

How can we calculate studentized residuals in Python?

Authored by
stats writer

April 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How can we calculate studentized residuals in Python?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=138278>

Studentized residuals are a statistical measure used to assess the relative difference between observed and predicted values in a regression model. In Python, this can be calculated by first obtaining the residuals of the model, then dividing each residual by its corresponding standard error. This process is known as studentization and can be easily implemented using built-in functions and libraries such as statsmodels and scipy. By calculating and analyzing studentized residuals, we can evaluate the significance and influence of individual data points on the overall regression model, providing valuable insights for data analysis and model improvement.

Calculate Studentized Residuals in Python

A studentized residual is simply a residual divided by its estimated standard deviation.

In practice, we typically say that any observation in a dataset that has a studentized residual greater than an absolute value of 3 is an outlier.

We can quickly obtain the studentized residuals of a regression model in Python by using the OLSResults.outlier_test() function from statsmodels, which uses the following syntax:

OLSResults.outlier_test()

where OLSResults is the name of a linear model fit using the ols() function from statsmodels.

Example: Calculating Studentized Residuals in Python

Suppose we build the following simple linear regression model in Python:

```
#import necessary packages and functions
```

```
import numpy as np
```

```
import pandas as pd
```

```
import statsmodels.api as sm
```

```
from statsmodels.formula.api import ols
```

```
#create dataset
```

```
df = pd.DataFrame({'rating': ,  
'points': })
```

```
#fit simple linear regression model
```

```
model = ols('rating ~ points', data=df).fit()
```

We can use the `outlier_test()` function to produce a DataFrame that contains the studentized residuals for each observation in the dataset:

```
#calculate studentized residuals
```

```
stud_res = model.outlier_test()
```

```
#display studentized residuals
```

```
print(stud_res)
```

```
student_resid unadj_p bonf(p)
0 -0.486471 0.641494 1.000000
1 -0.491937 0.637814 1.000000
2 0.172006 0.868300 1.000000
3 1.287711 0.238781 1.000000
4 0.106923 0.917850 1.000000
5 0.748842 0.478355 1.000000
6 -0.968124 0.365234 1.000000
7 -2.409911 0.046780 0.467801
8 1.688046 0.135258 1.000000
9 -0.014163 0.989095 1.000000
```

This DataFrame displays the following values for each observation in the dataset:

The studentized residual
The unadjusted p-value of the studentized residual
The Bonferroni-corrected p-value of the studentized residual

We can see that the studentized residual for the first observation in the dataset is -0.486471, the studentized residual for the second observation is -0.491937, and so on.

We can also create a quick plot of the predictor variable

values vs. the corresponding studentized residuals:

```
import matplotlib.pyplot as plt
```

```
#define predictor variable values and studentized residuals
```

```
x = df
```

```
y = stud_res
```

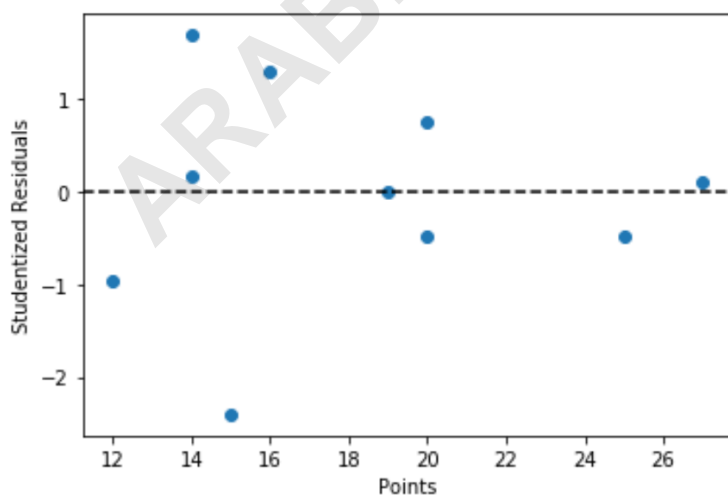
```
#create scatterplot of predictor variable vs. studentized residuals
```

```
plt.scatter(x, y)
```

```
plt.axhline(y=0, color='black', linestyle='--')
```

```
plt.xlabel('Points')
```

```
plt.ylabel('Studentized Residuals')
```



From the plot we can see that none of the observations have a studentized residual with an absolute value greater than 3, thus there are no clear outliers in the dataset.

[How to Perform Simple Linear Regression in Python](#)

[How to Perform Multiple Linear Regression in Python](#)

[How to Create a Residual Plot in Python](#)

ARABPSYCHOLOGY.COM