

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ: СЛУЧАЙ ИЗВЕСТНЫХ И РАВНЫХ ДИСПЕРСИЙ

Пусть из генеральных совокупностей с известными стандартными отклонениями $\sigma_x = \sigma_y = 10$ извлечены выборки объемом $N_1 = N_2 = 100$.

Пусть стандартные отклонения помещены в ячейки A1 и B1, варианты заполняют строки со второй по сто первую (ячейки A2:A101 и B2:B101, соответственно).

Найдем выборочные средние. В ячейку C1 введем формулу

$$=СУММ(A2:A101)/100$$
 и переместим маркер автозаполнения до ячейки D1.

	A	B	C	D
1	10	10	=СУММ(A2:A101)/100	
2	96,99768	110,8106		

Вычислим значение связанной с разностью средних статистики, подчиненной стандартному нормальному распределению. В ячейку G3 введем

$$=(C1-D1)/((A1^2/100+B1^2/100)^{0,5})$$

	A	B	C	D	E	F	G	H	I
1	10	10	99,59515	102,9814					
2	96,99768	110,8106							
3	87,22317	103,1786					Статистика: =(C1-D1)/((A1^2/100+B1^2/100)^0,5)		

Вероятность критического события можно найти, воспользовавшись функцией рабочего листа НОРМСТРАСП, возвращающей функцию распределения стандартного нормального закона. В ячейку G4 введем

$$=2-2*НОРМСТРАСП(ABS(G3))$$

	A	B	C	D	E	F	G	H	I
1	10	10	99,59515	102,9814					
2	96,99768	110,8106							
3	87,22317	103,1786					Статистика: -2,39441		
4	102,4426	97,91674					Вероятность критического события: =2-2*НОРМСТРАСП(ABS(G3))		

Окончательный результат:

	A	B	C	D	E	F	G
1	10	10	99,59515	102,9814			
2	96,99768	110,8106					
3	87,22317	103,1786					Статистика: -2,39441
4	102,4426	97,91674					Вероятность критического события: 0,016647

Пусть уровень значимости выбран равным $\alpha = 0,05$. В данном примере вероятность критического события оказалась меньше уровня значимости

$$P(A) \approx 0,017 < 0,05,$$

поэтому на уровне значимости $\alpha = 0,05$ гипотеза о равенстве математических ожиданий должна быть отвергнута.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ: СЛУЧАЙ НЕИЗВЕСТНЫХ РАВНЫХ ДИСПЕРСИЙ

Пусть из генеральных совокупностей извлечены выборки объемом $N_1 = N_2 = 100$. Пусть варианты заполняют строки с первой по сотую (ячейки A1:A100 и B1:B100, соответственно).

Для нахождения выборочных средних и выборочных дисперсий можно воспользоваться пакетом анализа. Из меню Сервис следует выбрать Анализ данных, далее – Описательная статистика. В качестве входного интервала следует указать два первых столбца (ячейки \$A\$1:\$B\$100). Результаты анализа можно поместить в ячейки начиная с C1; следует также установить флажок Описательная статистика.

	A	B	C	D	E	F	G	H
1	96,99768	110,8106						
2	87,22317	103,1786						
3	102,4426	97,91674						
4	112,7647	95,4788						
5	111,9835	98,36867						
6	117,3313	104,3061						
7	78,16412	104,1827						
8	97,65819	97,46845						
9	110,9502	81,56139						
10	89,13299	86,81963						
11	93,09796	104,5031						
12	83,09568	116,0176						
13	81,53089	106,3167						
14	90,22371	106,1668						
15	92,26493	97,17089						
16	78,82069	100,9208						
17	94,32075	91,59398						
18	95,95952	112,1863						
19	101,3485	94,81334						
20	96,34507	99,11591						

Описательная статистика ? x

Входные данные

Входной интервал:

Группирование: по столбцам по строкам

Метки в первой строке

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Итоговая статистика

Уровень надежности: %

К-ый наименьший:

К-ый наибольший:

OK Отмена Справка

Найдем смешанную оценку

$$D[\bar{x} - \bar{y}] = \frac{1}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \left(\sum_{i=1}^{N_1} (x_i - \bar{x})^2 + \sum_{i=1}^{N_2} (y_i - \bar{y})^2 \right)$$

дисперсии разности выборочных средних. В ячейку G16 введем

$$=1/(D15+F15-2) * (1/D15+1/F15) * ((D15-1)*D8+(F15-1)*F8)$$

8	97,65819	97,46845	Дисперси	117,8588	Дисперси	98,5838		
9	110,9502	81,56139	Эксцесс	-0,47571	Эксцесс	0,514209		
10	89,13299	86,81963	Асимметр	0,090701	Асимметр	0,182492		
11	93,09796	104,5031	Интервал	49,53235	Интервал	54,74467		
12	83,09568	116,0176	Минимум	74,22419	Минимум	79,41971		
13	81,53089	106,3167	Максимум	123,7565	Максимум	134,1644		
14	90,22371	106,1668	Сумма	9959,515	Сумма	10298,14		
15	92,26493	97,17089	Счет	100	Счет	100		
16	78,82069	100,9208						

Для вычисления статистики

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{D[\bar{x} - \bar{y}]}}$$

в ячейку G17 введем

$$=(D3-F3)/G16^0,5$$

	A	B	C	D	E	F	G	H
1	96,99768	110,8106	Столбец1		Столбец2			
2	87,22317	103,1786						
3	102,4426	97,91674	Среднее	99,59515	Среднее	102,9814		
4	112,7647	95,4788	Стандарт	1,085628	Стандарт	0,992894		
5	111,9835	98,36867	Медиана	99,15099	Медиана	102,7777		
6	117,3313	104,3061	Мода	#Н/Д	Мода	#Н/Д		
7	78,16412	104,1827	Стандарт	10,85628	Стандарт	9,928938		
8	97,65819	97,46845	Дисперси	117,8588	Дисперси	98,5838		
9	110,9502	81,56139	Эксцесс	-0,47571	Эксцесс	0,514209		
10	89,13299	86,81963	Асимметр	0,090701	Асимметр	0,182492		
11	93,09796	104,5031	Интервал	49,53235	Интервал	54,74467		
12	83,09568	116,0176	Минимум	74,22419	Минимум	79,41971		
13	81,53089	106,3167	Максимум	123,7565	Максимум	134,1644		
14	90,22371	106,1668	Сумма	9959,515	Сумма	10298,14		
15	92,26493	97,17089	Счет	100	Счет	100		
16	78,82069	100,9208						
17	94,32075	91,59398						

Вычислим вероятность критического события, состоящего в том, что при верной нулевой гипотезе (при равных неизвестных истинных математических ожиданиях) значение статистики по модулю окажется не меньшим, чем реально полученное. В ячейку G18 введем

$$=СТЮДРАСП(ABS(G17);D15+F15-2;2)$$

	A	B	C	D	E	F	G	H	I	J
1	96,99768	110,8106	Столбец1		Столбец2					
2	87,22317	103,1786								
3	102,4426	97,91674	Среднее	99,59515	Среднее	102,9814				
4	112,7647	95,4788	Стандарт	1,085628	Стандарт	0,992894				
5	111,9835	98,36867	Медиана	99,15099	Медиана	102,7777				
6	117,3313	104,3061	Мода	#Н/Д	Мода	#Н/Д				
7	78,16412	104,1827	Стандарт	10,85628	Стандарт	9,928938				
8	97,65819	97,46845	Дисперси	117,8588	Дисперси	98,5838				
9	110,9502	81,56139	Эксцесс	-0,47571	Эксцесс	0,514209				
10	89,13299	86,81963	Асимметр	0,090701	Асимметр	0,182492				
11	93,09796	104,5031	Интервал	49,53235	Интервал	54,74467				
12	83,09568	116,0176	Минимум	74,22419	Минимум	79,41971				
13	81,53089	106,3167	Максимум	123,7565	Максимум	134,1644				
14	90,22371	106,1668	Сумма	9959,515	Сумма	10298,14				
15	92,26493	97,17089	Счет	100	Счет	100				
16	78,82069	100,9208	Смешанная оценка дисперсии:			2,164426				
17	94,32075	91,59398	Статистика:			-2,30167				
18	95,95952	112,1863	Вероятность критического события:			=СТЮДРАСП(ABS(G17);D15+F15-2;2)				

Окончательный результат:

16	78,82069	100,9208	Смешанная оценка дисперсии:	2,164426
17	94,32075	91,59398	Статистика:	-2,30167
18	95,95952	112,1863	Вероятность критического события:	0,022394

Найденная в примере вероятность критического события

$$P(A) \approx 0,022 < 0,05,$$

поэтому на выбранном уровне значимости $\alpha = 0,05$ гипотеза о равенстве математических ожиданий должна быть отвергнута.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ МАТЕМАТИЧЕСКИХ ОЖИДАНИЙ: СЛУЧАЙ НЕИЗВЕСТНЫХ ДИСПЕРСИЙ

Все действия до вычисления статистики

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{D[\bar{x} - \bar{y}]}}$$

включительно аналогичны действиям предыдущего пункта. Следует лишь принять во внимание, что число степеней свободы вычисляется не как разность $k = N_1 + N_2 - 2$, а более сложным образом:

$$k = \frac{\left(\frac{s_x}{N_1} + \frac{s_y}{N_2} \right)^2}{\frac{s_x^2}{N_1^2(N_1 - 1)} + \frac{s_y^2}{N_2^2(N_2 - 1)}}.$$

Вычислим это значение в ячейке G18:

$$=(D4/D15+F4/F15)^2/(D4^2/(D15^2*(D15-1))+F4^2/(F15^2*(F15-1)))$$

	A	B	C	D	E	F	G	H	I	J	K	L
1	96,99768	110,8106	Столбец1	Столбец2								
2	87,22317	103,1786										
3	102,4426	97,91674	Среднее	99,59515	Среднее	102,9814						
4	112,7647	95,4788	Стандартн	1,085628	Стандартн	0,992894						
5	111,9835	98,36867	Медиана	99,15099	Медиана	102,7777						
6	117,3313	104,3061	Мода	#Н/Д	Мода	#Н/Д						
7	78,16412	104,1827	Стандартн	10,85628	Стандартн	9,928938						
8	97,65819	97,46845	Дисперсия	117,8588	Дисперсия	98,5838						
9	110,9502	81,56139	Эксцесс	-0,47571	Эксцесс	0,514209						
10	89,13299	86,81963	Асимметр	0,090701	Асимметр	0,182492						
11	93,09796	104,5031	Интервал	49,53235	Интервал	54,74467						
12	83,09568	116,0176	Минимум	74,22419	Минимум	79,41971						
13	81,53089	106,3167	Максимум	123,7565	Максимум	134,1644						
14	90,22371	106,1668	Сумма	9959,515	Сумма	10298,14						
15	92,26493	97,17089	Счет	100	Счет	100						
16	78,82069	100,9208		Смешанная оценка дисперсии:		2,164426						
17	94,32075	91,59398		Статистика:		-2,30167						
18	95,95952	112,1863		Число степеней свободы:		362,8961						
19	101,3485	94,81334		Вероятность критического события:		=СТЮДРАСП(ABS(G17);G18;2)						

Вероятность критического события вычислим в ячейке G19:

=СТЮДРАСП (ABS (G17) ; G18 ; 2)

	A	B	C	D	E	F	G	H	I
1	96,99768	110,8106	Столбец1	Столбец2					
2	87,22317	103,1786							
3	102,4426	97,91674	Среднее	99,59515	Среднее	102,9814			
4	112,7647	95,4788	Стандартн	1,085628	Стандартн	0,992894			
5	111,9835	98,36867	Медиана	99,15099	Медиана	102,7777			
6	117,3313	104,3061	Мода	#Н/Д	Мода	#Н/Д			
7	78,16412	104,1827	Стандартн	10,85628	Стандартн	9,928938			
8	97,65819	97,46845	Дисперсия	117,8588	Дисперсия	98,5838			
9	110,9502	81,56139	Эксцесс	-0,47571	Эксцесс	0,514209			
10	89,13299	86,81963	Асимметр	0,090701	Асимметр	0,182492			
11	93,09796	104,5031	Интервал	49,53235	Интервал	54,74467			
12	83,09568	116,0176	Минимум	74,22419	Минимум	79,41971			
13	81,53089	106,3167	Максимум	123,7565	Максимум	134,1644			
14	90,22371	106,1668	Сумма	9959,515	Сумма	10298,14			
15	92,26493	97,17089	Счет	100	Счет	100			
16	78,82069	100,9208		Смешанная оценка дисперсии:		2,164426			
17	94,32075	91,59398		Статистика:		-2,30167			
18	95,95952	112,1863		Число степеней свободы:		362,8961			
19	101,3485	94,81334		Вероятность критического события:		=СТЮДРАСП(ABS(G17);G18;2)			

	A	B	C	D	E	F	G
1	96,99768	110,8106	Столбец1	Столбец2			
2	87,22317	103,1786					
3	102,4426	97,91674	Среднее	99,59515	Среднее	102,9814	
4	112,7647	95,4788	Стандартн	1,085628	Стандартн	0,992894	
5	111,9835	98,36867	Медиана	99,15099	Медиана	102,7777	
6	117,3313	104,3061	Мода	#Н/Д	Мода	#Н/Д	
7	78,16412	104,1827	Стандартн	10,85628	Стандартн	9,928938	
8	97,65819	97,46845	Дисперсия	117,8588	Дисперсия	98,5838	
9	110,9502	81,56139	Эксцесс	-0,47571	Эксцесс	0,514209	
10	89,13299	86,81963	Асимметр	0,090701	Асимметр	0,182492	
11	93,09796	104,5031	Интервал	49,53235	Интервал	54,74467	
12	83,09568	116,0176	Минимум	74,22419	Минимум	79,41971	
13	81,53089	106,3167	Максимум	123,7565	Максимум	134,1644	
14	90,22371	106,1668	Сумма	9959,515	Сумма	10298,14	
15	92,26493	97,17089	Счет	100	Счет	100	
16	78,82069	100,9208		Смешанная оценка дисперсии:		2,164426	
17	94,32075	91,59398		Статистика:		-2,30167	
18	95,95952	112,1863		Число степеней свободы:		362,8961	
19	101,3485	94,81334		Вероятность критического события:		0,021921	

В рассмотренном примере вероятности критического события оказались близкими вне зависимости от предположения о равенстве дисперсий; и в том, и в другом случае на уровне значимости $\alpha = 0,05$ гипотеза о равенстве математических ожиданий должна быть отвергнута.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДИСПЕРСИЙ

Пусть требуется сравнить дисперсии двух генеральных совокупностей на основе извлеченных выборок объемами $N_1 = N_2 = 20$. Пусть варианты помещены в первые два столбца и заполняют строки со второй по двадцатую (ячейки A2:A20 и B2:B101, соответственно).

Для нахождения оценок дисперсий воспользуемся средствами пакета анализа (аналогично предыдущей задаче). Из меню Сервис следует выбрать Анализ данных, далее – Описательная статистика. В качестве входного интервала следует указать два первых столбца (ячейки \$A\$1:\$B\$20). Результаты анализа можно поместить начиная с ячейки C1; установить флажок Описательная статистика.

Вычислим подчиненную F -распределению статистику, характеризующую различие между дисперсиями. В ячейку G16 поместим формулу

$$=D8/F8$$

8	-97,1876	78,76266	Дисперси	129,2361	Дисперси	114,2354	
9	-92,9408	81,49287	Эксцесс	0,291126	Эксцесс	0,114196	
10	-101,148	98,96504	Асимметр	-0,80963	Асимметр	-0,80253	
11	-94,6158	96,11427	Интервал	43,41709	Интервал	37,67082	
12	-109,395	113,1854	Минимум	-125,431	Минимум	75,51458	
13	-84,5413	109,8481	Максимум	-82,0142	Максимум	113,1854	
14	-111,39	111,3724	Сумма	-1960,64	Сумма	1972,348	
15	-97,1629	110,9127	Счет	20	Счет	20	
16	-125,431	96,70991			Статистика:	=D8/F8	

	A	B
1	-93,8348	75,51458
2	-115,126	97,10953
3	-92,3503	87,93137
4	-91,7102	103,9361
5	-88,7351	98,03636
6	-82,0142	108,3103
7	-111,252	102,4355
8	-97,1876	78,76266
9	-92,9408	81,49287
10	-101,148	98,96504
11	-94,6158	96,11427
12	-109,395	113,1854
13	-84,5413	109,8481
14	-111,39	111,3724
15	-97,1629	110,9127
16	-125,431	96,70991
17	-91,6994	103,162
18	-100,871	96,92318
19	-83,4188	100,4649
20	-95,8188	101,1612

Для вычисления величины

$$P' = 2Q\left(\frac{\sigma_x^2}{\sigma_y^2}, N_1 - 1, N_2 - 1\right)$$

в ячейку G17 поместим формулу

$$=FРАСП(G16;D15-1;F15-1)$$

14	-111,39	111,3724	Сумма	-1960,64	Сумма	1972,348	
15	-97,1629	110,9127	Счет	20	Счет	20	
16	-125,431	96,70991			Статистика:	1,131314	
17	-91,6994	103,162			Односторонняя вероятность:	=FРАСП(G16;D15-1;F15-1)	

Для нахождения вероятности критического события в ячейку G18 введем

$$=ЕСЛИ(G17<1;G17;2-G17)$$

16	-125,431	96,70991			Статистика:	1,131314	
17	-91,6994	103,162			Односторонняя вероятность:	0,395388	
18	-100,871	96,92318			Вероятность критического события:	=ЕСЛИ(G17<1;G17;2-G17)	

Окончательный результат:

	A	B	C	D	E	F	G
1	-93,8348	75,51458	Столбец1		Столбец2		
2	-115,126	97,10953					
3	-92,3503	87,93137	Среднее	-98,0321	Среднее	98,61741	
4	-91,7102	103,9361	Стандарт	2,542008	Стандарт	2,389931	
5	-88,7351	98,03636	Медиана	-95,2173	Медиана	99,71496	
6	-82,0142	108,3103	Мода	#Н/Д	Мода	#Н/Д	
7	-111,252	102,4355	Стандарт	11,36821	Стандарт	10,68809	
8	-97,1876	78,76266	Дисперси	129,2361	Дисперси	114,2354	
9	-92,9408	81,49287	Эксцесс	0,291126	Эксцесс	0,114196	
10	-101,148	98,96504	Асимметр	-0,80963	Асимметр	-0,80253	
11	-94,6158	96,11427	Интервал	43,41709	Интервал	37,67082	
12	-109,395	113,1854	Минимум	-125,431	Минимум	75,51458	
13	-84,5413	109,8481	Максимум	-82,0142	Максимум	113,1854	
14	-111,39	111,3724	Сумма	-1960,64	Сумма	1972,348	
15	-97,1629	110,9127	Счет	20	Счет	20	
16	-125,431	96,70991			Статистика:	1,131314	
17	-91,6994	103,162			Односторонняя вероятность:	0,395388	
18	-100,871	96,92318			Вероятность критического события:	0,395388	

В данном примере $P(A) \approx 0,4 > 0,05$, поэтому на уровне значимости $\alpha = 0,05$ нет оснований отвергать гипотезу о равенстве дисперсий в пользу двусторонней альтернативы.