Теория вероятностей и математическая статистика
Лектор А.А. Лобузов
Семестр 6 Лекция 11

Проверка статистических гипотез о равенстве математических ожиданий

Рассматриваются нормально распределенные случайные величины ξ_1 и ξ_2: $\xi_i \sim N(a_i, \sigma_i^2)$, $\mathrm{M}\xi_i = a_i$, $\mathrm{D}\xi_i = \sigma_i^2$. По случайным выборкам $\mathbf{X} = (X_1, X_2, ..., X_N)$ и $\mathbf{Y} = (Y_1, Y_2, ..., Y_M)$ из распределений с.в. ξ_1 и ξ_2 соответственно нужно проверить гипотезу $\mathbf{H}_0 = \{\mathrm{M}\xi_1 = \mathrm{M}\xi_2\}$ при конкурирующих гипотезах $\mathbf{H}_1 = \{\mathrm{M}\xi_1 \neq \mathrm{M}\xi_2\}$, $\mathbf{H}_1^+ = \{\mathrm{M}\xi_1 > \mathrm{M}\xi_2\}$, $\mathbf{H}_1^- = \{\mathrm{M}\xi_1 < \mathrm{M}\xi_2\}$.

Рассмотрим проверку при разных предположениях.

I. Если дисперсии известны и равны $\sigma_1^2 = \sigma_2^2 = \sigma^2$ рассматривается статистика

$$T_{N,M}(\mathbf{X},\mathbf{Y}) = \frac{\overline{\mathbf{X}} - \overline{\mathbf{Y}}}{\sigma\sqrt{N^{-1} + M^{-1}}}$$

При указанных предположениях выборочные средние

$\overline{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N} X_i$ и $\overline{\mathbf{Y}} = \frac{1}{M}\sum_{i=1}^{M} Y_i$ имеют нормальные распределения $N(a_1, \frac{\sigma^2}{N})$ и $N(a_2, \frac{\sigma^2}{M})$ соответственно, так как $\mathbf{D}\overline{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{D}X_i = \frac{1}{N^2}N\sigma^2 = \frac{\sigma^2}{N}$.

Статистика $T_{N,M}(\mathbf{X},\mathbf{Y})$, являясь линейной комбинацией нормально распределенных случайных величин, тоже будет иметь нормальное распределение, и при выполнении гипотезы $\mathbf{H}_0 = \{a_1 = a_2\}$ имеем $T_{N,M}(\mathbf{X},\mathbf{Y}) \sim N(0,1)$.

Поэтому получаем следующую схему проверки гипотезы $\mathbf{H}_0$ при альтернативной гипотезе $\mathbf{H}_1$ по выборкам $\mathbf{x} = (x_1, x_2, ..., x_N)$ и $\mathbf{y} = (y_1, y_2, ..., y_M)$, полученных при наблюдениях случайных величин ξ_1 и ξ_2 соответственно, при уровне значимости α:

1. По числовым выборкам $\mathbf{x} = (x_1, x_2, ..., x_N)$ и $\mathbf{y} = (y_1, y_2, ..., y_M)$ находим выборочные средние $\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} x_i$ и $\overline{\mathbf{y}} = \frac{1}{M}\sum_{i=1}^{M} y_i$.

2. Вычисляем значение статистики

$$T_{N,M} = \frac{\overline{\mathbf{x}} - \overline{\mathbf{y}}}{\sigma\sqrt{\frac{1}{N} + \frac{1}{M}}}.$$

3. По заданному значению уровня значимости α берем $u_{1-\frac{\alpha}{2}}$ квантиль уровня $(1-\frac{\alpha}{2})$ стандартного нормального распределения $N(0,1)$ и делаем вывод о справедливости гипотезы: если $|T_{N,M}| \le u_{1-\frac{\alpha}{2}}$, то при уровне значимости α принимается гипотеза $\mathbf{H}_0$; если $|T_{N,M}| > u_{1-\frac{\alpha}{2}}$, то при уровне значимости α принимается альтернативная гипотеза $\mathbf{H}_1$.

В схеме проверки гипотезы $\mathbf{H}_0$ при конкурирующих гипотезах $\mathbf{H}_1^+ = \{\mathrm{M}\xi_1 > \mathrm{M}\xi_2\}$ и $\mathbf{H}_1^- = \{\mathrm{M}\xi_1 < \mathrm{M}\xi_2\}$ отличается лишь пункт 3.

При проверке гипотезы $\mathbf{H}_0$ при конкурирующей гипотезе $\mathbf{H}_1^+=\{\mathrm{M}\xi_1>\mathrm{M}\xi_2\}$ по заданному значению уровня значимости α берем $u_{1-\alpha}$ квантиль уровня $(1-\alpha)$ стандартного нормального распределения $N(0,1)$ и делаем вывод о справедливости гипотезы: если $T_{N,M}\le u_{1-\alpha}$, то при уровне значимости α принимается гипотеза $\mathbf{H}_0$; если $T_{N,M}>u_{1-\alpha}$, то при уровне значимости α принимается гипотеза $\mathbf{H}_1^+$.

При проверке гипотезы $\mathbf{H}_0$ при конкурирующей гипотезе $\mathbf{H}_1^-=\{\mathrm{M}\xi_1<\mathrm{M}\xi_2\}$ по заданному значению уровня значимости α берем u_α квантиль уровня α стандартного нормального распределения $N(0,1)$ и делаем вывод о справедливости гипотезы: если $T_{N,M}\ge u_\alpha$, то при уровне значимости α принимается гипотеза $\mathbf{H}_0$; если $T_{N,M}<u_\alpha$, то при уровне значимости α принимается гипотеза $\mathbf{H}_1^-$.

II. Если дисперсии известны, но не равны $\sigma_1^2\ne\sigma_2^2$ рассматривается статистика $T_{N,M}(\mathbf{X},\mathbf{Y})=\dfrac{\overline{\mathbf{X}}-\overline{\mathbf{Y}}}{\sqrt{\dfrac{\sigma_1^2}{N}+\dfrac{\sigma_2^2}{M}}}$, которая при выполнении гипотезы $\mathbf{H}_0=\{a_1=a_2\}$ тоже имеет стандартное нормальное распределение $N(0,1)$ и схемы проверки гипотезы $\mathbf{H}_0$ будут как в **I.**

III. Если дисперсии неизвестны, но предполагается, что они равны, рассматривается статистика

$$T_{N,M}(\mathbf{X},\mathbf{Y}) = \frac{\overline{\mathbf{X}}-\overline{\mathbf{Y}}}{\sqrt{S_1^2(N-1)+S_2^2(M-1)}}\sqrt{\frac{MN(N+M-2)}{N+M}},$$

которая при выполнении гипотезы $\mathbf{H}_0 = \{a_1 = a_2\}$ имеет распределение Стьюдента с числом степеней свободы $N+M-2$.

Получаем следующую схему проверки гипотезы $\mathbf{H}_0 = \{a_1 = a_2\}$ при альтернативной гипотезе $\mathbf{H}_1 = \{a_1 \neq a_2\}$ по выборкам $\mathbf{x} = (x_1, x_2, ..., x_N)$ и $\mathbf{y} = (y_1, y_2, ..., y_M)$, полученных при наблюдениях случайных величин ξ_1 и ξ_2 соответственно, при уровне значимости α:

1. По числовым выборкам $\mathbf{x} = (x_1, x_2, ..., x_N)$ и $\mathbf{y} = (y_1, y_2, ..., y_M)$ находим

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} x_i \,,\; \overline{\mathbf{y}} = \frac{1}{M}\sum_{i=1}^{M} y_i \,,\; S_1^2(N-1) = \sum_{i=1}^{N}(x_i - \overline{\mathbf{x}})^2 \,,\; S_2^2(M-1) = \sum_{i=1}^{M}(y_i - \overline{\mathbf{y}})^2 .$$

2. Вычисляем значение статистики

$$T_{N,M} = \frac{\overline{\mathbf{x}}-\overline{\mathbf{y}}}{\sqrt{S_1^2(N-1)+S_2^2(M-1)}}\sqrt{\frac{MN(N+M-2)}{N+M}} .$$

3. По заданному значению уровня значимости α берем $t_{1-\frac{\alpha}{2}}(N+M-2)$ квантиль уровня $(1-\frac{\alpha}{2})$ распределения Стьюдента с числом степеней свободы $N+M-2$ и делаем вывод о справедливости гипотезы: если $|T_{N,M}| \le t_{1-\frac{\alpha}{2}}(N+M-2)$, то при уровне значимости α принимается гипотеза $\mathbf{H}_0$; если $|T_{N,M}| > t_{1-\frac{\alpha}{2}}(N+M-2)$, то при уровне значимости α принимается альтернативная гипотеза $\mathbf{H}_1$.

В схеме проверки гипотезы $\mathbf{H}_0 = \{a_1 = a_2\}$ при конкурирующих гипотезах $\mathbf{H}_1^+ = \{a_1 > a_2\}$ и $\mathbf{H}_1^- = \{a_1 < a_2\}$ отличается лишь пункт 3.

При проверке гипотезы $\mathbf{H}_0 = \{a_1 = a_2\}$ при конкурирующей гипотезе $\mathbf{H}_1^+ = \{a_1 > a_2\}$ по заданному значению уровня значимости α берем $t_{1-\alpha}(N+M-2)$ квантиль уровня $(1-\alpha)$ распределения Стьюдента с числом степеней свободы $N+M-2$ и делаем вывод о справедливости гипотезы: если $T_{N,M} \le t_{1-\alpha}(N+M-2)$, то при уровне значимости α принимается гипотеза $\mathbf{H}_0$; если $T_{N,M} > t_{1-\alpha}(N+M-2)$, то при уровне значимости α принимается гипотеза $\mathbf{H}_1^+$.

При проверке гипотезы $\mathbf{H}_0 = \{a_1 = a_2\}$ при конкурирующей гипотезе $\mathbf{H}_1^- = \{a_1 < a_2\}$ по заданному значению уровня значимости α берем $t_{\alpha}(N+M-2)$ квантиль уровня α распределения Стьюдента с числом степеней свободы $N+M-2$ и делаем вывод о справедливости гипотезы: если $T_{N,M} \ge t_{\alpha}(N+M-2)$, то при уровне значимости α принимается гипотеза $\mathbf{H}_0$; если $T_{N,M} < t_{\alpha}(N+M-2)$, то при уровне значимости α принимается гипотеза $\mathbf{H}_1^-$.