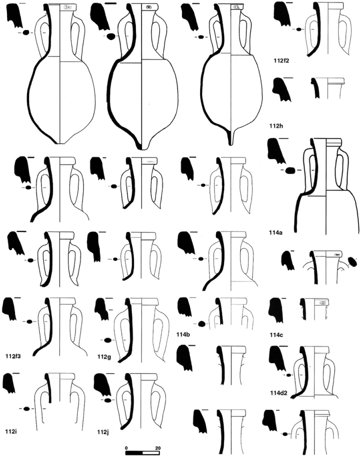
Summary statistics

A dataset

Measures of amphorae from Baetica
`https://github.comMcotsar`
`LearningBaetica/blob/master/Data/dataDressel.csv`



A dataset

We select only the Dressel A from Delicias, the four variates are exterior diameter, interior diameter, rim height, and rim width.

$$X = \begin{pmatrix} 150 & 120 & 44 & 24 \\ 140 & 102 & 46 & 30 \\ 150 & 110 & 45 & 25 \\ 130 & 86 & 44 & 27 \\ 150 & 88 & 30 & 32 \\ 150 & 110 & 44 & 23 \\ 165 & 100 & 41 & 27 \\ 150 & 100 & 53 & 21 \\ 140 & 105 & 56 & 25 \end{pmatrix}$$

Descriptive statistics

The mean vector is:

$$\bar{x} = (147.2,\ 102.3,\ 44.8,\ 26)^\top$$

With matrix notations:

$$\bar{x} = X^\top \mathbf{1}_n / n$$

This corresponds to the mean of the observed population.

Sample covariance matrix

The covariance matrix is the matrix with elements:

$$s_{jj} = s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$$

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

That is exactly the definitions in 1D for all pairs of variates.
With Matrix Notation:

$$S_u = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^\top = \frac{1}{n-1} \left(\sum_{i=1}^{n} x_i x_i^\top - n\bar{x}\bar{x}^\top \right)$$

Sample covariance matrix

- The first term is the *raw sums of squares and products* (SSP):

$$\sum_i x_i x_i^\top = X^\top X = \left[\sum_i x_{ir} x_{is}\right]_{r,s}.$$

For the second term:

$$\bar{x} = X^\top \mathbf{1}_n / n,$$

so that we can write:

$$n\bar{x}\bar{x}^\top = n(X^\top \mathbf{1}_n/n)(X^\top \mathbf{1}_n/n)^\top = X^\top \mathbf{1}_n \mathbf{1}_n^\top X$$

Sample covariance matrix

$$S_u = \frac{1}{n-1}(X^\top X - X^\top(\mathbf{1}_n\mathbf{1}_n^\top)X) = \frac{1}{n-1}X^\top(I_n - \mathbf{1}_n\mathbf{1}_n^\top/n)X$$

Which allows us to define the *corrected SSP* matrix:

$$A = X^\top H X,$$

where H is the *centering matrix*: $H = (I_n - \mathbf{1}_n\mathbf{1}_n^\top/n)$

Mean Squares and Products

The mean squares and products is then:

$$S = A/n.$$

The sample covariance matrix was:

$$S_u = \frac{A}{n-1}$$

so that we have:

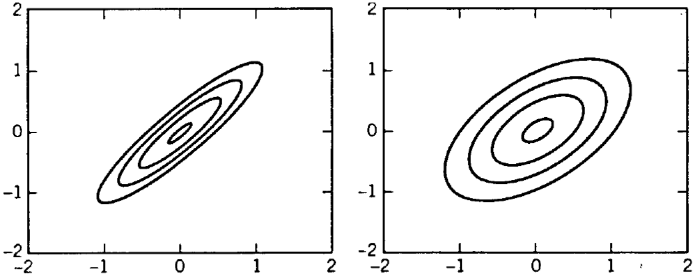
$$S_u = \frac{nS}{n-1}$$

Link with dimension 1?

Descriptive?

What can be said of these quantities ? What do they represent ?

S represents how the data is spread. An interesting quantity:
$|S|$ represents the "volume" of the data (c.f. Jacobian).



Amphorae Dataset

What code to use in `R` ?

$$S_u = \begin{pmatrix} 94.4 & 33.5 & -20.7 & -4.38 \\ 33.5 & 115 & 28.2 & -20.4 \\ -20.7 & 28.2 & 53.7 & -17.1 \\ -4.38 & -20.4 & -17.1 & 11.8 \end{pmatrix}$$

We can then compute the determinant of this matrix, plot ellipsoid based on it, etc.

Correlation matrix

Reminder: the correlation between two variables X and Y is

$$r_{XY} = \frac{s_{XY}}{\sqrt{s_{XX}s_{YY}}}.$$

This is for a random variable. In our case, we want to compute the correlation between two *samples*.

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}.$$

That is with a matrix notation:

$$R = D_s^{-1/2} S_u D_s^{-1/2}$$

Amphorae dataset

$$R = \begin{pmatrix} 1 & 0.32 & -0.29 & -0.13 \\ 0.32 & 1 & 0.36 & -0.55 \\ -0.29 & 0.36 & 1 & -0.68 \\ -0.13 & -0.55 & -0.68 & 1 \end{pmatrix}.$$

What can you say about this matrix ?

What are these quantities?

Covariance, mean, correlation?

Under which condition on the dataset these statistics have meaning?

What can you say if the dataset comes from a discrete distribution? From a normal distribution? From a distribution that has no mean?

Inferential statistics

Estimation in Multivariate normal models

- Estimation is more complex: In dimension p, $m \in \mathbb{R}^p$, p variances to estimate, $\binom{p}{2}$ covariances.
  $\Rightarrow$ without assumptions of independence it becomes quickly computationally infeasible.
- testing is more complex:
  - We can use p tests (one per dimension) but in this case it inflates Type 1 error, and reject if at least one should be rejected.
  - a Mutlivariate test is more adapted (fits the geometry of the problem)

Multivariate Gaussian Model

$$\mathcal{M}_{p,n}(\mathbb{R}) \ni X \sim \mathcal{N}(\mu, \Sigma)^{\otimes n}$$

where $X = [x_i]_i,\ x_i \in \mathcal{M}_{p,1}(\mathbb{R})$.
i.e. iid observations of a p dimensional Gaussian.
The model writes:

$$\{\mathcal{N}(\mu, \Sigma)^{\otimes n} \mid \mu \in \mathbb{R}^p, \Sigma \in \mathcal{M}_p(\mathbb{R})\}$$

How to estimate the value of the true parameters? Several approaches:

- Moment method (link with descriptive statistics);
- maximum likelihood;
- Bayesian methods [not studied here, but very simple in this case].

In our case, both Moment method and ML give the same result.

Maximum likelihood estimation

Observations x_i are independent:

$$
\begin{aligned}
&\mathcal{L}(x_1, \dots, x_n; \mu, \Sigma) \\
&= \prod_i f(x_i; \mu, \Sigma) \\
&= \prod_i \frac{1}{(\sqrt{2\pi})^2 |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu)^\top \Sigma^{-1}(x_i - \mu)\right) \\
&= \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} \exp\left(-\frac{1}{2}\sum_i (x_i - \mu)^\top \Sigma^{-1}(x_i - \mu)\right).
\end{aligned}
$$

We write $A = \sum_i (x_i - \bar{x})(x_i - \bar{x})^\top$
Indication: $A \in \mathcal{M}_n(\mathbb{R}),\ \frac{\partial}{\partial A} \log |A| = (A^{-1})^\top$.
Clue: $x = Tr(x),\ x \in \mathbb{R}$.

Maximum likelihood estimation

Maximising this likelihood leads to:

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\Sigma} = \frac{1}{n} \sum_i (x_i - \bar{x})(x_i - \bar{x})^\top = S. \end{cases}$$

$\Rightarrow A$ needs to be positive definite, it is possible to have degenerate cases in which the data lies in a subspace of measure 0.

Dykstra's Theorem

If $x_1, \ldots, x_n \sim \mathcal{N}(\mu, \Sigma),\ \Sigma > 0$, then:

$$A \text{ is positive definite} \Leftrightarrow n > p$$

- The covariance estimator is biased. There exists an unbiased version of the estimator $S_u = \frac{n}{n-1} S$.

Invariance principle of MLE

Zehna's Theorem

Let $x_1, \ldots, x_n$ be a random sample from a distribution $F(x; \theta_1, \ldots, \theta_n)$, and $\hat{\theta_1}, \ldots, \hat{\theta_n}$ the corresponding MLE. Then for ϕ any bijective function (bijection, one-to-one), the MLE for $\phi(\theta_1, \ldots, \theta_n)$ is $\phi(\hat{\theta_1}, \ldots, \hat{\theta_n})$.

A one page article [Zehna, 1966], you can have a look.

Distributions of empirical mean and covar

As seen previously: $\bar{x} = (1/n, \dots, 1/n) \cdot x$ thus:

$$\bar{X} \sim \mathcal{N}(\mu, \Sigma/n)$$

(Strong) Law of Large Numbers

If the x_i are independent,

$$\bar{x} \Rightarrow^{a.s.}_{n \to \infty} \mathbb{E}[X].$$

If the x_i are non-correlated,

$$\bar{x} \Rightarrow^{\mathbb{P}}_{n \to \infty} \mathbb{E}[X].$$

Multivariate Central Limit Theorem

Let $(x_i)_i$ be an iid sample from $\mathcal{N}(\mu, \Sigma)$, then:

$$\sqrt{n}(\bar{x} - \mu) \Rightarrow^{\mathcal{L}}_{n \to \infty} \mathcal{N}(0, \Sigma).$$

This is exactly the same as in 1-D, with matrices and vectors instead of reals.

Question

You use Monte Carlo methods to estimate a complex integral:

$$z = \int_{[0,1]^d} f(x)\mathrm{d}x,$$

where $= f : \mathbb{R}^d \to \mathbb{R}$.

1. How do you proceed?
2. what is the convergence speed of the estimator you got?
3. any ideas on how to improve the convergence of the method?
4. how to do if you integrate over $A \subset \mathbb{R}^d$?
5. how does the convergence change when d increase?

Multivariate Fisher's Theorem

Theorem

Let (x_i) be an iid sample from $\mathcal{N}(\mu, \Sigma)$. Then:
The distribution of $\bar{x}$ and A (the SSP matrix) are independent and,

$$A \sim \sum_{i=1}^{n-1} z_i z_i^\top,$$

where (z_i) are iid $\mathcal{N}_p(0_p, \Sigma)$.

Wishart's distribution

What is the distribution of $\sum_{i=1}^{n-1} z_i z_i^\top$?
Let $Z = [z_i]_i$ then $Z^\top Z = \sum_{i=1}^{n-1} z_i z_i^\top$, is said to follow the Wishart distribution $\mathcal{W}_p(n, \Sigma)$ or $\mathcal{W}_p(\Sigma, n)$, with n degrees of freedom and parameters p and Σ.

- There are several results on this distribution that appears on plenty of problems.
- It is analogous to the χ^2 distribution in dimension 1: for $p = 1$, $\mathcal{W}_1(n, 1) = \chi^2(n)$.

Wishart's distribution

- For $(x_i)_{i=1}^n$ sample from $\mathcal{N}(\mu, \Sigma)$:

$$nS = A = X^\top HX \sim \mathcal{W}_p(n-1, \Sigma);$$

- if $A \sim \mathcal{W}_p(n, \Sigma)$, then $\mathbb{E}[A] = n\Sigma$;
- if $A_1, \dots, A_k$ independent with distribution $W_p(n_1, \Sigma), \dots, W_p(n_k, \Sigma)$, then:

$$A_1 + \cdots + A_k \sim W_p(n_1, \dots, n_k, \Sigma);$$

- if $A \sim \mathcal{W}_p(n, \Sigma)$ and $C \in \mathcal{M}_{p\times q}(\mathbb{R})$, then:

$$C^\top AC \sim \mathcal{W}_q(n, C^\top \Sigma C).$$

  Special case if $C = (1, \dots, 1)^\top$:

$$C^\top AC \sim \mathcal{W}_1(n, C^\top \Sigma C) = C^\top \Sigma C \chi_n^2$$

Hotelling's T^2 distribution

Analogous to Student's t-distribution.
Reminder: in 1-D, $\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim \mathcal{T}(k)$

If $z \sim \mathcal{N}_p(0, I)$, and $\mathcal{M}_{p \times p} \ni A \sim \mathcal{W}_p(n, I)$ are independent, then,

$$t^2 = n z^\top A^{-1} z$$

follows Hotelling's T^2 distribution with parameters p and n: $T^2_{p,n}$.
Note, the same result holds if $z \sim \mathcal{N}_p(0, \Sigma)$ and $A \sim \mathcal{W}_p(n, \Sigma)$.

Generalised Multivariate Normal Results

Theorem 1

If $z \sim \mathcal{N}_p(\mu, \Sigma)$ is independent from $A \sim \mathcal{W}_p(n, \Sigma)$, then:

$$t^2 = n(x - \mu)^\top A^{-1}(x - \mu) \sim T^2_{p,n}.$$

Corollary 2

If $\bar{x}$ is the estimated mean, and S the estimated covariance, then:

$$t^2 = (n-1)(\bar{x} - \mu)^\top S^{-1}(\bar{x} - \mu) \sim T^2_{p,n-1}.$$

With $S_u = nS/(n-1)$, we have:

$$t^2 = n(\bar{x} - \mu)^\top S_u^{-1}(\bar{x} - \mu) \sim T^2_{p,n-1}.$$

Hotelling's T^2 and F-distribution

Hotelling's T^2 is not a standard distribution, but it is linked with the F-distribution:

$$\frac{n-p+1}{np} t^2 \sim F_{p,n-p+1}.$$

Equivalently

$$T^2_{p,n} = \frac{np}{n-p+1} F_{p,n-p+1}.$$

Example 3

In dimension 1 ($p = 1$):

$$\left(\frac{\bar{x}-\mu}{\sqrt{S_u}/\sqrt{n}}\right)^2 \sim T^2_{1,n-1} = F_{1,n-1} = t^2_{n-1}.$$

Linear transformations

Corollary 4

Let $x \sim \mathcal{N}_p(\mu, \Sigma)$, and $y = Cx$, where C is $q \times p$, with $q \leq p$. If $\bar{x}$ and S_x are the estimated mean and covariance, then:

$$
\begin{aligned}
\bar{y} &= C\bar{x} = \mathcal{N}_q(C\mu, C\Sigma C^\top / n) \\
nS_y &= nCS_xC^\top \sim \mathcal{W}_q(n-1, C\Sigma C^\top) \\
&(n-1)(C\bar{x} - C\mu)^\top (CS_xC^\top)^{-1}(C\bar{x} - C\mu) \sim T^2_{q,n-1}.
\end{aligned}
$$

Example of use: test

Let x_i be an iid sample from $\mathcal{N}_p(\mu, \Sigma)$. We write $\bar{x}$ and S_u the sample mean and covariance.

- Hypotheses, for a μ_0 fixed and specified:

$$
\begin{aligned}
H_0 :& \mu = \mu_0 \\
H_1 :& \mu \neq \mu_0
\end{aligned}
$$

- we compute the test statistic:

$$
t^2 = n(\bar{x} - \mu)^\top S_u^{-1}(\bar{x} - \mu) \sim T^2_{p,n-1},
$$

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

Example of use: test

- For a specified α (typically 0.05), we define a region of μ_0 in which under H_0 the data observed would be generated with a probability $1 - \alpha$. It has the form:

$$n(\bar{x} - \mu)^\top S_u^{-1}(\bar{x} - \mu) \leq \frac{(n-1)p}{n-p} F_{p,n-p;\alpha}.$$

  This defines an ellipsoid in the space of μ's centered on $\bar{x}$.

- outside of this region, the probability of observing the data is deemed too small and H_0 is rejected.

Similar formulae can be written for S the biased estimator of the covariance.

Example of one sided test

Ex. 3 of the Problem Sheet:
Let $x \sim \mathcal{N}_p(\mu, \Sigma)$, with μ and Σ unknown, let's assume that:

$$\bar{x} = \begin{pmatrix} 1 \\ 1/2 \end{pmatrix}, S = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \ n = 6.$$

Solve the test problem: $H_0 : \mu = (2, 2/3)^\top$ vs $H_1 : \mu \neq (2, 2/3)^\top$ for $\alpha = 0.05$.

The test statistic writes:

$$\frac{n-p}{p}(\bar{x} - \mu_0)^\top S^{-1}(\bar{x} - \mu_0)$$

$$= 2\,(-1, -1/6) \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} -1 \\ -1/6 \end{pmatrix}$$

$$= 43/27 \simeq 1.6$$

Example of one sided test

We have to compare this value with $F_{n,n-p;\alpha} = 6.944$ (check function `qf` in `R`).
As our statistic is smaller than this value:
We cannot reject H_0 we cannot reject H_0.

Note that we cannot show that H_0 is true under this setting.
Note also that exchanging H_0 and H_1 will not help.

Two sided test

In this case, we have two independent samples:

$$x_1, \ldots, x_n \sim^{iid} \mathcal{N}_p(\mu_1, \Sigma),$$
$$y_1, \ldots, y_m \sim^{iid} \mathcal{N}_p(\mu_2, \Sigma).$$

These samples have the same covariance but a different mean (*a priori*). The estimated means and covariance are $\bar{x}$ and $\bar{y}$, the estimated covariances are S_x and S_y.

We want to test $H_0 : \mu_1 = \mu_2$.

We define $\delta = \mu_1 - \mu_2$, the test hypothesis is then $\delta = 0$.

Two sided test

- Distribution of the difference:

$$(\bar{x} - \bar{y}) \sim \mathcal{N}_p(\delta, (n+m)\Sigma/(nm)).$$

- We know that:

$$nS_x + mS_y \sim \mathcal{W}_p(n+m-2, \Sigma).$$

- We define $S = (nS_x + mS_y)/(n+m)$. Because S_x (resp. S_y) is independent of $\bar{x}$ (resp. $\bar{y}$), we have that S is independent of $\bar{x} - \bar{y}$.

Two sided test

- We compute t^2:

$$t^2 = \frac{nm(n+m-2)}{(n+m)^2}((\bar{x}-\bar{y})-\delta)^\top S^{-1}((\bar{x}-\bar{y})-\delta) \sim T^2_{p,n+m-2}.$$

- Which leads to the test statistic:

$$F = \frac{nm(n+m-p-1)}{p(n+m)^2}((\bar{x}-\bar{y})-\delta)^\top S^{-1}((\bar{x}-\bar{y})-\delta) \\ \sim F_{p,n+m-p-1}.$$

- A confidence region of order $1-\alpha$ is then:

$$\left\{\delta \,\middle|\, ((\bar{x}-\bar{y})-\delta)^\top S^{-1}((\bar{x}-\bar{y})-\delta) \leq \frac{p(n+m)^2}{nm(n+m-p-1)} F_{p,n+m-p-1;\alpha}\right\}.$$

Two sided test

Similarly for a rejection region ($>$ instead of $\leq$).

$$\left\{\delta \,\middle|\, ((\bar{x}-\bar{y})-\delta)^\top S^{-1}((\bar{x}-\bar{y})-\delta) > \frac{p(n+m)^2}{nm(n+m-p-1)} F_{p,n+m-p-1;\alpha}\right\}.$$

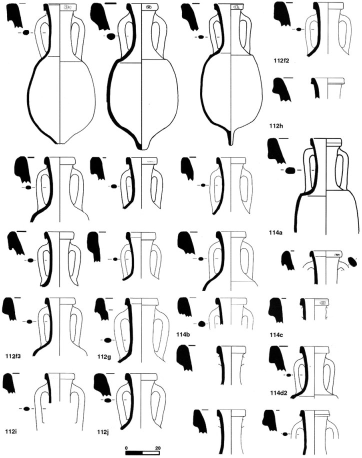
Now it is enough to check if $\delta = 0$ is in the rejection region.

Example of Multidimensional normal regression and tests

Measures of amphorae from Baetica

`https://github.comMcotsar`

`LearningBaetica/blob/master/Data/dataDressel.csv`



Dataset

```
     number        excavation            type            chronology         exterior_diam    inside_diam
 Min.   :  1.00   Length:470         Length:470         Length:470         Min.   :120.0   Min.   : 70.00
 1st Qu.: 24.00   Class :character   Class :character   Class :character   1st Qu.:160.0   1st Qu.: 90.00
 Median : 47.50   Mode  :character   Mode  :character   Mode  :character   Median :170.0   Median : 90.00
 Mean   : 50.55                                                            Mean   :166.9   Mean   : 94.47
 3rd Qu.: 74.00                                                            3rd Qu.:175.0   3rd Qu.:100.00
 Max.   :128.00                                                            Max.   :210.0   Max.   :140.00
     rim_h          rim_w          shape_w        rim_inside_h      rim_w_2      protruding_rim      site
 Min.   :25.00   Min.   :15.00   Min.   : 5.000   Min.   :18.00   Min.   : 7.00   Min.   : 5.00   Length:470
 1st Qu.:34.00   1st Qu.:34.00   1st Qu.: 9.000   1st Qu.:26.00   1st Qu.:28.00   1st Qu.:14.00   Class :character
 Median :35.00   Median :37.00   Median :10.000   Median :28.00   Median :31.00   Median :18.00   Mode  :character
 Mean   :35.39   Mean   :36.29   Mean   : 9.513   Mean   :28.38   Mean   :30.35   Mean   :18.89
 3rd Qu.:37.00   3rd Qu.:40.00   3rd Qu.:10.000   3rd Qu.:30.00   3rd Qu.:34.00   3rd Qu.:23.00
 Max.   :56.00   Max.   :48.00   Max.   :14.000   Max.   :45.00   Max.   :44.00   Max.   :42.00
> |
```

We reduce our study to the Dressel D from Delicias and Malpica.
The measures are in cm without uncertainty, I added a Gaussian noise with sd 1 independent to each observation to ease analysis.

Estimators

```
> colMeans(Delicias)
 exterior_diam    inside_diam          rim_h          rim_w        shape_w   rim_inside_h        rim_w_2 protruding_rim
    170.843955      92.838863      34.348870      37.990818       8.774759      27.403727      31.878133      15.166147
> colMeans(Malpica)
 exterior_diam    inside_diam          rim_h          rim_w        shape_w   rim_inside_h        rim_w_2 protruding_rim
    165.014347      95.203953      35.638047      36.902172       9.389259      28.367357      29.679128      22.441306
> cov(Malpica)
               exterior_diam inside_diam      rim_h      rim_w    shape_w rim_inside_h    rim_w_2 protruding_rim
exterior_diam    102.3574656  65.9881182  0.9216169  9.3842732  1.7286023     2.753100  7.9076654    -17.1853208
inside_diam       65.9881182  95.8744146 -0.7890684  5.5033154  0.3140279     4.865154  6.7865203     -2.4544259
rim_h              0.9216169  -0.7890684  8.6323077  3.1542544  0.7135214     3.934514  1.3098974      7.6117989
rim_w              9.3842732   5.5033154  3.1542544 11.8317916  1.5733241     3.152703  5.2554679     -0.3993738
shape_w            1.7286023   0.3140279  0.7135214  1.5733241  2.9069854     1.896117 -0.1021811      1.6085155
rim_inside_h       2.7530996   4.8651541  3.9345144  3.1527035  1.8961167    10.939538  2.6628437      8.8650598
rim_w_2            7.9076654   6.7865203  1.3098974  5.2554679 -0.1021811     2.662844 11.2686268     -2.6118438
protruding_rim   -17.1853208  -2.4544259  7.6117989 -0.3993738  1.6085155     8.865060 -2.6118438     48.6039097
> cov(Delicias)
               exterior_diam inside_diam       rim_h      rim_w    shape_w rim_inside_h    rim_w_2 protruding_rim
exterior_diam      51.155588   -5.998031 -10.1057172  9.9225463  1.6136606   -5.5584878 10.4480883    -21.6959500
inside_diam        -5.998031   73.139637   4.2297274 -3.2584924 -3.8087616    3.1318831 -6.4447282     13.5464299
rim_h             -10.105717    4.229727   8.9412875 -0.9358570 -0.6890488    3.2393062 -1.7956373      4.2332126
rim_w               9.922546   -3.258492  -0.9358570 12.5092371  0.8515161   -3.0604697  9.2884143     -9.0185737
shape_w             1.613661   -3.808762  -0.6890488  0.8515161  2.5173116   -0.3358484  0.9532381      0.2561426
rim_inside_h       -5.558488    3.131883   3.2393062 -3.0604697 -0.3358484   11.8165644 -0.5243239      3.4597304
rim_w_2            10.448088   -6.444728  -1.7956373  9.2884143  0.9532381   -0.5243239 15.8358794     -9.6494098
protruding_rim    -21.695950   13.546430   4.2332126 -9.0185737  0.2561426    3.4597304 -9.6494098     36.8424190
```

Tests

We want to know if there are significant differences between the amphorae from both sites. We follow the previous strategy with $\delta = 0$.

```
> S = (58*cov(Delicias) + 52*cov(Malpica))/(110)
> F=(52*58*(101))/(8*(110)*110) * (colMeans(Delicias)-colMeans(Malpica))%*%solve(S)%*%(colMeans(Delicias)-colMeans(Malpic
a))
> print(F)
         [,1]
[1,] 5.276236
> print(qf(0.95,8,101))
[1] 2.03138
```

We conclude that the sites provide significantly different Dressel D amphorae.