Статистические критерии

1 t-критерии

Пусть у нас есть две независимые *нормально распределённые* выборки $X_1, ..., X_m$ и $Y_1, ..., Y_n$ *с одинаковыми дисперсиями* и мы хотим проверить отличие от нуля разности средних этих выборок.

При нулевой гипотезе (то есть если средние равны) статистика

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(m-1)\sigma_X^2 + (n-1)\sigma_Y^2}{m+n-2}}\sqrt{\frac{1}{m} + \frac{1}{n}}}$$

имеет распределение Стьюдента с $m + n - 2$ степенями свободы. Здесь $\bar{X}$ и $\bar{Y}$ — выборочные средние, σ_X^2 и σ_Y^2 — выборочные дисперсии

$$\sigma_X^2 = \frac{1}{m-1}\sum_j (X_j - \bar{X})^2.$$

В случае, когда дисперсии выборок различны, применяется критерий Велша со статистикой

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}}},$$

которая при нулевой гипотезе имеет распределение Стьюдента с количеством степеней свободы

$$\frac{\left(\frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}\right)^2}{\frac{\sigma_X^4}{m^2(m-1)} + \frac{\sigma_Y^4}{n^2(n-1)}}.$$

2 Множественные сравнения. Поправки Бонферрони

α - вероятность ошибки первого рода для используемого критерия. Мы считаем результат статистически значимым (отвергаем нулевую гипотезу), если p-значение $< \alpha$.

Для множественных сравнений нужно делать поправки. Пусть у нас есть m независимых испытаний, для которых мы получили р-значения $p_1, ..., p_m$. Тогда, при выполнении нулевой гипотезы, вероятность того, что хотя бы одно $p_j < \alpha$, будет больше, чем

$$> m\alpha - \frac{m(m-1)}{2}\alpha^2$$

по формуле включения и исключения.

Даже без предположения независимости эта вероятность не превосходит $m\alpha$. На этом основаны *поправки Бонферрони*. Мы считаем результат статистически значимым, если хотя бы в одном сравнении $p_j < \alpha/m$. Вероятность этого события при нулевой гипотезе уже не превосходит α.

Также применяются немного более сложные поправки Холма-Бонферрони.

3 Дисперсионный анализ

Сравнение средних у нескольких групп. Название «дисперсионный анализ» (ANOVA) дано из-за того, что здесь используется F-критерий Фишера.

Можно рассматривать дисперсионный анализ как линейную модель.

3.1 Односторонний дисперсионный анализ

Используется модель

$$Y_{ij} = \mu + \alpha_j + e_{ij}, \tag{1}$$

где $\mu + \alpha_j$ — средние в группах, α_j удовлетворяют ограничению

$$\sum_j \alpha_j = 0,$$

чтобы они определялись однозначно. Предполагаем, что все ошибки e_{ij} независимы и одинаково распределены, имеют нормальное распределение с нулевым средним.

Дизайн *сбалансирован*, если количества наблюдений во всех группах одинаковы. На это условие всегда нужно обращать внимание при расчетах. Если дизайн не сбалансирован, то можно использовать веса у наблюдений, как и в других линейных и обобщенных линейных моделях.

Нулевая гипотеза: $\alpha_1 = \alpha_2 = \cdots = \alpha_J = 0$. Альтернативная гипотеза $\alpha_j \neq 0$ для некоторого j.

Для проверки этой гипотезы используется F-критерий Фишера, сравнивающий модель (3) с упрощенной моделью

$$Y_{ij} = \mu + e_{ij}. \tag{2}$$

Выборочная сумма квадратов ошибок для модели (2)

$$SSE_2 = \sum_{i,j} Y_{ij}^2 - \frac{1}{N}\left(\sum_{i,j} Y_{ij}\right)^2,$$

где N — общее количество наблюдений. То же для модели (3)

$$SSE_1 = \sum_{j=1}^{J}\left(\sum_i Y_{ij}^2 - \frac{1}{N_j}\left(\sum_i Y_{ij}\right)^2\right),$$

где N_j — количество наблюдений в j-й группе.

Статистика

$$F = \frac{(SSE_2 - SSE_1)/(J-1)}{SSE_1/(N-1)}$$

имеет распределение Фишера с количеством степеней свободы $(J-1, N-1)$.

3.2 Двусторонний дисперсионный анализ

Модель

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + e_{ijk}, \tag{3}$$

где делаются те же предположения относительно ошибок. В некоторых случаях рассматривается модель без «взаимодействий» γ_{jk}.

4 Критерий Краскела-Уоллиса

Это аналог критерия Манна-Уитни для сравнения распределений нескольких независимых выборок. В английской литературе иногда этот тест называют one-way ANOVA on ranks.

Используется статистика

$$\frac{12}{N(N+1)} \sum_{j=1}^{J} N_j \left(\bar{R}_j - \frac{N+1}{2} \right)^2,$$

где $\bar{R}_j$ — средний ранг наблюдений в j-й группе.

Нулевая гипотеза: распределения всех групп наблюдений совпадают. При нулевой гипотезе предельное распределение этой статистики есть распределение χ^2 с $J-1$ степенью свободы.

5 Задача о сопряженных признаках. Критерий χ^2 Пирсона

Пусть у нас есть выборка $(X_1, Y_1), \ldots, (X_N, Y_N)$. Величины X принимают значения $1, ..., k$, а величины Y — значения $1, ..., l$. Нам нужно установить, связаны ли между собой эти признаки, то есть являются ли случайные величины X_1 и Y_1 независимыми.

Статистика

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(O_{ij} - E_{ij})^2}{E_{ij}},$$

где

$$O_{ij} = \#\{r : X_r = i, Y_r = j\}$$

— наблюдаемые количества значений $(X, Y) = (i, j)$,

$$E_{ij} = \hat{p}_{i*}\, \hat{p}_{*j}$$

— ожидаемые количества значений $(X, Y) = (i, j)$, где

$$\hat{p}_{i*} = \frac{\#\{r : X_r = i\}}{N} = \frac{\sum_{j=1}^{l} O_{ij}}{N},$$

$$\hat{p}_{*j} = \frac{\#\{r : Y_r = j\}}{N} = \frac{\sum_{i=1}^{k} O_{ij}}{N}.$$

При выполнении нулевой гипотезы (величины X_1 и Y_1 независимы) распределение статистики χ^2 слабо сходится к распределению χ^2 с $(k-1)(l-1)$ степенью свободы.

Если нужно сравнить две независимые группы наблюдений $0, 1$, то критерий χ^2 применяется следующим образом. Величина X соответствует группе наблюдений. Величина Y соответствует самим наблюдениям.

6 Точный критерий Фишера

(exact Fisher's test, не путать с F-критерием Фишера). У нас опять есть выборка $(X_1, Y_1), \ldots, (X_N, Y_N)$. Величины X принимают значения $1, ..., k$, а величины Y — значения $1, ..., l$. Нам нужно установить, связаны ли между собой эти признаки, то есть являются ли случайные величины X_1 и Y_1 независимыми.

Пусть

$$p_{i*} = \mathbf{P}(X_1 = i),$$
$$p_{*j} = \mathbf{P}(Y_1 = j).$$

Тогда при нулевой гипотезе вероятность

$$p(C_{ij},\ i = 1, ..., k,\ j = 1, ..., l) := \mathbf{P}(\{O_{ij} = C_{ij} \text{ для всех } i, j\})$$

равна

$$p(C_{ij}) = \frac{N!}{\prod_{i,j} C_{ij}!} \prod_{i,j} (p_{i*} p_{*j})^{C_{ij}}$$

(это полиномиальное распределение), а вероятности получить маргинальные распределения исходов

$$p(C_{i*}) := \mathbf{P}\left(\sum_j O_{ij} = C_{i*} \text{ для всех } i\right),$$

$$p(C_{*j}) := \mathbf{P}\left(\sum_i O_{ij} = C_{*j} \text{ для всех } j\right)$$

равны

$$p(C_{i*}) := \frac{N!}{\prod_i C_{i*}!} \prod_i (p_{i*})^{C_{i*}},$$

$$p(C_{*j}) := \frac{N!}{\prod_j C_{*j}!} \prod_j (p_{*j})^{C_{i*}}.$$

Поэтому если мы рассмотрим условную вероятность при фиксации маргинальных значений, то получим

$$p(C_{ij}|C_{i*}, C_{*j}) := \mathbf{P}\left(\left\{O_{ij} = C_{ij}\ \forall i, j \,\middle|\, \sum_j O_{ij} = C_{i*}\ \forall i;\ \sum_i O_{ij} = C_{*j}\ \forall j\right\}\right)$$

$$= \frac{\prod_i C_{i*}!\ \prod_j C_{*j}!}{N! \prod_{i,j} C_{ij}!}.$$

P-значение вычисляется по формуле

$$p = \sum_{C_{ij}:\ C_{i*}=O_{i*}\forall i;\ C_{*j}=O_{*j}\forall j;\ p(C_{ij}|C_{i*},C_{*j}) \le p(O_{ij}|O_{i*},O_{*j})} p(C_{ij}|C_{i*}, C_{*j}).$$

В этом тесте существенным образом используется предположение о заданных маргинальных значениях, которое на практике обычно никогда не выполняется.

Обычно точный критерий Фишера применяют тогда, когда не менее $1/4$ всех ожидаемых значений E_{ij} меньше 5. В остальных случаях обычно используют критерий χ^2 (с поправкой непрерывности Йейтса).

7 Критерий МакНемара

Пусть у нас есть выборка парных наблюдений $0, 1$: $(X_1, Y_1), \ldots, (X_N, Y_N)$. Например, мы можем делать измерения для одних и тех объектов в два разных момента времени (X и Y). И нам нужно понять, так же ли часто 1 меняется на 0, как 0 меняется на 1, другими словами, различаются ли вероятности выпадения $(1, 0)$ и выпадения $(0, 1)$.

Матрица наблюдаемых количеств (contingency table)

S =

	Y=0	Y=1
X=0	a	b
X=1	c	d

.

При выполнении нулевой гипотезы и при фиксации $b+c$, величина b будет иметь биномиальное распределение с параметрами $b + c, 1/2$. Величина b и является статистикой этого теста.

Хотя этот тест часто применяется на практике, нужно быть очень осторожным с его применением и смотреть, действительно ли предположения этого теста соответствуют условиям эксперимента. Часто лучше использовать условную логистическую регрессию (conditional logistic regression).