3. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

В прикладных задачах часто возникает необходимость проверки суждений о генеральной совокупности на основе опытных данных (на основе выборки). Например, распространен вопрос о наличии *эффекта обработки*: действительно ли смена технологии позволяет улучшить качество продукции (предполагается, что сформулирован какой-либо критерий качества) и/или уменьшить разброс (коэффициент вариации) показателя качества. Если данные, на основе которых решаются подобные вопросы, получают в результате выборочного обследования, то возможен только вероятностный ответ на вопрос. В математической статистике разрабатываются методы, которые позволяют оценить соответствующие вероятности.

Статистической гипотезой H называется утверждение о виде неизвестного распределения или о параметрах известного распределения генеральной совокупности, подлежащее проверке на основе выборки (в первом случае гипотезу называют непараметрической, во втором – параметрической). *Конкурирующей* (*альтернативной*) гипотезой H_1 называют гипотезу, противоположную гипотезе H_0. При этом гипотезу H_0 называют *исходной* (*нулевой*).

Большинство статистических гипотез можно сформулировать в одной из двух форм:

1) выборка извлечена из генеральной совокупности, подчиненной указанному распределению;

2) выборки извлечены из генеральных совокупностей с равными числовыми характеристиками (математическими ожиданиями, дисперсиями).

Проверка статистической гипотезы H_0 – это выяснение того, насколько гипотеза согласуется с опытными данными (с выборкой). Содержание операции проверки гипотезы состоит в *неформальном* выборе связанного с выборкой $\mathbf{x}$ события A и последующем *формальном* отыскании условной вероятности $P(A|H)$ при проверяемой гипотезе H. Событие A принято выбирать так, чтобы его вероятность $P(A|H)$ оказалась малой; в этом случае его называют *критическим событием* для гипотезы H. *Статистическим критерием*, или *статистикой*, называют случайную величину с известным распределением, которая служит для проверки гипотезы.

Если вероятность $P(A|H)$ *реально наблюдаемого* в опыте критического события A оказывается меньше некоторого заранее заданного *уровня значимости* α, то *гипотеза H отвергается на уровне значимости* α.

Выбор уровня значимости является *выбором вероятности события, которое на практике считается невозможным*. В большинстве поисковых исследований уровень значимости выбирают равным $\alpha = 0{,}05$. Критические задачи (например, связанные с оценкой надежности транспортных средств) требуют выбора существенно меньших значений: $\alpha \sim 10^{-6}$. Однако при столь малых уровнях значимости большинство методов математической статистики становятся непригодными.

Уровень значимости – это вероятность того, что *верная гипотеза будет ошибочно отвергнута*. Соответствующую ошибку, состоящую в непринятии правильной на самом деле гипотезы, называют *ошибкой первого рода*.

Ошибкой второго рода называют ошибку, состоящую в том, что принимается альтернатива для истинной нулевой гипотезы. Вероятность ошибки второго рода дополняет до единицы число, называемое *мощностью* статистического критерия. Обычно одна и та же гипотеза может быть проверена при помощи различных критериев. Среди этих критериев следует по возможности выбирать тот, который обладает наибольшей мощностью: *значения критерия при нулевой гипотезе и альтернативе должны отличаться как можно больше*.

В простейших задачах критическое событие A можно отождествить с непоявлением наблюдения **x**. Если для подобной формулировки удается отыскать вероятность $P(A|H)$, то говорят, что гипотеза проверяется *непосредственно*. Во всех остальных случаях для косвенной проверки гипотезы привлекаются статистические критерии, выбор которых обычно отражает компромисс между сложностью вычисления критерия и необходимым объемом экспериментальных данных.

Мощность критерия – не единственная характеристика его пригодности. Наряду с мощностью, для описания критериев привлекают состоятельность и несмещенность. Критерий называют *состоятельным*, если при увеличении объема выборки вероятность отвергнуть нулевую гипотезу при верной альтернативе стремится к единице. Критерий называют *несмещенным*, если вероятность отвергнуть верную гипотезу меньше вероятности отвергнуть ее альтернативу.

Пример: проверка статистической гипотезы о симметрии монеты.

В теории доказывается неравенство, определяющее вероятность заданного отклонения частоты события (относительного числа его появлений) от вероятности этого события в каждом опыте. Именно, если в каждом из N опытов вероятность наступления события постоянна и равна p, то вероятность отклонения частоты события от p (в любую сторону) на величину ε равна

$$P\left(\left|\frac{m}{N}-p\right|\geq\varepsilon\right)\leq\frac{p(1-p)}{N\varepsilon^2},$$

где m – число появлений события в серии из N опытов.

Статистическая гипотеза H_0 – «монета симметрична» тождественна гипотезе H_0 – «вероятности появления герба и цифры равны между собой и равны 1/2». Пусть после $N=1000$ бросаний монеты число выпадений цифры составило $m=600$. Найденное на опыте отклонение вероятности от частоты составляет:

$$\left|\frac{600}{1000}-\frac{1}{2}\right|=0{,}1.$$

Найдем вероятность критического события A, состоящего в том, что при верной гипотезе H_0 отклонение вероятности от частоты («маловероятное» для верной гипотезы!) окажется не меньшим, чем в проведенном эксперименте:

$$P(A\,|\,H_0) = P\left(\left|\frac{m}{N} - p\right| \geq 0{,}1\right) \leq \frac{1/2(1-1/2)}{1000 \cdot 0{,}1^2} = 0{,}025\,.$$

Таким образом, если гипотеза о симметрии монеты верна, то *на практике реально осуществилось событие, вероятность которого не превышает 0,025*. Эта вероятность весьма мала (в сравнении, например, с уровнем значимости $\alpha = 0{,}05$), и на этой основе гипотеза о симметрии монеты должна быть отвергнута (на уровнях значимости $\alpha > 0{,}025$). В то же время, на уровне значимости $\alpha = 0{,}01$ экспериментальные результаты не противоречат гипотезе о симметрии монеты: для того, чтобы сделать обоснованный вывод для меньшего уровня значимости, нужно или увеличивать число испытаний, или использовать иное критическое событие (использовать статистических критерий большей мощности).

Нетрудно заметить, что при том же отклонении частоты вероятность критического события становится равной 0,05 при количестве повторений, равном 500. Это значение позволяет составить представление об объемах выборки, требуемых для суждений на уровне значимости 0,05 (в действительности для рассматриваемой задачи можно использовать иные критерии и сократить требуемый объем экспериментальных исследований примерно на порядок).

3.1 Гипотеза о нормальном распределении

Одной из наиболее распространенных *одновыборочных* задач является проверка гипотезы о нормальном распределении генеральной совокупности. Эта задача возникает, в частности, при проверке выполнения необходимых условий возможности применения других статистических методов обработки данных.

При использовании χ^2-*статистики* после построения непрерывного вариационного ряда вычисляется значение статистики (случайной величины, связанной с эмпирическими данными)

$$\chi^2 = \frac{1}{n}\sum_{j=1}^{l} \frac{(n_j - np_j)^2}{p_j}\,, \tag{3.1}$$

где n – объем выборки (как правило – не менее 200); l – число разрядов непрерывного вариационного ряда (не менее 8); n_j – частота; p_j – вероятность, найденная расчетом по нормальной кривой, выравнивающей выборку:

$$p_j = \frac{1}{s\sqrt{2\pi}} \int_{x_j}^{x_{j+1}} e^{-\frac{(x-\bar{x})^2}{2s^2}}\,dx = \Phi\left(\frac{x_{j+1} - \bar{x}}{s}\right) - \Phi\left(\frac{x_j - \bar{x}}{s}\right),\ j = \overline{1,l}\,. \tag{3.2}$$

В последнем соотношении $\bar{x}$ – оценка математического ожидания, s – оценка среднего квадратичного отклонения, $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ – функция Лапласа (или функция стандартного нормального распределения).

Статистика (3.1) подчинена χ^2-распределению с числом степеней свободы $l-3$. С учетом этого ищется вероятность критического события, состоящего в том, что для выборки из нормально распределенной генеральной совокупности истинное (неизвестное) значение статистики окажется столь же большим (т.е., большим или равным), как и наблюдаемое на опыте значение. Если указанная вероятность близка к нулю (меньше выбранного уровня значимости α), то нулевая гипотеза о нормальном распределении генеральной совокупности отвергается.

Вместо нахождения вероятности критического события можно сравнить найденное значение статистики (3.1) с квантилью χ^2-распределения $\chi^2_{l-3,\alpha}$ для $l-3$ степеней свободы и выбранного уровня значимости α. При выполнении неравенства

$$\chi^2 < \chi^2_{l-3,\alpha}$$

считают, что на уровне значимости α нет оснований отвергать нулевую гипотезу о нормальном распределении генеральной совокупности (вновь отметим, что истинность гипотезы этим не доказывается).

3.2 Проверка статистических гипотез: основные двухвыборочные задачи

На практике часто встречается случай, когда средний результат одной серии экспериментов отличается от среднего результата другой серии. При этом возникает вопрос, является обнаруженное расхождение средних статистически значимым – можно ли объяснить его случайными ошибками или же оно имеет закономерное происхождение. В промышленности такая задача сравнения возникает при контроле качества продукции, изготовленной при различных технологических режимах.

Значимость различия средних (оценок неизвестных математических ожиданий) зависит от генеральных дисперсий – малость различия в сравнении со стандартным отклонением указывает на его незначимость. Если математические ожидания (генеральные средние) оцениваются по результатам эксперимента (заменяются выборочными средними), то различие между средними может быть значимым даже в том случае, если оно незначительно по сравнению со стандартным отклонением (это имеет место для больших выборок). Поэтому «количественной характеристикой» различия между средними является стандартная ошибка – частное от деления оценки стандартного отклонения на корень из объема выборки.

Задача сравнения математических ожиданий решается различно в зависимости от того, являются ли известными дисперсии двух генеральных совокупностей, из которых извлечены выборки. Если дисперсии неизвестны, то способ решения зависит от того, считаются ли дисперсии равными. Наиболее просто задача сравнения математических ожиданий $M[X]=\bar{x}_0$ и $M[Y]=\bar{y}_0$ решается в первом случае.

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