This is an introductory and revision sheet which you can practise on as the course begins. These exercises will not be covered in problems classes, mostly they should be revision. A couple of things below may be written in a different way to what you are used to, but the notation is explained.

1. **Likelihood ratio tests.**

   Suppose $y_1, \dots, y_n \overset{\text{iid}}{\sim} f(y;\theta)$ where θ is an element of the parameter space Θ. Let Θ_0 be a subset of Θ, where $\dim\Theta = p$, $\dim\Theta_0 = q$ and $q < p$. Here $\dim\Theta$ denotes the dimension of Θ, i.e. the number of free parameters in Θ, and similarly for $\dim\Theta_0$.

   Consider testing the null hypothesis $H_0\colon \theta \in \Theta_0$ against the general alternative $H_1\colon \theta \in \Theta$.

   (a) What is the definition of the *log-likelihood ratio statistic* for testing H_0? Denote this statistic by $\Lambda(y)$.

   Solution: Let $L(\theta; y) = \prod_{i=1}^n f(y_i;\theta)$ denote the likelihood and $\ell(\theta;y) = \log L(\theta;y)$ the log-likelihood. Then any of the following are possible expressions for $\Lambda(y)$:

   $$\Lambda(y) = 2\log\left(\frac{\max_{\theta\in\Theta} L(\theta;y)}{\max_{\theta\in\Theta_0} L(\theta;y)}\right) = 2\log\left(\frac{L(\widehat{\theta};y)}{L(\widehat{\theta}_0;y)}\right) = 2\left[\ell(\widehat{\theta};y) - \ell(\widehat{\theta}_0;y)\right]$$

   where $\widehat{\theta}$ is the MLE, and where $\widehat{\theta}_0$ is the θ at which the likelihood is maximised over $\theta \in \Theta_0$.

   (b) What is the approximate distribution of $\Lambda(y)$ under H_0? What can you say about the conditions required for this to be a good approximation?

   Solution: We have $\Lambda(y) \approx \chi^2_{p-q}$ under H_0. The most important condition is that the sample size n is large (the associated limiting result is that $\Lambda(y)$ converges in distribution to χ^2_{p-q} as $n \to \infty$, under certain conditions).
   It is assumed that the other parts of the setup, e.g. Θ, Θ_0, p, q, do not depend on n. Further regularity conditions are also required, including: the true value of θ is in the interior of Θ_0; we can differentiate $f(y;\theta)$ sufficiently often with respect to θ; in $\int f(y;\theta)\,dy$ we can differentiate under the integral sufficiently often with respect to θ. In particular it is assumed that the support of f does not depend on θ, i.e. the set $\{x : f(x;\theta) > 0\}$ does not depend on θ.

2. **Standard distributions.**

   Let $z_1, z_2, \dots \overset{\text{iid}}{\sim} N(0,1)$. Write down, in terms of $z_1, z_2, \dots$, a random variable whose distribution is:

   (a) The chi-squared distribution with r degrees of freedom.

   Solution: A χ^2_r random variable can be defined as the sum of r squared standard normals: so answer $= z_1^2 + \cdots + z_r^2$.

(b) The t distribution with r degrees of freedom.

Solution: A t_r random variable can be defined as the ratio

$$\frac{N(0,1)}{\sqrt{\chi_r^2/r}}$$

where the normal in the numerator and the chi-squared in the denominator are independent. So

$$\text{answer} = \frac{z_{r+1}}{\sqrt{(z_1^2 + \cdots + z_r^2)/r}}.$$

Note: it is important that the z in the numerator (z_{r+1} here) does not appear in the denominator – because the normal in the numerator and the chi-squared in the denominator must be independent.

(c) The F distribution with m and n degrees of freedom (if you are not familiar with the F distribution, look it up).

Solution: An $F_{m,n}$ random variable can be defined as the ratio

$$\frac{\chi_m^2/m}{\chi_n^2/n}$$

where the chi-squared random variables in the numerator and denominator are independent. So

$$\text{answer} = \frac{(z_1^2 + \cdots + z_m^2)/m}{(z_{m+1}^2 + \cdots + z_{m+n}^2)/n}.$$

Note: again it is important that the set of z_i in the numerator do not overlap with the set of z_j in the denominator – again to ensure independence.

3. Linear regression models.

Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n \tag{1}$$

where $\epsilon_1, \ldots, \epsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$.

(a) Model (1) can also be written

$$y_i = \gamma_0 + \gamma_1(x_i - \overline{x}) + \epsilon_i, \quad i = 1, \ldots, n$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$. Write down expressions for β_0 and β_1 in terms of γ_0 and γ_1. How do the MLEs $\widehat{\beta}_0$ and $\widehat{\beta}_1$ relate to the MLEs $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$?

Solution: We have

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \epsilon_i \\ &= (\gamma_0 - \gamma_1 \overline{x}) + \gamma_1 x_i + \epsilon_i. \end{aligned}$$

Equating coefficients of x_i, and the constant term, gives

$$\beta_1 = \gamma_1, \qquad \beta_0 = \gamma_0 - \gamma_1 \overline{x}.$$

By the invariance property of MLEs, the same relationships hold between the MLEs: $\widehat{\beta_1} = \widehat{\gamma_1}$ and $\widehat{\beta_0} = \widehat{\gamma_0} - \widehat{\gamma_1}\overline{x}$.

(b) Let

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \qquad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Note that (1) represents n equations: write down the appropriate matrix X so that these equations, written in matrix form, are

$$y = X\beta + \epsilon.$$

Solution: We have

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \qquad \text{hence} \quad X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

(c) Write down the likelihood for model (1), and show that the log-likelihood $\ell(\beta, \sigma^2; y)$ can be written

$$\begin{aligned} \ell(\beta, \sigma^2; y) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} SS(\beta) + \text{constant} \\ &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)^T (y - X\beta) + \text{constant} \end{aligned}$$

where $SS(\beta) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$, and where a superscript of T denotes transpose.

Solution: We have $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$, so the likelihood is

$$\begin{aligned} L(\beta, \sigma^2; y) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - \beta_0 - \beta_1 x_i)^2 \right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} SS(\beta) \right). \end{aligned}$$

Taking logs gives the first expression for $\ell(\beta, \sigma^2; y)$, where the additive constant is $-\frac{n}{2}\log 2\pi$.
Now $y - X\beta$ is an n-component vector:

$$y - X\beta = \begin{pmatrix} y_1 - \beta_0 - \beta_1 x_1 \\ \vdots \\ y_n - \beta_0 - \beta_1 x_n \end{pmatrix}.$$

But for any n-component vector:

$$\text{if} \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \qquad \text{then} \quad u^T u = \begin{pmatrix} u_1, & \dots, & u_n \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = \sum_{i=1}^{n} u_i^2.$$

Hence $(y - X\beta)^T(y - X\beta) = SS(\beta)$, giving the second expression for the log-likelihood.

(d) Consider the multiple regression model

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \dots, n.$$

If this model is written as $y = X\beta + \epsilon$, what is X?

Solution: X is the $n \times p$ matrix given by

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}.$$

4. Mean vectors and covariance matrices.

Let $y = (y_1, \dots, y_n)^T$ be an $n \times 1$ vector of random variables. Recall that the $n \times 1$ mean vector $\mu = (\mu_i)$ and the $n \times n$ covariance matrix $\Sigma = (\Sigma_{ij})$ of y are defined by

$$\begin{aligned} \mu_i &= E(y_i), \quad i = 1, \dots, n \\ \Sigma_{ii} &= \mathrm{var}(y_i), \quad i = 1, \dots, n \\ \Sigma_{ij} &= \mathrm{cov}(y_i, y_j), \quad i \neq j = 1, \dots, n. \end{aligned}$$

We write $E(y) = \mu$ and $\mathrm{var}(y) = \Sigma$.

If A is a matrix of constants with n columns, show that

(a) $\mathrm{var}(y) = E[(y - \mu)(y - \mu)^T]$

Solution:

$$
\begin{aligned}
&(y-\mu)(y-\mu)^T \\
&= \begin{pmatrix} y_1-\mu_1 \\ \vdots \\ y_n-\mu_n \end{pmatrix} \left(y_1-\mu_1,\ \ldots,\ y_n-\mu_n\right) \\
&= \begin{pmatrix} (y_1-\mu_1)^2 & (y_1-\mu_1)(y_2-\mu_2) & \ldots & (y_1-\mu_1)(y_n-\mu_n) \\ (y_2-\mu_2)(y_1-\mu_1) & (y_2-\mu_2)^2 & \ldots & (y_2-\mu_2)(y_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (y_n-\mu_n)(y_1-\mu_1) & (y_n-\mu_n)(y_2-\mu_2) & \ldots & (y_n-\mu_n)^2 \end{pmatrix}.
\end{aligned}
$$

The result follows on taking expectations of both sides because $E[(y_i-\mu_i)^2] = \text{var}(y_i)$, and $E[(y_i-\mu_i)(y_j-\mu_j)] = \text{cov}(y_i, y_j)$ for $i \neq j$.

(b) $E(Ay) = AE(y)$

Solution: Recall the linearity property of expectation: if $a_1, \ldots, a_n$ are constants, then $E(\sum_{i=1}^n a_i y_i) = \sum_{i=1}^n a_i E(y_i)$.
Hence

$$
\begin{aligned}
E[Ay] &= AE[y] \qquad \text{by the linearity property of expectation} \\
&= A\mu.
\end{aligned}
$$

(c) $\text{var}(Ay) = A\,\text{var}(y)A^T$.

Solution:

$$
\begin{aligned}
\text{var}(Ay) &= E[A(y-\mu)\{A(y-\mu)\}^T] && \text{using (a)} \\
&= E[A(y-\mu)(y-\mu)^T A^T] && \text{since } (AB)^T = B^T A^T \\
&= AE[(y-\mu)(y-\mu)^T]A^T && \text{by linearity of expectation} \\
&= A\,\text{var}(y)A^T.
\end{aligned}
$$

5. Fitting a linear regression in R.

Work through the following example in R and think about what this simple analysis is doing.

Old Faithful is a geyser in Yellowstone National Park, USA. The dataset `faithful` is built-in to R. To look at the first few rows of the dataset use:

```
head(faithful)
```

To see what the data represent, use `?faithful` and read the first few lines of the help page.

Note: the function `head()` shows only the first part of a data frame or vector. We use it here to avoid getting too much output. Type `faithful` if you want to see the whole data frame (too much information). To get summaries use:

```
str(faithful)
summary(faithful)
```

Consider using the duration of the current eruption to predict the length of time until the next eruption takes place. Plot the data, fit a simple linear regression and draw the regression line on the plot:

```
plot(waiting ~ eruptions, data = faithful,
     xlab = "duration of current eruption (minutes)",
     ylab = "time until next eruption (minutes)")
fit1 <- lm(waiting ~ eruptions, data = faithful)
abline(fit1, col = "blue")
```

Summarise the fitted model:

```
fit1
summary(fit1)
```

A couple of residual plots:

```
plot(resid(fit1) ~ fitted(fit1), main = "Residuals vs Fitted values")
qqnorm(resid(fit1), main = "Normal Q-Q plot of residuals")
qqline(resid(fit1))
```

Confidence and prediction intervals at durations of $1.6, 2.1, \ldots, 5.1$ minutes:

```
new <- data.frame(eruptions = seq(1.6, 5.1, by = 0.5))
predict(fit1, newdata = new)
predict(fit1, newdata = new, interval = "confidence")
predict(fit1, newdata = new, interval = "prediction")
```

Add confidence and prediction intervals to the original plot:

```
new <- data.frame(eruptions = seq(1.6, 5.1, by = 0.5))
p.conf <- predict(fit1, newdata = new, interval = "confidence")
p.pred <- predict(fit1, newdata = new, interval = "prediction")
erup <- new$eruptions
plot(waiting ~ eruptions, data = faithful,
     xlab = "duration of current eruption (minutes)",
     ylab = "time until next eruption (minutes)")
lines(p.conf[, 1] ~ erup, col = "blue") # fitted values are 1st column of p.conf
lines(p.conf[, 2] ~ erup, lty = 2) # lwr conf values are 2nd column
lines(p.conf[, 3] ~ erup, lty = 2) # upr conf values are 3rd column
lines(p.pred[, 2] ~ erup, lty = 2, col = "red")
lines(p.pred[, 3] ~ erup, lty = 2, col = "red")
legend("topleft",
       c("fitted line", "confidence intervals", "prediction intervals"),
       lty = c(1, 2, 2), col = c("blue", 1, "red"))
```

Solution: See `http://www.stats.ox.ac.uk/~laws/LMs/OldFaithful.html`