

p -value: то, что вы всегда хотели узнать, но боялись спросить

МАРИЯ ЛЫСЮК

Аннотация. Аннотация должна передавать краткое содержание работы. Она должна быть ясной, содержательной, релевантной и короткой (не более 150 слов). Аннотация должна содержать информацию, необходимую для поиска по базам научных работ. В аннотации не должно быть математических формул.

Этот файл является образцом. Сравните его исходный код с финальным PDF-файлом, чтобы получить представление о том, как написать статью по данному шаблону.

Ключевые слова: p -value, уровень значимости, гипотезы, интерпретация.

С завидным постоянством хотя бы раз в жизни студент, слушающий курс статистики, сталкивается с вопросом экзаменатора (который обычно ещё надеется, что вопрос очевиден и «вытягивает» с помощью него студента): «Мистер X , что показывает p -value?»

И тут для многих наступает этот неловкий момент, и лицо выглядит примерно вот так (Эдвард Мунк, видимо, тоже не знал):



Рис. 1. Лицо обычного человека, у которого спросили, что такое p -value

Для того, чтобы осознать сие, безусловно, великое понятие, мы должны, как Будда, пройти 7 ступеней познания. Как водится, примеры красноречивее всего доносят нужную информацию до мозга, так что поговорим сегодня про машинки:)

Вкратце о ходе эксперимента. Мы будем узнавать, есть ли какая-то зависимость между тем, зависит ли лихачество водителей от цвета их машины. И гипотеза H_0 будет выглядеть следующим образом:

H_0 : Выдача штрафа не зависит от цвета машины.

H_1 : Водители с красными машинами чаще получают штрафы за превышение скорости по сравнению с синими машинами.

Итак, в добрый путь!

Ступень 1. Выберите уровень значимости. Начнем со знакомого до боли:) Строго говоря, уровень значимости- это мера, которая отражает наше желание относительно точности результатов- низкие уровни значимости говорят о маленькой вероятности того, что полученные экспериментальным путем результаты- случайны, и наоборот. Согласно конвенции, как правило, используется 5% уровень значимости. Это означает, что вероятность того, что наши результаты случайны- 0.05, и 0.95, что мы сами повлияли на результат.

- Пример: Возьмем и мы уровень значимости в 5% :)

Ступень 2. Определите ожидаемые результаты эксперимента. Как правило, ученые, проводя эксперимент и наблюдая впоследствии результаты, имеют представление о том, какие результаты являются «типичными» до начала эксперимента. Это может быть основано на результатах прошлых исследований, достоверных источников, научной литературы, и т.д. Для вашего эксперимента определите ваши ожидаемые результаты любым из способов.

- Пример: Пусть предыдущие исследования показали, что штрафы за превышение скорости чаще получают водители красных машин по сравнению с синими. Также пусть результаты по всей стране показывают превышение красными в отношении 2:1 в сравнении с синими. Мы же хотим узнать, применимы ли результаты, характерные ко всей стране, к нашему городу. Если мы возьмем рандомную выборку из 150 машинок, которым выписали штрафы, мы будем ожидать, что 100 машин будут красными, а 50 синими, *если наша полиция выписывает штрафы согласно национальной тенденции.*

Ожидаемое значение

Красная машинка	Синяя машинка
100	50

Ступень 3. Определите наблюдаемые результаты эксперимента.

После того как мы определили ожидаемые результаты, проводим реальный эксперимент и получаем наблюдаемые результаты. Если мы каким-либо образом повлияли и наблюдаемые результаты *отличаются от ожидаемых*, возможны 2 ситуации:

1) это произошло случайно

2) или те условия, в которых мы проводили эксперимент, повлияли на исход.

Как правило, цель нахождения р-значения- определить, правда ли что наблюдаемые результаты отличаются от ожидаемых настолько, что мы не можем отвергнуть нулевую гипотезу (гипотезу о том, что нет связи между переменными и наблюдаемым результатом).

- Пример: Пусть уже в нашем городе мы произвольно выбрали 150 красных и синих машин-нарушителей. Оказалось, что 90 штрафов выписали красным машинам и 60 голубым. Это отличается от ожидаемых 100 и 50 соответственно. Правда ли, что те условия, в которых мы проводили эксперимент (в нашем случае смена источника данных с национальных на местные) послужила причиной изменения результатов, или действия городской полиции также смещены, как предсказывает национальная средняя оценка, и мы просто наблюдаем случайную вариацию? Р-значение спешит на помощь!

Наблюдаемые результаты:

Красная машинка	Синяя машинка
90	60

Ступень 4. Определите степени свободы в вашем эксперименте. Степени свободы отражают меру изменчивости, характерную для исследования, которая определяется количеством переменных, которые вы изучаете. Степени свободы определяются как $n - 1$, где n - это количество переменных, используемых в эксперименте.

- Пример: У нас есть две переменных: красные машины и синие машины. Поэтому количество степеней свободы равно $2 - 1 = 1$.

Ступень 5. Сравните наблюдаемые результаты с ожидаемыми с помощью χ^2 . χ^2 - число, измеряющее разницу между ожидаемыми и наблюдаемыми результатами. Уравнение:

$$\chi^2 = \sum_{i=0}^n \frac{(h_i - e_i)^2}{e_i},$$

где h -значение наблюдаемой переменной, а e - ожидаемой.

- Пример: Мы должны просуммировать значения для всех возможных переменных, то есть в нашем случае для синих и красных машинок.



$$\begin{aligned}
 \chi^2 &= \sum_{i=0}^1 \frac{(h_i - e_i)^2}{e_i} \\
 \chi^2 &= \frac{(90 - 100)^2}{100} + \frac{(60 - 50)^2}{50} \\
 &= \frac{(-10)^2}{100} + \frac{(10)^2}{50} = \frac{100}{100} + \frac{100}{50} \\
 &= 1 + 2 = \boxed{3}
 \end{aligned}$$

Ступень 6. Используем таблицу χ^2 -распределения, чтобы аппроксимировать значение p-value . Скрестила пальцы: надеюсь, что все умеют пользоваться таблицами распределений:)

- Пример: Наше значение $\chi^2 = 3$. Далее пользуемся таблицей для нахождения значения p-значения. У нас одна степень свободы, берем эту строку, и ищем там первое значение, превышающее значение нашего $\chi^2 = 3$. Итак, первое это 3,84. Соответствующее значение p-значения равно 0,05. Это означает, что наше p-value располагается в границах между 0.05 и 0.1.

	p value											
df	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
	25%	20%	15%	10%	5%	2.5%	2%	1%	0.05%	0.025%	0.01%	0.005%
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.6	11.98	13.82	15.2
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.23	14.86	16.42	18.47	20
5	6.63	7.29	8.12	9.24	11.07	12.83	13.33	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.53	14.45	15.03	16.81	18.55	20.25	22.46	24.1
7	9.04	9.8	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.63	21.67	23.59	25.46	27.83	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42

Ступень 7. Вот мы и добрались до конца! Осталось решить, отвергается или нет нулевая гипотеза. Если значение р-значения меньше, чем уровень значимости, то мои поздравления (можете отсылать вашу работу в топовые журналы!), вы доказали, что вероятность того, что есть высокая корреляция между переменными, которыми вы манипулируете, и наблюдаемыми результатами, высока. Если ли же значение р-значение больше выбранного уровня значимости, вы не можете с точностью сказать, что полученные вами результаты случайны или являются результатом ваших действий.

- Пример: Наше значение р-значение находится в границах от 0.05 до 0.1. Это означает, что это определенно меньше, чем выбранный уровень значимости, равный 0,05, поэтому, к сожалению, мы не можем отвергнуть нулевую гипотезу. Другими словами, мы не достигли желаемого уровня в 95%, чтобы мы могли с точностью сказать, что в нашем городе полиция выдает штрафы красным и синим машинам в пропорции, значительно отличающейся от национального уровня. Иначе говоря, есть 5-10 % вероятность того, что изменения в выдаче штрафов красным и синим машинам связано не со сменой локации, а по чистой случайности. В виду того, что мы ищем вероятность, меньшую чем 0.05, то мы не можем быть *уверны, что полиция нашего города более склонна выдавать штрафы красным машинам, есть маленькая, но статистически значимая вероятность того, что это не так.*

