ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет информатики, математики и компьютерных наук

**Программа подготовки бакалавров по направлению
01.03.02 Прикладная математика и информатика**

Антонов Илья Витальевич

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

Проверка условной независимости в трехмерном распределении Бернулли

Научный руководитель
д.ф.-м.н., проф.
П.А. Колданов

Нижний Новгород, 2024

Содержание

1 Трехмерное распределение Бернулли

Определение 1.1. Случайный вектор $(X, Y, Z)^T$ имеет трехмерное распределение Бернулли, если множество его возможных значений

$$\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \dots, \begin{pmatrix}1\\1\\1\end{pmatrix}$$

и заданы вероятности

$$P(X = x, Y = y, Z = z) = p_{xyz} \geq 0, \quad \sum_{x=0}^{1}\sum_{y=0}^{1}\sum_{z=0}^{1} p_{xyz} = 1$$

Определение 1.2. Пусть $(X, Y, Z)^T$ – дискретный случайный вектор. Говорят, что случайные величины X и Y условно независимы при условии Z, и пишут $X \perp\!\!\!\perp Y \mid Z$ если для любых x, y и z, такого что $P(Z = z) > 0$, выполнено:

$$P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z)P(Y = y \mid Z = z)$$

Теорема 1.1. Пусть $(X, Y, Z)^T$ – случайный вектор, имеющий трехмерное распределение Бернулли, и $P(Z = 0) > 0$. Случайные величины X и Y условно независимы при условии Z тогда и только тогда, когда $p_{00z}p_{11z} = p_{01z}p_{10z}$, где $z = 0, 1$.

Доказательство. Пусть $X \perp\!\!\!\perp Y \mid Z$. Значит, для любых $x = 0, 1$, $y = 0, 1$ и $z = 0, 1$ выполнено условие:

$$P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z)P(Y = y \mid Z = z)$$

После домножения на $P(Z = z)^2$ получаем эквивалентное условие:

$$P(X = x, Y = y, Z = z)P(Z = z) = P(X = x, Z = z)P(Y = y, Z = z)$$

Найдем маргинальное распределение случайной величины Z:

$$P(Z = z) = \sum_{x=0}^{1}\sum_{y=0}^{1} p_{xyz} = p_{00z} + p_{01z} + p_{10z} + p_{11z}$$

Найдем маргинальные распределения $(X, Z)^T$ и $(Y, Z)^T$:

$$P(X = x, Z = z) = \sum_{y=0}^{1} p_{xyz} = p_{x0z} + p_{x1z}$$

$$P(Y = y, Z = z) = \sum_{x=0}^{1} p_{xyz} = p_{0yz} + p_{1yz}$$

Тогда условие условной независимости перепишем в следующем виде:

$$p_{xyz}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{x0z} + p_{x1z})(p_{0yz} + p_{1yz})$$

Это условие выполняется для всех $x = 0, 1$, $y = 0, 1$, $z = 0, 1$. Пусть z фиксировано. Если $x = 0$ и $y = 0$, то:

$$p_{00z}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{00z} + p_{01z})(p_{00z} + p_{10z})$$

$$p_{00z}p_{00z}+p_{00z}p_{01z}+p_{00z}p_{10z}+p_{00z}p_{11z} = p_{00z}p_{00z}+p_{00z}p_{10z}+p_{01z}p_{00z}+p_{01z}p_{10z}$$

$$p_{00z}p_{11z} = p_{01z}p_{10z}$$

Если $x = 0$ и $y = 1$, то:

$$p_{01z}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{00z} + p_{01z})(p_{01z} + p_{11z})$$

$$p_{01z}p_{00z}+p_{01z}p_{01z}+p_{01z}p_{10z}+p_{01z}p_{11z} = p_{00z}p_{01z}+p_{00z}p_{11z}+p_{01z}p_{01z}+p_{01z}p_{11z}$$

$$p_{01z}p_{10z} = p_{00z}p_{11z}$$

Если $x = 1$ и $y = 0$, то:

$$p_{10z}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{10z} + p_{11z})(p_{00z} + p_{10z})$$

$$p_{10z}p_{00z}+p_{10z}p_{01z}+p_{10z}p_{10z}+p_{10z}p_{11z} = p_{10z}p_{00z}+p_{10z}p_{10z}+p_{11z}p_{00z}+p_{11z}p_{10z}$$

$$p_{10z}p_{01z} = p_{11z}p_{00z}$$

Если $x = 1$ и $y = 1$, то:

$$p_{11z}(p_{00z} + p_{01z} + p_{10z} + p_{11z}) = (p_{10z} + p_{11z})(p_{01z} + p_{11z})$$

$$p_{11z}p_{00z}+p_{11z}p_{01z}+p_{11z}p_{10z}+p_{11z}p_{11z} = p_{10z}p_{01z}+p_{10z}p_{11z}+p_{11z}p_{01z}+p_{11z}p_{11z}$$

$$p_{11z}p_{00z} = p_{10z}p_{01z}$$

Таким образом, из условной независимости X и Y при условии Z следует $p_{00z}p_{11z} = p_{01z}p_{10z}$, где $z = 0, 1$.

Доказательство в обратную сторону проводится аналогично. $\square$

Пример 1.1. Пусть $(X, Y, Z)^T$ имеет трехмерное распределение Бернулли с вероятностями $p_{000} = 0.15$, $p_{001} = 0.1$, $p_{010} = 0.3$, $p_{011} = 0.1$, $p_{100} = 0.05$, $p_{101} = 0.1$, $p_{110} = 0.1$, $p_{111} = 0.1$. Заметим, что:

$$p_{000}p_{110} = p_{010}p_{100} = 0.015$$

$$p_{001}p_{111} = p_{011}p_{101} = 0.01$$

Значит из теоремы 1.1 следует, что $X \perp\!\!\!\perp Y \mid Z$.

Трехмерное распределение Бернулли можно представить в экспоненциальном виде:

$$P(X = x, Y = y, Z = z) = p_{000}^{(1-x)(1-y)(1-z)} \dots p_{111}^{xyz}$$

Докажем лемму касательно вида $P(X = x, Y = y, Z = z)$.

Лемма 1.1.

$$P(X = x, Y = y, Z = z) = \exp\left\{ \ln p_{000} \right\} \exp\left\{ xyz \ln\left(\frac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}}\right) + \right.$$

$$+x \ln\left(\frac{p_{100}}{p_{000}}\right) + y \ln\left(\frac{p_{010}}{p_{000}}\right) + z \ln\left(\frac{p_{001}}{p_{000}}\right) +$$

$$\left. +xy \ln\left(\frac{p_{000}p_{110}}{p_{010}p_{100}}\right) + xz \ln\left(\frac{p_{000}p_{101}}{p_{001}p_{100}}\right) + yz \ln\left(\frac{p_{000}p_{011}}{p_{001}p_{010}}\right) \right\}$$

Доказательство.

$$P(X = x, Y = y, Z = z) = p_{000}^{(1-x)(1-y)(1-z)} \dots p_{111}^{xyz} =$$

$$= \exp\left\{ (1-x)(1-y)(1-z) \ln p_{000} + (1-x)(1-y)z \ln p_{001} + \right.$$

$$+(1-x)y(1-z) \ln p_{010} + (1-x)yz \ln p_{011} + x(1-y)(1-z) \ln p_{100} +$$

$$+x(1-y)z\ln p_{101} + xy(1-z)\ln p_{110} + xyz\ln p_{111}\Bigg\} =$$

$$= \exp\Bigg\{(1-y-x+xy-z+yz+xz-xyz)\ln p_{000}+$$

$$+(z-yz-xz+xyz)\ln p_{001} + (y-yz-xy+xyz)\ln p_{010}+$$

$$+(yz-xyz)\ln p_{011} + (x-xz-xy+xyz)\ln p_{100}+$$

$$+(xz-xyz)\ln p_{101} + (xy-xyz)\ln p_{110} + xyz\ln p_{111}\Bigg\} =$$

$$= \exp\Bigg\{\ln p_{000}\Bigg\}\exp\Bigg\{xyz\ln\left(\frac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}}\right)+$$

$$+x\ln\left(\frac{p_{100}}{p_{000}}\right) + y\ln\left(\frac{p_{010}}{p_{000}}\right) + z\ln\left(\frac{p_{001}}{p_{000}}\right)+$$

$$+xy\ln\left(\frac{p_{000}p_{110}}{p_{010}p_{100}}\right) + xz\ln\left(\frac{p_{000}p_{101}}{p_{001}p_{100}}\right) + yz\ln\left(\frac{p_{000}p_{011}}{p_{001}p_{010}}\right)\Bigg\}$$

$\square$

Сформулируем теорему касательно параметра трехмерного распределения Бернулли, связанного с условной независимостью.

Теорема 1.2. Пусть $\theta = \ln\left(\dfrac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}}\right)$. Если выполнено одно из условий:

- $X \perp\!\!\!\perp Y \mid Z$
- $X \perp\!\!\!\perp Z \mid Y$
- $Y \perp\!\!\!\perp Z \mid X$

то параметр θ принимает значение 0.

Доказательство. Результаты теоремы 1.1 можно обобщить следующим образом:

$$X \perp\!\!\!\perp Z \mid Y \Leftrightarrow p_{000}p_{101} = p_{001}p_{100} \text{ и } p_{010}p_{111} = p_{011}p_{110}$$

$$Y \perp\!\!\!\perp Z \mid X \Leftrightarrow p_{000}p_{011} = p_{001}p_{010} \text{ и } p_{100}p_{111} = p_{101}p_{110}$$

1. Пусть $X \perp\!\!\!\perp Y \mid Z$, тогда по теореме 1.1 выполнено: $p_{000}p_{110} = p_{010}p_{100}$ и $p_{001}p_{111} = p_{011}p_{101}$. Отсюда следует, что $\theta = \ln(1) = 0$.

2. Пусть $X \perp\!\!\!\perp Z \mid Y$, тогда из вышеприведенных соображений $p_{000}p_{101} = p_{001}p_{100}$ и $p_{010}p_{111} = p_{011}p_{110}$. Отсюда следует, что $\theta = \ln(1) = 0$.

3. Пусть $Y \perp\!\!\!\perp Z \mid X$, тогда из вышеприведенных соображений $p_{000}p_{011} = p_{001}p_{010}$ и $p_{100}p_{111} = p_{101}p_{110}$. Отсюда следует, что $\theta = \ln(1) = 0$.

$\square$

2 Равномерно наиболее мощный несмещенный тест

Пусть

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \\ Z_n \end{pmatrix}$$

повторная выборка из распределения случайного вектора $(X, Y, Z)^T$.

Из леммы 1.1 непосредственно следует, что совместное распределение повторной выборки имеет вид:

$$P(X_1 = x_1, Y_1 = y_1, Z_1 = z_1, \ldots, X_n = x_n, Y_n = y_n, Z_n = z_n) =$$

$$= \prod_{i=1}^{n} P(X_i = x_i, Y_i = y_i, Z_i = z_i) =$$

$$= \exp\left\{ n \ln p_{000} \right\} \exp\left\{ \ln \left(\frac{p_{001}p_{111}p_{010}p_{100}}{p_{011}p_{101}p_{000}p_{110}} \right) \sum_{i=1}^{n} x_i y_i z_i + \right.$$

$$+ \ln \left(\frac{p_{100}}{p_{000}} \right) \sum_{i=1}^{n} x_i + \ln \left(\frac{p_{010}}{p_{000}} \right) \sum_{i=1}^{n} y_i + \ln \left(\frac{p_{001}}{p_{000}} \right) \sum_{i=1}^{n} z_i +$$

$$\left. + \ln \left(\frac{p_{000}p_{110}}{p_{010}p_{100}} \right) \sum_{i=1}^{n} x_i y_i + \ln \left(\frac{p_{000}p_{101}}{p_{001}p_{100}} \right) \sum_{i=1}^{n} x_i z_i + \ln \left(\frac{p_{000}p_{011}}{p_{001}p_{010}} \right) \sum_{i=1}^{n} y_i z_i \right\}$$

Из вышеприведенной теоремы следует, что интересующее нас значение параметра θ равно $\theta_0 = 0$. Для проверки гипотезы $H : \theta = \theta_0$ против альтернативы $K : \theta \neq \theta_0$ можно использовать равномерный наиболее мощный в классе несмещенных тест:

$$\varphi(u,t) = \begin{cases} 1,\ u < C_1(t) \text{ или } u > C_2(t) \\ \gamma_i,\ u = C_i(t),\ i = 1,2 \\ 0,\ C_1(t) < u < C_2(t) \end{cases}$$

где константы C_i и γ_i определяются из системы уравнений:

$$\begin{cases} E_{\theta_0}[\varphi(U,T) \mid T = t] = \alpha \\ E_{\theta_0}[U\varphi(U,T) \mid T = t] = \alpha E_{\theta_0}[U \mid T = t] \end{cases}$$

а статистиками являются:

$$U = \sum_{i=1}^{n} X_i Y_i Z_i, T_1 = \sum_{i=1}^{n} X_i Y_i, T_2 = \sum_{i=1}^{n} X_i Z_i, T_3 = \sum_{i=1}^{n} Y_i Z_i,$$

$$T_4 = \sum_{i=1}^{n} X_i, T_5 = \sum_{i=1}^{n} Y_i, T_6 = \sum_{i=1}^{n} Z_i$$

Приведем две технические леммы, для того, чтобы найти условное распределение статистики U при условии $T_1 = t_1, \dots, T_6 = t_6$.

Лемма 2.1.

$$P(U = u, T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) =$$

$$= \frac{n!}{\prod_{i=1}^{8} k_i(u)!} \left(\frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}}\right)^u \left(\frac{p_{000}p_{110}}{p_{010}p_{100}}\right)^{t_1} \left(\frac{p_{000}p_{101}}{p_{001}p_{100}}\right)^{t_2} \left(\frac{p_{000}p_{011}}{p_{001}p_{010}}\right)^{t_3} \cdot$$

$$\cdot \left(\frac{p_{100}}{p_{000}}\right)^{t_4} \left(\frac{p_{010}}{p_{000}}\right)^{t_5} \left(\frac{p_{001}}{p_{000}}\right)^{t_6} p_{000}^n$$

где $k_1(u) = u$, $k_2(u) = t_1 - u$, $k_3(u) = t_2 - u$, $k_4(u) = t_3 - u$, $k_5(u) = t_4 - t_1 - t_2 + u$, $k_6(u) = t_5 - t_1 - t_3 + u$, $k_7(u) = t_6 - t_2 - t_3 + u$, $k_8(u) = n - u + t_1 + t_2 + t_3 - t_4 - t_5 - t_6$

Доказательство.

$$P(U = u, T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) =$$

$$= P\Bigg(\sum_{i=1}^n X_iY_iZ_i = u, \sum_{i=1}^n X_iY_i = t_1, \sum_{i=1}^n X_iZ_i = t_2, \sum_{i=1}^n Y_iZ_i = t_3,$$

$$\sum_{i=1}^n X_i = t_4, \sum_{i=1}^n Y_i = t_5, \sum_{i=1}^n Z_i = t_6\Bigg) =$$

$$= P\Bigg(\sum_{i=1}^n X_iY_iZ_i = u, \sum_{i=1}^n X_iY_i(1-Z_i) = t_1 - u, \sum_{i=1}^n X_i(1-Y_i)Z_i = t_2 - u,$$

$$\sum_{i=1}^n (1-X_i)Y_iZ_i = t_3 - u, \sum_{i=1}^n X_i(1-Y_i)(1-Z_i) = t_4 - t_1 - t_2 + u,$$

$$\sum_{i=1}^n (1-X_i)Y_i(1-Z_i) = t_5 - t_1 - t_3 + u, \sum_{i=1}^n (1-X_i)(1-Y_i)Z_i = t_6 - t_2 - t_3 + u,$$

$$\sum_{i=1}^n (1-X_i)(1-Y_i)(1-Z_i) = n - u + t_1 + t_2 + t_3 - t_4 - t_5 - t_6\Bigg) =$$

$$= \frac{n!}{\prod_{i=1}^8 k_i(u)!} p_{111}^u p_{110}^{t_1-u} p_{101}^{t_2-u} p_{011}^{t_3-u} p_{100}^{t_4-t_1-t_2+u} p_{010}^{t_5-t_1-t_3+u} p_{001}^{t_6-t_2-t_3+u} \cdot$$

$$\cdot p_{000}^{n-u+t_1+t_2+t_3-t_4-t_5-t_6} =$$

$$= \frac{n!}{\prod_{i=1}^8 k_i(u)!} \left(\frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}}\right)^u \left(\frac{p_{000}p_{110}}{p_{010}p_{100}}\right)^{t_1} \left(\frac{p_{000}p_{101}}{p_{001}p_{100}}\right)^{t_2} \left(\frac{p_{000}p_{011}}{p_{001}p_{010}}\right)^{t_3} \cdot$$

$$\cdot \left(\frac{p_{100}}{p_{000}}\right)^{t_4} \left(\frac{p_{010}}{p_{000}}\right)^{t_5} \left(\frac{p_{001}}{p_{000}}\right)^{t_6} p_{000}^n$$

□

Лемма 2.2.

$$P_{\theta_0}(U = u \mid T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) = \frac{\frac{1}{\prod_{i=1}^8 k_i(u)!}}{\sum_s \frac{1}{\prod_{i=1}^8 k_i(s)!}}$$

где в знаменателе вышеприведенной формулы суммирование ведется по таким s, что $0 \le k_i(s) \le n$ для всех $i = 1\dots, 8$.

Доказательство. Найдем маргинальное распределение вектора $(T_1, \dots, T_6)^T$:

$$P(T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) =$$

$$= \sum_s P(U = s, T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6)$$

Тогда условное распределение статистики U при условии $T_1 = t_1, \dots, T_6 = t_6$ можно записать в виде:

$$P(U = u \mid T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) =$$

$$= \frac{\frac{n!}{\prod_{i=1}^{8} k_i(u)!} \left(\frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}} \right)^u}{\sum_s \frac{n!}{\prod_{i=1}^{8} k_i(s)!} \left(\frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}} \right)^s}$$

При истинности гипотезы $\theta = \theta_0 = 0$ параметр $\frac{p_{001}p_{010}p_{100}p_{111}}{p_{000}p_{011}p_{101}p_{110}} = 1$.

$$P_{\theta_0}(U = u \mid T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = t_4, T_5 = t_5, T_6 = t_6) = \frac{\frac{1}{\prod_{i=1}^{8} k_i(u)!}}{\sum_s \frac{1}{\prod_{i=1}^{8} k_i(s)!}}$$

$\square$

Условную вероятность из леммы 2.2 можно эффективно вычислить на ЭВМ. Пусть $f(i) = \sum_{j=1}^{i} \ln(j)$. Тогда значение условной вероятности можно переписать в виде:

$$\frac{\frac{1}{\prod_{i=1}^{8} k_i(u)!}}{\sum_s \frac{1}{\prod_{i=1}^{8} k_i(s)!}} = \frac{\exp\left\{ \ln\left(\frac{1}{\prod_{i=1}^{8} k_i(u)!} \right) \right\}}{\sum_s \exp\left\{ \ln\left(\frac{1}{\prod_{i=1}^{8} k_i(s)!} \right) \right\}} = \frac{\exp\left\{ -\sum_{i=1}^{8} \ln(k_i(u)!) \right\}}{\sum_s \exp\left\{ -\sum_{i=1}^{8} \ln(k_i(s)!) \right\}} =$$

$$= \frac{\exp\left\{ -\sum_{i=1}^{8} f(k_i(u)) \right\}}{\sum_s \exp\left\{ -\sum_{i=1}^{8} f(k_i(s)) \right\}}$$

Полученное выражение удобно с позиции того, что современные ЭВМ умеют вычислять функцию

$$\varphi(x, i) = \frac{\exp\{x_i\}}{\sum_{j=1}^{n} \exp\{x_j\}}, \ x = (x_1, \dots, x_n)$$

За счет свойства

$$\varphi(x, i) = \frac{\exp\{x_i\}}{\sum_{j=1}^{n} \exp\{x_j\}} = \frac{\exp\{x_i - C\}}{\sum_{j=1}^{n} \exp\{x_j - C\}}, \text{ где } C = \max_{1 \le j \le n} x_j$$

удается избежать переполнения вещественного типа данных, связанного с экспонентой.

3 Тест проверки условной независимости в трехмерном нормальном распределении

Пусть

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix}, \begin{pmatrix} X_2 \\ Y_2 \\ Z_2 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \\ Z_n \end{pmatrix}$$

повторная выборка из распределения $(X, Y, Z)^T$ с трехмерным нормальным распределением $N(\mu, \Sigma)$, где μ – вектор математических ожиданий, а Σ – ковариационная матрица:

$$\mu = \begin{pmatrix} EX \\ EY \\ EZ \end{pmatrix}, \ \Sigma = \begin{pmatrix} \sigma_{XX} & \sigma_{XY} & \sigma_{XZ} \\ \sigma_{YX} & \sigma_{YY} & \sigma_{YZ} \\ \sigma_{ZX} & \sigma_{ZY} & \sigma_{ZZ} \end{pmatrix}$$

а также $\sigma_{XY} = E((X - EX)(Y - EY))$.

Определение 3.1. В трехмерном нормальном распределении частным коэффициентом корреляции Пирсона называется:

$$\rho^{XY \cdot Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{YZ}}{\sqrt{1 - \rho_{XY}^2}\sqrt{1 - \rho_{YZ}^2}}$$

Известно, что в трехмерном нормальном распределении частный коэффициент корреляции совпадает с условным коэффициентом корреляции и выполнено:

$$X \perp\!\!\!\perp Y \mid Z \text{ тогда и только тогда, когда } \rho^{XY \cdot Z} = 0$$

Определение 3.2. Выборочным коэффициентом корреляции Пирсона называется:

$$r_{XY} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

где

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Определение 3.3. Выборочным частным коэффициентом корреляции Пирсона называется:

$$r^{XY\cdot Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{1 - r_{XY}^2}\sqrt{1 - r_{YZ}^2}}$$

Известно, что при истинности гипотезы $\rho^{XY\cdot Z} = 0$ статистика

$$T = \sqrt{n-3}\frac{r^{XY\cdot Z}}{\sqrt{1 - (r^{XY\cdot Z})^2}}$$

имеет распределение Стьюдента с $n-3$ степенями свободы. Тогда тест уровня α проверки гипотезы $H : \rho^{XY\cdot Z} = 0$ против альтернативы $K : \rho^{XY\cdot Z} \neq 0$ имеет вид:

$$\varphi(t) = \begin{cases} 1, \ t < C_1 \text{ или } t > C_2 \\ 0, \ C_1 \leq t \leq C_2 \end{cases}$$

где константы C_1 и C_2, удовлетворяющие уравнениям $P(T < C_1) = \alpha/2$ и $P(T > C_2) = 1 - \alpha/2$, берутся из таблиц квантилей распределения Стьюдента с $n-3$ степенями свободы.