Общие правила: N – размер объединенной выборки, m – размер выборки X(первой), n – размер выборки Y(второй), s_x – несмещенная (исправленная) дисперсия, $\widehat{d_x}$ – смещенная (обычная) выборочная дисперсия. R_i – ранг элемента (ранг – это номер элемента (нумерация с 1) в объеденной выборке. При этом если есть несколько одинаковых элементов (связка), то у них один ранг, равный среднему арифметическому их номеров). l – число связок, t_k – длина k-ой связки. Гипотезы H_1, H_2, H_3 – альтернативные, выбирается одна.
Ранговые критерии сводятся к нормальному (X – EX) / $\sqrt{DX}$

Меры Гутмана: $p_1 = 1 - \frac{max_B}{n}, p_2 = 1 - \frac{\sum max_B}{n}, \lambda_B = \frac{p_1 - p_2}{p_1}$, (сумма по local max)

Проверка гипотез о сдвиге («изменении мат. ожидания»)

$$H_0: \theta = 0$$

$H_1: \theta < 0$ (уменьшилось), $H_2: \theta > 0$ (увеличилось), $H_3: \theta \neq 0$ (изменилось)

(Дополнительно: $F_y = F_x(t - \theta)$, $\theta = EY - EX$)

$\overline{G}$ $< 0: (-\infty; Z_\alpha)$
$> 0: (Z_{1-\alpha}; +\infty)$
$\neq 0: \left(-\infty; Z_{\frac{\alpha}{2}}\right) \cup \left(Z_{1-\frac{\alpha}{2}}; +\infty\right)$

Критерий Стьюдента (нормальное распределение)

$$T = \frac{\overline{Y} - \overline{X}}{S_{xy}\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n + m - 2)$$

$$S_{xy} = \sqrt{\frac{(m-1) \cdot s_x^2 + (n-1) \cdot s_y^2}{n + m - 2}}$$

Условие: равенство теоретических дисперсий

Критерий Вилкоксона (неизвестное распределение)

$$W_{m,n} = \sum_{j=1}^{n} R_j \text{ (ранги второй выборки)}$$

$$E(W_{m,n}) = \frac{n}{2}(N + 1)$$

$$D(W_{m,n}) = \frac{m \cdot n}{12}(N + 1) \text{ (без связок)}$$

$$D_{\text{св}}(W_{m,n}) = D(W_{m,n}) - \frac{m \cdot n \sum_{k=1}^{l} \left(t_k(t_k^2 - 1)\right)}{12N(N-1)}$$

Проверка сжатия («изменения дисперсии»)

$$H_0: \triangle = \frac{\sigma_y}{\sigma_x} = 1$$

Критерий Фишера (нормальное распределение)

$$T = \frac{S_Y^2}{S_X^2} \sim F(n - 1, m - 1) \text{ (если } S_Y > S_X)$$

(если нет, то числитель из знаменатель меняются местами, n – 1 и m – 1 тоже меняются)

$\overline{G}$: (ОСОБЫЕ) $(f_{1-\alpha}; +\infty)$ (если $H_1 > 1 \mid < 1$) или $(f_{1-\frac{\alpha}{2}}; +\infty)$ (если $H_1 \neq 1$)

Критерий Ансари-Брэдли (неизвестное распределение)

$$A_{m,n} = \sum_{i=1}^{m} q_i = \begin{cases} R_i, \text{если } R_i \leq \frac{N+1}{2} \\ N + 1 - R_i, \text{иначе} \end{cases}$$

$$E(A_{m,n}) = \begin{cases} \frac{m \cdot (N+2)}{4}, & N = 2k \\ \frac{m \cdot (N+1)^2}{4N}, & N = 2k + 1 \end{cases}$$

$$D(A_{m,n}) = \begin{cases} \frac{m \cdot n \cdot (N+2) \cdot (N-2)}{48(N-1)}, & N = 2k \\ \frac{m \cdot n \cdot (N^2+3)(N+1)}{48N^2}, & N = 2k + 1 \end{cases}$$

$$D(A_{m,n}) = \begin{cases} \frac{mn\left(16\sum_{j=1}^{k} t_j R_j^2 - N(N+2)^2\right)}{16N(N-1)}, & N = 2k \\ \frac{mn\left(16\sum_{j=1}^{k} t_j R_j^2 - (N+1)^4\right)}{16N^2(N-1)}, & N = 2k + 1 \end{cases}$$

Критерий Колмогорова-Смиронова (проверка соответствия распределений)

$$D_{m,n} = sup\left|\widehat{F_X}(t) - \hat{F}_Y(t)\right|$$

$$D^*_{m,n} = \sqrt{\frac{nm}{n+m}} D_{m,n} \sim K \text{ (Колмогорова)}$$

$$H_0: \forall t\ F_X(t) = F_Y(t) \quad H_1: \exists t\ F_X(t) \neq F_Y(t)$$

Однофакторный дисперсионный анализ. (Проверяет несколько выборок). Модель: $X_{ij} = \theta + \tau_j + \varepsilon_{ij}$. j – номер выборки, i – номер элемента, τ_j – отколонение от общего среднего, ε_{ij} – независимые ненаблюдаемые погрешности, k – число выборок. $H_0: \tau_1 = \ldots = \tau_k = 0$ $H_1: \exists j\ \tau_j \neq 0$. $\overline{X}_{.j}$ – среднее по выборке j, $\overline{X}_N$ – среднее по общей выборке.

Классический F- критерий (норм. распред. с один. дисп.) $T(Z_N) = \frac{\frac{1}{k-1}\sum_{j=1}^{k} n_j\left(\overline{X}_{\bullet j} - \overline{X}_N\right)^2}{\underbrace{\frac{1}{N-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(X_{ij} - \overline{X}_{\bullet j}\right)^2}_{\tilde{S}_N^2}} \sim F(k-1, N-k)$

Критическая область: $(f_{1-\alpha}(k-1; N-k); +\infty)$ $\tilde{S}_N^2$

Доверительный интервал $\theta_j = \theta + \tau_j$: $\theta_j \in \left[\overline{X}_{\cdot j} \pm \frac{1}{\sqrt{n_j}}\sqrt{\tilde{S}_N^2}\ t_{1-\frac{\alpha}{2}, N-k}\right]$ // Доверительный интервал контраста: $\gamma = \sum_{j=1}^{k} c_j\theta_j$, где $\sum_{j=1}^{k} c_j = 0$ $\gamma \in \left[\sum_{j=1}^{k} c_j \overline{X}_{\cdot j} \pm t_{1-\frac{\alpha}{2}, N-k}\sqrt{\tilde{S}_N^2 \sum_{j=1}^{k}\frac{c_j^2}{n_j}}\right]$

Критерий Краскела-Уолиса (любое распред.) H_0: $\theta_1 = \dots = \theta_k = \theta$ H_1: $\exists i \neq j\ \theta_i \neq \theta_j$ Крит. Обл правая

$$H = \frac{12}{N(N+1)}\sum_{j=1}^{k}\frac{1}{n_j}\left(\sum_{i=1}^{n_j} R_{ij}\right)^2 - 3(N+1) \sim \chi^2(k-1)$$

$$H' = \frac{H}{1 - \frac{1}{(N^3-N)}\sum_{i=1}^{l}(t_i^3 - t_i)} \text{ (связки)}$$

Критерий Джонкхиера (любое распред.) H_a: $\theta_1 \leq \dots \leq \theta_k$ $\phi(y,z) = 1\ (y<z), 0.5(y=z), 0(y>z)$. $\mathbf{J} = \sum_{1\leq l<m\leq k} \mathbf{U}_{l,m}$

$U_{lm} = \sum_{i=1}^{n_l}\sum_{j=1}^{n_m}\phi(X_{il}, X_{jm})$ $E(J) = \frac{1}{4}\left(N^2 - \sum_{j=1}^{k} n_j^2\right)$ $D(J) = \frac{1}{72}\left(N^2(2N+3) - \sum_{j=1}^{k} n_j^2(2n_j+3)\right)$

Крит область правая

Проверка независимости номинальных признаков

Критическая область правая m, k – размеры ↓

Теорема Фишера-Пирсона H_0: $\forall i, j\ p_{ij} = p_{\cdot i}p_{\cdot j}$ H_a: $\exists$i, j $p_{ij} \neq p_{\cdot i}p_{\cdot j}$ $\widehat{\chi^2} = n\sum_{i=1}^{m}\sum_{j=1}^{k}\left(\frac{n_{ij}^2}{n_{i\cdot}n_{\cdot j}} - 1\right)$ Для m = k = 2 формула: $\widehat{\chi^2} = \frac{n(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}$. Коэффициент контингенции (двухстор. связь) $\Phi = \frac{ad-bc}{\sqrt{(a+b)(d+c)(a+c)(b+d)}}$, ассоциации Юла (A => B, если Q = 1) $Q = \frac{ad-bc}{ad+bc}$, взаим. связи Пирсона: $P = \sqrt{\frac{\widehat{\chi^2}}{\widehat{\chi^2}+n}}$, Крамера $C = \sqrt{\frac{\widehat{\chi^2}}{n\cdot\min\{(m-1),(k-1)\}}}$ $\widehat{\chi^2} \sim \chi^2\left((k-1)(m-1)\right)$

Критерии проверки связи двумерных выборок: (X_1, Y_1) ... (X_n, Y_n). Коэффициент ранговой корреляции Спирмена: $\widehat{\rho_s} = 1 - 6\cdot\frac{\sum_{i=1}^{n}(R_i - S_i)^2}{n^3 - n}$, если связки: $\widehat{\rho_S} = 1 - \frac{\sum_{i=1}^{n}(R_i - S_i)^2}{\frac{1}{6}(n^3-n) - (u_1+u_2)}$ $\sqrt{n-1}\cdot\widehat{\rho_S} \sim N(0,1)$, где $u_1 = \frac{1}{12}\sum_{k=1}^{q}\left(u_{1k}^3 - u_{1k}\right)$, u_{1k} – размер k-ой связки, q – количество связок, для u_2 выборки Y ситуация симметрична, согласованности Кендалла $\widehat{\tau_{XY}} = 1 - \frac{4K}{n(n-1)}$, где K – количество пар, где i < j и sign{$(X_i - X_j)(Y - Y_i)$} = -1. $\frac{3\sqrt{n}\widehat{\tau_{XY}}}{2} \sim N(0,1)$. H_a: $\rho > 0, \rho < 0, p \neq 0$

Кр. осн. на выбор. коэфф. корр. (норм. распред.) (H_a совпадает с пред.) $\widehat{\rho_{XY}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{S_X S_Y}$ $\frac{\sqrt{n-2}\ \widehat{\rho_{XY}}}{\sqrt{1-\widehat{\rho_{XY}^2}}} \sim t(n-2)$

ДИ: (обыч. норм) $\widehat{\rho_{XY}} \in \left[\widehat{\rho_{XY}}\left(1 - \frac{1-\hat{\rho}_{XY}^2}{2n}\right) \pm Z_{1-\frac{\alpha}{2}}\frac{\left(1-\hat{\rho}_{XY}^2\right)^2}{n}\right]$, преобр. Фишера:$[z_1, z_2] = \left[\frac{1}{2}\ln\left(\frac{1+\widehat{\rho_{XY}}}{1-\widehat{\rho_{XY}}}\right) - \frac{\widehat{\rho_{XY}}}{2(n-1)} \pm \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right]$, $\widehat{\rho_{XY}} \in [th\ z_1, th\ z_2]$, $th\ x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$; $\widehat{\rho_{xy}} = \frac{\frac{1}{n}\sum X_iY_i - \bar{X}\bar{Y}}{\sqrt{\left(\frac{1}{n}\sum X_i^2 - (\bar{X})^2\right)\left(\frac{1}{n}\sum Y_i^2 - (\bar{Y})^2\right)}}$

Частный коэффициент корреляции (норм распред). Убирает влияние других факторов.

$\rho_{12;3,\dots,l} = \frac{-R_{12}}{\sqrt{R_{11}R_{22}}}$, где l – число элементов, $R_{ij} = (-1)^{i+j}M_{ij}$, M_{ij} – det M без i j. Для l = 3, $\rho_{12;3} = \frac{\rho_{12} - \rho_{13}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}$. При использ. критер. выше $\frac{\sqrt{n-d-2}\ \widehat{\rho_{XY}}}{\sqrt{1-\widehat{\rho_{XY}^2}}} \sim t(n-d-2)$. d – число зафиксированных (после ;

Коэфф. конкордации Кендалла (независ. нескол.)

$\widehat{W}_n(m) = \frac{12}{m^2(n^3-n)}\sum_{i=1}^{n}\left(\sum_{j=1}^{m} R_{ij} - \frac{m(n+1)}{2}\right)^2$; $m(n-1)\widehat{W}_n(m) \sim \chi^2(n-1)$ m – число ранжировок, n – длина выборки. (крит. обл. правая) H_0: $F_\xi(x_1, \dots, x_m) = F_{\xi_1}(x_1)\cdot\ldots\cdot F_{\xi_m}(x_m)$, H_a ≠. крит. нужен для проверки независ. ранж. выборок.

Линейная регрессия. $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$; $\hat{\beta} = (X^TX)^{-1}X^TY$, $\hat{\beta}$ – вектор коэфф., X – матр. перв. столбец 1, другие – значения X_{ip}, p – кол-во регрессоров (X). $TSS = \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = ESS + RSS$; $RSS = \sum_{i=1}^{n}\left(Y_i - \widehat{Y_i}\right)^2$; $ESS = \sum_{i=1}^{n}\left(\widehat{Y_i} - \overline{Y}\right)^2$. $\sigma_\varepsilon^2 = \frac{RSS}{n-p-1}$. $\sigma_{\widehat{\beta_i}} = \sqrt{\sigma_\varepsilon^2 c_{ii}}$, где c_{ii} – элемент ii матрицы $(X^TX)^{-1}$. F-тест (знач. всей регресс) $\hat{F} = \frac{\frac{ESS}{p}}{\frac{RSS}{n-p-1}} \sim F(p, n-p-1)$ (H_0 ∀ β =0, H_a ∃ ≠ 0). Т-тест (знач. 1 факт.)(H_0: $\beta_i = 0$, H_a ≠) $\frac{\widehat{\beta_i} - \beta_i}{\sigma_{\widehat{\beta_i}}} \sim t(n-p-1)$ (можно пров. не только = 0, если просят). Коэфф. детерминации. (насколько хороша регресс.) $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$, $\widehat{R^2_{\text{кор.}}} = 1 - \left(1 - \widehat{R^2}\right)\frac{n-1}{n-p-1}$. Тест Чоу (пр. что две модели совпадают) H_0: $\beta_0^{(1)} = \beta_0^{(2)}; \dots; \beta_p^{(1)} = \beta_p^{(2)}$,$H_a$: ∃ ≠. Крит. Обл. правая. $\frac{\frac{RSS_{\text{общ}} - RSS_1 - RSS_2}{p+1}}{\frac{RSS_1 + RSS_2}{N-2p-2}} \sim F(p+1, N-2p-2)$. ДИ для E[Y|$X_0$]: $\left(\hat{Y}(X_0) \pm t_{1-\frac{\alpha}{2}; n-2}\sigma_\varepsilon\sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}}\right)$, для Y($X_0$) (+1 под корень), для σ_ε $\frac{RSS}{\sigma_\varepsilon} \sim \chi^2(n-p-1)$. Тест на огранич.$H_0$: $\widehat{\beta_{q+1}} = \dots = \widehat{\beta_p} = 0 (q<p)$, H_a: ∃ ≠ $\frac{\frac{RSS^2_{\text{урез}} - RSS^2_{\text{полн}}}{p-q}}{\frac{RSS^2_{\text{полн}}}{n-p-1}} \sim F(p-q, n-p-1)$. Крит. обл. правая. **Оценки МНК:** (Y = a + e; $\hat{a} = \overline{Y}$) (Y = bX_i + e; $\hat{b} = \frac{\sum X_iY_i}{\sum X_i}$) (Y = a + b$X_i$ + e)($\hat{a} = \overline{Y} - \hat{b}\overline{X}$ $\hat{b} = \sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) / \sum\left(X_i - \overline{X}\right)^2$). $\sigma_a^2 = \sigma_\varepsilon^2\ (1/n + \overline{X}^2 / \sum\left(X_i - \overline{X}\right)^2)$, $\sigma_b^2 = \sigma_\varepsilon^2 / \sum\left(X_i - \overline{X}\right)^2$, (Y = a + bX + e) HELP