Поряд з коефіцієнтом детермінації $I^2_{\eta\vec{\xi}}$, використовується також характеристика, яку розкриває таке визначення.

Означення. Нехай η і $\vec{\xi}$ – випадкові величина та вектор розмірності q, відповідно, причому $0 < D\eta < \infty$. Тоді ***індексом кореляції*** η ***щодо*** $\vec{\xi}$ називається величина

$$I_{\eta\vec{\xi}} = \sqrt{\frac{Df(\vec{\xi})}{D\eta}} = \sqrt{1 - \frac{D\varepsilon}{D\eta}} = \sqrt{1 - \frac{M\varepsilon^2}{D\eta}}.$$

Властивості індексу кореляції $I_{\eta\vec{\xi}}$:

1) $0 \le I_{\eta\vec{\xi}} \le 1$;
2) якщо $I_{\eta\vec{\xi}} = 0$, то відсутній вплив $\vec{\xi}$ на η;
3) якщо $I_{\eta\vec{\xi}} = 1$, то існує функціональний зв'язок між η та $\vec{\xi}$, а саме, з ймовірністю 1 справедливо $\eta = f\left(\vec{\xi}\right)$.

Аналіз статистичних зв'язків кількісних змінних у загальному випадку

Проведемо за допомогою коефіцієнта детермінації $I^2_{\eta\vec{\xi}}$ дослідження питання про істотність зв'язку між скалярною кількісною змінною η та кількісним вектором $\vec{\xi} = \left(\xi_1, \xi_2, \ldots, \xi_q\right)^T$, зі статистичної точки зору з деяким малим рівнем значущості $\alpha > 0$.

Розглянемо **випадки**:

I. випадок обробки згрупованих даних,

II. випадок наявності можливості апроксимації функції регресії на деякому класі параметричних функцій.

I. випадок обробки згрупованих даних

Нехай отримано n спостережень над скалярною випадковою величиною η і випадковим вектором $\vec{\xi} = \left(\xi_1, \xi_2, \ldots, \xi_q\right)^T$.

Проведемо *групування даних за вектором незалежних змінних* $\vec{\xi}$.

Нехай область значень скалярної змінної ξ_i розбилася на s_i інтервалів групування $\left(i=\overline{1,q}\right)$. Тоді область значень вектора $\vec{\xi}$, в свою чергу, розбивається на $s=\prod_{i=1}^{q} s_i$ гіперпаралелепіпедів групування.

Позначимо спостереження над залежною змінною η, які відповідають вимірам над вектором незалежних змінних $\vec{\xi}$ з i-го гіперпаралелепіпеда групування як:

$$y_{i1}, y_{i2}, \ldots, y_{in_i}, \quad i=\overline{1,s},$$

де n_i – кількість вимірів, які потрапили в i-й гіперпаралелепіпед групування, а $n=\sum_{i=1}^{s} n_i$.

Зауваження. У подальшому скрізь, як правило, в оцінках різних характеристик будемо опускати аргумент n, який вказує на об'єм вибірки, якщо кількість спостережень незмінна.

Тоді загальне середнє можна обчислити таким чином

$$\overline{y}=\frac{1}{n}\sum_{i=1}^{s} n_i \overline{y}_{i\bullet},$$

де $\overline{y}_{i\bullet}$ – середнє на i-му гіперпаралелепіпеду групування, яке обчислюється таким чином:

$$\overline{y}_{i\bullet}=\frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}, \ i=\overline{1,s}.$$

Нагадування. Крапка в $\overline{y}_{i\bullet}$ є позначенням сумування за тим індексом, замість якого вона фігурує.

Тоді для підрахування вибіркового значення $D\eta$ можна скористатися таким виразом:

$$s_\eta^2=\frac{1}{n}\sum_{i=1}^{s}\sum_{j=1}^{n_i}\left(y_{ij}-\overline{y}\right)^2,$$

а в свою чергу, оцінку для $Df(\xi)$ можна визначити згідно формули:

$$s_f^2 = \frac{1}{n}\sum_{i=1}^{s} n_i \left(\overline{y}_{i\bullet} - \overline{y}\right)^2.$$

Це дає змогу обчислити вибіркове значення коефіцієнта детермінації $I_{\eta\vec{\xi}}^2$ таким чином:

$$\hat{I}_{\eta\vec{\xi}}^2 = \frac{s_f^2}{s_\eta^2}.$$

Оцінку

$$\hat{I}_{\eta\vec{\xi}} = \sqrt{\frac{s_f^2}{s_\eta^2}}$$

будемо називати ***кореляційним відношенням*** η *щодо* $\vec{\xi}$.

Процедура використання отриманої статистики $I_{\eta\vec{\xi}}^2$ є традиційною:

1) якщо $\hat{I}_{\eta\vec{\xi}}^2 = 0$, то вважаємо, що зв'язок між η і $\vec{\xi}$ відсутній;
2) якщо $\hat{I}_{\eta\vec{\xi}}^2 = 1$, то вважаємо, що зв'язок між η і $\vec{\xi}$ функціональний;
3) якщо $\hat{I}_{\eta\vec{\xi}}^2 \in (0,1)$, то істотність зв'язку з'ясовується шляхом *перевірки на значимість* коефіцієнта детермінації η щодо $\vec{\xi}$, тобто перевірки гіпотези:

$$H_0 : I_{\eta\vec{\xi}}^2 = 0,$$

з деяким малим рівнем значущості $\alpha > 0$. Якщо гіпотеза H_0 не справедлива, то статистичний зв'язок між η і $\vec{\xi}$ вважається істотним, інакше зв'язок вважається не істотним.

Для перевірки останньої гіпотези скористаємося тим, що при справедливості гіпотези H_0, виявляється, що розподіл статистики

$$F = \frac{\hat{I}^2_{\eta\vec{\xi}}}{1-\hat{I}^2_{\eta\vec{\xi}}} \cdot \frac{n-s}{s-1}$$

можна наблизити F-розподілом з $(s-1)$ та $(n-s)$ ступенями свободи.

Тоді, з огляду на структуру статистики F, в якості критичної області для гіпотези H_0 потрібно взяти *область великих значень*, а відповідна область прийняття для нашої гіпотези матиме вигляд:

$$F < F_\alpha(s-1, n-s),$$

де $F_\alpha(m,n)$ – 100α відсоткова точка F-розподілу з m та n ступенями свободи.

II. випадок наявності можливості апроксимації функції регресії на деякому класі параметричних функцій.

Розглянемо ситуацію, коли функцію регресії η щодо $\vec{\xi}$ можна апроксимувати на деякому класі параметричних функцій $\tilde{f}(\vec{x},\vec{\lambda})$, $\vec{x} \in \mathbb{R}^q$, $\vec{\lambda} \in \mathbb{R}^p$.

Нехай доступні такі спостереження над скалярною η і вектором $\vec{\xi}$:

$$\eta : y_1, y_2, \ldots, y_n,$$
$$\vec{\xi} : \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n.$$

На основі цих спостережень згідно деякого відомого методу знаходимо оцінку $\hat{\vec{\lambda}}$ для вектора невідомих параметрів $\vec{\lambda}$. А також приймаємо до уваги, що:

$$I^2_{\eta\vec{\xi}} = 1 - \frac{M\varepsilon^2}{D\eta} = 1 - \frac{M\left(\eta - f\left(\vec{\xi}\right)\right)^2}{D\eta}.$$

Ці міркування дозволяють підрахувати *емпіричне (вибіркове) значення коефіцієнта детермінації* $I^2_{\eta\vec{\xi}}$ таким чином:

$$\hat{I}^2_{\eta\vec{\xi}} = 1 - \frac{\frac{1}{n-p}\sum_{i=1}^{n}\left(y_i - \tilde{f}\left(\vec{x}_i, \hat{\vec{\lambda}}\right)\right)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2},$$

де $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Процедура використання $\hat{I}^2_{\eta\vec{\xi}}$:

1) якщо $\hat{I}^2_{\eta\vec{\xi}} = 0$, то вважаємо, що зв'язок між η і $\vec{\xi}$ відсутній;

2) якщо $\hat{I}^2_{\eta\vec{\xi}} = 1$, то вважаємо, що зв'язок між η і $\vec{\xi}$ функціональний;

3) якщо $\hat{I}^2_{\eta\vec{\xi}} \in (0,1)$, то істотність зв'язку між η і $\vec{\xi}$ з'ясовується шляхом *перевірки на значимість* $I^2_{\eta\vec{\xi}}$, тобто гіпотези:

$$H_0 : I^2_{\eta\vec{\xi}} = 0,$$

з деяким малим рівнем значущості $\alpha > 0$.

Якщо гіпотеза H_0 справедлива, то статистичний зв'язок між η і $\vec{\xi}$ вважається не істотним, інакше зв'язок вважається істотним.

Перевірку цієї гіпотези будемо проводити за допомогою статистики

$$F = \frac{\hat{I}^2_{\eta\vec{\xi}}}{1 - \hat{I}^2_{\eta\vec{\xi}}} \cdot \frac{n-p}{p-1},$$

розподіл якої, при справедливості гіпотези H_0, можна наблизити F-розподілом з $(p-1)$ та $(n-p)$ ступенями свободи.

Це дозволяє область прийняття для гіпотези H_0 записати у вигляді

$$F < F_{\alpha}(p-1, n-p),$$

де $F_{\alpha}(m,n)$ – 100α відсоткова точка F-розподілу з m та n ступенями свободи.

У підсумку, у розглянутому випадку для коефіцієнта детермінації η щодо $\vec{\xi}$ отримано формулу його емпіричного значення та описано процедуру використання вибіркового значення $\hat{I}^2_{\eta\vec{\xi}}$ для з'ясування істотності статистичного зв'язку між η і $\vec{\xi}$.

Дослідження статистичних зв'язків кількісних змінних у нормальному випадку

Розглянемо на ймовірнісному просторі $(\Omega, \mathfrak{A}, P)$ скалярну залежну кількісну змінну η та вектор незалежних кількісних змінних $\vec{\xi} = \left(\xi_1, \xi_2, \ldots, \xi_q\right)^T$, які **нормально розподілені** з деякими параметрами. У разі потреби для уніфікації позначень у подальшому будемо використовувати нотацію ξ_0 для змінної η, тобто $\xi_0 \equiv \eta$.

Аналіз наявності статистичного зв'язку між гаусівськими змінними η та $\vec{\xi}$ має свою специфіку, а саме виявилось, що замість коефіцієнта детермінації η щодо $\vec{\xi}$ можна використовувати інші, більш зручні характеристики, а саме: *коефіцієнт кореляції*, *частинний коефіцієнтом кореляції* або *множинний коефіцієнт кореляції змінних* η *та* $\vec{\xi}$.

Розглянемо **випадки**:

I. випадок $q = 1$,

II. випадок $q > 1$.

I. випадок $q = 1$.

Тобто потрібно проаналізувати *парний статистичний зв'язок* між скалярними <u>нормально</u> розподіленими змінними η та ξ. Припустимо, що:

$$\eta \sim \mathcal{N}\left(m_\eta, \sigma_\eta^2\right),\ \xi \sim \mathcal{N}\left(m_\xi, \sigma_\xi^2\right),\ \sigma_\eta^2, \sigma_\xi^2 > 0. \qquad (*)$$

Згадаємо позначення:

$$r_{\eta\xi} = \frac{M\left(\eta - m_\eta\right)\left(\xi - m_\xi\right)}{\sqrt{D\eta D\xi}},$$

$$f(x) = M(\eta / \xi = x),\ g(x) = D(\eta / \xi = x).$$

Для $f(\bullet)$ та $g(\bullet)$ справедливе таке твердження.

Теорема 1. Нехай для скалярних змінних η та ξ має місце припущення (*), тоді справедливо

$$f(x) = m_\eta + r_{\eta\xi}\frac{\sigma_\eta}{\sigma_\xi}\left(x - m_\xi\right), \qquad g(x) = \sigma_\eta^2\left(1 - r_{\eta\xi}^2\right).$$

Остання теорема дозволяє переконатися, що справедлива.

Теорема 2. Нехай для скалярних змінних η та ξ має місце припущення (*), тоді для індексу кореляції справедливо $I_{\eta\xi} = \left|r_{\eta\xi}\right|$, а відповідно для коефіцієнту детермінації $I_{\eta\xi}^2 = r_{\eta\xi}^2$.

Доведення. Згідно з попередньою теоремою індекс кореляції $I_{\eta\xi}$ можна обчислити таким чином:

$$I_{\eta\xi} = \sqrt{\frac{Df(\xi)}{D\eta}} = \sqrt{\frac{D\left(m_\eta + r_{\eta\xi}\frac{\sigma_\eta}{\sigma_\xi}\left(\xi - m_\xi\right)\right)}{\sigma_\eta^2}} = \sqrt{\frac{r_{\eta\xi}^2\frac{\sigma_\eta^2}{\sigma_\xi^2}D\left(\xi - m_\xi\right)}{\sigma_\eta^2}} = \left|r_{\eta\xi}\right|. \quad \blacksquare$$

Останнє дозволяє у цій ситуації замість коефіцієнта детермінації $I_{\eta\xi}^2$ використовувати коефіцієнт кореляції $r_{\eta\xi}$.

Властивості $r_{\eta\xi}$.

1) $\left|r_{\eta\xi}\right| \le 1$;
2) якщо $r_{\eta\xi} = 0$, то η не залежить від ξ;
3) якщо $\left|r_{\eta\xi}\right| = 1$, то з ймовірністю 1 існує лінійний зв'язок між η та ξ, а саме, справедливо

$$\eta = m_\eta + r_{\eta\xi}\frac{\sigma_\eta}{\sigma_\xi}\left(\xi - m_\xi\right);$$

4) $r_{\eta\xi} = r_{\xi\eta}$.

$$\frac{\eta - m_\eta}{\sigma_\eta} = r_{\eta\xi}\frac{1}{\sigma_\xi}\left(\xi - m_\xi\right)$$

Доведення. Скористаємося властивостями індексу кореляції:

1) дійсно, з останньої теореми та першої властивості $I_{\eta\xi}$ випливає
$$\left|r_{\eta\xi}\right| = I_{\eta\xi} \le 1 \Rightarrow \left|r_{\eta\xi}\right| \le 1;$$
2) оскільки $r_{\eta\xi} = 0$, то остання теорема дозволяє стверджувати, що:
$$0 = \left|r_{\eta\xi}\right| = I_{\eta\xi} \Rightarrow I_{\eta\xi} = 0.$$
Згідно з 2-ою властивістю $I_{\eta\xi}$ це означає, що η не залежить від ξ;
3) так як $\left|r_{\eta\xi}\right| = 1$, то остання теорема приводить до висновку, що
$$1 = \left|r_{\eta\xi}\right| = I_{\eta\xi} \Rightarrow I_{\eta\xi} = 1.$$
Тоді 3-тя властивість $I_{\eta\xi}$ дозволяє стверджувати, що з ймовірністю 1 справедливо $\eta = f(\xi)$. Але в нормальному випадку
$$f(x) = m_\eta + r_{\eta\xi}\frac{\sigma_\eta}{\sigma_\xi}\left(x - m_\xi\right) \Rightarrow \eta = m_\eta + r_{\eta\xi}\frac{\sigma_\eta}{\sigma_\xi}\left(\xi - m_\xi\right).$$
4) очевидно. ■

Зауваження. Вигляд функції регресії $f(x)$, який наведено в передостанній теоремі вказує на те, що зв'язок між η і ξ має монотонний характер, а знак коефіцієнта кореляції $r_{\eta\xi}$ конкретизує його, а саме: якщо $r_{\eta\xi} > 0$ $\left(r_{\eta\xi} < 0\right)$, то функція залежності буде зростаючою (спадною).

Перейдемо до практичного використання коефіцієнта кореляції $r_{\eta\xi}$ як характеристики парного статистичного зв'язку для скалярних гаусівських змінних. Нехай доступні спостереження:

$$\eta : y_1, y_2, \ldots, y_n,$$
$$\xi : x_1, x_2, \ldots, x_n.$$

Тоді, на основі цих вимірів ***вибіркове (емпіричне) значення*** $\hat{r}_{\eta\xi}$ ***(парного) коефіцієнта кореляції*** $r_{\eta\xi}$ обчислюється за відомою формулою:

$$\hat{r}_{\eta\xi} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)\left(x_i - \overline{x}\right)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2 \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}},$$

де $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Процедура використання $\hat{r}_{\eta\xi}$ для аналізу статистичного зв'язку між скалярними змінними η та ξ має вигляд:

1) якщо $\hat{r}_{\eta\xi} = 0$, то η не залежить від ξ;
2) якщо $\left|\hat{r}_{\eta\xi}\right| = 1$, то зв'язок між η та ξ функціональний (точніше лінійний) і має вигляд

$$\eta = \hat{m}_\eta + \hat{r}_{\eta\xi}\frac{\hat{\sigma}_\eta}{\hat{\sigma}_\xi}\left(\xi - \hat{m}_\xi\right);$$

3) якщо $\left|\hat{r}_{\eta\xi}\right| \in \left(0,1\right)$, то потрібно здійснити *перевірку на значимість* $r_{\eta\xi}$, тобто перевірити гіпотезу:

$$H_0 : r_{\eta\xi} = 0,$$

з деяким рівнем значущості $\alpha > 0$.

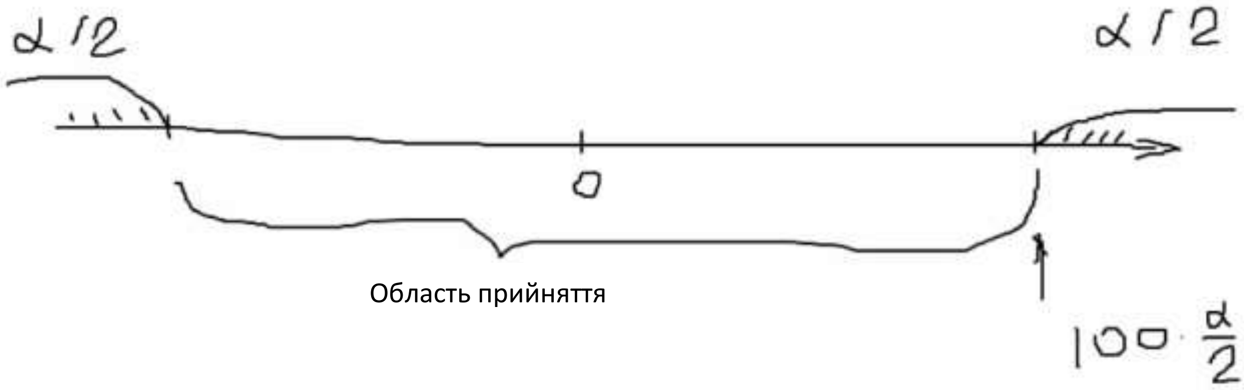
Скористаємося тим фактом, що, при справедливості гіпотези H_0 при великих n, розподіл статистики

$$t(n,2) = \frac{\sqrt{n-2}\ \hat{r}_{\eta\xi}}{\sqrt{1-\hat{r}_{\eta\xi}^2}}$$

можна наблизити t-розподілом Стьюдента з $(n-2)$ ступенями свободи. Тоді логічно в якості критичної області цієї гіпотези взяти області набуття статистикою $t(n,2)$ своїх екстремальних значень. А область прийняття гіпотези H_0 відповідно матиме вигляд:

$$\left|t(n,2)\right| < t_{\frac{\alpha}{2}}(n-2),$$

де $t_{\alpha}(n)$ – 100α відсоткова точка t-розподілу Стьюдента з n ступенями свободи.



Зауваження. Скрізь у подальшому, як правило, будемо вважати справедливим припущення про нормальність відповідних спостережень у разі необхідності перевірки гіпотез або побудови довірчих інтервалів, якщо не обумовлено інше.

Таким чином, у нормальному випадку в якості характеристики парного статистичного зв'язку можна використовувати звичайний коефіцієнт кореляції $r_{\eta\xi}$, який має прозору інтерпретацію й дозволяє стверджувати про наявність лінійного зв'язку, коли він набуває значення ± 1. Тому коефіцієнт кореляції $r_{\eta\xi}$ ще називають *характеристикою парного статистичного лінійного зв'язку*. У загальному випадку коефіцієнт кореляції вже не має такої яскравої інтерпретації і тому виникає потреба у зверненні до універсальної характеристики парного статистичного зв'язку – коефіцієнта детермінації $I^2_{\eta\xi}$.

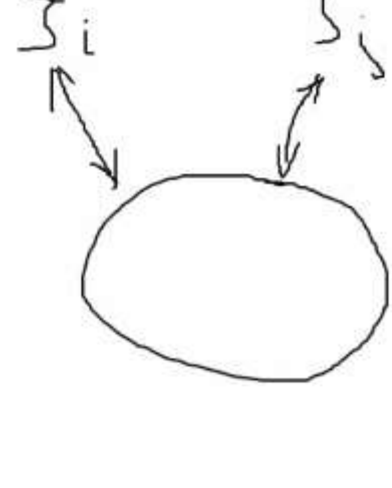
II. випадок $q > 1$.

а) *аналіз парного статистичного зв'язку* між двома скалярними нормально розподіленими змінними при наявності сторонніх змінних

Розглянемо скалярну залежну кількісну змінну η та вектор незалежних кількісних змінних $\vec{\xi} = \left(\xi_1, \xi_2, \ldots, \xi_q\right)^T$, причому $\xi_0 \equiv \eta$ тобто маємо справу з набором скалярних змінних

$$\xi_0, \xi_1, \xi_2, \ldots, \xi_q.$$

Необхідно з'ясувати, чи є істотним статистичний зв'язок між деякою парою змінних ξ_i та ξ_j у такій ситуації.



Практика дослідження зв'язку між змінними ξ_i та ξ_j за допомогою коефіцієнту кореляції $r_{\xi_i \xi_j}$ (або скорочено r_{ij}) показує, що коефіцієнт кореляції $r_{\xi_i \xi_j}$ може неадекватно віддзеркалювати наявний зв'язок, коли існують інші (сторонні) змінні, як у цьому випадку ξ_l, $l \in I$, $\left(I = \{0,1,2,\ldots,q\} \setminus \{i,j\}\right)$, які мають суттєвий вплив на статистичний зв'язок між змінними ξ_i та ξ_j. Тому виникає бажання удосконалити звичайний коефіцієнт кореляції $r_{\xi_i \xi_j}$ таким чином, щоб модифікована характеристика парного статистичного зв'язку була не чутлива до впливу сторонніх змінних.

Для цього було запропоновано підраховувати коефіцієнт кореляції для ξ_i та ξ_j не за їх звичайним сумісним розподілом, як це робилося раніше для парного коефіцієнта кореляції $r_{\xi_i \xi_j}$, а за їх умовним сумісним розподілом, де в умові сторонні змінні, наприклад ξ_l, $l \in I$, приймають фіксовані значення.

Означення (у загальному випадку). ***Частинним коефіцієнтом кореляції випадкових величин*** ξ_i ***та*** ξ_j ***при сторонніх змінних*** ξ_l, $l \in I$ називається числова характеристика $r_{ij(I)}$, яка підрахована за формулою парного коефіцієнта кореляції для ξ_i та ξ_j, але за умовним сумісним розподілом випадкових величин ξ_i та ξ_j при умові, що змінні ξ_l, $l \in I$ набувають фіксованих значень, де I – множина індексів сторонніх змінних.

Зауваження. Множина індексів сторонніх змінних I не містить у собі індекси i та j.

У загальному випадку, така модифікована характеристика парного статистичного зв'язку для ξ_i та ξ_j залежить від вибраних постійних значень, яких набувають сторонні змінні, і тому породжує певні труднощі в її використанні. Але з'ясувалося, що у нормальному випадку частинний коефіцієнт кореляції для випадкових величин ξ_i та ξ_j не залежить від фіксованих значень, яких набувають інші змінні ξ_l, $l \in I$, і тому знайшов своє широке застосування.

Якщо випадкові величини $\xi_0, \xi_1, \xi_2, \ldots, \xi_q$ мають невироджений сумісний гаусівський розподіл ($\xi_0 \equiv \eta$), а множина індексів сторонніх змінних $I=\{0,1,2,\ldots,q\}\setminus\{i,j\}$, то частинний коефіцієнт кореляції $r_{ij(I)}$ можна обчислити згідно з

$$r_{ij(I)}=-\frac{R_{ij}}{\sqrt{R_{ii}R_{jj}}},$$

де R_{kl} – алгебраїчне доповнення до елемента r_{kl} у матриці парних коефіцієнтів кореляції для змінних $\xi_0,\xi_1,\xi_2,\ \ldots,\xi_q$, а саме:

$$R=\begin{pmatrix} 1 & r_{01} & r_{02} & \cdots & r_{0q} \\ r_{10} & 1 & r_{12} & \cdots & r_{1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{q0} & r_{q1} & r_{q2} & \cdots & 1 \end{pmatrix}.$$

Зауваження. У позначенні $r_{ij(I)}=r_{ij(\{0,1,2,\ldots,q\}\setminus\{i,j\})}$ у дужках вказується множина індексів усіх сторонніх змінних, які приймають фіксовані значення.

Зауваження. З огляду на означення $r_{ij(I)}$, частинного коефіцієнта кореляції випадкових величин ξ_i та ξ_j при сторонніх змінних ξ_l, $l\in I$, можна стверджувати, що його властивості перекликаються з властивостями парного коефіцієнта кореляції для випадкових величин ξ_i та ξ_j, тобто r_{ij}.

Перейдемо до практичного використання частинного коефіцієнта кореляції $r_{ij(I)}$. Нехай доступні спостереження:

$$\eta: y_1,y_2,\ \ldots,y_n,$$
$$\xi_i: x_{i1},x_{i2},\ \ldots,x_{in},\ i=\overline{1,q}.$$

Для обчислення емпіричного значення частинного коефіцієнта кореляції $\hat{r}_{ij(I)}$, $I=\{0,1,2,\ldots,q\}\setminus\{i,j\}$ можна скористатися виразом

$$\hat{r}_{ij(I)}=-\frac{\hat{R}_{ij}}{\sqrt{\hat{R}_{ii}\hat{R}_{jj}}},$$

де $\hat{R}_{kl}$ – алгебраїчне доповнення до елемента $\hat{r}_{kl}$ у матриці вибіркових парних коефіцієнтів кореляції для змінних $\xi_0, \xi_1, \xi_2, \ldots, \xi_q$, а саме:

$$\hat{R} = \begin{pmatrix} 1 & \hat{r}_{01} & \hat{r}_{02} & \ldots & \hat{r}_{0q} \\ \hat{r}_{10} & 1 & \hat{r}_{12} & \ldots & \hat{r}_{1q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{r}_{q0} & \hat{r}_{q1} & \hat{r}_{q2} & \ldots & 1 \end{pmatrix}.$$

Нехай $I = \{0,1,2,\ldots,q\} \setminus \{i,j\}$ – множина індексів сторонніх змінних, які зафіксовані на деякому рівні при обчисленні частинного коефіцієнта кореляції $r_{ij(I)}$ для змінних ξ_i та ξ_j.

Процедура використання $\hat{r}_{ij(I)}$ аналогічна процедурі використання парного коефіцієнта кореляції $\hat{r}_{ij}$ в усьому, крім етапу перевірки цієї характеристики на значимість, в який потрібно внести незначні корекції.

В результаті **процедура використання** $\hat{r}_{ij(I)}$ набуває вигляду:

1) якщо $\hat{r}_{ij(I)} = 0$, то ξ_i не залежить від ξ_j;
2) якщо $\left|\hat{r}_{ij(I)}\right| = 1$, то зв'язок між ξ_i та ξ_j функціональний (точніше лінійний);
3) якщо $\left|\hat{r}_{ij(I)}\right| \in (0,1)$, то здійснюємо *перевірку на значимість* частинного коефіцієнта кореляції $r_{ij(I)}$:

$$H_0 : r_{ij(I)} = 0$$

з рівнем значущості $\alpha > 0$.

Виявляється, що при справедливості гіпотези H_0, розподіл статистики

$$t(n, q+1) = \frac{\sqrt{n-(q+1)}\ \hat{r}_{ij(I)}}{\sqrt{1-\hat{r}^2_{ij(I)}}} = \frac{\sqrt{n-q-1}\ \hat{r}_{ij(I)}}{\sqrt{1-\hat{r}^2_{ij(I)}}}$$

можна наблизити t-розподілом Стьюдента з $(n-q-1)$ ступенями свободи. Тоді до області відхилення гіпотези H_0 потрібно віднести екстремальні значення статистики $t(n,q+1)$, а це дозволяє записати область прийняття гіпотези H_0 у такому вигляді:

$$\left|t(n,q+1)\right| < t_{\frac{\alpha}{2}}(n-q-1),$$

де $t_{\alpha}(n)$ – 100α відсоткова точка t-розподілу Стьюдента з n ступенями свободи.

[redacted]