Tutorial 12

1. Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ random samples respectively.

   (a) Find the MLE of $\theta := (\mu_1, \mu_2, \sigma^2)$. Let c_n be the value such that $S^2 = c_n\widehat{\sigma}^2$ is an unbiased estimator of σ^2. What is c_n? What is S^2?

   (b) Consider testing $H_0 : \mu_1 \leq \mu_2$ versus $H_1 : \mu_1 > \mu_2$. Assume that $\alpha < \frac{1}{2}$. Show that the likelihood ratio test is equivalent to the test with critical (rejection) region

   $$\overline{x} - \overline{y} \geq s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}t_{n_1+n_2-2,\alpha}.$$

   Here $t_{p,\alpha}$ is the value such that $\mathbb{P}(T > t_{p,\alpha}) = \alpha$ for $T \sim t_p$.

   (c) Compute a normal approximation to the power function and use it to find the sample size n needed for the level 0.01 test to have power 0.95 when $n_1 = n_2 = \frac{n}{2}$ and $\frac{\mu_1-\mu_2}{\sigma} = \frac{1}{2}$.

2. Consider the linear Gaussian model $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ where $\underline{\epsilon} \sim N(\underline{0}, \sigma^2 I_n)$, put into canonical coordinates via an orthonormal transform $\underline{U} = A\underline{Y}$ where $U_i \sim N(\eta_i, \sigma^2)$ for $i = 1, \ldots, r$ and $U_i \sim N(0, \sigma^2)$ for $i = r+1, \ldots, n$ with unknown parameters $\underline{\eta} = (\eta_1, \ldots, \eta_r)^t$ and σ^2, and log likelihood function:

   $$\log L(\underline{\eta}, \sigma^2; \underline{u}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{r}(u_i - \eta_i)^2 - \frac{1}{2\sigma^2}\sum_{i=r+1}^{n} u_i^2 - \frac{n}{2}\log(2\pi\sigma^2).$$

   Show that the MLE for $(\underline{\eta}, \sigma^2)$ does not exist if $n = r$ and that it is given by $(U_1, \ldots, U_r, \frac{1}{n}\sum_{i=r+1}^{n} U_i^2)$ if $n \geq r+1$. Show, in particular, that $\widehat{\sigma}^2_{ML} = \frac{1}{n}\left|\underline{Y} - \widehat{\mu}\right|^2$

3. Consider a Gaussian linear model $Y = X\beta + \epsilon$, where Y is an n-vector, X is $n \times r$ of rank r ($r < n$) and $\epsilon \sim N(0, \sigma^2 I)$ and β is an r-vector of unknown parameters. σ^2 is unknown. Recall (from lectures) that the OLS estimator of $\widehat{\beta}$ is:

   $$\widehat{\beta} = (X^t X)^{-1} X^t Y.$$

   Show that $\widehat{\beta}_i$ is UMVU for each $i = 1, \ldots, r$ and that $S^2 = \frac{1}{n-r}\sum_{j=1}^{n}(Y_j - \widehat{Y}_j)^2$ is an UMVU estimator of σ^2, where $\widehat{Y} = X(X^t X)^{-1} X^t Y$.

4. Consider *simple* linear regression; there is one explanatory variable and

   $$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma^2) \quad \text{i.i.d.} \quad i = 1, \ldots, n$$

   where $x_1, \ldots, x_n$ are not all equal. Express this as a Gaussian linear model

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

   identifying X and $\underline{\beta}$.

(a) Show that

$$(X^tX)^{-1} = \frac{1}{\sum_{j=1}^n (x_j - \overline{x})^2} \begin{pmatrix} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{pmatrix}$$

where $\overline{x} = \frac{1}{n}\sum_{j=1}^n x_j$ and $\overline{x^2} = \frac{1}{n}\sum_{j=1}^n x_j^2$.

(b) Let $\binom{\widehat{\beta}_0}{\widehat{\beta}_1}$ denote the maximum likelihood estimator of $\beta = \binom{\beta_0}{\beta_1}$. What is the distribution of $\binom{\widehat{\beta}_0}{\widehat{\beta}_1}$?

(c) Let

$$S^2 = \frac{1}{n-2}\sum_{j=1}^n (Y_j - \widehat{\beta}_0 - \widehat{\beta}_1 x_j)^2.$$

What is the distribution of $\frac{(n-2)S^2}{\sigma^2}$?

(d) Suppose

$$Y(z) = \beta_0 + \beta_1 z + \epsilon \qquad \epsilon \sim N(0, \sigma^2).$$

Let s denote the observed value of S. Using $t_{p,\alpha}$ to denote the value such that $\mathbb{P}(T > t_{p,\alpha}) = \alpha$ for $T \sim t_p$, show that a symmetric confidence interval for $\mathbb{E}[Y(z)]$ is given by:

$$\left(\widehat{\beta}_0 + \widehat{\beta}_1 z \pm s\sqrt{\frac{1}{n} + \frac{(\overline{x} - z)^2}{\sum_{j=1}^n (x_j - \overline{x})^2}}\, t_{n-2,\alpha/2}\right).$$

(e) Let $Y_* = \beta_0 + \beta_1 z + \epsilon_*$ where $\epsilon_* \sim N(0, \sigma^2)$ is independent of $\epsilon_1, \ldots, \epsilon_n$ (Y_* is a new observation with explanatory variable set at z). Let $\overline{Y} = \frac{1}{n}\sum_{j=1}^n Y_j$ and let $\widehat{Y}^* = \widehat{\beta}_0 + \widehat{\beta}_1 z$ (the predictor of Y_* based on $Y_1, \ldots, Y_n$). Show that, if $\beta_1 = 0$, then

$$\mathbb{E}[(Y^* - \widehat{Y}^*)^2] \geq \mathbb{E}[(Y^* - \overline{Y})^2]$$

5. Consider the one way layout problem

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

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

(a) Show that

$$S^2 = \frac{\sum_{i=1}^p \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2}{\sum_{i=1}^p (n_i - 1)}$$

is an unbiased estimator of σ^2 and that

$$\frac{\sum_{i=1}^p \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i.})^2}{\sigma^2} \sim \chi^2_{n-p}.$$

(b) Show that a level $1-\alpha$ confidence intervals for $\beta_j - \beta_i$ is:

$$\beta_j - \beta_i \in \left(\overline{Y}_{j.} - \overline{Y}_{i.} \pm S t_{n-p;\alpha/2} \sqrt{\frac{n_i + n_j}{n_i n_j}} \right)$$

where S^2 is the unbiased estimator of σ, $\overline{Y}_{k.} = \frac{1}{n_k} \sum_{i=1}^{n_k} Y_{ki}$ and $t_{p,\alpha}$ denotes the value such that $\mathbb{P}(T > t_{p,\alpha}) = \alpha$ if $T \sim t_p$. Show that the level $1-\alpha$ confidence interval for σ^2 is given by:

$$\frac{(n-p)s^2}{k_{n-p;(\alpha/2)}} \leq \sigma^2 \leq \frac{(n-p)s^2}{k_{n-p;1-(\alpha/2)}}$$

where $k_{q,\beta}$ is the value such that $\mathbb{P}(V \geq k_{q,\beta}) = \beta$ if $V \sim \chi^2_q$.

(c) Find confidence intervals for $\psi = \frac{1}{2}(\beta_2 + \beta_3) - \beta_1$ and $\sigma^2_\psi := \mathbf{V}(\widehat{\psi})$ where $\widehat{\psi} = \frac{1}{2}(\widehat{\beta}_2 + \widehat{\beta}_3) - \widehat{\beta}_1$.

6. Show that if C is an $n \times r$ matrix of full rank r, $r \leq n$, then the $r \times r$ matrix $C^t C$ is of rank r and hence non singular.

   Hint: Because C^t is of rank r, it follows that for any r-vector x, $x^t C^t = 0$ implies $x = 0$. Use this to show that if x is a non zero r-vector, then $x^t C C^t x > 0$.

7. Consider the one-way layout model: k groups of observations, all random variables independent. For group j, $Y_{1,j}, \ldots, Y_{n_j,j} \sim N(\mu_j, \sigma^2)$. Let $n = n_1 + \ldots + n_k$ denote the total number of observations.

   (a) Compute the likelihood ratio test statistic for $H_0 : \mu_1 = \ldots = \mu_k$ versus $H_1 : \mu_i \neq \mu_j$ for some $i \neq j$.

   (b) Let $Q_{\text{res}} = \sum_{j=1}^k \sum_{i=1}^{n_j} (Y_{ij} - \overline{Y}_{.j})^2$ where $\overline{Y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}$, the sample average from group j. Let $Q_M = \sum_{j=1}^k n_j (\overline{Y}_{.j} - \overline{Y}_{..})^2$ where $\overline{Y}_{..} = \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^{n_j} Y_{ij}$ (the overall average). Here Q_{res} denotes the *residual* sum of squares, while Q_M denotes the *model* sum of squares. Show that the likelihood ratio test is equavalent to reject H_0 for $F := \frac{Q_M/(k-1)}{Q_{\text{res}}/(n-k)} > c$ for some $c > 0$.

   (c) Show that the statistic F has $F_{k-1,n-k}$ distribution.

Answers

1. (a) Computing maximum likelihood estimators for normal distribution parameters should be straightforward. The log-likelihood function is:

$$\log L(\mu_1, \mu_2, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(\sum_{j=1}^{n_1}(x_j-\mu_1)^2 + \sum_{j=1}^{n_2}(y_j-\mu_2)^2\right).$$

This is maximised for:
$\widehat{\mu}_1 = \overline{X}$, $\widehat{\mu}_2 = \overline{Y}$,

$$\widehat{\sigma}^2_{ML} = \frac{1}{n_1+n_2}\left(\sum_{j=1}^{n_1}(X_j-\overline{X})^2 + \sum_{j=1}^{n_2}(Y_j-\overline{Y})^2\right)$$

This estimator is biased; recall that

$$\frac{\sum_{j=1}^{n_1}(X_j-\overline{X})^2}{\sigma^2} \sim \chi^2_{n_1-1} \qquad \frac{\sum_{j=1}^{n_2}(Y_j-\overline{Y})^2}{\sigma^2} \sim \chi^2_{n_2-1}$$

and that, for $V \sim \chi^2_m$, $\mathbb{E}[V] = m$. Therefore:

$$\mathbb{E}[\widehat{\sigma}^2_{ML}] = \frac{n_1+n_2-2}{n_1+n_2}\sigma^2 \Rightarrow c_n = \frac{n_1+n_2}{n_1+n_2-2}$$

The unbiased estimator required in the question is therefore:

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

(b) Recall $H_0 : \mu_1 \leq \mu_2$ versus $H_1 : \mu_1 > \mu_2$. The log likelihood ratio test statistic is:

$$\lambda(x,y) = \frac{\sup_{\mu_1,\mu_2,\sigma\in H_0} L(\mu_1,\mu_2,\sigma;x,y)}{\sup_{\mu_1,\mu_2,\sigma} L(\mu_1,\mu_2,\sigma;x,y)} = \frac{L(\widehat{\mu}_{0,1},\widehat{\mu}_{0,2},\widehat{\sigma}_0)}{L(\widehat{\mu}_1,\widehat{\mu}_2,\widehat{\sigma})}$$

where $(\widehat{\mu}_1,\widehat{\mu}_2,\widehat{\sigma})$ are the MLE estimators for the full space

$$\Theta = \{(\mu_1,\mu_2,\sigma^2) : (\mu_1,\mu_2,\sigma^2) \in \mathbb{R}^2 \times \mathbb{R}_+\} = \mathbb{R}^2 \times \mathbb{R}_+.$$

These were computed in the previous part of the exercise. The values $(\widehat{\mu}_{0,1},\widehat{\mu}_{0,2},\widehat{\sigma}_0)$ are the values which maximise the likelihood over the null hypothesis space

$$\Theta_0 = \{(\mu_1,\mu_2,\sigma^2) \in \mathbb{R}^2 \times \mathbb{R}_+ : \mu_1 \leq \mu_2\}.$$

If $\overline{X} \leq \overline{Y}$, then (clearly) $(\widehat{\mu}_{01},\widehat{\mu}_{02},\widehat{\sigma}^2_0) = (\widehat{\mu}_1,\widehat{\mu}_2,\widehat{\sigma}^2)$ and hence $\lambda(x,y) = 1$ for $\overline{x} < \overline{y}$.

Now consider the other case, where $\overline{x} > \overline{y}$. The maximiser clearly does not lie in the interior of the space; in this case there are no solutions to the likelihood equations $\nabla_\theta \log L(\theta) = 0$ in the space Θ_0. Therefore the maximiser lies on the boundary.

Clearly, as $\mu_1 \to -\infty$ or $\mu_2 \to +\infty$, $\log L(\mu_1, \mu_2, \sigma) \to -\infty$, so the maximiser does not lie on the part of the boundary where parameter values are $\pm\infty$. Therefore, the maximiser lies on the boundary $\mu_1 = \mu_2$. Therefore, for $\overline{x} > \overline{y}$, $\widehat{\mu}_{0,1} = \widehat{\mu}_{02} = \widehat{\mu}_0$ where $(\widehat{\mu}_0, \widehat{\sigma}_0^2)$ are the values which maximise

$$\log L(\mu, \sigma) = -\frac{(n_1+n_2)}{2}\log(2\pi) - \frac{(n_1+n_2)}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\left(\sum_{j=1}^{n_1}(x_j-\mu)^2 + \sum_{j=1}^{n_2}(y_j-\mu)^2\right).$$

From this,

$$\begin{aligned}\widehat{\mu}_0 &= \tfrac{1}{n_1+n_2}\left(\textstyle\sum_{j=1}^{n_1} X_j + \sum_{j=1}^{n_2} Y_j\right)\\ \widehat{\sigma}_0^2 &= \tfrac{1}{n_1+n_2}\left(\textstyle\sum_{j=1}^{n_1}(X_j-\widehat{\mu}_0)^2 + \sum_{j=1}^{n_2}(Y_j-\widehat{\mu}_0)^2\right)\end{aligned}$$

To compute the likelihood ratio:

$$\begin{aligned} L(\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\sigma}^2) &= \frac{1}{(2\pi)^{(n_1+n_2)/2}\widehat{\sigma}^{n_1+n_2}} \exp\left\{-\frac{1}{2\widehat{\sigma}^2}\left(\sum_{j=1}^{n_1}(x_j-\widehat{\mu}_1)^2 + \sum_{j=1}^{n_2}(y_j-\widehat{\mu}_2)^2\right)\right\}\\ &= \frac{1}{(2\pi)^{(n_1+n_2)/2}\widehat{\sigma}^{n_1+n_2}} \exp\left\{-\frac{(n_1+n_2)}{2}\right\}\end{aligned}$$

The last simplification comes from the formula for $\widehat{\sigma}^2$. Similarly, for the case $\overline{x} > \overline{y}$,

$$\begin{aligned} L(\widehat{\mu}_{0,1}, \widehat{\mu}_{0,2}, \widehat{\sigma}_0^2) &= \frac{1}{(2\pi)^{(n_1+n_2)/2}\widehat{\sigma}_0^{n_1+n_2}} \exp\left\{-\frac{1}{2\widehat{\sigma}_0^2}\left(\sum_{j=1}^{n_1}(x_j-\widehat{\mu}_0)^2 + \sum_{j=1}^{n_2}(y_j-\widehat{\mu}_0)^2\right)\right\}\\ &= \frac{1}{(2\pi)^{(n_1+n_2)/2}\widehat{\sigma}_0^{n_1+n_2}} \exp\left\{-\frac{n_1+n_2}{2}\right\}\end{aligned}$$

using $\widehat{\mu}_{01} = \widehat{\mu}_{02} = \widehat{\mu}_0$.

The LRT is therefore:

$$\lambda(x,y) = \begin{cases} 1 & \overline{x} \leq \overline{y} \\ \left(\frac{\widehat{\sigma}}{\widehat{\sigma}}\right)^{n_1+n_2} & \overline{x} > \overline{y}\end{cases}$$

Test: reject H_0 for $\lambda(x,y) < c$ where $c < 1$, (so a necessary condition for rejection is: $\overline{x} > \overline{y}$). To get it into the format required in the question, use:

$$\begin{aligned}(n_1+n_2)\widehat{\sigma}_0^2 &= \sum_{j=1}^{n_1}(X_j-\widehat{\mu}_0)^2+\sum_{j=1}^{n_2}(Y_j-\widehat{\mu}_0)^2\\ &= \sum_{j=1}^{n_1}(X_j-\overline{X})^2+n_1(\overline{X}-\widehat{\mu}_0)^2+\sum_{j=1}^{n_2}(Y_j-\widehat{\mu}_0)^2+n_2(\overline{Y}-\widehat{\mu}_0)^2\\ &= (n_1+n_2)\widehat{\sigma}^2+\frac{n_1n_2}{n_1+n_2}(\overline{X}-\overline{Y})^2\end{aligned}$$

so that

$$\widehat{\sigma}_0^2=\widehat{\sigma}^2+\frac{n_1n_2}{(n_1+n_2)^2}(\overline{X}-\overline{Y})^2.$$

Therefore:

$$\lambda(x,y)<c\Leftrightarrow\frac{\widehat{\sigma}_0^2}{\widehat{\sigma}^2}>\frac{1}{c^{2/(n_1+n_2)}}\Leftrightarrow\left(1+\frac{n_1n_2}{(n_1+n_2)^2}\frac{(\overline{X}-\overline{Y})^2}{\widehat{\sigma}^2}\right)>\frac{1}{c^{2/(n_1+n_2)}}$$

Since $\widehat{\sigma}^2=\frac{n_1+n_2-2}{n_1+n_2}S^2$ also need $\overline{X}-\overline{Y}>0$ to reject H_0, this gives a test of reject H_0 if and only if

$$\frac{\overline{X}-\overline{Y}}{S}>k$$

for a suitable value of k, which depends on the significance level α. Since

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{n_1+n_2-2}$$

it follows that $\mathbb{P}_{\mu_1,\mu_2}\left(\frac{\overline{X}-\overline{Y}}{S}>k\right)$ is increasing as $\mu_1-\mu_2$ increases and the result follows.

(c) The test is: Reject H_0 for $\frac{(\overline{x}-\overline{y})}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}>t_{n_1+n_2-2;\alpha}$, where $t_{n_1+n_2;\alpha}$ is the value such that $\mathbb{P}(T>t_{n_1+n_2-2;\alpha})=\alpha$.

Let $\theta=\mu_2-\mu_1$, then $\overline{X}-\overline{Y}\sim N\left(\theta,\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$ and hence

$$Z:=\frac{(\overline{X}-\overline{Y})-\theta}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1).$$

The *power* of the test is

$$\begin{aligned}\beta(\theta) &:= \mathbb{P}\left(\left.\frac{(\overline{X}-\overline{Y})}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}>t_{n_1+n_2-2;\alpha}\right|\mu_2-\mu_1=\theta\right)\\ &= \mathbb{P}\left(\left.\frac{(\overline{X}-\overline{Y})-\theta}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}>t_{n_1+n_2-2;\alpha}-\frac{\theta}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right|\mu_2-\mu_1=\theta\right)\end{aligned}$$

For large n_1, n_2, $S \simeq \sigma$ (law of large numbers) and $t_{n_1+n_2;\alpha} \simeq z_\alpha$ where $\mathbb{P}(Z > z_\alpha) = \alpha$ for $Z \sim N(0,1)$, so

$$\beta(\theta) \simeq \mathbb{P}(Z \geq z_\alpha - \frac{\theta}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}})$$

For the numbers given, $\alpha = 0.01$ and

$$0.95 = \beta(\frac{\sigma}{2}) \simeq 1 - \Phi(z_{0.01} - \frac{\sqrt{n}}{4})$$

Using $z_{0.01} = 2.33$ and $z_{0.05} = 1.64$, we have:

$$-1.64 = 2.33 - \frac{\sqrt{n}}{4} \Rightarrow n = 253$$

2. Likelihood equations obtained by: $\frac{\partial}{\partial \eta_i} \log L = 0$, $i = 1, \ldots, r$ and $\frac{\partial}{\partial \sigma} \log L = 0$. These give directly that the ML estimate has to satisfy:

$$\begin{cases} \widehat{\eta}_i = U_i & i = 1, \ldots, r \\ \frac{1}{\widehat{\sigma}^2} \sum_{j=r+1}^n U_j^2 = n \end{cases}$$

For $r = n$, $\widehat{\eta}_i = U_i$ so that the log likelihood evaluated at $\widehat{\eta}$ is:

$$\log L(\widehat{\eta}, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2)$$

which is maximised for $\sigma = 0$, which is not in the (open) parameter space $(0, +\infty)$, hence $\widehat{\sigma}_{ML}$ does not exist. Hence no solution for $n = r$.

For $n \geq r + 1$,

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{j=r+1}^n U_j^2.$$

Let $U = \binom{U^{(1)}}{U^{(2)}}$ where $U^{(1)} = (U_1, \ldots, U_r)^t$ and $U^{(2)} = (U_{r+1}, \ldots, U_n)^t$. Let $A = \binom{A^{(1)}}{A^{(2)}}$ where $A^{(1)}$ is $r \times n$ and $A^{(2)}$ is $n - r \times n$. Note that $\widehat{\mu} = A^{(1)t}U^{(1)}$ so that $Y - \widehat{\mu} = A^{(2)t}U^{(2)}$. It follows that

$$\sum_{j=r+1}^n U_j^2 = U^{(2)t}U^{(2)} = U^{(2)t}A^{(2)}A^{(2)t}U^{(2)} = |Y - \widehat{\mu}|^2.$$

3. Unbiased follows directly from lectures:

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

so that

$$\mathbb{E}[\widehat{\beta}] = (X^tX)^{-1}X^t\mathbb{E}[Y] = (X^tX)^{-1}X^tX\beta = \beta.$$

For the sample standard deviation, let $H = X(X^tX)^{-1}X^t$ then H is idempotent, of rank r and hence $H = PDP^t$ where P is orthonormal and $D = \text{diag}(1, \ldots, 1, 0, \ldots, 0)$ where 1 appears with multiplicity r. Hence

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

Let $\eta = P^t\epsilon$ then $\eta \sim N(0, \sigma^2 I)$. Also,

$$(Y - \widehat{Y})^t(Y - \widehat{Y}) = \eta^t(I - D)\eta = \sum_{r+1}^{n} \eta_j^2$$

so that

$$\frac{(n - r)S^2}{\sigma^2} = \sum_{r+1}^{n} \left(\frac{\eta_j}{\sigma}\right)^2 \sim \chi^2_{n-r}.$$

From this, $\mathbb{E}[S^2] = \sigma^2$ so that the estimator is unbiased.

Now to show that the estimators are UMVU:

$$p(y_1, \ldots, y_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^t(y - X\beta)\right\}$$

and the argument inside $\exp\{-\frac{1}{2}(.)\}$ is:

$$\frac{1}{\sigma^2}(y^ty - y^tX\beta - \beta^tX^ty + \beta^tX^tX\beta).$$

The sufficient statistic is therefore:

$$T(y) = (y^ty, \sum_{j=1}^{n} X_{ji}y_j : i = 1, \ldots, r).$$

$\widehat{\beta}_i = \sum_{jk}(X^tX)^{-1}_{ij}X_{kj}y_k$ is clearly a linear function of the sufficient statistics. For the standard deviation:

$$(Y - \widehat{Y}^t)(Y - \widehat{Y}) = Y^tY - \widehat{Y}^t\widehat{Y}$$

This holds since

$$Y^t\widehat{Y} = Y^tHY = Y^tH^tHY = \widehat{Y}^t\widehat{Y}$$

Now $\widehat{Y} = X(X^tX)^{-1}X^tY$ which is a (linear) function of the sufficient statistics and hence

$$\mathbb{E}[S^2|T(Y)] = S^2.$$

The result follows by the Lehman-Scheffé theorem.

4. The purpose of this question is to see all the abstract results for $Y = X\beta + \epsilon$ in the concrete setting of a single explanatory variable. Here the formulae are more transparent and we can see (for example) what happens when there is ill-conditioning in the X matrix.

   (a) The matrix X is:

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

   and the parameter vector is:

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

   To get $(X'X)^{-1}$ (so that - for example - we can compute the covariance of the parameter vector estimator):

$$(X^tX) = \left(\begin{array}{cc} n & \sum_{j=1}^n x_j \\ \sum_{j=1}^n x_j & \sum_{j=1}^n x_j^2 \end{array} \right) = n \left(\begin{array}{cc} n & \overline{x} \\ \overline{x} & \overline{x^2} \end{array} \right)$$

   Using the usual formula for inverting a 2×2 matrix together with the obvious identity:

$$\det(X'X) = n(\overline{x^2} - \overline{x}^2) = \sum_{j=1}^n (x_j - \overline{x})^2$$

   gives:

$$(X^tX)^{-1} = \frac{1}{\sum_{j=1}^n (x_j - \overline{x})^2} \left(\begin{array}{cc} \overline{x^2} & -\overline{x} \\ -\overline{x} & 1 \end{array} \right)$$

   (b) The MLE is equal to the least squares estimator. From lectures,

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

   Plugging in $(X^tX)^{-1}$ which has been computed gives:

$$\begin{aligned} \left(\begin{array}{c} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{array} \right) &= (X^tX)^{-1} \left(\begin{array}{c} n\overline{Y} \\ n\overline{xY} \end{array} \right) \\ &= \frac{1}{\frac{1}{n}\sum_{j=1}^n (x_j - \overline{x})^2} \left(\begin{array}{c} \overline{x^2}\overline{Y} - \overline{x}\overline{xY} \\ \overline{xY} - \overline{x}\overline{Y} \end{array} \right) \\ &= \left(\begin{array}{c} \overline{Y} - \overline{x}\frac{\sum_{j=1}^n (x_j - \overline{x})(Y_j - \overline{y})}{\sum_{j=1}^n (x_j - \overline{x})^2} \\ \frac{\sum_{j=1}^n (x_j - \overline{x})(Y_j - \overline{y})}{\sum_{j=1}^n (x_j - \overline{x})^2} \end{array} \right). \end{aligned}$$

This gives the best fitting straight line in the least squares sense. Note that

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{x}.$$

(c) For the standard deviation estimate,

$$\frac{(n-2)S^2}{\sigma^2} = \frac{|Y - \widehat{\mu}|^2}{\sigma^2} \sim \chi^2_{n-2}.$$

Note: $n-2$ degrees of freedom is obtained from the previous exercise.

We may also see it directly: the argument goes as follows: $\widehat{Y} = X(X^tX)^{-1}X^tY$ so that the residuals are:

$$Y - \widehat{Y} = (I - X(X^tX)^{-1}X^t)Y = (I - H)\epsilon$$

where $H = X(X^tX)^{-1}X^t$ and $\epsilon \sim N(0, \sigma^2 I)$. This is because $Y = X\beta + \epsilon$ and $HX = X$. Note that $H^2 = H$ (straightforward computation). It therefore follows that all the eigenvalues are either 0 or 1. Therefore, since X is rank 2 it follows that H is of rank 2; 2 e-values are 1, the remaining are 0 and it is straightforward that that $I - H$ is rank $n-2$; the eigenvalues of matrix $I - H$ are $n-2$ 1's and 2 0'2. Let $D = \text{diag}(1, \ldots, 1, 0, 0)$ and let $I - H = PDP^t$ where P is orthonormal. Then

$$\sum (Y_i - \widehat{\beta}_0 - x_i\widehat{\beta}_1)^2 = (Y - \widehat{Y})^t(Y - \widehat{Y}) = \epsilon^t PDP^t \epsilon = \sum_{j=1}^{n-2} \eta_j^2$$

where $\eta = P^t\epsilon$. Since P is orthonormal, it follows that $\eta \sim N(0, \sigma^2 I)$.
Therefore, it follows that:

$$\frac{(n-2)S^2}{\sigma^2} \sim \chi^2_{n-2}.$$

$$\underline{\widehat{\beta}} \sim N(\underline{\beta}, (X^tX)^{-1}\sigma^2)$$

(d) Let $\underline{v} = (1, z)^t$ then

$$\mathbb{E}[Y(z)] = \underline{v}^t\underline{\beta}$$

$$\frac{\underline{v}^t\underline{\widehat{\beta}} - \underline{v}^t\underline{\beta}}{\sigma\sqrt{\underline{v}^t(X^tX)^{-1}\underline{v}}} \sim N(0,1)$$

$$\frac{\underline{v}^t\underline{\widehat{\beta}} - \underline{v}^t\underline{\beta}}{S\sqrt{\underline{v}^t(X^tX)^{-1}\underline{v}}} \sim t_{n-2}$$

with $1 - \alpha$ confidence,

$$\underline{v}^t\underline{\beta} \in \left(\underline{v}^t\underline{\widehat{\beta}} \pm s\sqrt{\underline{v}^t(X^tX)^{-1}\underline{v}} t_{n-2;\alpha/2}\right)$$

and

$$\underline{v}^t(X^tX)^{-1}\underline{v} = \frac{\overline{x^2} - 2z\overline{x} + z^2}{\sum_{j=1}^n (x_j - \overline{x})^2} = \frac{\frac{1}{n}\sum_{j=1}^n (x_j - \overline{x})^2 + (\overline{x} - z)^2}{\sum_{j=1}^n (x_j - \overline{x})^2}$$

and the result follows.

(e) From the previous part,

$$\mathbb{E}[(Y^* - \widehat{Y}^*)^2] = \mathbf{V}(Y^* - \widehat{Y}^*) = \mathbf{V}(Y^*) + \mathbf{V}(\widehat{Y}^*) = \sigma^2\left(1 + \frac{1}{n} + \frac{(\overline{x} - z)^2}{\sum_{j=1}^n (x_j - \overline{x})^2}\right)$$

while, under the assumption $\beta_1 = 0$,

$$\mathbb{E}[(Y^* - \overline{Y})^2] = \mathbf{V}(Y^*) + \mathbf{V}(\overline{Y}) = \sigma^2\left(1 + \frac{1}{n}\right)$$

and the result is clear.

5. (a)

$$\overline{Y}_{j.} - \overline{Y}_{i.} \sim N(\beta_j - \beta_i, \sigma^2(\frac{1}{n_j} + \frac{1}{n_i}))$$

$$S^2 = \frac{1}{n-p}\sum_{i=1}^p \sum_{j=1}^n (Y_{ij} - \overline{Y}_{i.})^2 \qquad n - p \qquad d.f.$$

is the unbiased estimator of σ^2. Then

$$\frac{(\overline{Y}_{j.} - \overline{Y}_{i.}) - (\beta_j - \beta_i)}{S\sqrt{\frac{n_i+n_j}{n_in_j}}} \sim t_{n-p}$$

and the confidence interval follows. The confidence interval for σ follows from:

$$\frac{(n-p)S^2}{\sigma^2} \sim \chi^2_{n-p}$$

hence the $1 - \alpha$ confidence bound is given by:

$$k_{n-p;1-(\alpha/2)} \leq \frac{(n-p)s^2}{\sigma^2} \leq k_{n-p;(\alpha/2)}$$

from which the result follows.

(b)

$$\widehat{\psi} \sim N\left(\psi, \sigma^2\left(\frac{1}{4n_2} + \frac{1}{4n_3} + \frac{1}{n_1}\right)\right)$$

the estimator of σ^2 is $S^2 = Q_{\text{res}} n - p$ given above with $n - p$ degrees of freedom and hence

$$\frac{1}{2}(\beta_2 + \beta_3) - \beta_1 \in \left(\frac{1}{2}\left(\overline{Y}_{2.} + \overline{Y}_{3.}\right) - \overline{Y}_{1.} \pm st_{n-p,\alpha/2}\sqrt{\frac{n_1n_3 + n_1n_2 - 4n_2n_3}{4n_1n_2n_3}}\right)$$

Similarly,

$$\mathbf{V}(\widehat{\psi}) = \frac{n_1n_3 + n_1n_2 + 4n_2n_3}{4n_1n_2n_3}\sigma^2$$

hence the confidence interval is:

$$\frac{n_1n_3+n_1n_2+4n_2n_3}{4n_1n_2n_3}\frac{(n-p)s^2}{k_{n-p;(\alpha/2)}} \leq \mathbf{V}(\widehat{\psi}) \leq \frac{n_1n_3+n_1n_2+4n_2n_3}{4n_1n_2n_3}\frac{(n-p)s^2}{k_{n-p;1-(\alpha/2)}}$$

6. $x^tC^tCx = 0$ implies that $x^tC^t = 0$ which implies that $x = 0$ so that if $x \neq 0$ then $x^tC^tCx \neq 0$ hence C^tC is (strictly) positive definite.

7. (a) Let $n = n_1 + \ldots + n_k$ denote the total number of experimental units. For $H_0 : \mu_1 = \ldots = \mu_k = \mu$, we have the maximiser $\widetilde{\mu} = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j} Y_{ij}$ and

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij} - \widetilde{\mu})^2$$

and the maximum likelihood under the constraint H_0 is: $\frac{1}{(2\pi)^{n/2}\widetilde{\sigma}^n}e^{-n/2}$.
For the unconstrained problem, the likelihood is maximised at $\widehat{\mu}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ij}$ and

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij} - \widehat{\mu}_j)^2.$$

The maximum likelihood for the unconstrained problem is: $\frac{1}{(2\pi)^{n/2}\widehat{\sigma}^n}e^{-n/2}$ and hence the likelihood ratio statistic is:

$$\lambda(y) = \left(\frac{\widehat{\sigma}}{\widetilde{\sigma}}\right)^n.$$

(b) Pythagorean identity: note that $\overline{Y}_{.j} = \widehat{\mu}_j$ and $\overline{Y}_{..} = \widetilde{\mu}$ from previous part.

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\widetilde{\mu})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{.j}+\overline{Y}_{.j}-\overline{Y}_{..})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(Y_{ij}-\overline{Y}_{.j})^2+\sum_{j=1}^{k}n_j(\overline{Y}_{.j}-\overline{Y}_{..})^2$$

so:

$$n\widetilde{\sigma}^2 = Q_{\text{res}} + Q_M \qquad n\widehat{\sigma}^2 = Q_{\text{res}}.$$

Therefore, the likelihood ratio test is:

$$\lambda(y) < c \Leftrightarrow \frac{Q_{res}}{Q_M + Q_{res}} < c^{2/n} \Leftrightarrow \frac{Q_M/(k-1)}{Q_{\text{res}}/(n-k)} > \left(\frac{n-k}{k-1}\right)\left(\frac{1-c^{2/n}}{c^{2/n}}\right) = k$$

establishing the result.

(c) It follows from the canonical representation (lectures) that $Q_M \perp Q_{\text{res}}$. Under $H_0 : \mu_1 = \ldots = \mu_k$, it follows that $\frac{Q_M}{\sigma^2} \sim \chi^2_{k-1}$ since the parameter space for $\mu_1, \ldots, \mu_k$ is k-dimensional and the parameter space for the mean under the null hypothesis is 1-dimensional, and $\frac{Q_{\text{res}}}{\sigma^2} \sim \chi^2_{n-k}$. The result follows from Proposition 11.4.