

ОСНОВНИ ФОРМУЛИ ПО МЕДИЦИНСКА СТАТИСТИКА, 2025 Г.

ДЕСКРИПТИВНА СТАТИСТИКА ПРИ КОЛИЧЕСТВЕНИ ПРОМЕНЛИВИ

Средна аритметична	
При негрупирани данни $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$	При групирани данни $\bar{X}_w = \frac{\sum (x_i \times w_i)}{\sum w_i}$
Стандартно отклонение	
При негрупирани данни $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2}$	При групирани данни $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2 \times w_i}$

Правило на 3-те сигми

- Интервалът $\mu \pm 1\sigma$ около средната аритметична съдържа приблизително 68% от наблюденията;
- Интервалът $\mu \pm 2\sigma$ около средната аритметична съдържа приблизително 95% от наблюденията;
- Интервалът $\mu \pm 3\sigma$ около средната аритметична съдържа приблизително 99.7% от наблюденията.

Правило на Tukey

- Долни аутлайъри $< Q_1 - (1.5 \times IQR)$
- Горни аутлайъри $> Q_3 + (1.5 \times IQR)$

ИНФЕРЕНТНА СТАТИСТИКА

Средна грешка	
На средната аритметична $S_{\bar{x}} = \frac{S}{\sqrt{n}}$	На относителния дял $S_{\hat{p}} = \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}}$
Доверителен интервал	
За популационна средна μ При известно σ (или много голяма извадка): $CI = \bar{X} \pm Z_{\alpha/2} \times S_{\bar{x}}$ При неизвестно σ (и малка до средна извадка): $CI = \bar{X} \pm t_{\alpha/2, df} \times S_{\bar{x}}$	За популационен относителен дял P $CI = \hat{p} \pm Z_{\alpha/2} \times S_{\hat{p}}$
Необходим брой единици на наблюдение	
При количествени променливи $n = \frac{Z_{\alpha/2}^2 \times S^2}{d^2}$	При качествени променливи $n = \frac{Z_{\alpha/2}^2 \times \hat{p} \times (1 - \hat{p})}{d^2}$

t-разпределение

df	$t_{0.05}$ (90%)	$t_{0.025}$ (95%)	$t_{0.005}$ (99%)
1	6.314	12.706	63.657
2	2.920	4.303	9.925
3	2.353	3.182	5.841
4	2.132	2.776	4.604
5	2.015	2.571	4.032
6	1.943	2.447	3.707
7	1.895	2.365	3.499
8	1.860	2.306	3.355
9	1.833	2.262	3.250
10	1.812	2.228	3.169
11	1.796	2.201	3.106
12	1.782	2.179	3.055
13	1.771	2.160	3.012
14	1.761	2.145	2.977
15	1.753	2.131	2.947
16	1.746	2.120	2.921
17	1.740	2.110	2.898
18	1.734	2.101	2.878
19	1.729	2.093	2.861
20	1.725	2.086	2.845
21	1.721	2.080	2.831
22	1.717	2.074	2.819
23	1.714	2.069	2.807
24	1.711	2.064	2.797
25	1.708	2.060	2.787
26	1.706	2.056	2.779
27	1.703	2.052	2.771
28	1.701	2.048	2.763
29	1.699	2.045	2.756
30	1.697	2.042	2.750
40	1.684	2.021	2.704
50	1.676	2.009	2.678
100	1.660	1.984	2.626
∞ (съвпада със Z)	1.645	1.960	2.576

ПРОВЕРКА НА ХИПОТЕЗА

t-тест за две независими извадки с различни дисперсии	Z-тест за два относителни дяла
$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ $df \approx \min(n_1 - 1, n_2 - 1)$	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \times (1 - \hat{p}) \times (\frac{1}{n_1} + \frac{1}{n_2})}}$ $\hat{p} = \frac{m_1 + m_2}{n_1 + n_2}$

Избор на тест за проверка на хипотеза

Вид променлива	Брой групи	Независими / свързани групи	Проверка за нормалност	Препоръчителен тест
Количествена	1	–	Нормално разпределена	t-тест за една извадка
			Не е нормално разпределена	Рангов тест на Wilcoxon (за медиана)
Количествена	2	Независими	Нормално разпределени, хомогенни дисперсии	t-тест за две независими извадки
			Не са нормално разпределени	Mann–Whitney U тест
Количествена	2	Свързани	Нормално разпределени разлики	t-тест за две свързани извадки
			Разликите не са нормално разпределени	Рангов тест на Wilcoxon
Количествена	≥3	Независими	Нормално разпределени, хомогенни дисперсии	ANOVA
			Не са нормално разпределени	Тест на Kruskal–Wallis
Количествена	≥3	Свързани	Нормално разпределени разлики	ANOVA с повторни измервания
			Разликите не са нормално разпределени	Тест на Friedman
Качествена	2	–	–	χ^2 -тест за независимост Точен тест на Fisher (при дихотомни променливи и малки извадки)
Качествена	≥3	–	–	χ^2 -тест за независимост
Относителен дял	1	–	–	Z-тест за един дял
Относителен дял	2	–	–	Z-тест за два дяла
Ординална	2	Независими	–	Mann–Whitney U тест
Ординална	2	Свързани	–	Рангов тест на Wilcoxon
Ординална	≥3	Независими	–	Тест на Kruskal–Wallis
Ординална	≥3	Свързани	–	Тест на Friedman

КОРЕЛАЦИЯ И РЕГРЕСИЯ

Връзка между две качествени променливи

χ^2 -тест за независимост	
$E_{ij} = \frac{R_i \times C_j}{n}$ $\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$ $df = (R - 1) \times (C - 1)$	
Коефициент ϕ	Коефициент V на Cramer
$\phi = \sqrt{\frac{\chi^2}{n}}$	$V = \sqrt{\frac{\chi^2}{n \times (k - 1)}}$

χ^2 -разпределение

df	$\chi^2_{0.10}$	$\chi^2_{0.05}$	$\chi^2_{0.01}$
1	2.706	3.841	6.635
2	4.605	5.991	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086
6	10.645	12.592	16.812
7	12.017	14.067	18.475
8	13.362	15.507	20.090
9	14.684	16.919	21.666
10	15.987	18.307	23.209
15	22.307	24.996	30.578
20	28.412	31.410	37.566
30	40.256	43.773	50.892
40	51.805	55.758	63.691
50	63.167	67.505	76.154
100	118.498	124.342	135.807

Връзка между две количествени променливи

Коефициент на корелация r на Pearson	
$r = \frac{\sum (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \times \sum (Y_i - \bar{Y})^2}}$ $t = \frac{r \times \sqrt{n - 2}}{\sqrt{1 - r^2}}$ $df = n - 2$	
Проста линейна регресия	
$b = \frac{n \times \sum (X \times Y) - \sum X \times \sum Y}{n \times \sum X^2 - (\sum X)^2}$	$a = \bar{Y} - b \times \bar{X}$