Государственный Медицинский Университет г.Семей
Кафедра общественного здравоохранения
по дисциплине “Биостатистика”.

СРС:

Критерий согласия. Практический пример применения критерия согласия. Закон Менделя.

ВЫПОЛНИЛ: ЕСЕНТАЕВ Н.А.
217ГР.ОМФ
ПРОВЕРИЛА: К.М.Н. ГОРЕМЫКИНА М.В.

Семей 2017г.

План:

- Введение;
- Закон распределения;
- Критерии согласия:
1) Пирсона;
2) Колмогорова;
3) Романовского;
- Практическое применение критерия согласия;
- Заключение;
- Список литературы.

Введение:

- **Статистический критерий** — строгое математическое правило, по которому принимается или отвергается та или иная статистическая гипотеза с известным уровнем значимости. Построение критерия представляет собой выбор подходящей функции от результатов наблюдений (ряда эмпирически полученных значений признака), которая служит для выявления меры расхождения между эмпирическими значениями и гипотетическими. Теоретическая модель в этом случае описывает закон распределения.
- **Теоретическое распределение** – это то распределение вероятностей, которое управляет случайным выбором [1].

Закон распределения

- Самая известная статистическо-вероятностная модель – это **закон нормального распределения**. Нормальный закон, как и другие теоретические распределения, не является фиксированным уравнением зависимости одной переменной от другой. Фиксируется только характер этой зависимости.
- Важное значение нормального распределения во многих областях науки вытекает из центральной предельной теоремы теории вероятностей. Если результат наблюдения является суммой многих случайных слабо взаимозависимых величин, каждая из которых вносит малый вклад относительно общей суммы, то при увеличении числа слагаемых распределение центрированного и нормированного результата стремится к нормальному. Этот закон теории вероятностей имеет следствием широкое распространение нормального распределения, что и стало одной из причин его наименования[2].

- Говоря о теоретическом законе распределения, которому гипотетически должны бы следовать элементы данной выборки, надо различать простые и сложные гипотезы об этом законе:

1) простая гипотеза - прямо указывает некий определенный закон вероятностей (распределение вероятностей), по которому возникли выборочные значения;
2) сложная гипотеза - указывает на единственное распределение, а какое-то их множество (например, параметрическое семейство)[3].

Критерий согласия

- Критерием согласия называют критерий, который позволяет установить, является ли расхождение эмпирического и теоретического распределений случайным или значимым, т. е. согласуются ли данные наблюдений с выдвинутой статистической гипотезой или не согласуются. Распределение генеральной совокупности, которое она имеет в силу выдвинутой гипотезы, называют теоретическим.
- Возникает необходимость установить критерий (правило), которое позволяло бы судить, является ли расхождение между эмпирическим и теоретическим распределениями случайным или значимым. Если расхождение окажется случайным, то считают, что данные наблюдений (выборки) согласуются с выдвинутой гипотезой о законе распределения генеральной совокупности и, следовательно, гипотезу принимают; если же расхождение окажется значимым, то данные наблюдений не согласуются с гипотезой и ее отвергают [4].

- Обычно эмпирические и теоретические частоты различаются в силу того, что:

1) расхождение случайно и связано с ограниченным количеством наблюдений;
2) расхождение неслучайно и объясняется тем, что статистическая гипотеза о том, что генеральная совокупность распределена нормально — ошибочна [4].

- Таким образом, критерии согласия позволяют отвергнуть или подтвердить правильность выдвинутой при выравнивании ряда гипотезы о характере распределения в эмпирическом ряду. Эмпирические частоты получают в результате наблюдения. Теоретические частоты рассчитывают по формулам[4].
- Для закона нормального распределения их можно найти следующим образом:

$$f_T = \varphi(t) \cdot \frac{n \cdot h}{\sigma} \rightarrow n = \sum f_i$$

$$\varphi(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{t^2}{2}} \rightarrow \pi \approx 3.1415,\ e \approx 2.7182$$

$$t = \frac{x_i - \bar{x}}{\sigma} \qquad \sigma = \sqrt{\frac{(x_i - \bar{x})^2 \cdot f_i}{\sum f_i}}$$

$$h(i) = x_{\max} - x_{\min}$$

- Имеется несколько критериев согласия, наиболее распространенными из которых являются: критерий Пирсона, критерий Колмогорова, критерий Романовского.

Критерий согласия Пирсона

- Критерий согласия Пирсона, или критерий согласия (Хи-квадрат) — наиболее часто употребляемый критерий для проверки гипотезы о принадлежности наблюдаемой выборки X_1. $X_{2,}$ $_{...,}$ X_n объёмом n некоторому теоретическому закону распределения F(x,θ). Свойства критерия были впервые исследованы Карлом Пирсоном в 1900 году[5].

Критерий согласия Пирсона

Критерий согласия Пирсона X^2 – один из основных, который можно представить как сумму отношений квадратов расхождений между теоретическими (f_T) и эмпирическими (f) частотами к теоретическим частотам:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - f_T)^2}{f_T}$$

- k–число групп, на которые разбито эмпирическое распределение,
- fi–наблюдаемая частота признака в i-й группе,
- fT–теоретическая частота.

- При проверке сложных гипотез, если параметры закона $F(x,\theta)$ по этой же выборке оцениваются в результате минимизации статистики X^2_n или по сгруппированной выборке методом максимального правдоподобия, то статистика X^2_n при справедливости проверяемой гипотезы подчиняется X^2_r-распределению с $r = k-m-1$ степенями свободы, где m — количество оцененных по выборке параметров.
- Если параметры оцениваются по исходной негруппированной выборке, то распределение статистики не будет являться X^2_{k-m-1} -распределением. Более того, распределения статистики при справедливости гипотезы H_0 будут зависеть от способа группирования, то есть от того, как область определения разбивается на интервалы.
- При оценивании методом максимального правдоподобия параметров по негруппированной выборке можно воспользоваться модифицированными критериями типа [6].

Критерий согласия Колмогорова

- **Критерий согласия Колмогорова** предназначен для проверки гипотезы о принадлежности выборки некоторому закону распределения, то есть проверки того, что эмпирическое распределение соответствует предполагаемой модели[7].
- Пусть эмпирическая функция распределения (ЭФР) $\boldsymbol{F_n}$ случайной величины $\boldsymbol{\xi}$, построенная по выборке $\boldsymbol{X = (X_{1,} \ldots, X_n)}$ $\boldsymbol{(X_i\ X)}$, имеет вид:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{X_i \leqslant x},$$

- Статистика критерия для эмпирической функции распределения $\boldsymbol{F_n(x)}$ определяется следующим образом:

$$D_n = \sup_x |F_n(x) - F(x)|,$$

- **Критерий Романовского** с основан на использовании критерия Пирсона, т.е. уже найденных значений, и числа степеней свободы df:

$$c = \frac{\left|\chi^2 - df\right|}{\sqrt{2df}}$$

- Он удобен при отсутствии таблиц для.
- Если c<3, то расхождения распределений случайны, если же c>3, то не случайны и теоретическое распределение не может служить моделью для изучаемого эмпирического распределения[7].

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ КРИТЕРИЯ СОГЛАСИЯ

- Все рассмотренные до сих пор критерии принято относить к группе так называемых параметрических критериев. Применение этих критериев требует знания типа распределения наблюдаемых случайных величин (нормальное, биномиальное, пуассоновское, двумерное нормальное или какое-либо иное) и проверяемая гипотеза касается параметров данных распределений. Прежде чем применять параметрические методы, необходимо убедиться в том, что мы действительно имеем дело с распределением требуемого типа.
- Предположение о виде распределения случайной величины – это статистическая гипотеза, которую можно проверить с помощью экспериментальных данных. Критерии, позволяющие решать такого рода задачи, называются критериями согласия – согласия выборочных данных некоторому наперед заданному теоретическому распределению.
- При проверке гипотезы о нормальности распределения с неизвестными средним и дисперсией критерий Колмогорова-Смирнова является более мощным, чем критерий χ^2 [8].

- При проведении данных исследований, в которых реализован ряд критериев проверки согласия эмпирического распределения с теоретической моделью: χ^2 Пирсона, отношения правдоподобия, Колмогорова, Смирнова и Мизеса, Никулина. Здесь и ниже, когда мы употребляем словосочетание “хорошее согласие”, то подразумеваем, что по всем критериям достигнутый уровень значимости, определяемый соотношением:

$$P\{S > S^*\} = \int_{S^*}^{\infty} g(s|H_0)ds$$

- где S^* - значение статистики критерия, вычисленное по наблюдаемой выборке, $g(s|H_0)$ - плотность предельного распределения статистики соответствующего критерия при справедливости гипотезы, H_0 был очень высок:

$$P\{S > S^*\}$$

- Зная предельные распределения $G(S \mid H_0)$ и $G(S \mid H_1)$ статистики S, для любого заданного уровня значимости α можно оценить мощность соответствующего критерия, рассматривая её как функцию от числа интервалов $x(\alpha, r)$ при заданном объеме выборки . Было проведено исследование мощности критериев Пирсона и Никулина как функции от n и $x(\alpha, r)$ аналитически и методами статистического моделирования. Причем результаты аналитических вычислений оказались полностью подтвержденными оценками мощности, полученными на основании моделирования[8].

- Пример. В некоторых классических экспериментах с селекцией гороха Мендель наблюдал частоты различных видов семян, получаемых при скрещивании растений с круглыми желтыми семенами и растений с морщинистыми зелеными семенами. Они приводятся ниже вместе с теоретическими вероятностями, вычисленными в соответствии с теорией наследственности Менделя.
- В этом случае теоретическое распределение дискретно и известно полностью. Для проверки согласия экспериментальных данных теоретическому распределению используем критерий χ^2 для простой гипотезы. Значение статистики, вычисленное по выборке равно:

$$\sum_{i=1}^{4} \frac{(n_i - np_i)^2}{np_i} = 0.47$$

- что меньше 5%-ного критического значения:

$$\chi^2_{3;0.95} = 7.81$$

- Следовательно, теория наследственности Менделя не противоречит полученным экспериментальным данным[8].

- Наряду с количественными статистическими критериями для определения типа распределения по выборочным данным используются графические методы.

- Простейший способ – построение по имеющейся выборке гистограммы относительных частот и на том же графике и в том же масштабе, кривой плотности нормального распределения с выборочным средним и выборочной дисперсией в качестве параметров. Значительные отклонения от нормальности (сильная асимметрия, бимодальность) легко обнаруживаются на графике[8].

Заключение:

- Критерии согласия основаны на использовании различных мер расстояния между анализируемым эмпирическим распределением и функцией распределения признака в генеральной совокупности. Критериями согласия называют статистические критерии, предназначенные для проверки согласия опытных данных и теоретической модели.
- Существует несколько критерий согласия: критерий согласия Колмогорова и омега-квадрат, χ^2 Пирсона, χ^2 Фишера и другие. Состоятельность критериев Колмогорова и омега-квадрат означает, что любое отличие распределения выборки от теоретического будет с их помощью обнаружено, если наблюдения будут продолжаться достаточно долго. Практическую значимость свойства состоятельности не велика, так как трудно рассчитывать на получение большого числа наблюдений в неизменных условиях, а теоретическое представление о законе распределения, которому должна подчиняться выборка, всегда приближённое. Поэтому точность статистических проверок не должна превышать точность выбранной модели[8].

Список литературы:

1. Рекомендации по стандартизации. Прикладная статистика: Правила проверки согласия опытного распределения с теоретическим. Часть II: Непараметрические критерии. — М.: Госстандарт РФ, 2002.;
2. Вентцель Е. С. Теория вероятностей. — 10-е изд., стер.. — М.: Academia, 2005.
3. Ширяев, А. Н. Вероятность. — М.: Наука, 1980.
4. Кендалл М., Стьюарт А. Статистические выводы и связи. — М.: Наука, 1973.
5. ''Pearson, Karl'' (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine Series 5 50 (302): 157—175.
6. Лемешко Б. Ю., Постовалов С. Н. О зависимости предельных распределений статистик X^2 Пирсона и отношения правдоподобия от способа группирования данных // Заводская лаборатория. 1998. Т. 64. — № 5. — С. 56-63.
7. Критерий Колмогорова на сайте Новосибирского государственного университета.
8. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. — М.: Изд-во стандартов. 2002. — 87 с.