## ПРОВЕРКА ГИПОТЕЗ О РАВЕНСТВЕ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ

*Статистическая гипотеза* $H_0$ – предположение о виде неизвестного распределения или о параметрах известного распределения.

*Конкурирующая* (*альтернативная*) гипотеза – гипотеза, противоположная гипотезе $H_0$. Гипотезу $H_0$ называют *исходной* (*нулевой*).

Проверка статистической гипотезы $H_0$ состоит в нахождении вероятности получить результаты, не лучшие реально полученных в эксперименте.

Содержание сводится к неформальному выбору связанного с результатами измерений критического события $A$, и последующему отысканию его условной вероятности $P(A\,|\,H)$ при гипотезе $H$. *Статистическим критерием*, или *статистикой* называют случайную величину с известным распределением, которая позволяет найти вероятность критического события. Если найденная вероятность оказывается меньше некоторого заранее заданного числа – *уровня значимости* $\alpha$ – то гипотеза $H$ отвергается на уровне значимости $\alpha$. Выбор уровня значимости является *выбором вероятности события, которое на практике считается невозможным*. В большинстве поисковых исследований уровень значимости выбирают равным $\alpha = 0{,}05$.

На практике часто возникает вопрос, является ли обнаруженное расхождение средних двух серий экспериментов *статистически значимым* – можно ли объяснить его случайными ошибками или же оно имеет закономерное происхождение (например, контроль качества продукции, изготовленной при различных технологических режимах). Решение этого вопроса сводится к сравнению неизвестных математических ожиданий.

Задача сравнения решается по разному в зависимости от того, являются ли известными дисперсии двух генеральных совокупностей, из которых для анализа извлечены выборки. Если дисперсии неизвестны – то в зависимости от того, равны ли они.

Если дисперсии подчиненных нормальному закону генеральных совокупностей, из которых извлечены выборки

$$\{x_i\},\ \{y_j\},\ i = \overline{1, N_1},\ \ j = \overline{1, N_2},$$

известны и равны $\sigma_x^2$ и $\sigma_y^2$, то связанная с разностью $\bar{x} - \bar{y}$ средних статистика

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{D[\bar{x} - \bar{y}]}} = (\bar{x} - \bar{y})\left(\frac{\sigma_x^2}{N_1} + \frac{\sigma_y^2}{N_2}\right)^{-\frac{1}{2}}$$

подчинена стандартному нормальному закону $N(0{,}1)$, что позволяет вычислить вероятность критического события

$$P(A\,|\,H_0) = 1 - \frac{1}{\sqrt{2\pi}}\int_{-|t|}^{-|t|} e^{-\frac{x^2}{2}}dx = 1 - 2\Phi(|t|) = 2 - 2\Phi^*(|t|),$$

где $\Phi$ – функция Лапласа, $\Phi^*$ – функция стандартного нормального распределения. Если вероятность $P(A\,|\,H_0)$ оказывается меньше уровня значимости, то гипотеза $H_0$ о равенстве математических ожиданий отвергается.

Если дисперсии генеральных совокупностей неизвестны, но предполагаются равными, то сравнение математических ожиданий начинается с отыскания смешанной оценки дисперсии разности средних:

$$D[\bar{x} - \bar{y}] = \frac{1}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)\left(\sum_{i=1}^{N_1}(x_i - \bar{x})^2 + \sum_{i=1}^{N_2}(y_i - \bar{y})^2\right).$$

После этого находится значение статистики

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{D[\bar{x} - \bar{y}]}}.$$

Найденная статистика подчинена распределению Стьюдента с $k = N_1 + N_2 - 2$ степенями свободы.

Если дисперсии генеральных совокупностей неизвестны и не предполагаются равными, то можно приближенно считать, что статистика

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{D[\bar{x} - \bar{y}]}},$$

также подчинена распределению Стьюдента. Однако число степеней свободы уже не является целым числом:

$$k = \frac{\left(\dfrac{s_x}{N_1} + \dfrac{s_y}{N_2}\right)^2}{\dfrac{s_x^2}{N_1^2(N_1 - 1)} + \dfrac{s_y^2}{N_2^2(N_2 - 1)}},$$

где

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{N_1 - 1}\sum_{i=1}^{N_1}(x_i - \bar{x})^2}, \quad s_y = \sqrt{s_y^2} = \sqrt{\frac{1}{N_2 - 1}\sum_{i=1}^{N_2}(y_i - \bar{y})^2}.$$

– оценки стандартных отклонений.

## ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДИСПЕРСИЙ

Дисперсия и связанные с ней характеристики – стандартное отклонение и коэффициент вариации – характеризуют такие показатели, как точность приборов и технологических процессов. Пусть имеются две нормально распределенные генеральные совокупности с неизвестными дисперсиями $\sigma_x^2$ и $\sigma_y^2$. Необходимо проверить нулевую гипотезу $H_0$: $\sigma_x^2 = \sigma_y^2$ о равенстве дисперсий.

Задача проверки сводится к сравнению оценок дисперсий

$$s_x^2 = \frac{1}{N_1 - 1}\sum_{i=1}^{N_1}(x_i - \bar{x})^2, \quad s_y^2 = \frac{1}{N_2 - 1}\sum_{i=1}^{N_2}(y_i - \bar{y})^2.$$

Статистика

$$F = \frac{s_x^2}{s_y^2},$$

равная отношению этих оценок, подчинена распределению Фишера с числом степеней свободы $N_1 - 1$, $N_2 - 1$.

Если проверка выполняется относительно двусторонней альтернативы $H_1$: $\sigma_x^2 \neq \sigma_y^2$, то вероятность критического события находится по формуле

$$P(A \mid H_0) = \begin{cases} P', P' \leq 1 \\ 2 - P', P' > 1 \end{cases},$$

где $P' = 2F\left(\dfrac{s_x^2}{s_y^2}, N_1 - 1, N_2 - 1\right)$, $F$ – функция $F$-распределения.