Testovanie štatistických hypotéz II.

Pravdepodobnosť a štatistika pre informatikov 2025

Príklad 1

Pre body z písomky bolo na prednáške ukázané, že ich môžeme považovať za normálne rozdelené. Okrem toho sa zamietla hypotéza o strednej hodnote 10 aj 20. Otestujte teraz hypotézu, že sa stredná hodnota rovná 15.

```
body = [0.5, 1.0, 1.0, 2.0, 2.0, 2.0, 2.0, 2.5, 4.0, 4.0, 5.0, 5.0, 5.5,
6.0, 6.5, 7.0, 7.0, 7.5, 7.5, 7.5, 7.5, 8.0, 9.0, 9.0, 10.0, 10.0, 10.5,
10.5, 10.5, 11.0, 11.0, 11.0, 11.0, 11.5, 12.0, 12.0, 13.0, 14.5, 14.5,
15.0, 15.0, 16.0, 16.5, 17.0, 17.0, 17.5, 17.5, 18.0, 18.0, 18.5, 19.0,
19.0, 20.5, 20.5, 21.0, 21.0, 21.0, 21.5, 21.5, 24.0, 24.5, 25.5, 26.0,
26.0, 26.0, 27.0, 30.0, 30.0]
```

Príklad 2

Magel, R. C. (1998). *Testing for differences between two brands of cookies.* Teaching Statistics, 20(3), 81–83.

Assume the students in the class represent a random sample of professional taste testers. (In this case we are testing cookies.) Each student will rate two brands of cookies on a scale of "0"= terrible to "10"=excellent, with "5"=OK. Use whole numbers only.

(Data will be in pairs. Record the rating each student gave the Brand A cookie first, and the Brand B cookie second.)

NOTE: The data actually collected in class were as follows:

(5,7) (4,6) (5,8) (5,7) (6,8) (3,9)
(5,10) (4,7) (5,7) (5,7) (4,7) (5,8)
(6,5) (7,9) (5,3) (1,5) (4,8)
(5,7) (3,7) (8,7) (4,6) (3,6)

Pre pohodlnejšiu prácu dáta, ktoré sa dajú kopírovať:

```
brand_A = [5, 4, 5, 5, 6,
3, 5, 4, 5, 5, 4, 5, 6, 7,
5, 1, 4, 5, 3, 8, 4, 3]
```

```
brand_B = [7, 6, 8, 7, 8,
9, 10, 7, 7, 7, 7, 8, 5, 9,
3, 5, 8, 7, 7, 7, 6, 6]
```

Otestujeme normalitu – kvôli menšiemu počtu dát Shapiro-Wilkovym testom. Na základe výsledku vyberieme vhodný test na porovnanie datasetov A a B.

Príklad 3

Magel, R. C. (1998). *Testing for differences between two brands of cookies.* Teaching Statistics, 20(3), 81–83.

Let μ_A = mean number of chocolate chips in each Brand A cookie.
Let μ_B = mean number of chocolate chips in each Brand B cookie.
We would like to test (at $\alpha = 0.05$) to see if there is evidence that the two means differ. A random sample of Brand A cookies was taken and the number of chocolate chips in each cookie was counted with the following results:

Similarly, a random sample of Brand B cookies gave the following results:

NOTE: The class obtained the following data:

Sample A:
21, 14, 23, 15, 7, 16, 15, 15, 21, 13, 12, 15, 12, 9, 19, 7, 11, 21, 16, 16, 15, 12.
Sample B:
15, 21, 22, 24, 20, 14, 12, 21, 20, 9, 20, 17, 30, 23, 19, 20, 20, 20, 22, 18, 22, 25.

Pre pohodlnejšiu prácu:

```
sample_A = [21, 14, 23, 15, 7,
16, 15, 15, 21, 13, 12, 15,
12, 9, 19, 7, 11, 21, 16, 16,
15, 12]

sample_B = [15, 21, 22, 24,
20, 14, 12, 21, 20, 9, 20, 17,
30, 23, 19, 20, 20, 20, 22,
18, 22, 25]
```

Otestujeme normalitu – kvôli menšiemu počtu dát Shapiro-Wilkovym testom. Na základe výsledku vyberieme vhodný test na porovnanie datasetov A a B.

Príklad 4

https://istats.shinyapps.io/ANOVA/ (vizualizácie k učebnici Agresti, Franklin, Klingenberg: Statistics: The Art and Science of Learning from Data), dáta o osteoporóze

- Zrekonštrujeme výsledok testovania hypotézy o zhode všetkých stredných hodnôt.
- Pri t-testoch robených pre dvojice dát si vysvelíme úpravu p-hodnôt pomocou https://www.statsmodels.org/dev/generated/statsmodels.stats.multitest.multipletests.html
- Úpravu p-hodnôt použijeme aj pri testovaní normality (dát je viac ako v predchádzajúcich príkladoch, použijeme Lillieforsov test)

```
A = [-0.46, -0.57, -0.70, -0.61, -0.65, -0.62, -0.93, -0.83, -0.70, -0.65,
       -0.62, -0.92, -0.33, -0.70, -0.47, -0.71, -0.94, -0.47, -0.52, -0.85,
       -0.38, -0.76, -0.71, -0.77, -0.73, -0.67, -0.68, -0.67, -0.68, -0.52,
       -0.18, -0.80, -0.99, -0.49, -0.81, -0.80, -0.66, -0.84, -0.83, -0.86,
       -0.31, -0.82]

B= [-0.02, -0.43, -0.24, 0.05, -0.44, -0.10, -0.35, -0.02, -0.21, 0.31,
```

```
        -0.06, -0.40, -0.04, -0.14, -0.39, 0.05, -0.05, 0.06, -0.35, -0.40,
        -0.34, -0.30, -0.35, -0.20, -0.38, -0.27, 0.30, -0.29, -0.25, -0.10,
        -0.34, -0.36, -0.22, -0.06, -0.26, -0.27, 0.10, 0.03, -0.18, -0.25,
        -0.20, -0.37]

C = [-0.20, -0.07, -0.30, -0.26, -0.38, 0.01, -0.23, 0.11, 0.10, 0.07,
        -0.28, -0.19, 0.02, -0.14, -0.25, -0.21, 0.00, -0.54, -0.20, -0.06,
        -0.27, 0.06, -0.14, -0.33, -0.40, -0.21, -0.23, -0.38, -0.43, -0.27,
        0.10, -0.21, -0.34, -0.13, -0.18, 0.00, -0.37, 0.07, 0.19, -0.42,
        -0.17, -0.47]

D = [0.53, 0.54, 0.11, 0.42, 0.73, 1.21, 0.67, 0.63, 0.25, 0.07,
        0.34, 0.32, 0.18, -0.02, 0.53, 0.00, 0.34, 0.04, 0.60, 0.00,
        0.35, -0.45, -0.26, 0.33, -0.05, -0.12, 0.51, 0.29, 0.32, 0.39,
        0.38, -0.32, 0.36, 0.39, 0.58, -0.07, 0.37, 0.61, 0.04, 0.50,
        -0.34, 0.47]
```