

```
In [ ]: import pandas as pd
        from scipy.stats import ttest_ind
        from scipy import stats
        import matplotlib as plt
```

```
In [ ]: df = pd.read_csv(r"C:\Users\Mamzi\Desktop\stat\stat_hw.csv")
        df
```

```
Out[ ]:
```

	gender	stress_first	num_ill	grade_pre	final_grade	stress_final
0	boy	30	24	13	16	20
1	boy	27	20	16	20	26
2	boy	9	8	7	18	9
3	boy	20	12	30	35	19
4	boy	3	5	5	5	2
5	boy	15	24	10	15	14
6	boy	5	0	9	10	6
7	boy	10	8	15	24	10
8	boy	24	20	18	30	20
9	boy	34	24	20	27	19
10	girl	5	7	18	17	4
11	girl	15	30	14	10	14
12	girl	5	2	20	20	5
13	girl	10	10	9	10	8
14	girl	25	18	13	9	21
15	girl	8	9	13	16	11
16	girl	14	20	10	1	12
17	girl	11	10	16	14	12
18	girl	24	15	5	6	18
19	girl	5	8	7	14	6

```
In [ ]: df.describe()
```

Out[]:

	stress_first	num_ill	grade_pre	final_grade	stress_final
count	20.000000	20.000000	20.000000	20.000000	20.000000
mean	14.950000	13.700000	13.400000	15.850000	12.800000
std	9.495013	8.329403	6.09918	8.554931	6.685412
min	3.000000	0.000000	5.000000	1.000000	2.000000
25%	7.250000	8.000000	9.000000	10.000000	7.500000
50%	12.500000	11.000000	13.000000	15.500000	12.000000
75%	24.000000	20.000000	16.500000	20.000000	19.000000
max	34.000000	30.000000	30.000000	35.000000	26.000000

In []:

```
boys = df[df['gender'] == 'boy']
girls = df[df['gender'] == 'girl']
```

جدول آمار توصیفی گروه پسران :

این جدول حاوی خلاصه‌ای از آمار توصیفی پنج مجموعه داده برای گروه پسران است. در ادامه، توضیح هر یک از سطرهای این جدول ارائه شده است:

- **count (تعداد):** تعداد مشاهدات در هر مجموعه داده. در این مثال، برای همه ستون‌ها، تعداد مشاهدات برابر با ۱۰ است.
- **mean (میانگین):** میانگین مقادیر هر مجموعه داده. میانگین ستون‌های اول تا پنجم به ترتیب ۱۲.۲، ۱۲.۹، ۱۲.۵، ۱۱.۷ و ۱۱.۱ است.
- **std (انحراف معیار):** انحراف معیار مقادیر هر مجموعه داده. انحراف معیار ستون‌های اول تا پنجم به ترتیب ۷.۴۱، ۸.۰۵، ۴.۷۹، ۵.۶۴ و ۵.۵۷ است.
- **min (کمینه):** کمترین مقدار در هر مجموعه داده. کمینه ستون‌های اول تا پنجم به ترتیب ۵، ۲، ۱ و ۴ است.
- **۲۵٪ (صدک ۲۵):** مقداری که ۲۵ درصد از داده‌ها کمتر از آن و ۷۵ درصد بیشتر از آن هستند. صدک ۲۵ ستون‌های اول تا پنجم به ترتیب ۵.۷۵، ۸.۲۵، ۹.۲۵، ۹.۲۵ و ۶.۵ است.
- **۵۰٪ (میانۀ صدک ۵۰):** مقداری که ۵۰ درصد از داده‌ها کمتر از آن و ۵۰ درصد بیشتر از آن هستند (میانۀ). میانۀ ستون‌های اول تا پنجم به ترتیب ۱۰.۵، ۱۰، ۱۳، ۱۲ و ۱۱.۵ است.

- **۷۵٪ (صدک ۷۵) :** مقداری که ۷۵ درصد از داده‌ها کمتر از آن و ۲۵ درصد بیشتر از آن هستند. صدک ۷۵ ستون‌های اول تا پنجم به ترتیب ۱۴.۷۵، ۱۷.۲۵، ۱۵.۵، ۱۵.۵ و ۱۳.۵ است.
- **max (بیشینه) :** بیشترین مقدار در هر مجموعه داده. بیشینه ستون‌های اول تا پنجم به ترتیب ۲۵، ۳۰، ۲۰، ۲۰ و ۲۱ است.

In []: `boys.describe()`

Out []:

	stress_first	num_ill	grade_pre	final_grade	stress_final
count	10.000000	10.000000	10.000000	10.000000	10.000000
mean	17.700000	14.500000	14.300000	20.000000	14.500000
std	10.893933	8.959787	7.334091	9.189366	7.546154
min	3.000000	0.000000	5.000000	5.000000	2.000000
25%	9.250000	8.000000	9.250000	15.250000	9.250000
50%	17.500000	16.000000	14.000000	19.000000	16.500000
75%	26.250000	23.000000	17.500000	26.250000	19.750000
max	34.000000	24.000000	30.000000	35.000000	26.000000

جدول آمار توصیفی گروه دختران :

In []: `girls.describe()`

Out []:

	stress_first	num_ill	grade_pre	final_grade	stress_final
count	10.000000	10.000000	10.000000	10.000000	10.000000
mean	12.200000	12.900000	12.500000	11.700000	11.100000
std	7.405704	8.047774	4.790036	5.638164	5.566766
min	5.000000	2.000000	5.000000	1.000000	4.000000
25%	5.750000	8.250000	9.250000	9.250000	6.500000
50%	10.500000	10.000000	13.000000	12.000000	11.500000
75%	14.750000	17.250000	15.500000	15.500000	13.500000
max	25.000000	30.000000	20.000000	20.000000	21.000000

آزمون t مستقل :

در اینجا یک آزمون تی-مستقل بین دو گروه مختلف از داده‌ها انجام می‌شود. فرض صفر در اینجا این است که میانگین دو گروه برابر است. به طور خاص، از تابع `ttest_ind` از کتابخانه `scipy.stats` استفاده شده است.

- **test**: نتیجه آزمون t را نشان می‌دهد.
- **pval**: مقدار P-Value مربوط به آزمون t را نشان می‌دهد.

نتیجه آزمون t در `test` ذخیره می‌شود و مقدار P-Value مربوطه در `pval` ذخیره می‌شود. این دو مقدار به طور معمول برای ارزیابی تفاوت معنادار بین میانگین دو گروه استفاده می‌شوند. اگر مقدار P-Value کمتر از سطح معناداری مشخص شده باشد (مثلاً ۰.۰۵)، ما می‌توانیم فرض صفر را رد کنیم و به این نتیجه برسیم که تفاوت معناداری بین میانگین دو گروه وجود دارد.

$$H_0 : \mu_A = \mu_B$$

$$H_1 : \mu_A \neq \mu_B$$

```
In [ ]: test,pval = ttest_ind(boys['stress_first'],girls['stress_first'])
        print('test : ' , test)
        print('P-value : ' , pval)
```

```
test : 1.3203380498166697
P-value : 0.20326609342697904
```

```
In [ ]: test,pval = ttest_ind(boys['num_ill'],girls['num_ill'])
        print('test : ' , test)
        print('P-value : ' , pval)
```

```
test : 0.42011620445891606
P-value : 0.6793728859666204
```

```
In [ ]: test,pval = ttest_ind(boys['grade_pre'],girls['grade_pre'])
        print('test : ' , test)
        print('P-value : ' , pval)
```

```
test : 0.649801145817031
P-value : 0.5240275078282306
```

```
In [ ]: test,pval = ttest_ind(boys['final_grade'],girls['final_grade'])
        print('test : ' , test)
        print('P-value : ' , pval)
```

```
test : 2.434516632931371
P-value : 0.025546913188794387
```

```
In [ ]: test,pval = ttest_ind(boys['stress_final'],girls['stress_final'])
print('test : ' , test)
print('P-value : ' , pval)
```

```
test : 1.1465741568668137
P-value : 0.2665713911906977
```

آزمون t زوجی :

در اینجا به دلیل اینکه همان دانشجویان استرسشان در ابتدای سال و انتهای سال اندازه گرفته شده است مستقل نیستند و باید از آزمون t زوجی استفاده کنیم

- `df['stress_first']` : نشان دهنده داده‌های گروه یا شرایط قبلی است.
- `df['stress_final']` : نشان دهنده داده‌های گروه یا شرایط بعدی است.

و نتیجه این آزمون به صورت دو مقدار `test` و `pval` بازگردانده می‌شود:

- **test**: نتیجه آزمون t را نشان می‌دهد.
- **pval**: مقدار P-Value مربوط به آزمون t را نشان می‌دهد.

مقدار P-Value معمولاً برای ارزیابی معناداری تفاوت بین میانگین دو گروه یا شرایط متناظر استفاده می‌شود. اگر P-Value کمتر از سطح معناداری مشخص شده باشد (مثلاً ۰.۰۵)، می‌توانیم فرض صفر را رد کنیم و به این نتیجه برسیم که تفاوت معناداری بین میانگین دو گروه یا شرایط وجود دارد.

```
In [ ]: test , pval = stats.ttest_rel(df['stress_first'], df['stress_final'])
print('test : ' , test)
print('P-value : ' , pval)
```

```
test : 2.319291915660352
P-value : 0.031673369402023775
```