

Прикладная статистика. Занятие 4. Критерии согласия.
Анализ таблиц сопряжённости.

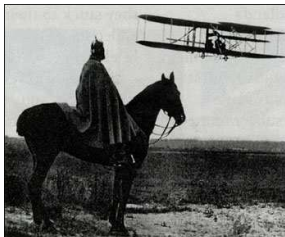
6 марта 2012 г.

Гибель кавалеристов

Bortkiewicz L., Das Gesetz der kleinen Zahlen, Leipzig, 1898: в Генеральный штаб прусской армии ежегодно в течение 20 лет поступали от 10 кавалерийских корпусов данные о количестве кавалеристов, погибших в результате гибели под ними коня (всего 200 донесений).

Количество погибших	0	1	2	3	4	5	6	Всего
Количество донесений	109	65	22	3	1	0	0	200

Можно ли описать эти данные распределением Пуассона?



Гибель кавалеристов

Распределение Пуассона:

$$\mathbb{P}(Y = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Оценим параметр λ методом максимального правдоподобия:

$$L(\lambda, Y) = e^{-n\lambda} \prod_{i=1}^n \frac{\lambda^{Y_i}}{Y_i!},$$

$$LL(\lambda, Y) = -n\lambda + \sum_{i=1}^n Y_i \ln \lambda - \sum_{i=1}^n \ln Y_i!,$$

$$\frac{\partial LL}{\partial \lambda} = -n + \sum_{i=1}^n \frac{Y_i}{\lambda},$$

$$\hat{\lambda} = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \text{ — оценка максимального правдоподобия.}$$

$$\bar{Y} = 0.61.$$

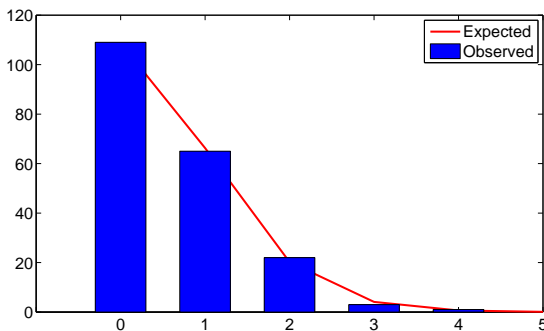
Гибель кавалеристов

Проверим гипотезу $H_0: Y \sim P(0.61)$ против общей альтернативы.

Критерий хи-квадрат:

$$O: [109, 65, 26],$$

$$E: [108.6702, 66.28, 25.0410].$$

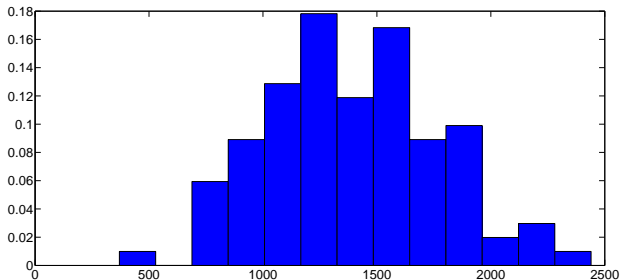


$$p = 0.7734.$$

Циклическая долговечность алюминиевых листов

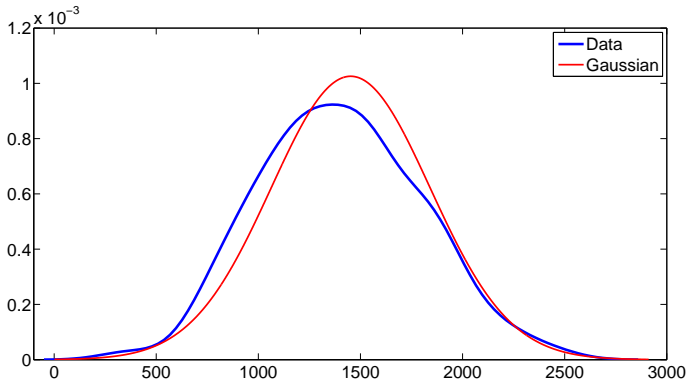
Birnbaum, Saunders, A Statistical Model for Life-Length of Materials, 1958:
измерена циклическая долговечность (в тысячах циклов напряжения до разрушения) 101 прямоугольной полосы алюминиевого листа 6061-T6.

Каким распределением можно описать эту величину?



Циклическая долговечность алюминиевых листов

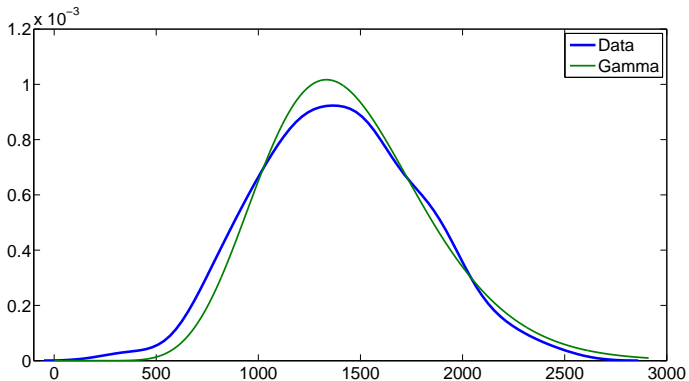
Нормальное распределение со средним 1401 и стандартным отклонением 389



Критерий хи-квадрат: $p = 0.6021$.

Циклическая долговечность алюминиевых листов

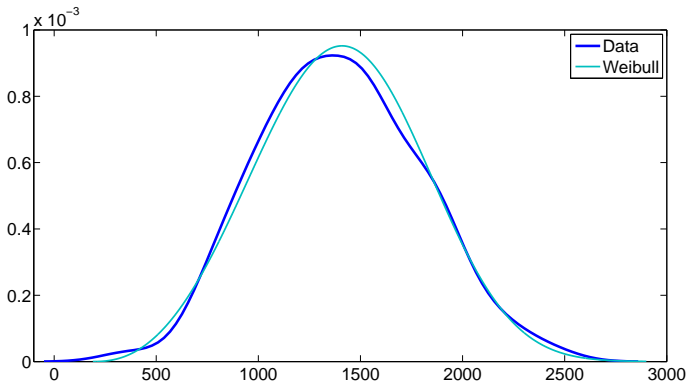
Гамма-распределение с параметром формы 11.85 и коэффициентом масштаба 0.00846



Критерий хи-квадрат: $p = 0.2472$.

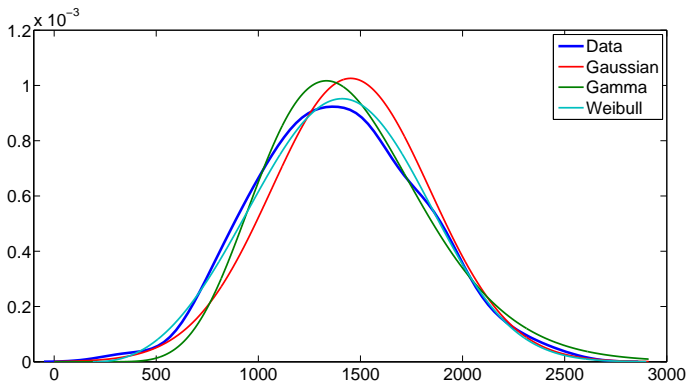
Циклическая долговечность алюминиевых листов

Трёхпараметрическое распределение Вейбулла со значениями параметра сдвига 181, формы — 3.34, масштаба — 1357



Критерий хи-квадрат: $p = 0.4365$.

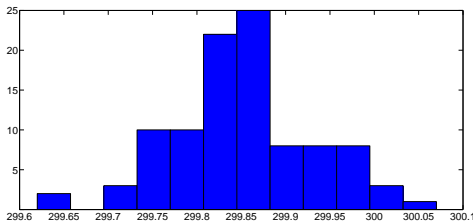
Циклическая долговечность алюминиевых листов



Критерии согласия не всегда помогут выбрать модельное распределение.

Измерения скорости света

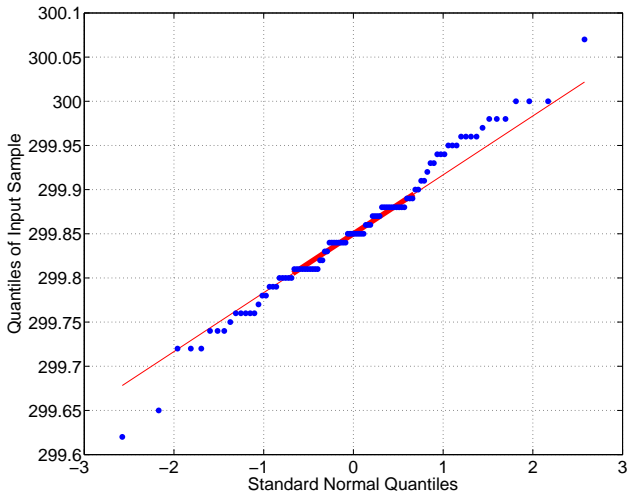
Данные классического эксперимента Михельсона по измерению скорости света (1879), 100 наблюдений.



Подчиняются ли измерения нормальному распределению?

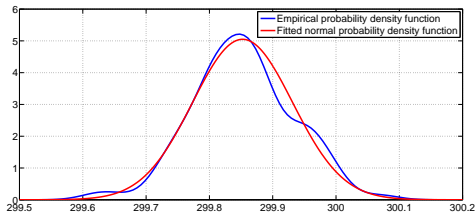


Измерения скорости света

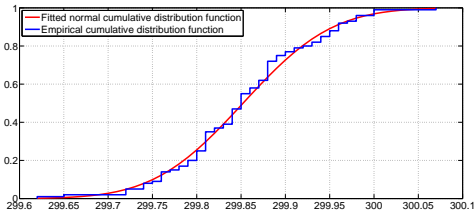


Q-Q plot

Измерения скорости света



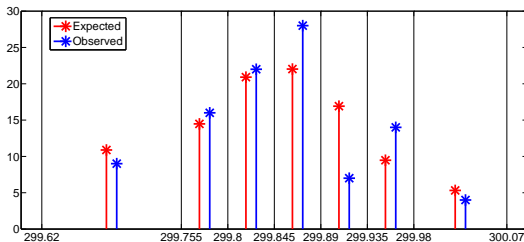
Оценки плотности вероятности



Оценки функции распределения

Измерения скорости света

Критерий хи-квадрат: $p = 0.0333$.



Чему равно число степеней свободы?

Критерий Колмогорова-Смирнова (Лиллиефорса): $p = 0.0860$.

Критерий Жарка-Бера (основан на выборочных коэффициентах асимметрии и эксцесса): $p = 0.5$ / $p = 0.8533$.

Критерий Андерсона-Дарлинга: $p = 0.2550$.

Критерий Крамера-фон Мизеса: $p = 0.2227$.

Критерий Шапиро-Уилка: $p = 0.5138$.

Итого о проверке нормальности

Главный вопрос: действительно ли нужно проверять нормальность?

- **очень маленькие выборки:** любой критерий может пропустить отклонения от нормальности, графические методы тоже часто бесполезны;
- **очень большие выборки:** любой критерий может выявлять небольшие статистически, но не практически значимые отклонения от нормальности (например, данные имеют распределение $N(0, 1.1)$, а не $N(0, 1)$); значительная часть методов, предполагающих нормальность, демонстрируют устойчивость к отклонениям;
- **выбросы:** сильно влияют на выборочные коэффициенты асимметрии и эксцесса;
- **связки:** критерии Шапиро-Уилка и, в особенности, Андерсона-Дарлинга, чувствительны к повторяющимся элементам;
- **критерий Колмогорова-Смирнова (Лиллиефорса):** представляет только исторический интерес (Agostino, Goodness-of-fit techniques);
- **критерий хи-квадрат:** слишком общий, не самый мощный, потеря информации из-за разбиения на интервалы.

Итого о проверке нормальности

Сравнение критериев проверки нормальности распределения случайных величин

Наименование критерия (раздел)	Характер альтернативного распределения					Ранг
	асимметричное		симметричное		≈ нормальное	
	$\alpha_4 < 3$	$\alpha_4 > 3$	$\alpha_4 < 3$	$\alpha_4 > 3$	$\alpha_4 \approx 3$	
Критерий Шапиро-Уилка (3.2.2.1)	1	1	3	2	2	1
Критерий K^2 (3.2.2.16)	7	8	10	6	4	2
Критерий Дарбина (3.1.2.7)	11	7	7	15	1	3
Критерий Д'Агостино (3.2.2.14)	12	9	4	5	12	4
Критерий α_4 (3.2.2.16)	14	5	2	4	18	5
Критерий Васичека (3.2.2.2)	2	14	8	10	10	6
Критерий Дэвида-Хартли-Пирсона (3.2.2.10)	21	2	1	9	1	7
Критерий χ^2 (3.1.1.1)	9	20	9	8	3	8
Критерий Андерсона-Дарлинга (3.1.2.4)	18	3	5	18	7	9
Критерий Филлибена (3.2.2.5)	3	12	18	1	9	10
Критерий Колмогорова-Смирнова (3.1.2.1)	16	10	6	16	5	11
Критерий Мартинеса-Иглевича (3.2.2.14)	10	16	13	3	15	12
Критерий Лина-Мудхолкара (3.2.2.13)	4	15	12	12	16	13
Критерий α_3 (3.2.2.16)	8	6	21	7	19	14
Критерий Шпигельхальтера (3.2.2.11)	19	13	11	11	8	15
Критерий Саркади (3.2.2.12)	5	18	15	14	13	16
Критерий Смирнова-Крамера-фон Мизеса (3.1.2.2)	17	11	20	17	6	17
Критерий Локка-Спурье (3.2.2.7)	13	4	19	21	17	18
Критерий Оя (3.2.2.8)	20	17	14	13	14	19
Критерий Хегази-Грина (3.2.2.3)	6	19	16	19	21	20
Критерий Муроты-Такеучи (3.2.2.17)	15	21	17	20	20	21

Кобзарь, Прикладная математическая статистика, 2006.

Смертность среди королевских пингвинов

Descamps et al., Relating demographic performance to breeding-site location in the king penguin, 2009: было помечено 50 королевских пингвинов в каждой из трёх областей гнездования на острове Владение архипелага Крозе; через год были собраны данные о том, сколько пингвинов погибло в каждой из групп.

Одинакова ли смертность пингвинов, обитающих в различных областях гнездования?



Смертность среди королевских пингвинов

	Выжило	Погибло
Lower area	43	7
Middle area	44	6
Upper area	49	1

Критерий хи-квадрат: $\chi = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$, $p = 0.0872$.

Ожидаемое число наблюдений во всех ячейках второго столбца меньше 5
⇒ применимость критерия под вопросом.

G-критерий независимости: $G = 2 \sum_{ij} O_{ij} \cdot \ln(O_{ij}/E_{ij})$, $p = 0.04827$.

Точный критерий Фишера: $p = 0.0869$.

Камни в почках

Charig et al., Comparison of treatment of renal calculi by operative surgery, percutaneous nephrolithotomy, and extracorporeal shock wave lithotripsy, 1986: по числу успешных исходов сравнивались два метода удаления камней в почках.

Пациенты с одним камнем <2 см

	Открытая хирургия	Чрескожная нефролитотомия
Успешный исход	81	234
Неудачный исход	6	36
Доля успешных исходов	93%	87%

Критерий хи-квадрат: $p = 0.1051$.

Пациенты с крупным камнем или несколькими камнями

	Открытая хирургия	Чрескожная нефролитотомия
Успешный исход	192	55
Неудачный исход	71	25
Доля успешных исходов	73%	69%

Критерий хи-квадрат: $p = 0.4580$.

Камни в почках

Все пациенты вместе:

	Открытая хирургия	Чрескожная нефролитотомия
Успешный исход	273	289
Неудачный исход	77	61
Доля успешных исходов	78%	83%

Критерий хи-квадрат: $p = 0.1285$.

Камни в почках

	Открытая хирургия	Чрескожная нефролитотомия	χ^2 p-value
Один камень <2 см	<i>Группа1</i> 93% (81/87)	<i>Группа2</i> 87% (234/270)	0.1051
Крупный камень или несколько камней	<i>Группа3</i> 73% (192/263)	<i>Группа4</i> 69% (55/80)	0.4580
Суммарно	78% (273/350)	83% (289/350)	0.1285

Причины несогласованности выводов:

- размеры групп 1-4 и 2-3 чересчур сильно отличаются, суммарный вывод определяется в основном вкладом групп 2-3;
- дополнительная переменная — размер камня — оказывает большее влияние на результат операции, чем выбор метода;
- все различия статистически незначимы, перед нами случайные колебания.

Абитуриенты Беркли

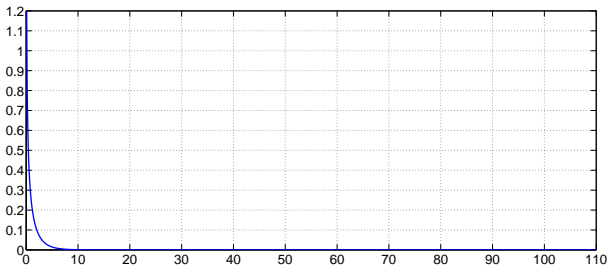
Bickel et al., Sex Bias in Graduate Admissions: Data from Berkeley, 1975: в 1973 году на университет Беркли, Калифорния, подали в суд: доля поступивших абитуриентов мужского пола была выше, чем доля поступивших женского пола.

	Не поступили	Поступили	Доля поступивших
Мужчины	4704	3738	44.3%
Женщины	2827	1494	34.6%



Абитуриенты Беркли

Критерий хи-квадрат, односторонняя гипотеза: $df = 1$; $\chi = 108.1$



$p \approx 0$.

	Наблюдаемые		Ожидаемые		Разности	
	-	+	-	+	-	+
Мужчины	4704	3738	4981.3	3460.7	-277.3	277.3
Женщины	2827	1494	2549.7	1771.3	277.3	-277.3

Абитуриенты Беркли

Будем искать виноватых: посмотрим детализированную статистику по 85 факультетам.

Значимо (на уровне $\alpha = 0.05$) меньше женщин прошли отбор на 4 факультета, суммарный дефицит — 26.

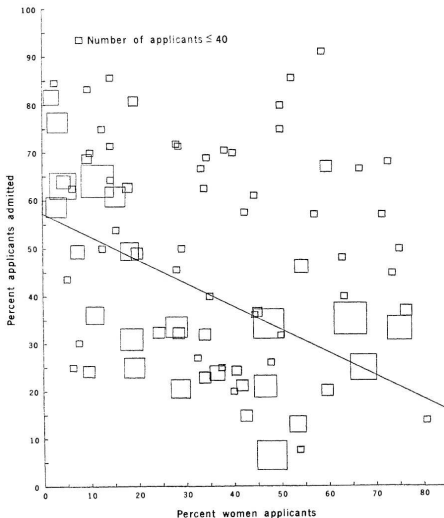
На 6 факультетов поступило значимо меньше мужчин, суммарный дефицит — 64.

Данные по шести крупнейшим факультетам:

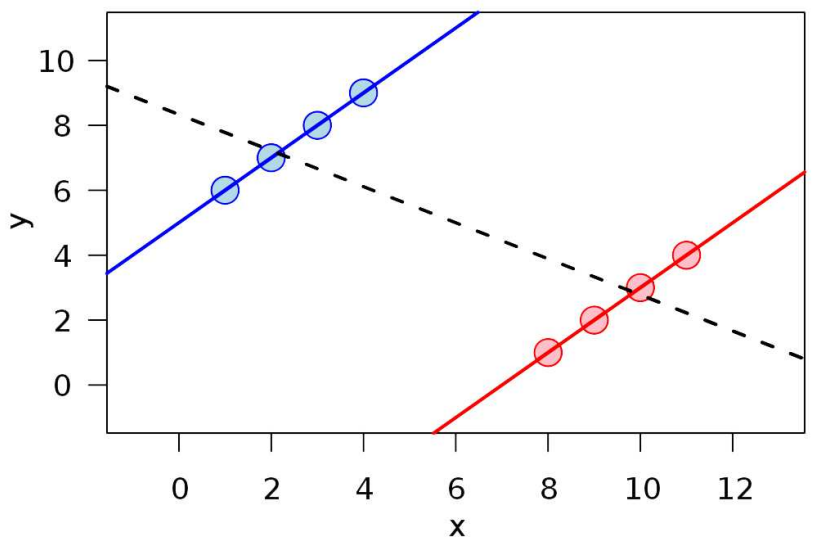
	Мужчины		Женщины	
	Σ	+	Σ	+
1	825	62%	108	82%
2	560	63%	25	68%
3	325	37%	593	34%
4	417	33%	375	35%
5	191	28%	393	24%
6	272	6%	341	7%

Абитуриенты Беркли

Ответ: женщины чаще пытаются поступить на факультеты с большим конкурсом.



Парадокс Симпсона для непрерывных данных



Прикладная статистика
Семинар 4. Критерии согласия. Анализ таблиц сопряжённости.

Рябенко Евгений
riabenko.e@gmail.com