

# Derivation of a test for equality of means with unequal variances

Brenton Horne

28th September 2021

## Contents

Hypotheses	2
Definitions	2
Derivation of the maximum likelihood under the null	2
Numerical approximation of maximum likelihood estimators . . . . .	4
Derivation of the unrestricted maximum likelihood	7
Likelihood ratio	8

# Hypotheses

Let  $Y_{ij}$  denote the  $j$ th observation of the  $i$ th treatment group. Where  $i = 1, 2, 3, \dots, m$  and  $j = 1, 2, 3, \dots, n_i$ . Under the null hypothesis:  $Y_{ij} \sim N(\mu, \sigma_i^2)$ . Under the alternative hypothesis:  $Y_{ij} \sim N(\mu_i, \sigma_i^2)$ , where  $\mu_i \neq \mu_k$  for at least one pair of  $i$  and  $k$  values.

## Definitions

$$\begin{aligned} n &= \sum_{i=1}^m n_i \\ \bar{Y} &= \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} Y_{ij} \\ \bar{Y}_i &= \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij} \end{aligned}$$

We will later use  $\delta_{ik}$ , which is the Kronecker delta symbol. It equals 0 if  $i \neq k$  and 1 otherwise.

## Derivation of the maximum likelihood under the null

Let  $\Omega_0$  denote the parameter space under the null hypothesis.  $\Omega_0 = \{(\mu, \sigma_i^2) : -\infty < \mu < \infty, \sigma_i^2 > 0\}$ .

$\Omega_a = \{(\mu_i, \sigma_i^2) : \mu_i \neq \mu_j \text{ for at least one pair of } i \text{ and } j \text{ values, } -\infty < \mu_i < \infty, \sigma_i^2 > 0\}$ . The unrestricted parameter space is thus:  $\Omega = \Omega_0 \cup \Omega_a$ .

$$\begin{aligned} L(\Omega_0) &= \prod_{i=1}^m \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(Y_{ij} - \mu)^2\right) \\ &= (2\pi)^{-n/2} \left(\prod_{i=1}^m \sigma_i^{-n_i}\right) \exp\left(-\frac{1}{2} \sum_{i=1}^m \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \mu)^2\right). \end{aligned} \tag{1}$$

Taking the natural logarithm yields:

$$\ln L(\Omega_0) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^m n_i \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^m \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \mu)^2.$$

Differentiating the log-likelihood with respect to  $\mu$  and setting to zero to maximize the likelihood:

$$\begin{aligned}
\frac{\partial \ln L(\Omega_0)}{\partial \mu} \Big|_{\mu=\hat{\mu}, \sigma_i^2=\hat{\sigma}_i^2} &= -\frac{1}{2} \sum_{i=1}^m \frac{1}{\hat{\sigma}_i^2} \sum_{j=1}^{n_i} 2(-1)(Y_{ij} - \hat{\mu}) \\
&= 0 \\
\sum_{i=1}^m \frac{1}{\hat{\sigma}_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}) &= 0 \\
\sum_{i=1}^m \frac{1}{\hat{\sigma}_i^2} (n_i \bar{Y}_i - n_i \hat{\mu}) &= 0 \\
\left( \sum_{i=1}^m \frac{n_i \bar{Y}_i}{\hat{\sigma}_i^2} \right) - \left( \sum_{i=1}^m \frac{n_i}{\hat{\sigma}_i^2} \right) \hat{\mu} &= 0 \\
\hat{\mu} &= \frac{\sum_{i=1}^m \frac{n_i \bar{Y}_i}{\hat{\sigma}_i^2}}{\sum_{i=1}^m \frac{n_i}{\hat{\sigma}_i^2}}. \tag{2}
\end{aligned}$$

Differentiating the log-likelihood with respect to  $\sigma_k^2$  and setting to zero to maximize the likelihood:

$$\begin{aligned}
\frac{\partial \ln L(\Omega_0)}{\partial \sigma_k^2} \Big|_{\mu=\hat{\mu}, \sigma_i^2=\hat{\sigma}_i^2} &= -\frac{1}{2} \sum_{i=1}^m \frac{n_i}{\hat{\sigma}_i^2} \delta_{ik} - \frac{1}{2} \sum_{i=1}^m -\frac{1}{\hat{\sigma}_i^4} \delta_{ik} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \\
&= -\frac{1}{2} \frac{n_k}{\hat{\sigma}_k^2} + \frac{1}{2\hat{\sigma}_k^4} \sum_{j=1}^{n_k} (Y_{kj} - \hat{\mu})^2 \\
&= 0.
\end{aligned}$$

Multiplying by  $2\hat{\sigma}_k^4$  yields:

$$\begin{aligned}
-n_k \hat{\sigma}_k^2 + \sum_{j=1}^{n_k} (Y_{kj} - \hat{\mu})^2 &= 0 \\
\hat{\sigma}_k^2 &= \frac{1}{n_k} \sum_{j=1}^{n_k} (Y_{kj} - \hat{\mu})^2. \tag{3}
\end{aligned}$$

Some simplification of our likelihood can be done now using our maximum likelihood estimators (MLEs), although unfortunately Equations 2 and 3 cannot be analytically solved, they must be numerically solved to yield values for

$\hat{\mu}$  and  $\hat{\sigma}_k^2$ .

$$\begin{aligned}
L(\widehat{\Omega}_0) &= (2\pi)^{-n/2} \left( \prod_{i=1}^m \widehat{\sigma}_i^2^{-n_i/2} \right) \exp \left( -\frac{1}{2} \sum_{i=1}^m \frac{1}{\widehat{\sigma}_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \right) \\
&= (2\pi)^{-n/2} \left( \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \right)^{-n_i/2} \right) \exp \left( -\frac{1}{2} \sum_{i=1}^m \frac{1}{\frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \right) \\
&= (2\pi)^{-n/2} \left( \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \right)^{-n_i/2} \right) \exp \left( -\frac{1}{2} \sum_{i=1}^m n_i \right) \\
&= (2\pi)^{-n/2} \left( \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \right)^{-n_i/2} \right) \exp \left( -\frac{n}{2} \right) \\
&= (2\pi e)^{-n/2} \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2 \right)^{-n_i/2}.
\end{aligned}$$

## Numerical approximation of maximum likelihood estimators

So here we must numerically approximate  $\hat{\mu}$  and  $\hat{\sigma}_i^2$  using a technique like Newton's method and substitute this into our expression for  $L(\Omega_0)$  to get the maximum likelihood under the null. If we use Newton's method, we must have functions we are finding the zeros of. Let:

$$\begin{aligned}
f(\hat{\mu}, \hat{\sigma}_k^2) &= \hat{\mu} - \frac{\sum_{k=1}^m \frac{n_k \bar{Y}_k}{\widehat{\sigma}_k^2}}{\sum_{k=1}^m \frac{n_k}{\widehat{\sigma}_k^2}} \\
g_i(\hat{\mu}, \hat{\sigma}_k^2) &= \hat{\sigma}_i^2 - \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu})^2.
\end{aligned}$$

Therefore:

$$\begin{aligned}
\frac{\partial f}{\partial \widehat{\mu}} &= 1 \\
\frac{\partial f}{\partial \widehat{\sigma}_i^2} &= -\frac{\sum_{k=1}^m -\frac{n_k \overline{Y}_k}{\widehat{\sigma}_k^4} \delta_{ik}}{\sum_{k=1}^m \frac{n_k}{\widehat{\sigma}_k^2}} - \frac{\sum_{k=1}^m \frac{n_k \overline{Y}_k}{\widehat{\sigma}_k^2}}{\left(\sum_{k=1}^m \frac{n_k}{\widehat{\sigma}_k^2}\right)^2} \sum_{k=1}^m -\frac{n_k}{\widehat{\sigma}_k^4} \delta_{ik} \\
&= \frac{\frac{n_i \overline{Y}_i}{\widehat{\sigma}_i^4}}{\sum_{k=1}^m \frac{n_k}{\widehat{\sigma}_k^2}} - \frac{\sum_{k=1}^m \frac{n_k \overline{Y}_k}{\widehat{\sigma}_k^2}}{\left(\sum_{k=1}^m \frac{n_k}{\widehat{\sigma}_k^2}\right)^2} \frac{n_i}{\widehat{\sigma}_i^4} \\
&= \frac{n_i}{\widehat{\sigma}_i^4 \sum_{k=1}^m \frac{n_k}{\widehat{\sigma}_k^2}} \left( \overline{Y}_i - \frac{\sum_{i=1}^m \frac{n_i \overline{Y}_i}{\widehat{\sigma}_i^2}}{\sum_{i=1}^m \frac{n_i}{\widehat{\sigma}_i^2}} \right).
\end{aligned}$$

As for the derivatives of  $g_i(\widehat{\mu}, \widehat{\sigma}_k^2)$ :

$$\begin{aligned}
\frac{\partial g_i(\widehat{\mu}, \widehat{\sigma}_k^2)}{\partial \widehat{\mu}} &= -\frac{1}{n_i} \sum_{j=1}^{n_i} 2(-1)(Y_{ij} - \widehat{\mu}) \\
&= \frac{2}{n_i} (n_i \overline{Y}_i - n_i \widehat{\mu}) \\
&= 2(\overline{Y}_i - \widehat{\mu}) \\
\frac{\partial g_i(\widehat{\mu}, \widehat{\sigma}_k^2)}{\partial \widehat{\sigma}_k^2} &= \delta_{ik}.
\end{aligned}$$

This thus gives us a Jacobian:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f}{\partial \widehat{\mu}} & \frac{\partial f}{\partial \widehat{\sigma}_1^2} & \frac{\partial f}{\partial \widehat{\sigma}_2^2} & \frac{\partial f}{\partial \widehat{\sigma}_3^2} & \dots & \frac{\partial f}{\partial \widehat{\sigma}_m^2} \\ \frac{\partial g_1}{\partial \widehat{\mu}} & \frac{\partial g_1}{\partial \widehat{\sigma}_1^2} & \frac{\partial g_1}{\partial \widehat{\sigma}_2^2} & \frac{\partial g_1}{\partial \widehat{\sigma}_3^2} & \dots & \frac{\partial g_1}{\partial \widehat{\sigma}_m^2} \\ \frac{\partial g_2}{\partial \widehat{\mu}} & \frac{\partial g_2}{\partial \widehat{\sigma}_1^2} & \frac{\partial g_2}{\partial \widehat{\sigma}_2^2} & \frac{\partial g_2}{\partial \widehat{\sigma}_3^2} & \dots & \frac{\partial g_2}{\partial \widehat{\sigma}_m^2} \\ \frac{\partial g_3}{\partial \widehat{\mu}} & \frac{\partial g_3}{\partial \widehat{\sigma}_1^2} & \frac{\partial g_3}{\partial \widehat{\sigma}_2^2} & \frac{\partial g_3}{\partial \widehat{\sigma}_3^2} & \dots & \frac{\partial g_3}{\partial \widehat{\sigma}_m^2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial g_m}{\partial \widehat{\mu}} & \frac{\partial g_m}{\partial \widehat{\sigma}_1^2} & \frac{\partial g_m}{\partial \widehat{\sigma}_2^2} & \frac{\partial g_m}{\partial \widehat{\sigma}_3^2} & \dots & \frac{\partial g_m}{\partial \widehat{\sigma}_m^2} \end{bmatrix}.$$

Interestingly, this Jacobian's main diagonal (left to right, top to bottom) consists entirely of 1s. The only off-

diagonal elements that are non-zero are in the first row and first column. We also have this function vector:

$$\mathbf{F} = \begin{bmatrix} f(\widehat{\mu}, \widehat{\sigma}_k^2) \\ g_1(\widehat{\mu}, \widehat{\sigma}_k^2) \\ g_2(\widehat{\mu}, \widehat{\sigma}_k^2) \\ g_3(\widehat{\mu}, \widehat{\sigma}_k^2) \\ \dots\dots\dots \\ g_m(\widehat{\mu}, \widehat{\sigma}_k^2) \end{bmatrix}.$$

And we use the algorithm:

$$\begin{bmatrix} \widehat{\mu} \\ \widehat{\sigma}_1^2 \\ \widehat{\sigma}_2^2 \\ \widehat{\sigma}_3^2 \\ \dots \\ \widehat{\sigma}_m^2 \end{bmatrix}_{\text{New}} = \begin{bmatrix} \widehat{\mu} \\ \widehat{\sigma}_1^2 \\ \widehat{\sigma}_2^2 \\ \widehat{\sigma}_3^2 \\ \dots \\ \widehat{\sigma}_m^2 \end{bmatrix}_{\text{Old}} - \mathbf{J}^{-1}\mathbf{F}$$

to numerically approximate our MLEs under the null. What should we use as our initial guess for  $\begin{bmatrix} \widehat{\mu} \\ \widehat{\sigma}_1^2 \\ \widehat{\sigma}_2^2 \\ \widehat{\sigma}_3^2 \\ \dots \\ \widehat{\sigma}_m^2 \end{bmatrix}$ ? Well,

how about the overall mean for  $\widehat{\mu}$  and the  $k$ th sample variance for  $\widehat{\sigma}_k^2$ ?

# Derivation of the unrestricted maximum likelihood

We will use  $L(\Omega)$  to denote the unrestricted likelihood function.

$$\begin{aligned} L(\Omega) &= \prod_{i=1}^m \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(Y_{ij} - \mu_i)^2\right) \\ &= (2\pi)^{-n/2} \left(\prod_{i=1}^m \sigma_i^{-n_i}\right) \exp\left(-\frac{1}{2} \sum_{i=1}^m \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2\right). \end{aligned}$$

Hence the unrestricted log-likelihood is:

$$\ln L(\Omega) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^m n_i \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^m \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2.$$

Setting the derivative with respect to  $\mu_l$  to zero to find the MLE for  $\mu_l$ :

$$\begin{aligned} \frac{\partial \ln L(\Omega)}{\partial \mu_l} \Big|_{\mu_l = \hat{\mu}_l, \sigma_k^2 = \hat{\sigma}_k^2} &= -\frac{1}{2} \sum_{i=1}^m \frac{1}{\hat{\sigma}_i^2} \sum_{j=1}^{n_i} 2(-1)(Y_{ij} - \hat{\mu}_i) \delta_{il} \\ &= \frac{1}{\hat{\sigma}_l^2} \sum_{j=1}^{n_l} (Y_{lj} - \hat{\mu}_l) \\ &= \frac{1}{\hat{\sigma}_l^2} (n_l \bar{Y}_l - n_l \hat{\mu}_l) \\ &= 0. \\ \implies \hat{\mu}_l &= \bar{Y}_l. \end{aligned}$$

Setting the derivative with respect to  $\sigma_k^2$  to zero to find the MLE for  $\sigma_k^2$ :

$$\begin{aligned} \frac{\partial \ln L(\Omega)}{\partial \hat{\sigma}_k^2} \Big|_{\mu_l = \hat{\mu}_l, \sigma_k^2 = \hat{\sigma}_k^2} &= -\frac{1}{2} \sum_{i=1}^m \frac{n_i}{\hat{\sigma}_i^2} \delta_{ik} - \frac{1}{2} \sum_{i=1}^m -\frac{1}{\hat{\sigma}_i^4} \delta_{ik} \sum_{j=1}^{n_i} (Y_{ij} - \hat{\mu}_i)^2 \\ &= -\frac{n_k}{2\hat{\sigma}_k^2} + \frac{1}{2\hat{\sigma}_k^4} \sum_{j=1}^{n_k} (Y_{kj} - \hat{\mu}_k)^2 \\ &= 0. \\ \implies \hat{\sigma}_k^2 &= \frac{1}{n_k} \sum_{j=1}^{n_k} (Y_{kj} - \hat{\mu}_k)^2 \\ &= \frac{1}{n_k} \sum_{j=1}^{n_k} (Y_{kj} - \bar{Y}_k)^2. \end{aligned}$$

Substituting these MLEs into our expression for  $L(\Omega)$  to get the unrestricted maximum likelihood:

$$\begin{aligned}
L(\widehat{\Omega}) &= (2\pi)^{-n/2} \left( \prod_{i=1}^m (\widehat{\sigma}_i^2)^{-n_i/2} \right) \exp \left( -\frac{1}{2} \sum_{i=1}^m \frac{1}{\widehat{\sigma}_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \widehat{\mu}_i)^2 \right) \\
&= (2\pi)^{-n/2} \left( \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right)^{-n_i/2} \right) \exp \left( -\frac{1}{2} \sum_{i=1}^m \frac{1}{\frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right) \\
&= (2\pi)^{-n/2} \left( \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right)^{-n_i/2} \right) \exp \left( -\frac{1}{2} \sum_{i=1}^m n_i \right) \\
&= (2\pi)^{-n/2} \left( \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right)^{-n_i/2} \right) \exp \left( -\frac{n}{2} \right) \\
&= (2\pi e)^{-n/2} \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right)^{-n_i/2}.
\end{aligned}$$

## Likelihood ratio

Hence our likelihood ratio is:

$$\begin{aligned}
\lambda &= \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} \\
&= \frac{(2\pi e)^{-n/2} \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \widehat{\mu})^2 \right)^{-n_i/2}}{(2\pi e)^{-n/2} \prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right)^{-n_i/2}} \\
&= \frac{\prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \widehat{\mu})^2 \right)^{-n_i/2}}{\prod_{i=1}^m \left( \frac{1}{n_i} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 \right)^{-n_i/2}} \\
&= \prod_{i=1}^m \left( \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}{\sum_{j=1}^{n_i} (Y_{ij} - \widehat{\mu})^2} \right)^{n_i/2}.
\end{aligned}$$

And our test statistic is:

$$-2 \ln \lambda \sim \chi_{m-1}^2.$$

As the unrestricted maximum likelihood has  $m - 1$  more parameters than the maximum likelihood under the null.