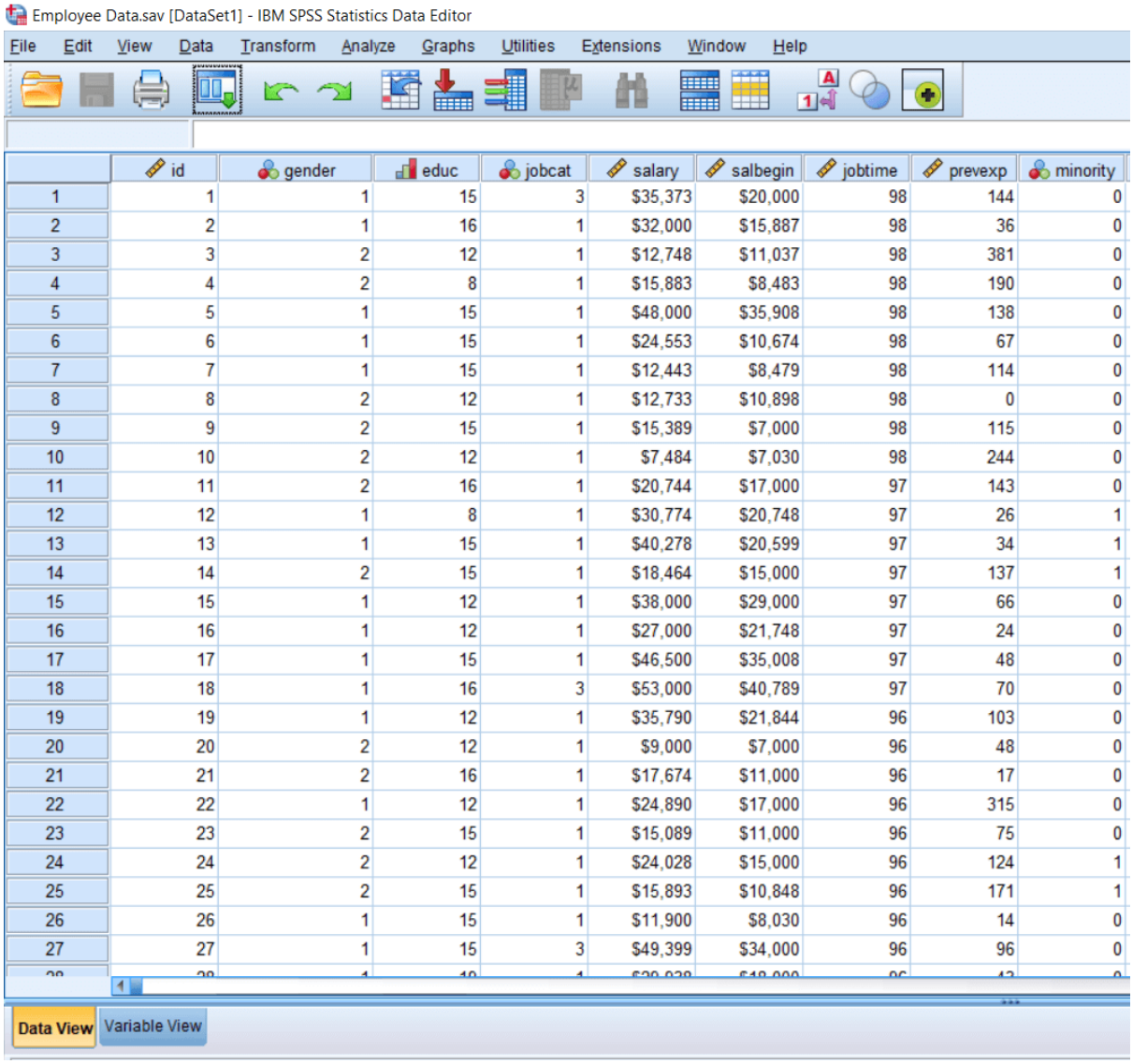
# Paired Sample T-test

- In this section, we will learn about **Paired Sample T-test**. A **Paired sample t-test** is also known as a **correlated** sample T-test or **dependent** sample T-test. So it presumed that either repeated measure design had been used in the study or we are taking the same subject across two different situations. So it is a very useful test to conduct **repeated** measure inference analysis and that often we do.
- **For example**, suppose the **researcher** wants to understand whether a student by a particular method leads to significant improvement in his **performance**. For that, he takes **10** students or **15** students or **30** We will take atleast **30 students** because that is recommended for doing a parametric test. So we took 30 students and measured their **performance** at the **beginning** of training without any administration of training. Now we administrate a training program, and after the training, we again measure their **performance**. Now we want to **compare** the performance of students **before** the training program and **after** the training program. In that case, he can use a **paired sample t-test**.
- Similarly, a **Government scan** uses a **Paired sample t-test** for measuring the effect of a tourism promotion for the **tourist**. So tourists come to the country, and when they are in the country, we can measure how they feel about the country, do they like the country and they are familiar with the country's culture. After measuring the attitude towards the country or destination, once they have been through the country and they are living in the country. At that time, we can again ask them to fill a survey. Now we can find out whether there have been significant differences in tourist's attitude when they arrived in the country and when they are living in the country.
- Similarly, when we visit the **hotels** or go for a **party** or **picnic**, we want to know whether going to a **picnic** or **party** or visiting a **hostel** leads to any improvement in our mood. So we can measure our **mood** at the **beginning**, and once we are there, we finished our vacation and again rate our **mood** and find out whether there have been any significant improvements in our mood. **Hospitals** or **hostel** can use them. It can be used at any place where we measure the difference in the same individual's condition after they have undergone some training or experiences.

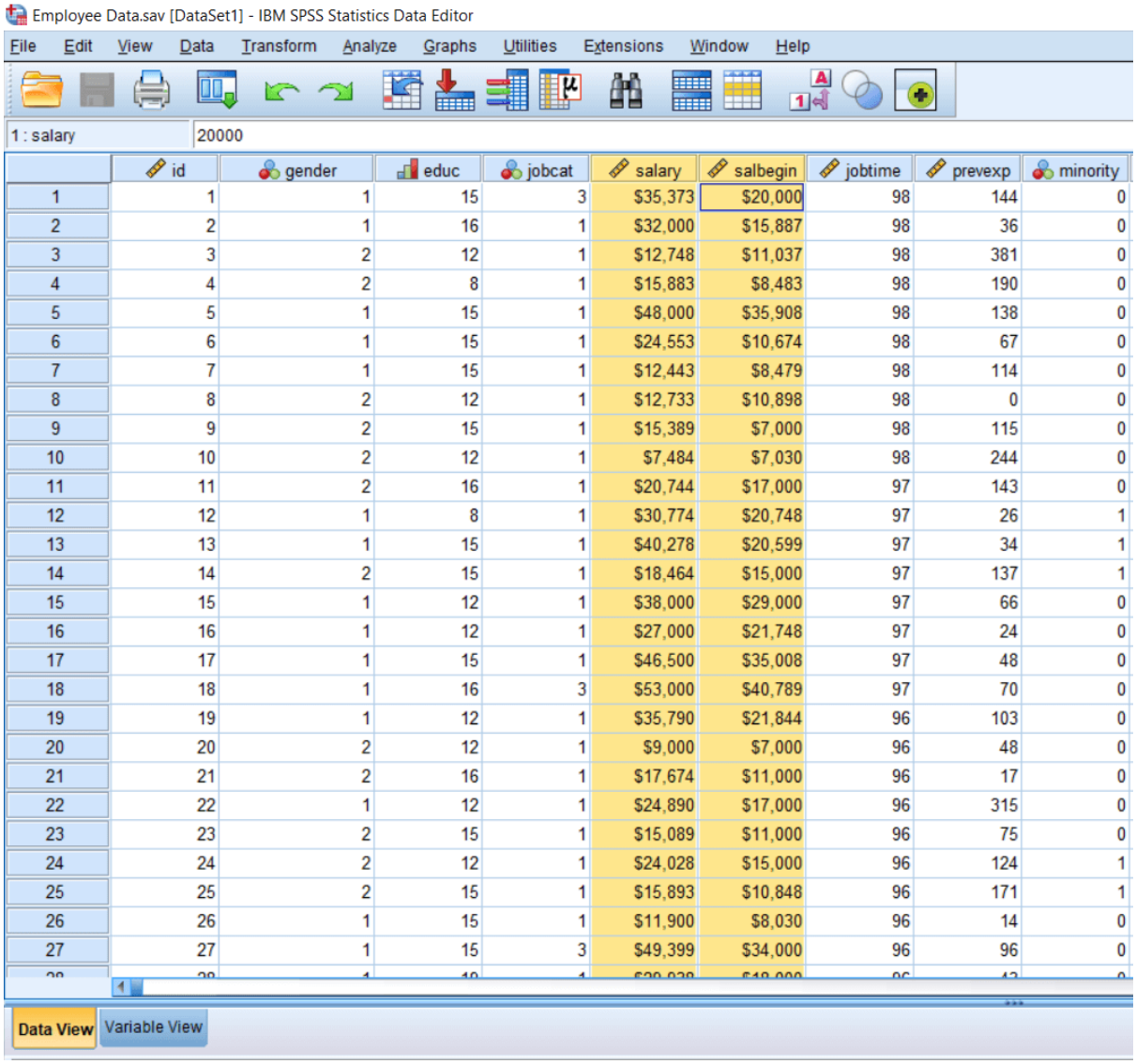
# Calculating Paired Sample t-test

In this section, we will learn about calculating the **Paired Sample T-test**. To calculate it, we will take the **Employee data set**. In this data set, we have the id of **employees**, their **gender, education, job category, current salary**, and **beginning salary**.



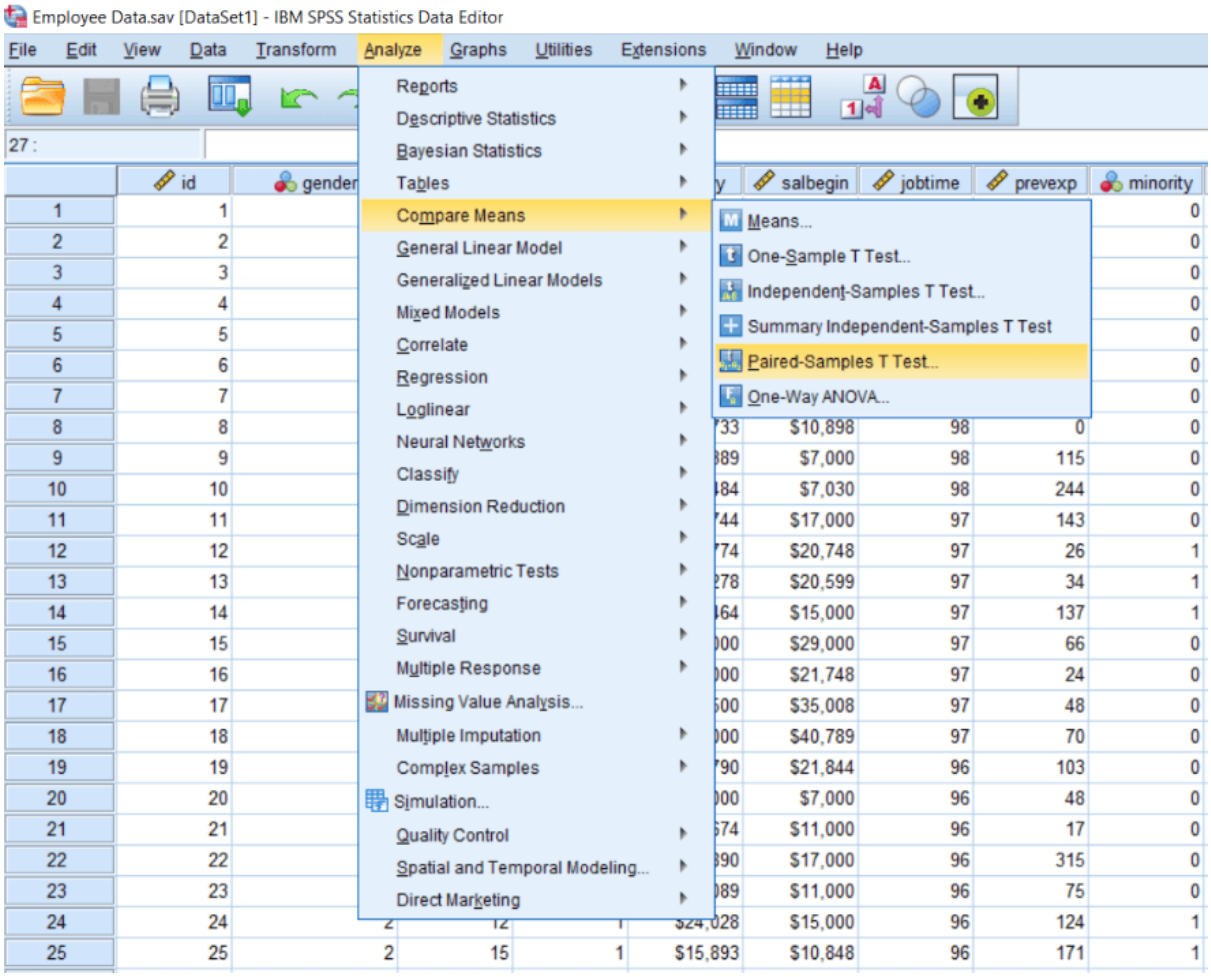

| | id | gender | educ | jobcat | salary | salbegin | jobtime | prevexp | minority |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 15 | 3 | $35,373 | $20,000 | 98 | 144 | 0 |
| 2 | 2 | 1 | 16 | 1 | $32,000 | $15,887 | 98 | 36 | 0 |
| 3 | 3 | 2 | 12 | 1 | $12,748 | $11,037 | 98 | 381 | 0 |
| 4 | 4 | 2 | 8 | 1 | $15,883 | $8,483 | 98 | 190 | 0 |
| 5 | 5 | 1 | 15 | 1 | $48,000 | $35,908 | 98 | 138 | 0 |
| 6 | 6 | 1 | 15 | 1 | $24,553 | $10,674 | 98 | 67 | 0 |
| 7 | 7 | 1 | 15 | 1 | $12,443 | $8,479 | 98 | 114 | 0 |
| 8 | 8 | 2 | 12 | 1 | $12,733 | $10,898 | 98 | 0 | 0 |
| 9 | 9 | 2 | 15 | 1 | $15,389 | $7,000 | 98 | 115 | 0 |
| 10 | 10 | 2 | 12 | 1 | $7,484 | $7,030 | 98 | 244 | 0 |
| 11 | 11 | 2 | 16 | 1 | $20,744 | $17,000 | 97 | 143 | 0 |
| 12 | 12 | 1 | 8 | 1 | $30,774 | $20,748 | 97 | 26 | 1 |
| 13 | 13 | 1 | 15 | 1 | $40,278 | $20,599 | 97 | 34 | 1 |
| 14 | 14 | 2 | 15 | 1 | $18,464 | $15,000 | 97 | 137 | 1 |
| 15 | 15 | 1 | 12 | 1 | $38,000 | $29,000 | 97 | 66 | 0 |
| 16 | 16 | 1 | 12 | 1 | $27,000 | $21,748 | 97 | 24 | 0 |
| 17 | 17 | 1 | 15 | 1 | $46,500 | $35,008 | 97 | 48 | 0 |
| 18 | 18 | 1 | 16 | 3 | $53,000 | $40,789 | 97 | 70 | 0 |
| 19 | 19 | 1 | 12 | 1 | $35,790 | $21,844 | 96 | 103 | 0 |
| 20 | 20 | 2 | 12 | 1 | $9,000 | $7,000 | 96 | 48 | 0 |
| 21 | 21 | 2 | 16 | 1 | $17,674 | $11,000 | 96 | 17 | 0 |
| 22 | 22 | 1 | 12 | 1 | $24,890 | $17,000 | 96 | 315 | 0 |
| 23 | 23 | 2 | 15 | 1 | $15,089 | $11,000 | 96 | 75 | 0 |
| 24 | 24 | 2 | 12 | 1 | $24,028 | $15,000 | 96 | 124 | 1 |
| 25 | 25 | 2 | 15 | 1 | $15,893 | $10,848 | 96 | 171 | 1 |
| 26 | 26 | 1 | 15 | 1 | $11,900 | $8,030 | 96 | 14 | 0 |
| 27 | 27 | 1 | 15 | 3 | $49,399 | $34,000 | 96 | 96 | 0 |

Suppose the employees or managements want to know whether there has been any significant **improvement** in the employee's **salary** since they joined this company. There have been significant improvements in their **current salary** compared to their baseline **salary** or the salary they took at the time of joining. So, in this case, we have repeated measure kind of situation where we have an **id,** and we have **two observations** for that id. One is for the cur**rent salary,** and another is for the **beginning salary**. So for all ids, we have two observations.

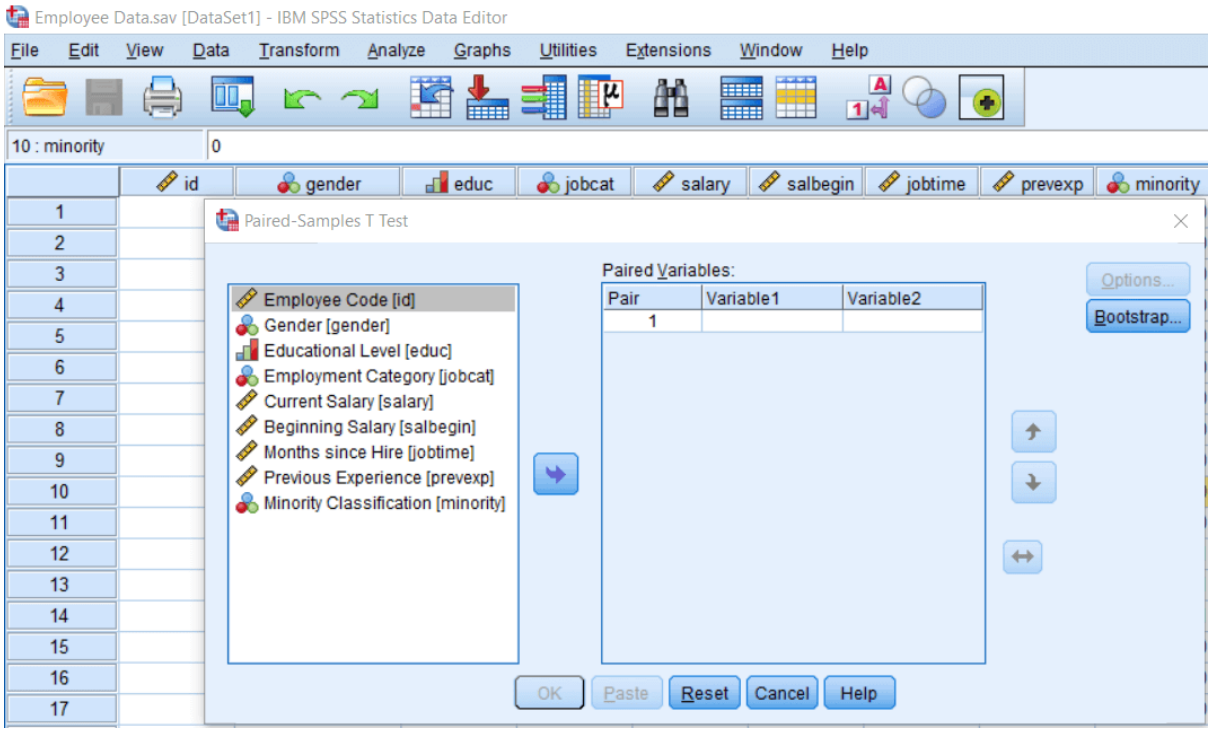



| | id | gender | educ | jobcat | salary | salbegin | jobtime | prevexp | minority |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 15 | 3 | $35,373 | $20,000 | 98 | 144 | 0 |
| 2 | 2 | 1 | 16 | 1 | $32,000 | $15,887 | 98 | 36 | 0 |
| 3 | 3 | 2 | 12 | 1 | $12,748 | $11,037 | 98 | 381 | 0 |
| 4 | 4 | 2 | 8 | 1 | $15,883 | $8,483 | 98 | 190 | 0 |
| 5 | 5 | 1 | 15 | 1 | $48,000 | $35,908 | 98 | 138 | 0 |
| 6 | 6 | 1 | 15 | 1 | $24,553 | $10,674 | 98 | 67 | 0 |
| 7 | 7 | 1 | 15 | 1 | $12,443 | $8,479 | 98 | 114 | 0 |
| 8 | 8 | 2 | 12 | 1 | $12,733 | $10,898 | 98 | 0 | 0 |
| 9 | 9 | 2 | 15 | 1 | $15,389 | $7,000 | 98 | 115 | 0 |
| 10 | 10 | 2 | 12 | 1 | $7,484 | $7,030 | 98 | 244 | 0 |
| 11 | 11 | 2 | 16 | 1 | $20,744 | $17,000 | 97 | 143 | 0 |
| 12 | 12 | 1 | 8 | 1 | $30,774 | $20,748 | 97 | 26 | 1 |
| 13 | 13 | 1 | 15 | 1 | $40,278 | $20,599 | 97 | 34 | 1 |
| 14 | 14 | 2 | 15 | 1 | $18,464 | $15,000 | 97 | 137 | 1 |
| 15 | 15 | 1 | 12 | 1 | $38,000 | $29,000 | 97 | 66 | 0 |
| 16 | 16 | 1 | 12 | 1 | $27,000 | $21,748 | 97 | 24 | 0 |
| 17 | 17 | 1 | 15 | 1 | $46,500 | $35,008 | 97 | 48 | 0 |
| 18 | 18 | 1 | 16 | 3 | $53,000 | $40,789 | 97 | 70 | 0 |
| 19 | 19 | 1 | 12 | 1 | $35,790 | $21,844 | 96 | 103 | 0 |
| 20 | 20 | 2 | 12 | 1 | $9,000 | $7,000 | 96 | 48 | 0 |
| 21 | 21 | 2 | 16 | 1 | $17,674 | $11,000 | 96 | 17 | 0 |
| 22 | 22 | 1 | 12 | 1 | $24,890 | $17,000 | 96 | 315 | 0 |
| 23 | 23 | 2 | 15 | 1 | $15,089 | $11,000 | 96 | 75 | 0 |
| 24 | 24 | 2 | 12 | 1 | $24,028 | $15,000 | 96 | 124 | 1 |
| 25 | 25 | 2 | 15 | 1 | $15,893 | $10,848 | 96 | 171 | 1 |
| 26 | 26 | 1 | 15 | 1 | $11,900 | $8,030 | 96 | 14 | 0 |
| 27 | 27 | 1 | 15 | 3 | $49,399 | $34,000 | 96 | 96 | 0 |

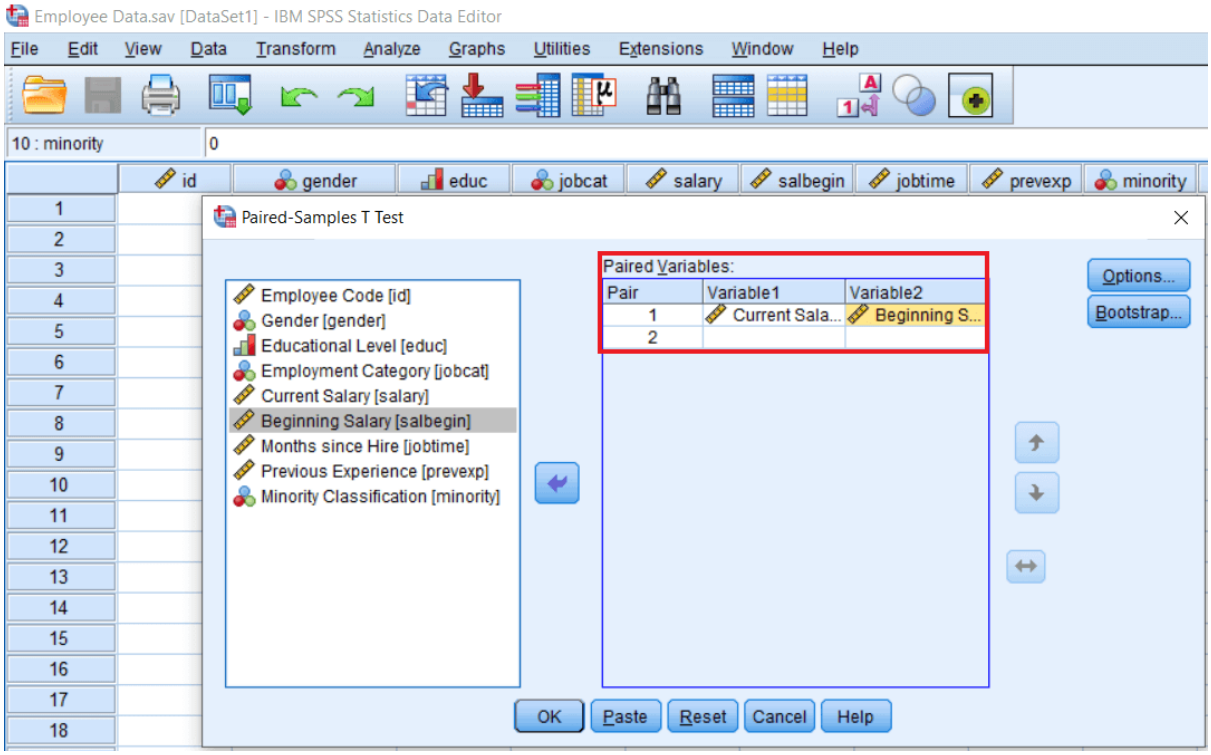
To conduct a **Paired sample t-test**, we will go to the **Analyze menu** and **Compare Means** and look at the **Paired Sample T-test**.




The symbol of **Paired sample T-test** reads as **t A1-A2**. It means the groups A1 and A2 are part of the **same group**. They are the **same individuals** just they have been measured in two different situations. If we look at the symbol of the **Independent sample t-test**, we find t A-B. It means they are **different individuals,** as people from Delhi and Mumbai. Now we will select the **Paired sample t-test** option, and the following dialog box will open:

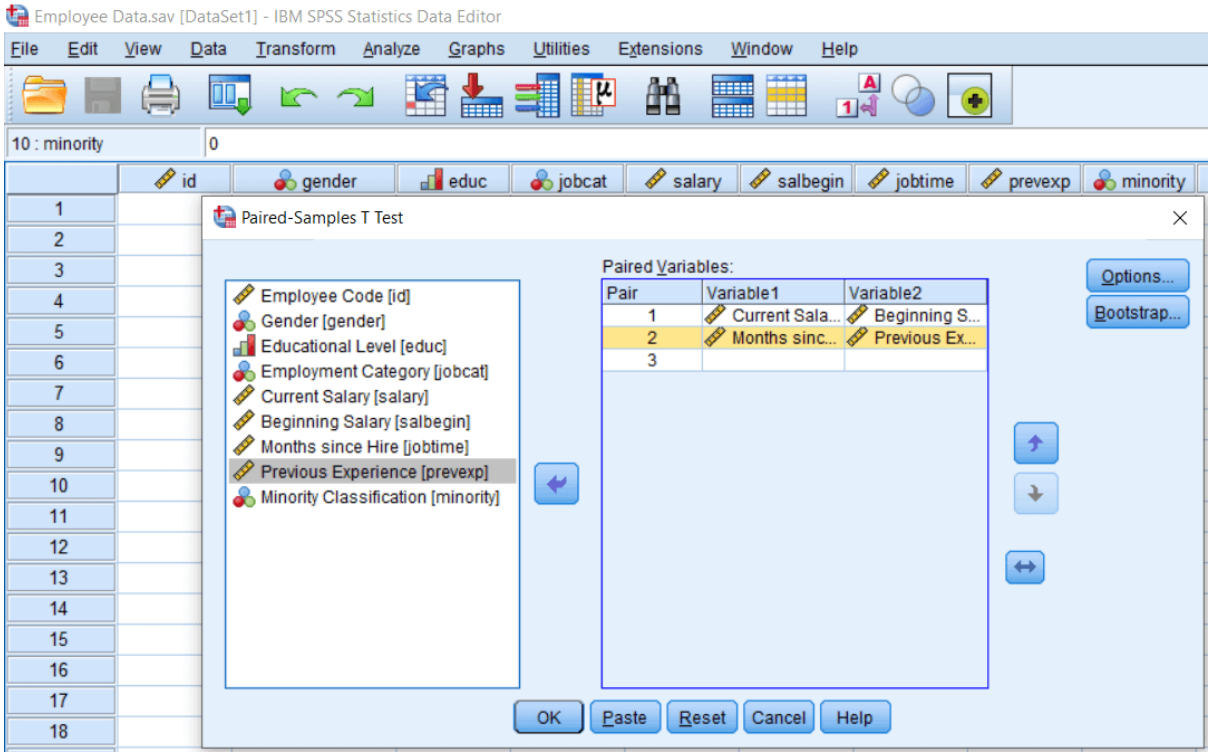



Now we have to define our **pairs**. So our **first pairs** are the **Current salary** of the employees and the **Beginning salary** of the employees. We will shift the **Current salary** in **Variable 1** and **Beginning salary** in **Variable 2** like this:

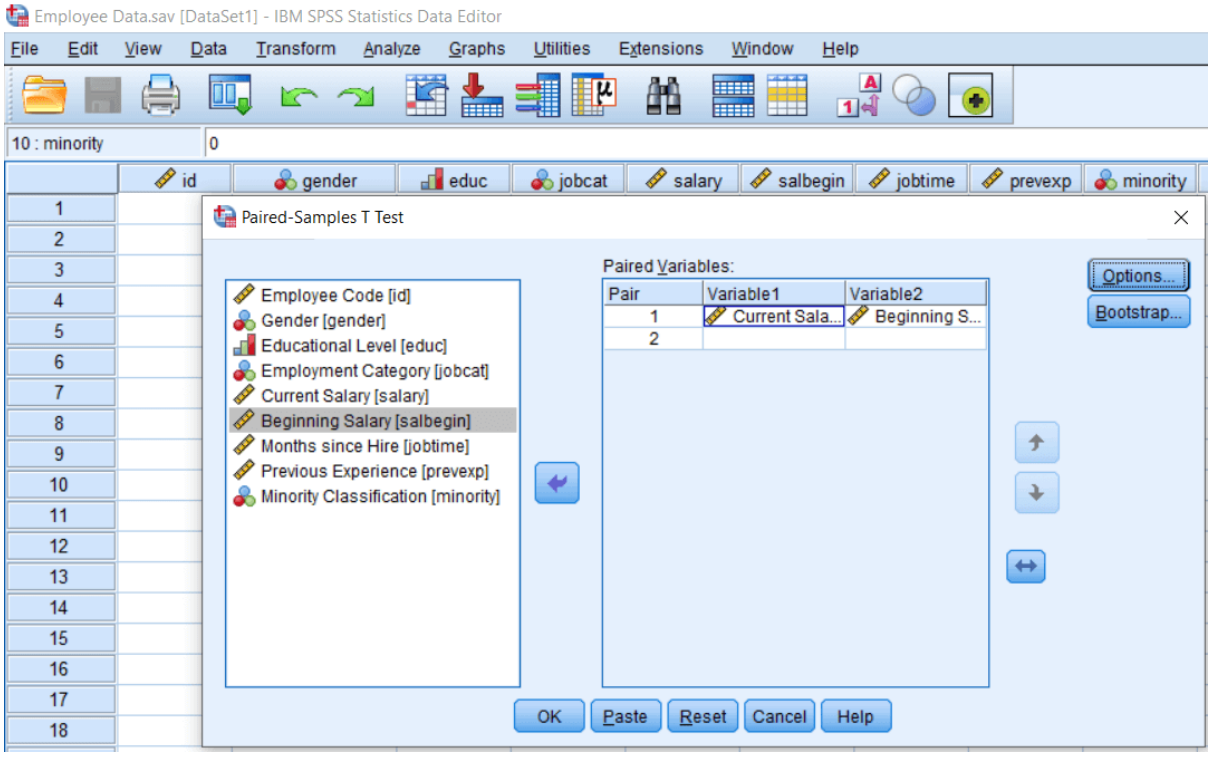



We are finding the differences between the **Current salary** and the **Beginning salary**. So that's our **Pair 1**. If we have more pairs, we can make multiple pair comparison by

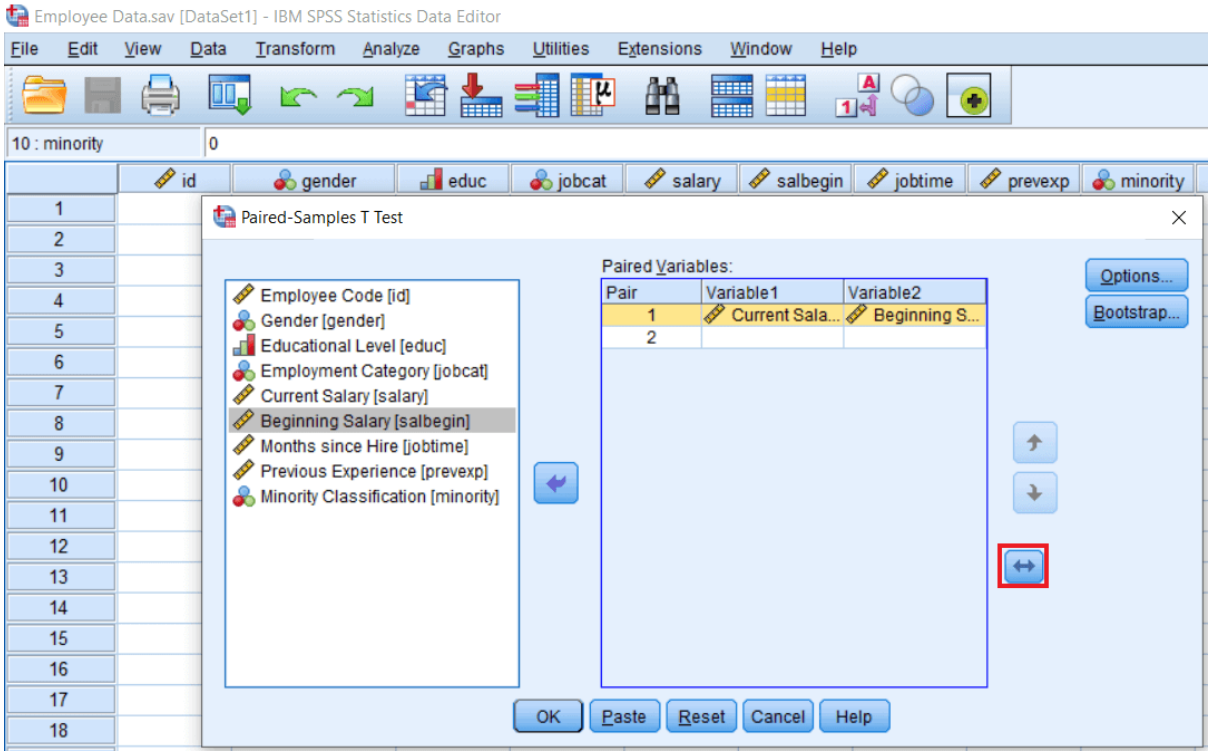
defining our pairs in **Pair 2**. In Pair 2, we are defining joining month in variable 1 and previous experience in variable 2 only for the learning purpose.



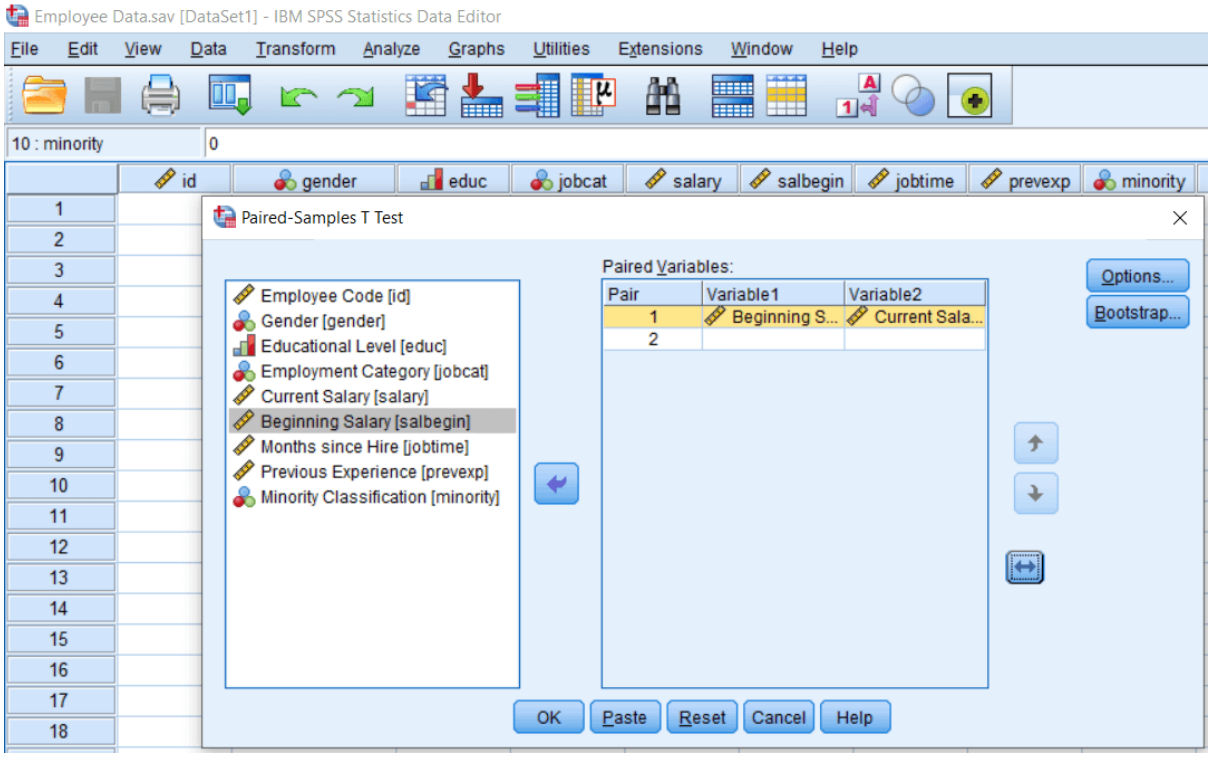

These are not meaningful **Pairs**. Once we select pair 2, we can see that pair 3 is automatically activated. Pair 2 is not meaningful, so we are going to remove it. So we are taking just one pair to compare.



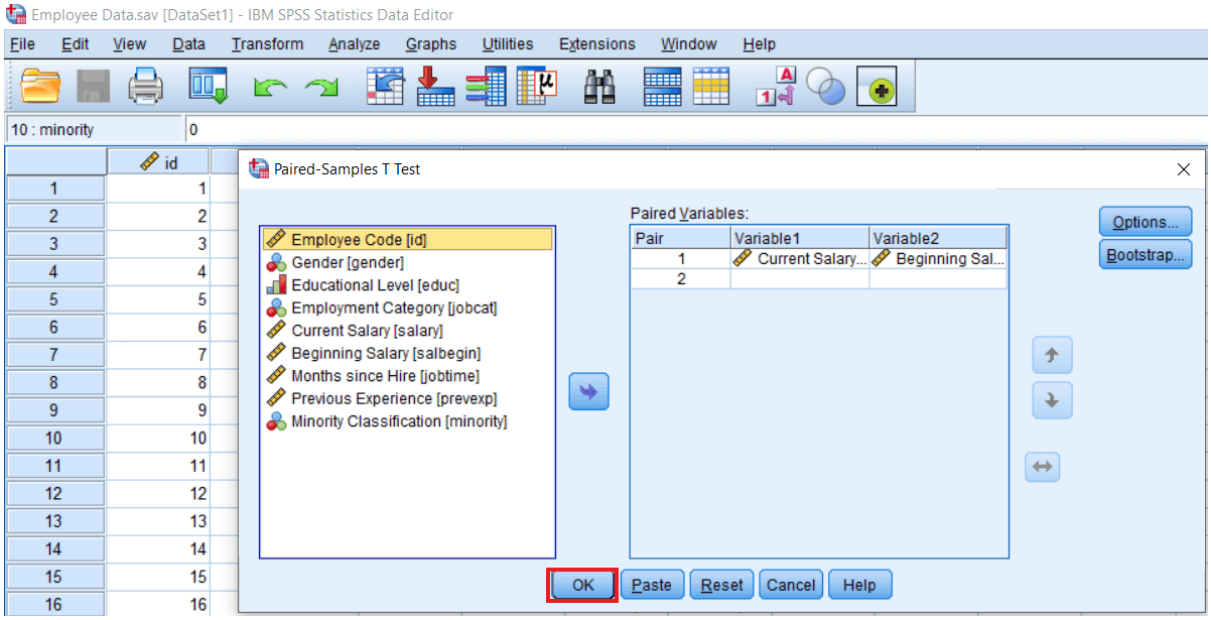
Rests of the **Options** are the same as the **Independent sample t-test**. If we want to know more about them, we can look at the **Independent sample t-test file**. In the above image, an **arrow** can be used to change the **position** of our variables. **For example**, the current salary is our first variable, and beginning salary is our second variable. But if we want to change their position, we can **select** the **pair** and click on this **arrow** tab like this:




When we click on the **arrow tab**, we can see the **beginning salary** as our **first** variable, and the **current salary** has become our **second** variable like this:

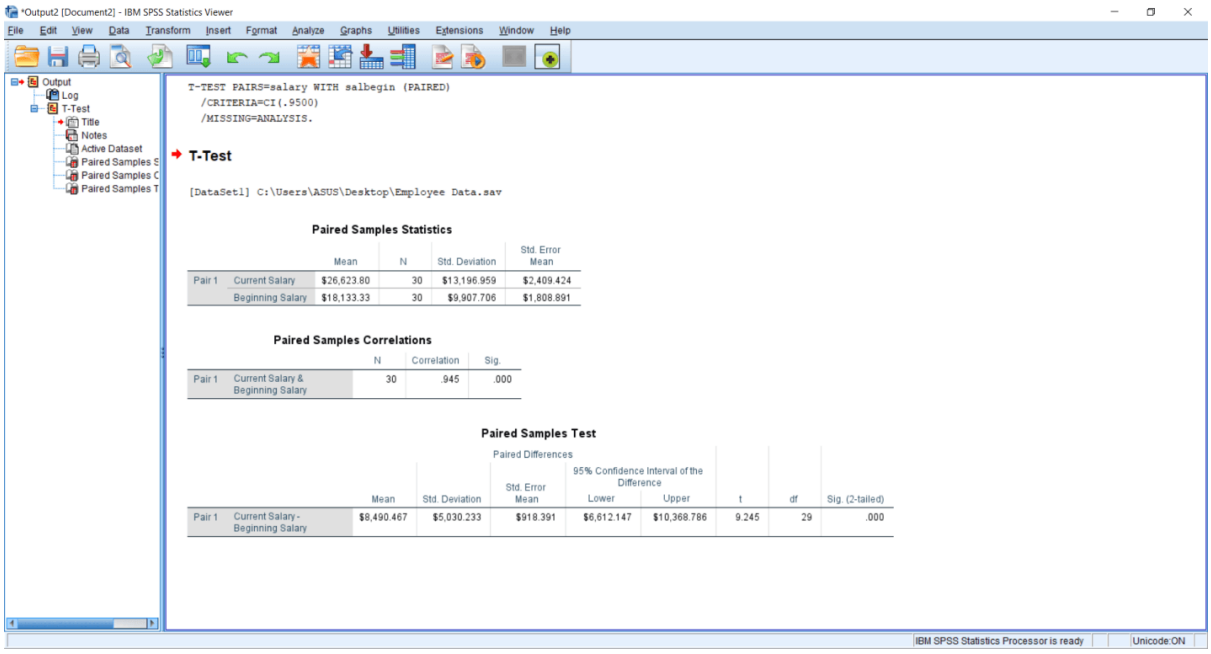



The only difference is going to make in the sign of our data. It's going to give us either a **positive** value or **negative** value depending upon the **situation. For example**, we know that **current employees** must be drawing **more salary** than their **beginning salary**. So if we take the **beginning salary** as our **first variable**, we will get a **negative** test. Remember, the formula of **t** is the difference between **Group Means** divided by **Standard Error**. So the difference between **Group Means** is going to a **negative** value because the **beginning salary** was **lesser** than the **current salary**. We want to avoid the **negativity** value, so in that case, we can take our **current salary** as the **first** variable. It is not going to have any implication for our result or actual conclusion that we draw. It is only going to **change** the **sign** of our t-**test value**. Now we will click on the Ok button:



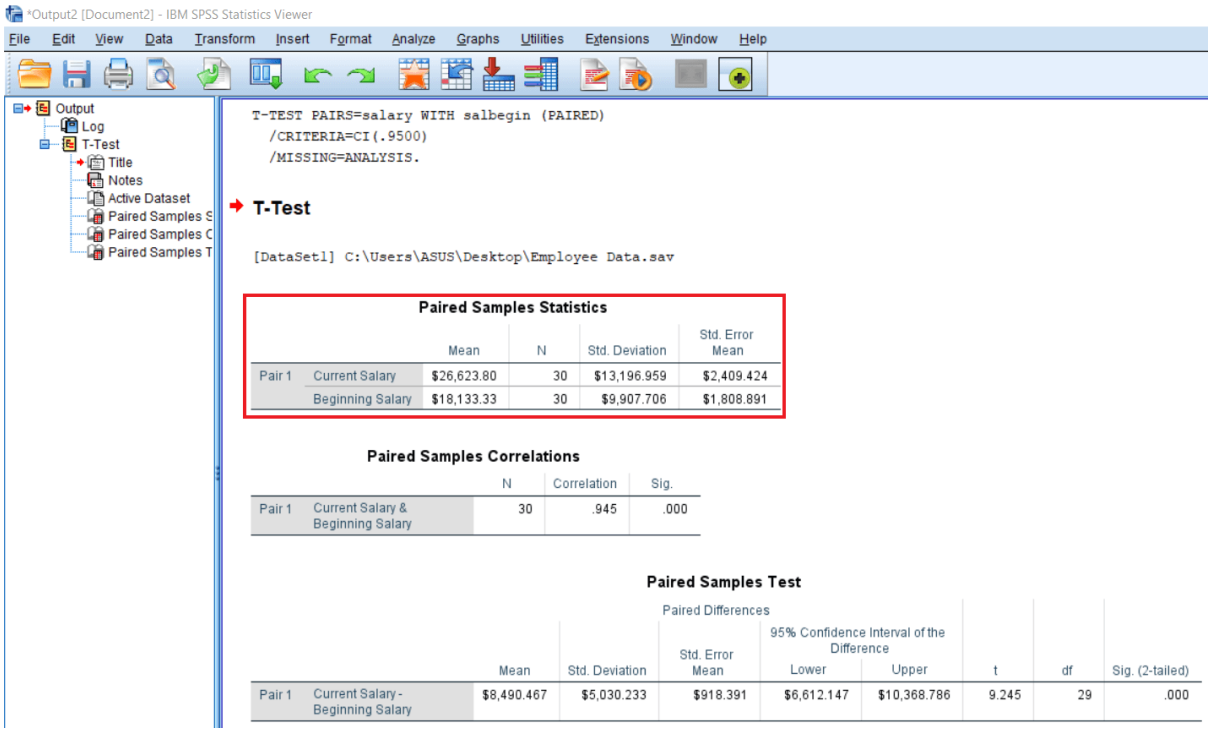

# Output of Paired Sample T-test

In this section, we will discuss the **Output** of the **Paired sample t-test**. Output of the Paired sample t-test is given below, which is the output of the previous **Calculating Paired sample T-test** file:

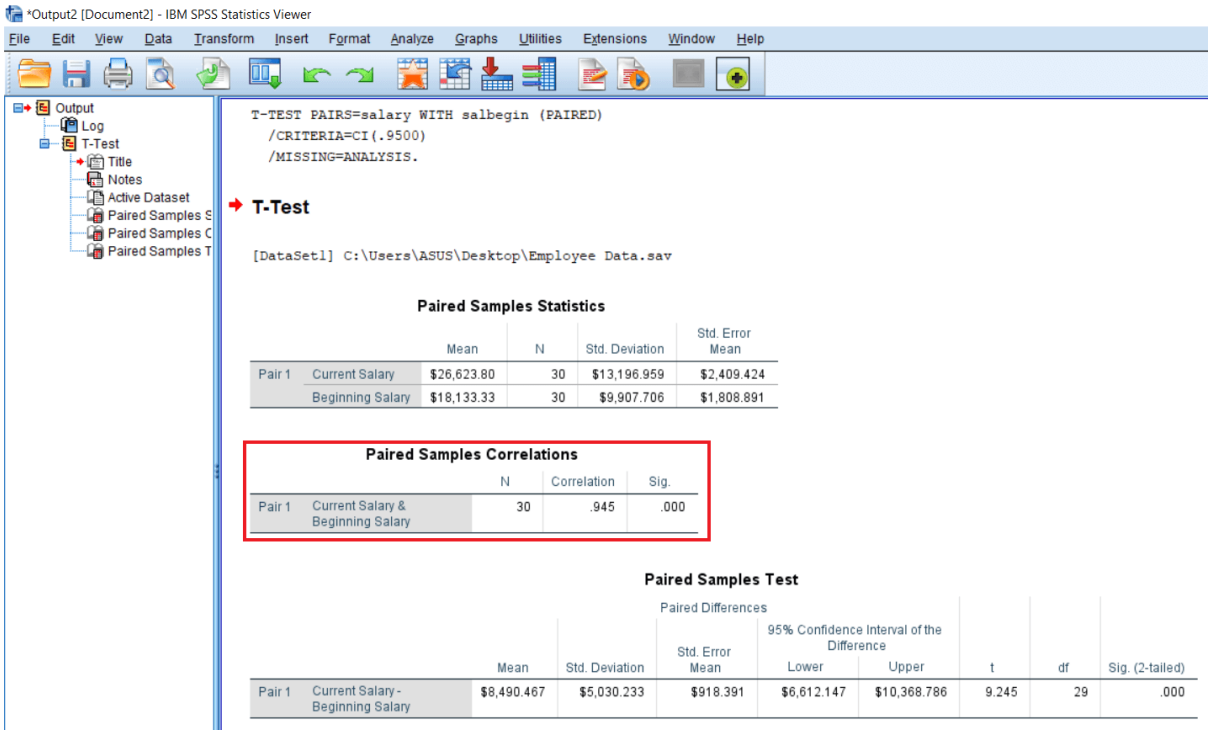



This is the **descriptive output**. So we can see the **average salary.** Currently, it is **26** thousand **623.80** dollars while it was **18** thousand **133.33** dollars at the **beginning**. So the salary has approximately **doubled** since the employee joined

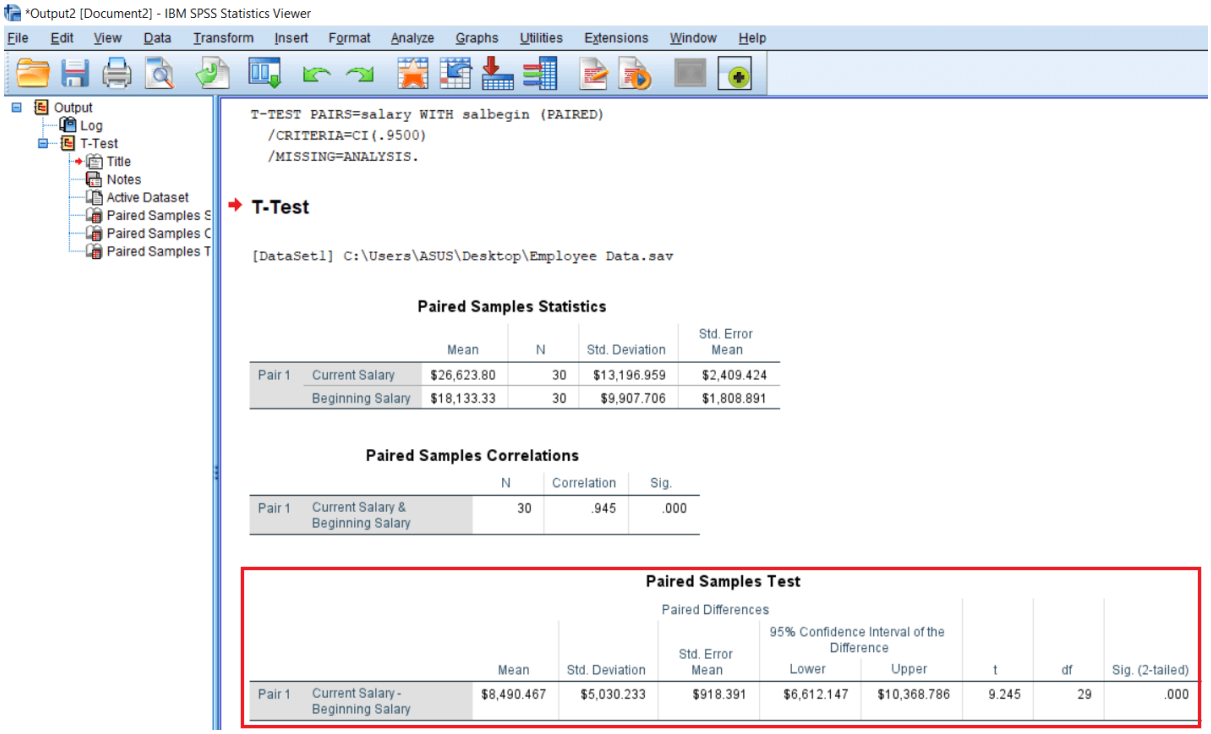
this company. We have **30** in the **first** group and **30** in the **second** group because they are the same individuals in two different situations. In **Standard Deviation**, the current salary is **13 thousand 196.959** dollars, and the beginning salary is **9 thousand 907.706** dollars. The **Current salary** is much **higher** as compared to their **beginning salary**. It means the salary has increased, but there is a huge variation in the salary. The **current salary** is **2** thousand **409.424** dollars in the **Standard error,** and the **beginning salary** is **1** thousand **808.891** dollars, as shown below:




When we calculate the **Paired sample T-test**, we also get a **correlation** because since we are taking the subject into different situations. So there is bound to be a significant correlation between them. In this case, the **correlation** is **.945,** and that's a very strong correlation. **Significant data** is equal to **.001** labels. So we are getting a good correlation between the salary scores. So it means those persons who are getting a higher salary earlier also getting a higher salary now. So there is a pattern in the increase or decrease in salary.




The following table is our final table. So, the average salary for the pair **Current Vs. Beginning** is **8 thousand 490.467** dollars. This **value** shows the **difference** between salaries. **Mean**s refers to the main difference between salaries, i.e., **26,623.80-18,133.33**. The **Standard Deviation** is very high, which is **5 thousand 030.233**. That is the cause of concern, but it's much below the average salary, so we can accept that. The **Standard error** of mean is **918.391. 95% confidential** interval is **6 thousand 612** to **10 thousand 368**. Both the confidence intervals are **positive**. So we can consider our outcome reliable. We were not expecting a zero value in this 95% confidence interval. The **t** value is **9.245**, the **degree** of freedom is **29**, and the result is a **significant tailed** file equal to **.001** labels. So it means there has been significant improvement in the salary of employees since they joined the company.




**Paired Samples Statistics**

| | | Mean | N | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Pair 1 | Current Salary | $26,623.80 | 30 | $13,196.959 | $2,409.424 |
| | Beginning Salary | $18,133.33 | 30 | $9,907.706 | $1,808.891 |

**Paired Samples Correlations**

| | | N | Correlation | Sig. |
|---|---|---|---|---|
| Pair 1 | Current Salary & Beginning Salary | 30 | .945 | .000 |

**Paired Samples Test**

| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference Lower | Upper | | | |
| Pair 1 | Current Salary - Beginning Salary | $8,490.467 | $5,030.233 | $918.391 | $6,612.147 | $10,368.786 | 9.245 | 29 | .000 |