Прикладной статистический анализ данных.
2. Проверка параметрических гипотез.

Михаил Хальман
psad.homework@gmail.com

2017

Виды задач: одновыборочные

- $X^n \sim N(\mu, \sigma^2)$
  - $H_0\colon \mu = \mu_0$
    - σ известна 1
    - σ неизвестна 3
  - $H_0\colon \sigma = \sigma_0$ 2

Виды задач: двухвыборочные

- $X_1^{n_1} \sim N(\mu_1, \sigma_1^2), \quad X_2^{n_2} \sim N(\mu_2, \sigma_2^2)$
  - $H_0\colon \mu_1 = \mu_2$
    - X_1, X_2 независимые
      - σ_1, σ_2 известны 4
      - σ_1, σ_2 неизвестны 5
    - X_1, X_2 связанные 6
  - $H_0\colon \sigma_1 = \sigma_2$ 7

1 Z-критерий

выборка: $X^n = (X_1, \ldots, X_n)\,, X \sim N\left(\mu, \sigma^2\right)$

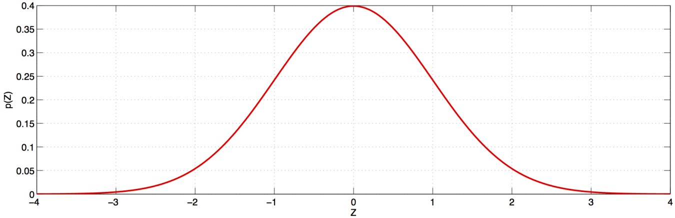
σ известна

нулевая гипотеза: $H_0 : \mu = \mu_0$

альтернатива: $H_1 : \mu <\neq> \mu_0$

статистика: $Z\left(X^n\right) = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

нулевое распределение: $N(0, 1)$



достигаемый уровень значимости:

$$p\left(Z\right) = \begin{cases} 1 - F_{N(0,1)}(Z), & H_1 : \theta > \theta_0, \\ F_{N(0,1)}(Z), & H_1 : \theta < \theta_0, \\ 2\left(1 - F_{N(0,1)}(|Z|)\right), & H_1 : \theta \neq \theta_0. \end{cases}$$

Z-критерий

Пример (Kanji, критерий 1): линия по производству пудры должна обеспечивать средний вес пудры в упаковке 4 грамма, заявленное стандартное отклонение — 1 грамм.
В ходе инспекции выбрано 9 упаковок, средний вес продукта в них составляет 4.6 грамма.

$H_0:$ средний вес пудры в упаковке соответствует норме.
$H_1:$ средний вес пудры в упаковке не соответствует норме $\Rightarrow p = 0.0719$, 95% доверительный интервал для среднего веса — $[3.95, 5.25]$ г.
$H_1:$ средний вес пудры в упаковке превышает норму $\Rightarrow p = 0.0359$, нижний 95% доверительный предел для среднего веса — 4.05 г.

2 Критерий хи-квадрат

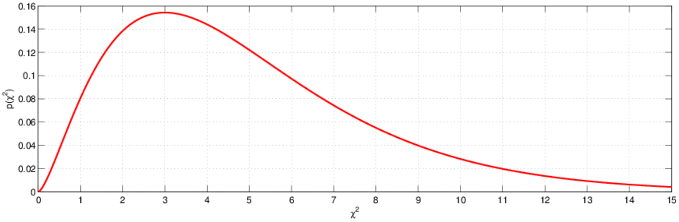
выборка: $X^n = (X_1, \dots, X_n), X \sim N(\mu, \sigma^2)$

нулевая гипотеза: $H_0\colon \sigma = \sigma_0$

альтернатива: $H_1\colon \sigma <\neq> \sigma_0$

статистика: $\chi^2(X^n) = \frac{(n-1)S^2}{\sigma_0^2}$

нулевое распределение: χ^2_{n-1}



достигаемый уровень значимости:

$$p(\chi^2) = \begin{cases} 1 - F_{\chi^2_{n-1}}(\chi^2), & H_1\colon \sigma > \sigma_0, \\ F_{\chi^2_{n-1}}(\chi^2), & H_1\colon \sigma < \sigma_0, \\ 2\min\left(1 - F_{\chi^2_{n-1}}(\chi^2), F_{\chi^2_{n-1}}(\chi^2)\right), & H_1\colon \sigma \neq \sigma_0. \end{cases}$$

2 Критерий хи-квадрат

Пример (Kanji, критерий 15): при производстве микрогидравлической системы делается инъекция жидкости. Дисперсия объёма жидкости — критически важный параметр, установленный стандартом на уровне 9 кв. мл. В выборке из 25 микрогидравлических систем выборочная дисперсия объёма жидкости составляет 12 кв. мл.

$H_0:$ дисперсия объёма жидкости соответствует стандарту.
$H_1:$ дисперсия объёма жидкости не соответствует стандарту $\Rightarrow p = 0.254$, 95% доверительный интервал для дисперсии — $[7.3, 23.2]$ кв. мл.
$H_1:$ дисперсия объёма жидкости превышает допустимое значение $\Rightarrow p = 0.127$, односторонний нижний 95% доверительный предел — 7.9 кв. мл.

t-критерий Стьюдента

выборка: $X^n = (X_1, \ldots, X_n), X \sim N(\mu, \sigma^2)$

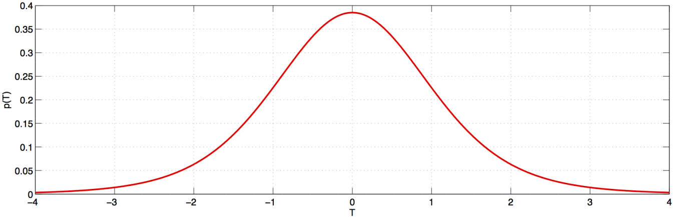
σ неизвестна

нулевая гипотеза: $H_0\colon \mu = \mu_0$

альтернатива: $H_1\colon \mu <\neq> \mu_0$

статистика: $T(X^n) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$

нулевое распределение: $St(n-1)$



С ростом объёма выборки разница между t- и z-критериями уменьшается.

3 t-критерий Стьюдента

Пример: средний вес детей при рождении составляет 3300 г. В то же время, если мать ребёнка живёт за чертой бедности, то средний вес таких детей — 2800 г.
С целью увеличить вес тех детей, чьи матери живут за чертой бедности, разработана экспериментальная программа ведения беременности. Чтобы проверить ее эффективность, проводится эксперимент. В нём принимают участие 25 женщин, живущих за чертой бедности. У всех них рождаются дети, и их средний вес составляет 3075 г, выборочное стандартное отклонение — 500 г.
Эффективна ли программа?

$H_0:$ программа не влияет на вес детей, $\mu = 2800$
$H_1:$ программа как-то влияет на вес детей, $\mu \neq 2800 \Rightarrow p = 0.0111$, 95% доверительный интервал для изменения веса — $[68.6, 481.4]$ г.
$H_1:$ программа увеличивает вес детей, $\mu > 2800 \Rightarrow p = 0.0056$, 95% нижний доверительный предел для увеличения веса — 103.9 г.

Выбор альтернативы

Одностороннюю альтернативу можно использовать, если знак изменения среднего известен заранее.

Альтернатива должна выбираться до получения данных!

Z-критерий

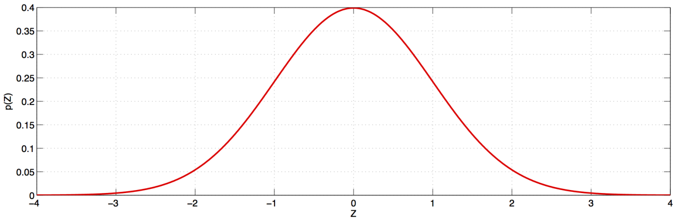
выборки: $X_1^{n_1} = \left(X_{11}, \dots, X_{1n_1}\right), X_1 \sim N\left(\mu_1, \sigma_1^2\right)$
$X_2^{n_2} = \left(X_{21}, \dots, X_{2n_2}\right), X_2 \sim N\left(\mu_2, \sigma_2^2\right)$
σ_1, σ_2 известны

нулевая гипотеза: $H_0\colon \mu_1 = \mu_2$

альтернатива: $H_1\colon \mu_1 <\neq> \mu_2$

статистика: $Z\left(X_1^{n_1}, X_2^{n_2}\right) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

нулевое распределение: $N(0,1)$



4 Z-критерий

Пример (Kanji, критерий 3): известно, что одна из линий по расфасовке чипсов даёт упаковки с более вариабельным весом продукта, чем вторая. Дисперсии равны 0.000576 г^2 и 0.001089 г^2 соответственно, средние значения веса в выборках из 13 и 8 элементов — 80.02 г и 79.98 г.

$H_0:$ средний вес продукта в упаковках, произведённых на двух линиях, совпадает.
$H_1:$ средние веса продукта в упаковках, произведённых на двух линиях, различаются $\Rightarrow p = 0.001$, 95% доверительный интервал для разности — $[0.039, 0.041]$.

5 t-критерий Стьюдента / Аспина-Уэлша (проблема Беренца-Фишера)

выборки: $X_1^{n_1} = (X_{11}, \ldots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma_1^2)$
$X_2^{n_2} = (X_{21}, \ldots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma_2^2)$
σ_1, σ_2 неизвестны

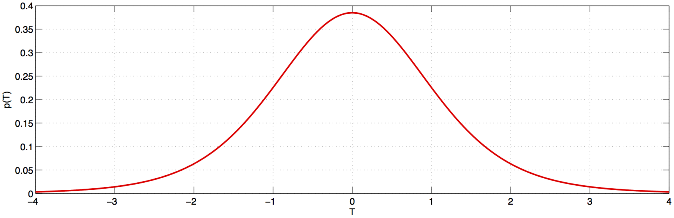
нулевая гипотеза: $H_0\colon \mu_1 = \mu_2$

альтернатива: $H_1\colon \mu_1 <\neq> \mu_2$

статистика: $T(X_1^{n_1}, X_2^{n_2}) = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{S_1^4}{n_1^2(n_1-1)} + \frac{S_2^4}{n_2^2(n_2-1)}}$$

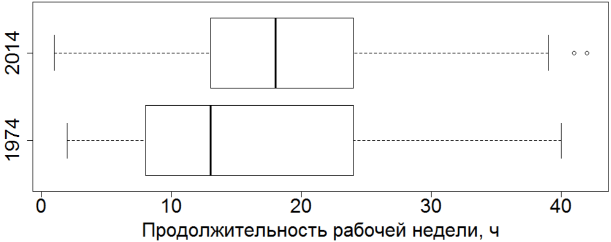
нулевое распределение: $\approx St(\nu)$



Приближение достаточно точно при $n_1 = n_2$ или $[n_1 > n_2] = [\sigma_1 > \sigma_2]$.

5 t-критерий Стьюдента / Аспина-Уэлша (проблема Беренца-Фишера)

Пример: в 1974 году 108 респондентов GSS работали неполный день, в 2014 — 196. Для каждого из них известно количество рабочих часов за неделю, предшествующую опросу.



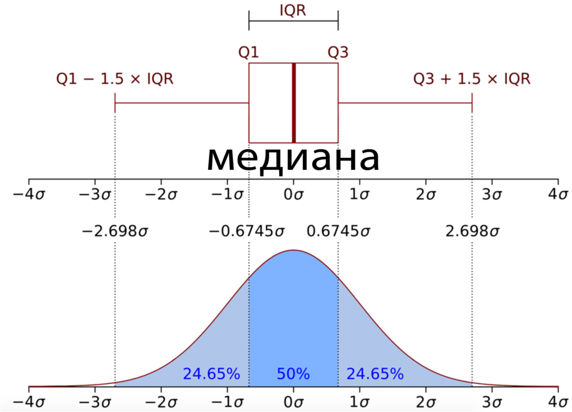
Изменилось ли среднее время работы у работающих неполный день?

$H_0:$ среднее время работы не изменилось, $\mu_1 = \mu_2$.
$H_0:$ среднее время работы изменилось, $\mu_1 \neq \mu_2$.

t-критерий: $p = 0.02707$, средняя продолжительность рабочей недели увеличилась на 2.57 часов (95% доверительный интервал — $[0.29, 4.85]$ ч).

Боксплот

Ящик с усами — способ визуализации основных характеристик распределения:



Длина усов отличается в разных реализациях.

t-критерий Стьюдента для связанных выборок

выборки: $X_1^n = (X_{11}, \ldots, X_{1n}), X_1 \sim N(\mu_1, \sigma_1^2)$
$X_2^n = (X_{21}, \ldots, X_{2n}), X_2 \sim N(\mu_2, \sigma_2^2)$
выборки связанные

нулевая гипотеза: $H_0\colon \mu_1 = \mu_2$

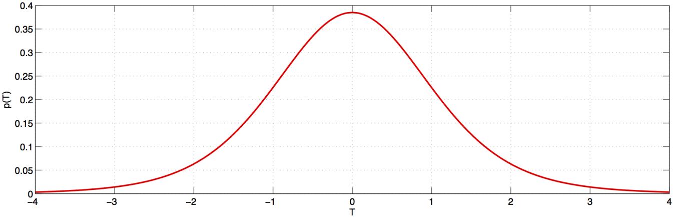
альтернатива: $H_1\colon \mu_1 <\neq> \mu_2$

статистика: $T(X_1^n, X_2^n) = \frac{\bar{X}_1 - \bar{X}_2}{S/\sqrt{n}}$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2}$$

$$D_i = X_{1i} - X_{2i}$$

нулевое распределение: $St(n-1)$



6 t-критерий Стьюдента для связанных выборок

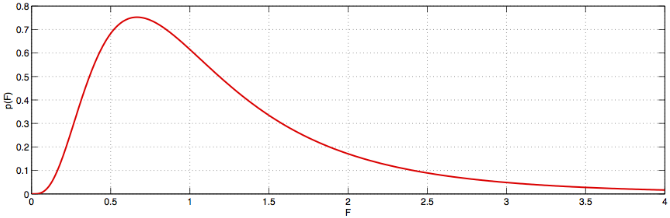
Пример (Kanji, критерий 10): на 10 испытуемых сравниваются два лекарства против респираторного заболевания. Каждый из испытуемых вдыхает первое лекарство с помощью ингалятора, после чего проходит упражнение беговой дорожке. Измеряется время достижения максимальной нагрузки. Затем после периода восстановления эксперимент повторяется со вторым лекарством.

$H_0:$ время достижения максимальной нагрузки не отличается для исследуемых лекарств.

$H_1:$ время достижения максимальной нагрузки для исследуемых лекарств отличается $\Rightarrow p = 0.916$; 95% доверительный интервал для разницы — $[-2.1, 0.9]$.

7 F-критерий Фишера

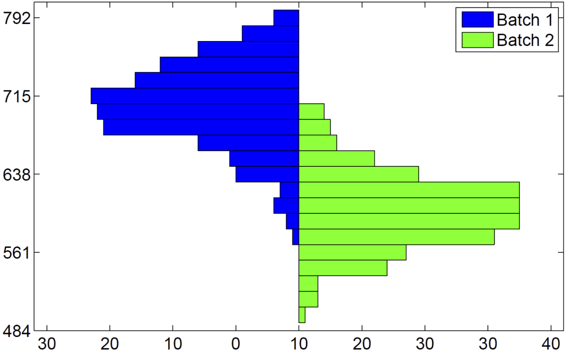
выборки:	$X_1^{n_1} = (X_{11}, \dots, X_{1n_1}), X_1 \sim N(\mu_1, \sigma_1^2)$
	$X_2^{n_2} = (X_{21}, \dots, X_{2n_2}), X_2 \sim N(\mu_2, \sigma_2^2)$
нулевая гипотеза:	$H_0\colon \sigma_1 = \sigma_2$
альтернатива:	$H_1\colon \sigma_1 <\neq> \sigma_2$
статистика:	$F(X_1^{n_1}, X_2^{n_2}) = \frac{S_1^2}{S_2^2}$
нулевое распределение:	$F(n_1 - 1, n_2 - 1)$



Критерий Фишера неустойчив к отклонениям от нормальности даже асимптотически.

7 F-критерий Фишера

Пример (NIST/industry ceramics consortium for strength optimization of ceramic, 1996): собраны данные о прочности материала 440 керамических изделий из двух партий по 220 в каждой.
Одинакова ли дисперсия прочности в разных партиях?



Критерий Фишера: $p = 0.1721, \quad [C_L, C_U] = [0.9225, 1.5690]$.

Литература

Критерии нормальности:

- Харке-Бера (Jarque–Bera) — Кобзарь, 3.2.2.16
- Шапиро-Уилка (Shapiro-Wilk) — Кобзарь, 3.2.2.1
- хи-квадрат (chi-square) — Кобзарь, 3.1.1.1, 3.2.1.1
- согласия (goodness-of-fit), основанные на эмпирической функции распределения — Кобзарь, 3.1.2, 3.2.1.2

Для нормальных распределений:

- Z-критерии (Z-tests) — Kanji, №№ 1, 2, 3
- t-критерии Стьюдента (t-tests) — Kanji, №№ 7, 8, 9
- критерий хи-квадрат (chi-square test) — Kanji, №15
- критерий Фишера (F-test) — Kanji, №16

Критерии, основанные на правдоподобии: Bilder, раздел B.5

Литература

Для распределения Бернулли:

- всё про одновыборочную задачу — Agresti, 1.3, 1.4
- Z-критерии (Z-tests) — Kanji, №№ 4, 5
- точный критерий (exact binomial test) — McDonald, `http://www.biostathandbook.com/exactgof.html`
- доверительные интервалы Уилсона (score confidence intervals) — Newcombe, 1998a, 1998b, 1998c

Agresti A. *Categorical Data Analysis*, 2013.

Bilder C.R., Loughin T.M. *Analysis of Categorical Data with R*, 2013.

Kanji G.K. *100 statistical tests*, 2006.

McDonald J.H. *Handbook of Biological Statistics*, 2008.

Newcombe R.G. (1998). *Two-sided confidence intervals for the single proportion: comparison of seven methods*. Statistics in Medicine, 17, 857–872.

Newcombe R.G. (1998). *Interval estimation for the difference between independent proportions: comparison of eleven methods*. Statistics in Medicine, 17, 873–890.

Newcombe R.G. (1998). *Improved confidence intervals for the difference between binomial proportions based on paired data*. Statistics in Medicine, 17, 2635–2650.

Литература

Кобзарь А.И. *Прикладная математическая статистика*, 2006.

Королёв В.Ю. *Теория вероятностей и математическая статистика*, 2008.

NIST/SEMATECH. *e-Handbook of Statistical Methods*.
`http://www.itl.nist.gov/div898/handbook/`