ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ. ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ

Проверка статистических гипотез

- **Статистическая гипотеза** –гипотеза о виде неизвестного распределения или о параметрах известных распределений.
- **Нулевая (основная)** – выдвинутая гипотеза H_0.
- **Конкурирующая (альтернативная)** – гипотеза H_1, которая противоречит нулевой.

Пример:

$H_0 : a = 10$; $H_1 : a \neq 10$ или $H_1 : a > 10$

- **Простая** гипотеза содержит только одно предположение.

  Пример: H_0 :математическое ожидание нормального распределения равно 3.

- **Сложная** гипотеза состоит из конечного или бесконечного числа простых гипотез.

  Пример: H_0:математическое ожидание нормального распределения не равно 3.

Ошибки проверки статистических гипотез

- **Ошибка первого рода** состоит в том, что будет отвергнута правильная гипотеза.
- **Ошибка второго рода** состоит в том, что будет принята неправильная гипотеза.
- **Уровень значимости** α - вероятность совершения ошибки первого рода.

- **Статистический критерий** – случайная величина K, с известным законом распределения для проверки нулевой гипотезы.
- **Критическая область** – совокупность значений критерия, при которых нулевую гипотезу отвергают.
- **Область принятия гипотезы (область допустимых значений)** – совокупность значений критерия, при которых гипотезу принимают.
- **Критические точки (границы)** $\kappa_{кр}$ – точки, отделяющие критическую область от области принятия гипотезы.
- **Мощность критерия** - вероятность попадания критерия в критическую область при условии, что верна конкурирующая гипотеза.

Алгоритм проверки гипотезы

- 1) выбирается статистический критерий K ;
- 2) вычисляется его наблюдаемое значение $K_{набл}$;
- 3) определяется по известному уровню значимости α критическое значение $K_{кр}$;
- 4) если вычисленное значение $K_{набл}$ попадает в область принятия гипотезы, то нулевая гипотеза принимается, если в критическую область – нулевая гипотеза отвергается.

Критические области

- **Правосторонняя** критическая область, определяется неравенством: $K > \kappa_{кр}$.

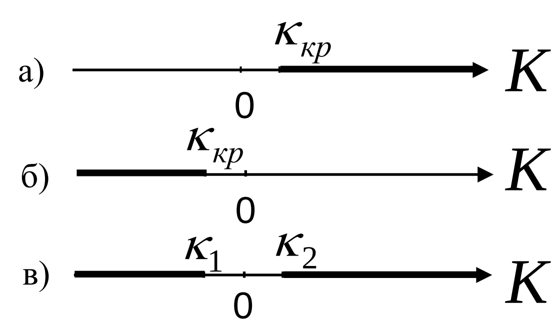
  рис. а)

- **Левосторонняя** – неравенством $K < \kappa_{кр}$.

  рис. б)

- **Двусторонняя** – неравенствами $K < \kappa_1$, $K > \kappa_2$ $(\kappa_2 > \kappa_1)$

  рис. в)



Сравнение двух средних нормальных генеральных совокупностей, дисперсии которых известны (независимые выборки)

- **Гипотезы:** $H_0 : M(X) = M(Y);\ H_1 : M(X) > M(Y),$

X, Y – нормальные генеральные совокупности

- **Критерий:** $Z = \dfrac{\overline{x}_B - \overline{y}_B}{\sqrt{D(X)/n + D(Y)/m}}$

$D(X),\ D(Y)$ – генеральные дисперсии,

$\overline{x_B}, \overline{y_B}$ – выборочные средние,

n, m – объёмы выборок из X и Y соответственно.

Критическая область – правосторонняя

$P\left(Z > z_{кр}\right) = \alpha,\ \alpha -$ уровень значимости,

$\Phi\left(z_{кр}\right) = 0{,}5\left(1 - 2\alpha\right)$

(таблица функции Лапласа).

При $Z_{набл} > z_{кр}$ – гипотеза H_0 отвергается.

При $Z_{набл} < z_{кр}$ – гипотеза принимается.

- Если $H_1: M(X) < M(Y)$,

то критическая область – левосторонняя

$$\Phi\left(z_{кр}\right) = 0{,}5(1-2\alpha)$$

при $Z_{набл} < -z_{кр}$ нулевую гипотезу отвергают.

- Если $H_1: M(X) \neq M(Y)$,

то критическая область двухсторонняя

$$\Phi\left(z_{кр}\right) = 0{,}5(1-\alpha)$$

при $Z_{набл} \in \left(-z_{кр}; z_{кр}\right)$ гипотезу принимают.

Пример

- В результате исследования накопления содержания натрия в водоёме по двум независимым выборкам, объёмы которых соответственно равны $n = 9$ и $m = 10$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние $\overline{x}_B = 25$ (мг/л) (выборка взята в области питания) и $\overline{y}_B = 38$(мг/л) (выборка взята ниже по течению). Генеральные дисперсии известны: $D(X) = 3{,}6$ (мг/л)^2; $D(Y) = 6$ (мг/л)^2. При уровне значимости $\alpha = 0{,}01$ проверить нулевую гипотезу H_0: $M(X) = M(Y)$, при конкурирующей гипотезе H_1: $M(X) < M(Y)$.

Выборочный коэффициент корреляции. Выборочные уравнения прямых среднеквадратической регрессии

Корреляционная таблица:

Y	X			
	x_1	x_2	…	x_k
y_1	n_{11}	n_{21}	…	n_{k1}
y_2	n_{12}	n_{22}	…	n_{k2}
…	…	…	…	…
y_m	n_{1m}	n_{2m}	…	n_{km}

n_{ij} – число появлений в выборке пары чисел (x_i, y_j).

Выборочный коэффициент корреляции

- $$r_B = \frac{\overline{(xy)_B} - \overline{x_B} \cdot \overline{y_B}}{\sigma_x \cdot \sigma_y}$$, где $$\overline{(xy)_B} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} x_i y_j}{n}$$,

$\overline{x}_B$, $\overline{y}_B$ – выборочные средние признаков X и Y, σ_x, σ_y – выборочные средние квадратические отклонения.

- **!!!** Выборочный коэффициент корреляции находится в пределах от -1 до 1.
- Чем ближе коэффициент корреляции к нулю, тем хуже зависимость выражается линейно.

Проверка гипотезы о значимости коэффициента корреляции

- **Гипотезы:** $H_0 : r_B = 0,\ H_1 : r_B \neq 0$.
- **Критерий:** $T = \dfrac{r_B \cdot \sqrt{n-2}}{\sqrt{1 - r_B^2}}$,

n – объём выборки.

Критическая область двусторонняя.

$t_{кр}(\alpha, к)$ находим по таблице распределения Стьюдента для двусторонней критической области ($k = n - 2$– число степеней свободы,

α – уровень значимости).

Уравнения линейной регрессии

- **Уравнение прямой регрессии Y на X:**

$$\overline{y_x} = r_B \cdot \frac{\sigma_y}{\sigma_x}\left(x - \overline{x_B}\right) + \overline{y_B}$$

- **Уравнение прямой регрессии X на Y:**

$$\overline{x_y} = r_B \cdot \frac{\sigma_x}{\sigma_y}\left(y - \overline{y_B}\right) + \overline{x_B}$$

- **Коэффициент регрессии Y на X:**

$$\rho_{yx} = r_B \cdot \frac{\sigma_y}{\sigma_x}$$

- **Коэффициент регрессии X на Y:**

$$\rho_{xy} = r_B \cdot \frac{\sigma_x}{\sigma_y}$$