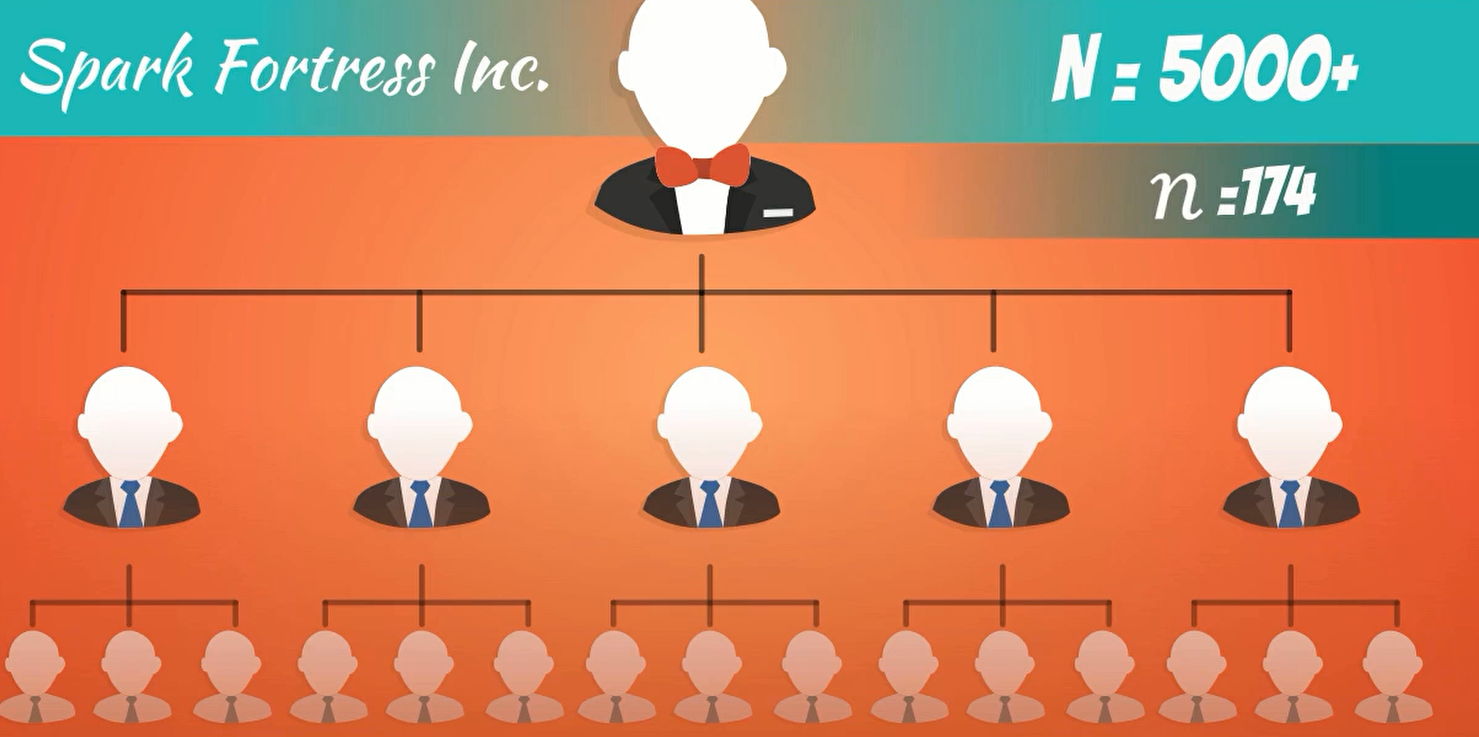
Practical Example. Hypothesis Testing

We explore the topic of gender pay gap.



we will test whether a particular company is discriminating against some of its employees on a gender basis. Our Fictitious company is called Sparke fortress incorporated it is a big company with more than 5000 employees and here will work with a sample of 174 of them.



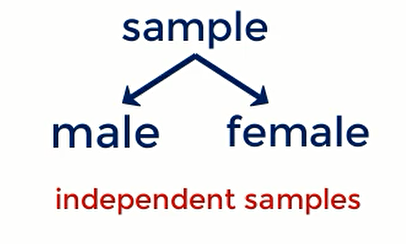
we have tata showing us their detail.





we are going to test if there is a significant difference in the salaries of employees are paid based on their gender.

Our 174-employee sample could be divided into two sub samples one that is exclusively male and one female. So, we have two samples drawn from the same population that are independent.



although so far, we have worked with different populations only. if the values in one sample reveal no information about the other sample, then they are considered independent.



There are different methodologies to conduct a study and while regression analysis is my preferred one, we will have to wait until next section for that here.



let's State that two hypotheses

**H0**: The average male salary is equal to the average female salary.

H0: µm = µf

or H0: µm - µf = 0

**H1**: The average male salary differs from the average female salary.

H0: µm ≠ µf

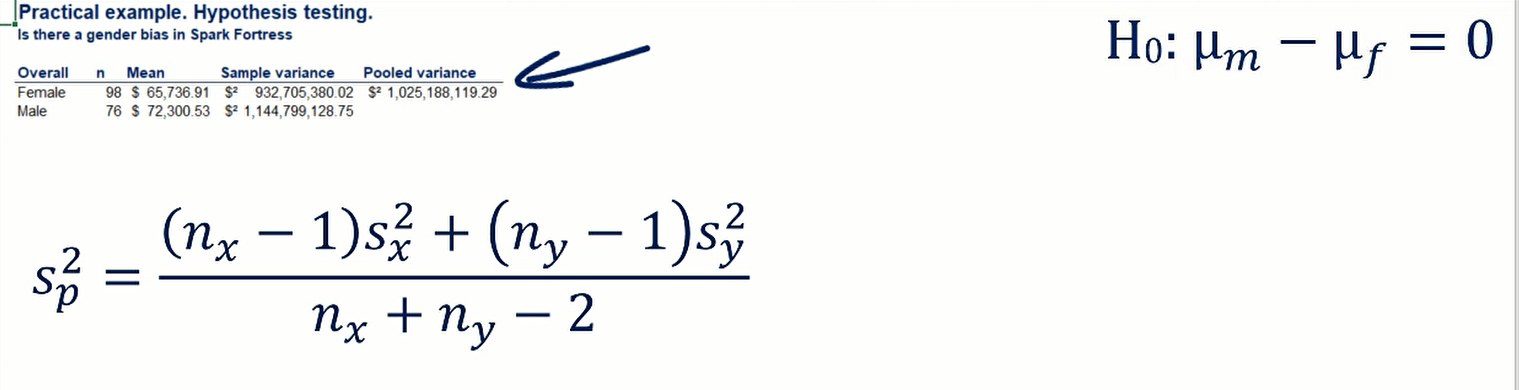
Or H0: µm -µf ≠ 0

The test should use is t-test for the independent samples. salary population variance, it is surely unknown and we can assume it is equal.

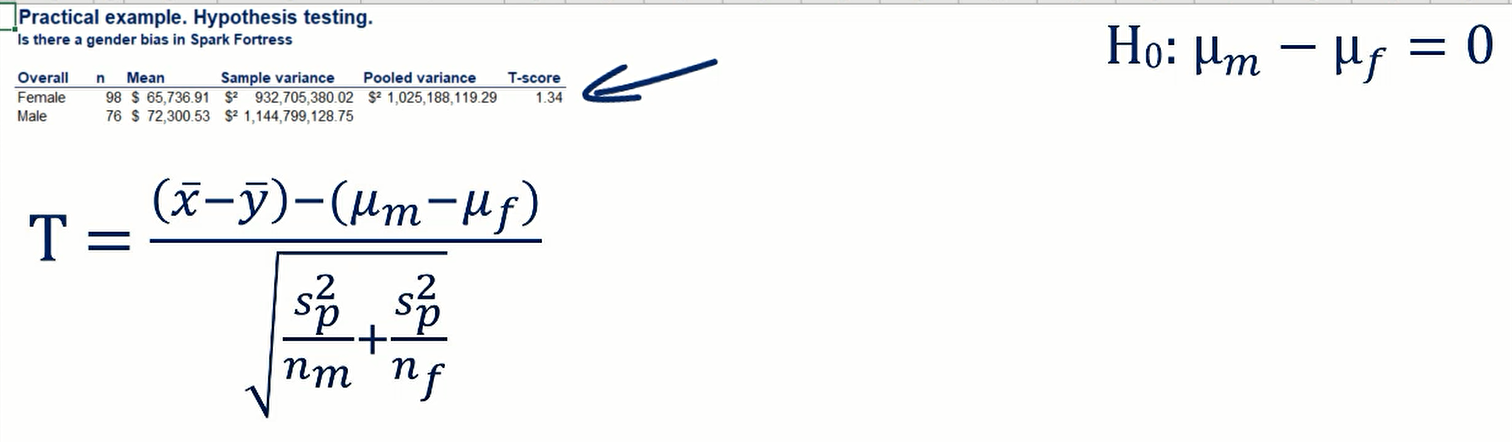


let's construct a frequency distribution table.

we have 98 females and 76 males. These are our sample sizes. assume that the population variances are equal. We should also compute the Pooled variance.



Pooled variance formula



t-score formula

Here, t-score =1.34

Degree of freedom = 98 +76-2 =172

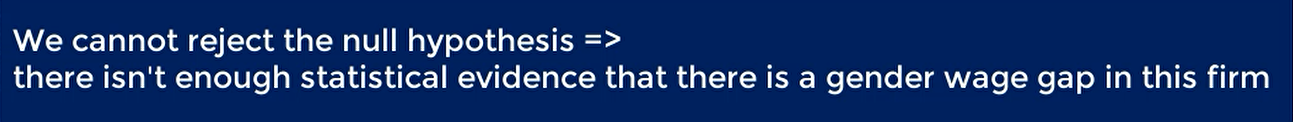
Once we have surpassed 50 degrees of freedom. student's t distribution almost completely overlaps with the normal distribution.



Thus, the p-values for t-score of 1.34 and Z -score of 1.34 it will be Virtually the same. So, I will just give you the P-value is 0.182. The p-value is much greater than all common levels of significance.

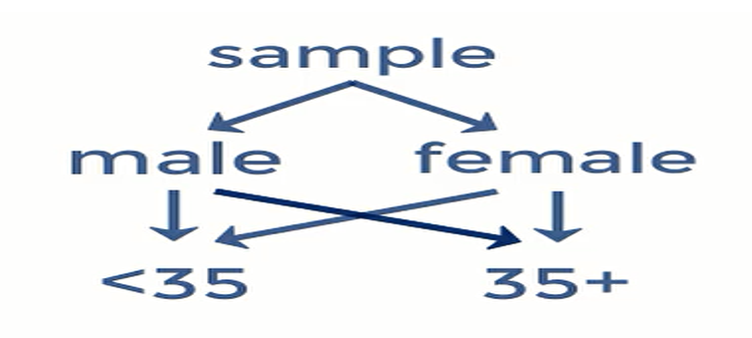


we conclude that we cannot reject the null hypothesis.



let's order the salaries from largest to the smallest. We can see that the highest paid employee is actually the president and CEO of the company. Caroline bolds Who is female.

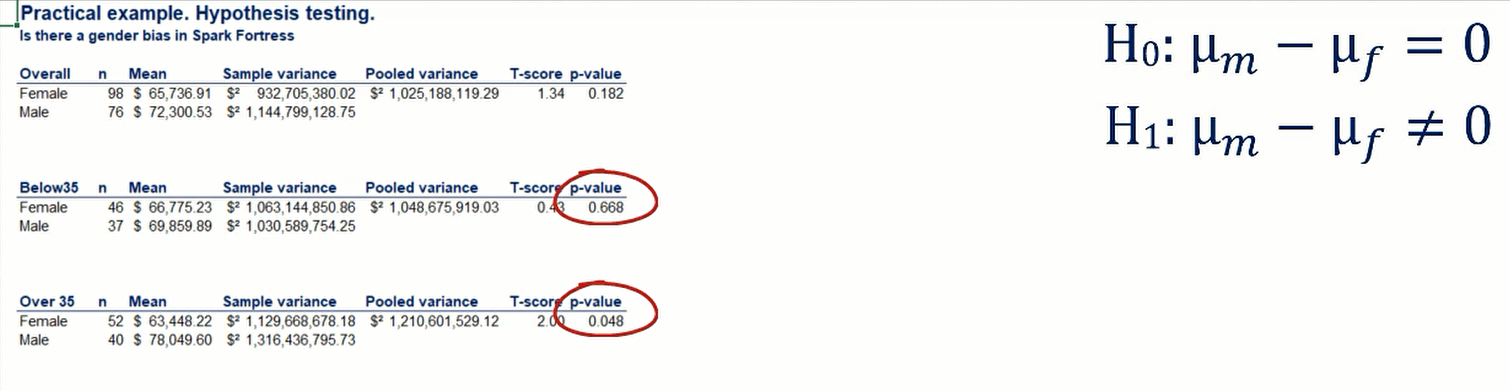
I would normally for the segment the data. let's divided the employees into two more groups below 35 and above 35. This will give us valuable information about the wage equality of younger and older staff.



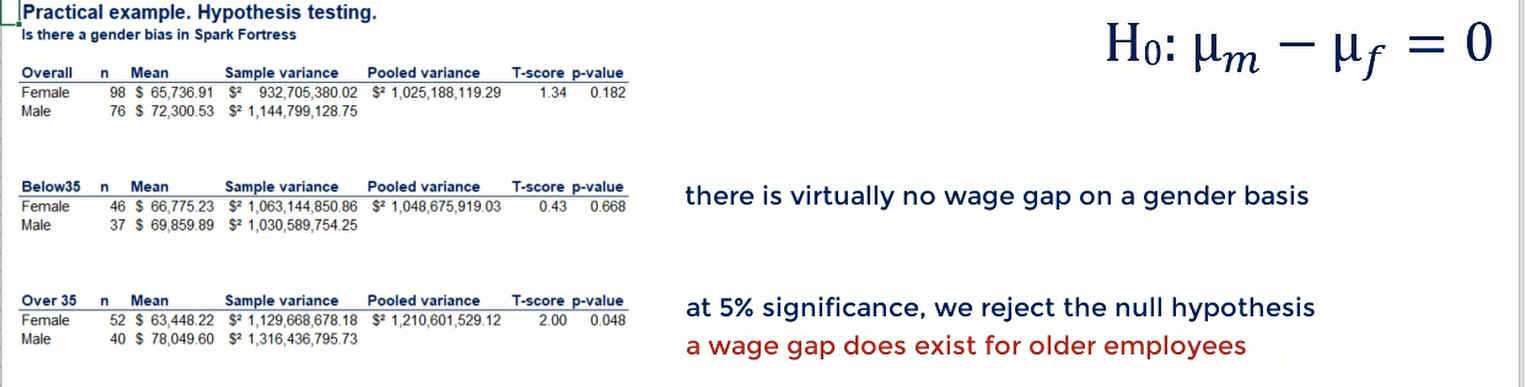
I have created two more data sets that are based on the original one. Let's run the same tests as before, but this time we will do it in our segmented data.

The hypothesis is the same.

We get for these true tests is a test score of 0.43 for employees below 35 and 2.0 for employees over 35. the corresponding p-values are 0.668 and 0.048.



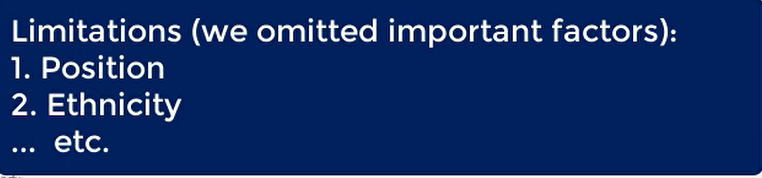
what these numbers mean that is the group below 35 there is Virtually no wage gap on a gender basis. In the older group P value is 0.048 this is very close to 0.05 but still below it. This implies that at 95% significance we reject the null hypothesis. therefore, a wage gap does exist for older employees.



this is a two-sided test so we are not sure who gets more money. Well do you remember the nifty track the t score of 2 is positive. therefore, the difference in pay is positive in favor of males.



The limitation of this analysis.



so, we are not completely sure what's going on in the frame but we can say that overall, there is no wage kept in Sparke Fortress and this is driven by wage equality among young employees.

