

ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ И ПРОВЕРКА ГИПОТЕЗ

*Элементы теории вероятностей и
математической статистики*

Результаты расчетов

математик: $x = 25$

инженер: $x = 25 \pm 3$

статистик: $x = 25 \pm 3$ с вероятностью 99%

Выборка

- *Цель выборочного наблюдения:*
- **оценить** неизвестные «истинные» показатели *генеральной совокупности* по известным наблюдаемым *выборочным* показателям

Стандартное распределение

- Статистический показатель
 - оценка
 - результат выборочного наблюдения
 - случайная величина со своим распределением
- Метод вычисления показателя →
 - стандартное распределение
 - доверительные интервалы
 - проверка гипотез
 - вероятность → квантиль
 - квантиль → вероятность

Построение доверительных интервалов

Доверительные интервалы

- confidence intervals, confidence bands
- Зная общую форму распределения и его числовые характеристики, можно оценить диапазон значений, в котором окажется большинство единиц совокупности

Доверительный интервал

- Диапазон значений, в котором с заданной вероятностью окажется случайная величина
 - вероятность (два определения)
 - доля (процент) единиц совокупности с заданными свойствами
 - уровень субъективной уверенности
- Вероятность этого события – *доверительная* вероятность
 - уровень нашего доверия к утверждению по поводу попадания в интервал
 - степень достоверности события
 - возможная частота (процент случаев), с которой произойдет событие

Доверительный интервал для генерального среднего

- По известным выборочным показателям (среднее и СКО) строится диапазон возможных значений неизвестного генерального среднего

Распределение выборочного среднего

- Среднее значение
 - сумма случайных величин

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

- нормальное распределение

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Свойства случайных величин

	$x + y$	$x - y$
E	$E(x + y) =$ $E(x) + E(y)$	$E(x - y) =$ $E(x) - E(y)$
D	$D(x + y) =$ $D(x) + D(y)$	$D(x - y) =$ $D(x) + D(y)$

Правило сложения дисперсий

Дисперсии *складываются*
и при сложении, и при вычитании
независимых случайных величин

неопределенность возрастает

М.О. выборочного среднего

$$\begin{aligned} E[\bar{x}] &= \frac{1}{n} E[x_1 + \dots + x_n] = \\ &= \frac{1}{n} (E[x_1] + \dots + E[x_n]) = \\ &= \frac{1}{n} (\mu + \dots + \mu) = \mu \end{aligned}$$

Дисперсия выборочного среднего

$$\begin{aligned} D[\bar{x}] &= D\left[\frac{x_1}{n} + \dots + \frac{x_n}{n}\right] = \\ &= D\left[\frac{x_1}{n}\right] + \dots + D\left[\frac{x_n}{n}\right] = \\ &= \frac{1}{n^2} (D[x_1] + \dots + D[x_n]) = \\ &= \frac{1}{n^2} (\sigma^2 + \dots + \sigma^2) = \frac{\sigma^2}{n} \end{aligned}$$

Свойства выборочного среднего

– МО выборочного среднего

$$\mu_{\bar{x}} = \mu_x$$

– СКО выборочного среднего

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

Двусторонний интервал

симметричный доверительный интервал

$$\bar{x} - \Delta \leq \mu \leq \bar{x} + \Delta$$

$\bar{x}$ – выборочное среднее

Δ – ошибка выборки
(ошибка выборочного наблюдения)

μ – генеральное среднее (М.О.)

Ошибка выборочного наблюдения

разность между генеральным средним и
выборочным средним

$$\Delta_{\bar{x}} = |\bar{x} - \mu|$$

«ошибка выборки»

Уровень значимости

- вероятность того, что значение μ окажется за пределами доверительного интервала

– significance level

– p -level

$$\alpha = 1 - p$$

Расположение внутри и вне интервала

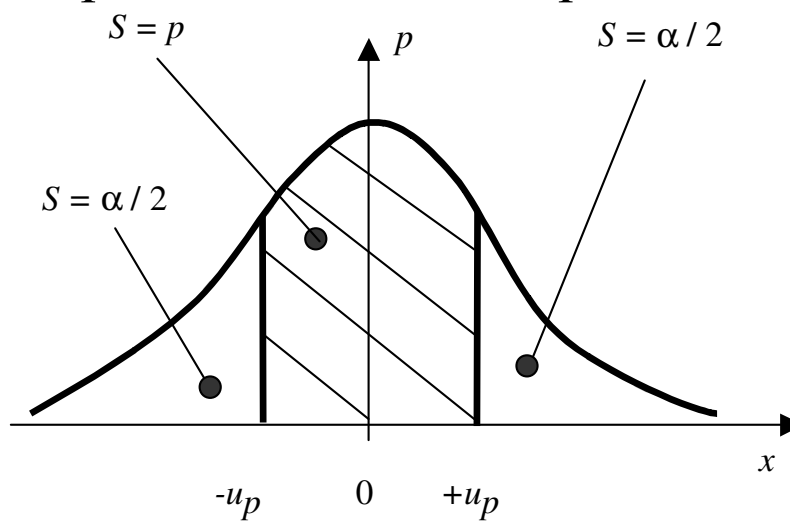
– *полная система событий*

– вероятность = 1

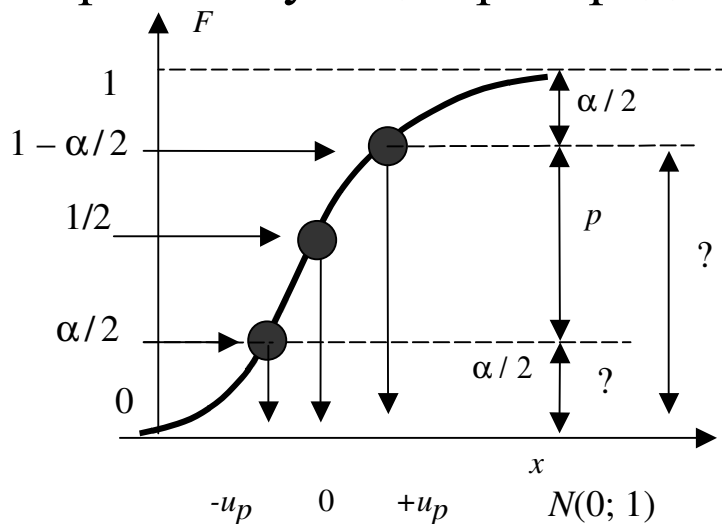
$$p + \alpha = 1$$

- доверительная вероятность + уровень значимости = 1

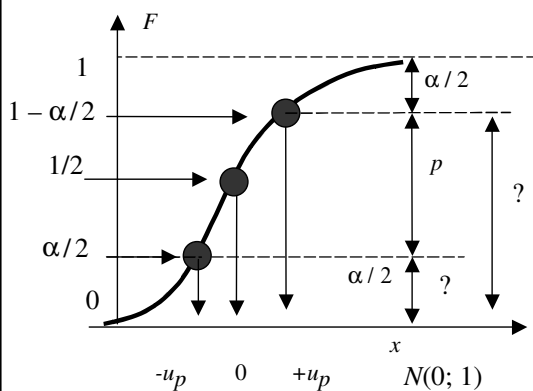
Двусторонний доверительный интервал: Плотность вероятности



Двусторонний доверительный интервал: Функция распределения



Двусторонний доверительный интервал: Квантили распределения

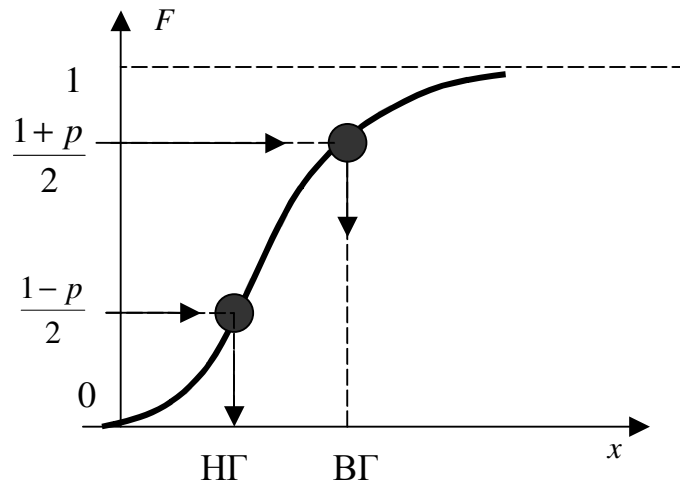


$$\alpha = 1 - p$$

$$\frac{\alpha}{2} = \frac{1 - p}{2}$$

$$\begin{aligned} 1 - \frac{\alpha}{2} &= 1 - \frac{1 - p}{2} = \\ &= \frac{2 - 1 + p}{2} = \frac{1 + p}{2} \end{aligned}$$

Двусторонний доверительный интервал: Квантили распределения



Доверительный интервал для генерального среднего

$$\bar{x} - \Delta \leq \mu \leq \bar{x} + \Delta$$

$$\bar{x} - u_p \sigma_{\bar{x}} \leq \mu \leq \bar{x} + u_p \sigma_{\bar{x}}$$

$$\bar{x} - u_p \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + u_p \frac{\sigma}{\sqrt{n}}$$

u_p – квантиль стандартного нормального распределения (сколько «сигм» брать?)

Три «стандартных» квантиля нормального распределения

<i>Квантиль</i>	<i>Ошибка выборки</i>		<i>Вероятность</i>	
1	$\pm 1\sigma$	одна сигма	68%	0,6827
2	$\pm 2\sigma$	две сигмы	95%	0,9545
3	$\pm 3\sigma$	три сигмы	99,7%	0,9973

«Таблица умножения»

<i>Квантиль</i>	<i>Вероятность</i>
1	68%
2	95%
3	99,7%

Правило трех сигм

- В диапазон от $\mu - 3\sigma$ до $\mu + 3\sigma$ попадает 99,7% всех значений случайной величины, распределенной по нормальному закону = «почти все»
- Появление значений за пределами интервала $\mu \pm 3\sigma$ практически невозможно ($\alpha = 0,003$)
- Наблюдения за пределами диапазона можно считать ошибками (аномалиями) и отбрасывать
- Много аномальных наблюдений = выборка является неоднородной

Доверительный интервал для генерального среднего

- Отношение среднего к СКО ~ распределение Стьюдента
- При малом объеме выборки используют распределение Стьюдента
- При большом объеме выборки $n > 100$ t -распределение приближенно заменяют нормальным

Распределение Стьюдента

- Среднее на сигму ~ распред. Стьюдента

$$\frac{\bar{x}}{s} = \frac{\frac{\sum x}{n}}{\sqrt{\frac{\sum (x - \bar{x})^2}{n}}} \sim \frac{z}{\sqrt{\chi_n^2/n}} \sim t_n\left(\mu, \frac{s}{\sqrt{n}}\right)$$

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_n(0; 1)$$

Доверительный интервал для Г.С.: t -распределение

$$\bar{x} + t_{\frac{1-p}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{1+p}{2}} \frac{s}{\sqrt{n}}$$

$$\bar{x} - t_{\frac{1+p}{2}} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{1+p}{2}} \frac{s}{\sqrt{n}}$$

Ошибка выборки

$$\Delta = t_n \left(\frac{1+p}{2} \right) \cdot \frac{s}{\sqrt{n}}$$

$$\Delta = t \cdot \frac{s}{\sqrt{n}}$$

t – коэффициент доверия, квантиль
(процентная точка) распределения Стьюдента

t -распределение

- t -распределение дает более широкие доверительные интервалы, чем нормальное
- Чем меньше объем выборки (число степеней свободы), тем шире интервал
- Пример:

$$t\{p=0,9545; n=10\}=2,28$$

$$t\{p=0,9545; n=20\}=2,13$$

$$z\{p=0,9545\}=2,00$$

Объем выборки

- Дано: требуемая погрешность Δ
- Найти: необходимый объем выборки n

$$\Delta = t \cdot \frac{s}{\sqrt{n}}$$

$$n = ?$$

Упрощенное построение доверительного интервала

- Заменяем распределение Стьюдента нормальным распределением и используем приближенные значения трех стандартных квантилей – одна, две или три «СИГМЫ»
 $\bar{x} = 10; \quad \sigma = 5; \quad n = 100; \quad p = 95\%$

$$\sigma(\bar{x}) = \frac{5}{\sqrt{100}} = \frac{5}{10} = 0,5$$

$$10 - 1 \leq \mu \leq 10 + 1; \quad 9 \leq \mu \leq 11$$

- С вероятностью 95% можно утверждать, что генеральное среднее находится в диапазоне от 9 до 11
- Рискнем ошибиться в 5% случаев

Проверка статистических гипотез

Проверка гипотез

- hypothesis testing
- статистические гипотезы – предположения (утверждения) о свойствах случайных величин
 - стат. гипотезу можно «принять» или «отвергнуть»
 - стат. гипотезу нельзя доказать
 - любые заключения носят приближенный характер
- «фактические данные говорят в пользу утверждения...» либо «факты не говорят в пользу ...»

Примеры статистических гипотез

- среднее значение находится в диапазоне от 20 до 25
- «истинное» среднее значение равно 76
- дисперсия больше 2,5
- показатель имеет нормальное распределение

Методы *проверки гипотез*

Статистические выводы относительно свойств совокупности делают с помощью методов *проверки гипотез*:

- вначале выдвигается предположение
- затем предположение проверяется

По результатам проверки его принимают или отвергают с заданной вероятностью

- Условия проверки статистических выводов:
критерий согласия – правило, по которому можно *согласиться* с определенным предположением

Критерий согласия

Критерий – правило для принятия вывода; признак, на основе которого производится оценка

- [Gr. *kriterion* – средство для суждения, принятия решения
- ← *krites* – судить]

Критерий согласия позволяет сравнить теоретическое и эмпирическое распределения:

«Можно ли согласиться с гипотезой, что эти распределения одинаковы?»

Порядок проверки гипотезы

- задают уровень значимости
- строят критическую область (теоретическую статистику)
- проверяют, куда попала изучаемая величина (фактическая статистика)
- сравнивают критическое (табличное, теоретическое значение) с фактическим
- если фактическая статистика попадает в критическую область, то гипотеза отвергается

Свойства статистических гипотез

- статистическую гипотезу нельзя доказать
- статистическую гипотезу можно отвергнуть

Свойства гипотез

- Для опровержения гипотезы достаточно одного противоречащего примера
- Доказать правильность гипотезы невозможно даже с миллионом примеров
- Для «доказательства» утверждения выдвигается противоположная гипотеза и затем отвергается

Статистическая значимость

statistical significance

существенная (значимая) разница между фактическим значением и гипотезой

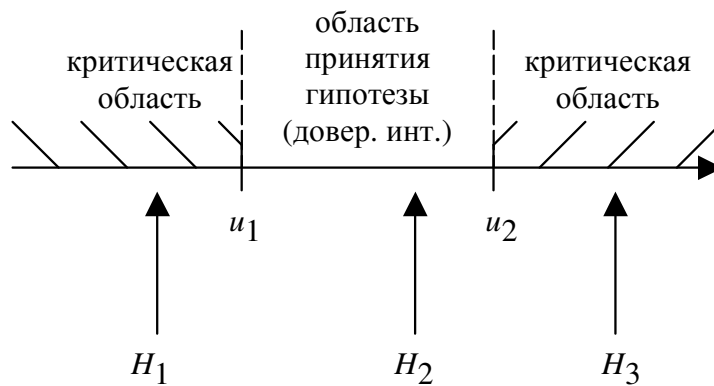
существенное, осязаемое, значимое отличие изучаемого показателя от нуля

Разница показателей либо *статистически значима*, либо укладывается в рамки допустимого отклонения, т.е. *не является значимой*

Критическая область

- противоположность доверительному интервалу
- строится с помощью квантилей
- Если значение попало в критическую область, гипотезу отвергают
 - Это маловероятное, практически невозможное событие
 - При проверке гипотезы это нежелательное событие считается невозможным
 - Если значение параметра выйдет за границы доверительного интервала, гипотезу отвергают, и говорят, что это событие произошло не случайно

Критическая область



Уровень значимости

significance level, p-level

уровень значимости α определяет мощность критерия

вероятность нежелательного события

маловероятное, практически невозможное событие

Если невозможное событие произойдет:

исходную гипотезу придется отвергнуть

это явление не случайно

Уровень значимости в процентах:

сколько раз из 100 испытаний мы рискуем ошибиться, объявив событие неслучайным

на какой риск мы согласны

Ошибки

Решение после проверки гипотез:

- ошибка первого рода: отвергнуть верную гипотезу;
- ошибка второго рода: принять неверную гипотезу

Уровень значимости зависит от *тяжести последствий* ошибки

- Чем серьезней последствия ошибки, тем ниже должен быть уровень значимости
 - сломанная парта
 - сломанное крыло самолета

Гипотеза о генеральном среднем

$$H_0: \mu = A$$

- Гипотеза о том, что Г.С. = A
- Находим табличное (критическое) значение статистики
- Строим фактическую статистику
- Если фактическая статистика больше критической, гипотезу отвергаем

Гипотеза о генеральном среднем

$$H_0: \mu = 5$$

$$\bar{x} = 10; \quad n = 100; \quad \sigma = 5; \quad \alpha = 0,05$$

$$t_{\text{кр}} = 1,984$$

$$t_{\text{ф}} = \left| \frac{10 - 5}{0,5} \right| = \frac{5}{0,5} = 10$$

$$t_{\text{ф}} > t_{\text{кр}}$$

Гипотеза о генеральном среднем

- Фактическая t -статистика = сравнение
 - разности выборочного и генерального среднего
 - с возможной ошибкой выборки
 - СКО среднего, умноженное на квантиль
- Если модуль отклонения больше допустимого уровня ошибки, то гипотезу отклоняют

Гипотеза о генеральном среднем



Упрощенная проверка гипотезы

- Гипотеза о том, что генеральное среднее равно пяти. Строим доверительный интервал упрощенным способом:
- $H_0: \mu = 5; \quad 9 \leq \mu \leq 11$
- Проверяемое значение не попадает в доверительный интервал; оно относится к критической области. Гипотезу следует отвергнуть.

Упрощенная проверка гипотезы



Сравнение средних

- Имеется две разных выборки, нужно проверить возможность того, что их генеральные средние равны
- Насколько существенна (статистически значима) разница средних значений
- Насколько ощутимы результаты управленческого решения, воздействия, изменения.

Необходимость сравнения средних

Продажи до рекламной кампании и после нее

Выздоровление при обычном лечении и с новым лекарством

Стабильность технологического процесса во времени

Эффективность реорганизации компании

Сравнение средних: Постановка задачи

- | | | |
|-------------------------------|-------------|-------------|
| • Даны две выборки объемом | n_1 | n_2 |
| • Генеральные средние выборок | μ_1 | μ_2 |
| • Генеральные СКО выборок | σ_1 | σ_2 |
| • Выборочные средние | $\bar{x}_1$ | $\bar{x}_2$ |

- Проверить гипотезу о равенстве генеральных средних

$$H_0 : \mu_1 = \mu_2$$

Гипотеза о равенстве средних

$$H_0 : \mu_1 = \mu_2$$

Будет ли различие средних статистически значимым?

$$H_0 : \mu_1 - \mu_2 = 0$$

Значимое отличие разности от нуля
(по сравнению с погрешностью)

Разность средних

$$\Delta = \bar{x}_1 - \bar{x}_2$$

- Выборочные средние – случайные величины
- Разность выборочных средних – случайная величина

Разность средних

$$\Delta = \bar{x}_1 - \bar{x}_2$$

$$E[\Delta] = E[\bar{x}_1] - E[\bar{x}_2] = \mu_1 - \mu_2$$

$$D[\Delta] = D[\bar{x}_1] + D[\bar{x}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Доверительный интервал для разности средних

$$(\bar{x}_1 - \bar{x}_2) - u_{\frac{1+p}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + u_{\frac{1+p}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Проверка гипотезы

- Гипотеза о равенстве средних отвергается, если фактическая статистика больше критического значения

$$\frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > u_{\frac{1+p}{2}}$$

Средневзвешенная дисперсия

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- с числом степеней свободы

$$df = n_1 + n_2 - 2$$

Критерий согласия

- Сколько сигм составляет разница?
 - больше трех сигм: статистически значимо
 - меньше одной сигмы: статистически незначимо
 - от одной до трех сигм: неопределенность

Критерий согласия

$$|\bar{x}_1 - \bar{x}_2| > t_{\frac{1+p}{2}}(n_1 + n_2 - 2) \cdot s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\frac{|\bar{x}_1 - \bar{x}_2|}{s \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{\frac{1+p}{2}}(n_1 + n_2 - 2)$$

- Сколько сигм составляет разница?

Число степеней свободы n

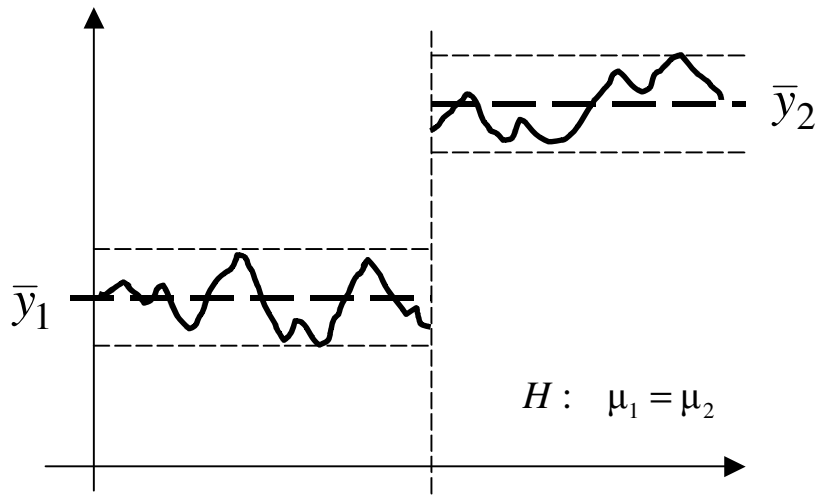
- degrees of freedom (d.f.)
- число независимых координат минус число связей (ограничений, условий)
 - точка на плоскости: $n = 2$
 - точка в пространстве: $n = 3$
 - предмет в пространстве: $n = 6$ (6 DOF)
 - дверь: $n = ?$
 - выборка с учетом среднего: $n = ?$
 - две выборки и два средних: $n = ?$

Определенность выводов

При проверке статистических гипотез возможны 3 варианта:

- (1) гипотеза отвергается с любым уровнем значимости – «УВЕРЕННОЕ НЕТ»
- (2) гипотеза принимается с любым уровнем значимости – «УВЕРЕННОЕ ДА»
- (3) гипотеза отвергается с одним уровнем значимости и отвергается с другим – «НЕУВЕРЕННОСТЬ, НЕОПРЕДЕЛЕННОСТЬ»

Пример



Пример

