Assignment 6

Due: Tuesday December 4

Note: You will get another assignment on November 27 that will also be due on the 4th

1. Suppose we have data $\boldsymbol{y} = (\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_n})^T \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu}, \Sigma)$. Thus, for each $\boldsymbol{y_i}$ the density of the distribution is given as:

$$f(\boldsymbol{y_i}) = (\sqrt{2\pi})^{-p}|\Sigma|^{-1/2}exp(-\frac{1}{2}(\boldsymbol{y_i} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{y_i} - \boldsymbol{\mu})).$$

We will work out the likelihood ratio test for testing:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$$
$$H_0 : \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$$

This will lead to a test statistic:

$$\Lambda = \frac{\max\ L(\Sigma|\boldsymbol{y})}{\max\ L(\mu, \Sigma|\boldsymbol{y})}.$$

(a) Write out the likelihood function under the alternative hypothesis (unconstrained).

(b) What are the maximum likelihood estimates for μ and Σ under the alternative hypothesis?

(c) Write out the likelihood function under the null hypothesis (constrained).

(d) What is the maximum likelihood estimate for Σ under the null hypothesis?

(e) Write out Λ and simplify if you can?

(f) Assuming a large sample, what is the asymptotic distribution of $-2ln\Lambda$?

(g) Using the bird data available on the website test whether $\boldsymbol{\mu'} = (200, 250)$.

2. Suppose we have data $\boldsymbol{y_1} = (\boldsymbol{y_{1,1}}, \boldsymbol{y_{1,2}}, \ldots, \boldsymbol{y_{1,n_1}})^T \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu_1}, \Sigma_1)$ and $\boldsymbol{y_2} = (\boldsymbol{y_{2,1}}, \boldsymbol{y_{2,2}}, \ldots, \boldsymbol{y_{2,n_2}})^T \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu_2}, \Sigma_2)$. Thus, for the two groups $k = \{1, 2\}$, for each $\boldsymbol{y_{k,i}}$ the density of the distribution is given as:

$$f(\boldsymbol{y_{k,i}}) = (\sqrt{2\pi})^{-p}|\Sigma_k|^{-1/2}exp(-\frac{1}{2}(\boldsymbol{y_{k,i}} - \boldsymbol{\mu_k})^T\Sigma_k^{-1}(\boldsymbol{y_{k,i}} - \boldsymbol{\mu_k})).$$

We will work out the likelihood ratio test for testing:

$$\begin{aligned} H_0 : \Sigma_1 &= \Sigma_2 = \Sigma \\ H_0 : \Sigma_1 &\neq \Sigma_2 \end{aligned}$$

This will lead to a test statistic:

$$\Lambda = \frac{\max\ L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma|\boldsymbol{y}_1, \boldsymbol{y}_2)}{\max\ L(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1, \Sigma_2|\boldsymbol{y}_1, \boldsymbol{y}_2)}.$$

(a) Write out the likelihood function under the alternative hypothesis (unconstrained).

(b) What are the maximum likelihood estimates for $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma_1$, and Σ_2 under the alternative hypothesis?

(c) Write out the likelihood function under the null hypothesis (constrained).

(d) What is the maximum likelihood estimate for Σ under the null hypothesis?

(e) Write out Λ and simplify if you can?

(f) Assuming a large sample, what is the asymptotic distribution of $-2ln\Lambda$?

(g) Using the milk transportation costs data available on the website test whether $\Sigma_{gasoline} = \Sigma_{diesal}$.

3. Jolicoeur and Mosimann studies the relationship of size and shape for painted turtles. The turtle data is on the website and consists of measurements (length, width, height) made on 24 male turtles and 24 female turtles. Assume $\boldsymbol{y_f} = (\boldsymbol{y_{f,1}}, \boldsymbol{y_{f,2}}, \ldots, \boldsymbol{y_{f,24}})^T \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu_1}, \Sigma)$ and $\boldsymbol{y_m} = (\boldsymbol{y_{m,1}}, \boldsymbol{y_{m,2}}, \ldots, \boldsymbol{y_{m,24}})^T \overset{\text{iid}}{\sim} MVN_p(\boldsymbol{\mu_2}, \Sigma)$. Examine the following two test of hypotheses (hint: it would be a good idea to examine result 6.2 in section 6.3):

(a) Test 1:

$$\begin{aligned} H_0 : \boldsymbol{\mu_f} &= \boldsymbol{\mu_m} \\ H_0 : \boldsymbol{\mu_f} &\neq \boldsymbol{\mu_m} \end{aligned}$$

i. What $\boldsymbol{C}$ would I need to make exactly the same test as Test 1. (hint: What matrix $\boldsymbol{Cx} = \boldsymbol{x}$?):

$$\begin{aligned} H_0 : \boldsymbol{C\mu_f} &= \boldsymbol{C\mu_m} \\ H_0 : \boldsymbol{C\mu_f} &\neq \boldsymbol{C\mu_m} \end{aligned}$$

ii. What are the number of linearly independent rows of $\boldsymbol{C}$?

iii. (True/False) The number of linearly independent rows of $\boldsymbol{C}$ is the number of parameters p that we use in the calculation of the critical value:

$$\frac{(n_f + n_m - 2)p}{(n_f + n_m - p - 1)} F(\alpha)_{p, n_f + n_m - p - 1}.$$

(b) Suppose an investigator wanted to test that the $\mu_{length} + \mu_{width} + \mu_{height}$ is different for male and female turtles. Thus the test is:

$$\begin{aligned} H_0 : (\mu_{length} + \mu_{width} + \mu_{height})_f &= (\mu_{length} + \mu_{width} + \mu_{height})_m \\ H_0 : (\mu_{length} + \mu_{width} + \mu_{height})_f &\neq (\mu_{length} + \mu_{width} + \mu_{height})_m \end{aligned}$$

i. What is the $\boldsymbol{C}$ for this test such that:

$$\begin{aligned} H_0 : \boldsymbol{C\mu_f} &= \boldsymbol{C\mu_m} \\ H_0 : \boldsymbol{C\mu_f} &\neq \boldsymbol{C\mu_m} \end{aligned}$$

ii. What is the distribution of $\boldsymbol{Cy_{f,i}}$ (just analytically present the result)?

iii. What is the distribution of $\boldsymbol{Cy_{m,i}}$ (just analytically present the result)?

iv. What is p for this test?

v. Conduct the test and make a decision at the $\alpha = 0.05$ level.

4. For this question, use the biting-fly data which is on the website and assume that $\Sigma_1 \neq \Sigma_2$. Just use the large sample results on pages 291-294.

   (a) Construct a large sample test at the $\alpha = 0.05$ level for the following:

$$\begin{aligned} H_0 : \boldsymbol{\mu_1} &= \boldsymbol{\mu_2} \\ H_0 : \boldsymbol{\mu_1} &\neq \boldsymbol{\mu_2} \end{aligned}$$

   (b) Construct the 95% simultaneous large sample confidence interval for the means.