Tutorial 12

1. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be two independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively. All parameters are assumed unknown. Let

$$R = \frac{\sum_{j=1}^{n_2} (Y_j - \overline{Y})^2}{\sum_{j=1}^{n_1} (X_j - \overline{X})^2}$$

and $F = \frac{(n_1-1)}{(n_2-1)} R$.

   (a) Show that $\frac{\sigma_1^2}{\sigma_2^2} F$ has an F_{n_2-1,n_1-1} distribution.

   (b) Compute the LR test of $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$ and show that it satisfies: reject H_0 for $F > x_1$ or $F < x_2$ where (x_1, x_2) satisfy:

$$\begin{cases} F_{n_2-1,n_1-1}(x_2) - F_{n_2-1,n_1-1}(x_1) = \alpha \\ \frac{x_1}{(1+\frac{n_2-1}{n_1-1}x_1)^{1+(n_1/n_2)}} = \frac{x_2}{(1+\frac{n_2-1}{n_1-1}x_2)^{1+(n_1/n_2)}}. \end{cases}$$

   Here $F_{v,w}(x) = \mathbb{P}(X \leq x)$ for $X \sim F_{v,w}$.

   (c) Can you show that the LR test with significance α is asymptotically equivalent to: reject H_0 for $F > F_{n_2-1,n_1-1;\alpha/2}$ or $F > \frac{1}{F_{n_1-1,n_2-1;\alpha/2}}$?

2. Consider the regression problem

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}$$

where $\underline{Y}$ is an n vector, X is an $n \times (p+q+1)$ matrix, $\underline{\beta} = \binom{\beta^{(1)}}{\beta^{(2)}}$, $\beta^{(1)}$ is a $p+1$ vector and $\beta^{(2)}$ is a q vector. Let X_1 be the matrix with the first $p+1$ columns of X and X_2 the matrix with the remaining q columns. Consider the hypothesis test $H_0 : \beta^{(2)} = 0$ versus $H_1 : \beta^{(2)} \neq 0$. Suppose that X has full rank.

   (a) Let $\widehat{\underline{\mu}}$ denote the ML estimator of $X\underline{\beta}$ for the full model and let $\widehat{\underline{\mu}}_0$ the estimator of $X_1\underline{\beta}^{(1)}$ under the null hypothesis. Show that

$$\mathbb{E}[\widehat{\mu} - \widehat{\mu}_0] = (I - X_1(X_1^t X_1)^{-1} X_1^t) X_2 \beta^{(2)}.$$

   (b) Let

$$F = \frac{(Q_{\text{res},I} - Q_{\text{res},II})/q}{Q_{\text{res},II}/(n-(p+q+1))}.$$

   Show that this has $F_{q,n-(p+q-1)}(\theta^2)$ distribution, where the non-centrality parameter θ^2 is:

$$\theta^2 = \frac{1}{\sigma^2} \beta^{(2)t} (X_2^t X_2 - X_2^t X_1 (X_1^t X_1)^{-1} X_1^t X_2) \beta^{(2)}.$$

3. Consider the one-way layout model

$$Y_{ij} = \alpha + \beta_i + \epsilon_{ij}, \quad i = 1, \ldots, p, \quad j = 1, \ldots, n_i$$

where ϵ_{ij} are i.i.d. $N(0, \sigma^2)$ and $\sum_{i=1}^{p} n_i \beta_i = 0$. Let $n = n_1 + \ldots + n_p$.

(a) Find the MLE $(\widehat{\alpha}, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)^t$ of the parameter vector $(\alpha, \beta_1, \ldots, \beta_p)$.

(b) Compute the covariance matrix for $(\widehat{\alpha}, \widehat{\beta}_1, \ldots, \widehat{\beta}_p)^t$.

(c) Give symmetric confidence intervals for α and β_k.

4. Consider again the one-way layout model of the previous exercise. Consider the two models:

$$\begin{cases} I & Y_{ij} = \alpha + \epsilon_{ij} \\ II & Y_{ij} = \alpha + \beta_i + \epsilon_{ij} \end{cases}$$

where Model II is the full model and Model I is the reduced model. Let $Q_{\text{res},I}$ and $Q_{\text{res},II}$ be the residual sums of squares of the two models. Show that

$$\frac{(Q_{\text{res},I} - Q_{\text{res},II})/(p-1)}{Q_{\text{res},II}/(n-p)} \sim F_{p-1,n-p}(\delta^2)$$

where the non-centrality parameter is:

$$\delta^2 = \frac{1}{\sigma^2} \sum_{k=1}^{p} n_k \beta_k^2.$$

5. Let $X = (X_1|X_2)$ where X_1 is $n \times p$, X_2 is $n \times q$, X is $n \times p + q$ and $X^t X$ is invertible. Show that

$$X(X^t X)^{-1} X^t X_1 = X_1.$$

6. Consider the linear model $Y = X\beta + \epsilon$ where $\epsilon \sim N(0, \sigma^2 I)$. Let $\widehat{Y} = X\widehat{\beta}$ denote the fitted values, where $\widehat{\beta}$ is the least squares estimator of β. Assume that $X_{.1} = \mathbf{1}_n$ (the n-vector with each entry 1). We use $\text{Var}(Z)$ to denote the *covariance* matrix of a random vector Z. Show that

(a) $\text{Var}(Y) = \text{Var}(\widehat{Y}) + \text{Var}(Y - \widehat{Y})$.

(b) $\sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2$.

7. Consider the linear model $Y = X\beta + \epsilon$ where the first column of X is a column of 1s. (This corresponds to multiple linear regression). Suppose that $\epsilon \sim N(0, \sigma^2 I)$. Define the *hat matrix* H as $H = X(X^t X)^{-1} X^t$. β is a p-vector of parameters. Show that:

(a) $\frac{1}{n} \leq H_{ii} \leq 1$ for all $i = 1, \ldots, p$,

(b) $\text{tr}(H) = p$,

(c) $H_{ii} = \text{Cor}(Y_i, \widehat{Y}_i)^2$.

You may use the fact that if $X = (X^{(1)}|X^{(2)})$, $H^{(1)} = X^{(1)}(X^{(1)\prime}X^{(1)})^{-1}X^{(1)\prime}$ and $H = X(X'X)^{-1}X'$ then $HH^{(1)} = H^{(1)}H = H^{(1)}$.

8. Consider again the regression model

$$Y = X\beta + \epsilon$$

where all elements of the first column of X are 1 and $\epsilon \sim N(0, \sigma^2 I)$. Define

$$R^2 = 1 - \frac{Q\text{res}}{Q_T}$$

where $\widehat{Y}_j$ are the fitted values, $\overline{Y} = \frac{1}{n}\sum_{j=1}^n Y_j$, $Q\text{res} = \sum_{j=1}^n (Y_j - \widehat{Y}_j)^2$ (the residual sum of squares) and $Q_T = \sum_{j=1}^n (Y_j - \overline{Y})^2$ (the total sum of squares).

(a) Show that

$$R^2 = \left(\frac{\sum_{i=1}^n (Y_i - \overline{Y})(\widehat{Y}_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (Y_i - \overline{Y})^2 \sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2}} \right)^2.$$

(b) Show that the test with critical region $R^2 > c$ is equivalent to the LRT test for testing the null model (where only β_0 is non-zero) against the full model (where all coefficients are non-zero).

(c) Show that R^2 is distributed according to a Beta $\left(\frac{p-1}{2}, \frac{n-p}{2}\right)$ distribution.

9. Let $Y = X\beta + \epsilon$ where $\epsilon \sim N(0, \sigma^2 I_n)$, X is $n \times p$ of full rank, $p < n$ and let $\widehat{Y} = X(X^tX)^{-1}X^tY$, the projection onto $\mathcal{S} = \{\mu : \mu = X\beta \quad \beta \in \mathbb{R}^p\}$. Let $H = X(X^tX)^{-1}X^t$, the projection matrix. Let Y^* be independent and indentically distributed with Y. Show that:

$$\mathbb{E}\left[\left|Y^* - \widehat{Y}\right|^2\right] = \mathbb{E}\left[\left|Y - \widehat{Y}\right|^2\right] + 2\sigma^2 \text{tr}(H).$$

Answers

1. (a)

$$W := \frac{\sum_{j=1}^{n_2}(Y_j - \overline{Y})^2}{\sigma_2^2} \sim \chi^2_{n_2-1}, \qquad V := \frac{\sum_{j=1}^{n_1}(X_j - \overline{X})^2}{\sigma_1^2} \sim \chi^2_{n_1-1}, \qquad V \perp W.$$

From the definition of an F distribution,

$$G := \frac{W/(n_2-1)}{V/(n_1-1)} \sim F_{n_2-1,n_1-1}.$$

Therefore

$$G = \frac{\sigma_1^2}{\sigma_2^2}\frac{(n_1-1)}{n_2-2)}\frac{\sum_{j=1}^{n_2}(Y_j - \overline{Y})^2}{\sum_{j=1}^{n_1}(X_j - \overline{X})^2} = \frac{\sigma_1^2}{\sigma_2^2}F \sim F_{n_2-1,n_1-1}.$$

as required.

(b) The likelihood is:

$$L(\mu_1, \mu_2, \sigma_1, \sigma_2) = \frac{1}{(2\pi)^{(n_1+n_2)/2}\sigma_1^{n_1}\sigma_2^{n_2}} \exp\{-\frac{1}{2\sigma_1^2}\sum_{j=1}^{n_1}(x_j - \mu_1)^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{n_2}(y_j - \mu_2)^2\}$$

The likelihood ratio statistic is:

$$\lambda(x,y) = \frac{\sup_{\mu_1,\mu_2,\sigma} L(\mu_1, \mu_2, \sigma, \sigma)}{\sup_{\mu_1,\mu_2,\sigma_1,\sigma_2} L(\mu_1, \mu_2, \sigma_1, \sigma_2)}$$

For the numerator (restriction to H_0 true), the likelihood, subject to the constraint that $\sigma_1 = \sigma_2 = \sigma$ is maximised for $(\mu_1, \mu_2, \sigma^2) = (\widehat{\mu}_{10}, \widehat{\mu}_{20}, \widehat{\sigma}_0^2)$ where

$$(\widehat{\mu}_{10}, \widehat{\mu}_{20}, \widehat{\sigma}_0^2) = (\overline{x}, \overline{y}, \frac{1}{n_1+n_2}(\sum_{j=1}^{n_1}(x_j - \overline{x})^2 + \sum_{j=1}^{n_2}(y_j - \overline{y})^2))$$

For the denominator (no restrictions on parameter space) the likelihood is maximised for $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = (\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1^2, \widehat{\sigma}_2^2)$, where

$$(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}_1^2, \widehat{\sigma}_2^2) = (\overline{x}, \overline{y}, \frac{1}{n_1}\sum_{j=1}^{n_1}(x_j - \overline{x})^2, \frac{1}{n_2}\sum_{j=1}^{n_2}(y_j - \overline{y})^2).$$

Note:

$$\widehat{\sigma}_0^2 = \frac{n_1}{n_1+n_2}\widehat{\sigma}_1^2 + \frac{n_2}{n_1+n_2}\widehat{\sigma_2}^2$$

then, using the usual trick of

$$\begin{array}{l}\left(\sum_{j=1}^{n_1}(x_j-\overline{x})^2+\sum_{j=1}^{n_2}(y_j-\overline{y})^2\right)=(n_1+n_2)\widehat{\sigma}_0^2,\\ \sum_{j=1}^{n_1}(x_j-\overline{x})^2=n_1\widehat{\sigma}_1^2,\\ \sum_{j=1}^{n_2}(y_j-\overline{y})^2=n_2\widehat{\sigma}_2^2\end{array}$$

gives:

$$\begin{aligned}\lambda(x,y) &= \frac{\frac{1}{(2\pi)^{(n_1+n_2)/2}\widehat{\sigma}_0^{n_1+n_2}}\exp\left\{-\frac{1}{2\widehat{\sigma}_0^2}\left(\sum_{j=1}^{n_1}(x_j-\overline{x})^2+\sum_{j=1}^{n_2}(y_j-\overline{y})^2\right)\right\}}{\frac{1}{(2\pi)^{(n_1+n_2)/2}\widehat{\sigma}_1^{n_1}\widehat{\sigma}_2^{n_2}}\exp\left\{-\frac{1}{2\widehat{\sigma}_1^2}\sum_{j=1}^{n_1}(x_j-\overline{x})^2-\frac{1}{2\widehat{\sigma}_2^2}\sum_{j=1}^{n_2}(y_j-\overline{y})^2\right\}}\\ &= \frac{\widehat{\sigma}_1^{n_1}\widehat{\sigma_2}^{n_2}}{\widehat{\sigma}_0^{n_1+n_2}}=\frac{(n_1+n_2)^{n_1+n_2}}{n_1^{n_1/2}n_2^{n_2/2}}\left(\frac{n_1\widehat{\sigma}_1^2}{n_1\widehat{\sigma}_1^2+n_2\widehat{\sigma}_2^2}\right)^{n_1/2}\left(\frac{n_2\widehat{\sigma}_2^2}{n_1\widehat{\sigma}_1^2+n_2\widehat{\sigma}_2^2}\right)^{n_2/2}\\ &= \frac{(n_1+n_2)^{(n_1+n_2)/2}}{n_1^{n_1/2}n_2^{n_2/2}}\left(\frac{1}{1+\frac{\sum_j(y_j-\overline{y})^2}{\sum_j(x_j-\overline{x})^2}}\right)^{(n_1+n_2)/2}\left(\frac{\sum_j(y_j-\overline{y})^2}{\sum_j(x_j-\overline{x})^2}\right)^{n_2/2}\\ &= \frac{(n_1+n_2)^{(n_1+n_2)/2}}{n_1^{n_1/2}n_2^{n_2/2}}\frac{R^{n_2/2}}{(1+R)^{(n_1+n_2)/2}}\end{aligned}$$

Then

$$\lambda(x,y)=\frac{(n_1+n_2)^{(n_1+n_2)/2}}{n_1^{n_1/2}n_2^{n_2/2}}\frac{(\frac{n_2-1}{n_1-1}F)^{n_2/2}}{(1+\frac{n_2-1}{n_1-1}F)^{(n_1+n_2)/2}}.$$

Reject H_0 if and only if $\lambda(x,y)<c$. We would like to show that this implies: reject H_0 for $F<k_1$ and $F>k_2$ for some k_1 and k_2, which we will then compute (or at least find an expression for).

Note that $\lambda=\lambda(F)$ (it is a function of F). As a function of F,

$$\frac{d}{dF}\log\lambda(F)=0\Leftrightarrow F=\frac{1-\frac{1}{n_1}}{1-\frac{1}{n_2}}.$$

Therefore $\lambda(0)=\lambda(+\infty)=0$, $\lambda(F)$ increases from 0 to a unique maximum at $F=\frac{1-\frac{1}{n_1}}{1-\frac{1}{n_2}}$ and then decreases to 0. The rejection region therefore has the form $F<k_1$, $F>k_2$ as required. Since $F\sim F_{n_2-1,n_1-1}$, k_1 and k_2 satisfy the following two equations: with confidence level $1-\alpha$,

$$\begin{cases}F_{n_2-1,n_1-1}(k_2)-F_{n_2-1,n_1-1}(k_1)=1-\alpha\\ \frac{k_1}{(1+\frac{n_2-1}{n_1-1}k_1)^{1+(n_1/n_2)}}=\frac{k_2}{(1+\frac{n_2-1}{n_1-1}k_2)^{1+(n_1/n_2)}}\end{cases}$$

For the first of these, $F_{n_2-1,n_1-1}(x)=\mathbb{P}(F\leq x)$ for $F\sim F_{n_2-1,n_1-1}$. For the second of these, since we reject for $\lambda\leq k$ for some value k, we have $\lambda(k_1)=\lambda(k_2)=k$.

2. (a) For the reduced model, $\beta^{(1)}$ is estimated by

$$\beta^{(1)*} = (X_1^t X_1)^{-1} X_1^t Y$$

so that (using $\mathbb{E}[Y] = X\beta$)

$$\mathbb{E}[\beta^{(1)*}] = (X_1^t X_1)^{-1} X_1^t (X_1 | X_2) \begin{pmatrix} \beta^{(1)} \\ \beta^{(2)} \end{pmatrix} = \beta^{(1)} + (X_1^t X_1)^{-1} X_1^t X_2 \beta^{(2)}$$

and hence, using $\mathbb{E}[\widehat{\mu}_0] = X_1 \mathbb{E}[\beta^{(1)*}]$ and $\mathbb{E}[\widehat{\mu}] = X\beta$, we have:

$$\begin{aligned} \mathbb{E}\left[\widehat{\mu} - \widehat{\mu}_0\right] &= (X_1|X_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} - X_1 \beta^{(1)} - X_1 (X_1^t X_1)^{-1} X_1^t X_2 \beta^{(2)} \\ &= \left(I - X_1 (X_1^t X_1)^{-1} X_1^t\right) X_2 \beta^{(2)}. \end{aligned}$$

(b) The non-centrality parameter is:

$$\theta^2 = \frac{1}{\sigma^2} \left(\mathbb{E}\left[\widehat{\mu} - \widehat{\mu}_0\right]^t \mathbb{E}\left[\widehat{\mu} - \widehat{\mu}_0\right] \right)$$

and, using the previous part,

$$\begin{aligned} \mathbb{E}\left[\widehat{\mu} - \widehat{\mu}_0\right]^t \mathbb{E}\left[\widehat{\mu} - \widehat{\mu}_0\right] &= \beta^{(2)t} X_2^t (I - X_1 (X_1^t X_1)^{-1} X_1^t)(I - X_1 (X_1^t X_1)^{-1} X_1^t) X_2 \beta^{(2)} \\ &= \beta^{(2)t} (X_2^t X_2 - X_2^t (X_1^t X_1)^{-1} X_1^t X_2) \beta^{(2)} \end{aligned}$$

From lectures (moving to canonical co-ordinates): $(Q_{\text{res},I} - Q_{\text{res},II}) \sim \chi_q^2(\theta^2)$ and $Q_{\text{res},II} \sim \chi^2_{n-(p+q+1)}$. These are independent. The result follows from the definition of the non-central F distribution.

A quick reminder of lectures: consider a linear model $Y = (X_1|X_2) \left(\frac{\beta^{(1)}}{\beta^{(2)}} \right) + \epsilon$ where $\beta^{(1)}$ is a $p+1$ vector, $\beta^{(2)}$ is a q vector and let

$$\mathcal{S}_1 = \{X_1 \beta : \beta \in \mathbb{R}^{p+1}\} \qquad \mathcal{S} = \{(X_1|X_2)\gamma : \gamma \in \mathbb{R}^{p+q+1}\}$$

then we can find an $n \times n$ orthonormal matrix $V = \begin{pmatrix} V^{(1)} \\ \hline V^{(2)} \\ \hline V^{(3)} \end{pmatrix}$ where $V^{(1)}$ spans the space $\mathcal{S}_1$ and $\left(\frac{V^{(1)}}{V^{(2)}} \right)$ spans the space $\mathcal{S}$. Let $U = VY$. Then

$$U \sim N(VX\beta, \sigma^2 VIV') = N(VX\beta, \sigma^2 I).$$

This is a vector of n independent random variables, where $U_i \sim N(\eta_i, \sigma^2)$ and $\eta_{p+q+2} = \ldots = \eta_n = 0$ for some $\eta_1, \ldots, \eta_{p+q+1}$.

Let

$$U^{(1)} = \begin{pmatrix} U_1 \\ \vdots \\ U_{p+1} \end{pmatrix} \qquad U^{(2)} = \begin{pmatrix} U_{p+2} \\ \vdots \\ U_{p+q+1} \end{pmatrix}$$

We have

$$\widehat{\mu} = (V^{(1)\prime}|V^{(2)\prime}) \begin{pmatrix} U^{(1)} \\ \hline U^{(2)} \end{pmatrix} \qquad \widehat{\mu}_0 = V^{(1)}U^{(1)}$$

so that

$$\begin{aligned} Q_{\text{res},I} - Q_{\text{res},II} &= |Y - \widehat{\mu}_0|^2 - |Y - \widehat{\mu}|^2 \\ &= |V'U - V^{(1)\prime}U^{(1)}|^2 - |V'U - (V^{(1)\prime}|V^{(2)\prime}) \begin{pmatrix} U^{(1)} \\ \hline U^{(2)} \end{pmatrix} |^2 \\ &= |(V^{(2)\prime}|V^{(3)\prime}) \begin{pmatrix} U^{(2)} \\ \hline U^{(3)} \end{pmatrix} |^2 - |V^{(3)\prime}U^{(3)}|^2 \\ &= U^{(2)\prime}V^{(2)}V^{(2)\prime}U^{(2)} = \sum_{j=p+2}^{p+q+1} U_j^2. \end{aligned}$$

while

$$\mathbb{E}[\widehat{\mu} - \widehat{\mu}_0] = V^{(2)\prime} \begin{pmatrix} \eta_{p+2} \\ \vdots \\ \eta_{p+q+1} \end{pmatrix}$$

so that

$$|\mathbb{E}[\widehat{\mu} - \widehat{\mu}_0]|^2 = \sum_{j=p+2}^{p+q+1} \eta_j^2$$

$\frac{Q_{\text{res},II}}{\sigma^2} = \sum_{j=p+q+2}^{n} U_j^2 \sim \chi^2_{n-(p+q+1)}$. Furthermore,

$$\frac{Q_{\text{res},I} - Q_{\text{res},II}}{\sigma^2} = \sum_{j=p+2}^{p+q+1} \left(\frac{U_j}{\sigma}\right)^2 \sim \chi_q^2 \left(\sum_{j=p+2}^{q} \left(\frac{\eta_j}{\sigma}\right)^2 \right)$$

and $\frac{Q_{\text{res},I} - Q_{\text{res},II}}{\sigma^2} \perp \frac{Q_{\text{res},II}}{\sigma^2}$ and the result follows by the definition of the non-central F distribution.

3. (a) The MLE for $(\mu_1, \ldots, \mu_p)$ is $(\overline{Y}_{1.}, \ldots, \overline{Y}_{p.})$.

$$\overline{Y}_{i.} = \widehat{\alpha} + \widehat{\beta}_i$$

$$\sum_i n_i \overline{Y}_{i.} = n\widehat{\alpha} \Rightarrow \widehat{\alpha} = \overline{Y}_{..}$$

$$\widehat{\beta}_i = \overline{Y}_{i.} - \overline{Y}_{..}$$

(b)

$$\mathrm{Var}(\widehat{\alpha}) = \frac{\sigma^2}{n}$$

$$\begin{aligned}
\mathrm{Var}(\widehat{\beta}_i) &= \mathrm{Var}(\overline{Y}_{i.} - \overline{Y}_{..}) \\
&= \mathrm{Var}((1 - \frac{n_i}{n})\overline{Y}_{i.} - \sum_{j \neq i} \frac{n_j}{n} \overline{Y}_{j.}) \\
&= \left(1 - \frac{n_i}{n}\right)^2 \frac{\sigma^2}{n_i} + \sum_{j \neq i} \frac{n_j^2}{n^2} \frac{\sigma^2}{n_j} \\
&= (1 - \frac{n_i}{n})^2 \frac{\sigma^2}{n_i} + \left(1 - \frac{n_i}{n}\right) \frac{\sigma^2}{n} \\
&= \left(\frac{1}{n_i} - \frac{2}{n} + \frac{n_i}{n^2} + \frac{1}{n} - \frac{n_i}{n^2}\right) \sigma^2 = \left(\frac{1}{n_i} - \frac{1}{n}\right) \sigma^2
\end{aligned}$$

$$\mathrm{Cov}(\widehat{\alpha}, \widehat{\beta}_i) = \mathrm{Cov}(\overline{Y}_{..}, \overline{Y}_{i.}) - \mathrm{Var}(\overline{Y}_{..}) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$$

$$i \neq j: \qquad \mathrm{Cov}(\widehat{\beta}_i, \widehat{\beta}_j) = -\mathrm{Cov}(\overline{Y}_{i.}, \overline{Y}_{..}) - \mathrm{Cov}(\overline{Y}_{j.}, \overline{Y}_{..}) + \mathrm{Var}(Y_{..}) = -\frac{\sigma^2}{n}$$

(c)

$$\alpha \in \left(\widehat{\alpha} \pm \frac{s}{\sqrt{n}} t_{n-p,a/2}\right)$$

(d)

$$\beta_j \in \left(\widehat{\beta}_j \pm s\sqrt{\frac{1}{n_j} - \frac{1}{n}} t_{n-p,a/2}\right)$$

where a is the significance and

$$s = \sqrt{\frac{Q\mathrm{res}}{n-p}} \qquad Q\mathrm{res} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (Y_{ij} - \widehat{\alpha} - \widehat{\beta}_j)^2$$

4.

$$\delta^2 = \frac{1}{\sigma^2} |\mu - \mu_0|^2 = \frac{1}{\sigma^2} \sum_{i=1}^{p} n_i \beta_i^2$$

5.

$$X(X^tX)^{-1}X^tX = X \Rightarrow X(X^tX)^{-1}X^t(X_1|X_2) = (X_1|X_2) \Rightarrow X(X^tX)^{-1}X_1 = X_1.$$

6. (a) Using $\widehat{\epsilon} = Y - \widehat{Y}$ and $\text{Var}(Z)$ to denote the covariance matrix of a random vector Z,

$$\text{Var}(Y) = \text{Var}(\widehat{Y} + Y - \widehat{Y}) = \text{Var}(\widehat{Y}) + \text{Var}(\widehat{\epsilon}) + 2\text{Cov}(\widehat{Y}, Y - \widehat{Y})$$

Now, $\widehat{\beta} = (X'X)^{-1}X'Y$, so that $\widehat{Y} = X\widehat{\beta} = X(X'X)^{-1}X'Y$ and

$$\text{Cov}(\widehat{Y}, Y - \widehat{Y}) = \text{Cov}(X(X'X)^{-1}X'Y, (I - X(X'X)^{-1}X'Y) = X(X'X)^{-1}X'\text{Cov}(Y)(I - X(X'X)^{-1}X')'$$

Now use: $\text{Var}(Y) = \text{Var}(\epsilon) = \sigma^2 I$ and

$$X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X'$$

so that

$$\text{Cov}(\widehat{Y}, Y - \widehat{Y}) = \sigma^2 X(X^tX)^{-1}X^t(I - X(X^tX)^{-1}X^t) = 0$$

as required.

(b) We'll consider this in two ways. Firstly, directly and secondly, by putting into canonical variables. Directly:

$$\sum_{i=1}^n (Y_i - \overline{Y})^2 = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2 + 2\sum_{i=1}^n (Y_i - \widehat{Y}_i)(\widehat{Y}_i - \overline{Y})$$

$$\sum_{i=1}^n (Y_i - \widehat{Y}_i)(\widehat{Y}_i - \overline{Y}) = Y^t(I - X(X^tX)^{-1}X^t)(X(X^tX)^{-1}X^t - X_1(X_1^tX_1)^{-1}X_1^t)Y$$

where $X_1 = (1, \ldots, 1)^t$. From above (previous exercise, taking $X = (X_1|X_2)$), it follows that:

$$\sum_{i=1}^n (Y_i - \widehat{Y}_i)(\widehat{Y}_i - \overline{Y}) = 0$$

and the result follows.

Canonical variables: Assume X is $n \times r$, of rank r $U = A^tY$ where A is an orthonormal $n \times n$ matrix. We let $A_{.1} = (\frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})^t$ so that $U_1 = \sqrt{n}\overline{Y}$. We let $A_{.2}, \ldots, A_{.r}$ be the $r-1$ unit vectors, orthogonal to each other and to $A_{.1}$, so that $A_{.1}, \ldots, A_{.r}$ are an orthonormal basis for the space $\mathcal{S} = \{X\beta : \beta \in \mathbb{R}^r\}$. Let $B_1 = (A_{.1}|0|\ldots|0)$ (the first column $A_{.1}$ the other columns 0), $B_1 = (0|A_{.2}|\ldots|A_{.r}|0|\ldots|0)$ (the $n \times n$ matrix with the first column 0s and the subsequent $r-1$ columns $A_{.1}, \ldots, A_{.r}$ and the remaining columns 0. Let $B_3 = A - B_2 - B_1$. Then

$$\begin{aligned}
\sum (Y_i - \overline{Y})^2 &= (Y - \overline{Y}\mathbf{1}_n)^t(Y - \overline{Y}\mathbf{1}_n) \\
&= U^t(A - B_1)^t(A - B_1)U = U^tB_2^tB_2U + U^tB_3^tB_3U \\
&= (\widehat{Y} - \overline{Y}\mathbf{1}_n)^t(\widehat{Y} - \overline{Y}\mathbf{1}_n) + (Y - \widehat{Y})^t(Y - \widehat{Y})
\end{aligned}$$

as required.

7. (a) Firstly, to show that $H_{ii} \leq 1$ for each i: $I - H$ is non-negative definite, since $\widehat{\epsilon}^t\widehat{\epsilon} = Y^t(I - H)Y$, hence $H_{ii} \leq 1$. (otherwise there would exist a diagonal element of $I - H$ which was negative, say i. Take vector $v = (0, \ldots, 0, 1, 0, \ldots, 0)^t$ to get $v^t(I - H)v = 1 - H_{ii} < 1$). Secondly, to show that $H_{ii} \geq \frac{1}{n}$ for each i: let $X^{(1)} = (1, \ldots, 1)^t$ and $X = (X^{(1)}|X^{(2)})$ then

$$G := X(X^tX)^{-1}X^t - X^{(1)}(X^{(1)t}X^{(1)})^{-1}X^{(1)t}$$

is positive definite, since

$$G'G = G^2 = (H - H^{(1)})^2 = H^2 - HH^{(1)} - H^{(1)}H + H^{(1)2} = H - H^{(1)} = G$$

using the hint. Now,

$$H^{(1)} = X^{(1)}(X^{(1)\prime}X^{(1)})^{-1}X^{(1)\prime} = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$$

where $\mathbf{1}$ is the n vector with each entry 1. Therefore $H_{ii} = \frac{1}{n} + G_{ii}$ for all $1 \leq i \leq p$. By the argument for the previous part, $G_{ii} \geq 0$, from which the result follows.

(b) Since H is idempotent, it has eigenvalues 1 or 0. Since it is of rank p, it has p eigenvalues 1 and $n - p$ eigenvalues 0. The trace is the sum of the eigenvalues, hence the result follows.

(c)

$$\text{Cov}(Y, \widehat{Y}) = \text{Var}(\widehat{Y})$$

$$\text{Cor}(Y_i, \widehat{Y}_i) = \frac{\text{Var}(\widehat{Y}_i)}{\sqrt{\text{Var}(Y_i)\text{Var}(\widehat{Y}_i)}} = \frac{H_{ii}}{\sqrt{H_{ii}}} = \sqrt{H_{ii}}.$$

8. (a) Follows directly from previous exercise;

$$\sum_{i=1}^{n}(Y_i - \overline{Y})(\widehat{Y}_i - \overline{Y}) = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2.$$

(b)

$$\begin{aligned}\lambda(y) &= \frac{\sup_\sigma \sup_{\beta_0} \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^n (y_j - \beta_0)^2\right\}}{\sup_\sigma \sup_\beta \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^t(y - X\beta)\right\}} \\ &= \frac{\widetilde{\sigma}^{-n}}{\widehat{\sigma}^{-n}}\end{aligned}$$

where

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(y_j - \overline{y})^2 \qquad \widehat{\sigma}^2 = \frac{1}{n}(y - X\widehat{\beta})^t(y - X\widehat{\beta})$$

$$\widehat{\beta} = (X^tX)^{-1}X^tY \qquad \widehat{Y} = X(X^tX)^{-1}X^tY$$

It now follows that

$$\lambda(y) = \left(\frac{Q_{\text{res}}}{Q_T}\right)^{n/2}$$

where $Q_{\text{res}} = \sum_{j=1}^n (Y_j - \widehat{Y})^2$ and $Q_T = \sum_{j=1}^n (Y_j - \overline{Y})^2$. The likelihood ratio test is: reject H_0 for $\lambda(y) < k$ for some k which is equivalent to: reject H_0 for $R^2 > c$ for some c.

(c) Use the result: if $X \sim \text{Gamma}(a, \lambda)$ and $Y \sim \text{Gamma}(b, \lambda)$ then

$$\frac{X}{X+Y} \sim \text{Beta}(a, b).$$

This is basic calculus: let $V = \frac{X}{X+Y}$, then for $t \in (0, 1)$,

$$\mathbb{P}(V \le t) = \mathbb{P}(X \le \frac{t}{1-t} Y) = \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} \int_0^\infty dy y^{b-1} e^{-\lambda y} \int_{ty/(1-t)}^\infty dx x^{a-1} e^{-\lambda x}$$

Take derivative with respect to t to get the density:

$$f_V(t) = \frac{\lambda^{a+b}}{\Gamma(a)\Gamma(b)} \frac{t^{a-1}}{(1-t)^{a+1}} \int_0^\infty y^{a+b-1} e^{-\lambda y/(1-t)} dy$$

Now use: $z = \frac{\lambda y}{1-t}$ to get:

$$f_V(t) = \frac{1}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1} \int_0^\infty z^{a+b-1} e^{-z} dz = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1} \qquad t \in (0, 1)$$

as required.

Let $Q_M = \sum_{j=1}^n (\widehat{Y}_j - \overline{Y})^2$, then $Q_M \perp Q_{\text{res}}$ and

$$\frac{Q_M}{\sigma^2} \sim \chi^2_{p-1} = \Gamma(\frac{p-1}{2}, \frac{1}{2}) \qquad \frac{Q_{\text{res}}}{\sigma^2} \sim \chi^2_{n-p} = \Gamma(\frac{n-p}{2}, \frac{1}{2})$$

so that

$$R^2 = \frac{Q_T - Q_{\text{res}}}{Q_T} = \frac{Q_M}{Q_M + Q_{\text{res}}} \sim \text{Beta}\left(\frac{p-1}{2}, \frac{n-p}{2}\right).$$

9. Firstly, $HX\beta = X(X^tX)^{-1}X^tX\beta = X\beta$ so that

$$\widehat{Y} - X\beta = HY - X\beta = H(Y - X\beta) = H\epsilon.$$

Therefore $\mathbb{E}[\widehat{Y}] = X\beta$ and

$$\begin{aligned}
\mathbb{E}[|Y^* - \widehat{Y}|^2] &= \mathbb{E}[|Y^* - X\beta + X\beta - \widehat{Y}|^2] \\
&= \mathbb{E}[|Y^* - X\beta|^2] + \mathbb{E}[|\widehat{Y} - X\beta|^2] \\
&= n\sigma^2 + \text{tr}\text{Var}(\widehat{Y}) \\
&= n\sigma^2 + \sigma^2 \text{tr}(H)
\end{aligned}$$

using $\text{Var}(\widehat{Y}) = \text{Var}(HY) = \sigma^2 HIH^t = \sigma^2 H$.

For the right hand side, using $Y - \widehat{Y} = (I - H)Y$, $\mathbb{E}[Y] = \mathbb{E}[\widehat{Y}]$ and $(I - H)^2 = I - H$ gives:

$$\mathbb{E}[|(I - H)Y|^2] = \text{tr}\text{Var}((I - H)Y) = \sigma^2 \text{tr}(I - H) = n\sigma^2 - \sigma^2 \text{tr}(H)$$

so that

$$\mathbb{E}[|Y^* - \widehat{Y}|^2] = \mathbb{E}[|Y - \widehat{Y}|^2] + 2\sigma^2 \mathrm{tr}(H).$$