

15_two_way_anova

August 1, 2025

1 Two way Anova

Kategorička nezavisna varijabla Department ((Marketing, Logistics, Sales)

Kategorička nezavisna varijabla Experience Level (Junior, Mid, Senior)

Numerička zavisna varijabla Sales

Također zanima nas postoji li interakcija između dvije nezavisne varijable, odnosno varira li utjecaj odjela na plaću ovisno o iskustvu

```
[1]: import pandas as pd
import numpy as np
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

```
[2]: np.random.seed(42)
```

```
[3]: departments = np.random.choice(['Marketing', 'Logistics', 'Sales'], 50)
experience_levels = np.random.choice(['Junior', 'Mid', 'Senior'], 50)
```

```
[4]: salary = np.random.normal(loc=50000, scale=5000, size=50) + \
      (np.where(departments == 'Marketing', 2000, 0)) + \
      (np.where(departments == 'Sales', -3000, 0)) + \
      (np.where(experience_levels == 'Senior', 5000, 0)) + \
      (np.where(experience_levels == 'Junior', -5000, 0))
```

```
[5]: df = pd.DataFrame({
    'Department': departments,
    'Experience': experience_levels,
    'Salary': salary
})
```

```
[6]: df.head()
```

```
[6]:   Department Experience      Salary
0      Sales      Mid  49910.613974
1  Marketing  Junior  51438.742298
2      Sales      Mid  51471.661651
3      Sales  Junior  45774.988986
```

4 Marketing Mid 50964.170549

```
[7]: df_depart=df.groupby("Department")
```

```
[8]: df_depart['Salary'].mean()
```

```
[8]: Department
Logistics    48198.185959
Marketing    54369.758458
Sales        46454.917331
Name: Salary, dtype: float64
```

```
[9]: model = ols('Salary ~ C(Department) + C(Experience) + C(Department):
↳C(Experience)', data=df).fit()
```

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

```
[11]: anova_table
```

```
[11]:
```

	sum_sq	df	F	PR(>F)
C(Department)	3.544608e+08	2.0	7.097703	0.002255
C(Experience)	5.870328e+08	2.0	11.754713	0.000092
C(Department):C(Experience)	1.016781e+08	4.0	1.017999	0.409391
Residual	1.023774e+09	41.0	NaN	NaN

Za odjel F-statistics iznosi 7.097703, a P_value iznosi 0.002255 = značajan utjecaj odjela na iznos plaće postoji statistički značajna razlika u prosječnim plaćama između različitih odjela, neovisno o razini iskustva

Za iskustvo F-statistics iznosi 11.754713, a P_value iznosi 0.000092 = značajan utjecaj iskustva na iznos plaće

postoji statistički značajna razlika u prosječnim plaćama između različitih razina iskustva (Junior, Mid, Senior), neovisno o odjelu

Odjel vs iskustvo F-statistics iznosi 1.017999, a P_value iznosi 0.409391 = nemamo značajan utjecaj iskustva na odjel

Što to znači? To znači da efekt iskustva na plaću je relativno sličan u svim odjelima, i obrnuto, efekt odjela na plaću je relativno sličan za sve razine iskustva. Drugim riječima, ne postoji poseban “sinergijski” ili “antagonistički” efekt gdje bi se utjecaj iskustva na plaću dramatično mijenjao ovisno o odjelu (ili obrnuto).

1.1 “Na temelju ove dvosmjerne ANOVA analize, možemo s visokom sigurnošću zaključiti da:

- Odjel značajno utječe na iznos plaće. Postoje statistički značajne razlike u prosječnim plaćama između različitih odjela.
- Razina iskustva značajno utječe na iznos plaće. Postoje statistički značajne razlike u prosječnim plaćama između Junior, Mid i Senior pozicija.

- Nema značajne interakcije između odjela i iskustva. To znači da utjecaj iskustva na plaću je sličan u svim odjelima, i obrnuto. Nema specifičnih kombinacija odjela i iskustva koje bi rezultirale neočekivano visokim ili niskim plaćama izvan onoga što bi se očekivalo na temelju individualnih efekata odjela i iskustva.
- U budućim analizama i modelima za predviđanje plaće, i ‘Department’ i ‘Experience’ su vrlo važne varijable koje treba uključiti, jer svaka od njih neovisno objašnjava značajan dio varijabilnosti u plaćama.”