高级统计 Week2（缺失值/极端值）作业

_教师：刘红云

姓名：黄婷 学号：201521430033 院系：脑与认知科学研究院

第一部分（共两部分）：

Data 文件：HBAT.sav

Code 文件：CH2_DATA_EXAMINATION_SYXTAX.SPS

主要结果输出+说明：

- TITLE 'BASIC DESCRIPTIVE STATISTICS AND GRAPHICS -- X6' .

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
X6 - Product Quality	100	100.0%	0	0.0%	100	100.0%

从此图中可以看出该数据文件中的“X6 - Product Quality”变量中，有效数据 100 个，占比例为 100%；不存在缺失数据；总数据个数同有效数据为 100 个。

Descriptives

			Statistic	Std. Error
X6 - Product Quality	Mean		7.810	.1396
	95% Confidence Interval for Mean	Lower Bound	7.533	
		Upper Bound	8.087	
	5% Trimmed Mean		7.833	
	Median		8.000	
	Variance		1.950	
	Std. Deviation		1.3963	
	Minimum		5.0	
	Maximum		10.0	
	Range		5.0	
	Interquartile Range		2.6	
	Skewness		-.245	.241
	Kurtosis		-1.132	.478

此图为“X6 - Product Quality”变量的统计描述表格，包括：

（此部分有参考张文彤-spss 统计分析基础教程）

1）：集中趋势指标：产品质量平均得分 7.810（Mean），去掉两侧各 5%的极端值后，截尾均数为 7.833（5% Trimmed Mean）；中位数为 8（Median）。均数、截尾均数和中位数基本相同，推测数据可能是对称分布。

2）：离散趋势指标：产品质量得分方差为 1.950（Variance），对其开根号得到标准差 1.3963（Std.Deviation）。最低（Minimum）得分为 5，最高（Maximum）得分为 10（Excellent），两数之差表示全距为 5（Range），四分位数间距为 2.6（Interquartile Range）.

3）：分布特征指标：表示数据偏离正态分布程度的偏度系数为-0.245（Skewness），峰度系数为-1.132（Kurtosis）；以及各自的标准误。

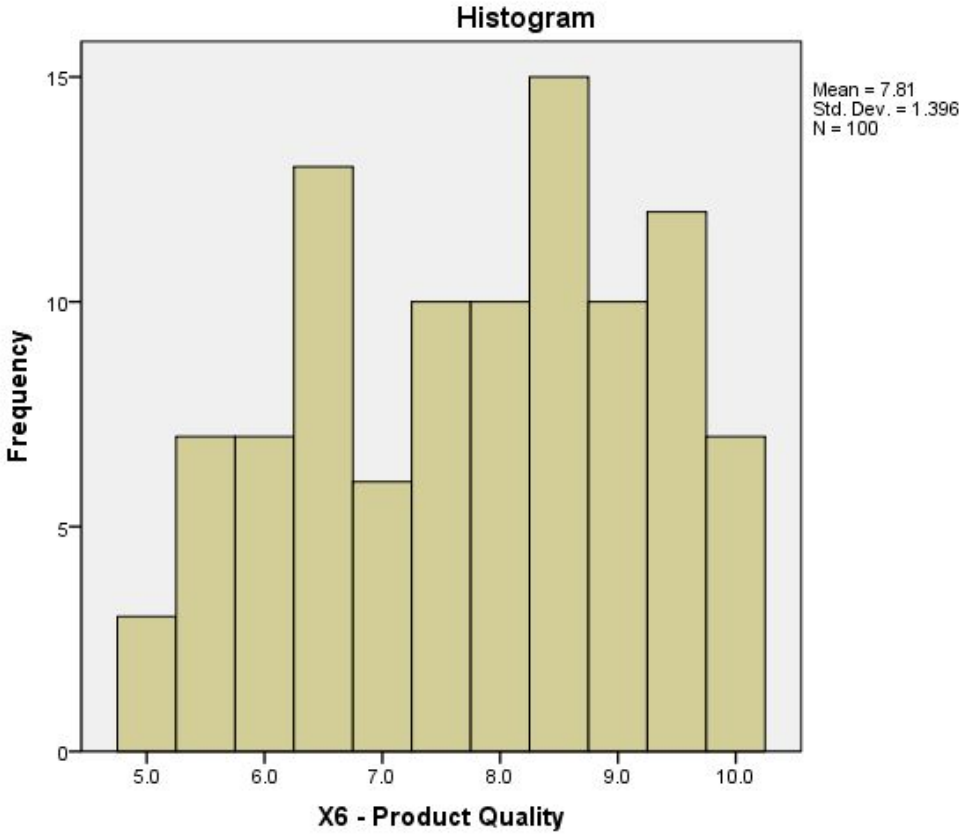
4）：参数估计：均数的标准误为 0.1396，总体均数 95%置信区间的边界值为 7.533-8.087（95%Confidencen interval for mean）

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
X6 - Product Quality	.109	100	.005	.950	100	.001

a. Lilliefors Significance Correction

对“X6 - Product Quality”该项变量数据进行正态性检验，两种检验方法的 SIg 值<0.05，该变量数据集不符合正态分布。



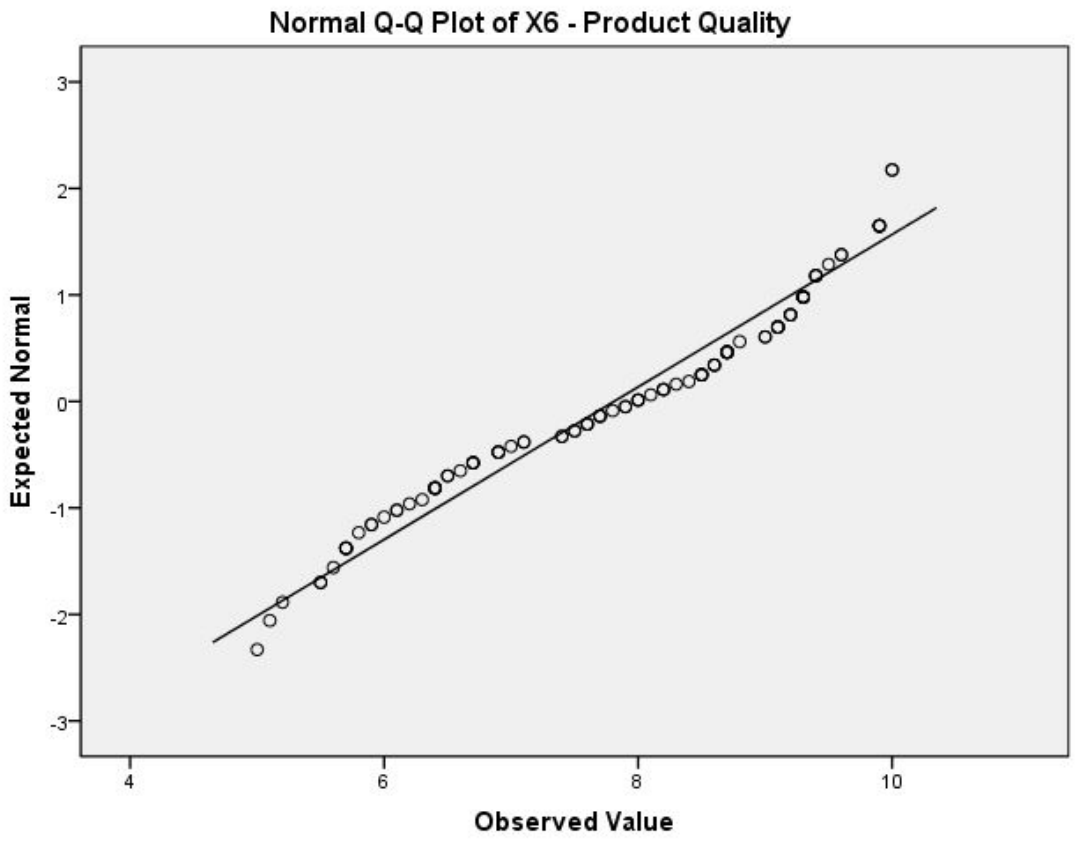
此图为“X6 - Product Quality”变量下数据的直方图，x 轴表示得分，y 轴表示该得分出现的频数，从该图可以看出数据多集中于右侧，呈负偏态分布。

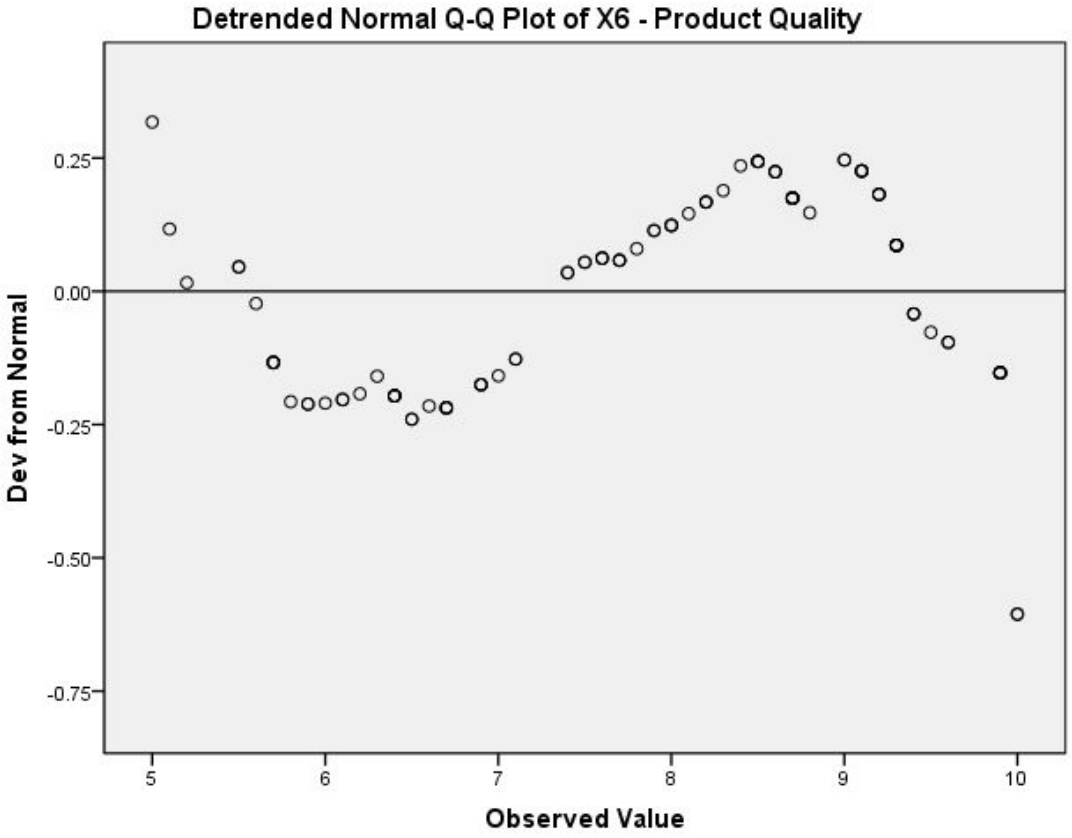
X6 - Product Quality Stem-and-Leaf Plot

Frequency	Stem & Leaf
3.00	5 . 012
10.00	5 . 5567777899
10.00	6 . 0112344444
10.00	6 . 5567777999
5.00	7 . 01144
11.00	7 . 55666777899
9.00	8 . 000122234
14.00	8 . 55556667777778
18.00	9 . 001111222333333444
8.00	9 . 56699999
2.00	10 . 00

Stem width: 1.0
Each leaf: 1 case(s)

此图为“X6 - Product Quality”变量下数据的茎叶图，同上面的直方图一样，可以看出产品质量得分多分布于 8,9 附近，呈负偏态分布。





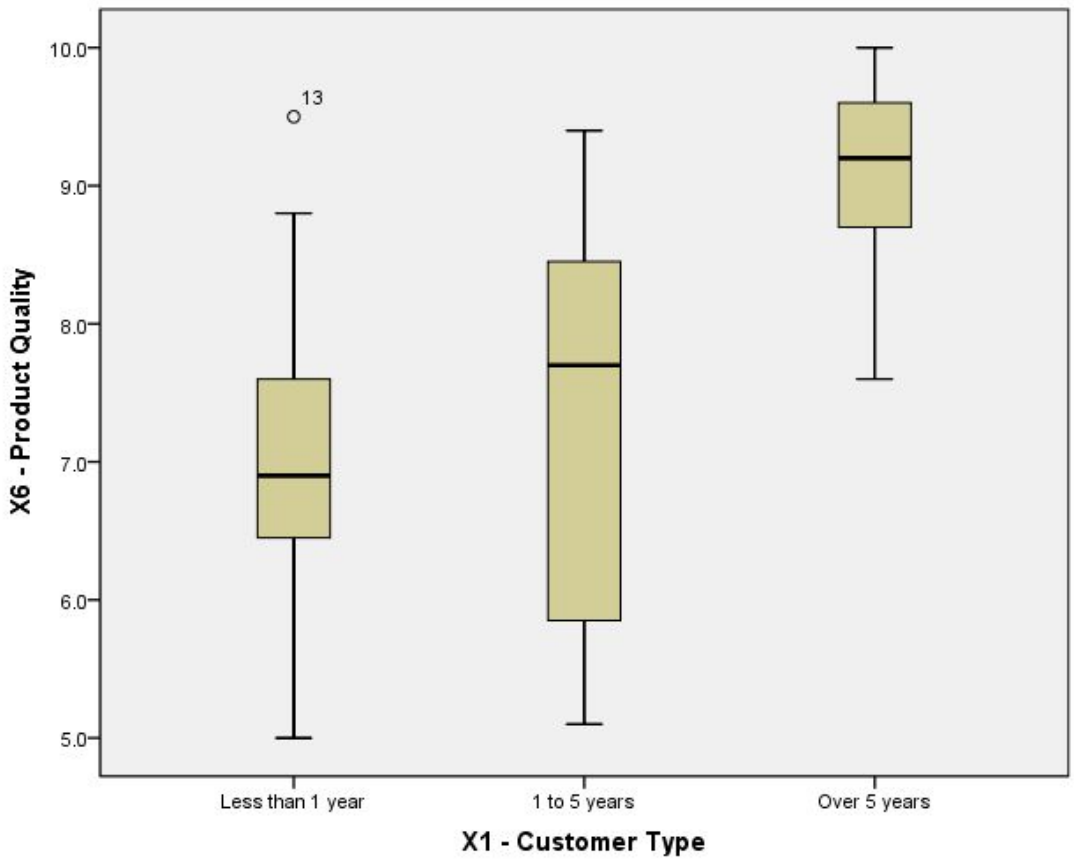
第一张 Q-Q 图，两坐标表示理论累积概率和实际累积概率，如果数据呈正态分布，则数据点应和理论直线（即对角线）基本重合，从此图来看有数据部分偏离；进一步看第二张趋势 Q-Q 图，该图反应的是按正态分布计算的理论值与实际值之差的分布情况，如果数据呈正态分布则数据点应均匀分布在 y=0 这条直线上下，此图中有明显的起伏波动，故可判断为不符合正态分布。

- TITLE 'BASIC DESCRIPTIVE STATISTICS AND GRAPHICS -- X6 BY X1' .

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	X1 - Customer Type	N	Percent	N	Percent	N	Percent
X6 - Product Quality	Less than 1 year	32	100.0%	0	0.0%	32	100.0%
	1 to 5 years	35	100.0%	0	0.0%	35	100.0%
	Over 5 years	33	100.0%	0	0.0%	33	100.0%

此表表示将变量 X6 的数据按照 X1 的标准分成三类，分别对有效数据、无效数据和总数据个数进行描述性统计。

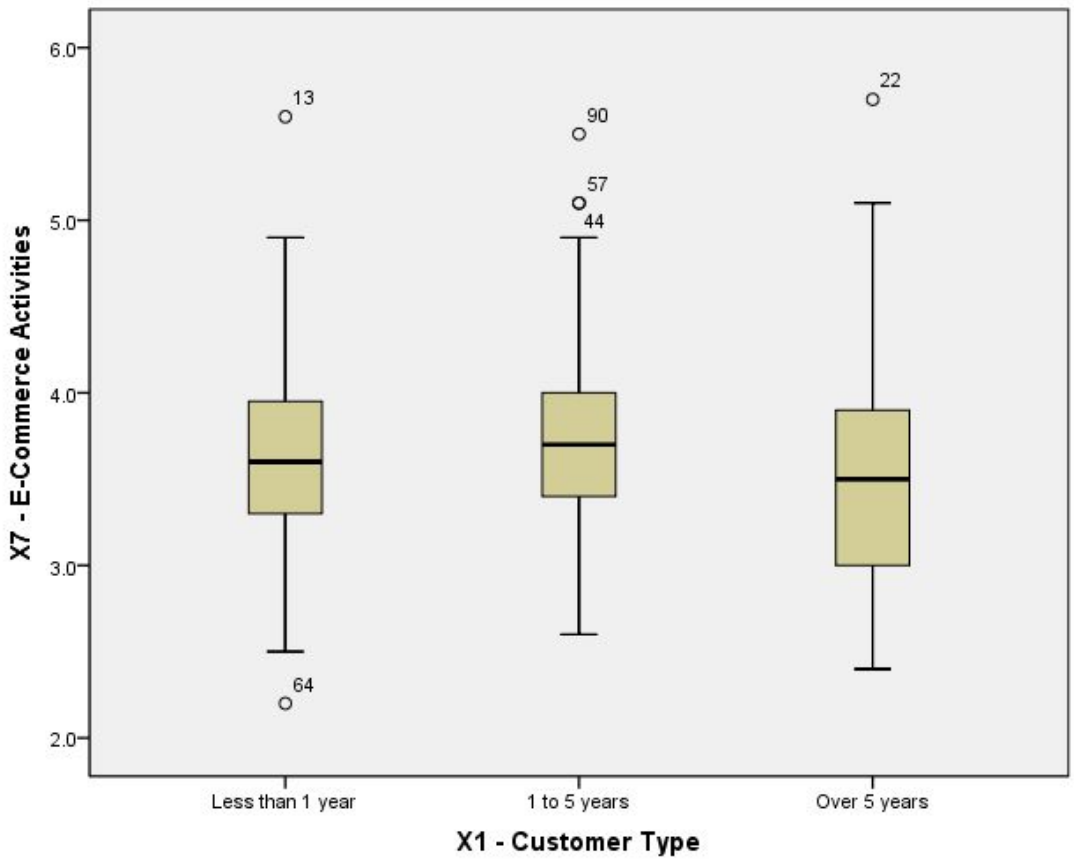


此图也是将变量 X6 的数据按照 X1 的标准分成三类，分别对分数进行描述，横坐标表示根据 X1 变量将客户类型（X1：customer type）分为三类，分别是小于 1 年、1-5 年、大于 5 年；纵坐标表示三个类别下各自产品得分情况（X6:product quality）。此图用箱图表示，箱图的特点是：每个箱图都由中间最粗的一条线、一个方框、外延出两条细线和单独散点构成，其中中间粗线表示中位数，方框的两段表示上四分位数和下四分位数，两者距离为四分位数，外延的细线表示抛去极端值之后的最大值和最小值，单独散落的点表示极端值，即超过 1.5 倍四分位数间距的值。

- TITLE 'BASIC DESCRIPTIVE STATISTICS AND GRAPHICS -- X7 BY X1' .

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	X1 - Customer Type	N	Percent	N	Percent	N	Percent
X7 - E-Commerce Activities	Less than 1 year	32	100.0%	0	0.0%	32	100.0%
	1 to 5 years	35	100.0%	0	0.0%	35	100.0%
	Over 5 years	33	100.0%	0	0.0%	33	100.0%



上边一表一图解释可参考上文相同部分，两者不同之处在于这里分析的变量是 X1 和 X7。

- 单因素方差分析-- X6 、X7 BY X1' .

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
X6 - Product Quality	Between Groups	83.078	2	41.539	36.652	.000
	Within Groups	109.932	97	1.133		
	Total	193.010	99			
X7 - E-Commerce Activities	Between Groups	.864	2	.432	.878	.419
	Within Groups	47.718	97	.492		
	Total	48.582	99			

词表为分别对 X6/X7 与 X1 进行单因素方差分析的结果表，第一行为 X6 by X1：其中 between group 表示组间变异，within group 表示组内变异，total 表示总变异，第 2、3、4 列分别表示离均差平方和（sum of aquares）、自由度（df）、均方（mean square），检验统计量 F 为 36.652，再去根据自由度查表得到，p 值显著，则可说明 X6 变量在 X1 变量三个水平上有差异，即不同 customer type 的人消费的 product quality 存在差异；同理分析第二行，X7 在 X1 各水平上不存在差异。

Multiple Comparisons

Scheffe

Dependent Variable	(I) X1 - Customer Type	(J) X1 - Customer Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
X6 - Product Quality	Less than 1 year	1 to 5 years	-.1431	.2604	.860	-.790	.504
		Over 5 years	-2.0092*	.2641	.000	-2.666	-1.353
	1 to 5 years	Less than 1 year	.1431	.2604	.860	-.504	.790
		Over 5 years	-1.8661*	.2583	.000	-2.508	-1.224
	Over 5 years	Less than 1 year	2.0092*	.2641	.000	1.353	2.666
		1 to 5 years	1.8661*	.2583	.000	1.224	2.508
X7 - E-Commerce Activities	Less than 1 year	1 to 5 years	-.1050	.1715	.829	-.531	.321
		Over 5 years	.1205	.1740	.787	-.312	.553
	1 to 5 years	Less than 1 year	.1050	.1715	.829	-.321	.531
		Over 5 years	.2255	.1702	.419	-.198	.649
	Over 5 years	Less than 1 year	-.1205	.1740	.787	-.553	.312
		1 to 5 years	-.2255	.1702	.419	-.649	.198

*. The mean difference is significant at the .05 level.

词表在方差分析的基础上进行事后多重比较，用于变量包含有大于 2 个水平的分析中，即可以看到是哪些水平上存在差异。从上图中可以看出 X6:product quality 与 X1：customer type 在 1、2 水平上差异不显著，在第 3 水平（over 5 year）差异显著；对于变量 X7 在 X1 各个水平上均不显著。

Homogeneous Subsets

X6 - Product Quality

Scheffe[a,b]

X1 - Customer Type	N	Subset for alpha = .05	
		1	2
Less than 1 year	32	7.097	
1 to 5 years	35	7.240	
Over 5 years	33		9.106
Sig.		.861	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 33.287.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

X7 - E-Commerce Activities

Scheffe[a,b]

X1 - Customer Type	N	Subset for alpha = .05
		1
Over 5 years	33	3.555
Less than 1 year	32	3.675
1 to 5 years	35	3.780
Sig.		.426

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 33.287.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

上图说明的问题同上文，是通过寻找同志子集的方式检验各组之间的差别。即 X6:product quality 与 X1：customer type 在 1、2 水平上差异不显著，在第 3 水平（over 5 year）差异显著；对于变量 X7 在 X1 各个水平上均不显著。

第二部分：

Data 文件：HBAT_MISSING
Code 文件：Missing data syntax.sps

- Step1：Determine the Type of Missing Data

以下分析基于缺失值不能被直接忽视的情况

- Step 2：Determine the Extent of Missing Data

Assessing the extent and patterns of missing data

The percentage of variables with missing data for each case;

The number of cases with missing data for each variable.

The number of cases with no missing data on any of the variables.

Univariate Statistics

	N	Mean	Std. Deviation	Missing		No. of Extremes[a]	
				Count	Percent	Low	High
v1	49	4.008	.9318	21	30.0	0	0
v2	57	1.944	.8751	13	18.6	0	0
v3	53	8.062	1.4072	17	24.3	0	0
v4	63	5.168	1.1714	7	10.0	0	0
v5	61	2.856	.7760	9	12.9	0	0
v6	64	2.611	.7174	6	8.6	0	0
v7	61	6.823	1.6809	9	12.9	1	0
v8	61	46.033	9.3559	9	12.9	0	0
v9	63	4.759	.8319	7	10.0	0	0
v10	68			2	2.9		
v11	68			2	2.9		
v12	68			2	2.9		
v13	69			1	1.4		
v14	68			2	2.9		

a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

此图统计了各类变量下的有效值和缺失值，及其缺失值个数所占比例。从图中可以看出 v1-v9 存在不同比例的缺失值；v10-v14 中存在极少的缺失值。

Data Patterns (all cases)

Case	# Missing	% Missing	Missing and Extreme Value Patterns														Variable Values								
			v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v1	v2	v3	v4	v5	v6	v7	v8	v9
201	0	.0															3.3	.9	8.6	4.0	2.1	1.8	6.3	41.0	4.5
202	2	14.3	S		S												.	.4	.	2.5	1.2	1.7	5.2	35.0	3.3
203	2	14.3		S					S								3.0	.	9.1	7.1	3.5	3.4	.	55.0	5.2
204	3	21.4	S		S						S						.	1.5	.	4.8	1.9	2.5	7.2	36.0	.
205	1	7.1			S												5.1	1.4	.	4.8	3.3	2.6	3.8	49.0	4.9
206	0	.0															4.6	2.1	7.9	5.8	3.4	2.8	4.7	49.0	5.9
207	3	21.4	S		S						S						.	1.5	.	4.8	1.9	2.5	7.2	36.0	.
208	0	.0															5.2	1.3	9.7	6.1	3.2	3.9	6.7	54.0	5.8
209	0	.0															3.5	2.8	9.9	3.5	3.1	1.7	5.4	49.0	5.4
210	7	50.0				S	S	S	S	S	S	S					4.1	3.7	5.9	.	.	.	.	.	.
211	0	.0															3.0	2.8	7.8	7.1	3.0	3.8	7.9	49.0	4.4
212	0	.0															4.8	1.7	7.6	4.2	3.3	1.4	5.8	39.0	5.5
213	2	14.3		S	S												3.1	.	.	7.8	3.6	4.0	5.9	43.0	5.2
214	7	50.0	S			S		S	S	S			S		S		.	2.7	5.0	.	2.2	.	.	.	3.6
215	0	.0															4.0	.5	6.7	4.5	2.2	2.1	5.0	31.0	4.0
216	2	14.3	S				S										.	1.6	6.4	5.0	.	2.1	8.4	25.0	3.4
217	0	.0															6.1	.5	9.2	4.8	3.3	2.8	7.1	60.0	5.2
218	2	14.3	S				S										.	2.8	5.2	5.0	.	2.7	8.4	38.0	3.7
219	2	14.3							S	S							3.1	2.2	6.7	6.8	2.6	2.9	.	.	4.3
220	1	7.1		S													6.5	.	9.0	7.0	3.2	3.7	8.0	33.0	5.4
221	3	21.4	S		S				S								.	1.6	.	4.8	2.0	2.8	.	32.0	4.3
222	2	14.3			S		S										3.9	2.2	.	4.6	.	2.5	8.3	47.0	5.0
223	0	.0															2.8	1.4	8.1	3.8	2.1	1.4	6.6	39.0	4.4
224	3	21.4	S	S						S							.	.	8.6	5.7	2.7	3.7	6.7	.	5.0
225	2	14.3			S	S											4.7	1.3	.	.	3.0	2.6	6.8	54.0	5.9
226	0	.0															3.4	2.0	9.7	4.7	2.7	1.7	4.8	49.0	4.7
227	2	14.3		S						S							3.2	.	5.7	5.1	3.6	2.9	6.2	.	4.4
228	2	14.3	S			S											.	1.8	7.7	.	3.4	1.5	5.9	40.0	5.6
229	1	7.1					S										5.3	1.4	9.7	6.1	.	3.9	6.8	54.0	5.9
230	0	.0															4.7	1.3	9.9	6.7	3.0	2.6	6.8	55.0	6.0
231	1	7.1							S								3.7	.7	8.2	6.0	2.1	2.5	.	41.0	5.0
232	2	14.3	S	S													.	.	8.2	5.0	3.6	2.5	9.0	53.0	5.2
233	7	50.0		S	S		S	S			S			S		S	4.5	.	.	5.9	.	.	8.8	50.0	.
234	0	.0															2.8	2.4	6.7	4.9	2.5	2.6	9.2	32.0	3.7
235	2	14.3						S			S						3.8	.8	8.7	2.9	1.6	.	5.6	39.0	.
236	0	.0															2.9	2.6	7.7	7.0	2.8	3.6	7.7	47.0	4.2
237	1	7.1		S													4.9	.	7.4	6.9	4.6	4.0	9.6	62.0	6.2
238	1	7.1	S														.	2.5	9.6	5.5	4.0	3.0	7.7	65.0	6.0
239	0	.0															4.3	1.8	7.6	5.4	3.1	2.5	4.4	46.0	5.6
240	1	7.1	S						-								.	1.5	9.9	2.7	1.3	1.2	1.7	50.0	5.0
241	2	14.3			S		S										3.1	1.9	.	4.5	.	3.1	3.8	54.0	4.8
242	0	.0															5.1	1.9	9.2	5.8	3.6	2.3	4.5	60.0	6.1
243	0	.0															4.1	1.1	9.3	5.5	2.5	2.7	7.4	47.0	5.3
244	1	7.1								S							3.0	3.8	5.5	4.9	3.4	2.6	6.0	.	4.2
245	7	50.0	S		S		S		S	S				S		S	.	2.0	.	4.7	.	3.2	.	.	3.4
246	1	7.1				S											3.7	1.4	9.0	.	2.6	2.3	6.8	45.0	4.9
247	0	.0															4.2	2.5	9.2	6.2	3.3	3.9	7.3	59.0	6.0
248	2	14.3	S	S													.	.	6.4	5.3	3.0	2.5	7.1	46.0	4.5
249	1	7.1		S													5.3	.	8.5	3.7	3.5	1.9	4.8	58.0	4.3
250	2	14.3	S		S												.	3.7	.	5.2	3.0	2.3	9.1	49.0	4.8
251	0	.0															3.0	3.2	6.0	5.3	3.1	3.0	8.0	43.0	3.3
252	0	.0															2.8	3.8	8.9	6.9	3.3	3.2	8.2	53.0	5.0
253	1	7.1	S														.	2.0	9.3	5.9	3.7	2.4	4.6	60.0	6.1
254	0	.0															3.4	3.7	6.4	5.7	3.5	3.4	8.4	47.0	3.8
255	2	14.3	S		S												.	1.0	.	3.4	1.7	1.1	6.2	35.0	4.1
256	1	7.1	S														.	3.3	7.5	4.5	2.5	2.4	7.6	39.0	3.6
257	2	14.3		S	S												3.6	.	.	5.8	3.7	2.5	9.3	44.0	4.8
258	0	.0															4.0	.9	9.1	5.4	2.4	2.6	7.3	46.0	5.1
259	1	7.1	S														.	2.1	6.9	5.4	1.1	2.6	8.9	29.0	3.9
260	1	7.1	S														.	2.0	6.4	4.5	2.1	2.2	8.8	28.0	3.3
261	7	50.0		S	S			S	S	S	S		S				3.6	.	.	6.2	4.5	.	.	.	.
262	0	.0															5.6	2.2	8.2	3.1	4.0	1.6	5.3	55.0	3.9
263	7	50.0		S		S	S	S	S	S		S					3.6	.	9.9	.	.	.	.	.	4.9
264	0	.0															5.2	1.3	9.1	4.5	3.3	2.7	7.3	60.0	5.1
265	0	.0															3.0	2.0	6.6	6.6	2.4	2.7	8.2	41.0	4.1
266	0	.0															4.2	2.4	9.4	4.9	3.2	2.7	8.5	49.0	5.2
267	2	14.3			S	S											3.8	.8	.	.	2.2	2.6	5.3	42.0	5.1
268	1	7.1									S						3.3	2.6	9.7	3.3	2.9	1.5	5.2	47.0	.
269	2	14.3	S		S												.	1.9	.	4.5	1.5	3.1	9.9	39.0	3.3
270	0	.0															4.5	1.6	8.7	4.6	3.1	2.1	6.8	56.0	5.1

- indicates an extreme low value, while + indicates an extreme high value. The range used is (Q1 - 1.5*IQR, Q3 + 1.5*IQR).

此图表（Data patterns all case）记录了每个被试在所有变量条件下的缺失情况，第一列是被试编号，第二列开始分别记录了缺失变量数量、缺失比例和具体的缺失情况。在输出结果中紧排列在此表值后的一个表（Missing pattern case with missing values）（由于过大+实际也看不太清就不贴出来了）与此表类似，不同之处在只截取了含有缺失值的记录的部分。在下面一个标为 Tabulated patterns 的表算是对以上两个表的一个总结。

- Step3：Diagnosing the randomness of the missing data process

Separate Variance t Tests[a]

		v1	v2	v3	v4	v5	v6	v7	v8	v9
v1	t	.	-.3	1.6	2.4	2.9	1.7	-1.0	2.6	2.6
	df	.	38.2	17.5	45.0	24.1	41.1	25.9	24.4	27.2
	P(2-tail)	.	.757	.126	.019	.008	.101	.316	.016	.015
	# Present	49	39	40	44	43	44	43	43	44
	# Missing	0	18	13	19	18	20	18	18	19
	Mean(Present)	4.008	1.921	8.255	5.373	3.056	2.707	6.665	48.209	4.948
	Mean(Missing)	.	1.994	7.469	4.695	2.378	2.400	7.200	40.833	4.321
v2	t	-.4	.	-.1	-2.7	-4.5	-2.3	-1.5	-1.2	-1.5
	df	11.9	.	12.0	16.4	18.1	11.9	12.9	11.4	23.1
	P(2-tail)	.700	.	.950	.015	.000	.038	.155	.255	.147
	# Present	39	57	44	51	50	54	51	52	52
	# Missing	10	0	9	12	11	10	10	9	11
	Mean(Present)	3.977	1.944	8.057	4.982	2.694	2.519	6.682	45.462	4.706
	Mean(Missing)	4.130	.	8.089	5.958	3.591	3.110	7.540	49.333	5.009
v3	t	.3	1.6	.	.8	1.2	.0	-.2	1.7	1.1
	df	16.1	21.4	.	22.7	15.3	25.9	18.3	31.9	19.3
	P(2-tail)	.748	.132	.	.434	.250	.978	.841	.100	.269
	# Present	40	44	53	48	48	49	47	46	50
	# Missing	9	13	0	15	13	15	14	15	13
	Mean(Present)	4.025	2.036	8.062	5.235	2.931	2.612	6.796	47.022	4.818
	Mean(Missing)	3.933	1.631	.	4.953	2.577	2.607	6.914	43.000	4.531
v4	t	.1	.0	.7	.	.7	1.4	1.5	.3	-.8
	df	9.3	5.8	4.4	.	5.8	3.8	5.7	4.1	6.2
	P(2-tail)	.902	.988	.542	.	.484	.239	.177	.814	.465
	# Present	44	51	48	63	56	60	57	57	57
	# Missing	5	6	5	0	5	4	4	4	6
	Mean(Present)	4.011	1.943	8.121	5.168	2.871	2.635	6.867	46.088	4.733
	Mean(Missing)	3.980	1.950	7.500	.	2.680	2.250	6.200	45.250	5.000
v5	t	-.2	-1.0	.7	.2	.	-1.2	-.8	.3	.9
	df	7.4	8.3	4.3	12.6	.	6.5	5.9	5.8	7.1
	P(2-tail)	.810	.344	.518	.838	.	.260	.447	.763	.378
	# Present	43	50	48	56	61	58	55	55	56
	# Missing	6	7	5	7	0	6	6	6	7
	Mean(Present)	3.998	1.904	8.129	5.175	2.856	2.579	6.758	46.182	4.798
	Mean(Missing)	4.083	2.229	7.420	5.114	.	2.917	7.417	44.667	4.443
v6	t	.4	-.6	.6	.2	.1	.	-.2	.3	.8
	df	11.5	2.1	3.2	2.1	2.0	.	1.0	1.1	1.1
	P(2-tail)	.672	.629	.568	.883	.926	.	.848	.822	.566
	# Present	44	54	49	60	58	64	59	59	61
	# Missing	5	3	4	3	3	0	2	2	2
	Mean(Present)	4.018	1.919	8.118	5.177	2.860	2.611	6.810	46.085	4.775
	Mean(Missing)	3.920	2.400	7.375	5.000	2.767	.	7.200	44.500	4.250
v7	t	2.5	-.5	.8	-1.9	.1	-2.1	.	.5	1.5
	df	14.9	5.9	5.6	6.5	5.7	7.5	.	2.1	8.4
	P(2-tail)	.024	.613	.437	.097	.921	.076	.	.652	.179
	# Present	43	51	47	57	55	59	61	58	56
	# Missing	6	6	6	6	6	5	0	3	7
	Mean(Present)	4.077	1.920	8.138	5.088	2.860	2.581	6.823	46.207	4.805
	Mean(Missing)	3.517	2.150	7.467	5.933	2.817	2.960	.	42.667	4.386
v8	t	2.9	-2.6	2.1	-1.2	-1.0	-2.3	1.8	.	2.2
	df	14.4	4.8	6.9	7.5	6.0	6.3	9.0	.	9.2
	P(2-tail)	.011	.049	.073	.271	.371	.056	.107	.	.052
	# Present	43	52	46	57	55	59	58	61	56
	# Missing	6	5	7	6	6	5	3	0	7
	Mean(Present)	4.088	1.854	8.261	5.126	2.822	2.573	6.850	46.033	4.821
	Mean(Missing)	3.433	2.880	6.757	5.567	3.167	3.060	6.300	.	4.257
v9	t	.7	-.2	.0	1.0	.6	1.3	.0	1.5	.
	df	8.9	4.4	2.1	5.8	4.3	2.3	5.1	5.7	.
	P(2-tail)	.531	.880	.975	.351	.582	.294	.972	.182	.
	# Present	44	52	50	57	56	61	56	56	63
	# Missing	5	5	3	6	5	3	5	5	0
	Mean(Present)	4.025	1.937	8.060	5.223	2.882	2.633	6.825	46.429	4.759
	Mean(Missing)	3.860	2.020	8.100	4.650	2.560	2.167	6.800	41.600	.

For each quantitative variable, pairs of groups are formed by indicator variables (present, missing).

上页图表表示，将 v1-v9 共九个变量进行两两比较，即某一变量按照另一个变量的缺失情况分为两组，即缺失部分在此变量上对应的值为一组；未缺失部分在此变量上对应的值为另一组，然后将这两组数据集进行差异性检验。图中标注绿色背景色部分为，缺失值部分的值与未缺失部分的值存在差异，可诊断为这些缺失属于随机缺失（MAR），其他差异不显著部分，属于完全随机缺失（MCAR）。

- Step 4: Select the Imputation Method（选择插补方法）

  4-1--Imputation Using Only Valid Data

  -Complete case approach---LISTWISE

Listwise Statistics

Listwise Means

Number of cases	v1	v2	v3	v4	v5	v6	v7	v8	v9
26	4.019	1.950	8.354	5.269	2.981	2.600	6.754	48.308	4.896

Listwise Covariances

	v1	v2	v3	v4	v5	v6	v7	v8	v9
v1	.9184								
v2	-.4266	.7850							
v3	.4813	-.3052	1.3682						
v4	-.2594	.3132	-.0787	1.2166					
v5	.2644	.1818	.0895	.0246	.2376				
v6	-.0668	.2344	-.0572	.6596	.0772	.5536			
v7	-.5519	.4344	-.3730	.5837	-.1009	.5452	1.9178		
v8	4.6218	.3440	6.0988	1.6938	2.6782	1.8040	-1.1012	64.7815	
v9	.4221	-.1978	.6802	.1507	.1187	.0384	-.4498	3.6572	.6436

Listwise Correlations

	v1	v2	v3	v4	v5	v6	v7	v8	v9
v1	1								
v2	-.502	1							
v3	.429	-.294	1						
v4	-.245	.320	-.061	1					
v5	.566	.421	.157	.046	1				
v6	-.094	.356	-.066	.804	.213	1			
v7	-.416	.354	-.230	.382	-.150	.529	1		
v8	.599	.048	.648	.191	.683	.301	-.099	1	
v9	.549	-.278	.725	.170	.304	.064	-.405	.566	1

以上三个表格（listwise means，listwise covariaances，listwise correlation）输出的是全部无缺失的被试对应数据的均数、协方差和相关系数。

4-2--Imputation Using Only Valid Data

-Using all-available data---PAIRWISE

Pairwise Statistics

转向数据

Pairwise Frequencies

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14
v1	49													
v2	39	57												
v3	40	44	53											
v4	44	51	48	63										
v5	43	50	48	56	61									
v6	44	54	49	60	58	64								
v7	43	51	47	57	55	59	61							
v8	43	52	46	57	55	59	58	61						
v9	44	52	50	57	56	61	56	56	63					
v10	47	56	51	63	61	64	61	61	62	68				
v11	48	56	52	62	59	64	61	61	62	66	68			
v12	48	56	53	61	61	63	60	60	62	66	66	68		
v13	49	56	52	63	60	64	61	61	62	67	68	67	69	
v14	48	56	53	61	61	63	60	60	62	66	66	68	67	68

Pairwise Means

	v1	v2	v3	v4	v5	v6	v7	v8	v9
v1	4.008	1.921	8.255	5.373	3.056	2.707	6.665	48.209	4.948
v2	3.977	1.944	8.057	4.982	2.694	2.519	6.682	45.462	4.706
v3	4.025	2.036	8.062	5.235	2.931	2.612	6.796	47.022	4.818
v4	4.011	1.943	8.121	5.168	2.871	2.635	6.867	46.088	4.733
v5	3.998	1.904	8.129	5.175	2.856	2.579	6.758	46.182	4.798
v6	4.018	1.919	8.118	5.177	2.860	2.611	6.810	46.085	4.775
v7	4.077	1.920	8.138	5.088	2.860	2.581	6.823	46.207	4.805
v8	4.088	1.854	8.261	5.126	2.822	2.573	6.850	46.033	4.821
v9	4.025	1.937	8.060	5.223	2.882	2.633	6.825	46.429	4.759
v10	4.015	1.913	8.069	5.168	2.856	2.611	6.823	46.033	4.756
v11	4.017	1.930	8.121	5.152	2.839	2.611	6.823	46.033	4.777
v12	3.998	1.943	8.062	5.164	2.856	2.602	6.790	45.967	4.781
v13	4.008	1.930	8.121	5.168	2.867	2.611	6.823	46.033	4.777
v14	3.998	1.943	8.062	5.164	2.856	2.602	6.790	45.967	4.781

Mean of quantitative variable when other variable is present.

Pairwise Standard Deviations

	v1	v2	v3	v4	v5	v6	v7	v8	v9
v1	.9318	.9200	1.3032	1.2185	.6242	.7321	1.5357	7.7477	.7096
v2	.8845	.8751	1.4334	1.4118	.7243	.6796	1.6686	9.4027	.8768
v3	.9834	.8832	1.4072	1.3	.6956	.7412	1.6007	9.8036	.8366
v4	.9751	.8624	1.3397	1.1	.7963	.7255	1.7230	9.5754	.8384
v5	.9605	.8864	1.3133	1.2257	.7760	.7235	1.6616	9.2298	.8161
v6	.9770	.8454	1.3305	1.1522	.7424	.7174	1.6821	9.4564	.8320
v7	.9661	.8656	1.3377	1.1665	.7603	.7343	1.6809	9.3166	.8411
v8	.9582	.8318	1.2421	1.2002	.7697	.7268	1.7189	9.3559	.8396
v9	.9729	.8593	1.3939	1.1526	.7383	.7204	1.7120	9.5114	.8319
v10	.9498	.8500	1.3778	1.1714	.7760	.7174	1.6809	9.3559	.8385
v11	.9397	.8770	1.3533	1.1734	.7541	.7174	1.6809	9.3559	.8253
v12	.9389	.8829	1.4072	1.1855	.7760	.7192	1.6751	9.4204	.8202
v13	.9318	.8770	1.3533	1.1714	.7778	.7174	1.6809	9.3559	.8253
v14	.9389	.8829	1.4072	1.1855	.7760	.7192	1.6751	9.4204	.8202

双击以激活

Standard deviation of quantitative variable when other variable is present.

Pairwise Covariances

	v1	v2	v3	v4	v5	v6	v7	v8	v9
v1	.8683								
v2	-.3632	.7657							
v3	.4750	-.5293	1.9801						
v4	-.1181	.2860	-.1020	1.3722					
v5	.1901	.2704	.0897	.4329	.6022				
v6	.0222	.1491	-.0346	.6679	.1848	.5146			
v7	-.1775	.5022	-.7668	.8155	.0839	.4964	2.8255		
v8	2.8001	1.1668	7.3231	2.5994	5.0552	1.8403	-3.0579	87.5322	
v9	.3548	-.1512	.8244	.3692	.3272	.1227	-.3681	5.3434	.6921

Pairwise Correlations

	v1	v2	v3	v4	v5	v6	v7	v8	v9
v1	1								
v2	-.446	1							
v3	.371	-.418	1						

以上 5 个表格（Pairwise frequencies，Pairwise means，Pairwise standard deviations，covariances，correlation）输出的是在变量两两匹配的情况下数据的频数、均数、标准差、协方差、相关系数。

4-3--使用回归算法进行填充：

前提：XY 有较强的相关性

Regression Estimated Statistics

Regression Means[a]

v2	v3	v4	v5	v6	v7	v8	v9
1.982	8.006	5.175	2.834	2.593	6.962	45.634	4.740

a. Random normal variate is added to each estimate.

Regression Covariances[a]

	v2	v3	v4	v5	v6	v7	v8	v9
v2	.8489							
v3	-.4829	1.6753						
v4	.2952	-.0535	1.3206					
v5	.3051	.1586	.3659	.5764				
v6	.1522	.0365	.6274	.1985	.5024			
v7	.4563	-.6927	.8578	.0258	.5028	2.8948		
v8	.9556	7.2704	2.5862	5.1006	1.5293	-2.6505	95.3681	
v9	-.0579	.7216	.3192	.3899	.1318	-.3905	5.7215	.7178

a. Random normal variate is added to each estimate.

Regression Correlations[a]

	v2	v3	v4	v5	v6	v7	v8	v9
v2	1							
v3	-.405	1						
v4	.279	-.036	1					
v5	.436	.161	.419	1				
v6	.233	.040	.770	.369	1			
v7	.291	-.315	.439	.020	.417	1		
v8	.106	.575	.230	.688	.221	-.160	1	
v9	-.074	.658	.328	.606	.220	-.271	.692	1

a. Random normal variate is added to each estimate.

以上三个表格（Regression means， Regression covariaances， Regression correlation）输出的是根据算法随机填充残差后重新进行分析的均数、协方差和相关系数。

4-4--使用 EM 算法进行填充：

EM Estimated Statistics

EM Means[a]

v2	v3	v4	v5	v6	v7	v8	v9
2.047	8.037	5.162	2.845	2.613	6.918	45.748	4.738

a. Little's MCAR test: Chi-Square = 121.820, DF = 120, Sig. = .436

EM Covariances[a]

	v2	v3	v4	v5	v6	v7	v8	v9
v2	.8026							
v3	-.4441	1.7975						
v4	.3299	-.1676	1.3176					
v5	.3340	.0768	.3665	.5909				
v6	.2015	-.0967	.6629	.2075	.5298			
v7	.5537	-.7441	.7808	.0405	.5208	2.8575		
v8	1.2423	7.0178	2.1485	5.0084	1.7461	-3.2114	89.4697	
v9	-.0675	.6926	.3292	.3573	.1212	-.3607	5.3744	.6904

a. Little's MCAR test: Chi-Square = 121.820, DF = 120, Sig. = .436

EM Correlations[a]

	v2	v3	v4	v5	v6	v7	v8	v9
v2	1							
v3	-.370	1						
v4	.321	-.109	1					
v5	.485	.075	.415	1				
v6	.309	-.099	.793	.371	1			
v7	.366	-.328	.402	.031	.423	1		
v8	.147	.553	.198	.689	.254	-.201	1	
v9	-.091	.622	.345	.559	.200	-.257	.684	1

a. Little's MCAR test: Chi-Square = 121.820, DF = 120, Sig. = .436

以上三个表格（EMmeans，EMcovariaances，EMcorrelation），第一个表格给出的是 EM 估计的各变量均值，后两个图表是 8 个变量之间的协方差矩阵和相关系数矩阵，最后一行小字给出的是 LIttle’s MCAR 检验，sig=0.436>0.05，即不能拒绝原假设，接受原假设，数据是完全随机缺失。