Критерии проверки статистических гипотез

1. Сравнение двух дисперсий нормальных генеральных совокупностей

Краткая теория:

По независимым выборкам, объемы которых n_1, n_2, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии s_X^2 и s_Y^2. Требуется сравнить эти дисперсии.

Правило 1. *Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу H_0: $D(X)=D(Y)$ о равенстве генеральных дисперсий нормальных совокупностей при конкурирующей гипотезе H_1: $D(X) > D(Y)$, надо вычислить наблюдаемое значение критерия (отношение большей исправленной дисперсии к меньшей)*

$$F_{\text{набл}} = s_{\text{Б}}^2 / s_{\text{М}}^2$$

и по таблице критических точек распределения Фишера—Снедекора, по заданному уровню значимости α и числам степеней свободы $k_1 = n_1 - 1$, $k_2 = n_2 - 1$ (k_1 — число степеней свободы большей исправленной дисперсии) найти критическую точку $F_{\text{кр}}(\alpha; k_1, k_2)$. Если $F_{\text{набл}} < F_{\text{кр}}$ — нет оснований отвергнуть нулевую гипотезу. Если $F_{\text{набл}} > F_{\text{кр}}$ — нулевую гипотезу отвергают.

Правило 2. *При конкурирующей гипотезе H_1: $D(X) \neq D(Y)$ критическую точку $F_{\text{кр}}(\alpha/2; k_1, k_2)$ ищут по уровню значимости $\alpha/2$ (вдвое меньшему заданного) и числам степеней свободы k_1 и k_2 (k_1 — число степеней свободы большей дисперсии). Если $F_{\text{набл}} < F_{\text{кр}}$ — нет оснований отвергнуть нулевую гипотезу. Если $F_{\text{набл}} > F_{\text{кр}}$ — нулевую гипотезу отвергают.*

Задания: сгенерировать две выборки из нормального распределения с произвольными значениями параметров μ и σ (на ваш выбор) объемом 100 и 150 элементов. Решить задачу с использованием Правила 1 и Правила 2 (см.выше), то есть подтвердить или опровергнуть основную гипотезу при уровне значимости 0,05.

2. Сравнение двух средних генеральных совокупностей, дисперсии которых известны (большие выборки)

Краткая теория:

Обозначим через n и m объемы больших ($n > 30$, $m > 30$) независимых выборок, по которым найдены соответствующие выборочные средние $\bar{x}$ и $\bar{y}$. Генеральные дисперсии $D(X)$ и $D(Y)$ известны.

Правило 1. *Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу H_0: $M(X)=M(Y)$ о равенстве математических ожиданий (генеральных средних) двух нормальных генеральных совокупностей с известными дисперсиями (в случае больших выборок) при конкурирующей гипотезе H_1: $M(X) \neq M(Y)$, надо* *вычислить наблюдаемое значение критерия*

$$Z_{набл} = \frac{\bar{x} - \bar{y}}{\sqrt{D(X)/n + D(Y)/m}}$$

и по таблице функции Лапласа найти критическую точку $z_{кр}$ из равенства

$$\Phi(z_{кр}) = (1 - \alpha)/2.$$

Если $|Z_{набл}| < z_{кр}$ — нет оснований отвергнуть нулевую гипотезу. Если $|Z_{набл}| > z_{кр}$ — нулевую гипотезу отвергают.

Правило 2. *При конкурирующей гипотезе H_1: $M(X) > M(Y)$ находят критическую точку $z_{кр}$ по таблице функции Лапласа из равенства*

$$\Phi(z_{кр}) = (1 - 2\alpha)/2.$$

Если $Z_{набл} < z_{кр}$ — нет оснований отвергнуть нулевую гипотезу. Если $Z_{набл} > z_{кр}$ — нулевую гипотезу отвергают.

Правило 3. *При конкурирующей гипотезе H_1: $M(X) < M(Y)$ находят «вспомогательную точку» $z_{кр}$ по правилу 2. Если $Z_{набл} > -z_{кр}$ — нет оснований отвергнуть нулевую гипотезу. Если $Z_{набл} < -z_{кр}$ — нулевую гипотезу отвергают.*

Задания: сгенерировать две выборки из нормального распределения с произвольными значениями параметров μ и σ (на ваш выбор) объемом 100 и 150 элементов. Решить задачу с использованием Правила 1 и Правила 2 и Правила 3 (см.выше), то есть подтвердить или опровергнуть основную гипотезу при уровне значимости 0,05.

3. Сравнение нескольких дисперсий нормальных генеральных совокупностей по выборкам различного объема. Критерий Бартлетта

Краткая теория:

Пусть генеральные совокупности $X_1, X_2, \ldots, X_l$ распределены нормально. Из этих совокупностей извлечены независимые выборки, вообще говоря, различных объемов n_i (некоторые объемы могут быть одинаковыми; если все выборки имеют одинаковый объем, то предпочтительнее пользоваться критерием Кочрена, который приведен в следующем параграфе). По выборкам найдены исправленные выборочные дисперсии $s_1^2, s_2^2, \ldots, s_l^2$. Требуется при уровне значимости α проверить нулевую гипотезу об однородности дисперсий, т. е. гипотезу о равенстве между собой генеральных дисперсий:

$$H_0:\ D(X_1) = D(X_2) = \ldots = D(X_l).$$

Введем обозначения: $k_i = n_i - 1$ — число степеней свободы дисперсии s_i^2; $k = \sum_{i=1}^{l} k_i$ — сумма чисел степеней свободы; $\overline{s^2} = \left(\sum_{i=1}^{l} k_i s_i^2\right)\Big/ k$ — средняя арифметическая исправленных дисперсий, взвешенная по числам степеней свободы;

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

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

$B = V/C$ — случайная величина (критерий Бартлетта), которая при условии справедливости гипотезы об однородности дисперсий распределена приближенно как χ^2 с $l-1$ степенями свободы, если объем каждой выборки $n_i \geqslant 4$.

Правило. *Для того чтобы при заданном уровне значимости α проверить нулевую гипотезу об однородности дисперсий нормальных совокупностей, надо вычислить наблюдаемое значение критерия Бартлетта $B_{набл} = V/C$ и по таблице критических точек распределения χ^2, по уровню значимости α и числу степеней свободы $l-1$ (l — число выборок) найти критическую точку $\chi^2_{кр}(\alpha;\ l-1)$ правосторонней критической области). Если $B_{набл} < \chi^2_{кр}$ — нет оснований отвергнуть нулевую гипотезу. Если $B_{набл} > \chi^2_{кр}$ — нулевую гипотезу отвергают.*

Замечание 1. Не следует торопиться вычислять постоянную C. Сначала надо найти V и сравнить с $\chi^2_{кр}$; если окажется, что $V < \chi^2_{кр}$, то подавно (так как $C > 1$) $B = V/C < \chi^2_{кр}$ и, следовательно, C вычислять не нужно. Если же $V > \chi^2_{кр}$, то надо вычислить C и затем сравнить B с $\chi^2_{кр}$.

Замечание 2. Критерий Бартлетта весьма чувствителен к отклонениям распределений от нормального, поэтому к выводам, полученным по этому критерию, надо относиться осторожно.

Замечание 3. При условии однородности дисперсий в качестве оценки генеральной дисперсии принимают среднюю арифметическую исправленных дисперсий, взвешенную по числам степеней свободы:

$$\overline{s^2} = \sum k_i s_i^2 / k.$$

Задания: сгенерировать пять выборок из нормального распределения так, чтобы параметр σ у них был одинаковый, а параметр μ выберете произвольно, но так, чтобы для каждой выборки было свое значение. Объемы выборок 50, 60, 70, 80, 90. Подтвердить или опровергнуть основную гипотезу при уровне значимости 0,05.