3 Sampling Distributions for Hypothesis Testing

3.1 Introduction

We seek procedures for testing parameters and constructing confidence (region) estimates for unknown population parameters, such as the mean, $\mu$. These procedures will require distributional assumptions to be made, and so appropriate multivariate distributions need to be considered. In particular, we consider multi-dimensional analogues of the Normal, Chi-square, and Student-t distributions, these being the Multivariate Normal, Wishart, and Hotelling-$T^2$ distributions.

As a motivating example, consider the household spending data:

<table>
<thead>
<tr>
<th>Groceries (£)</th>
<th>Leisure (£)</th>
<th>Income (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>227.01</td>
<td>96.98</td>
<td>741.29</td>
</tr>
<tr>
<td>241.42</td>
<td>140.44</td>
<td>854.07</td>
</tr>
<tr>
<td>188.08</td>
<td>85.13</td>
<td>812.07</td>
</tr>
<tr>
<td>238.23</td>
<td>158.22</td>
<td>813.69</td>
</tr>
<tr>
<td>235.86</td>
<td>103.06</td>
<td>731.42</td>
</tr>
</tbody>
</table>

**Question:** Is there any evidence to reject the hypothesis that the mean amounts spent on groceries, leisure, and income are £180, £113, and £750, respectively in households across the population?

3.2 Multivariate Normal Distribution

The Normal distribution is of fundamental importance in univariate sampling theory.

To recall, suppose that $X \sim N(\mu, \sigma^2)$. Then $X$ has a p.d.f. given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}.$$

A generalization of this distribution for a $p \times 1$ random vector $X$ is:

**Definition 3.1 (MVN distribution)**

A $p \times 1$ random vector $X$ is said to have a multivariate normal (MVN) distribution if its joint p.d.f. is given by

$$f_X(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu) \right\}$$

(1)
for $\mathbf{x} \in \mathbb{R}^p$,

where $\Sigma$ is a $p \times p$, symmetric, positive-definite, matrix.

Write $\mathbf{X} \sim N_p(\mu, \Sigma)$.

**Proposition 3.2 (Mean, Variance of MVN)**
Suppose that $\mathbf{X} \sim N_p(\mu, \Sigma)$. Then
(i) $E[\mathbf{X}] = \mu$
(ii) $\text{var}(\mathbf{X}) = \Sigma$.

**Remarks 3.3**
Suppose that $\mathbf{X} \sim N_p(\mu, \Sigma)$ with $p = 1$. Then, in this case, $\Sigma = \sigma_{11} = \sigma_1^2$, and so

$$f_{\mathbf{X}}(\mathbf{x}) = f_{X_1}(x_1) = \frac{1}{(2\pi)^{1/2}\sigma_1^{1/2}} \exp \left\{ -\frac{1}{2\sigma_1^2}(x_1 - \mu_1)^2 \right\}$$

for $x_1 \in \mathbb{R}$ i.e. $X_1 \sim N(\mu_1, \sigma_1^2)$. Thus, the MVN really is a generalization of the univariate Normal.

### 3.3 Bivariate Normal Distribution

![Figure 1: Bivariate Normal density](image)

The Bivariate Normal distribution is just the MVN for $p = 2$. So, here, $\mu = (\mu_1, \mu_2)'$, and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{bmatrix}$$
where \( \rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2} \) (the correlation coefficient between \( X_1 \) and \( X_2 \)).

It can be shown that

\[
f_{(X_1,X_2)}(x_1,x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-\rho^2}} \times \exp \left\{ - \frac{1}{2(1-\rho^2)} \left[ \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) \right] \right\} \tag{2}
\]

for \((x_1,x_2) \in \mathbb{R}^2\) and provided that \( |\rho| < 1 \).

Remarks 3.4 (Comments on the Bivariate Normal)

(i) The p.d.f. of (2) is specified by 5 parameters, \( \mu_1, \mu_2, \sigma_1, \sigma_2 \) and \( \rho \).

(ii) \( x_1, x_2 \) only appear in the argument of the \( \exp(\cdot) \) function. So the contour lines of \( f_{(X_1,X_2)}(\cdot,\cdot) \) are given by

\[
\left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left( \frac{x_1 - \mu_1}{\sigma_1} \right) \left( \frac{x_2 - \mu_2}{\sigma_2} \right) = k > 0.
\]

These are ellipse equations.

If \( \rho < 0 \), then the major axis has negative slope, and for \( \rho > 0 \), a positive slope; e.g. for \( \Sigma = \begin{bmatrix} 1.5 & -1 \\ -1 & 2.5 \end{bmatrix}, \rho = -1/\sqrt{1.5 \times 2.5} < 0 \), and \( \Sigma = \begin{bmatrix} 1.5 & 1 \\ 1 & 2.5 \end{bmatrix}, \rho = 1/\sqrt{1.5 \times 2.5} > 0 \).

(iii) \( \Sigma \) is positive definite if, and only if, \( |\rho| < 1 \). If \( \rho = 1 \), then rows (or columns) of \( \Sigma \) are no longer linearly independent.
(iv) For this distribution, it is the case that $\rho = 0$ implies that $X_1$ and $X_2$ are independent, since

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2(1 - \rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2}\right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1}\right) \left(\frac{x_2 - \mu_2}{\sigma_2}\right)\right]\right\}$$

with $R_{X_1} = \mathbb{R}$ and $R_{X_2} = \mathbb{R}$, i.e. $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$.

### 3.4 More on the MVN distribution

Some notes on transformations.

**Proposition 3.5**
Suppose $X \sim N_p(\mu, \Sigma)$ and $Y = A'X$, where $A$ is a $p \times m$ matrix of constants. Then

$$Y \sim N_m(A'\mu, A'\Sigma A).$$

**Theorem 3.6**
$X \sim N_p(\mu, \Sigma)$, where rank($\Sigma$) = $p$, if, and only if, $X = \mu + BU$,

where $U \sim N_p(0, I)$, $BB' = \Sigma$, and $B$ is a $p \times p$ constant matrix of full rank $p$.

**Proposition 3.7**
Suppose $X \sim N_p(\mu, \Sigma)$. If $\Sigma$ is of full rank $p$, and $B$ is a $p \times p$ constant matrix of full rank $p$ such that $BB' = \Sigma$, then

$$U \sim B^{-1}(X - \mu) \sim N_p(0, I).$$

**Proposition 3.8** (Distribution of sample mean vector)
Suppose $X_r \sim N_p(\mu, \Sigma)$, $r = 1, \ldots, n$, and mutually independent. Then

$$\bar{X} := \frac{1}{n} \sum_{r=1}^{n} X_r \sim N_p\left(\mu, \frac{1}{n} \Sigma\right).$$

**Proposition 3.9**
Suppose $X \sim N_p(\mu, \Sigma)$ (where $\Sigma$ is of full rank $p$ so that $\Sigma^{-1}$ exists). Then

$$(X - \mu)'\Sigma^{-1}(X - \mu) \sim \chi^2_p.$$
3.5 Wishart Distribution

The Wishart distribution is a generalization of the $\chi^2$-distribution for random vectors.

**Definition 3.10 (Wishart distribution)**

Suppose $X_r \sim N_p(\mu_r, \Sigma)$, $r = 1, \ldots, f$, and mutually independent. Then the matrix

$$W = \sum_{r=1}^{f} X_r X'_r$$

is said to have a **Wishart** distribution on $f$ degrees of freedom.

Distinguish between 2 cases:

(i) If $\mu_r = 0$ for all $r = 1, \ldots, f$, then distribution is **central**, and we write $W \sim W_p(f, \Sigma)$ (with $M$ nominally equal to a $f \times p$ zero matrix).

(ii) Otherwise, we say that $W$ is **non-central**, and write $W \sim W_p(f, \Sigma; M)$, where $M = [\mu_1, \ldots, \mu_f]'$.

**Remarks 3.11**

(i) For $p = 1$, $\mu_r = 0$, $r = 1, \ldots, f$, $\Sigma = \sigma^2$, say, then $X_r \sim N(0, \sigma^2)$, and so

$$W = \sum_{r=1}^{f} X_r^2 \sim W_1(f, \sigma^2).$$

On the other hand, $\frac{X_r}{\sigma} \sim N(0, 1)$, which implies that

$$\frac{1}{\sigma^2} \sum_{r=1}^{f} X_r^2 \sim \chi^2_f \Rightarrow W = \sum_{r=1}^{f} X_r^2 \sim \sigma^2 \chi^2_f.$$

(ii)

$$E[W] = f \Sigma + M'M$$

**Proof**

From Remarks 1.16 (v), it can be deduced that

$$E[X_rX'_r] = \Sigma + \mu_r\mu'_r.$$ 

Hence

$$E[W] = \sum_{r=1}^{f} E[X_rX'_r] = \sum_{r=1}^{f} \{\Sigma + \mu_r\mu'_r\}$$

$$= f \Sigma + \sum_{r=1}^{f} \mu_r\mu'_r = f \Sigma + M'M \quad \square$$
(iii) Suppose \( W_1 \sim W_p(f_1, \Sigma; M_1) \) and \( W_2 \sim W_p(f_2, \Sigma; M_2) \) independently. Then

\[
W_1 + W_2 \sim W_p(f_1 + f_2, \Sigma; M)
\]

where \( M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} \).

(iv) Suppose \( W \sim W_p(f, \Sigma; M) \) and \( C \) is a \( p \times q \) matrix of constants. Then

\[
C'WC \sim W_q(f, C'\Sigma C; MC).
\]

### 3.6 Hotelling \( T^2 \)-distribution

This is a multivariate extension of the Student-\( t \) distribution.

**Definition 3.12 (Hotelling \( T^2 \))**

If \( Y \sim N_p(0, I) \) and \( W \sim W_p(f, I) \) independently, then

\[
\alpha = fY'W^{-1}Y
\]

is said to have a Hotelling \( T^2 \)-distribution.

Write \( \alpha \sim T^2_p(f) \).

The next proposition is particularly useful for deriving various tests for the population mean in the case where the population covariance matrix is unknown.

**Proposition 3.13 (Useful result for hypothesis testing)**

Suppose that \( Y \sim N_q(\mu_0, \frac{1}{f} \Sigma) \) with \( f > q - 1 \), and \( fS \sim W_q(f, \Sigma) \) independently. Then

(i) \( T^2 := k(Y - \mu_0)'S^{-1}(Y - \mu_0) \sim T^2_q(f) \)

(ii) \( \left( \frac{f - q + 1}{fq} \right) T^2 \sim F_{q,f-q+1} \).
3.7 Application: Test for the mean

Suppose that we have a random sample $X_r \sim N_p(\mu, \Sigma)$, $r=1, \ldots, n$, where both $\mu$ and $\Sigma$ are unknown.

We seek a test for $H_0 : \mu = \mu_0$

vs.

$H_1 : \mu \neq \mu_0$.

To this end, consider the statistic

$$T^2 = n(\overline{X} - \mu_0)'S^{-1}(\overline{X} - \mu_0).$$

Since $\overline{X}$ is an apparently ’good’ estimator of $\mu$, then if $H_0$ is true, $\overline{X}$ should be close to $\mu_0$, and so $T^2$ should be small.

On the other hand, if $H_0$ is not true, then $\overline{X}$ should be far from $\mu_0$, and so $T^2$ should be large.

So perhaps a reasonable decision procedure to adopt would be to:

Reject $H_0$ if $T^2 > k'$

Accept $H_0$ if $T^2 \leq k'$ for some constant $k' > 0$.

But how do we choose $k'$?

It can be shown that $\overline{X}$ and $S$ are independent, and that

$$\overline{X} \sim N_p(\mu, \frac{1}{n}\Sigma) \quad (n-1)S \sim W_p(n-1, \Sigma).$$

Thus, it follows from the previous proposition that under $H_0$, and for $n > p$,

$$T^2 = n(\overline{X} - \mu_0)'S^{-1}(\overline{X} - \mu_0) \sim T^2_p(n-1)$$

and

$$\frac{(n-p)}{p(n-1)}T^2 \sim F_{p,n-p}.$$

Details of the test are determined as follows:

$$\alpha = \mathbb{P}(\text{Reject } H_0|H_0 \text{ true}) = \mathbb{P}(T^2 > k'|H_0 \text{ true})$$

$$= \mathbb{P}\left(\frac{(n-p)}{p(n-1)}T^2 > k'|H_0 \text{ true}\right)$$

so we should select $k' = F_{p,n-p,\alpha}$, the upper $100\alpha\%$ point of the $F$-distribution with $p$ and $n-p$ degrees of freedom.

Thus we reject $H_0$ (in favour of $H_1$) if

$$\frac{(n-p)T^2_{\text{obs}}}{p(n-1)} > F_{p,n-p,\alpha}$$

7
i.e.

\[ T_{obs}^2 > \frac{p(n - 1)}{n - p} F_{p, n-p, \alpha} \]

and accept \( H_0 \) otherwise.

```r
> groceries <- c(227.01, 241.42, 188.08, 238.23, 235.86)
> leisure <- c(96.98, 140.44, 85.13, 158.22, 103.06)
> income <- c(741.29, 854.07, 812.07, 813.69, 731.42)
> spend <- data.frame(groceries, leisure, income)
> S <- var(spend)
> S.inv <- solve(S)
> n <- dim(spend)[1]
> p <- dim(spend)[2]
> n
[1] 5
> p
[1] 3
> m.spend <- apply(spend, 2, mean)
> m.spend
   groceries leisure income
      226.12  116.766  790.508
> mu0 <- c(180, 113, 750)
> mu0
[1] 180 113 750
> T2 <- n * t(m.spend - mu0) %*% S.inv %*% (m.spend - mu0)
> T2
[,1]
[1,] 126.1583
> T2 <- drop(T2)
> T2
[1] 126.1583
> F.obs <- ((n - p) * T2)/(p * (n - 1))
> F.obs
[1] 21.02638
> p.value <- 1 - pf(F.obs, p, n - p)
> p.value
[1] 0.04574171
```

**Conclusion:** There is evidence to reject the hypothesis at the 5% level of significance.