2 p}{n}.$$

Compare this to the *bias-variance tradeoff* in the lectures and identify the "bias^2" and "variance" terms.

9. Suppose we observe data (X, Y) generated from a normal linear model. Let $x^* \in \mathbb{R}^p$ be a fixed vector and $\epsilon^* \sim N(0, \sigma^2)$ be independent of (X, Y). Denote $Y^* = (x^*)^T\beta + \epsilon^*$. Construct $(1-\alpha)$-confidence intervals for $(x^*)^T\beta$ and Y^*. Which interval is shorter in length? Can you give an intuitive explanation to your answer?

10. Consider the normal linear model with a fixed design matrix X. Suppose only the first p_0 components of β are non-zero. Show that

$$\text{Var}(\hat{\beta}_{0,j}) \leq \text{Var}(\hat{\beta}_j) \qquad \text{for } j = 1, \dots, p_0.$$

Here $\hat{\beta}_{0,j}$ denotes the j^{th} component of $\hat{\beta}_0$. *Hint: Use the partial regression characterisation of $\hat{\beta}_j$.*

11. (Tripos 2022/II/13J) Consider the following `R` code:

```
> n <- 1000000
> sigma_z <- 1; sigma_x1 <- 0.5; sigma_x2 <- 1; sigma_y <- 2; beta <- 2
> Z <- sigma_z * rnorm(n)
> X1 <- Z + sigma_x1 * rnorm(n)
> X2 <- Z + sigma_x2 * rnorm(n)
> Y <- beta * Z + sigma_y * rnorm(n)
> lm(Y ~ Z)
Coefficients:
(Intercept)           Z
  -0.003089     1.999780
> lm(Y ~ X1)
Coefficients:
(Intercept)          X1
  -0.002904     1.600521
> lm(Y ~ X2)
Coefficients:
(Intercept)          X2
  -0.002672     0.997499
```

Describe the phenomenon you see from the output above, then give a mathematical explanation to this phenomenon. Do you expect the slope coefficient in the second model to be generally smaller than that in the first model? Do you think modifying (for example, doubling) the value of `sigma_y` will substantially alter the slope coefficient in the second model? Justify your answer.

12. This question is about understanding what can happen to the F-test when the linearity assumption does not hold. Consider the model in question 8 and let us assume ϵ follows the multivarite normal distribution. Define $\beta \in \mathbb{R}^p$ by $X\beta = P\mu$, so $Y = X\beta + (I-P)\mu + \varepsilon$, and partition $\beta = (\beta_0^T, \beta_1^T)^T$

as before. Suppose we try to test the hypothesis $H_0 : \beta_1 = 0$ against the alternative $H_1 : \beta_1 \neq 0$ by rejecting the null hypothesis when

$$F := \frac{\frac{1}{p-p_0}\|(P-P_0)Y\|^2}{\frac{1}{n-p}\|(I-P)Y\|^2} > F_{p-p_0,n-p}(\alpha).$$

We will show that the size of this test (the probability of rejecting H_0 when it is in fact true) is at most α.

(a) Show that the numerator and denominator of F are independent (no matter which hypothesis is true).

(b) What is the distribution of $\|(P-P_0)Y\|^2$ under the null hypothesis (i.e. when $Y = X_0\beta_0 + (I-P)\mu + \varepsilon$)?

(c) By considering the eigendecomposition of $I-P$, show that $\|(I-P)Y\|^2$ has the same distribution as

$$Z_1^2 + \cdots + Z_{n-p}^2,$$

where the Z_i are independent and $Z_i \sim N(\lambda_i, \sigma^2)$ for some λ_i such that

$$\sum_{i=1}^{n-p} \lambda_i^2 = \|(I-P)\mu\|^2.$$

(d) For any two real-valued random variables A and B, let us write $A \preceq B$ to mean $\mathbb{P}(A > x) \leq \mathbb{P}(B > x)$ for all $x \in (-\infty, \infty)$ (we say A is *stochastically less than* B). Now prove that if $A_1, \ldots, A_m$ and $B_1, \ldots, B_m$ are all independent real-valued random variables and $A_1 \preceq B_1, \ldots, A_m \preceq B_m$, then $A_1 + \cdots + A_m \preceq B_1 + \cdots + B_m$. *Hint: Use induction on* m *and recall that for real-valued random variables* U_1 *and* U_2*, the tower property of conditional expectation gives us that* $\mathbb{P}(U_1 + U_2 > x) = \mathbb{E}\{\mathbb{P}(U_1 > x - U_2 \mid U_2)\} := \mathbb{E}\{\mathbb{E}(\mathbb{1}_{\{U_1 > x - U_2\}} \mid U_2)\}$.

(e) Let $Z \sim \sigma^2\chi^2_{n-p}$. Show that

$$Z \preceq \|(I-P)Y\|^2.$$

Conclude that the size of the test mentioned at the beginning of this question is at most α.