Беседы о прикладной статистике

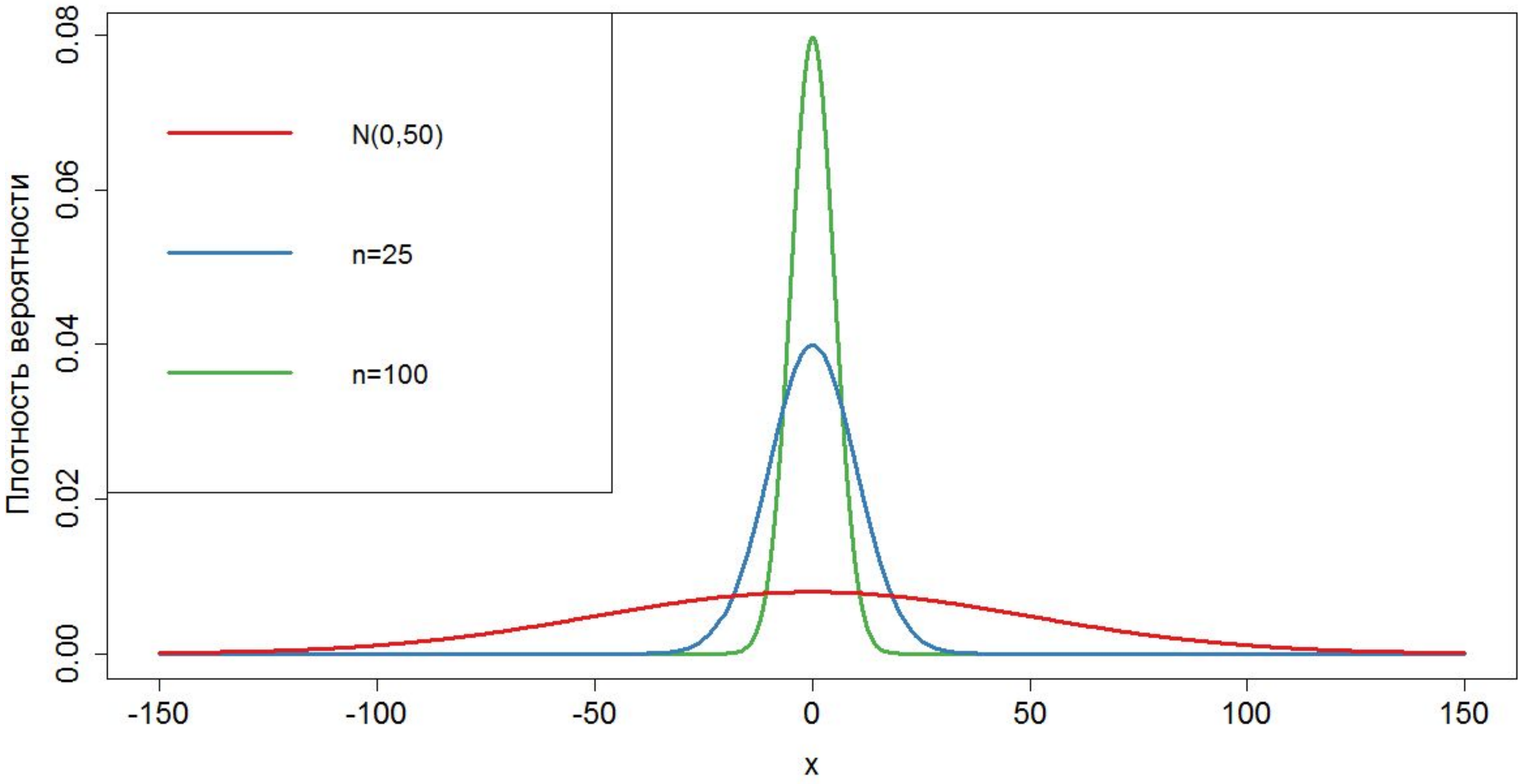
Семинар 5. Доверительные интервалы для среднего. Критерий t Стьюдента. Критерии Уилкоксона для ранговых сравнений

Фастовец И. А.

Стандартное отклонение распределения выборочных средних

- Если из нормального распределения N(μ,σ) взять выборку объемом n, тогда средние по выборкам будут иметь нормальное распределение N(μ, $\frac{\sigma}{\sqrt{n}}$)
- *Центральная предельная теорема*: если из любого распределения, имеющего среднее μ и конечное стандартное отклонение σ взять выборку критически большого объема n, тогда распределение средних по выборкам будет приблизительно нормальным N(μ, $\frac{\sigma}{\sqrt{n}}$)

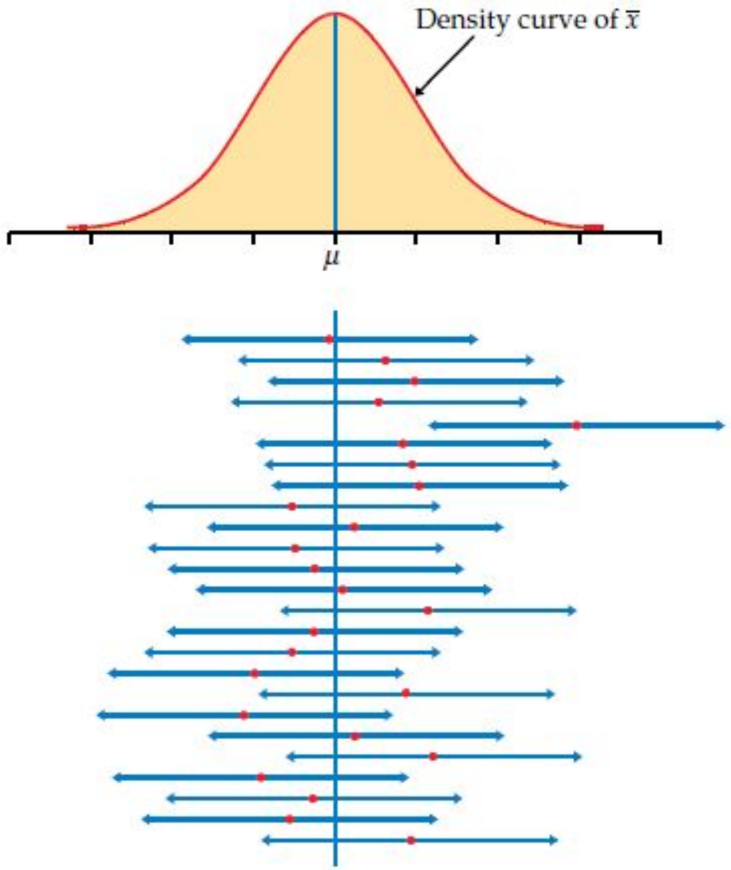
Стандартное отклонение распределения выборочных средних



z - доверительный интервал для среднего

- Среднее μ неизвестно, стандартное отклонение σ известно
- Предел погрешности для среднего $\bar{x}$ будет иметь вид $m = z' \frac{\sigma}{\sqrt{n}}$, где z' – значение на стандартной нормальной кривой с площадью С между критическими точками -z' и z' (берем из таблицы)
- Доверительный интервал на уровне доверительной вероятности С

  будет равен $\bar{x} \pm m$
- Объем выборки, необходимый для нужного предела погрешности $n = z' \left(\frac{z'\sigma}{m}\right)^2$



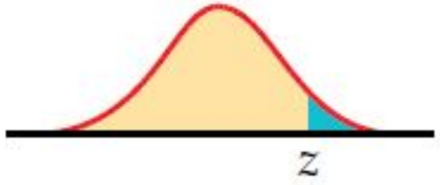
Статистические сравнения

- *Нулевая гипотеза* – утверждение, которое мы проверяем при помощи тестов значимости. В данном случае H_0 : сравниваемые средние генеральных совокупностей равны
- *P - значение* – расчетная вероятность того, что случайная величина с данным распределением (распределением тестовой статистики при нулевой гипотезе) примет значение, не меньшее, чем фактическое значение тестовой статистики. Чем меньше P - значение, тем лучше
- Если P - значение меньше заданного уровня *доверительной вероятности* α (например, 0.05), мы говорим, что результат *статистически значимый*

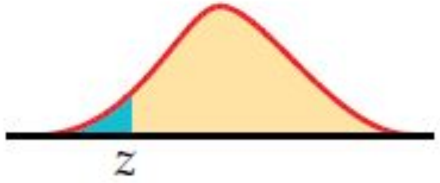
z - тест значимости разницы среднего распределения и заданной константы

- Тестируем H_0: $\mu = \mu_0$, где μ_0 – заданная константа, а μ – неизвестное среднее нормального распределения со стандартным отклонением σ.
- Взята выборка объемом n со средним $\bar{x}$. Тогда статистика z имеет вид: $z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$
- По таблице находим соответствующее P - значение, делаем вывод о значимости различий

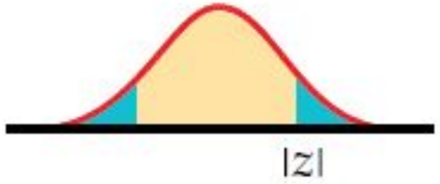
$H_a: \mu > \mu_0$ is $P(Z \geq z)$



$H_a: \mu < \mu_0$ is $P(Z \leq z)$



$H_a: \mu \neq \mu_0$ is $2P(Z \geq |z|)$



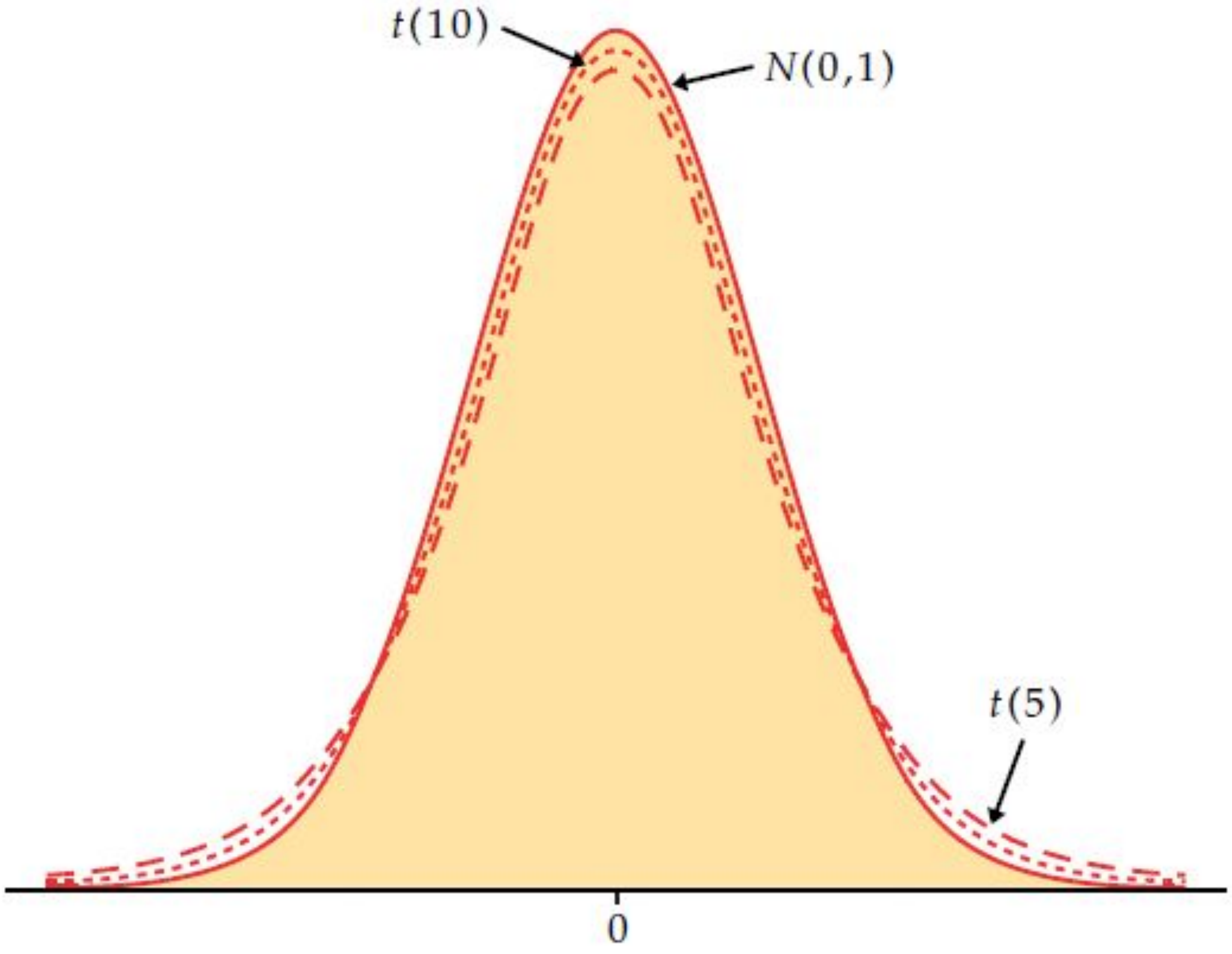
z тест значимости разницы двух средних

- Предположим, что $\overline{x}_1$– среднее значение выборки объемом n из распределения N(μ_1, σ_1), а $\overline{x}_2$ – среднее независимой выборки объемом n из распределения N(μ_2, σ_2). Тогда z статистика для двух выборок будет иметь вид
- $z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ и будет распределена в соответствии с N(0,1)

Семейство распределений t Стьюдента

- Если σ неизвестна, то для нахождения доверительного интервала или тестирования значимости различий применяют не $\frac{\sigma}{\sqrt{n}}$, а оценку стандартного отклонения распределения среднего по выборкам $\frac{s}{\sqrt{n}}$, где s – стандартное отклонение выборки. Эта величина именуется *стандартной ошибкой среднего*
- Возьмем выборку объемом n из генеральной совокупности N(μ,σ), тогда t статистика для одной выборки примет вид $t = \frac{\overline{x} - \mu}{s/_{\sqrt{n}}}$ и будет иметь t распределение с количеством степеней свободы n-1

Нормальное и t распределения



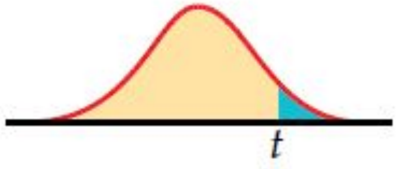
t доверительный интервал для одной выборки

- Возьмем случайную выборку из генеральной совокупности с неизвестным средним μ и неизвестным стандартным отклонением
- t доверительный интервал для μ на уровне доверительной вероятности С будет равен $x \pm t' \frac{s}{\sqrt{n}}$, где t' – значение на кривой t распределения t(n-1) площадью С между критическими точками -t' и t'
- $t' \frac{s}{\sqrt{n}}$ в данном случае будет пределом погрешности. Этот интервал является точным в случае нормального распределения генеральной совокупности и приблизительно корректным для больших n в других случаях

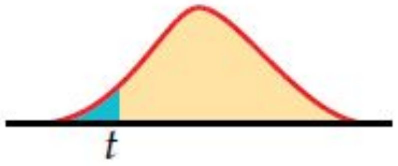
t тест для одной выборки (сравнение с константой)

- Случайную выборку объемом n взяли из распределения с неизвестным средним μ и неизвестным стандартным отклонением. Задача – сравнить μ с заданной константой μ_0
- Чтобы протестировать $H_0\colon\ \mu = \mu_0$ рассчитаем значение t статистики $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
- По таблице найдем P-значение
- Сделаем вывод о значимости различий

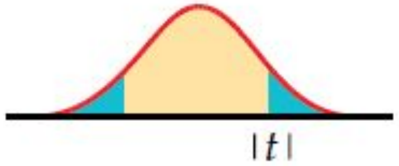
$H_a\colon \mu > \mu_0$ is $P(T \geq t)$



$H_a\colon \mu < \mu_0$ is $P(T \leq t)$



$H_a\colon \mu \neq \mu_0$ is $2P(T \geq |t|)$



t тест для парных выборок

- Когда некоторая характеристика измерена для одних и тех же объектов до и после воздействия
- Рассчитаем разницу между значениями для каждого объекта
- Проведем t тест для одной выборки для сравнения среднего разниц значений с 0
- $t = 6.46$, $df = 14$, $P<0.001$

Aggressive behaviors of dementia patients

Patient	Moon days	Other days	Difference	Patient	Moon days	Other days	Difference
1	3.33	0.27	3.06	9	6.00	1.59	4.41
2	3.67	0.59	3.08	10	4.33	0.60	3.73
3	2.67	0.32	2.35	11	3.33	0.65	2.68
4	3.33	0.19	3.14	12	0.67	0.69	−0.02
5	3.33	1.26	2.07	13	1.33	1.26	0.07
6	3.67	0.11	3.56	14	0.33	0.23	0.10
7	4.67	0.30	4.37	15	2.00	0.38	1.62
8	2.67	0.40	2.27				

Постановка задачи для двух независимых выборок

- Задача – сравнить эффект между двумя группами
- Каждая группа представляет из себя случайную выборку из двух различных распределений
- Эффект в одной группе не зависит от эффекта в другой группе

Population	Variable	Mean	Standard deviation
1	x_1	μ_1	σ_1
2	x_2	μ_2	σ_2

Population	Sample size	Sample mean	Sample standard deviation
1	n_1	$\bar{x}_1$	s_1
2	n_2	$\bar{x}_2$	s_2

t тест Уэлча для двух выборок

- Расчет t статистики будет похож на расчет z статистики, в котором мы заменяем σ на s:
- $t = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- Проблема заключается в расчете количества степеней свободы у этого распределения. Консервативный расчет: взять наименьшее из n_1-1 и n_2-1
- Более точная аппроксимация:

$$\mathrm{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

t тест Уэлча для двух выборок

- Допустим, что одна случайная выборка объемом n_1 взята из нормальной генеральной совокупности со средним μ_1, а вторая случайная выборка объемом n_2 взята из другой нормальной генеральной совокупности со средним μ_2
- Тогда t-статистика будет иметь вид $t = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
- Далее рассчитываем P-значение по таблице t критерия для количества степеней свободы выбранной аппроксимации

Доверительный интервал Уэлча для разницы средних

- Доверительный интервал для разницы средних на уровне доверительной вероятности C будет иметь вид:
- $(\overline{x}_1 - \overline{x}_1) \pm t'\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$, где t' – значение кривой t распределения плотности вероятности с площадью C между -t' и t'
- Количество степеней свободы аппроксимируется одним из двух способов

t тест с объединенной оценкой дисперсии

- Если мы предполагаем, что выборки взяты из генеральных совокупностей с одним стандартным отклонением σ_p , формулы приобретают вид:
- $t = \frac{(\overline{x}_1 - \overline{x}_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$
- $(\overline{x}_1 - \overline{x}_1) \pm t' s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$
- Количество степеней свободы в данном случае равно $n_1 + n_2 - 2$
- Для использования этого теста необходимо проверить однородность дисперсий. Кроме того, выборки должны быть похожими по объему

U тест суммы рангов Уилкоксона-Манна-Уитни

- Возьмем случайную выборку объемом n_1 из одной генеральной совокупности и вторую независимую выборку объемом n_2 из второй генеральной совокупности в сумме имеем $N = n_1 + n_2$ наблюдений. Проранжируем все N наблюдений. Тогда сумма W рангов первой выборки является статистикой суммы рангов Уилкоксона.
- Если две генеральных совокупности имеют одинаковое непрерывное распределение то W имеет среднее и стандартное отклонение:

$$\mu_W = \frac{n_1(N+1)}{2}$$

- Нахождение P-значения производим разными методами, лучше всего точным (exact)

$$\sigma_W = \sqrt{\frac{n_1 n_2 (N+1)}{12}}$$

Пример: урожай на двух площадках с сорняками и без

Weeds per meter	Yield (bu/acre)			
0	166.7	172.2	165.0	176.9
3	158.6	176.4	153.1	156.0

Yield	153.1	156.0	158.6	**165.0**	**166.7**	**172.2**	176.4	**176.9**
Rank	1	2	3	**4**	**5**	**6**	7	**8**

Treatment	Sum of ranks
No weeds	23
Weeds	13

Тест связанных рангов Уилкоксона

- Применяется для парных выборок
- Взять случайным образом пары выборок объемом n из генеральной совокупности, рассчитать разности значений между парами. Проранжировать эти разности
- Сумма рангов с положительными различиями W^+ является статистикой Уилкоксона для связанных рангов и имеет среднее и стандартное отклонение
- Нахождение P-значения производится разными методами. Лучше всего производить точный расчет (exact) в программе

$$\mu_{W^+} = \frac{n(n+1)}{4}$$

$$\sigma_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Пример: способность детей запоминать сказки с картинками и без

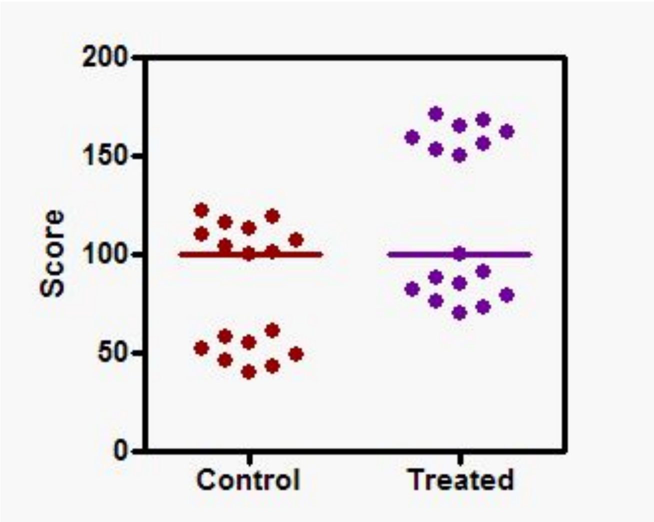
Child	1	2	3	4	5
Story 2	0.77	0.49	0.66	0.28	0.38
Story 1	0.40	0.72	0.00	0.36	0.55
Difference	0.37	−0.23	0.66	−0.08	−0.17

0.37 0.23 **0.66** 0.08 0.17

Absolute value	0.08	0.17	0.23	**0.37**	**0.66**
Rank	1	2	3	**4**	**5**

Что тестируют ранговые тесты?

- Ранговые тесты не тестируют различия в средних
- Можно трактовать как значимость различия медиан, но только в случае, если сравниваемые распределения идентичны, что крайне редко. Кроме того, так как ранговые тесты применяют для малых выборок, проверить это невозможно
- H_0 в данном случае формулируют как *отсутствие систематических различий между сравниваемыми группами*
- На примере представлены две выборки и медианы для них

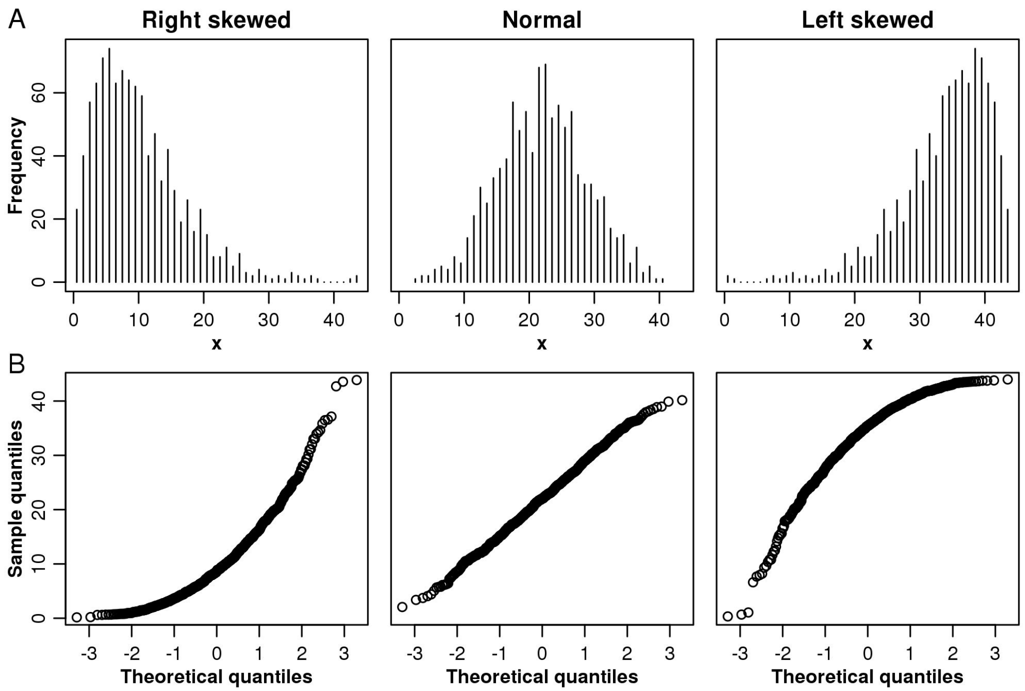


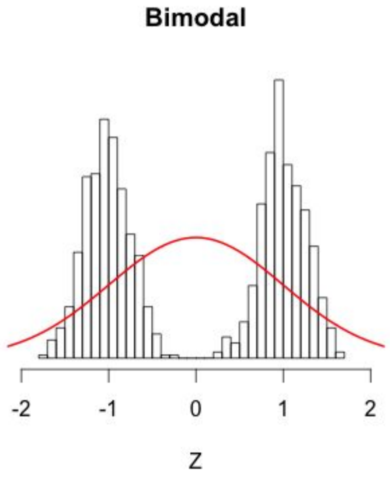
Методы расчета P-значения ранговых тестов

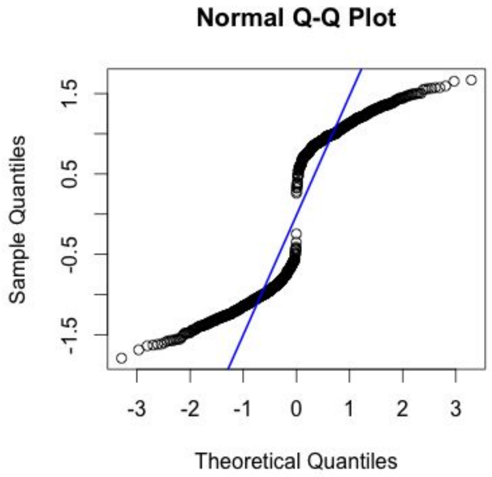
- Точный метод. Лучший. Перебор всех возможных комбинаций данных для построения точного распределения нужной статистики. При больших выборках не выполним из-за слишком большого времени вычисления
- Алгоритмы Монте Карло: пермутационные тесты сравнений и соответствующие им бутстреп – доверительные интервалы для любых статистик. По сути – перебор ограниченного случайного числа комбинаций. Позволяют получить P – значение с заданной точностью. Хороши для больших выборок
- Аппроксимация W статистики нормальным распределением – самый простой, но самый ненадежный способ. Тем не менее, зачастую хорошо работает при больших выборках

Как проверить нормальность распределения?

- Графики нормальных квантилей (Q-Q плоты, могут быть для разных распределений)
- Формальные тесты: их множество, но самые распространенные – это тесты Андерсона-Дарлинга и Шапиро-Уилка







Ошибки первого и второго рода

- Ошибка 1 рода происходит, когда мы отвергаем нулевую гипотезу (принимаем альтернативную), когда она правильная*
- Ошибка 2 рода происходит, когда мы не отвергаем нулевую гипотезу, когда альтернативная гипотеза правильная*

		Верная гипотеза	
		H_0	H_1
Результат применения критерия	H_0	H_0 верно принята	H_0 неверно принята (Ошибка *второго* рода)
	H_1	H_0 неверно отвергнута (Ошибка *первого* рода)	H_0 верно отвергнута

* Chihara Laura and Tim Hesterberg. *Mathematical statistics with resampling and R*. John Wiley & Sons, 2012.

Корректировка на множественные сравнения

- Если производить одновременно не одно, а много сравнений на уровне значимости α = 0.05, то вероятность ошибки первого рода становится выше 0.05
- Для корректировки α существует много различных методов
- Самые простые методы Бонферрони и Шидака
- Поправка Бонферрони: α’= α/m
- Поправка Шидака: $\alpha' = 1-(1-\alpha)^{1/m}$, где m – количество сравнений, а α’ – скорректированный уровень значимости для индивидуальных сравнений
- Количество попарных сравнений находят по формуле $\frac{N(N-1)}{2}$, где N – количество групп, которые нужно сравнить между собой

На следующем занятии

- Связь двух количественных переменных
- Меры связи количественных переменных: ковариация, коэффициенты корреляции Пирсона и Спирмена
- Введение в линейную регрессию. Коэффициенты линейной регрессии, проверка значимости
- Немного затронем робастные линейные оценки, устойчивые к выбросам