ПРИКЛАДНАЯ СТАТИСТИКА

Шпаргалка по параметрическим критериям*

1. Про единственную выборку

Математическое ожидание при большом числе наблюдений

а. Наблюдаем: $X_1, X_2, \ldots, X_n$;

б. Предполагаем: X_i независимы и одинаково распределены (не обязательно нормально), количество наблюдений n велико.

в. Проверяемая гипотеза: $H_0\colon \mu = \mu_0$ против $H_a\colon \mu \neq \mu_0$;

г. Статистика:

$$Z = \frac{\bar{X} - \mu_0}{se(\bar{X})} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

д. При верной H_0 оказывается, что $Z \to \mathcal{N}(0;1)$;

Математическое ожидание при нормальных наблюдениях

а. Наблюдаем: $X_1, X_2, \ldots, X_n$;

б. Предполагаем: X_i независимы и одинаково нормально распределены $\mathcal{N}(\mu;\sigma^2)$, количество наблюдений n может быть мало.

в. Проверяемая гипотеза: $H_0\colon \mu = \mu_0$ против $H_a\colon \mu \neq \mu_0$;

*Эта pdf-ка, по факту, представляет из себя немного дополненный конспект Бориса Демешева: https://github.com/bdemeshev/pr201/raw/master/probab_pset/new_el.pdf г. Статистика:

$$t = \frac{\bar{X} - \mu_0}{se(\bar{X})} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

д. При верной H_0 оказывается, что $t \sim t_{n-1}$;

Математическое ожидание при нормальных наблюдениях и известной дисперсии

а. Наблюдаем: $X_1, X_2, ..., X_n$, знаем величину σ^2;

б. Предполагаем: X_i независимы и одинаково нормально распределены $\mathcal{N}(\mu; \sigma^2)$, количество наблюдений n может быть мало.

в. Проверяемая гипотеза: $H_0: \mu = \mu_0$ против $H_a: \mu \neq \mu_0$;

г. Статистика:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

д. При верной H_0 оказывается, что $Z \sim \mathcal{N}(0; 1)$;

Гипотеза о вероятности при наблюдениях с распределением Бернулли (0 или 1)

а. Наблюдаем: $X_1, X_2, ..., X_n$;

б. Предполагаем: X_i независимы и имеют распределение Бернулли: равны 1 с вероятностью p и 0 с вероятностью $1 - p$. Количество наблюдений n велико.

в. Проверяемая гипотеза: $H_0: p = p_0$ против $H_a: p \neq p_0$;

г. Статистика:

$$Z = \frac{\hat{p} - p_0}{se(\hat{p})} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Возможен вариант этой статистики:

$$Z = \frac{\hat{p} - p_0}{se(\hat{p})} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

д. При верной H_0 оказывается, что $Z \to \mathcal{N}(0; 1)$;

е. Гипотеза о вероятностях является частным случаем гипотезы о математическом ожидании при большом количестве наблюдений. Можно заметить, что $\hat{p} = \bar{X}$ и $\hat{\sigma}^2 = \hat{p}(1 - \hat{p}) \cdot \frac{n}{n-1}$. И потому также корректен вариант статистики

$$Z = \frac{\bar{X} - \mu_0}{se(\bar{X})} = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}$$

Гипотеза о дисперсии при нормальных наблюдениях

а. Наблюдаем: $X_1, X_2, \ldots, X_n$;

б. Предполагаем: X_i независимы и одинаково нормально распределены $\mathcal{N}(\mu; \sigma^2)$, количество наблюдений n может быть мало.

в. Проверяемая гипотеза: H_0: $\sigma = \sigma_0$ против H_a: $\sigma \neq \sigma_0$;

г. Статистика:

$$S = \frac{\sum(X_i - \bar{X})^2}{\sigma_0^2} = \frac{(n-1)\hat{\sigma}^2}{\sigma_0^2}$$

д. При верной H_0 оказывается, что $S \sim \chi^2_{n-1}$;

Гипотеза о дисперсии при нормальных наблюдениях и известном математическом ожидании

а. Наблюдаем: $X_1, X_2, \ldots, X_n$, знаем величину μ;

б. Предполагаем: X_i независимы и одинаково нормально распределены $\mathcal{N}(\mu; \sigma^2)$, количество наблюдений n может быть мало.

в. Проверяемая гипотеза: H_0: $\sigma = \sigma_0$ против H_a: $\sigma \neq \sigma_0$;

г. Статистика:

$$S = \frac{\sum(X_i - \bar{X})^2}{\sigma_0^2} = \frac{n\hat{\sigma}^2}{\sigma_0^2}$$

д. При верной H_0 оказывается, что $S \sim \chi^2_n$;

2. Про пару выборок

Гипотеза о разнице ожиданий при большом количестве наблюдений

а. Наблюдаем: $X_1, X_2, ..., X_{n_x}, Y_1, Y_2, ..., Y_{n_y}$.

Возможно, что $n_x \neq n_y$. Дисперсии σ_x^2 и σ_y^2 не знаем и не уверены, что они равны.

б. Предполагаем: X_i одинаково распределены между собой (не обязательно нормально), Y_i одинаково распределены между собой, но возможно совсем не так, как X_i (не обязательно нормально). Все величины независимы. Количества n_x и n_y велики.

в. Проверяемая гипотеза: H_0: $\mu_x - \mu_y = \delta_0$ против H_a: $\mu_x - \mu_y \neq \delta_0$;

г. Статистика:

$$Z = \frac{\bar{X} - \bar{Y} - \delta_0}{se(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\hat{\sigma}_x^2}{n_x} + \frac{\hat{\sigma}_y^2}{n_y}}}$$

д. При верной H_0 оказывается, что $Z \to \mathcal{N}(0; 1)$;

Гипотеза о разнице ожиданий при нормальности распределения обеих выборок и известных дисперсиях

а. Наблюдаем: $X_1, X_2, ..., X_{n_x}, Y_1, Y_2, ..., Y_{n_y}$. Возможно, что $n_x \neq n_y$. Дисперсии σ_x^2 и σ_y^2 знаем. Возможно, что дисперсии не равны.

б. Предполагаем: X_i одинаково распределены между собой $\mathcal{N}(\mu_x, \sigma_x^2)$, Y_i одинаково распределены между собой $\mathcal{N}(\mu_y, \sigma_y^2)$. Все величины независимы. Количества n_x и n_y любые.

в. Проверяемая гипотеза: H_0: $\mu_x - \mu_y = \delta_0$ против H_a: $\mu_x - \mu_y \neq \delta_0$;

г. Статистика:

$$Z = \frac{\bar{X} - \bar{Y} - \delta_0}{\sigma_{\bar{X} - \bar{Y}}} = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}$$

д. При верной H_0 оказывается, что $Z \sim \mathcal{N}(0; 1)$;

Гипотеза о разнице ожиданий при нормальности распределения обеих выборок и неизвестных но равных дисперсиях

а. Наблюдаем: $X_1, X_2, ..., X_{n_x}, Y_1, Y_2, ..., Y_{n_y}$. Возможно, что $n_x \neq n_y$. Дисперсии σ_x^2 и σ_y^2 равны, но неизвестны.

б. Предполагаем: X_i одинаково распределены между собой $\mathcal{N}(\mu_x, \sigma^2)$, Y_i одинаково распределены между собой $\mathcal{N}(\mu_y, \sigma^2)$. Все величины независимы. Количества n_x и n_y любые.

в. Проверяемая гипотеза: H_0: $\mu_x - \mu_y = \delta_0$ против H_a: $\mu_x - \mu_y \neq \delta_0$;

г. Статистика:

$$t = \frac{\bar{X} - \bar{Y} - \delta_0}{se(\bar{X} - \bar{Y})} = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{\hat{\sigma}^2}{n_x} + \frac{\hat{\sigma}^2}{n_y}}},$$

где

$$\hat{\sigma}^2 = \frac{\sum(X_i - \bar{X})^2 + \sum(Y_i - \bar{Y})^2}{n_x + n_y - 2}$$

д. При верной H_0 оказывается, что $t \sim t_{n_x+n_y-2}$;

Гипотеза о разнице ожиданий в связанных парах

а. Наблюдаем: $X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_n$. Количество X_i и Y_i одинаковое.

б. Предполагаем: внутри пары X_i и Y_i зависимы, а наблюдения с разными номерами независимы. Рассматриваем разницу $D_i = X_i - Y_i$ и получаем одномерную выборку. Величины D_i независимы и одинаково распределены. Возможно три описанных ранее случая :) Здесь для примера рассмотрим случай, когда $D_i \sim \mathcal{N}(\mu_d, \sigma_d^2)$ с неизвестной дисперсией.

в. Проверяемая гипотеза: H_0: $\mu_d = \mu_0$ против H_a: $\mu_d \neq \mu_0$;

г. Статистика:

$$t = \frac{\bar{D} - \mu_d}{se(\bar{D})} = \frac{\bar{X} - \bar{Y} - \mu_d}{\sqrt{\frac{\hat{\sigma}_d^2}{n}}},$$

где

$$\hat{\sigma}_d^2 = \frac{\sum(D_i - \bar{D})^2}{n-1} = \frac{\sum(X_i - Y_i - (\bar{X} - \bar{Y}))^2}{n-1}$$

д. При верной H_0 оказывается, что $t \sim t_{n-1}$;

Гипотеза о равенстве дисперсий при нормальности распределения обеих выборок

а. Наблюдаем: $X_1, X_2, \ldots, X_{n_x}, Y_1, Y_2, \ldots, Y_{n_y}$. Возможно, что $n_x \neq n_y$. Дисперсии σ_x^2 и σ_y^2 не знаем. Возможно, что дисперсии не равны.

б. Предполагаем: X_i одинаково распределены между собой $\mathcal{N}(\mu_x, \sigma^2)$, Y_i одинаково распределены между собой $\mathcal{N}(\mu_y, \sigma^2)$. Все величины независимы. Количества n_x и n_y любые.

в. Проверяемая гипотеза: H_0: $\sigma_x = \sigma_y$ против H_a: $\sigma_x \neq \sigma_y$;

г. Статистика:

$$F = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_y^2}$$

д. При верной H_0 оказывается, что $F \sim F_{n_x-1, n_y-1}$;