Derivation of a test for equality of means with unequal variances

Brenton Horne

27th September 2021

Contents

Hypotheses

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

Definitions

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

We will later use δ_{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 Ω_0 denote the parameter space under the null hypothesis. $\Omega_0 = \{(\mu, \sigma_i^2) : -\infty < \mu < \infty,\ \sigma_i^2 > 0\}$. $\Omega_a = \{(\mu_i, \sigma_i^2) : \mu_i \neq \mu_j$ for at least one pair of i and j values, $-\infty < \mu_i < \infty,\ \sigma_i^2 > 0\}$. The unrestricted parameter space is thus: $\Omega = \Omega_0 \cup \Omega_a$.

$$
\begin{aligned}
L(H_0) &= \prod_{i=1}^{m} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2}(Y_{ij} - \mu)^2\right) \\
&= (2\pi)^{-n/2} \left(\prod_{i=1}^{m} \sigma_i^{-n_i}\right) \exp\left(-\frac{1}{2} \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \mu)^2\right). \qquad (1)
\end{aligned}
$$

Taking the natural logarithm yields:

$$
\ln L(H_0) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^{m} n_i \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \sum_{j=1}^{n_i} (Y_{ij} - \mu)^2.
$$

Differentiating the log-likelihood with respect to μ and setting to zero to maximize the likelihood:

$$\begin{aligned}
\frac{\partial \ln L(H_0)}{\partial \mu}\Big|_{\mu=\widehat{\mu},\ \sigma_i^2=\widehat{\sigma_i^2}} &= -\frac{1}{2}\sum_{i=1}^{m}\frac{1}{\widehat{\sigma_i^2}}\sum_{j=1}^{n_i}2(-1)(Y_{ij}-\widehat{\mu}) \\
&= 0 \\
\sum_{i=1}^{m}\frac{1}{\widehat{\sigma_i^2}}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu}) &= 0 \\
\sum_{i=1}^{m}\frac{1}{\widehat{\sigma_i^2}}(n_i\overline{Y}_i - n_i\widehat{\mu}) &= 0 \\
\left(\sum_{i=1}^{m}\frac{n_i\overline{Y}_i}{\widehat{\sigma_i^2}}\right) - \left(\sum_{i=1}^{m}\frac{n_i}{\widehat{\sigma_i^2}}\right)\widehat{\mu} &= 0 \\
\widehat{\mu} &= \frac{\sum_{i=1}^{m}\frac{n_i\overline{Y}_i}{\widehat{\sigma_i^2}}}{\sum_{i=1}^{m}\frac{n_i}{\widehat{\sigma_i^2}}}. \qquad (2)
\end{aligned}$$

Differentiating the log-likelihood with respect to σ_k^2 and setting to zero to maximize the likelihood:

$$\begin{aligned}
\frac{\partial \ln L(H_0)}{\partial \sigma_k^2}\Big|_{\mu=\widehat{\mu},\ \sigma_i^2=\widehat{\sigma_i^2}} &= -\frac{1}{2}\sum_{i=1}^{m}\frac{n_i}{\widehat{\sigma_i^2}}\delta_{ik} - \frac{1}{2}\sum_{i=1}^{m}-\frac{1}{\widehat{\sigma_i^4}}\delta_{ik}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu})^2 \\
&= -\frac{1}{2}\frac{n_k}{\widehat{\sigma_k^2}} + \frac{1}{2\widehat{\sigma_k^4}}\sum_{j=1}^{n_k}(Y_{kj}-\widehat{\mu})^2 \\
&= 0.
\end{aligned}$$

Multiplying by $2\widehat{\sigma_i^4}$ yields:

$$\begin{aligned}
-n_k\widehat{\sigma_k^2} + \sum_{j=1}^{n_k}(Y_{kj}-\widehat{\mu})^2 &= 0 \\
\widehat{\sigma_k^2} &= \frac{1}{n_k}\sum_{j=1}^{n_k}(Y_{kj}-\widehat{\mu})^2. \qquad (3)
\end{aligned}$$

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

$\widehat{\mu}$ and $\widehat{\sigma_k^2}$.

$$
\begin{aligned}
L(H_0) &= (2\pi)^{-n/2}\left(\prod_{i=1}^{m}\widehat{\sigma_i^2}^{-n_i/2}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{1}{\widehat{\sigma_i^2}}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu})^2\right)\\
&= (2\pi)^{-n/2}\left(\prod_{i=1}^{m}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu})^2\right)^{-n_i/2}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{1}{\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu})^2}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu})^2\right)\\
&= (2\pi)^{-n/2}\left(\prod_{i=1}^{m}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu})^2\right)^{-n_i/2}\right)\exp\left(-\frac{1}{2}\sum_{i=1}^{m}n_i\right)\\
&= (2\pi)^{-n/2}\left(\prod_{i=1}^{m}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu})^2\right)^{-n_i/2}\right)\exp\left(-\frac{n}{2}\right)\\
&= (2\pi e)^{-n/2}\prod_{i=1}^{m}\left(\frac{1}{n_i}\sum_{j=1}^{n_i}(Y_{ij}-\widehat{\mu})^2\right)^{-n_i/2}.
\end{aligned}
$$

Numerical approximation of maximum likelihood estimators

So here we must numerically approximate $\widehat{\mu}$ and $\widehat{\sigma_i^2}$ using a technique like Newton's method and substitute this into our expression for $L(H_0)$ 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:

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

Therefore:

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

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

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

This thus gives us a Jacobian:

$$
J = \begin{bmatrix}
\frac{\partial f}{\partial \widehat{\mu}} & \frac{\partial f}{\partial \widehat{\sigma^2_1}} & \frac{\partial f}{\partial \widehat{\sigma^2_2}} & \frac{\partial f}{\partial \widehat{\sigma^2_3}} & \cdots & \frac{\partial f}{\partial \widehat{\sigma^2_m}} \\
\frac{\partial g_1}{\partial \widehat{\mu}} & \frac{\partial g_1}{\partial \widehat{\sigma^2_1}} & \frac{\partial g_1}{\partial \widehat{\sigma^2_2}} & \frac{\partial g_1}{\partial \widehat{\sigma^2_3}} & \cdots & \frac{\partial g_1}{\partial \widehat{\sigma^2_m}} \\
\frac{\partial g_2}{\partial \widehat{\mu}} & \frac{\partial g_2}{\partial \widehat{\sigma^2_1}} & \frac{\partial g_2}{\partial \widehat{\sigma^2_2}} & \frac{\partial g_2}{\partial \widehat{\sigma^2_3}} & \cdots & \frac{\partial g_2}{\partial \widehat{\sigma^2_m}} \\
\frac{\partial g_3}{\partial \widehat{\mu}} & \frac{\partial g_3}{\partial \widehat{\sigma^2_1}} & \frac{\partial g_3}{\partial \widehat{\sigma^2_2}} & \frac{\partial g_3}{\partial \widehat{\sigma^2_3}} & \cdots & \frac{\partial g_3}{\partial \widehat{\sigma^2_m}} \\
\hdotsfor{6} \\
\frac{\partial g_m}{\partial \widehat{\mu}} & \frac{\partial g_m}{\partial \widehat{\sigma^2_1}} & \frac{\partial g_m}{\partial \widehat{\sigma^2_2}} & \frac{\partial g_m}{\partial \widehat{\sigma^2_3}} & \cdots & \frac{\partial g_m}{\partial \widehat{\sigma^2_m}}
\end{bmatrix}.
$$

Interestingly, this Jacobian's main diagonal (left to right, top to bottom) consists entirely of 1s. The only offdiagonal elements that are non-zero are in the first row and first column. We also have this function vector:

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

And we use the algorithm:

$$\begin{bmatrix} \widehat{\mu} \\ \widehat{\sigma_1^2} \\ \widehat{\sigma_2^2} \\ \widehat{\sigma_3^2} \\ \cdots \\ \widehat{\sigma_m^2} \end{bmatrix}_{\text{New}} = \begin{bmatrix} \widehat{\mu} \\ \widehat{\sigma_1^2} \\ \widehat{\sigma_2^2} \\ \widehat{\sigma_3^2} \\ \cdots \\ \widehat{\sigma_m^2} \end{bmatrix}_{\text{Old}} - J^{-1}\mathbf{F}$$

to numerically approximate our MLEs under the null. What should we use as our initial guess for $\begin{bmatrix} \widehat{\mu} \\ \widehat{\sigma_1^2} \\ \widehat{\sigma_2^2} \\ \widehat{\sigma_3^2} \\ \cdots \\ \widehat{\sigma_m^2} \end{bmatrix}$? Well, how about the overall mean for $\widehat{\mu}$ and the kth sample variance for $\widehat{\sigma_k^2}$?

Derivation of the unrestricted maximum likelihood

We will use L_u to denote the unrestricted likelihood function.

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

Hence the unrestricted log-likelihood is:

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

Setting the derivative with respect to μ_l to zero to find the MLE for μ_l:

$$
\begin{aligned}
\frac{\partial \ln L_u}{\partial \mu_l}\bigg|_{\mu_l=\widehat{\mu}_l,\ \sigma_k^2=\widehat{\sigma_k^2}} &= -\frac{1}{2}\sum_{i=1}^{m}\frac{1}{\widehat{\sigma_i^2}}\sum_{j=1}^{n_i} 2(-1)(Y_{ij}-\widehat{\mu}_i)\delta_{il} \\
&= \frac{1}{\widehat{\sigma_l^2}}\sum_{j=1}^{n_l}(Y_{lj}-\widehat{\mu}_l) \\
&= \frac{1}{\widehat{\sigma_l^2}}(n_l\overline{Y}_l - n_l\widehat{\mu}_l) \\
&= 0. \\
\implies \widehat{\mu}_l &= \overline{Y}_l.
\end{aligned}
$$

Setting the derivative with respect to σ_k^2 to zero to find the MLE for σ_k^2:

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

Substituting these MLEs into our expression for L_u to get the unrestricted maximum likelihood:

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

Likelihood ratio

Hence our likelihood ratio is:

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

And our test statistic is:

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

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