

Сравнение долей. Построение доверительных интервалов.

Используемая литература для подготовки задач к семинару:

Гусак А.А. «Теория вероятностей»

Гланц С. «Медико-биологическая статистика»



На препарате А положительный результат лечения наблюдается у 17 из 32 пациентов, а на препарате В у 9 из 22. Построить 95% доверительный интервал для разности долей. Обнаружены ли статистически значимые различия?



Решить задачу 1 через тестирование гипотезы.

На препарате А положительный результат лечения наблюдается у 17 из 32 пациентов, а на препарате В у 9 из 22. Являются ли различия статистически значимые между долями пациентов с положительным эффектом в этих двух группах.

Уровень статистической значимости принять за 0.05



Было проведено исследование научных статей на количество авторов в разные годы.
Построить 90% и 95% интервалы

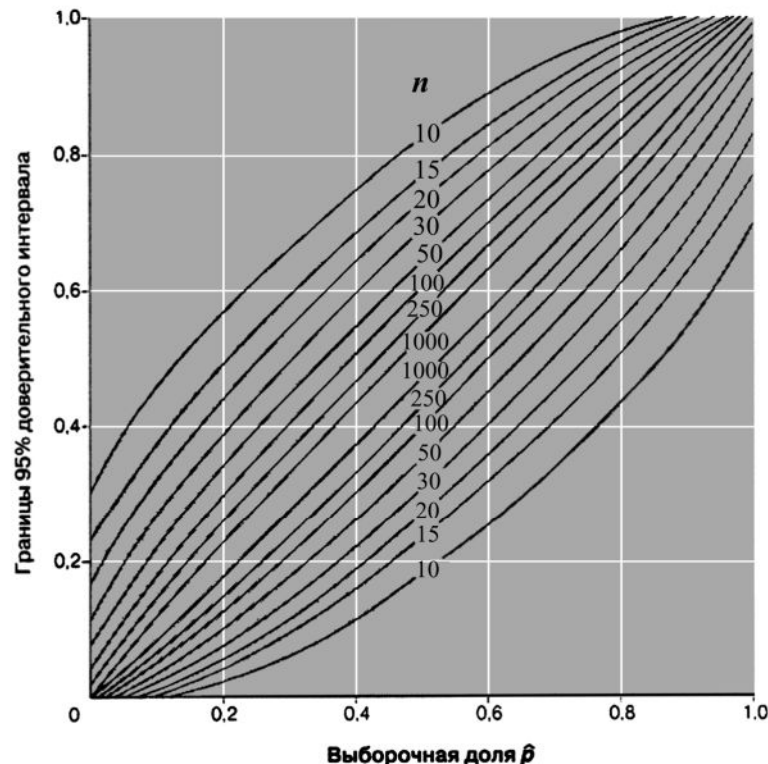
<i>Год</i>	<i>Число статей</i>	<i>Среднее число авторов</i>	<i>Стандартное отклонение</i>	<i>Интервал</i>
1946	151	2	1.4	
1956	149	2.3	1.6	
1966	157	2.8	1.2	
1976	155	4.9	7.3	



С помощью 90% доверительного интервала оценить средний вес нормально распределенной популяции, если дисперсия генеральной совокупности 3.6, а среднее арифметическое по выборке объемом 100 получилось равным 71.2.

Найдите 95% доверительные интервалы для долей больных, которые не чувствовали боли при включенном и выключенном приборе. Можно ли по этим интервалам оценить статистическую значимость различий?

	Прибор включен	Прибор выключен
Боли нет	24	3
Боль есть	6	17





Даны две группы студентов. В каждой по 21 человеку. Первая группа прошла недельную подготовку для сдачи тестов, другая группа приступила к тестам без подготовки. В 1й группе среднее время прохождения теста целиком требует 1.8 ч, во второй 2.3 ч. Постройте 95% интервал для разности средних арифметических, если известно, что стандартное отклонение 1 группы составило 0.4 ч, во 2й группе -0.6ч



Таблица 4.1. Критические значения t (двусторонний вариант)

v	Уровень значимости α								
	0,5	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	1,000	3,078	6,314	12,706	31,821	63,656	127,321	318,289	636,578
2	0,816	1,886	2,920	4,303	6,965	9,925	14,089	22,328	31,600
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,214	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,689
28	0,683	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,660
30	0,683	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
31	0,682	1,309	1,696	2,040	2,453	2,744	3,022	3,375	3,633
32	0,682	1,309	1,694	2,037	2,449	2,738	3,015	3,365	3,622
33	0,682	1,308	1,692	2,035	2,445	2,733	3,008	3,356	3,611
34	0,682	1,307	1,691	2,032	2,441	2,728	3,002	3,348	3,601
35	0,682	1,306	1,690	2,030	2,438	2,724	2,996	3,340	3,591
36	0,681	1,306	1,688	2,028	2,434	2,719	2,990	3,333	3,582
37	0,681	1,305	1,687	2,026	2,431	2,715	2,985	3,326	3,574
38	0,681	1,304	1,686	2,024	2,429	2,712	2,980	3,319	3,566
39	0,681	1,304	1,685	2,023	2,426	2,708	2,976	3,313	3,558
40	0,681	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551



Расчет критерия Манна – Уитни U

1. Берем 1й элемент из x1 и ставим его в начало x2

47 58 60 77

2. Присваиваем ранги этим значениям

1 2 3 4

3. Берем из массива рангов первое значение и вычитаем из этого значения 1

$1 - 1 = 0$

Теперь повторяем действия со 2м и 3м элементом из массива x1

1. 90 58 60 77

2. 4 1 2 3

3. $4 - 1 = 3$

И последний раз повторим действия для 75 (3й элемент)

1. 75, 58, 60, 77

2. 3 1 2 4

3. $3 - 1 = 2$

Сложим значения, которые мы получали в п.3

$0 + 3 + 2 = 5$ Это и будет расчетный статистик

```
import numpy as np
import scipy.stats as stats

x1= np.array([47, 90, 75])

x2 = np.array([ 58, 60, 77])

stats.mannwhitneyu(x1, x2)
MannwhitneyuResult(statistic=5.0, pvalue=1.0)
```



Рассчитать критерий в ручную

Рассмотрим пример из книги Стентона Гланца

```
group_1= np.array ([1000, 1380, 1200])
```

```
group_2 = np.array ([1400, 1600, 1180, 1220])
```

```
stats.mannwhitneyu(group_1, group_2)
```

```
MannwhitneyuResult(statistic=3.0, pvalue=0.4)
```



1000 1400 1600 1180 1220

1 4 5 2 3

$1-1=0$

1380 1400 1600 1180 1220

3 4 5 1 2

$3-1=2$

1200 1400 1600 1180 1220

2 4 5 1 3

$2-1=1$

$0+2+1=3$



Средние расходы на обследование одного больного до и после ознакомления с расходами коллег:

$X = \text{np.array}([20, 17, 14, 42, 50, 62, 8, 49, 81, 54, 48, 55, 56])$

$Y = \text{np.array}([20, 26, 1, 24, 1, 47, 15, 7, 65, 9, 21, 36, 30])$

Произошли ли изменения?

до	после	Delta	Ранг
20	20		
17	26		
14	1		
42	24		
50	1		
62	47		
8	15		
49	7		
81	65		
54	9		
48	21		
55	36		
56	30		



Средние расходы на лечение одного больного до и после ознакомления с расходами коллег

$X = \text{np.array}([32, 41, 51, 29, 76, 47, 60, 58, 40, 64, 73, 66, 73])$

$Y = \text{np.array}([42, 90, 71, 47, 56, 43, 137, 63, 28, 60, 87, 69, 50])$.

Произошли ли изменения?

до	после	Delta	Ранг
32	42		
41	90		
51	71		
29	47		
76	56		
47	43		
60	137		
58	63		
40	28		
64	60		
73	87		
66	69		
73	50		



```
X= np.array([20,17, 14, 42, 50, 62, 8, 49, 81, 54, 48, 55, 56])
Y= np.array ([20, 26, 1, 24, 1, 47, 15, 7, 65, 9, 21, 36, 30])

stats.wilcoxon(X,Y)
WilcoxonResult(statistic=3.0, pvalue=0.004741768038406972)
```



Даны значения проницаемости сосудов сетчатки gr1(здоровые пациенты), gr 2 (поражения в области центральной ямки), gr 3 (в области центральной ямки и на периферии)

```
gr1=np.array([0.5, 0.7, 1, 1.2, 1.4])
```

```
gr2= np.array([1.3, 1.45, 1.6, 1.7, 1.8])
```

```
gr3=np.array([6.2, 12.6, 13.2, 14.1, 14.2])
```

Сравнить данные, относящиеся к разным видам поражений



При исследовании препарата для снижения кровяного давления у больных 3 раза измерялся сердечный выброс. Менялся ли сердечный выброс? Найти критерий в ручную, проверьте значение функцией и интерпретируйте результат с использованием p-value

Больной	Измерение		
	1	2	3
1	3,5	8,6	5,1
2	3,3	5,4	8,6
3	4,9	8,8	6,7
4	3,6	5,6	5,0



Задачи на повторение