此笔记为excel数据分析功能详解

笔记本: 我的第一个笔记本

创建时间: 2020/1/11 23:21 **更新时间**: 2020/1/12 13:18

作者: 羊熊

URL: https://www.jianshu.com/p/1cf9cef9ede5

此笔记为excel数据分析功能详解

单因素方差分析

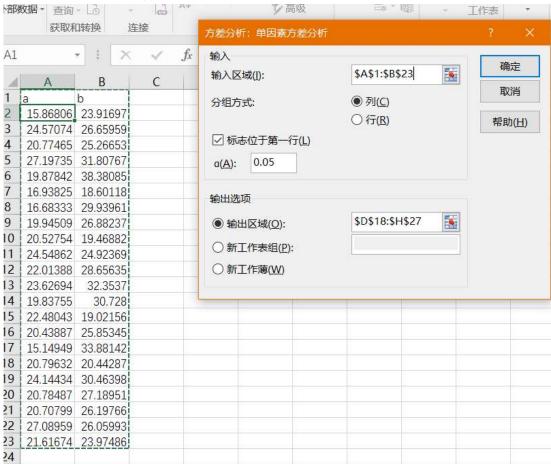
意义:

测试多列数据之间的差异显著度。差异主要指均值的差异。

h0: 差异显著 h1: 差异不显著

设置:

输入区域为两列或多列数据



结果为差异显著

看f值>f crit

p值<0.05就说明很显著了

SUMMAR'	Υ					
组	观测数	求和	平均	方差		
а	22	465.6191	21.1645	10.51458		
b	22	590.6706	26.84866	25.15819		
方差分析				_		
差异源	SS	df	MS	F	P-value	F crit
组间	355.4064	1	355.4064	19.92592	5.94E-05	4.072654
组内	749.1282	42	17.83639			
总计	1104.535	43				

无重复双因素分析

意义:

测试多行及多列数据间的差异显著度。

设置:

A	В	С	D	Е	F	G	H	I	J
	地区1	地区2	地区3	地区4	地区5				
品牌1	365	350	343	340	323				
品牌2	345	368	363	330	333				
品牌3	358	323	353	343	308				
品牌4	288	280	298	260	298				
				方差分析	:无重复双因	表分析			? ×
				输入 输入区 ¹	或([):	i i	\$A\$1:\$F\$5		确定
				☑标志				1	取消
				a(<u>A</u>):	0.05				帮助(<u>H</u>)
				输出选I	Д				
				● 輸出	·区域(<u>O</u>):		\$E\$6:\$J\$18		
				○新工	作表组(P):				
				○新工	作薄(<u>W</u>)				
					_				_

结果:

行的p值<0.05: 列差异显著 列的p值>0.05: 列差异不显著

方差分析	: 无重复双	因素分析				
SUMMARY	观测数	求和	平均	方差		
品牌1	5	1721	344. 2	233. 7		
品牌2	5	1739	347.8	295.7		
品牌3	5	1685	337	442.5		
品牌4	5	1424	284. 8	249. 2		
地区1	4	1356	339	1224. 667		
地区2	4	1321	330. 25	1464. 25		
地区3	4	1357	339. 25	822. 9167		
地区4	4	1273	318. 25	1538. 917		
地区5	4	1262	315. 5	241. 6667		
方差分析						
差异源	SS	df	MS	F	P-value	F crit
行	13004. 55	3	4334. 85	18. 10777	9. 46E-05	3. 490295
列	2011. 7	4	502. 925	2.100846	0.143665	3. 259167
误差	2872. 7	12	239. 3917			
总计	17888. 95	19				

可重复双因素分析

意义:

测试行和列对数据的影响显著度, 以及行列之间的交互

设置:



结果:

行和列的p值都很小,说明行间和列间差异都显著。 交互作用p值很大,说明行列之间交互作用很小。

观测数	3	3	6			
求和	12114.75	11925.84	24040.59			
平均	4038.249	3975.28	4006.765			
方差	1704.53	700.311	2151.488			
总ì	+					
观测数	15	15				
求和	60239	59294.55				
平均	4015.933	3952.97				
方差	921.2694	1222.76				
方差分析						
差异源	SS	df	MS	F	P-value	F crit
行	10901.61	4	2725.403	3.002395	0.043089	2.866081
列	29732.96	1	29732.96	32.75482	1.34E-05	4.351244
交互	959.9393	4	239.9848	0.264375	0.897343	2.866081
内部	18154.86	20	907.7431			
总计	59749.37	29				

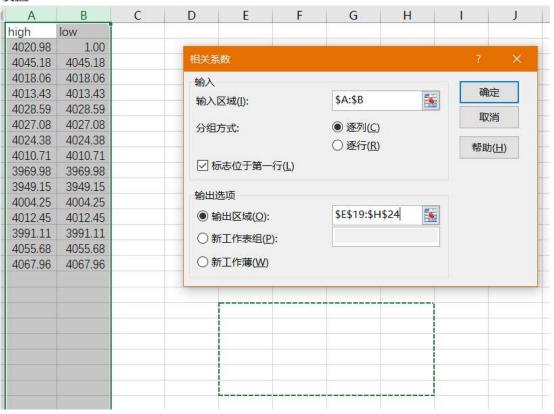
相关系数

意义:

测试两个样本之间的线性相关程度。

值为-1到1。

设置:



结果:

非常相关, 0.688的相关度。

4	Α	В	C	D	E	F	G	
	high	low						
	4020.98	3963.675						
	4045.18	3992.174						
	4018.06	3966.414						
	4013.43	3974.812						
	4028.59	3980.863						
	4027.08	3917.026						
	4024.38	3892.523						
	4010.71	3967.673						
P	3969.98	3897.352						
	3949.15	3900.244						
	4004.25	3952.878						
9	4012.45	3963.074						
	3991.11	3946.386						
00	4055.68	3981.114						
	4067.96	3998.339						
53								
8								
Ü						high	low	
ß					high	1		
					low	0.688146	1	
					11	The state of the s	(/	

协方差

意义:

两个样本之间的相关度。

设置:



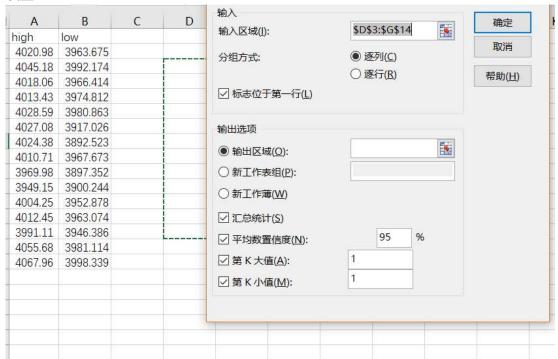
high	low				
4020.98	3963.675				
4045.18	3992.174				
4018.06	3966.414				
4013.43	3974.812				
4028.59	3980.863				
4027.08	3917.026				
4024.38	3892.523				
4010.71	3967.673				
3969.98	3897.352				
3949.15	3900.244				
4004.25	3952.878				
4012.45	3963.074				
3991.11	3946.386				
4055.68	3981.114				
4067.96	3998.339				
				high	low
			high	859.8514	
			low	681.681	1141.243

描述统计

意义:

单个样本的各项参数。

设置:



结果:

1	V		
high	low		
4020.98	3963.675		
4045.18	3992.174		
4018.06	3966.414	hig	gh
4013.43	3974.812		
4028.59	3980.863	平均	4015.933
4027.08	3917.026	标准误差	7.836961
4024.38	3892.523	中位数	4018.058
4010.71	3967.673	众数	#N/A
3969.98	3897.352	标准差	30.35242
3949.15	3900.244	方差	921.2694
4004.25	3952.878	峰度	0.807836
4012.45	3963.074	偏度	-0.52811
3991.11	3946.386	区域	118.8103
4055.68	3981.114	最小值	3949.15
4067.96	3998.339	最大值	4067.961
		求和	60239
		观测数	15
		最大(1)	4067.961
		最小(1)	3949.15
		置信度(95	16.80861

指数平滑

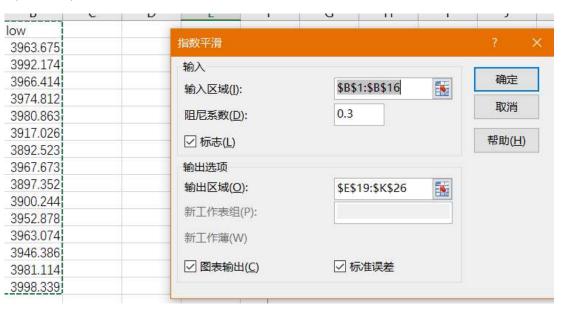
意义:

相当于股票指标的ema。

设置:

每一行ema相当于

(1 - 阻尼) * 新数据 + 阻尼 * 上一个ema



结果:



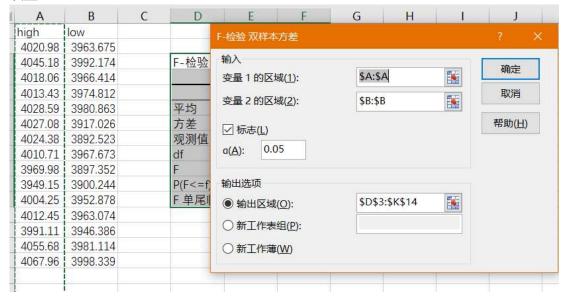
f检验,双样本方差

意义:

测试两个样本方差的差异显著度。

如果差异显著,之后t检验就最好用异方差t检验。否则就最好用等方差t检验。

设置:



结果:

p值大于0.05,说明方差差异不是很显著。

high	low				
4020.98	3963.675				
4045.18	3992.174	F-检验 双	样本方差分	析	
4018.06	3966.414		\$ 0.00 CA	0123	
4013.43	3974.812		high	low	
4028.59	3980.863	平均	4015.933	3952.97	
4027.08	3917.026	方差	921.2694	1222.76	
4024.38	3892.523	观测值	15	15	
4010.71	3967.673	df	14	14	
3969.98	3897.352	F	0.753434		
3949.15	3900.244	P(F<=f)	0.301726		
4004.25	3952.878	F单尾临界	0.402621		
4012.45	3963.074	22 (340.8)			
3991.11	3946.386				
4055.68	3981.114				
4067.96	3998.339				

傅里叶分析

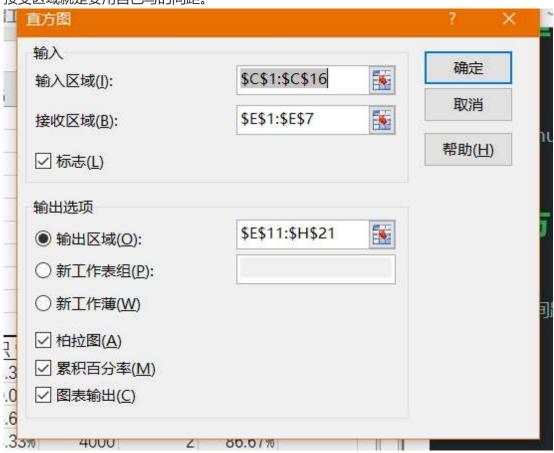
直方图

意义:

自己设置间距,来画直方图

设置:

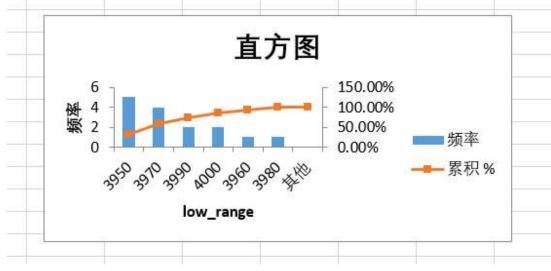
接受区域就是要用自己写的间距。



结果:

频率是按照 <=每个间距的数据来统计的。比如下图中low_range ==3950的频率,统计的是小于3950的数字总数。

low_range	频率	累积%	low_range	频率	累积%
3950	5	33.33%	3950	5	33.33%
3960	1	40.00%	3970	4	60.00%
3970	4	66.67%	3990	2	73.33%
3980	1	73.33%	4000	2	86.67%
3990	2	86.67%	3960	1	93.33%
4000	2	100.00%	3980	1	100.00%
其他	0	100.00%	其他	0	100.00%



移动平均

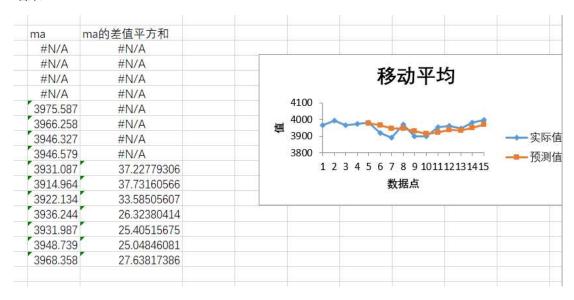
意义:

相当于股票的ma。

设置:

间隔相当于几日均线。





随机数发生器

意义:

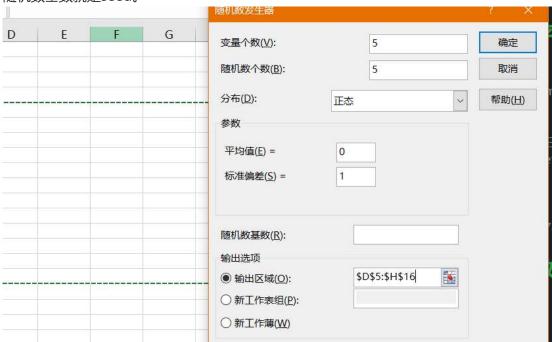
生成随机数

设置:

变量个数为每次随机几个变量。

随机数个数为生成几次。

随机数基数就是seed。



结果:

5 * 5 = 25个标准正态分布随机数

	-1.07804			
).464581	1.729713	-0.55013	2.067391	0.657562
).710735	1.111471	-0.41661	-0.54391	0.300312
).471409	-1.52572	0.373274	1.171338	-0.69838
-0.12768	1.889384	-0.28068	-0.09873	1.496728

排位与百分比排位

意义:

排位

设置:

俞入		
俞入区域(<u>l</u>):	\$C\$1:\$C\$16	确定
分组方式:	● 列(C)	取消
3-22-3-4-	○ 行(<u>R</u>)	帮助(<u>H</u>)
☑ 标志位于第一行(L)		
俞出选项		
動 輸出区域(○):	\$E\$5:\$I\$17	5
○ 新工作表组(<u>P</u>):		
○ 新工作薄(<u>W</u>)		

结果:

百分比排位为: 小于该值的个数/(小于该值的个数+大于该值的个数)

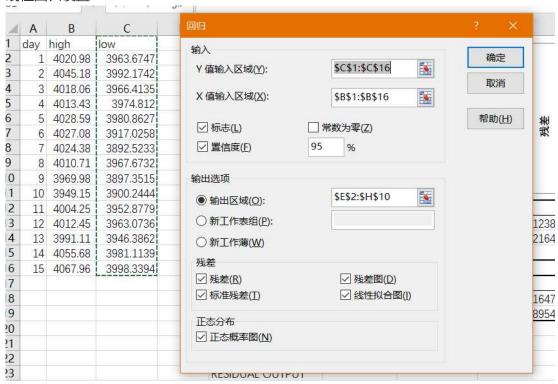
Α	В	С	D	E	F	G	Н
day	high	low					
1	4020.98	3963.675					
2	4045.18	3992.174					
3	4018.06	3966.414					
4	4013.43	3974.812		点	low	排位	百分比
5	4028.59	3980.863		15	3998.339	1	100.00%
6	4027.08	3917.026		2	3992.174	2	92.80%
7	4024.38	3892.523		14	3981.114	3	85.70%
8	4010.71	3967.673		5	3980.863	4	78.50%
9	3969.98	3897.352		4	3974.812	5	71.40%
10	3949.15	3900.244		8	3967.673	6	64.20%
11	4004.25	3952.878		3	3966.414	7	57.10%
12	4012.45	3963.074		1	3963.675	8	50.00%
13	3991.11	3946.386		12	3963.074	9	42.80%
14	4055.68	3981.114		11	3952.878	10	35.70%
15	4067.96	3998.339		13	3946.386	11	28.50%
				6	3917.026	12	21.40%
				10	3900.244	13	14.20%
				9	3897.352	14	7.10%
				7	3892.523	15	0.00%



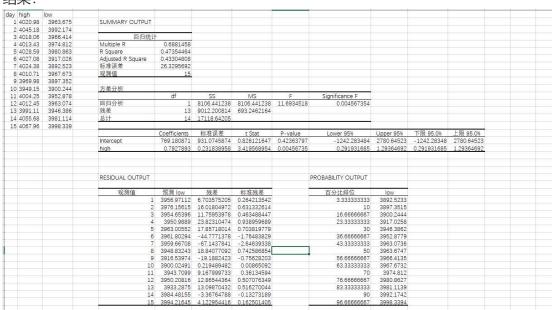
意义:

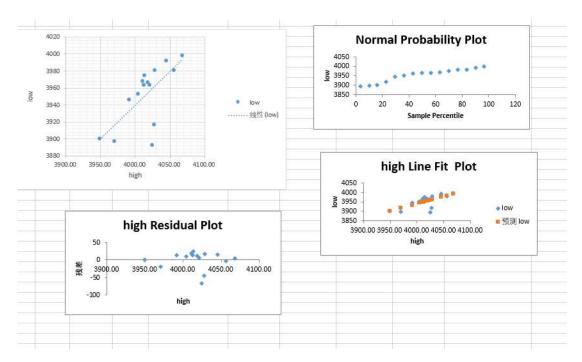
一元线性回归 或者 多元线性回归

线性回归设置:



结果:





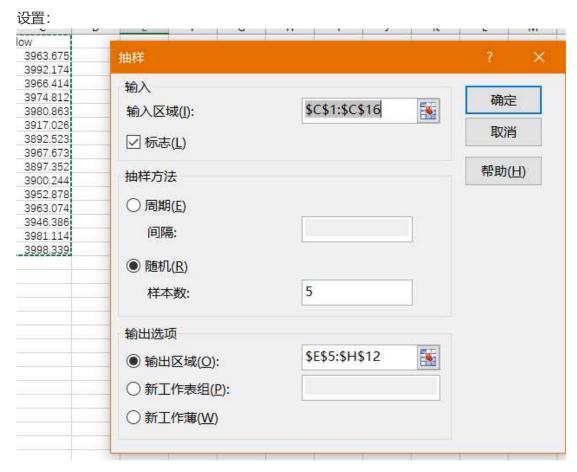
详细解释:

https://wenku.baidu.com/view/0996a441866fb84ae55c8d3c.html https://www.cnblogs.com/nick477931661/p/9113396.html https://blog.csdn.net/data_cola/article/details/87994083

抽样

意义:

从样本中抽n个值



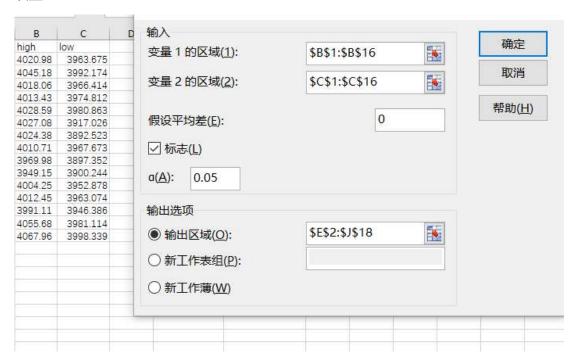
4	
3974.81	
3892.52	
3897.35	
3966.41	
3963.67	
	3892.52 3897.35 3966.41

t-检验: 平均值的成对二样本 分析

意义:

成对二样本分析检验的是成对的两个数值的差异是否不等于零,所谓成对的数值,一般是同一个被访者进行两项测试产生的两个值,例如试饮两个口味的饮品给出的两个评分。

设置:



结果:

p值小于0.05, t值大于t双尾临界,所以差异显著。

В	C	D	E	F	G	Н	- 1
high	low						
4020.98	3963.675		t-检验: 成对双样	本均值分析			
4045.18	3992.174						
4018.06	3966.414			high	low		
4013.43	3974.812		平均	4015.933193	3952.97		
4028.59	3980.863		方差	921.2693839	1222.76		
4027.08	3917.026		观测值	15	15		
4024.38	3892.523		泊松相关系数	0.688145801			
4010.71	3967.673		假设平均差	0			
3969.98	3897.352		df	14			
3949.15	3900.244		t Stat	9.328961306			
4004.25	3952.878		P(T<=t) 单尾	1.09641E-07			
4012.45	3963.074		t单尾临界	1.761310136			
3991.11	3946.386		P(T<=t) 双尾	2.19281E-07			
4055.68	3981.114		t 双尾临界	2.144786688			
4067.96	3998.339		-51-56W5-54V856				

t-检验: 双样本等方差假设

这个是相对于成对二样本分析而言。

用等方差还是异方差假设就是要用"方差齐性检验"方差齐性检验的原理,如果检验是显著的,说明两个独立总体方差不一样,这是用异方差假设,否则用等方差假设。

设置:



结果:

p双尾很小, 因此差异显著

万差 921.2693839 1222.76 见测值 15 15 合并方差 1072.014765 15 既设平均差 0 28 Stat 5.266458858 6.69488E-06 単尾临界 1.701130934 (T<=t) 双尾 1.33898E-05		high	low
加測値 15 15 15	平均	4015.933193	3952.97
合并方差 1072.014765 设设平均差 0 If 28 Stat 5.266458858 (T<=t) 单尾 6.69488E-06 单尾临界 1.701130934 (T<=t) 双尾 1.33898E-05	方差	921.2693839	1222.76
度设平均差 0 If 28 Stat 5.266458858 (T<=t) 単尾 6.69488E-06 単尾临界 1.701130934 (T<=t) 双尾 1.33898E-05	观测值	15	15
Stat 5.266458858 (T<=t) 单尾 6.69488E-06 单尾临界 1.701130934 (T<=t) 双尾 1.33898E-05	合并方差	1072.014765	
Stat 5.266458858 (T<=t) 单尾 6.69488E-06 单尾临界 1.701130934 (T<=t) 双尾 1.33898E-05	假设平均差	0	
(T<=t) 单尾 6.69488E-06 单尾临界 1.701130934 (T<=t) 双尾 1.33898E-05	df	28	
单尾临界 1.701130934 (T<=t) 双尾 1.33898E-05	Stat	5.266458858	
(T<=t) 双尾 1.33898E-05	P(T<=t) 单尾	6.69488E-06	
	单尾临界	1.701130934	
2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2 1	P(T<=t) 双尾	1.33898E-05	
双尾临界 2.048407142	: 双尾临界	2.048407142	

t-检验: 双样本异方差假设

意义:

异方差情况下的双样本t检验

輸入	ADA4 ADA4 C	800	确定
变量 1 的区域(<u>1</u>):	\$B\$1:\$B\$16		Drawe
变量 2 的区域(<u>2</u>):	\$C\$1:\$C\$16	*	取消
假设平均差(<u>E</u>):	0		帮助(<u>H</u>)
☑ 标志(<u>L</u>)		- ()	
a(<u>A</u>): 0.05			
前出选项	A STATE OF THE STA		
	\$D\$25:\$H\$35	Laborate 1	
俞出选项⑩ 输出区域(<u>O</u>):新工作表组(<u>P</u>):	\$D\$25:\$H\$35	L-EAST	

很接近双样本等方差t检验。

差异也是显著。

	high	low
平均	4015.933193	3952.969753
方差	921.2693839	1222.760147
观测值	15	15
假设平均差	0	
if	27	
Stat	5.266458858	
P(T<=t) 单尾	7.44744E-06	
单尾临界	1.703288446	
P(T<=t) 双尾	1.48949E-05	
双尾临界	2.051830516	

z-检验: 双样本平均差检验

意义:

已知两个样本总体方差的情况下才用这个检验。实际效果和t检验也差不多。

设置:



p都等于0了,差异非常显著。和t检验结果类似。

z-检验: 双样本:	51E73 (/)	
	high	low
平均	4015.93	3952.97
已知协方差	30.35	34.97
观测值	15	15
假设平均差	0	
z	30.1725	
P(Z<=z) 单尾	0	
z单尾临界	1.64485	
P(Z<=z) 双尾	0	
z双尾临界	1.95996	