Проверка однородности генеральных дисперсий

Лекция №6
для студентов 2 курса,
обучающихся по специальности 060609 – Медицинская кибернетика

доц. Шапиро Л.А.

Красноярск, 2015 г.

План лекции:

1. Актуальность темы.
2. Сравнение двух генеральных дисперсий по независимым выборкам из нормальных совокупностей.
3. Сравнение нескольких генеральных дисперсий. Критерии Кочрена.
4. Сравнение нескольких генеральных дисперсий. Критерий Бартлетта , Левене.
5. Заключение

Актуальность темы

На практике задача сравнений дисперсий возникает, если требуется сравнить точность приборов, инструментов, методов измерений и т.д. Лучше тот прибор, инструмент, метод, который обеспечивает наименьшее рассеяние результатов измерений, т.е. наименьшую дисперсию.

Сравнение двух генеральных дисперсий по независимым выборкам из нормальных совокупностей. Критерий Фишера-Снедекора.

Пусть есть две независимые выборки значений нормально распределенной величины X: $x_1, x_2, \ldots, x_n$ - всего n элементов, и нормально распределенной величины Y: $y_1, y_2, \ldots, y_m$ - m элементов.

Для этих выборок найдены исправленные выборочные дисперсии s^2_x и s^2_y.

Требуется по исправленным выборочным дисперсиям при заданном уровне значимости α проверить нулевую гипотезу, что генеральные дисперсии совокупностей X и Y равны между собой:

$$H_0: D[X] = D[Y]$$

Гипотеза проверяется по критерию Фишера:

$$F = \frac{S_1^2}{S_2^2}$$

Величина F при условии справедливости нулевой гипотезы имеет распределение Фишера-Снедекора, со степенями свободы

$k_1=n-1$, $k_2=m-1$, где n- объем выборки, по которой вычислена большая дисперсия, m - объем выборки, по которой вычислена меньшая дисперсия (распределение Фишера-Снедекора зависит только от числа степеней свободы и не зависит от других параметров).

Критическая область строится в зависимости от вида конкурирующей гипотезы.

1. Односторонняя критическая область.

H_0: D[X] = D[Y]

H_1: D[X] > D[Y]

Вероятность попадания критерия в эту область:

$$P\,[F>F_{кр}(\alpha, k_1, k_2)] = \alpha$$

Критическую точку $F_{кр}(\alpha, k_1, k_2)$ находим по таблице распределения Фишера.

При $F_{набл} > F_{кр}$ нулевая гипотеза отвергается, генеральные дисперсии различаются

При $F_{набл} < F_{кр}$ нулевая гипотеза принимается, генеральные дисперсии равны

Пример: По двум независимым выборкам $n_1=12$ и $n_2=15$ из нормально распределенных генеральных совокупностей X и Y найдены исправленные выборочные дисперсии $s^2_x=11{,}41$ и $s^2_y=6{,}52$.

При уровне значимости 0,05 проверить нулевую гипотезу H_0: D[X] = D[Y] о равенстве генеральных дисперсий при конкурирующей гипотезе

$$H_1: D[X] > D[Y].$$

Решение: Найдем отношение большей исправленной дисперсии к меньшей:

$$F_{набл}=11{,}41/6{,}52=1{,}75$$

$k_1=12-1=11$, $k_2=15-1=14$

$F_{кр}(0{,}05,\ 11,\ 14)=2{,}56$

Так как $F_{набл}<F_{кр}$ (1,75<2,56) нет оснований отвергнуть нулевую гипотезу о равенстве генеральных дисперсий.

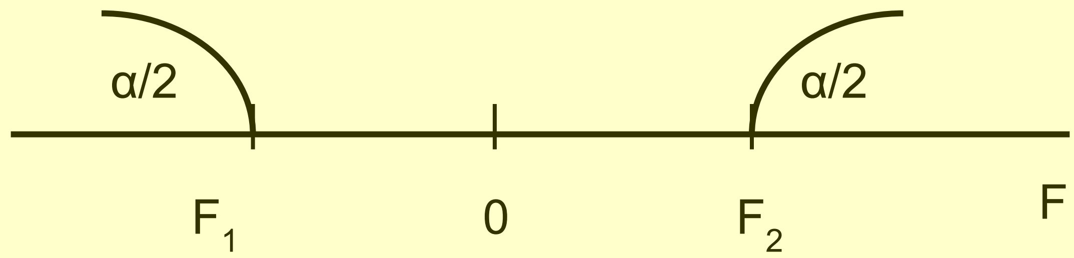
2. Двусторонняя критическая область.

H_0: $D[X] = D[Y]$

H_1: $D[X] \neq D[Y]$

Строим двустороннюю критическую область так, чтобы вероятность попадания критерия в эту область в предположении справедливости нулевой гипотезы равна α. Наибольшая мощность критерия (вероятность попадания в критическую область при справедливости конкурирующей гипотезы) достигается тогда, когда вероятность попадания критерия в каждый из двух интервалов равна α/2.

$P\ (F<F_1) = \alpha/2$ $P\ (F>F_2) = \alpha/2$



т.е. область принятия гипотезы будет $F_1<F<F_2$.

Правую точку F_2 находим по таблице. Левой точки таблица не содержит. Однако достаточно найти правую критическую точку F_2 при уровне значимости вдвое меньше заданного ($\alpha/2$). Вероятность попадания критерия в «левую часть» тоже равна $\alpha/2$. Так как эти события несовместны, то вероятность попадания критерия во всю двустороннюю область будет равна $\alpha/2+ \alpha/2= \alpha$

- Пример:По двум независимым выборкам $n_1=10$ и $n_2=18$ из нормально распределенных генеральных совокупностей X и Y найдены исправленные выборочные дисперсии $s^2_x=1,23$ и $s^2_y=0,41$.

При уровне значимости 0,1 проверить нулевую гипотезу H_0: $D[X] = D[Y]$ о равенстве генеральных дисперсий при конкурирующей гипотезе

$$H_1: D[X] \neq D[Y].$$

Решение: Найдем отношение большей исправленной дисперсии к меньшей:

$F_{набл}=1,23/0,41=3$

$k_1=10-1=9$, $k_2=18-1=17$

$F_{кр}(0{,}05,\ 9,\ 17)=2{,}5$

Так как $F_{набл}>F_{кр}$ (3>2,5) нулевая гипотеза о равенстве генеральных дисперсий отвергается.

Сравнение исправленной выборочной дисперсии с гипотетической генеральной дисперсией нормальной совокупности.

- Генеральная дисперсия хотя и неизвестна, но можно предполагать теоретически или из предыдущего опыта, что она равна σ_0^2.
- Имеется выборка с исправленной дисперсией S^2 с $k=n-1$ степенями свободы. Требуется проверить нулевую гипотезу, что при заданном уровне значимости генеральная дисперсия равна гипотетическому значению σ_0^2.
- Т.к. S^2 –несмещенная оценка генеральной дисперсии нулевую гипотезу можно записать в виде: H_0: $M(S^2) = \sigma_0^2$, т.е. требуется установить значимо или нет различаются выборочная и генеральная дисперсия.

Критерий принятия гипотезы:

$$\chi^2_{набл} = \frac{(n-1)S^2}{\sigma^2}$$

Критическая область строится в зависимости от конкурирующей гипотезы:

1. H_0: $\sigma^2 = \sigma_0^2$

   H_1: $\sigma^2 > \sigma_0^2$ правосторонняя область

$$P\,[\chi^2 > \chi^2_{кр}(\alpha, k)] = \alpha, \quad k=n-1$$

- Пример: По выборке n=13 из нормально распределенной генеральной совокупности найдена исправленная выборочная дисперсии $s^2=14{,}6$.

1. При уровне значимости 0,01 проверить нулевую гипотезу H_0: $\sigma^2 = \sigma_0^2=12$

   H_1: $\sigma^2 > 12$.

Решение:

$$\chi^2_{набл} = \frac{(n-1)S^2}{\sigma^2} = \frac{(13-1)\cdot 14{,}6}{12} = 14{,}6$$

$\chi^2_{кр}(0{,}01,\ 12)=26{,}2$ $\chi^2_{набл} < \chi^2_{кр}$ $(14{,}6 < 26{,}2)$

нулевая гипотеза не отвергается- различие между выборочной дисперсией (14,6) и гипотетической генеральной дисперсией (12)-незначимо.

2. При уровне значимости α проверить нулевую гипотезу H_0: $\sigma^2 = \sigma_0^{\ 2}$ $\quad$ H_1: $\sigma^2 \neq \sigma_0^{\ 2}$.

Решение: Находим двустороннюю критическую область:

$$P\,[\chi^2 < \chi^2_{\text{левкр}}(\alpha/2, k)] = \alpha/2, \quad k=n-1$$

$$P\,[\chi^2 > \chi^2_{\text{правкр}}(\alpha/2, k)] = \alpha/2,$$

В таблице есть только правосторонние критические точки.

т.к. события $\chi^2 < \chi^2_{\text{левкр}}$ и $\chi^2 > \chi^2_{\text{левкр}}$ противоположны сумма их вероятностей равна 1:

$$P\,(\chi^2 < \chi^2_{\text{левкр}}) + P\,(\chi^2 > \chi^2_{\text{левкр}}) = 1$$

$$P\,(\chi^2 > \chi^2_{\text{левкр}}) = 1 - P\,(\chi^2 < \chi^2_{\text{левкр}})$$

Следовательно, по таблице находим $\chi^2_{\text{правкр}}(\alpha/2, k)]$ и $\chi^2_{\text{левкр}}(1-\alpha/2, k)$

Если $\chi^2_{\text{левкр}} < \chi^2 < \chi^2_{\text{правкр}}$ нет оснований отвергнуть нулевую гипотезу

Если $\chi^2_{\text{набл}} < \chi^2_{\text{левкр}}$ или $\chi^2_{\text{набл}} > \chi^2_{\text{правкр}}$ нулевая гипотеза отвергается

- Пример: По выборке n=13 из нормально распределенной генеральной совокупности найдена исправленная выборочная дисперсия $s^2=10{,}3$.

При уровне значимости 0,02 проверить нулевую гипотезу H_0: $\sigma^2 = \sigma_0^2=12$

H_1: $\sigma^2 \neq 12$.

Решение:

$$\chi^2_{набл} = \frac{(n-1)S^2}{\sigma^2} = \frac{(13-1)\cdot 10{,}3}{12} = 10{,}3$$

$\chi^2_{крправ}(0{,}01,\ 12)=26{,}2$ $\chi^2_{левкр}(0{,}99,\ 12)=3{,}57$

$3{,}57 < 10{,}3 < 26{,}2$

нулевая гипотеза не отвергается - различие между выборочной дисперсией (10,3) и гипотетической генеральной дисперсией (12) - незначимо.

3. При уровне значимости α проверить нулевую гипотезу $H_0: \sigma^2 = \sigma_0^{\,2}$ $H_1: \sigma^2 < \sigma_0^{\,2}$.

Левосторонняя критическая область.

по таблице находим $\chi^2_{кр}(1-\alpha, k)$

Если $\chi^2_{набл} > \chi^2_{кр}(1-\alpha, k)$ нет оснований отвергнуть нулевую гипотезу

Если $\chi^2_{набл} < \chi^2_{кр}(1-\alpha, k)$ нулевая гипотеза отвергается

$\chi^2_{левкр}(0{,}98,\ 12) = 4{,}18$. $4{,}18 < 10{,}3$ нулевая гипотеза отвергается.

Сравнение нескольких генеральных дисперсий по независимым выборкам равного объема из нормальных совокупностей, критерий Кочрена.

Пусть генеральные совокупности X_1, X_2,..., X_l распределены нормально. Из этих совокупностей извлечены l независимых выборок одинакового объема n и по ним найдены исправленные выборочные дисперсии s^2_1 , $s^2_{2,}$ … s^2_l с числом степеней свободы k=n- l.

Требуется по исправленным выборочным дисперсиям при заданном уровне значимости α проверить нулевую гипотезу, что генеральные дисперсии совокупностей равны между собой: H_0: $D(X_1) = D(X_2) =\ldots = D(X_l)$

В качестве критерия проверки нулевой гипотезы примем критерий Кочрена - отношение максимальной исправленной дисперсии к сумме всех исправленных дисперсий:

$$G=S^2_{max}/(S^2_1+S^2_2+\ldots+S^2_l)$$

Распределение этой СВ зависит только от числа степеней свободы k и числа выборок *l*.

правосторонняя область

$P\,[G> G_{кр}(\alpha, k, l)] = \alpha,\ \ k=n-l$

Если $G_{набл} < G_{кр}(\alpha, k, l)$ нет оснований отвергнуть нулевую гипотезу

Если $G_{набл} > G_{кр}(\alpha, k, l)$ нулевая гипотеза отвергается

- Пример: По четырем независимым выборкам n=17 из нормально распределенной генеральной совокупности найдены исправленные выборочные дисперсии: 0,26, 0,36, 0,40, 0,42.

а) При уровне значимости 0,05 проверить нулевую гипотезу об однородности генеральных дисперсий

б) Оценить генеральную дисперсию

Решение:

$G_{набл}=0{,}42/(0{,}26+0{,}36+0{,}40+0{,}42)=0{,}2917$

$G_{крправ}(0{,}05, 16, 4)=0{,}4366$

$0{,}2917 < 0{,}4366$

нулевая гипотеза не отвергается - исправленные выборочные дисперсии различаются незначимо.

б)т.к. нулевая гипотеза принимается, в качестве оценки генеральной дисперсии примем среднее арифметическое исправленных дисперсий:

$\sigma^2 = (0,26 + 0,36 + 0,40 + 0,42)/4 = 0,36$

Сравнение нескольких генеральных дисперсий по независимым выборкам различного объема из нормальных совокупностей, критерий Бартлетта.

Пусть генеральные совокупности X_1, X_2,..., X_l распределены нормально. Из этих совокупностей извлечены l независимых выборок различного объема n_1, n_2, … n_l и по ним найдены исправленные выборочные дисперсии $s^2{}_1$, $s^2{}_{2,}$ … $s^2{}_l$.

Требуется по исправленным выборочным дисперсиям при заданном уровне значимости α проверить нулевую гипотезу, что генеральные дисперсии совокупностей равны между собой: H_0: $D(X_1) = D(X_2) =\ldots = D(X_l)$

(гипотеза об однородности дисперсий)

Число степеней свободы дисперсии s^2_i :

$k_i = n_i - 1$.

Обозначим $\overline{s}^2$ - среднюю арифметическую исправленных дисперсий, взвешенную по числам степеней свободы:

$$\overline{s}^2 = \frac{\left(\sum_{i=1}^{l} k_i s_i^2\right)}{k}$$

Критерий Бартлета: B=V/C, где

$$V = 2{,}303\left[k \lg \overline{s}^2 - \sum_{i=1}^{l} k_i \lg s_i^2\right]$$

$$C = 1 + \frac{1}{3(l-1)}\left[\sum_{i=1}^{l} \frac{1}{k_i} - \frac{1}{k}\right]$$

Бартлетт установил, что при условии справедливости нулевой гипотезы С.В. В распределена приближенно как χ^2 с l-1 степенями свободы, если все $k_i>2$, т.е. объем каждой выборки не меньше 4.

Критическую область строят правостороннюю:

$$P\,[B > \chi^2_{кр}(\alpha, l-1)] = \alpha$$

по таблице находим $\chi^2_{кр}(\alpha, l-1)$

Если $B_{набл} < \chi^2_{кр}$ - нет оснований отвергнуть нулевую гипотезу

Если $B_{набл} > \chi^2_{кр}$ нулевая гипотеза отвергается

Пример: по четырем независимым выборкам объемом $n_1=10$, $n_2=12$, $n_3=15$, $n_4=16$, извлеченных из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии: 0,25, 0,40, 0,36, 0,46.

При уровне значимости 0,05 проверить гипотезу об однородности дисперсий (критическая область-правосторонняя).

Решение:

$$\bar{s}^2 = \frac{\left(\sum_{i=1}^{l} k_i s_i^2\right)}{k} = 18{,}59/49 = 0{,}379$$

lg 0,379=-0,42

$V=2{,}303[49 \cdot (-0{,}42)-(-21{,}066)]=1{,}02$

По таблице находим $\chi^2_{кр}(0{,}05,\ 4\text{-}1)] = 7{,}8$

т.к. $V< \chi^2_{кр}$ $C=1{,}06>1$, то $B_{набл}=V/C < \chi^2_{кр}$

нет оснований отвергнуть нулевую гипотезу, исправленные дисперсии различаются незначимо

Критерий Левене (Levene)

$$W = \frac{(N-k)\cdot\sum_{i=1}^{k} N_i (Z_{i} - Z_{.....})^2}{(k-1)\cdot\sum_{i=1}^{k}\sum_{j=1}^{N_i}(Z_{ij} - Z_{i.})^2}$$

где:

W- критерий Левене

k- число различных групп

N- число случаев во всех группах

Ni- число случаев в i группе

Yij –значение j переменной в i группе

$$Z_{ij} = \left|Y_{ij} - \overline{Y}_i\right|$$

$\overline{Y}_i$ - средняя арифметическая i-й группы

Заключение

Нами рассмотрены:

Критерии проверки однородности дисперсий.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА:

Основная литература:

- Попов А.М. Теория вероятней и математическая статистика /А.М. Попов, В.Н. Сотников. – М.: ЮРАЙТ, 2011. – 440 с.
- Герасимов А. Н. Медицинская статистика: учебное пособие / А. Н. Герасимов. – М. : Мед. информ. агентство, 2007. – 480 с.
- Балдин К. В. Основы теории вероятностей и математической статистики : учебник / К. В. Балдин. – М. : Флинта, 2010. – 488с.

Учебно–методические пособия:

- Шапиро Л.А., Шилина Н.Г. Руководство к практическим занятиям по медицинской и биологической статистике Красноярск: ООО «Поликом». – 2003.

БЛАГОДАРЮ ЗА ВНИМАНИЕ