

MATH 203: PRINCIPLES OF STATISTICS 1
FALL 2024

HYPOTHESIS TESTING WITH UNEQUAL VARIANCES

To test the null hypothesis $H_0 : \mu_1 = \mu_2$ against a selected alternative

- when m or n is **small**, we must assume **Normality** of the two populations, and that the two population **standard deviations** are the **same**, that is $\sigma_1 = \sigma_2 = \sigma$ say. In this case, we use the test statistic

$$T = \frac{(\bar{X} - \bar{Y})}{S_P \sqrt{\frac{1}{m} + \frac{1}{n}}} \quad (1)$$

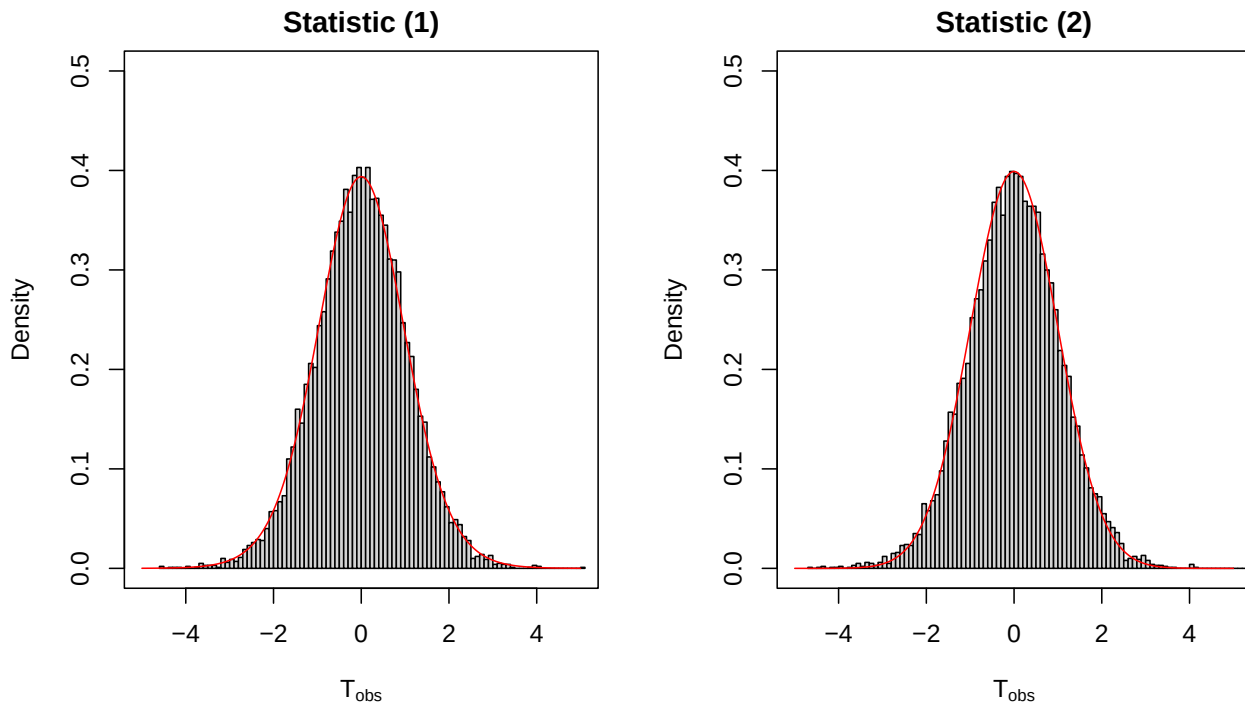
and compare this with the t-distribution with $m + n - 2$ degrees of freedom, where m and n are the two sample sizes.

- when m and n are **large**, we use the test statistic

$$T = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}} \quad (2)$$

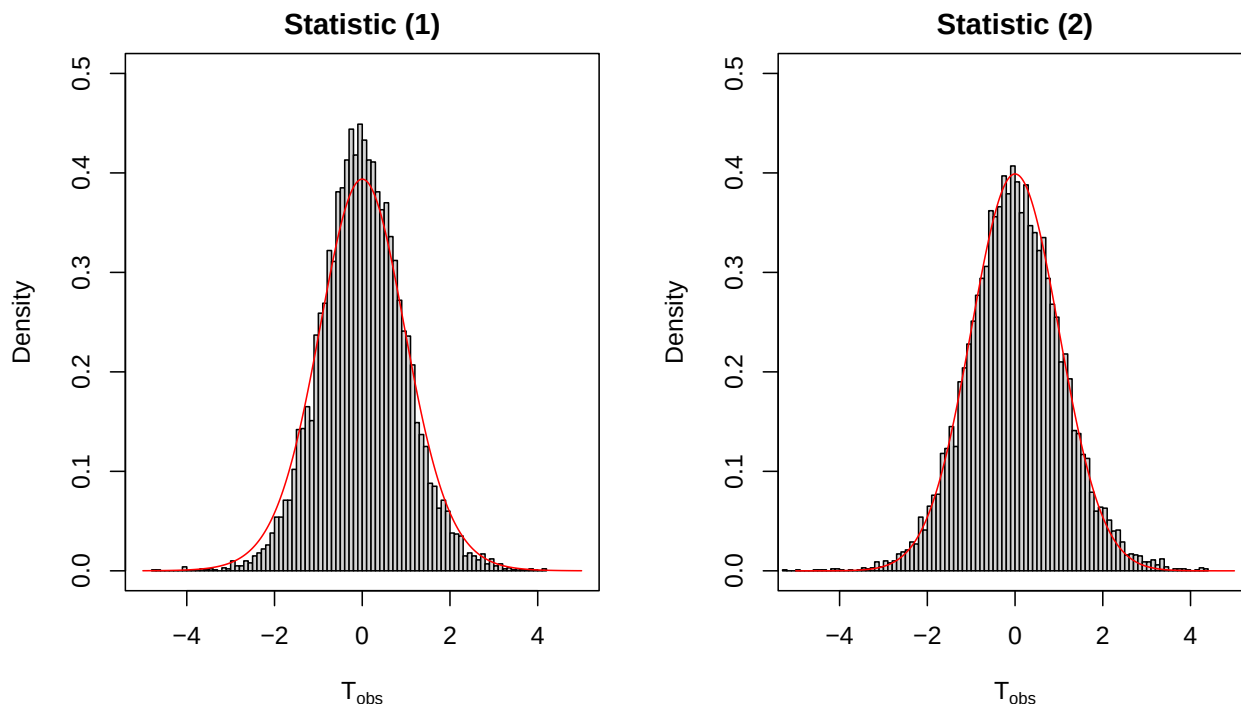
and compare this with the standard Normal distribution. We **do not** need to assume Normality of the original data, nor assume that the population variances are equal.

In the first simulation, $m = 9$ and $n = 12$, in $N = 10000$ replicated studies with $\mu_1 = \mu_2 = 10$ and $\sigma_1 = \sigma_2 = 2$, with the data Normally distributed.

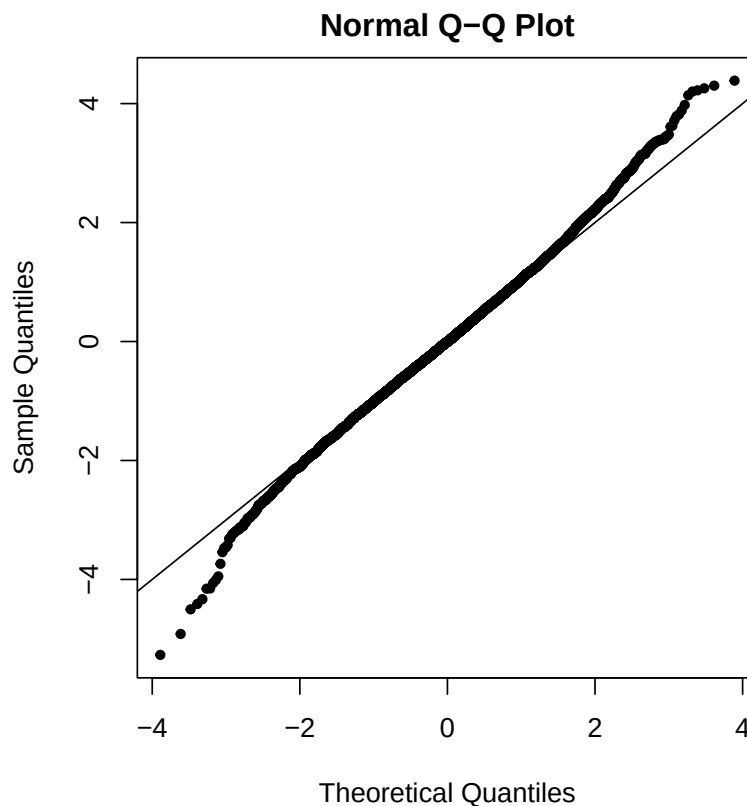


The distribution of the two statistics follow the presumed exact and approximate null distributions plotted in red: t_{19} for statistic (1), and standard Normal for statistic (2).

Now we repeat the calculation with $\sigma_1 = 2$ but $\sigma_2 = 5$, still with the data Normally distributed.



The null distribution of statistic (1) is no longer the t-distribution computed previously, as expected. The distribution of statistic (2) is still apparently well-approximated by the standard Normal. We can check this with a QQ plot:



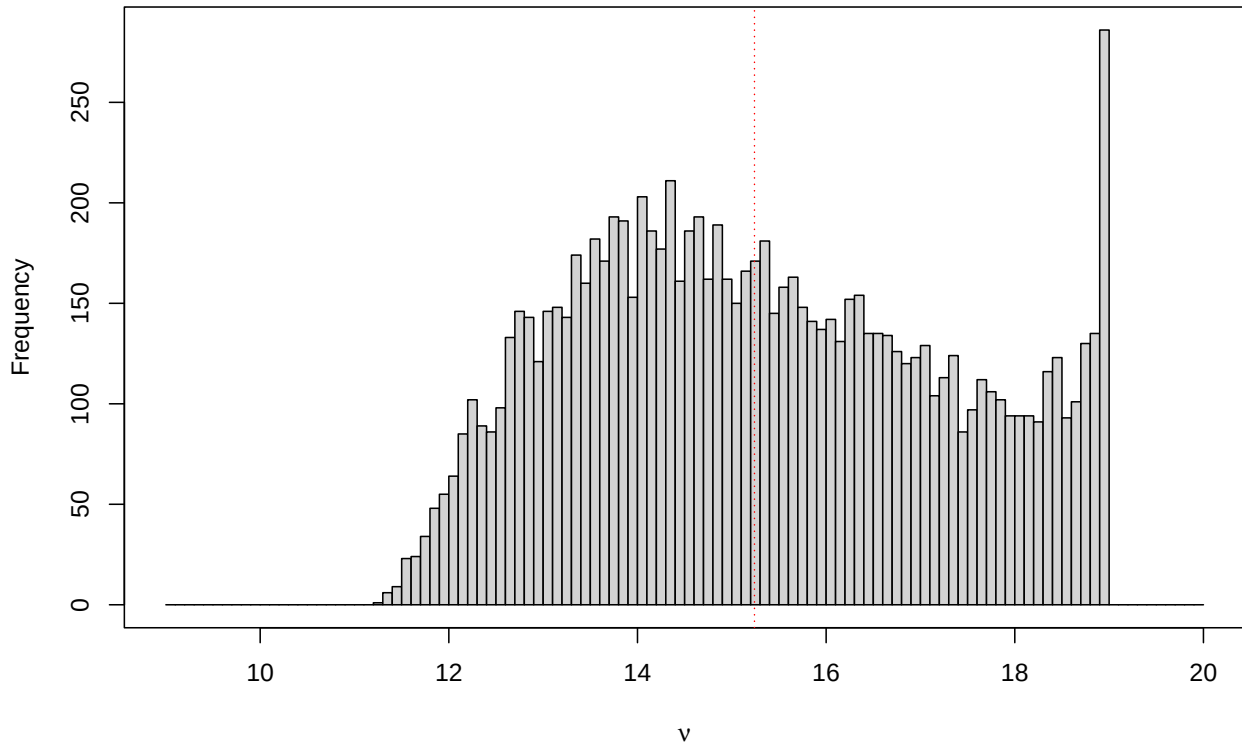
The QQ plot reveals that in fact the null distribution of statistic (2) is heavier tailed than the standard Normal. This is important, because when we carry out hypothesis tests, we are most interested in the tails of the null distribution. A standard approximation to the null distribution in this case is the t-distribution with

$$\nu = \frac{\left(\frac{s_1^2}{m} + \frac{s_2^2}{n}\right)^2}{\frac{1}{m-1} \left(\frac{s_1^2}{m}\right)^2 + \frac{1}{n-1} \left(\frac{s_2^2}{n}\right)^2} \quad (3)$$

Note that this degrees of freedom depends on m and n , but also the two observed sample variances (and therefore the on true values of σ_1 and σ_2) and is not in general a whole number. The maximum value ν can take is $m + n - 2$.

We can compute the value of ν for each of the previous replicates: in this case, the maximum value 19, and the value we obtain by using σ_1^2 and σ_2^2 into (3) is 15.2403.

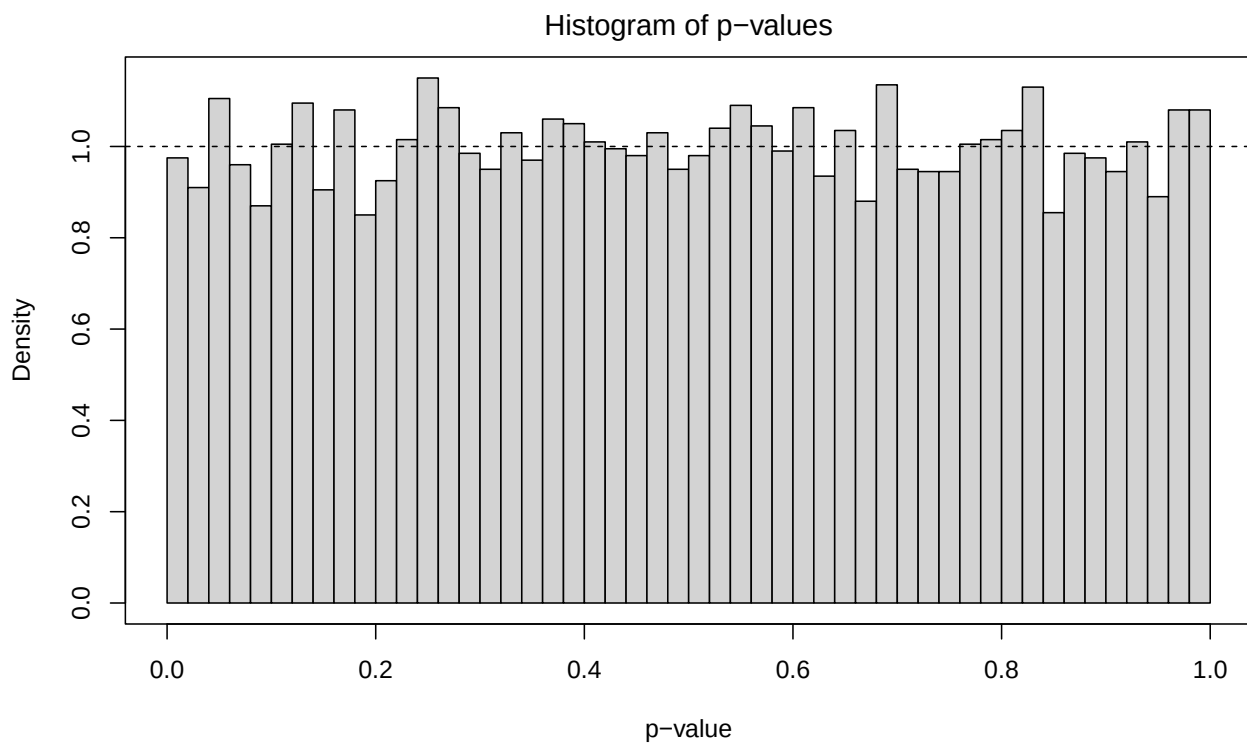
Histogram of ν ($m=9, n=12$)



As ν varies across replicates, we check the uniformity of p-values derived in a test against a two-sided alternative. That is, for each we compute

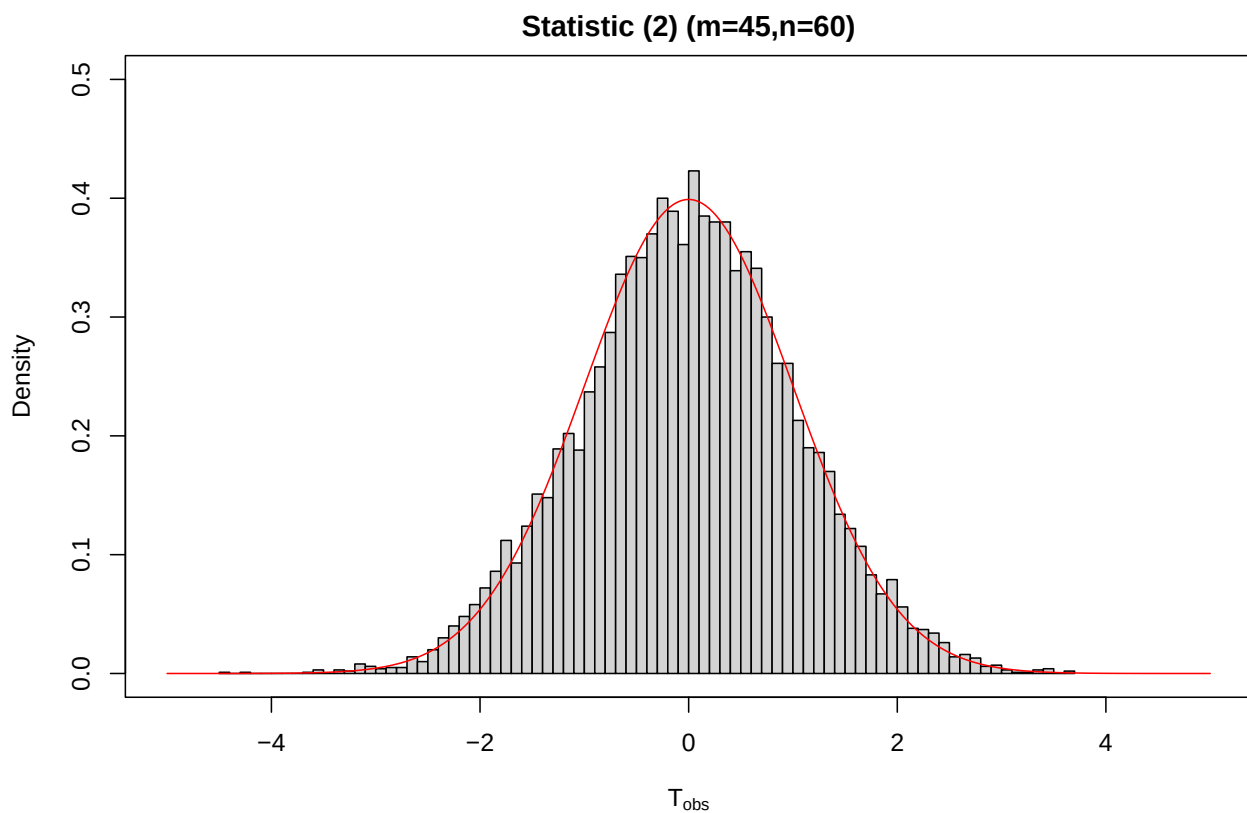
$$P_0(|T| \geq T_{obs})$$

with P_0 being the t-distribution with ν degrees of freedom, with ν computed using (3). This probability should be have a uniform distribution if the null hypothesis is true.

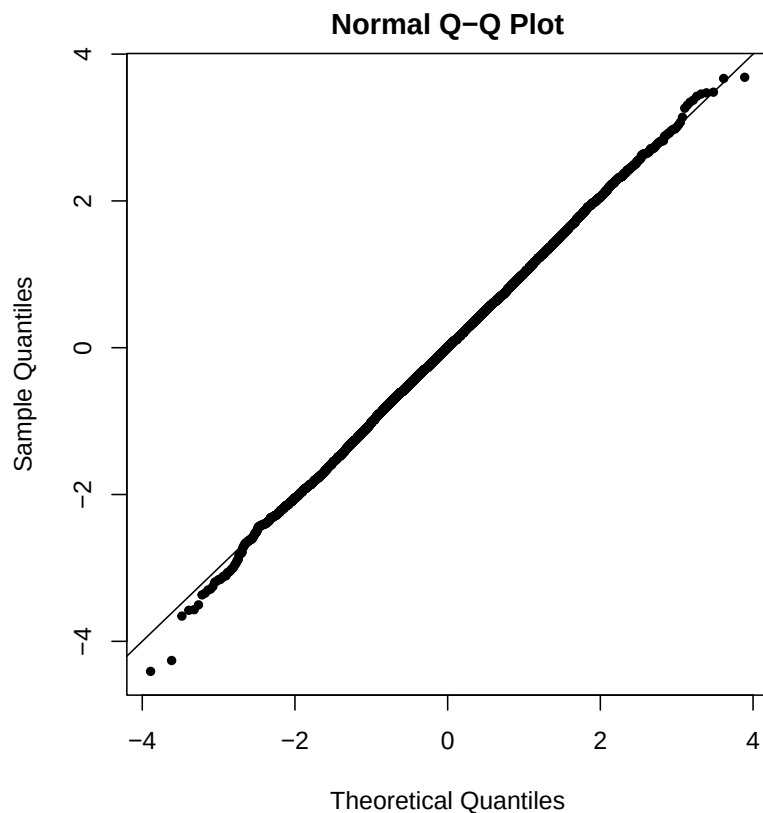


This confirms that the approximation is a good one at least for these values of m and n . We can therefore use this approximation to carry out the test, and also to compute confidence intervals.

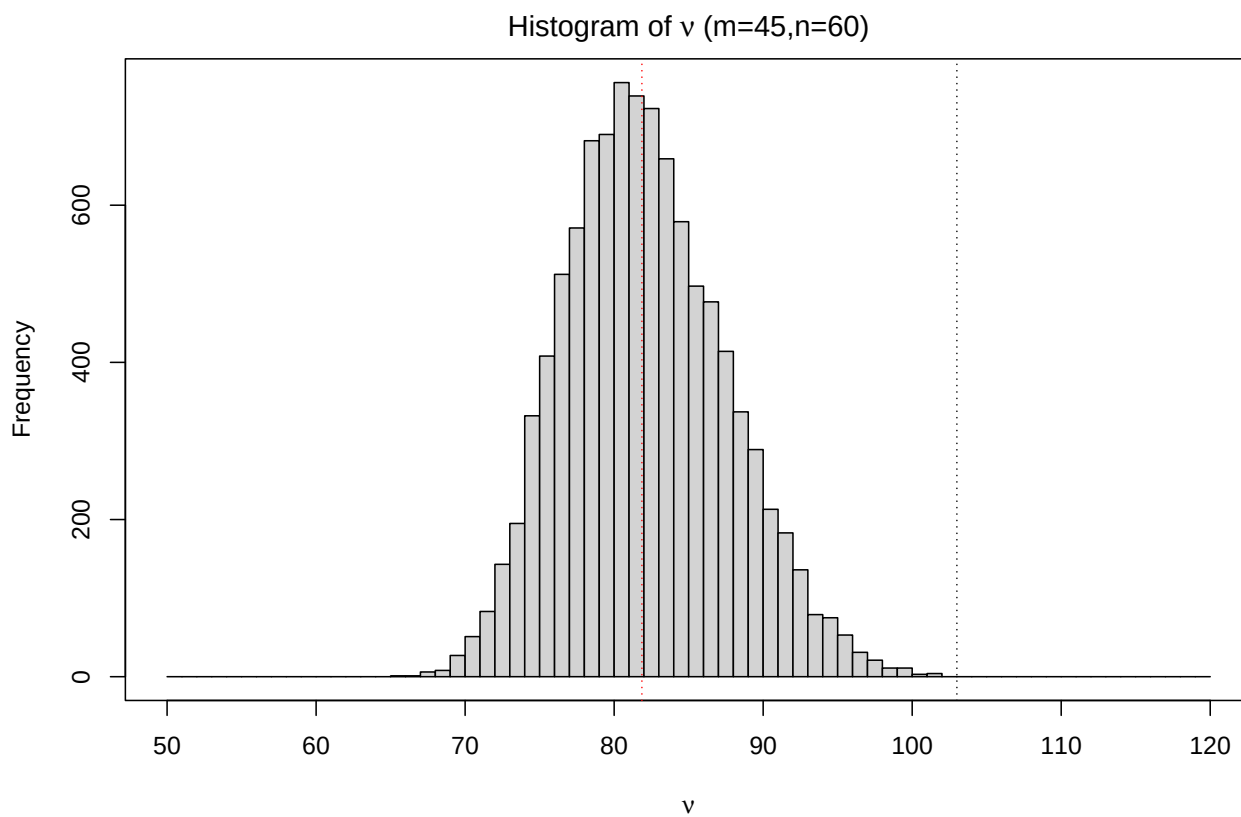
If the sample sizes increase, the distribution of test statistic (2) becomes more like the Normal distribution



We can verify the Normality with a QQ plot:

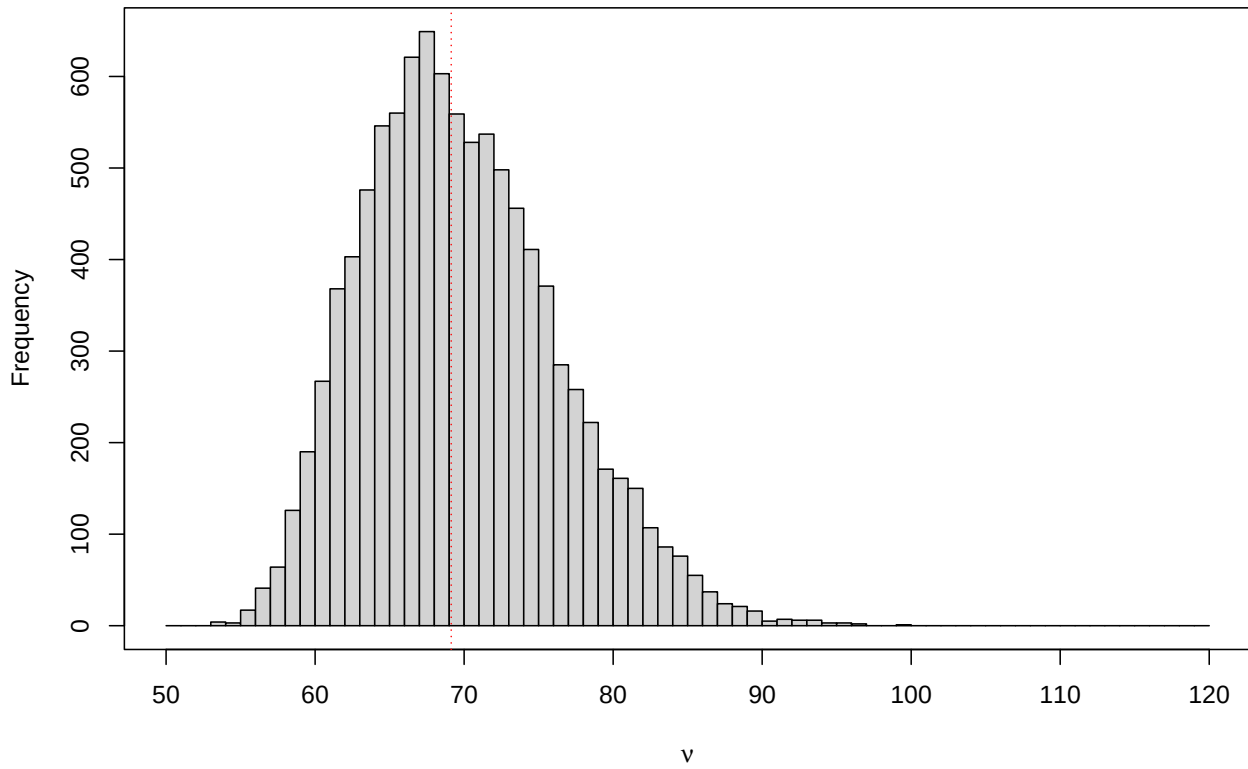


However, the degrees of freedom computed using (3) when $\sigma_1 = 2$ and $\sigma_2 = 5$ is still below $m + n - 2 = 103$. The value we obtain by using σ_1^2 and σ_2^2 in (3) is 81.8627.



Finally, we confirm that the null distribution of statistic (2) changes as σ_1 and σ_2 change: if $\sigma_1 = 8$ and $\sigma_2 = 5$, we obtain the following computed values of ν in the 10000 replicates:

Histogram of ν ($m=45, n=60$)



The degrees of freedom computed using (3) when $\sigma_1 = 8$ and $\sigma_2 = 5$ using σ_1^2 and σ_2^2 in (3) is 69.1327.