

Лабораторная работа №7

Анализ выборки. Критерий согласия Пирсона.

К.С. Пилипенко

2022

Критерий согласия Пирсона (Хи-квадрат) был придуман для проверки значимости расхождения эмпирических (наблюдаемых) и теоретических (ожидаемых) частот. Выражается следующей формулой:

$$\chi^2 = \sum_i^n \frac{(O_i - E_i)^2}{E_i}, \quad (1)$$

где O_i — наблюдаемые частоты (Observed), E_i — ожидаемые частоты (Expected).

Полученное значение χ^2 сравнивают с теоретически рассчитанным критическим значением $\chi_{\text{кр.}}^2$, которое зависит от значения доверительной вероятности (как правило принимается равным 95%) и числа степеней свободы k , которое на один меньше количества уникальных значений в выборке ($k = N - 1$).

Для расчета критического значения критерия $\chi_{\text{кр.}}^2$ можно воспользоваться специальной таблицей, но лучше и проще всего воспользоваться функцией `ХИ2.ОБР.ПХ`

Напомню, что в большинстве случаев распределение дискретной случайной величины (N) подчиняется формуле Пуассона:

$$P(N) = \frac{\bar{N}^N e^{-\bar{N}}}{N!}, \quad (2)$$

где \bar{N} — среднее значение. Для расчета этой вероятности можно вос-

пользовать функцией ПУАССОН.РАСП (x ; среднее; интегральная), где x — значение, для которого строится распределение, среднее — среднее арифметическое распределения, интегральная — логическое значение, определяющее форму функции. Если ИСТИНА функция возвращает вероятность того, что количество событий будет меньше, чем N то есть $P(n < N)$, если ЛОЖЬ то, возвращается значение функции взвешенной вероятности, то есть вероятность того, что количество происходящих событий будет равно N раз. Все аргументы являются обязательными.

Ход работы

Задание №1. Проверить выборку на соответствие распределению Пуассона

- | | N | O_i |
|---|----|-------|
| | 1 | 2 |
| | 2 | 2 |
| | 3 | 4 |
| | 4 | 8 |
| • Выделите текст таблицы справа, скопируйте и вставьте в пустое поле таблицы Excel; | 5 | 12 |
| | 6 | 22 |
| • Далее, кликнув по первому столбцу вставленного текста, переходим во вкладку «Данные». Там, в группе инструментов «Работа с данными» кликаем по кнопке «Текст по столбцам»; | 7 | 35 |
| | 8 | 60 |
| • В открывшемся окне нажимаем «Далее», в списке «Символов-разделителей» кликаем на чекбокс «пробел», далее нажимаем «Готово». | 9 | 75 |
| | 10 | 77 |
| | 11 | 80 |
| • Следующий шаг - получить частоты распределения Пуассона (E_i). Для этого необходимо воспользоваться функцией ПУАССОН.РАСП. Нужно рассчитать вероятность (P_i) для каждого N и получить столбец вероятностей. Остаётся открытым вопрос, откуда брать среднее (\bar{N})? Его можно подобрать вручную построив два графика в одних координатах $E_i(N)$ и $O_i(N)$, и добившись их наилучшего соответствия; | 12 | 65 |
| | 13 | 53 |
| | 14 | 43 |
| | 15 | 25 |
| | 16 | 23 |
| | 17 | 16 |
| | 18 | 12 |
- Чтобы найти ожидаемые (теоретические) частоты E_i нужно умножить соответствующие вероятности (P_i) на объём выборки, то есть на $\sum_{i=1}^{18} O_i$. Ож. частоты E_i должны быть целыми ненулевыми числами, поэтому стоит использовать функцию ОКРУГЛ;

- Используя формулу 1 посчитать критерий Пирсона (увы, функции Excel для этой формулы нет);
- Используя функцию ХИ2 . РАСП постройте график плотности распределения $\omega(\chi^2_{кр.})$ (последний параметр должен быть ЛОЖЬ);
- В отдельном столбце посчитать критическое значение пользуясь функцией ХИ2 . ОБР . ПХ. Указать это значение на графике $\omega(\chi^2_{кр.})$. Сравнить полученное значение с экспериментальным;
- Получить p-value используя функцию ХИ2 . РАСП . ПХ для посчитанного χ^2

Контрольные вопросы

1. Сформулируйте нулевую гипотезу H_0 . Назовите условие, при котором можно отклонить нулевую гипотезу.
2. Вывести дисперсию распределение Пуассона ($D[N]$).
3. Что такое p-value (p-значение)? Чему численно равно это значение?
4. В каких случаях используется функция ХИ2 . ОБР? Что она позволяет оценить