Формулы для лекции о проверке статистических гипотез

Обозначения

Имеем вектор данных, например столбец в таблице данных
$x_1, x_2, x_3, \ldots, x_n$

При изучении вероятностных свойств статистических критериев используем случайные величины $X_1, X_2, X_3, \ldots, X_n$

Комментарий.
Один кубик бросаем 5 раз
эквивалентно
подбрасываем 5 тождественных кубиков.

Обозначения

H_0 основная гипотеза
H_1 альтернативная гипотеза

Статистика критерия это функция данных

$T(x_1, x_2, x_3, \ldots, x_n)$

Примеры

Проверка простой гипотезы $H_0 : EX = 11$
$\bar{x}_n - 11$

Проверка гипотезы независимости.
Выборочный коэффициент корреляции Пирсона

$$\hat{\rho}_{Pearson} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)\cdot(y_i - \bar{y}_n)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2}} = \frac{cov(x, y)}{\sigma_x \cdot \sigma_y}$$

Правило 1

Основная гипотеза отвергается, если

$$T(x_1, x_2, x_3, \ldots, x_n) > c_\alpha$$

Основная гипотеза не отвергается, если

$$T(x_1, x_2, x_3, \ldots, x_n) \leqslant c_\alpha$$

Как находится критическое значение c_α ?

Вероятность ошибки первого рода

$$P_{H_0}(T(X_1, X_2, X_3, \ldots, X_n) > c_\alpha)$$

Вероятность ошибки второго рода

$$P_{H_1}(T(X_1, X_2, X_3, \ldots, X_n) \leqslant c_\alpha)$$

Мощность = 1 - вероятность ошибки второго рода =

$$P_{H_1}(T(X_1, X_2, X_3, \ldots, X_n) > c_\alpha)$$

Уравнение для определения порогового значения c_α

$$P_{H_0}(T(X_1, X_2, X_3, \ldots, X_n) > c_\alpha) = \alpha$$

Уровень значимости α
Обычно равен 0.05, 0.01 или 0.005

Обозначение.
Подставим наблюдения в формулу для статистики критерия. Результат вычислений обозначим $T_{эксп}$

$$T(x_1, x_2, x_3, \ldots, x_n) = T_{эксп}$$

Определение
p-значение (p-value, значимость)

$$P_{H_0}(T(X_1, X_2, X_3, \ldots, X_n) > T_{эксп}) = p - value$$

<u>Правило 2</u>

p-value $> \alpha$ <=> основная гипотеза не отвергается
p-value $< \alpha$ <=> основная гипотеза отвергается

Правило 1 и правило 2 эквивалентны.

Напоминание

Пусть $f(t)$ - плотность распределения случайной величины X, тогда

$$P(X \in [a, b]) = \int_a^b f(t) dt$$