Теория к экзамену

Авторы: Минец Максим и Федотова Дарья

СУПЕР ВАЖНО!

Вероятность лежит в границах от 0 до 1
Дисперсия не может быть отрицательной
Уровень квантиля лежит в границах от 0 до 1
Среднее квадратическое не может быть отрицательной
Функция распределения определена на всей числовой прямой, принимает значения от 0 до 1, а еще она монотонно неубывает.
Функция плотности всегда неотрицательна, а интеграл от $-\infty$ до $+\infty$ всегда равен 1

ЧТО БЫЛО РАНЬШЕ...

Свойства мат. ожидания
Ec = c
$\mathrm{E}(\mathrm{c}\cdot\xi) = c \cdot E\xi$
$\mathrm{E}(\mathrm{c} + \xi) = c + E\xi$
$\mathrm{E}(\xi_1 + \xi_2) = E\xi_1 + E\xi_2$

Свойства дисперсии
Dc = 0
$\mathrm{D}(\mathrm{c}\cdot\xi) = c^2 \cdot D\xi$
$\mathrm{D}(\mathrm{c} + \xi) = D\xi$
$\mathrm{D}(\xi_1 + \xi_2) =$
$= \mathrm{D}\xi_1 + D\xi_2 + 2cov(\xi_1;\xi_2)$

Квантиль
z_ϕ - квантиль уровня ϕ или ϕ-квантиль
$F(z_\phi) = \phi$
$\int_{-\infty}^{z_\phi} f(x)dx = \phi$

Биномиальное распределение $Bi(n,p)$
$\mathrm{f(x)} = \mathrm{P}(\xi = k) = C_n^k p^k (1-p)^{n-k}$
$\mathrm{E}\xi = np$
$\mathrm{D}\xi = np(1-p)$

Пуассоновское распределение $\Pi(\lambda)$
$*\lambda = n \cdot p$
$\mathrm{f(x)} = \mathrm{P}(\xi = k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$
$\mathrm{E}\xi = D\xi = \lambda$

Геометрическое распределение $G(p)$
*до первой неудачи
$\mathrm{f(x)} = \mathrm{P}(\xi = k) = p \cdot q^{k-1}$
$\mathrm{E}\xi = \frac{1}{p}$
$\mathrm{D}\xi = \frac{q}{p^2}$

Равномерное распределение $R(a,b)$
$\mathrm{E}\xi = \frac{a+b}{2}$; $D\xi = \frac{(b-a)^2}{12}$

Функция плотности: $\begin{cases} \frac{1}{b-a}, x \in [a;b] \\ 0, \text{иначе} \end{cases}$

Экспоненциальное распределение $E(\lambda)$
$\mathrm{E}\xi = \frac{1}{\lambda}$; $D\xi = \frac{1}{\lambda^2}$
Функция плотности: $\begin{cases} \lambda \cdot e^{-\lambda x}, x \geq 0 \\ 0, \text{иначе} \end{cases}$

Нормальное распределение $N(m, \sigma^2)$
$m -$ среднее отклонение ; $\sigma -$ среднеквадратическое (стандартное) отклонение
$M(x) = m$; $D(x) = \sigma^2$
Функция плотности: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-m)^2}{2\sigma^2}}$
Функция распределения: $F(x) = \Phi(\frac{x-m}{\sigma})$

ПОШЛА ЖАРА

Интегральная теорема Муавра-Лапласа
X $\sim Bi(a,b)$. Если число испытаний по схеме Бернулли велико, то:

$$P(a \le X \le b) = \Phi_0(\frac{b-np}{\sqrt{npq}}) - \Phi_0(\frac{a-np}{\sqrt{npq}})$$

$$P(|X-b| \le a) = P(-a+b \le X \le a+b)$$

$$P(|X-E(X)| \le a) = 2\Phi_0(\frac{a}{\sqrt{npq}})$$

Неравенства Чебышева
P$\{|\xi| \ge \epsilon\} \le \frac{E\xi}{\epsilon}$
P$\{|\xi| \ge \epsilon\} \le \frac{E|\xi|^2}{\epsilon^2}$
P$\{|\xi - E\xi| \ge \epsilon\} \le \frac{E(\xi-E\xi)^2}{\epsilon^2} = \frac{D\xi}{\epsilon^2}$

Локальная теорема Муавра-Лапласа
Только для $X \sim Bi(a,b)$. Если число испытаний по схеме Бернулли велико, то:

$$P(X=k) = \frac{1}{\sqrt{2\pi \cdot npq}} e^{\frac{-(k-np)^2}{2npq}}$$

Центральная предельная теорема
Если СВ X представляет собой сумму очень большого (ну или не очень большого) числа взаимно независимых СВ, влияние каждой из которых на всю сумму ничтожно мало, то X имеет распределение, близкое к нормальному.

Формула свертки: $f_\eta(z) = \int\limits_{-\infty}^{+\infty} f_{\xi_1}(x) \cdot f_{\xi_2}(z-x)dx$

Ковариация: $cov(\xi,\eta) = E((\xi - E\xi)(\eta - E\eta)) = E(\xi\eta) - E\xi E\eta$

Если ξ и η независимы $cov(\xi,\eta) = 0$, иначе:
$E(\xi\eta) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x \cdot y \cdot f(x,y)dxdy$

cov$(\xi,\xi) = D\xi$
cov$(\xi+a, \eta+b) = cov(\xi,\eta)$
cov$(\mathrm{a}\xi, b\eta) = a \cdot b \cdot cov(\xi,\eta)$

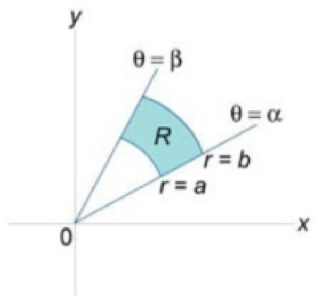
Коэффициент корреляции $\rho(\xi,\eta) = \frac{cov(\xi,\eta)}{\sqrt{D\xi \cdot D\eta}}$

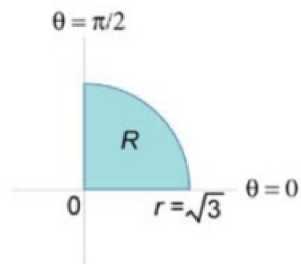
Многомерные СВ
Дано $f(x,y)$:
1. $f(x) = \int\limits_{-\infty}^{+\infty} f(x,y)dy$. Аналогично для $f(y)$.
2. $F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(t_1,t_2)dt_1 dt_2$
3. $F(x) = F(x,+\infty)$. Аналогично для $F(y)$.
4. $E(XY) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} x \cdot y \cdot f(x,y)dxdy$
5. $P(\xi \in D) = \int\limits_{x_1}^{x_2}\int\limits_{y_1}^{y_2} f(x,y)dxdy$

ТОЖЕ ВАЖНО

Переход к полярным координатам
$0 \le a \le r \le b, \alpha \le \theta \le \beta$, где $\beta - \alpha \le 2\pi$

$$\iint_R f(x,y)\,dx\,dy = \int\limits_{\alpha}^{\beta}\int\limits_{a}^{b} f(rcos\theta, rsin\theta)\,rd\,rd\theta$$





МАТСТАТ

Точечные оценки
Оценка мат. ожидания (выборочное среднее): $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Оценка дисперсии (смещенная, выборочная дисперсия): $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$
Оценка дисперсии (несмещенная, выборочная дисперсия): $\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$

Правило Стерджесса: $k = 1 + log_2 N$
Логарифм округляется в нижнюю сторону.

Свойства оценок
1. Несмещенность: $E\hat{\theta} = \theta$
Например, рассмотрим оценку математического ожидания, которая известна как выборочное среднее: $\hat{m} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$E\overline{X} = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n}nEx_1 = m$$

Существует еще одна известная смещенная оценка для дисперсии, которую называют выборочной дисперсией: $S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$

$$ES^2 = \frac{n-1}{n}\sigma^2$$

А также есть еще оценка для дисперсии, которая уже является несмещенной – это несмещенная выборочная дисперсия: $\tilde{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2$

$$E\tilde{S}^2 = \sigma^2$$

2. Ассимптотическая несмещенность:
Оценка $\hat{\theta}$ называется асимптотически несмещенной, если ее математическое ожидание стремится к параметру θ когда объем выборки стремится к бесконечности.

$$\lim_{n\to\infty} E\hat{\theta} = \theta$$

Рассмотрим выборочную дисперсию, выписанную немного выше. Она смещенная, но если n стремится к бесконечности, то $ES^2 = \frac{n-1}{n}\sigma^2 \to \sigma^2$, а значит, она является асимптотически несмещенной.

3. Состоятельность:
Оценка $\hat{\theta}$ называется состоятельной, если она сходится по вероятности к θ.

$$\hat{\theta} \xrightarrow{P} \theta, n \to \infty$$

$$\forall \varepsilon > 0 : P(|\hat{\theta} - \theta| > \varepsilon) \to 0, n \to \infty$$

$$\forall \varepsilon > 0 : P(|\hat{\theta} - \theta| \le \varepsilon) \to 1, n \to \infty$$

Если вспомнить здесь еще неравенство Чебышева, то можно получить достаточное условие состоятельности:

$$\forall \varepsilon > 0 : P(|\hat{\theta} - E\hat{\theta}| \le \varepsilon) \ge 1 - \frac{D\hat{\theta}}{\varepsilon^2}, n \to \infty$$

Два условия для состоятельности:
1. $\lim\limits_{n\to\infty} E\hat{\theta} = \theta$
2. $\lim\limits_{n\to\infty} D\hat{\theta} = 0$
Если все это выполняется, то при $n \to \infty$:

$$P(|\hat{\theta} - \theta| \le \varepsilon) \ge 1$$

А так как вероятность не может больше чем 1, то получим наше первоначальное определение состоятельности:

$$P(|\hat{\theta} - \theta| \le \varepsilon) \to 1, n \to \infty$$

4. Сильная состоятельность:
Оценка $\hat{\theta}$ называется сильно состоятельной, если она сходится почти наверное к θ.

$$\hat{\theta} \xrightarrow{\text{п.н.}} \theta, n \to \infty$$

5. Эффективность:
Оценка $\hat{\theta}$ называется эффективной, если она несмещенная и ее дисперсия является наименьшей среди всех несмещенных оценок для этого параметра θ.

Методы нахождения и построения оценок

Метод моментов
Рассморим начальные моменты $\mu_j = EX^j, j = \overline{1,k}$, где k — количество параметров, которые нужно оценить. X здесь — это случайная величина, порождающая выборку, то есть величина, у которой распределение такое же, как у всех случайных величин выборки. Понятно, что эти моменты будут зависеть от неизвестных θ_i. А теперь приравняем каждый момент соответствующему ему начальному выборочному моменту, которые равны $\hat{\mu}_j = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^j$. И получим следующую систему:

$$\begin{cases} \mu_1 = \hat{\mu}_1 \\ \vdots \\ \mu_k = \hat{\mu}_k \end{cases}$$

Если система решается и решается однозначно, то ее решение будет являться оценками параметров $\theta_1, \dots, \theta_k$.

И рассмотрим пример, как это используется. Пусть есть выборка из n случайных величин, которые имеют нормальное распределение с неивестными параметрами μ и σ: $X_1, \dots, X_n \sim N(\theta_1, \theta_2^2)$. Тогда составим систему для того, чтобы оценить θ_1 и θ_2.

$$\begin{cases} EX = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i \\ EX^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 \end{cases}$$

$EX = \theta_1$ из условия, также из условия известно $DX = \theta_2^2$. Тогда можно сказать, чему равно EX^2 из определения дисперсии. Так как $DX = EX^2 - (EX)^2$, то $EX^2 = DX + (EX)^2 = \theta_2^2 + \theta_1^2$. А также известно, чему равен первый выборочный момент — это выборочное среднее. Запишем все это в систему:

$$\begin{cases} \theta_1 = \bar{X} \\ \theta_2^2 + \theta_1^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 \end{cases} \Rightarrow \begin{cases} \theta_1 = \bar{X} \\ \theta_2^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 - \bar{X}^2 \end{cases}$$

Это и есть полученные оценки:

$$\begin{cases} \hat{\theta}_1 = \bar{X} \\ \hat{\theta}_2^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 - \bar{X}^2 \end{cases}$$

Метод максимального правдоподобия

Введем функцию правдоподобия

$$L(X_1, \ldots, X_n, \theta_1, \ldots, \theta_k) = \begin{cases} \displaystyle\prod_{i=1}^{n} f(x_i, \theta_1, \ldots, \theta_k), \text{если распределение непрерывное} \\ \displaystyle\prod_{i=1}^{n} P(X = x_i, \theta_1, \ldots, \theta_k), \text{если распределение дискретное} \end{cases}$$

Оценкой максимального правдоподобия называется максимум этой функции, то есть $\hat{\theta} = \underset{\theta \in \mathbb{R}^k}{\operatorname{argmax}} L(X_1, \ldots, X_n, \theta)$, где $\theta = (\theta_1, \ldots, \theta_k)^T$

Так как функция правдоподобия и логарифмическая функция правдоподобия ($\ln L(X_1, \ldots, X_n, \theta)$) имеют одинаковые точки максимума, то часто удобно использовать именно логарифмическую функцию правдоподобия. И тогда оценки можно найти, продифференцировав функцию относительно каждого оцениваемого параметра:

$$\begin{cases} \dfrac{\partial \ln L(X_1, \ldots, X_n, \theta)}{\partial \theta_1} = 0 \\ \vdots \\ \dfrac{\partial \ln L(X_1, \ldots, X_n, \theta)}{\partial \theta_k} = 0 \end{cases}$$

Если решить эту систему, получатся оценки максимального правдоподобия.

ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ

Теорема Фишера:
Если есть какая-то выборка $X_1, ..., X_n$, все элементы которой имеют нормальное распределение $N(\mu, \sigma^2)$, то:
1. среднее выборочное $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ или $\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} \sim N(0,1)$
2. $\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$, ($S^2$ – смещенная выборочная дисперсия)
3. $\overline{X}$ и S^2 независимы
4. $\frac{(\overline{X}-\mu)\sqrt{n-1}}{\sqrt{S^2}} \sim t(n-1)$

Интервалы для выборки из нормального распределения:
1. Оценка μ при известной σ^2:

$$G = \frac{(\overline{X}-\mu)\sqrt{n}}{\sigma^2} \sim N(0,1)$$

$$P\left(\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

2. Оценка σ^2 при известном μ:

$$G = \sum_{i=1}^{n}(\frac{X_i-\mu}{\sqrt{S}})^2 \sim \chi^2(n)$$

$$P\left(\frac{\sum_{i=1}^{n}(\frac{X_i-\mu}{\sqrt{S}})^2}{\chi^2_{1-\alpha/2;n}} < \sigma^2 < \frac{\sum_{i=1}^{n}(\frac{X_i-\mu}{\sqrt{S}})^2}{\chi^2_{\alpha/2;n}}\right) = 1-\alpha$$

3. Оценка μ при неизвестной σ^2:

$$G = \frac{(\overline{X}-\mu)\sqrt{n-1}}{S^2} \sim t(n-1)$$

$$P\left(\overline{X} - t_{1-\alpha/2;n-1}\frac{\sqrt{S^2}}{\sqrt{n-1}} < \mu < \overline{X} + t_{1-\alpha/2;n-1}\frac{\sqrt{S^2}}{\sqrt{n-1}}\right) = 1-\alpha$$

4. Оценка σ^2 при неизвестном μ:

$$G = \frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$P\left(\frac{nS^2}{\chi^2_{1-\alpha/2;n-1}} < \sigma^2 < \frac{nS^2}{\chi^2_{\alpha/2;n-1}}\right) = 1-\alpha$$

ДОПОЛНИТЕЛЬНО К ДОВЕРИТЕЛЬНЫМ ИНТЕРВАЛАМ

(РАЗНОСТЬ МАТ. ОЖИДАНИЙ)

1. Две выборки X, Y:
σ_1 и σ_2 известны и не равны.
m_1 и m_2 неизвестны.
Необходимо оценить $\theta = m_1 - m_2$.
$\hat{\theta} = \overline{X} - \overline{Y}$
$E\hat{\theta} = E(\overline{X} - \overline{Y}) = m_1 - m_2$
$D\hat{\theta} = D(\overline{X} - \overline{Y}) = D(\overline{X}) + D(\overline{Y}) = D\left(\frac{1}{n_1}\sum_{i=1}^{n_1} x_i\right) + D\left(\frac{1}{n_2}\sum_{i=1}^{n_2} y_i\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$
Центральная статистика: $G = \dfrac{\overline{X} - \overline{Y} - (m_1 - m_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$
Доверительный интервал:
$$P\left(\overline{X} - \overline{Y} - z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < m_1 - m_2 < \overline{X} - \overline{Y} + z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) = 1 - \alpha$$

2. Две выборки X, Y:
σ_1 и σ_2 неизвестны и $\sigma_1 = \sigma_2 = \sigma$.
m_1 и m_2 неизвестны.
Необходимо оценить $\theta = m_1 - m_2$.

$\hat{\theta} = \overline{X} - \overline{Y}$
$E\hat{\theta} = E(\overline{X} - \overline{Y}) = m_1 - m_2$
$D\hat{\theta} = D(\overline{X} - \overline{Y}) = D\left(\frac{1}{n_1}\sum_{i=1}^{n_1} x_i\right) + D\left(\frac{1}{n_2}\sum_{i=1}^{n_2} y_i\right) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_1}(x_i - \overline{X})^2 + \sum_{i=1}^{n_2}(y_i - \overline{Y})^2}{n_1 + n_2 - 2} = S_{XY}^2$$
Центральная статистика: $G = \dfrac{\overline{X} - \overline{Y} - (m_1 - m_2)}{S_{XY}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$

Примем $n_1 + n_2 - 2$ за n.

Доверительный интервал:
$$P\left(\overline{X} - \overline{Y} - t_{1-\alpha/2;n} \cdot S_{XY}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < m_1 - m_2 < \overline{X} - \overline{Y} + t_{1-\alpha/2;n} \cdot S_{XY}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 1 - \alpha$$

ЕЩЕ ДОПОЛНИТЕЛЬНО (БИНОМИАЛЬНОЕ РАСПРЕДЕЛЕНИЕ)

$$P\left(\hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1 - \alpha$$

Здесь $\hat{p} = \dfrac{m}{n}$

МЕТОД МОНТЕ-КАРЛО

Если функция $f(x)$ ограничена и интегрируема на $[0,1]$, то интеграл $I = \int\limits_0^1 f(x)dx$ можно рассматривать как мат. ожидание $I = Ef(\xi)$, где ξ - СВ, распределенная равномерно на $[0,1]$ $\Rightarrow$ по УЗБЧ: $\dfrac{1}{n}\sum\limits_{k=1}^{n} f(\xi_k) \xrightarrow{\text{п.н.}} I$ в предположении, что $\xi_1, \xi_2, ..., \xi_n$ - независимы и одинаково распределены по закону $R(0,1)$. Поэтому при достаточно большом n

$I = \int\limits_0^1 f(x)dx \approx \dfrac{1}{n}\sum\limits_{k=1}^{n} f(\xi_k)$

В случае $|f(x)| \leq c$ при $x \in [0,1]$ ЦПТ приводит к следующей приближенной оценке:

$$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} f(\xi_k) - I\right| < \Delta\right) \geq 2\Phi_0\left(\Delta\frac{\sqrt{n}}{c}\right)$$

ПРОВЕРКА ГИПОТЕЗ

Алгоритм проверки гипотез:

1) Формулируем основную гипотезу H_0 и альтернативную гипотезу H_1
2) Выбираем уровень значимости α
3) Выбрать статистику $T(x)$
4) Статистическое распределение $T(x)$ при справедливости гипотезы H_0
5) Строим доверительную и критическую области
6) Вычисляем $T(x)$
7) Принимаем решение: принимаем гипотезу H_0 или отвергаем гипотезу H_0 в пользу H_1

Проверка гипотез для математического ожидания

Случайная величина X имеет нормальное распределение. Требуется, используя реализацию выборки, проверить гипотезу H_0, состоящую в том, что $m_X =$ некоторому фиксированному числу, против гипотезы H1 о том, что $m_X \neq$ числу (в табличке $\gamma = 1 - \frac{\alpha}{2}$)

Предположение	**Статистика Z критерия**	**Распределение $F(z\|H_0)$**	**Доверительная область G критерия**
σ_X^2 известна	$\frac{(\overline{X}-m_0)\sqrt{n}}{\sigma_X}$	$\mathcal{N}(0;1)$	$[-u_\gamma, u_\gamma]$
σ_X^2 неизвестна	$\frac{(\overline{X}-m_0)\sqrt{n-1}}{\sqrt{s^2}}$	$\mathbf{t}(n-1)$	$[-t_{\gamma,n-1}, t_{\gamma,n-1}]$

Проверка гипотез для дисперсии

Пусть случайная величина X нормально распределена, а ее дисперсия неизвестна. Требуется на основе реализации выборки, порожденной случайной величиной, проверить гипотезу H_0 о том, что $\sigma_X^2 =$ некоторому фиксированному числу, против гипотезы H1 о том, что $\sigma_X^2 \neq$ числу (в табличке $\gamma = 1 - \frac{\alpha}{2}$)

Предположение	**Статистика Z критерия**	**Распределение $F(z\|H_0)$**	**Доверительная область G критерия**
m_X известно	$\frac{\sum_{k=1}^{n}(X_k-m_X)^2}{\sigma_0^2}$	$\chi^2(n)$	$[\chi^2_{1-\gamma,n}, \chi^2_{\gamma,n}]$
m_X неизвестно	$\frac{ns^2}{\sigma_0^2}$	$\chi^2(n-1)$	$[\chi^2_{1-\gamma,n-1}, \chi^2_{\gamma,n-1}]$

Критерий проверки некоррелированности двух случайных величин

Пусть есть выборка из пар случайных величин $(X_1, Y_1), \ldots, (X_n, Y_n)$, порожденная гауссовским случайным вектором (X, Y)

1. $H_0 : \rho_{xy} = 0$ - величины некорелированы
   $H_1 : \rho_{xy} < 0$ - величины корелированы отрицательно
   $H_2 : \rho_{xy} > 0$ - величины корелированы положительно
   $H_3 : \rho_{xy} \neq 0$ - величины корелированы
2. Выбираем уровень значимости: некоторое α
3. $T((X_1, Y_1), \ldots, (X_n, Y_n)) = \dfrac{\sqrt{n-2}\hat{\rho}_{xy}}{\sqrt{1-\hat{\rho}_{xy}^2}}$

$$\hat{\rho}_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

4. $T((X_1, Y_1), \ldots, (X_n, Y_n))|_{H_0} \sim t(n-2)$

5. Для H_1 доверительная область в границах $[t_{\alpha,n-2}; +\infty)$, критическая область в границах $(-\infty; t_{\alpha,n-2})$
Для H_2 доверительная область в границах $(-\infty; t_{1-\alpha,n-2}]$, критическая область в границах $(t_{1-\alpha,n-2}; +\infty)$
Для H_3 доверительная область в границах $[t_{\alpha/2,n-2}; t_{1-\alpha/2,n-2}]$, критическая область в границах
$(-\infty; t_{\alpha/2,n-2}) \cup (t_{1-\alpha/2,n-2}; +\infty)$

Критерий проверки на независимость
СВ X и Y являются независимы, когда $F(x,y) = F_X(x) \cdot F_Y(y)$, $\forall x, y$
Строим табличку (будет полезна в дальнейших вычислениях), где на пересечении строк и столбцов - количество наблюдений, которые попали в пересечение строка-столбец
Алгоритм для решения данной гипотезы:
1) Основная гипотеза H_0: $P(A = A_i, B = B_i) = P(A = A_i) \cdot P(B = B_i)$, то есть A и B - независимые. Альтернативная гипотеза H_1: A и B - зависимые, то есть $\neq$
2) Выбираем уровень значимости α
3) $\bar{\chi^2} = n\big(\sum\limits_{i=1}^{s}\sum\limits_{j=1}^{k} \frac{(n_{ij}^2)}{n_{i\cdot}n_{\cdot j}} - 1\big)$

s - число строк; k - число столбцов; $n_{i\cdot} = \sum\limits_{j=1}^{s} n_{ij}$; $n_{\cdot j} = \sum\limits_{i=1}^{k} n_{ij}$
4) $\bar{\chi^2}\big|_{H_0} \sim \chi^2((k-1)(s-1))$
5) Доверительная область лежит в границе $-\infty$ от до $\chi^2_{1-\alpha;(k-1)(s-1)}$, а критическая - от $\chi^2_{1-\alpha;(k-1)(s-1)}$ до ∞
6) Вычисляем $\bar{\chi^2}$
7) Принимаем или отвергаем гипотезу

НЕМНОГО ЗАДАЧЕК

Задача 1. Неравенство Чебышева. Муавр-Лаплас
В продукции цеха детали отличного качества составляют 80%. В каких пределах с вероятностью 0,99 будет находиться количество деталей отличного качества, если взять 10000 деталей? Построить оценку с помощью неравенства Чебышёва и по теореме Муавра- Лапласа.

Решение

ξ_i	0	1
p	0,2	0,8

$\xi \sim Bi(10000; 0,8)$
$E\xi = 8000$
$D\xi = 8000 \cdot 0,2 = 1600$
По неравенству Чебышева: $P(|\xi - E\xi| > \varepsilon) \leq \frac{D\xi}{\varepsilon^2}$
$P(|\xi - E\xi| \leq \varepsilon) \geq 1 - \frac{D\xi}{\varepsilon^2} = 0,99$
$\frac{D\xi}{\varepsilon^2} = 0,01 \Rightarrow \varepsilon = 10\sqrt{D\xi} = 400 \Rightarrow |\xi - E\xi| \leq 400 \Rightarrow$
$\Rightarrow$ т.к. $E\xi = 8000$, то получаем интервал [7600;8400].

Теорема Муавра-Лапласа:
$P(|\xi - E\xi| < \varepsilon) = 2\Phi_0(\frac{\varepsilon}{\sqrt{D\xi}}) = 2\Phi_0(\frac{\varepsilon}{40})$

$2\Phi_0(\frac{\varepsilon}{40}) = 0,99 \Rightarrow \Phi_0(\frac{\varepsilon}{40}) = 0,495 \Rightarrow \frac{\varepsilon}{40} = 2,6 \Rightarrow \varepsilon = 104$

$|\xi - E\xi| \leq 104 \Rightarrow$ получаем интервал [7896;8104]

Задача 2. Чебышев

Пусть ξ_1 - число выпадений герба при 10-ти подбрасываниях монеты, а ξ_2 - число выпавших очков на грани тетраэдра (грани перенумерованы числами 1, 2, 3, 4) при его однократном подбрасывании. Оценить вероятность осуществления неравенства $\xi_1 + \xi_2 < 10$. Решить задачу, используя 1-е и 2-е неравенства Чебышева.

Неравенство Чебышева
Пусть r-ый абсолютный момент случайной величины ξ конечен, т.е. $\mathrm{E}|\xi| < \infty$.
Тогда для всех $\varepsilon > 0$ выполняется неравенство
$\mathrm{P}\{|\xi| \geq \varepsilon\} \leq \frac{\mathrm{E}|\xi|^r}{\varepsilon^r}$

Решение

ξ_1 - число выпадений герба при 10-ти подбрасываниях монеты
$\xi_1 \sim \mathrm{Bi}(10; 0{,}5)$
$n = 10$; $p = 0{,}5$; $q = 1 - p = 0{,}5$
$\mathrm{E}\xi_1 = n \cdot p = 10 \cdot 0{,}5 = 5$
$\mathrm{D}\xi_1 = n \cdot p \cdot q = 5 \cdot 0{,}5 = 2{,}5$
$\mathrm{E}\xi_1^2 = \mathrm{D}\xi_1 + (\mathrm{E}\xi_1)^2 = np(q + np) = 27{,}5$
ξ_2 - число выпавших очков на грани тетраэдра (грани перенумерованы числами 1, 2, 3, 4) при его однократном подбрасывании

ξ_2	1	2	3	4
P	0,25	0,25	0,25	0,25

$\mathrm{E}\xi_2 = 1 \cdot 0,25 + 2 \cdot 0,25 + 3 \cdot 0,25 + 4 \cdot 0,25 = 0{,}25 + 0{,}5 + 0{,}75 + 1 = 2{,}5$
$\mathrm{E}\xi_2^2 = 1 \cdot 0,25 + 4 \cdot 0,25 + 9 \cdot 0,25 + 16 \cdot 0,25 = 0{,}25 + 1 + 2{,}25 + 4 = 7{,}5$
$\mathrm{D}\xi_2 = \mathrm{E}\xi_2^2 - (\mathrm{E}\xi_2)^2 = 7{,}5 - 6{,}25 = 1{,}25$

Найдем математическое ожидание и диспресию суммы ξ_1 и ξ_2:
$\mathrm{E}(\xi_1 + \xi_2) = \mathrm{E}\xi_1 + \mathrm{E}\xi_2 = 5 + 2{,}5 = 7{,}5$
$\mathrm{D}(\xi_1 + \xi_2) = \mathrm{D}\xi_1 + \mathrm{D}\xi_2 + 2cov(\xi_1; \xi_2)$
Здесь ковариация будет равна нулю, так как ξ_1 и ξ_2 являются независимыми событиями.
$\mathrm{D}(\xi_1 + \xi_2) = \mathrm{D}\xi_1 + \mathrm{D}\xi_2 + 0 = 2{,}5 + 1{,}25 = 3{,}75$

Таким образом, получаем, что $(\xi_1 + \xi_2) \sim \mathrm{N}(7{,}5; 3{,}75)$, так как сумма двух любых распределений дает гауссовское распределение по ЦПТ.

Найдем вероятность осуществления неравенства $\xi_1 + \xi_2 < 10$ по 1-ому и 2-ому неравенствам Чебышева.

По первому неравентсву Чебышева:
$\mathrm{P}\{\xi \geq \varepsilon\} \leq \frac{\mathrm{E}\xi}{\varepsilon}$
$\mathrm{P}\{(\xi_1 + \xi_2) \geq 10\} \leq \frac{\mathrm{E}(\xi_1+\xi_2)}{10} = \frac{7{,}5}{10} = 0{,}75$
$\mathrm{P}\{(\xi_1 + \xi_2) < 10\} = 1 - \mathrm{P}\{(\xi_1 + \xi_2) \geq 10\} = 1 - 0{,}75 = 0{,}25$

По второму неравенству Чебышева:
$\mathrm{P}\{|\xi| \geq \varepsilon\} \leq \frac{\mathrm{E}\xi^2}{\varepsilon^2}$

Вычислим $\mathrm{E}(\xi_1 + \xi_2)^2$:
$\mathrm{E}(\xi_1 + \xi_2)^2 = \mathrm{E}(\xi_1^2 + 2\xi_1\xi_2 + \xi_2^2) = \mathrm{E}\xi_1^2 + 2{\cdot}\mathrm{E}\xi_1\xi_2 + \mathrm{E}\xi_2^2$
Так как ξ_1 и ξ_2 независимые, то $\mathrm{E}\xi_1\xi_2 = \mathrm{E}\xi_1 \cdot \mathrm{E}\xi_2$.
$\mathrm{E}(\xi_1 + \xi_2)^2 = \mathrm{E}\xi_1^2 + 2{\cdot}\ \mathrm{E}\xi_1\ \mathrm{E}\xi_2 + \mathrm{E}\xi_2^2 = 27{,}5 + 2{\cdot}5{\cdot}2{,}5 + 7{,}5 = 27{,}5 + 25 + 7{,}5 = 60$
$\mathrm{P}\{|\xi_1 + \xi_2| \geq 10\} \leq \frac{\mathrm{E}(\xi_1+\xi_2)^2}{10^2} = \frac{60}{100} = 0{,}6$
$\mathrm{P}\{|\xi_1 + \xi_2| < 10\} = 1 - \mathrm{P}\{|\xi_1 + \xi_2| \geq 10\} = 1 - 0{,}6 = 0{,}4$

Ответ: 0,25 и 0,4

Задача 3. Муавр-Лаплас

Вероятность появления бракованной детали в партии из 1000 деталей равна 0,05. Найти нижнюю и верхнюю границы числа дефектных деталей в этой партии с вероятностью 0,9.

Решение

Определим событие A:
$A = \{$Появление бракованной детали в партии из 1000 деталей$\}$

По теореме о вероятности отклонения относительной частоты события от постоянной вероятности в независимых испытаниях: Если в схеме n независимых испытаний событие A наступает в каждом из них с вероятностью p, а $\frac{m}{n}$ относительная частота события A, то для любого заданного числа $\varepsilon > 0$:

$$P(|\frac{m}{n} - p| \leq \varepsilon) \approx 2\Phi(\varepsilon\sqrt{\frac{n}{pq}})$$

Формула выводится из интегральной теоремы Лапласа.
Так как вероятность появления случайного события A в каждом испытании постоянна ($p = 0.05$), то вероятность того, что в испытаниях событие A наступит не менее m_1 и не более m_2 раз (от m_1 до m_2 раз включительно), приблизительно равна:

$$2\Phi(\varepsilon\sqrt{\frac{1000}{0.05 * 0.95}}) = 0.9$$

$$2\Phi(145\varepsilon) = 0.9$$

$$\Phi(145\varepsilon) = 0.45$$

По таблице функции Лапласа $\Phi(1.65) = 0.45 \Rightarrow \varepsilon \approx 0.011$.
Таким орбазом, с вероятность 0.9 отклонение от частоты бракованных деталей от вероятности 0.05 удовлетворяет ранее упомянутому неравенству:

$$|\frac{m}{n} - p| \leq \varepsilon$$

$$|\frac{m}{1000} - 0.05| \leq 0.011$$

$$0.039 \leq \frac{m}{1000} \leq 0.061$$

$$39 \leq m \leq 61$$

Следовательно нижняя граница $m_1 = 39$, верхняя граница $m_2 = 61$
Ответ: $m_1 = 39$, $m_2 = 61$.

Задача 4. Вещественные корни
Вычислить вероятность того, что корни уравнения $x^2 + 2\xi x + \eta = 0$ вещественны, если случайный вектор (ξ, η) равномерно распределен на области $D = [-1, 1] \times [-1, 1]$.

Решение
Два случая с дискриминантом:
$D = 0 : x_0 = -\dfrac{2\xi}{2}$
$D > 0 : x_{1,2} = \dfrac{-2\xi \pm \sqrt{4\xi^2 - 4\eta}}{2}$

$\xi \sim R(-1, 1)$ и $\eta \sim R(-1, 1)$

$E\xi = E\eta = 0$ и $D\xi = D\eta = \dfrac{1}{3}$

$E(4\xi^2 - 4\eta) = 4E(\xi^2) - 4E\eta = 4(\dfrac{1}{3} + 0) - 4 \cdot 0 = \dfrac{4}{3}$
$D(4\xi^2 - 4\eta) = 16D(\xi^2) + 16D\eta = 16(E(\xi^4) - (E(\xi^2))^2) + 16D\eta = 16\left(\int\limits_{-1}^{1} \dfrac{x^4}{2}dx - (\dfrac{1}{3})^2\right) + 16 \cdot \dfrac{1}{3} =$
$16\left(\dfrac{x^5}{10}\Big|_{-1}^{1} - \dfrac{1}{9}\right) + \dfrac{16}{3} = 16 \cdot \dfrac{4}{45} + \dfrac{16}{3} = 6\dfrac{34}{45}$

$P(4\xi^2 - 4\eta \geq 0) = \Phi_0(+\infty) - \Phi_0\left(\dfrac{0 - \dfrac{4}{3}}{\sqrt{6\dfrac{34}{45}}}\right) = 0.5 + \Phi_0(0.5) = 0.5 + 0.1915 = 0.6915$

Задача 5. Похоже на номер 4 из ИДЗ
Вектор (ξ, η) имеет математическое ожиданиее $(1, 2)$ и ковариационную матрицу $\begin{pmatrix} 1.5 & 3 \\ 3 & 7.5 \end{pmatrix}$.
$\zeta = 4\xi - 3\eta$
Найти: МО и дисперсию ζ, $P\{0 < \zeta < 2\}$

Решение
Если вектор (ξ, η) распределен нормально, то каждая его компонента также имеет нормальное распределение. По условию нам дан вектор математических ожиданий и ковариационная матрица, из которой мы можем вычленить дисперсии случайных величин ξ и η, а также их ковариации:
$E\xi = 1; E\eta = 2; D\xi = 1.5; D\eta = 7.5$
$cov(\xi, \eta) = cov(\eta, \xi) = 3$
Теперь рассмотрим линейную комбинацию $L = 4\xi - 3\eta$. Так как ξ и η распределены нормально и коэффициенты при случайных величинах ненулевые, то можно утверждать, что и величина L распределена нормально (по свойству нормального распределения).
Чтобы найти вероятность попадания в интервал $(0, +\infty)$, необходимо найти математическое ожидание и дисперсию L.
$EL = E(4\xi - 3\eta) = 4E(\xi) - 3E(\eta) = 4 \cdot 1 - 3 \cdot 2 = -2$

$DL = D(4\xi - 3\eta) = D(3\xi) + D(-3\eta) + 2cov(4\xi, -3\eta) =$
$= 16D\eta + 9D\xi - 2 \cdot 4 \cdot 3 \cdot cov(\xi, \eta) = 16 \cdot 1.5 + 3 \cdot 7.5 - 24 \cdot 3 = 24 + 67.5 - 72 = 19.5$

$P\{0 < 4\xi - 3\eta < 2\} = \Phi_0(\frac{2-EL}{\sqrt{DL}}) - \Phi_0(\frac{0-EL}{\sqrt{DL}}) =$
$= \Phi_0(\frac{2+2}{\sqrt{19.5}}) - \Phi_0(\frac{0+2}{\sqrt{19.5}}) = \Phi_0(0.9) + \Phi_0(0.45) = 0,3159 + 0,1736 = 0,1423$

Задача 6. Среднее объема (мат ожидания) бочки

Бочка имеет форму цилиндра радиуса r и высоты h. При изготовлении бочки на размеры разрешен допуск, так что радиус равен $r + \xi$, а высота равна $h + \eta$, причем случайные ошибки ξ и η независимы и распределены равномерно на $[-\Delta, \Delta]$ и $[-\delta, \delta]$ соответсвенно. Найти среднее значение объема бочки.

Решение

$EV = E(\pi r^2 h)$, где $r = r + \xi, h = h + \eta$

$$E(\pi r^2 h) = E(\pi(r+\xi)^2(h+\eta)) = \pi E(r^2 + 2r\xi + \xi^2) \cdot E(h+\eta) =$$
$$= \pi(E(r^2) + E(2r\xi) + E(\xi^2))(Eh + E\eta) = \pi(r^2 + 2rE\xi + E(\xi^2))(h + E\eta) =$$
$$= \pi(r^2 + 2r\frac{\Delta + (-\Delta)}{2} + E(\xi^2))(h + \frac{\delta + (-\delta)}{2}) = \pi h(r^2 + E(\xi^2)) =$$
$$= \pi h(r^2 + \int\limits_{-\Delta}^{\Delta} \frac{x^2}{2\Delta} dx) = \pi h(r^2 + \frac{1}{2\Delta}\frac{x^3}{3}\Bigg|_{\Delta}^{\Delta}) = \pi h(r^2 + \frac{1}{2\Delta}\frac{2\Delta^3}{3}) = \pi h(r^2 + \frac{\Delta^2}{3})$$

Задача 7. ЦПТ

СВ $\xi_1, \cdots, \xi_36$ - независимы, и каждая из них распределена по экспоненциальному закону с параметром 2. Найдите $P(\sum\limits_{i=1}^{36} \xi_i > 20)$

Решение

$\xi_1, \cdots, \xi_36 \sim E(2)$ и независимые.

$E(\sum\limits_{i=1}^{36} \xi_i) = 36 \cdot E\xi_1 = 36 \cdot \frac{1}{2} = 18$

$D(\sum\limits_{i=1}^{36} \xi_i) = 36 \cdot D\xi_i = 36 \cdot \frac{1}{4} = 9$

$\sum\limits_{i=1}^{36} \xi_i \sim N(18; 9)$

$P(\sum\limits_{i=1}^{36} \xi_i > 20) = \Phi_0(+\infty) - \Phi_0(\frac{20-18}{3}) = 0,5 - \Phi_0(\frac{2}{3} = 0,5 - 0,2453 = 0,2547)$

Задача 8. Полярные координаты
Случайные величины X и Y независимы и имеют гауссовские распределения с нулевыми математическими ожиданиями и дисперсиями равными 4. Найти вероятность попадания точки с координатами (X, Y) в круг радиуса 3 с центром в начале координат.

Решение
$P((X,Y) \in K) = \iint_K f(x,y)\,dx\,dy = \iint_K \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x^2+y^2)-\mu}{2\sigma^2}} dxdy = *$
Переходим к полярным координатам:
$\begin{cases} x = \rho cos\varphi \\ y = \rho sin\varphi \end{cases}$
$= \int\limits_0^{2\pi}\int\limits_0^{3} \dfrac{1}{2\sqrt{2\pi}} e^{\frac{-\rho^2(cos^2(\varphi)+sin^2(\varphi))}{2\cdot 4}} d\rho d\varphi = \int\limits_0^{2\pi}\int\limits_0^{3} \dfrac{1}{2\sqrt{2\pi}} e^{\frac{-\rho^2}{8}} d\rho d\varphi =$ здесь нужно досчитать

Задача 9. Вероятность попадания в круг
Плотность распределения случайного вектора$\varepsilon = (\xi, \eta)$
$f(x,y) = \begin{cases} c \cdot (R - \sqrt{x^2+y^2}), x^2+y^2 \le R^2 \\ 0, \text{в остальных случаях} \end{cases}$
Найти: константу c и вероятность попадания точки с координатами (ξ, η) в круг радиуса 1 с центром в начале координат, если $R = 2$.

Решение
1) Найдем константу из условия нормировки:
$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)dxdy = 1$

$\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} c \cdot (R - \sqrt{x^2+y^2})dxdy = *$
Переходим к полярным координатам: $x = \rho \cdot \cos\phi; y = \rho \cdot \sin\phi$
$|J| = \rho$ - Якобиан перехода (первый интеграл - угол; второй интеграл - радиус)
$\int\limits_0^{2\pi} \cdots d\phi = 2\pi - 0$, так как функция от ϕ не зависит
$* = \int\limits_0^{2\pi}\int\limits_0^{R} c \cdot (R-\rho) \cdot J\ dxdy\ (J = \rho) = 2\pi \cdot c \int\limits_0^{R} (R \cdot \rho - \rho^2) d\rho = 2\pi c(\frac{R^3}{2} - \frac{R^3}{3}) = 2\pi c \cdot \frac{R^3}{6} = \frac{\pi c R^3}{3}$
$\frac{\pi c R^3}{3} = 1 \Rightarrow c = \frac{3}{R^3 \pi}$

2) Пусть $R = 2 \Rightarrow c = \frac{3}{8\pi}$
$P(\varepsilon \in K) = \int\int\limits_K f(x,y)dxdy = \int\limits_0^{2\pi}\int\limits_0^{1} (\frac{3}{8\pi}(2-\rho)\rho) d\rho d\phi = 2\pi\frac{3}{8\pi}\int\limits_0^{1}(2\rho - \rho^2)d\rho = \frac{3}{4}(1-\frac{1}{3}) = \frac{1}{2}$

Задача 10. Ковариационная матрица и коэф. корреляции

Плотность распределения случайного вектора $\zeta = (\xi, \eta)$ имеет вид $f(x,y) = \begin{cases} \dfrac{3(x^2+y^2)}{2}, 0 \le x \le 1, 0 \le y \le 1 \\ 0, \text{в остальных случаях} \end{cases}$.

Найти ковариационную матрицу вектора ζ и коэффициент корреляции случайных величин ξ и η.

Решение

Находим плотности распределения отдельно для x и y:

$f(x) = \int\limits_0^1 \dfrac{3(x^2+y^2)}{2} dy = \dfrac{3}{2}x^2 + \dfrac{1}{2}$

$f(x) = \int\limits_0^1 \dfrac{3(x^2+y^2)}{2} dx = \dfrac{3}{2}y^2 + \dfrac{1}{2}$

Проверяем x и y на независимость: $f(x) \cdot f(y) = \left(\dfrac{3}{2}x^2 + \dfrac{1}{2}\right)\left(\dfrac{3}{2}y^2 + \dfrac{1}{2}\right) = \dfrac{9}{4}x^2y^2 + \dfrac{3}{4}x^2 + \dfrac{3}{4}y^2 + \dfrac{1}{4}$

$f(x) \cdot f(y) \neq f(x,y) \Rightarrow$ зависимы.

$E\xi = \int\limits_0^1 x\left(\dfrac{3}{2}x^2 + \dfrac{1}{2}\right)dx = \int\limits_0^1 \left(\dfrac{3}{2}x^3 + \dfrac{1}{2}x\right)dx = \left(\dfrac{3x^4}{2 \cdot 4} + \dfrac{x^2}{2 \cdot 2}\right)\Bigg|_0^1 = \dfrac{3}{8} + \dfrac{1}{4} = \dfrac{5}{8}$

$E\eta = \int\limits_0^1 y\left(\dfrac{3}{2}y^2 + \dfrac{1}{2}\right)dy = \dfrac{5}{8}$

$D\xi = \int\limits_0^1 x^2\left(\dfrac{3}{2}x^2 + \dfrac{1}{2}\right)dx - \dfrac{25}{64} = \int\limits_0^1 \left(\dfrac{3}{2}x^4 + \dfrac{1}{2}x^2\right)dx - \dfrac{25}{64} = \left(\dfrac{3x^5}{2 \cdot 5} + \dfrac{x^3}{2 \cdot 3}\right)\Bigg|_0^1 - \dfrac{25}{64} =$

$= \dfrac{3}{10} + \dfrac{1}{6} - \dfrac{25}{64} = \dfrac{73}{960}$

$D\eta = \int\limits_0^1 y^2\left(\dfrac{3}{2}y^2 + \dfrac{1}{2}\right)yx - \dfrac{25}{64} = \dfrac{73}{960}$

$cov(\xi, \eta) = E(\xi\eta) - E\xi E\eta$

Т.к. величины зависимы:

$E(\xi\eta) = \int\limits_0^1 \int\limits_0^1 x \cdot y \cdot \dfrac{3(x^2+y^2)}{2} dxdy = \dfrac{3}{8}$

$\Rightarrow cov(\xi, \eta) = \dfrac{3}{8} - \dfrac{5}{8} \cdot \dfrac{5}{8} = -\dfrac{1}{64}$

$K = \begin{pmatrix} \dfrac{73}{960} & -\dfrac{1}{64} \\ -\dfrac{1}{64} & \dfrac{73}{960} \end{pmatrix}$

$\rho_{\xi\eta} = \dfrac{cov(\xi,\eta)}{\sqrt{D\xi D\eta}} = \dfrac{-\dfrac{1}{64}}{\dfrac{73}{960}} = -\dfrac{15}{73}$

Задача 11. Мат. ожидание и ков. матрица для вектора. Некоррелированность и независимость

Распределение дискретного случайного вектора $\xi = (\xi_1, \xi_2)^T$ задано таблицей. Найти математическое ожидание и ковариоционную матрицу вектора ξ. Являются ли случайные величины ξ_1 и ξ_2 некоррелированными? Являются ли СВ ξ_1 и ξ_2 независимыми?

	$\xi_2 =$ -2	$\xi_2 = 2$
$\xi_1 = -1$	2/15	1/5
$\xi_1 = 0$	1/15	3/10
$\xi_1 = 1$	1/5	1/10

Решение

ξ_1	-1	0	1
P	5/15	11/30	3/10

$E\xi_1 = -\frac{5}{15} + \frac{3}{10} = -\frac{1}{30}$

ξ_2	-2	2
P	6/15	6/10

$E\xi_2 = -\frac{12}{15} + \frac{12}{10} = \frac{12}{30} = \frac{2}{5}$

$E\xi = (-\frac{1}{30}; \frac{2}{5})$

$E(\xi_1\xi_2) = \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} x_i \cdot y_j \cdot p_{ij} = (-1)(-2)\frac{2}{15} + (-1)\frac{2}{5} + 0 + 0 + (-2)\frac{1}{5} + \frac{2}{10} = \cdots = -\frac{10}{30} = -\frac{1}{3}$

$D\xi_1 = E\xi_1^2 - (E\xi_1)^2 = \frac{19}{30} - (\frac{-1}{30})^2 = \frac{569}{900}$
$D\xi_2 = E\xi_2^2 - (E\xi_2)^2 = 4 - (\frac{2}{5})^2 = \frac{96}{25}$

$cov(\xi_1, \xi_2) = E(\xi_1\xi_2) - E\xi_1 E\xi_2 = -\frac{1}{3} + \frac{1}{30} \cdot \frac{2}{5} = -\frac{1}{3} + \frac{1}{75} = -\frac{24}{75}$ $\Rightarrow$коррелированы отрицательно

$$K = \begin{pmatrix} \dfrac{569}{900} & -\dfrac{24}{75} \\ -\dfrac{24}{75} & \dfrac{96}{25} \end{pmatrix}$$

$\rho_{\xi_1\xi_2} = \frac{-\frac{24}{75}}{\sqrt{\frac{569}{900} \cdot \frac{96}{25}}} = -\frac{24}{75\sqrt{\frac{4552}{1875}}} = 0,132$ (это считать необязательно)

Компоненты дискретного вектора независимы $\iff p_{ij} = p_{i\cdot} p_{\cdot j} \forall i = \overline{1,3}; j = \overline{1,2}$

Рассмотрим $p_{11} = \frac{2}{15}$, при этом $p_{1\cdot} = \frac{5}{15}$ $p_{\cdot 1} = \frac{6}{15}$
Т.к. $\frac{2}{15} \neq \frac{30}{225} \Rightarrow \xi_1$ и ξ_2 зависимы

Задача 12. ММП
Выборка $X_1, \cdots, X_n$ соответствует распределению Парето с функцией распределения
$F(x) = \begin{cases} 0, x < 1 \\ 1 - x^{-\alpha}, x \geq 1 \end{cases}$ Оценить параметр α методом максимального правдоподобия.

Решение

Зная функцию распредления, можем найти функцию плотности: $f(x) = \begin{cases} 0, x < 1 \\ \alpha \cdot x^{-\alpha-1}, x \geq 1 \end{cases}$

Так как распределение непрерывное, то $L(\alpha, x_1, \ldots, x_n) = \prod\limits_{i=1}^{n} f(\alpha, x_1, \ldots, x_n)$

$L(\alpha, x_1, \ldots, x_n) = \prod\limits_{i=1}^{n} \alpha \cdot x_i^{-\alpha-1} = \alpha^n \cdot \prod\limits_{i=1}^{n} x_i^{-\alpha-1}$
Логарифмируем функцию правдоподобия:
$\ln L(\alpha, x_1, \ldots, x_n) = \ln \alpha^n + \ln \prod\limits_{i=1}^{n} x_i^{-\alpha-1} = n \cdot \ln \alpha + \sum\limits_{i=1}^{n} \ln x_i^{-\alpha-1} = n \cdot \ln \alpha + \sum\limits_{i=1}^{n} (-\alpha - 1) \ln x_i =$
$= n \cdot \ln \alpha + (-\alpha - 1) \sum\limits_{i=1}^{n} \ln x_i$
Теперь продиффиринцируем функцию правдоподобия:
$\frac{d \ln L(\alpha, x_1, \ldots, x_n)}{d\alpha} = n \cdot \frac{1}{\alpha} - \sum\limits_{i=1}^{n} \ln x_i$
Приравнияем функцию правдоподобия к нулю и выразим α:
$\frac{n}{\hat{\alpha}} - \sum\limits_{i=1}^{n} \ln x_i = 0$
$\frac{n}{\hat{\alpha}} = \sum\limits_{i=1}^{n} \ln x_i \Rightarrow \hat{\alpha} = \frac{n}{\sum\limits_{i=1}^{n} \ln x_i}$

Задача 13. Метод моментов
Выборка $X_1, \cdots, X_n$ соответствует гауссовскому распределению $N(\theta_1, \theta_2^2)$. Построить оценку вектора (θ_1, θ_2^2) по методу моментов.

Решение
Немного из теории (да, она была выше, ну и что...):
k-ый момент: $\mu_k = E\xi^k$, то есть $\mu_k(\theta) = \int\limits_{-\infty}^{+\infty} x^k \cdot f(x, \theta) dx$

Приравниваем моменты выборочные к моментам теоретическим:
$\begin{cases} \overline{X} = \hat{\theta_1} \text{- 1ый выборочный момент равен второму (мат. ожиданию)} \\ \frac{1}{n} \sum\limits_{i=1}^{n} x_i^2 = \hat{\theta_2^2} + (\hat{\theta_1})^2 \text{- дисперсия + мат. ожидание}^2 \end{cases}$

$\begin{cases} \hat{\theta_1} = \overline{X} \text{- 1ый выборочный момент равен второму (мат. ожиданию)} \\ \hat{\theta_2^2} = \frac{1}{n} \sum\limits_{i=1}^{n} x_i^2 + (\hat{\theta_1})^2 \end{cases}$

Задача 14. Метод моментов
Выборка $X_1, \cdots, X_n$ соответствует гауссовскому распределению $R(0, \theta)$. Построить оценку по методу моментов для неизвестного параметра θ.

Решение
Одна неизвестная $\Rightarrow$ одно уравнене
1-ый выборочный момент $= \overline{X}$; мат. ожидание данного распределения $= \frac{\theta}{2}$
$\overline{X} = \sum\limits_{i=1}^{n} x_i = \frac{\theta}{2} \Rightarrow \hat{\theta} = 2\overline{X}$

Задача 15. ММП. Несмещенность и состоятельность
Выборка $X_1, \cdots, X_n$ соответствует распределению Релея, плотность которого имеет вид $f(x) = \frac{2x}{\theta} exp(\frac{-x^2}{\theta})$ при x>0. Найдите оценку максимального правдоподобия параметра θ. Докажите несмещённость и состоятельность этой оценки.

Решение

$L(x_1, \ldots, x_n, \theta) = \prod\limits_{i=1}^{n} f(x_i)$

$\ln L(x_1, \ldots, x_n, \theta) = \ln \prod\limits_{i=1}^{n} f(x_i) = \ln \left(\prod\limits_{i=1}^{n} \frac{2x_i}{\theta} e^{(\frac{-x_i^2}{\theta})} \right) = \ln \left(\frac{2^n}{\theta^n} \prod\limits_{i=1}^{n} x_i e^{(\frac{-x_i^2}{\theta})} \right) =$

$= n \cdot \ln 2 - n \cdot \ln \theta + \sum\limits_{i=1}^{n} \ln \left(x_i e^{(\frac{-x_i^2}{\theta})} \right) = n \cdot \ln 2 - n \cdot \ln \theta + \sum\limits_{i=1}^{n} \ln x_i - \sum\limits_{i=1}^{n} \frac{x_i^2}{\theta} =$

$= n \cdot \ln 2 - n \cdot \ln \theta + \sum\limits_{i=1}^{n} \ln x_i - \frac{1}{\theta} \cdot \sum\limits_{i=1}^{n} x_i^2$

$\frac{d \ln L(\alpha, x_1, \ldots, x_n)}{d\alpha} = 0 - n \cdot \frac{1}{\theta} + 0 - (-\frac{1}{\theta^2}) \cdot \sum\limits_{i=1}^{n} x_i^2$

$- n \cdot \frac{1}{\hat{\theta}} + \frac{1}{\hat{\theta}^2} \cdot \sum\limits_{i=1}^{n} x_i^2 = 0$

$- n \cdot \hat{\theta} + \sum\limits_{i=1}^{n} x_i^2 = 0$

$n \cdot \hat{\theta} = \sum\limits_{i=1}^{n} x_i^2$

$\hat{\theta} = \dfrac{\sum\limits_{i=1}^{n} x_i^2}{n}$

Несмещенность: $E\hat{\theta} = \theta$

$E\left(\dfrac{\sum\limits_{i=1}^{n} x_i^2}{n} \right) = \dfrac{\sum\limits_{i=1}^{n} E(x_i^2)}{n}$

$E(x_i^2) = \int\limits_{-\infty}^{+\infty} x^2 \cdot f(x) dx = \int\limits_{0}^{+\infty} x^2 \cdot \dfrac{2x}{\theta} \cdot e^{\frac{-x^2}{\theta}} dx = \frac{2}{\theta} \int\limits_{0}^{+\infty} x^3 \cdot e^{\frac{-x^2}{\theta}} dx = \frac{2}{\theta} \cdot \frac{1}{2} \int\limits_{0}^{+\infty} x^2 \cdot e^{\frac{-x^2}{\theta}} dx^2 =$

$= \frac{1}{\theta} \int\limits_{0}^{+\infty} y \cdot e^{\frac{-y}{\theta}} dy = - \int\limits_{0}^{+\infty} y \; de^{\frac{-y}{\theta}} = -e^{\frac{-y}{\theta}} \cdot y|_0^{+\infty} + \int e^{\frac{-y}{\theta}} dy = 0 + \int e^{\frac{-y}{\theta}} dy =$

$= -\theta \cdot (e^{\frac{-y}{\theta}}|_0^{+\infty}) = -\theta \cdot 0 + \theta \cdot 1 = \theta \Rightarrow$ несмещенность доказана

Состоятельность: $P(|\hat{\theta} - \theta| > \varepsilon) = 0$ при $n \to \infty$

$P(|\hat{\theta} - E\hat{\theta}| > \varepsilon) \leq \frac{D\hat{\theta}}{\varepsilon^2}$, то есть нужно доказать, что $D\hat{\theta} \to 0$ при $n \to \infty$

$D\hat{\theta} = E\hat{\theta}^2 - (E\hat{\theta})^2$
$E\hat{\theta} = Ex_i^2 \Rightarrow E\hat{\theta}^2 = Ex_i^4$

$E(x_i^4) = \int\limits_{0}^{+\infty} x^4 \cdot \dfrac{2x}{\theta} \cdot e^{\frac{-x^2}{\theta}} dx = \frac{2}{\theta} \int\limits_{0}^{+\infty} x^5 \cdot e^{\frac{-x^2}{\theta}} dx = \frac{1}{\theta} \int\limits_{0}^{+\infty} x^4 \cdot e^{\frac{-x^2}{\theta}} dx^2 =$

$= - \int\limits_{0}^{+\infty} x^4 \; de^{\frac{-x^2}{\theta}} = -x^4 \cdot e^{\frac{-x^2}{\theta}}|_0^{+\infty} + \int\limits_{0}^{+\infty} e^{\frac{-x^2}{\theta}} dx^4 = 0 + \int\limits_{0}^{+\infty} e^{\frac{-y}{\theta}} dy^2 = 2 \int\limits_{0}^{+\infty} e^{\frac{-y}{\theta}} y \; dy = -2\theta \int\limits_{0}^{+\infty} y \; de^{\frac{-y}{\theta}} =$

$= 2\theta(e^{\frac{-y}{\theta}} \cdot y|_0^{+\infty} + \int\limits_{0}^{+\infty} e^{\frac{-y}{\theta}} dy) = 2\theta(0 + -\theta) \cdot (e^{\frac{-y}{\theta}}|_0^{+\infty}) = -2\theta^2 \cdot (-1) = 2\theta^2$

$D\theta = \frac{1}{n}(2\theta^2 - \theta^2) = \frac{\theta^2}{n} \to 0$ при $n \to \infty \Rightarrow$ оценка несмещенная

Задача 16. Несмещенность и состоятельность
Выборка $X_1, ..., X_n$ соответствует распределению $R(0, \theta)$. Пусть $\theta = 2\overline{X}$. Является ли оценка несмещенной и состоятельной оценкой неизвестного параметра θ?

Решение
$X_1, ..., X_n \sim R(0, \theta)$
$\hat{\theta} = 2\overline{X}$

$$E\hat{\theta} = E(2\overline{X}) = 2E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{2}{n}\sum_{i=1}^{n} EX_i = \frac{2}{n}n\frac{\theta}{2} = \theta \Rightarrow \text{оценка несмещенная.}$$

По ЗБЧ: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{P}} EX_1 = \frac{\theta}{2} \Rightarrow 2\overline{X} \xrightarrow{P} \theta \Rightarrow \hat{\theta} \xrightarrow{P} \theta \Rightarrow$ оценка состоятельная.

Задача 17. Доверительный интервал для мат. ожидания при неизв. дисперсии
Имееются данные о потреблении мяса и мясопродукта на душу населения (в кг) по нескольким субъектам Северо-Кавказского федерального округа: 40; 54; 62; 66; 51; 76. Постройте доверительный интервал уровня надежности 0,95 для среднего значения потребления мяса и мясопродуктов на душу населения. Можно ли считать на уровне значимости 0,05, что потребление мяса и мясопродуктов на душу населения в Северо-Кавказском федеральном округе в среднем составляет 60 кг?

Решение
Найдем некоторые полезности:
$\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i = \frac{349}{6} = 58,17$
$s_X^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{6}(331,24 + 17,64 + 14,44 + 60,84 + 51,84 + 316,84) = \frac{792,84}{6} = 132,14$

Так как уровень надежности доверительного интервала равен 0,95 (по условию), то:
$1 - \alpha = 0,95 \Rightarrow \alpha = 0,05 \Rightarrow \frac{\alpha}{2} = 0,025 \Rightarrow 1 - \frac{\alpha}{2} = 0,975$

$G = \frac{(\bar{X} - m_X)\cdot\sqrt{n-1}}{s_X} \sim t(n-1)$

$P\big(t_{\alpha/2;n-1} < \frac{(\bar{X} - m_X)\cdot\sqrt{n-1}}{s_X} < t_{1-\alpha/2;n-1}\big) = 1 - \alpha \Rightarrow$
$\Rightarrow P\big(\bar{X} - \frac{s_X}{\sqrt{n-1}}\cdot t_{1-\alpha/2;n-1} < m_X < \bar{X} + \frac{s_X}{\sqrt{n-1}}\cdot t_{1-\alpha/2;n-1}\big) = 1 - \alpha$

Найдем значение квантиля: $t_{1-\alpha/2;n-1} = t_{0,975;5} = 2,571$

Доверительный интервал для математического ожидания:
$\big(\bar{X} - \frac{s_X}{\sqrt{n-1}}\cdot t_{1-\alpha/2;n-1}; \bar{X} + \frac{s_X}{\sqrt{n-1}}\cdot t_{1-\alpha/2;n-1}\big) =$
$= \big(58,17 - \frac{11,5}{\sqrt{5}}\cdot 2,571; 58,17 + \frac{11,5}{\sqrt{5}}\cdot 2,571\big) =$
$= \big(58,17 - \frac{29,57}{2,24}; 58,17 + \frac{29,57}{2,24}\big) = \big(44,97; 71,37\big)$

Задача 18. Доверительный интервал для разности мат. ожиданий
В случайной выборке из 316 семей Западного Мидленда выборочное среднее еженедельных доходов составило 25.85$, а среднеквадратическое отклонение еженедельных доходов - 12.7$. Для 351-й семьи для Йоркшира выборочное среднее составило 23.14$, а среднеквадратическое отклонение - 13.4$. Постройте доверительный интервал уровня надежности 0.95 для разности средних еженедельных доходов семей Западного Мидленда и Йоркшира. Можно ли считать (на уровне доверия 0.95), что еженедельные доходы семей Западного Мидленда и Йоркшира в среднем одинаковы?

Решение
$n_1 = 316, n_2 = 351$
$\overline{X} = 25.85, \overline{Y} = 23.14$
$\sigma_1 = 12.7, \sigma_2 = 13.4$
$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \Rightarrow 1 - \alpha/2 = 0.975$
$\theta = m_1 - m_2 = \overline{X} - \overline{Y}$ - оцениваемый параметр
$E(\overline{X} - \overline{Y}) = m_1 - m_2 \; D(\overline{X} - \overline{Y}) = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

Центральная статистика (известны мат. ожидания и дисперсии):
$$G = \frac{\overline{X} - \overline{Y} - (m_1 - m_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Доверительный интервал:
$$P\left(- z_{0.975} < \frac{\overline{X} - \overline{Y} - (m_1 - m_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} < z_{0.975} \right) = 0.95$$
$$P\left(\overline{X} - \overline{Y} - z_{0.975}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < m_1 - m_2 < \overline{X} - \overline{Y} + z_{0.975}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right) = 0.95$$
$$P\left(25.85 - 23.14 - 1.96\sqrt{\frac{(12.7)^2}{316} + \frac{(13.4)^2}{351}} < m_1 - m_2 < 2.71 + 1.96\sqrt{\frac{(12.7)^2}{316} + \frac{(13.4)^2}{351}} \right) = 0.95$$

$$P\Big(0.75 < m_1 - m_2 < 4.67 \Big) = 0.95$$
ДИ: $(0.75; 4.67)$

В среднем мат. ожидания не одинаковы, т.к. $0 \notin (0.75; 4.67)$.

Задача 19. Гипотеза для разности мат. ожиданий
Используя данные задачи №1 (ДИ), проверьте гипотезу о равенстве средних значений январских температур в г. Саратове и Алатырь. Уровень значимости считать равным 0.05.

Решение
$X_1, \cdots, X_n \sim N(m_1, \sigma^2)$
$Y_1, \cdots, Y_n \sim N(m_2, \sigma^2)$
σ^2 - неизвестен

1) $H_0 : m_1 = m_2 \Rightarrow m_1 - m_2 = 0$
$H_1 : m_1 - m_2 \neq 0$
2) $\alpha = 0,05$
3) $z = \dfrac{\overline{X} - \overline{Y} - (m_1 - m_2)}{s_{xy} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$
4) $z|_{H_0} \sim t(n_1 + n_2 - 2)$
5) Доверительная область лежит в границах от $t_{\alpha/2;n_1+n_2-2} = -t_{1-\alpha/2;n_1+n_2-2}$ до $t_{1-\alpha/2;n_1+n_2-2}$
6) $\cdots$

Задача 20. Гипотеза на значение вероятности
В первой 1000 новорожденных 2020 года оказалось 517 мальчиков и 483 девочки. Можно ли считать, опираясь на эти данные, что вероятность рождения мальчика больше, чем 0.5? Уровень значимости принять равным 0.05.

Решение
$X_1, \cdots, X_n \sim Bi(1, p)$
1) $H_0 : p = 0,5; H_1 : p > 0,5$
2) $\alpha = 0,05$
3) $z = \sum\limits_{i=1}^{n} x_i$ и $z|_{H_0} \sim Bi(n, p)$
Если же испытаний много ($n \to \infty$), то центрируем и нормируем:
$$\tilde{z} = \frac{\sum\limits_{i=1}^{n} x_i - np}{\sqrt{npq}} = \frac{\sum\limits_{i=1}^{n} x_i - 0,5p}{\sqrt{n \cdot 0,5 \cdot 0,5}}$$
4) $\tilde{z}|_{H_0} \sim N(0;1)$
5) Доверительная область от $-\infty$ до $z_{1-\alpha} = z_{0,95} = 1,65$
6) $\tilde{z} = \dfrac{517 - 0,5 \cdot 1000}{\sqrt{1000 \cdot 0,5 \cdot 0,5}} = 1,06$
7) $\tilde{z} = 1,06$ попадает в доверительную область $\Rightarrow$ гипотеза H_0 принимается

Задача 21. Среднее (мат. ожмдание) сократилось

Среднее время сборки изделия составляет 90 минут. Инженер предложил новый метод сборки изделий. В результате сборки 10 изделий новым способом время сборки составило 79,74, 112, 95,83. 96, 77, 84,70,90 минут. Можно ли считать, что время сборки в среднем сократилось. Уровень значимости считать равным 0.05

Решение

$\overline{X} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i = \frac{1}{10} \cdot 860 = 86$

$s^2 = \sum\limits_{i=1}^{n} (x_i - \overline{X})^2 = 141,6$

$\sqrt{s^2} = s = 11,9$

1) H_0: $m = 90$; H_1: $m < 90$ (альтернатива - время сократилось)

2) $\alpha = 0,05$

3) $z = \dfrac{(\overline{X} - 90)\sqrt{n-1}}{s}$

4) $z|_{H_0} \sim t(n-1)$

5) Доверительная область от $t_{\alpha;n-1}$ до $+\infty$

$t_{\alpha;n-1} = t_{0,05;9} = -1,83$

6) $z = \dfrac{(\overline{X} - 90)\sqrt{n-1}}{s} = \dfrac{(86-90)\sqrt{9}}{11,9} = -1,008$

7) На уровне значимости 0,05 гипотеза H_0 принимается.

Задача 22. Гипотеза на коррелированность

Изучается зависимость между показателями X (оценка студента за контрольную работу по теории вероятностей) и Y (оценка студента за экзамен по программированию). Используя данные о показателях X и Y для группы 182, в которой обучается 30 человек, Кристиан Бенуа вычислил выборочный коэффициент корреляции и получил значение 0.45. Можно ли, опираясь на эти данные, считать, что показатели X и Y являются положительно коррелированными? Уровень значимости принять равным 0.05. Опишите процедуру проверки соответствующей гипотезы.

Решение

1) Сформулируем основную и альтернативную гитпотезы:

$H_0 : \rho_{xy} = 0$ - некоррелированы

$H_1 : \rho_{xy} > 0$ - коррелированы положительно

2) Выбираем уровень значимости α

3. $T((X_1, Y_1), \dots, (X_n, Y_n)) = \dfrac{\sqrt{n-2}\hat{\rho}_{xy}}{\sqrt{1-\hat{\rho}_{xy}^2}}$

4. $T((X_1, Y_1), \dots, (X_n, Y_n))|_{H_0} \sim t(n-2)$

5. Доверительная область в границах $(-\infty; t_{1-\alpha,n-2}]$, критическая область в границах $(t_{1-\alpha,n-2}; +\infty)$

$t_{1-\alpha,n-2} = t_{0,95,28} = 1,7 \Rightarrow$ доверительная область: $(-\infty; 1,7]$; критическая область: $(1,7; +\infty)$

6) $T((X_1, Y_1), \dots, (X_n, Y_n)) = \dfrac{\sqrt{n-2}\hat{\rho}_{xy}}{\sqrt{1-\hat{\rho}_{xy}^2}} = \dfrac{\sqrt{28} \cdot 0,45}{\sqrt{1-0,45^2}} = 2,67$

7) $T((X_1, Y_1), \dots, (X_n, Y_n)) = 2,67$ попадает в критическую область $\Rightarrow$ на уровне значимости 0,05 гипотеза H_0 отвергается в пользу H_1 (величины положительны коррелированы) $\Rightarrow$ имеется зависимость между X_i и Y_i.

Задача 23. Независимость двух событий/СВ
Среди 70 опрошенных пациентов мужского пола 50 удовлетворены работой врачей поликлиники. Среди 140 опрошенных женщин работой врачей удовлетворены 65. Можно ли считать, что удовлетворенность врачей зависит от пола пациента?

Решение
Пусть A - мужчина, $\bar{A}$ - женщины, B - удовлетворены, $\bar{B}$ - не удовлетворены

A \ B	B	$\bar{B}$	Sum
A	50	20	70
$\bar{A}$	65	75	140
Sum	115	95	210

1) Основная гипотеза H_0: $P(A = A_i, B = B_i) = P(A = A_i) \cdot P(B = B_i)$, то есть A и B - независимые.
Альтернативная гипотеза H_1: $P(A = A_i, B = B_i) \neq P(A = A_i) \cdot P(B = B_i)$, то есть A и B - зависимые
2) Выбираем уровень значимости α
3) $T(x) = n\big(\sum\limits_{i=1}^{s}\sum\limits_{j=1}^{k}\dfrac{(n_{ij}^2)}{n_{i\cdot}n_{\cdot j}} - 1\big)$
4) $T(x)|_{H_0} \sim \chi^2((k-1)(s-1)) = \chi^2(1)$
5) $\chi^2_{1-\alpha;(k-1)(s-1)} = \chi^2_{0,95;1} = 3,84 \Rightarrow$ Доверительная область лежит в границе $-\infty$ от до 3,84, а критическая: от 3,84 до ∞
6) $T(x) = 210\big(\dfrac{50^2}{70 \cdot 115} + \frac{20^2}{70 \cdot 95} + \dfrac{65^2}{140 \cdot 115} + \dfrac{75^2}{140 \cdot 95} - 1\big) = 210\big(\frac{923}{874} - 1\big) = \dfrac{210 \cdot 49}{874} \approx 11,773$
7) $T(x) = 11,773$ попадает в критическую область $\Rightarrow$ гипотеза H_0 отвергается в пользу альтернативной гипотезы $H_1 \Rightarrow$ признаки зависимы