Application of linear regression in statistical problem programming

Eshboyev Shuhrat

Graduate of Termez State University

Annotation: this thesis considers the method of applying linear regression to programming a statistical issue using Python language libraries. In this case, an example of a model class is created, and the use of The Patsy formula is given.

Kalit so'zlar: *Statmodels, Patsy, Fit usuli.*

Statsmodels kutubxonasi turli vaziyatlarda qo'llaniladigan bir necha turdagi statistik modellarni qo'llab-quvvatlaydi, lekin deyarli barchasi bir xil foydalanish sxemasiga amal qiladi, bu esa turli modellar o'rtasida almashishni osonlashtiradi. Statistik modellardagi statistik modellar model sinflari bilan ifodalanadi. Ular chiziqli modelning javob va tushuntirish o'zgaruvchilari uchun dizayn matritsalari yoki Patsy formulasi va ma'lumotlar ramkasi (yoki lug'atga o'xshash boshqa ob'ekt) berilgan holda boshlanishi mumkin. Statistik modellar bilan statistik modelni o'rnatish va tahlil qilishda asosiy ish jarayoni quyidagi bosqichlarni o'z ichiga oladi:

1. Model sinfining namunasi yaratiladi, masalan, model = sm.MODEL(y, X) yoki model = smf.model(formula, ma'lumotlar) dan foydalanib, bu yerda MODEL va model ma'lum bir model nomidir, masalan, OLS, GLS, Logit va boshqalar. Bu yerda konvensiya shundan iboratki, dizayn matritsalarini argument sifatida qabul qiladigan sinflar uchun katta harflar va Patsy formulalarini hamda ma'lumotlar ramkalarini argument sifatida oladigan sinflar uchun kichik harflar nomlari ishlatiladi.

2. Model namunasini yaratish hech qanday hisob-kitoblarni amalga oshirmaydi. Modelni ma'lumotlarga moslashtirish uchun moslashtirish usulini

chaqirishimiz kerak, natija = model.fit(), bu moslikni bajaradi va keyingi tahlil qilish uchun usullar va atributlarga ega bo'lgan natija ob'ektini qaytaradi.

3. Fit usuli bilan qaytarilgan natija ob'ekti uchun umumiy statistikani chop etadi. Natija ob'ekti har bir statistik model uchun mazmunan bir oz farq qiladi, lekin ko'pchilik modellar statistik model ma'lumotlarni muvaffaqiyatli tushuntirib berishini baholash uchun foydali bo'lishi mumkin bo'lgan bir necha turdagi statistik ma'lumotlarni o'z ichiga olgan, moslik natijasini tavsiflovchi xulosa matnini ishlab chiqaradigan usul xulosasini amalga oshiradi. Xulosa usulidan olingan natijani ko'rish odatda moslashtirish jarayonining natijasini tahlil qilishda yaxshi boshlanish nuqtasidir.

4. Modelni moslashtirish natijalarini qayta ishlashdan keyingi jarayon: Xulosa usulidan tashqari, natija ob'ekti o'rnatilgan parametrlarni (paramlarni), model uchun qoldiqni va ma'lumotlarni (rezidni), o'rnatilgan qiymatlarni (o'rnatilgan qiymatlarni) va yangi mustaqil o'zgaruvchilar uchun javob o'zgaruvchilari qiymatini bashorat qilish usuli (prognoz qilish) va atributlarini o'z ichiga oladi.

5. Nihoyat, Matplotlib va Seaborn grafik kutubxonalari bilan to'g'ridan-to'g'ri statsmodels kutubxonasiga kiritilgan ko'plab grafik tartiblardan ba'zilari yordamida moslashtirish natijasini tasavvur qilish foydali bo'lishi mumkin (statsmodels.graphics moduli).

Ushbu ish jarayonini oddiy misol bilan ko'rsatish uchun, quyida biz haqiqiy qiymati $y = 1 + 2x_1 + 3x_2 + 4x_1x_2$ bo'lgan ma'lumotlarni yaratish uchun modelni moslashtirishni ko'rib chiqamiz. Biz ma'lumotlarni Pandas ma'lumotlar shabloni ob'ektida saqlashdan boshlaymiz:

```
In [1]: N = 100
In [2]: x1 = np.random.randn(N)
In [3]: x2 = np.random.randn(N)
```

```
In [4]: data = pd.DataFrame({"x1": x1, "x2": x2})
In [5]: def y_true(x1, x2):
   ...:     return 1 + 2 * x1 + 3 * x2 + 4 * x1 * x 2
In [6]: data["y_true"] = y_true(x1, x2)
```

Bu yerda biz DataFrame obyekti ma'lumotlaridagi y_true ustunida y ning haqiqiy qiymatini saqladik. Haqiqiy qiymatlarga oddiy taqsimlangan shovqinni qo'shish orqali y ning xatoli kuzatuvini simulyatsiya qilamiz va natijani y ustunida saqlaymiz:

```
In [7]: e = 0.5 * np.random.randn(N)
In [8]: data["y"] = data["y_true"] + e
```

Endi ma'lumotlardan ma'lumki, *y* natijadan tashqari ikkita tushuntiruvchi x1 va x2 o'zgaruvchilar bor. Biz boshlashimiz mumkin bo'lgan eng oddiy model chiziqli model $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ bo'lib, biz uni "y ~ x1 + x2" Patsy formulasi bilan aniqlashimiz mumkin. Javob o'zgaruvchisi uzluksiz bo'lgani uchun oddiy chiziqli kvadratlar yordamida modelni ma'lumotlarga moslashtirish yaxshi boshlanish nuqtasidir, buning uchun smf.ols sinfidan foydalanishimiz mumkin.

```
In [9]: model = smf.ols("y ~ x1 + x2", data)
In [10]: result = model.fit()
```

Esda tutish kerakki, oddiy eng kichik kvadrat regressiya o'rnatilgan modelning qoldiqlari va ma'lumotlar normal taqsimlanganligini taxmin qiladi. Biroq, ma'lumotlarni tahlil qilishdan oldin, biz ushbu shart bajariladimi yoki yo'qligini bilmasligimiz mumkin. Shunga qaramay, ma'lumotlarni modelga moslashtirishdan boshlashimiz va grafik usullar va statistik testlar yordamida qoldiqning taqsimlanishini tekshirishimiz mumkin (qoldiqlar haqiqatan ham normal taqsimlangan degan nol gipoteza bilan).

Foydalanilgan adabiyotlar

1. Robert Johansson. Numerical Python: Scientifi Computing and Data Science Applications with Numpy, SciPy and Matplotlib. Apress.
2. Normurodov C., Toyirov A., Yuldashev S. Numerical modeling of a wave in a nonlinear medium with dissipation //AIP Conference Proceedings. – AIP Publishing LLC, 2022. – T. 2637. – №. 1. – C. 040005.
3. Normurodov C. et al. Numerical simulation of the inverse problem for the vortex-current equation //AIP Conference Proceedings. – AIP Publishing LLC, 2022. – T. 2637. – №. 1. – C. 040018.
4. Normurodov C. B., Toyirov A. X., Yuldashev S. M. Numerical modeling of nonlinear wave systems by the spectral-grid method //International Scientific Journal Theoretical & Applied Science, Philadelphia, USA. – 2020. – T. 83. – №. 3. – C. 43-54.
5. Narmuradov C. B. et al. MATHEMATICAL MODELING OF MOVEMENT OF A VISCOUS INCOMPRESSIBLE LIQUID BY THE SPECTRAL-GRID METHOD //Theoretical & Applied Science. – 2020. – №. 4. – C. 252-260.
6. Begaliyevich N. C., Khasanovich T. A. Spectral-grid method for solving evolution problems with high gradients //EPRA International Journal of Multidisciplinary Research (IJMR). – T. 67.
7. Нармурадов Ч. Б., Тойиров А. Х. Математическое моделирование нелинейных волновых систем //Проблемы вычислительной и прикладной математики. – 2018. – №. 1. – С. 21-31.
8. BEGALIYEVICH N. C. et al. Mathematical Modeling of the Hydrodynamic Stability Problem by the Spectral-grid Method //International Journal of Innovations in Engineering Research and Technology. – T. 7. – №. 11. – C. 20-26.

9. Toyirov A. K., Yuldashev S. M., Abdullayev B. P. Numerical modeling the equations of heat conductivity and burgers by the spectral-grid method //НАУКА 2020. ТЕОРИЯ И ПРАКТИКА. – 2020. – С. 30-31.