

Piotr Szczypiński

Biostatystyka
Ćwiczenia laboratoryjne
część 6
ANOVA w języku Python

ANOVA jednoczynnikowa

Zadanie 1

Załaduj dane z tabeli 1 (próby czterech populacji) do interpretera Python. Stosując jednoczynnikową analizę wariancji (funkcja `f_oneway`) sprawdź, czy średnie rozkładów różnią się w sposób znaczący. Wyświetl histogramy i oceń czy wyniki uzyskane z analizy wariancji potwierdzają wizualną ocenę histogramów. Powtórz powyższe ćwiczenie dla danych z tabeli 2 a następnie tylko dla trzech pierwszych populacji (Y1, Y2, Y3). Na czym polega różnica między tymi danymi? Zinterpretuj uzyskane wyniki.

Tabela 1

```
X1 = [-2.20633294  2.37538572  1.34642554 -0.68241594 -0.74053563 -1.84714087
-0.8924559  0.76932116  2.75495351 -1.47531224  3.91324105  1.8188871
4.64454724 -2.41462404 -0.84693508 -0.71564497 -1.01525292 -0.33120378
-1.31745963  0.24921888 -1.55021395 -0.69615595 -0.56957039 -0.86376226
-0.23689064  4.05011086 -2.93303237 -0.40854696  1.30747435 -0.62049142]
X2 = [ 3.54075709  0.18756598  4.50881574  0.67402821  1.83138559  4.07621537
3.60551625  3.58619986  2.08994763  2.49899861  1.43777692  3.29018263
4.08456719  2.69561188  3.68467336  1.9713825  5.20263274  2.02814441
4.45346765  0.34402668  3.86497603  5.5813109  1.09682727  3.44165707
5.1235829  4.45326603  1.09574542  4.16160757  1.64773029 -1.06768588]
X3 = [ 6.68382158  5.81690198  2.20293949  5.46504364  6.61786535  5.42551135
1.98440097  5.88942657  5.85689432  6.01346683  6.53307527  8.44335662
7.22896858  5.59933061  7.07699566  4.12889546  4.29846479  3.79782556
7.70468161  6.87130702  5.15474931  5.57203824  5.36133618  7.80203348
4.24235508  4.18207132  9.48730351  5.80586579  9.5087635  4.47197423]
X4 = [ 11.58036832  11.60967327  5.88130927  3.94119576  11.38100467
8.46107991  6.70686981  10.29028155  5.79674354  7.77366476
3.91135053  8.65722743  5.62834415  9.97511081  12.30313004
10.47680646  10.18759656  8.21183547  11.88499854  12.23717209
6.99986302  10.42994918  8.43648903  8.14882794  6.16136479
13.13910427  11.21308487  5.40981639  11.66412644  4.89970616]
```

Tabela 2

```
Y1 = [-2.39171609 -2.83289054  0.16181899 -0.29781831  0.67461163 -1.83696967
1.26310692  1.5034946  0.43070269  0.05032675 -1.23924621 -0.05832621
-1.19992955  1.09951361  0.53711405  1.31141728  0.40531205 -3.68261161
-1.27034006  3.05972242 -3.35957777  2.70197823  0.07861712  4.70271213
-1.47620376 -1.14070776 -1.22071682 -1.72005498 -0.07108455 -3.12091824]
Y2 = [-0.89108977  4.32237562 -1.95179306 -0.82015877  0.17205561  3.5359783
-1.37709467  2.62097613  0.57924411 -1.83806465  3.39437457  2.64647146
1.31203107 -1.08713965 -0.01806057 -1.44385579  1.05591984 -1.00171331
1.81525717 -3.24502477  0.19653028  0.42466428  1.99078741 -3.20174614
-1.66598688  3.62671585  3.98938211  2.5531568  -0.72196291 -1.42602116]
Y3 = [-2.01843223 -2.49335659 -0.14978193  0.79167199 -0.98962022 -2.02705255
2.17186133  0.31395653 -0.58503247  0.76470067 -1.12474307 -0.13038176
1.76869079 -2.23247536 -1.28707082  3.57916539 -4.96404186 -0.94431805
3.50845545  0.05672964 -0.29522188  1.73613229 -1.59175857 -0.19037575
3.43441201  1.41237577 -1.21449425  0.74177592  2.3125995  -2.35962559]
Y4 = [ 9.28242643  7.81058947  7.10674617  13.05096686  8.4942901
12.22507158  10.61764339  10.39818433  6.2044844  9.0275227
10.36414393  6.82035455  10.29761735  9.44422605  9.52604743
9.48160993  6.46317089  11.77994151  8.55735575  12.89117636
8.11270311  8.82083459  11.39243549  6.50515229  5.40010858
8.89198772  10.2049745  7.21798525  8.23610182  5.7858507 ]
```

ANOVA wieloczynnikowa

W pakiecie `scipy.stats` brakuje narzędzi do przeprowadzenia wieloczynnikowej analizy wariancji. Analizy takiej można dokonać stosując narzędzia innych bibliotek (pakietów), takich jak `statsmodels` oraz `pandas`. Jeśli nie są one dostępne w używanej dystrybucji środowiska Python należy je doinstalować. W przypadku używanego na zajęciach systemu Linux w dystrybucji Kubuntu w celu instalacji tych pakietów należy uruchomić okno konsoli systemowej i zainstalować pakiety za pomocą programu `apt-get`. Instalowanie nowych pakietów wymaga uprawnień administratora systemu – komenda `sudo`.

```
sudo apt-get install python-pandas
sudo apt-get install python-statsmodels
```

Dokumentacja pakietów jest dostępna na stronach <http://statsmodels.sourceforge.net/> i <http://pandas.pydata.org/>.

Zadanie 2

Dla danych zamieszczonych na stronie <https://people.richland.edu/james/lecture/m170/ch13-2wy.html> zastosuj dwuczynnikową analizę wariancji. Dokonaj interpretacji wyników.

W celu wykonania ćwiczenia zapoznaj się z dokumentacją funkcji `anova_lm` http://statsmodels.sourceforge.net/devel/generated/statsmodels.stats.anova.anova_lm.html oraz przykładami <http://statsmodels.sourceforge.net/devel/anova.html>, http://statsmodels.sourceforge.net/devel/examples/notebooks/generated/interactions_anova.html.