Лекція 12

Аналіз вимірювання ПЗ: кореляційний аналіз

1. Оцінка парної кореляції.
2. Парна рангова кореляція.

Кореляційний аналіз

- Мета – виявлення наявності взаємозв’язку між досліджуваними величинами
- У випадку нормального розподілу досліджуваних величин розраховується парна кореляція Пірсона, в іншому – парна рангова кореляція Спірмена чи Кендала

Кореляційний аналіз

Властивості

- $|r| \leq 0$;
- якщо $r = 0$, то η та ξ — незалежні випадкові величини;
- якщо $|r| = 1$, то між η та ξ має місце функціональний зв'язок, у противному разі — випадковий лінійний регресійний
- $$\eta = \alpha + \beta\xi + \varepsilon,$$
- де ξ – вада.

Кореляційний аналіз

- Кореляція $\hat{r} = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{S_x S_y}$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

$$S_x{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad S_y{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

Кореляційний аналіз

- Статистичне значення $\hat{r}$ завжди є відмінним від нуля. Тому виникає задача перевірки значущості коефіцієнта кореляції

- Для перевірки якої реалізують *t*-тест на основі статистичної характеристики

$$t = \frac{\hat{r}\sqrt{n-2}}{\sqrt{1-\hat{r}^2}}$$

- Значення *t* порівнюють із $t_{\alpha/2,\nu}$.

$$|t| \le t_{\alpha/2,\nu}$$

Парна рангова кореляція

- Попередньо початковий масив даних $\{x_i, y_i;$
- $i = \overline{1,n}$ } переформовують у масив рангів

$$\{r_{xi}, r_{yi}, \ i = \overline{1,n}\},$$

-
- де r_{xi}, r_{yi} – порядкові номери варіант у варіаційних рядах за *x* та *y*. При цьому кожному r_{xi} надається номер r_{yi}, що відповідає значенню y_i

Парна рангова кореляція

- Значення оцінки рангового коефіцієнта кореляції Спірмена обчислюють за формулою

$$\hat{\tau}_c = 1 - \frac{6}{n\left(n^2-1\right)}\sum_{i=1}^{n} d_i ,$$

- де
- $d_i = r_{xi} - r_{yi}$

Парна рангова кореляція

- Коефіцієнт рангової кореляції Спірмена має такі властивості:
- $-1 \leq т_c \leq 1$;
- якщо $r_{xi} = r_{yi}$, $i = \overline{1,n}$, то $т_c = 1$, що означає повну узгодженість між X і *Y*;
- якщо $т_c = -1$, то має місце протилежне впорядкування послідовностей рангів, що означає повну неузгодженість (від’ємна кореляція);
- якщо $т_c = 0$, то має місце відсутність кореляції.

Парна рангова кореляція

- Для перевірки значущості вводиться статистична характеристика
- $$t = \frac{\hat{\tau}_c \sqrt{n-2}}{\sqrt{1-\hat{\tau}_c^2}}$$
- яка має *t*-розподіл з *v = n - 2* кількістю ступенів вільності.

Висновки

- Статистичний аналіз найбільш використовується при аналізі деяких вибірок даних