TWO WAY ANOVA CLLASIFICATION

The **"two-way ANOVA classification"** refers to the fact that the data is classified based on two categorical independent variables, allowing for the analysis of both main effects and interaction effects on the dependent variable.

[10]:
```
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

[11]:
```
model = ols('salary ~ C(gender) + C(ssc_b) + C(gender):C(ssc_b)', data=df).fit()
```

[12]:
```
anova_table = sm.stats.anova_lm(model, typ=2)
```

[13]:
```
print(anova_table)
```

```
                       sum_sq     df         F    PR(>F)
C(gender)        9.640996e+10    1.0  4.303801  0.039240
C(ssc_b)         7.570114e+09    1.0  0.337935  0.561644
C(gender):C(ssc_b)  2.950021e+09    1.0  0.131691  0.717049
Residual         4.726636e+12  211.0       NaN       NaN
```

C(gender):

Sum of Squares (sum_sq): 9.640996e+10

Degrees of Freedom (df): 1.0

F-statistic (F): 4.303801

p-value (Pr(>F)): 0.039240

Interpretation:

The **p-value of 0.039240 is less** than the commonly used significance level of 0.05.

This indicates that the **'gender' variable has a statistically significant effect on the response variable 'salary'**.

The null hypothesis that 'gender' has no effect on 'salary' can be rejected, and the **alternative hypothesis that 'gender' does have an effect on 'salary' is supported by the data.**

The F-statistic of 4.303801 represents the **ratio of the variance between the 'gender' groups to the variance within the 'gender' groups**. A larger F-statistic suggests a greater difference between the group means, which is the case here.

The **degrees of freedom (df)** of 1.0 for the 'gender' variable indicates that there are two levels or categories of the 'gender' variable in the analysis (e.g., male and female).

The sum of squares (sum_sq) of 9.640996e+10 represents the amount of variation in the 'salary' variable that is explained by the 'gender' variable.

In summary, the ANOVA output shows that **the 'gender' Independent variable has a statistically significant effect on the 'salary' Dependent variable**, as indicated by the p-value of 0.039240, which is less than the commonly used significance level of 0.05.

C(ssc_b):

Sum of Squares (sum_sq): 7.570114e+09

Degrees of Freedom (df): 1.0

F-statistic (F): 0.337935

p-value (Pr(>F)): 0.561644

Interpretation:

The p-value of 0.561644 is greater than the commonly used significance level of 0.05.

This indicates that **the 'ssc_b' (secondary school board) variable does not have a statistically significant effect on the response variable 'salary'**.

The null hypothesis that 'ssc_b' has no statistically significant effect on 'salary' cannot be rejected based on the provided data and analysis.

The **F-statistic of 0.337935** represents the ratio of the variance between the 'ssc_b' groups to the variance within the 'ssc_b' groups. **A smaller F-statistic suggests that the difference between the group means is not large enough to be considered statistically significant.**

The degrees of freedom (df) of 1.0 for the 'ssc_b' variable indicates that there are two levels or categories of the 'ssc_b' variable in the analysis (e.g., central board and others).

The sum of squares (sum_sq) of 7.570114e+09 represents the amount of variation in the 'salary' variable that is explained by the 'ssc_b' variable, which is relatively small compared to the total variation.

In summary, the ANOVA output shows that the **'ssc_b' variable does not have a statistically significant effect on the 'salary' variable**, as indicated by the p-value of 0.561644, which is greater than the commonly used significance level of 0.05.

C(gender):C(ssc_b):

Sum of Squares (sum_sq): 2.950021e+09

Degrees of Freedom (df): 1.0

F-statistic (F): 0.131691

p-value (Pr(>F)): 0.717049

Interpretation:

The p-value of 0.717049 is greater than the commonly used significance level of 0.05.

This suggests that the interaction effect between **'gender' and 'ssc_b' is not statistically significant.**

The **null hypothesis that the interaction effect is zero** cannot be rejected based on the provided data and analysis.

The **F-statistic of 0.131691** indicates a small ratio of the variance between the interaction effect to the variance within the interaction effect groups.

In summary, the ANOVA output suggests that the interaction effect between **'gender' and 'ssc_b' does not have a statistically significant impact on the response variable based on the p-value of 0.717049**, which is greater than the commonly used significance level of 0.05.