Two sample normal problem.

The parameters are μ_X, μ_Y and σ^2 (recall there is a common variance assumption. Thus the parameter space is

$$\Theta = \{\theta = (\mu_X, \mu_Y, \sigma^2) : \mu_X, \mu_Y \in R, \sigma^2 > 0\}.$$

The hypotheses to be tested are

$$H_0 : \mu_X - \mu_Y = \delta \text{ and } H_A : \mu_X - \mu_Y \neq \delta$$

where δ is a fixed (or given) number. The set of parameters satisfying the null hypothesis is

$$\omega_0 = \{\theta = (\mu + \delta, \mu, \sigma^2) : \mu \in R, \sigma^2 > 0\} .$$

The set ω_0 is a set of dimension 2. The set combining both the null and alternative θ is $\Omega = \Theta$.

The data is $X_i \sim N(\mu_X, \sigma^2)$, $i = 1, \ldots, n$ and $Y_j \sim N(\mu_Y, \sigma^2)$, $j = 1, \ldots, m$, where the X_i are iid, the Y_j are iid, and the X's and Y's are independent. The likelihood function is given by

$$\ell(\theta) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^{n+m}} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(X_i - \mu_X)^2 + \sum_{j=1}^{m}(Y_j - \mu_Y)^2\right)\right\} \tag{1}$$

We will also explicitly write (1) as

$$\ell(\mu_X, \mu_Y, \sigma^2)$$

when this notation is more clear.

The likelihood ratio statistic is

$$\Lambda = \frac{\max_{\theta \in \omega_0} \ell(\theta)}{\max_{\theta \in \Theta} \ell(\theta)} \tag{2}$$

For $\theta \in \omega_0$ the likelihood (1) is $\ell(\mu + \delta, \mu, \sigma^2)$, so that we must maximize this as a function of two variables μ and σ^2. It is more convenient to work with $\log(\ell(\mu + \delta, \mu, \sigma^2))$ to use calculus for this maximization. The student should work with this and obtain that the maximum of

$$\max_{\theta \in \omega_0} \ell(\theta)$$

occurs at

$$\hat{\mu}_0 = \frac{1}{n+m}\left\{n\bar{X} + m\bar{Y} - n\delta\right\}$$

and

$$\hat{\sigma}_0^2 = \frac{1}{n+m}\left\{\sum_{i=1}^{n}(X_i - \hat{\mu}_0 - \delta)^2 + \sum_{j=1}^{m}(Y_j - \hat{\mu}_0)^2\right\}$$

where the subscript 0 represents the fact that the maximization is over ω_0. Let $\hat{\theta}_0 = (\hat{\mu}_0 + \delta, \hat{\mu}_0, \hat{\sigma}_0^2)$. Therefore

$$\ell(\hat{\theta}_0) = \max_{\theta \in \omega_0} \ell(\theta)$$

There is a fair amount of simplification in the exponential part of the above, yielding

$$\ell(\hat{\theta}_0) = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}_0^2}\right)^{n+m}} \exp\left\{-\frac{n+m}{2}\right\}$$

For the denominator of (2) the maximization is over Ω. The likelihood (1) is $\ell(\mu_X, \mu_Y, \sigma^2)$, so that we must maximize this as a function of three variables μ_X, μ_Y and σ.

The student should work with this and obtain that the maximum of

$$\max_{\theta \in \Omega} \ell(\theta)$$

occurs at

$$\hat{\mu}_X = \bar{X}\ ,\ \hat{\mu}_Y = \bar{Y}$$

and

$$\hat{\sigma}^2 = \frac{1}{n+m}\left\{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{j=1}^{m}(Y_j - \bar{Y})^2\right\}$$

Let $\hat{\theta} = (\hat{\mu}_X, \hat{\mu}_Y, \sigma^2)$. Notice that $\hat{\theta}$ is not the same as $\hat{\theta}_0$. This is to be expected since the two maximizations are not the same.

Again there is a fair amount of simplification in the exponential part of the above, yielding

$$\ell(\hat{\theta}) = \frac{1}{\left(\sqrt{2\pi\hat{\sigma}^2}\right)^{n+m}} \exp\left\{-\frac{n+m}{2}\right\}$$

Thus the likelihood ratio becomes

$$\Lambda = \frac{\ell(\hat{\theta}_0)}{\ell(\hat{\theta})} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right)^{(n+m)/2} \tag{3}$$

Next we simplify this ratio. Towards this note

$$\begin{aligned}
(n+m)\hat{\sigma}_0^2 &= \sum_{i=1}^{n}(X_i - \hat{\mu}_0 - \delta)^2 + \sum_{j=1}^{m}(Y_j - \hat{\mu}_0)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \hat{\mu}_0 - \delta)^2 + \sum_{j=1}^{m}(Y_j - \bar{Y} + \bar{Y} - \hat{\mu}_0)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n\left(\bar{X} - \hat{\mu}_0 - \delta\right)^2 + \sum_{j=1}^{m}(Y_j - \bar{Y})^2 + m\left(\bar{Y} - \hat{\mu}_0\right)^2 \\
&= (n+m)\hat{\sigma}^2 + n\left(\bar{X} - \hat{\mu}_0 - \delta\right)^2 + m\left(\bar{Y} - \hat{\mu}_0\right)^2
\end{aligned}$$

Next we simplify the last two terms. Using the definition of $\hat{\mu}_0$ we obtain

$$\begin{aligned}
\bar{X} - \hat{\mu}_0 - \delta &= \bar{x} - \frac{n\bar{X} + m\bar{Y} - n\delta}{n+m} - \delta \\
&= \frac{m}{n+m}\left(\bar{X} - \bar{Y} - \delta\right) \\
\bar{Y} - \hat{\mu}_0 &= \bar{Y} - \frac{n\bar{X} + m\bar{Y} - n\delta}{n+m} \\
&= \frac{n}{n+m}\left(\bar{Y} - \bar{X} + \delta\right) \\
&= -\frac{n}{n+m}\left(\bar{X} - \bar{Y} - \delta\right)
\end{aligned}$$

Therefore

$$\begin{aligned}
n\left(\bar{X}-\hat{\mu}_0-\delta\right)^2+m\left(\bar{Y}-\hat{\mu}_0\right)^2 &= \left(n\frac{m^2}{(n+m)^2}+m\frac{n^2}{(n+m)^2}\right)\left(\bar{X}-\bar{Y}-\delta\right)^2 \\
&= \frac{nm}{n+m}\left(\frac{m}{(n+m)}+\frac{n}{(n+m)}\right)\left(\bar{X}-\bar{Y}-\delta\right)^2 \\
&= \frac{nm}{n+m}\left(\bar{X}-\bar{Y}-\delta\right)^2
\end{aligned}$$

Putting these together we find the likelihood ratio Λ is given by

$$\Lambda^{\frac{2}{n+m}}=\frac{\hat{\sigma}^2}{\hat{\sigma}^2+\frac{nm}{(n+m)^2}\left(\bar{X}-\bar{Y}-\delta\right)^2}$$

Notice that

$$\frac{1}{1+A}<c \Leftrightarrow A>c_1$$

for appropriate constant c_1. The student should find c_1 in terms of c.

The rejection region is then of the form

$$\frac{nm}{(n+m)^2}\frac{\left(\bar{X}-\bar{Y}-\delta\right)^2}{\hat{\sigma}^2}>c$$

The rejection region can be rewritten as

$$\frac{\left(\bar{X}-\bar{Y}-\delta\right)^2}{s_P^2\left(\frac{1}{n}+\frac{1}{m}\right)}>c_1 \tag{4}$$

by using the fact that

$$s_P^2=\frac{n+m}{n+m-2}\hat{\sigma}^2$$

The student should find the relation between c_1 and the constant c.

In equation (4) the statistic of interest is

$$\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{s_P^2\left(\frac{1}{n}+\frac{1}{m}\right)}} \tag{5}$$

Under the null hypothesis we can obtain the sampling distribution of (5).

Notice

$$\bar{X}-\bar{Y}\sim N\left(\delta,\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)\right)$$

and

$$\begin{aligned}
\frac{n+m-2}{\sigma^2}s_P^2 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2+\frac{1}{\sigma^2}\sum_{j=1}^{m}(Y_j-\bar{Y})^2 \\
&\sim \chi^2_{(n+m-2)}
\end{aligned}$$

and that these two normal and χ^2 random variables are independent. The student should consider why these are independent and why we have a $\chi^2_{(n+m-2)}$ distribution. The easiest way to do this is to return to one of our problems and handouts in Stat 3657. Recall our use of orthonormal matrices in the study of the sample mean and sample variance in the case of iid normal samples. A similar method will apply here; what will be the appropriate choice of an orthonormal matrix? We could also obtain these results as a direct application of the results in Chapter 6 or Chapter 3, and some additional results dealing with two functions, each using different independent r.v.s. Specifically

$$\bar{X}_n, \bar{Y}_m, S^2_{X,n}, S^2_{Y,m}$$

are independent. Therefore

$$\bar{X} - \bar{Y} - \delta$$

and

$$\frac{n+m-2}{\sigma^2} s_P^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 + \frac{1}{\sigma^2}\sum_{j=1}^{m}(Y_j - \bar{Y})^2$$

are independent. Using the method of moment generating functions we can then find their marginal distributions, and then use the properties of Chapter 6 to determine the distribution of the student's t ratio in (5).

Recalling that a t random variable is the ratio of an independent normal and the square root of a χ^2 divided by its degree of freedom. Thus it will follow that under the null hypothesis that (5) has a $t_{(n+m-2)}$ distribution, that is a t distribution with $n+m-2$ degrees of freedom.

Consider the test of size α given by (4). It is given by

$$\left| \frac{(\bar{X} - \bar{Y} - \delta)}{s_P\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}} \right| > c_2 = \sqrt{c_1}$$

and c_2 is given by

$$P(T > c_2) = \frac{\alpha}{2}$$

where $T \sim t_{(n+m-2)}$.

Confidence Interval for $\mu_X - \mu_Y$

How do we obtain the confidence interval for $\mu_X - \mu_Y$. Following our framework this 100(1-α)% confidence interval is the random set of possible values δ for which the null

$$H_0 : \mu_X - \mu_Y = \delta$$

is not rejected in favour of the alternative

$$H_A : \mu_X - \mu_Y \neq \delta$$

Specifically this is the random set

$$\begin{aligned}
D &= \left\{ \delta \ : \ \left| \frac{(\bar{X} - \bar{Y} - \delta)}{s_P \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}} \right| \leq t_{(n+m-2), \frac{\alpha}{2}} \right\} \\
&= \left\{ \delta \ : \ \bar{X} - \bar{Y} - t_{(n+m-2), \frac{\alpha}{2}} s_P \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)} \leq \delta \leq \bar{X} - \bar{Y} + t_{(n+m-2), \frac{\alpha}{2}} s_P \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)} \right\} \\
&= \left[\bar{X} - \bar{Y} - t_{(n+m-2), \frac{\alpha}{2}} s_P \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)} \ , \ \bar{X} - \bar{Y} + t_{(n+m-2), \frac{\alpha}{2}} s_P \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)} \right]
\end{aligned}$$

A more compact notation for the 100(1-α)% confidence interval is

$$\bar{X} - \bar{Y} \pm t_{(n+m-2), \frac{\alpha}{2}} s_P \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)}$$

One Sided Alternative

The student should work through the so called one sided alternative

$$H_0 : (\mu_X - \mu_Y) = \delta = \delta_0 \text{ versus } H_A : \delta > \delta_0$$

Recall in our notation $\delta = \mu_X - \mu_Y$.

Find the GLR $\Lambda(\mathbf{X})$. You should note how this compares with the GLR in the two sided alternative; see equation (3).

Manipulate, for a positive constant $c < 1$, the rejection region

$$R = \{\mathbf{x} : \Lambda(\mathbf{x}) < c\}$$

Work at this to show this is also of the form

$$R = \{\mathbf{x} : \frac{\bar{x} - \bar{y} - \delta_0}{\sqrt{S_p^2}} > c_1\}$$

or equivalently

$$R = \{\mathbf{x} : \frac{(\bar{x} - \bar{y} - \delta_0)}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} > c_2\}$$

for appropriate constants c_1 and c_2. Also find the relations between c, c_1 and c_2.

Notice we can now find the actual test, that is a formula involving the test statistic and constant to give the rejection region. For given δ_0 and $n, ,$, and given size α, we find c_2 by solving

$$P_0\left(\frac{(\bar{X} - \bar{Y} - \delta_0)}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} > c_2\right) = \alpha$$

In general this would be difficult to solve in an exact sense. However in the iid normal case here for any (μ, σ^2) satisfying the null hypothesis, that is $(\mu, \sigma^2) \in \Theta_0$, then

$$T = \frac{(\bar{X} - \bar{Y} - \delta_0)}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t_{(n-1+m-1)}$$

Thus

$$\begin{aligned} \alpha &= P_0\left(\frac{(\bar{X} - \bar{Y} - \delta_0)}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} > c_2\right) \\ &= P\left(T > c_2\right) \end{aligned}$$

Thus $c_2 = t_{(n-1+m-1),\alpha}$, the upper α critical value or $1 - \alpha$ quantile.

Test for Equality of Variances

In this problem the parameter space is

$$\Theta = \{(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2) : -\infty < \mu_X, \mu_Y < \infty, \sigma_X^2 > 0, \sigma_Y^2 > 0\}$$

We are interested in testing

$$H_0 : \sigma_X^2 = \sigma_Y^2 \text{ versus } H_A : \sigma_X^2 \neq \sigma_Y^2$$

Both the null and alternative are composite hypotheses. The null set is

$$\Theta_0 = \{(\mu_X, \mu_Y, \sigma^2, \sigma^2) : -\infty < \mu_X, \mu_Y < \infty, \sigma^2 > 0\}$$

It is a set of dimension 3. The alternative set is

$$\Theta_A = \Theta_0^c$$

Aside : It would be a different alternative set if we were interested in testing

$$H_0 : \sigma_X^2 = \sigma_Y^2 \text{ versus } H_A : \sigma_X^2 < \sigma_Y^2$$

that is the variance of the Y population is larger than the variance of the X population.