

项目一：描述统计分析

一、实验目的

掌握运用 SPSS 统计分析软件整理原始数据并分析其基本特征。

二、实验内容

- 1、对原始数据进行统计分组（简单分组、复合分组、单项分组和组距分组）和累计次数；
- 2、简单统计图绘制和分布曲线绘制；
- 3、基本指标计算（最大值、最小值、算术平均数、中位数、众数、全距、标准差、、峰度、偏度等）。

三、实验要求

数据：某班统计学期末考试成绩

序号	学号	性别	分数	序号	学号	性别	分数
1	0220111309	女	77	46	0220110932	女	74
2	0220111312	男	79	47	0220110944	女	66
3	0220111314	男	66	48	0220110945	男	79
4	0220111316	女	69	49	0220111003	男	61
5	0220111319	女	47	50	0220111008	男	76
6	0220111320	女	48	51	0220111014	男	84
7	0220111327	男	66	52	0220111015	男	83
8	0220111328	男	30	53	0220111018	女	74
9	0220111331	男	72	54	0220111022	女	78
10	0220111336	男	52	55	0220111024	女	62
11	0220111340	男	67	56	0220111025	男	72
12	0220111343	女	64	57	0220111032	女	75
13	0220111345	女	58	58	0220111034	女	82
14	0220111347	男	60	59	0220111035	男	80
15	0220111348	男	66	60	0220111038	男	33
16	0220111349	女	69	61	0220111041	女	76
17	0220111402	女	40	62	0220111042	男	70
18	0220111404	男	72	63	0220111045	男	73
19	0220111405	女	75	64	0220111046	女	73
20	0220111406	男	77	65	0220111101	男	81
21	0220111417	男	84	66	0220111107	女	60
22	0220111420	男	60	67	0220111109	男	75

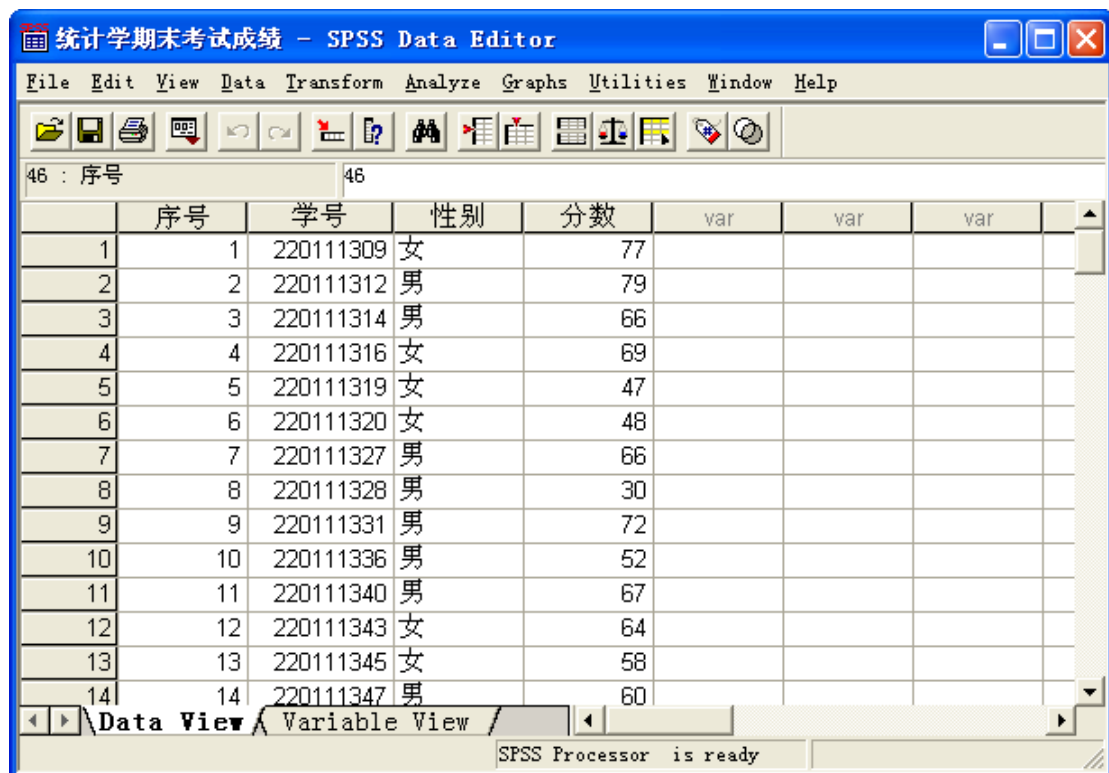
23	0220111424	女	59	68	0220111111	男	62
24	0220111426	女	70	69	0220111112	女	74
25	0220111429	女	74	70	0220111113	男	92
26	0220111435	男	39	71	0220111114	女	59
27	0220111438	男	72	72	0220111118	女	53
28	0220111440	女	69	73	0220111120	男	48
29	0220111445	男	65	74	0220111128	女	41
30	0220111504	女	48	75	0220111134	男	71
31	0220111526	女	60	76	0220111138	女	70
32	0220111527	女	72	77	0220111145	男	83
33	0220111533	男	42	78	0220111148	男	59
34	0220111536	男	58	79	0220111203	女	62
35	0220111610	男	51	80	0220111206	男	85
36	0220111612	男	56	81	0220111208	女	83
37	0220111617	女	62	82	0220111239	男	69
38	0220111621	女	68	83	0220111242	男	75
39	0220111622	女	71	84	0220111249	女	76
40	0220111625	男	48	85	0220111301	女	40
41	0220111626	男	56	86	0220111302	男	68
42	0220111639	女	76	87	0220111303	男	75
43	0220111641	男	59	88	0220111308	女	60
44	0220111642	男	47	89	0220111310	男	67
45	0220110930	男	62				

- 1、按照“性别”标志进行分组。
- 2、按照“分数”进行分组：60分以下、60-70分、70-80分、80-90分、90分以上；并绘制直方图和正态曲线。
- 3、进行分数和性别的双变量分析，并绘制对比条形图。
- 4、基本指标计算：
 - (1) 在原始数据基础上，计算分数的算术平均数、中位数、标准差等；
 - (2) 在组距数列基础上，计算统计学算术平均分数、中位数、众数、标准差等；
 - (3) 在性别分组数列基础上，比较男同学和女同学的成绩高低。

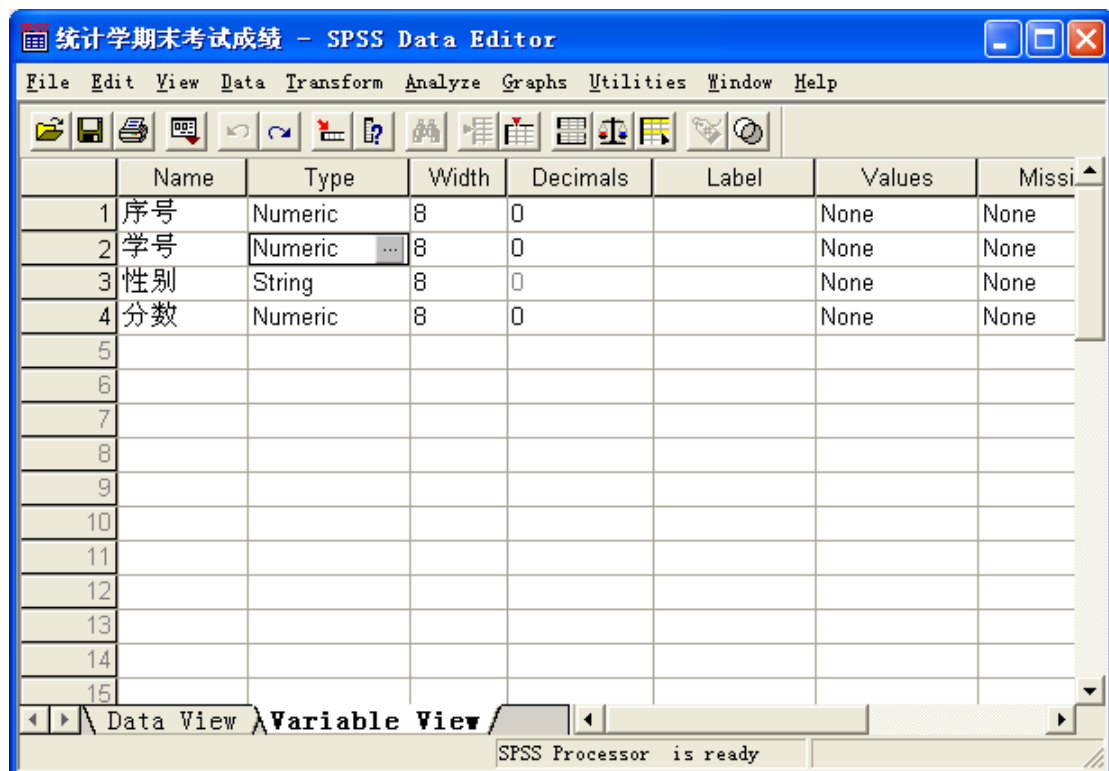
四、实验步骤

(一) 按性别分组

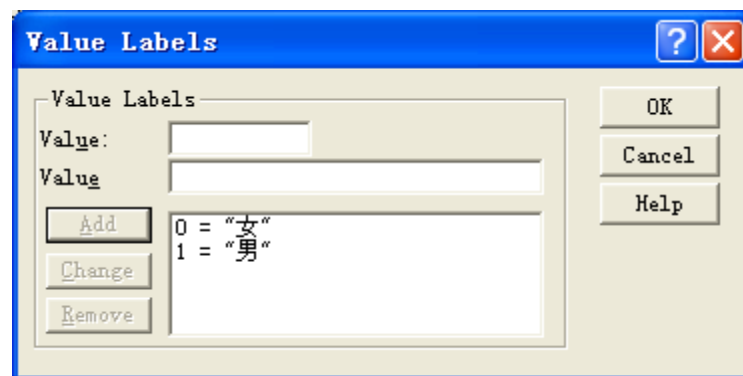
- 1、载入需要分析的数据；



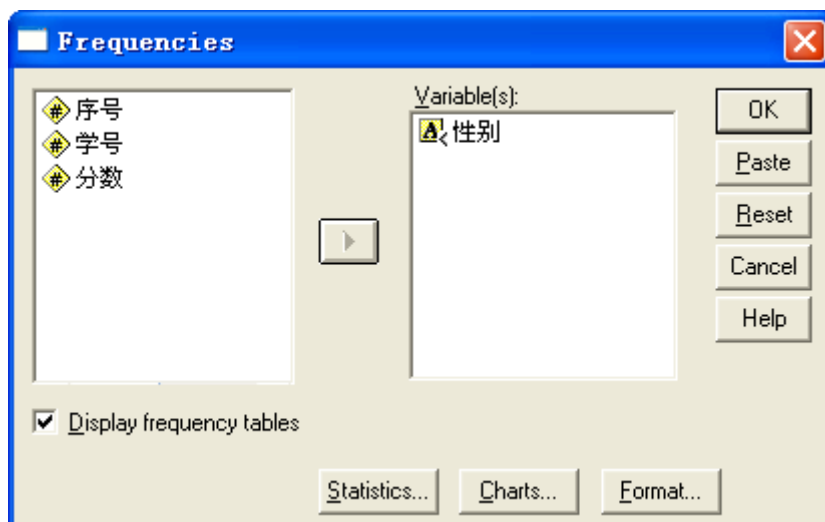
2、在 Variable View 中设置变量中文名称以及变量类型



3、将性别转化为数字变量；



4、Analyze → Descriptive Statistics → Frequencies:
Variable→性别



5、点击 OK，执行运算。输出结果如下：

性 别

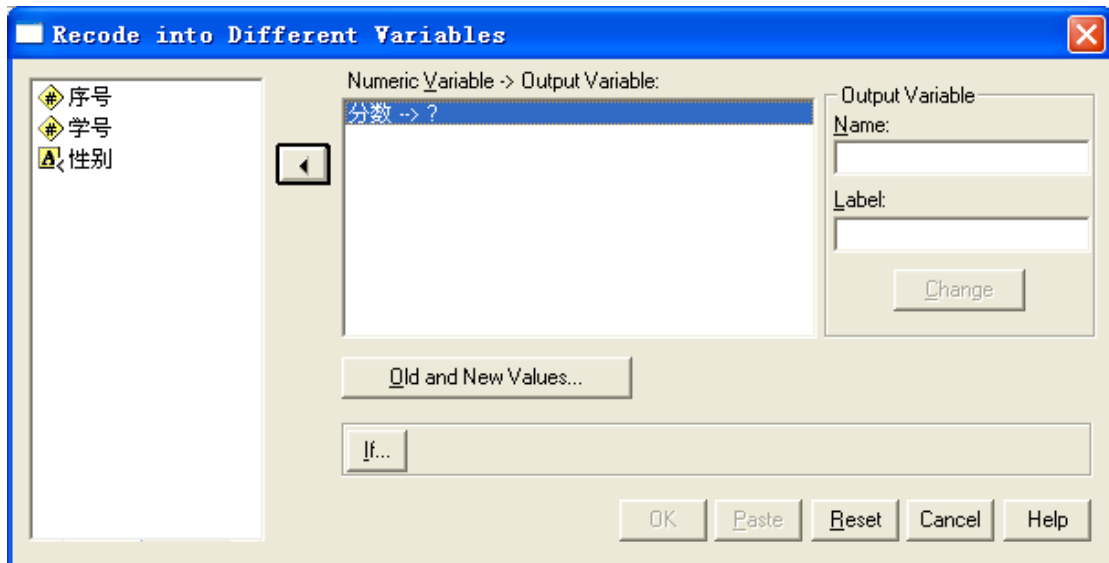
	Frequency	Percent	Valid Percent	Cumulative Percent
男	49	55.1	55.1	55.1
女	40	44.9	44.9	100.0
Total	89	100.0	100.0	

(二) 按分数进行分组

分两步进行，先对原变量重新编码，然后按照新的编码计算频数。

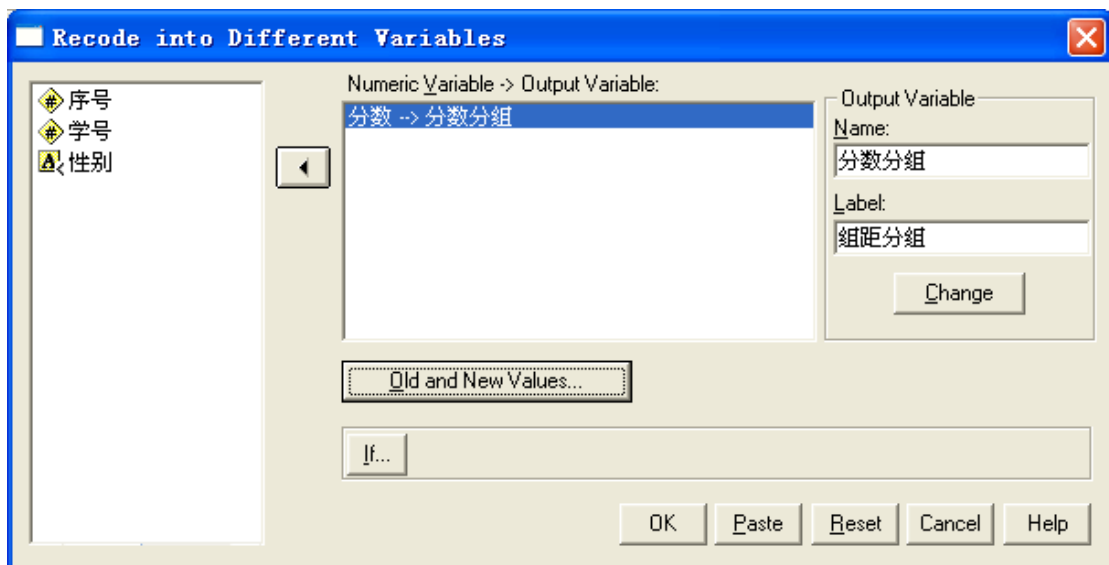
1、重新编码

(1) Transform → Recode → Recode Into Different Variables



(2) Numeric Variable -> Output Variable → 分数

- Output Variable 对话框：Name → 分数分组（输出变量名称）
Label → 组距分组
- 点击 Change 按钮确定输入。

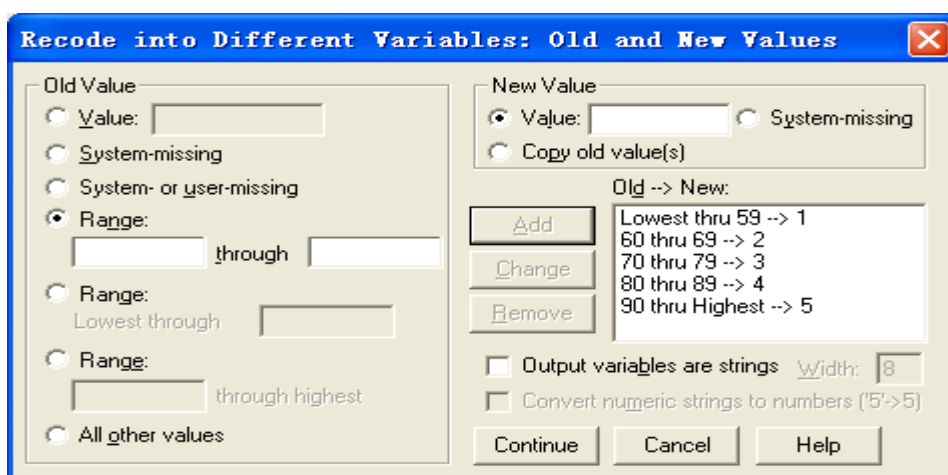


(3) 点击 Old and New Values

将原变量“分数”转变为新的变量“分数分组”。

- 在 Old Value 对话框中，分别选择三个“Range”单选项；
- 每次选择 Range 选项时，分别在填写框中填入需要转变为新变量值的原变量的变动范围，随后在 New Value 对话框中输入对应的新变量值；
- 每次选择和输入后，都需要点击 Old->New 对话框的 Add 按钮将新旧变量加入转换框中。
- 点击 Continue。

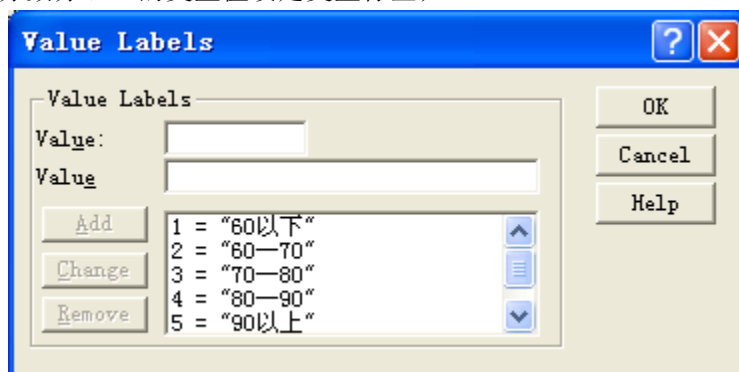
此例根据分数标志将原数列分为 5 组。



(4) 点击 OK，执行运算。新变量增加到数据表中。

序号	学号	性别	分数	分数分组	var	var
1	220111309	女	77	3		
2	220111312	男	79	3		
3	220111314	男	66	2		
4	220111316	女	69	2		
5	220111319	女	47	1		
6	220111320	女	48	1		
7	220111327	男	66	2		
8	220111328	男	30	1		
9	220111331	男	72	3		
10	220111336	男	52	1		
11	220111340	男	67	2		
12	220111343	女	64	2		
13	220111345	女	58	1		
14	220111347	男	60	2		

(5) 将“分数分组”的变量值设定变量标签：



2、按照新变量计算频数；

1) Analyze → Descriptive Statistics → Frequencies: Variable → 组距分组(分数分组)

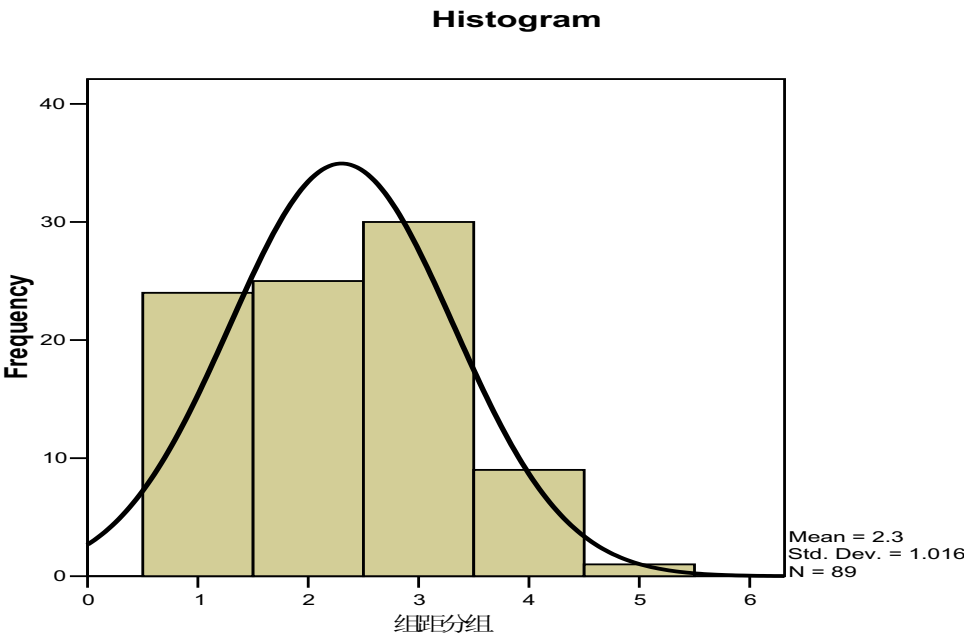
2) Charts 选项中选择直方图和正态曲线，其他选项保持默认；



3) 点击 OK，执行运算，结果如下：

×é³/à·Ö×é

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	60ÖÖiÄ	24	27.0	27.0	27.0
	60jª70	25	28.1	28.1	55.1
	70jª80	30	33.7	33.7	88.8
	80jª90	9	10.1	10.1	98.9
	90ÖÖÉi	1	1.1	1.1	100.0
	Total	89	100.0	100.0	

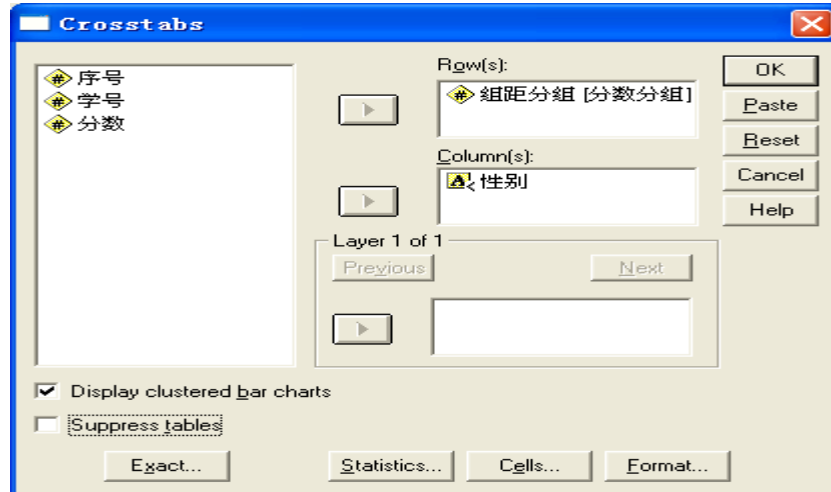


（三）进行分数和性别的双变量分析

1、对分数和性别进行交叉分析；

1) Analyze → Descriptive Statistics → Crosstabs

- Row（行）→ 分数分组
- Column（列）→ 性别

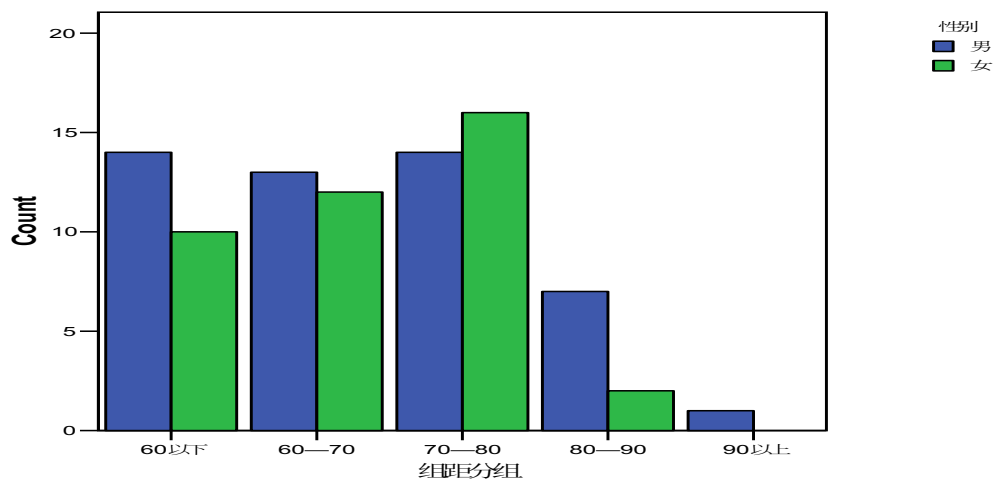


2) 选择“Display clustered bar charts”（条状图），其他选项保持默认，点击 OK，执行运算。

×é¼à·Ö×é * ÐÔ±ð Crosstabulation

Count		ÐÔ±ð		Total
		ÄÐ	Ä®	
×é¼à·Ö×é	60ÖÖİÄ	14	10	24
	60ıª70	13	12	25
	70ıª80	14	16	30
	80ıª90	7	2	9
	90ÖÖİİ	1	0	1
Total		49	40	89

Bar Chart



(四) 基本指标计算

1、未分组数列；

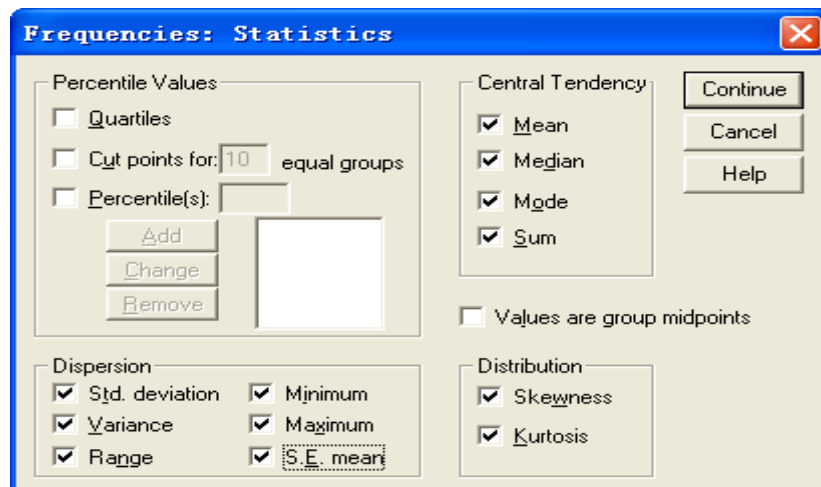
1) Analyze → Descriptive Statistics → Frequencies :

- Variable → 分数
- 取消 Display frequency tables 的默认选中

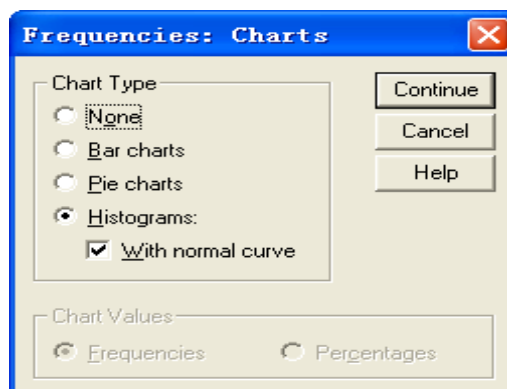


2) Statistic 选项中选择：

- Central Tendency (集中趋势) : 全选
- Dispersion (分散参数): 全选
- Distribution (分布参数): 全选



3) Chart 选择项中选择直方图和分布曲线；



4) 点击 OK，执行运算。Output 输出结果如下：

Statistics

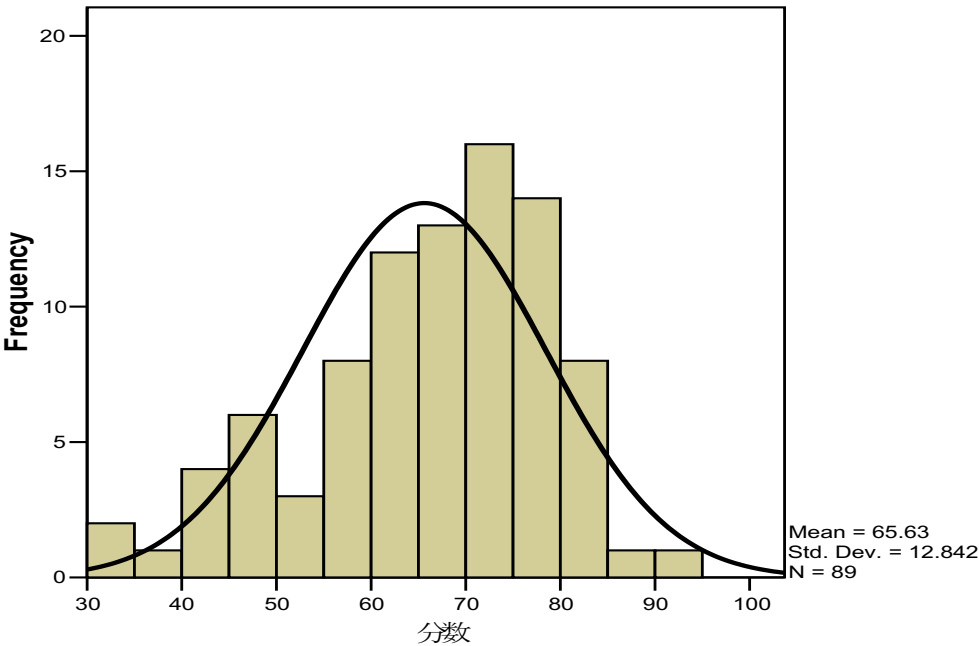
•

分数

N	Valid	89
	Missing	0
Mean (平均数)		65.63
Std. Error of Mean (平均数标准误)		1.361
Median (中位数)		68.00
Mode (众数)		60(a)
Std. Deviation (标准差)		12.842
Variance (方差)		164.918
Skewness (偏度)		-.660
Std. Error of Skewness (偏度标准误)		.255
Kurtosis (峰度)		.071
Std. Error of Kurtosis (峰度标准误)		.506
Range (全距)		62
Minimum (最小值)		30
Maximum (最大值)		92
Sum (合计)		5841

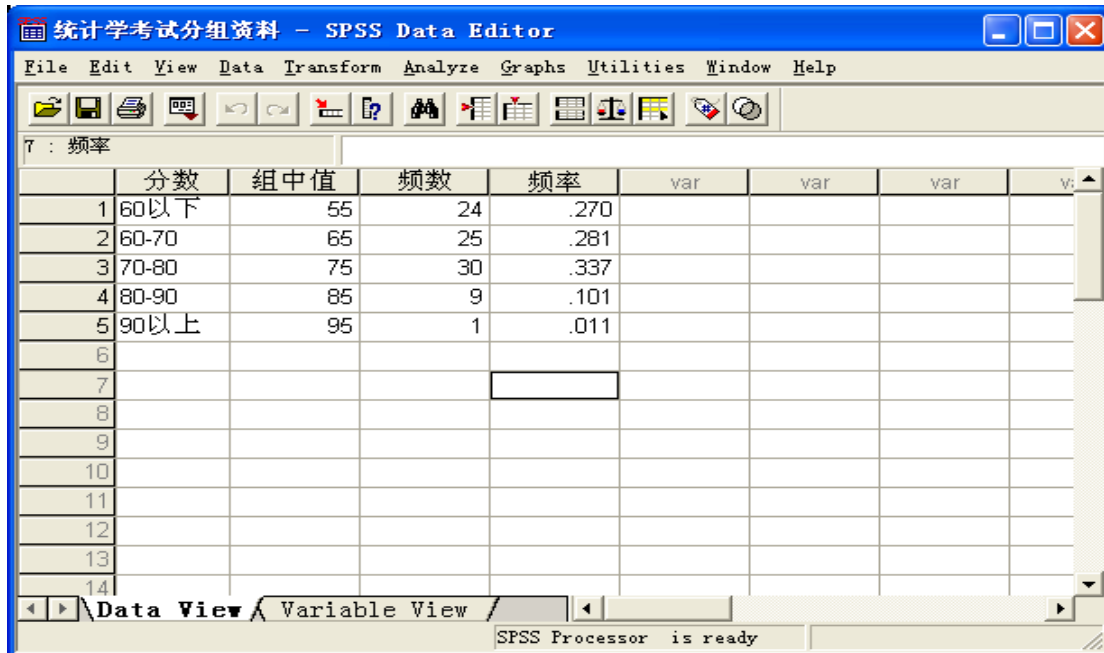
a. Multiple modes exist. The smallest value is shown

Histogram



2、分组数列基本指标计算

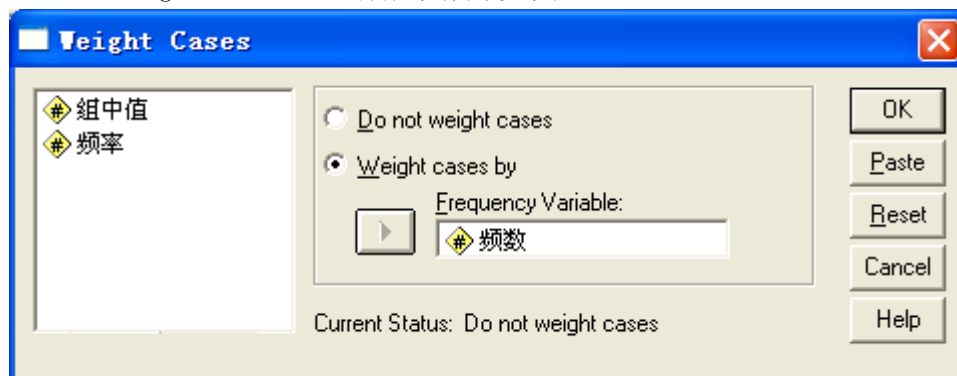
先将按分数分组的结果形成新的数据表（直接拷贝到新的数据表中），手工计算出组中值，再采用加权算术平均的形式计算。



SPSS Data Editor window titled "统计学考试分组资料 - SPSS Data Editor". The window shows a data table with the following columns: 分数 (Score), 组中值 (Midpoint), 频数 (Frequency), 频率 (Frequency), and several empty columns labeled 'var'. The data is as follows:

	分数	组中值	频数	频率	var	var	var	var
1	60以下	55	24	.270				
2	60-70	65	25	.281				
3	70-80	75	30	.337				
4	80-90	85	9	.101				
5	90以上	95	1	.011				
6								
7								
8								
9								
10								
11								
12								
13								
14								

1) Data → Weighted Case: 选择频数作为权数;



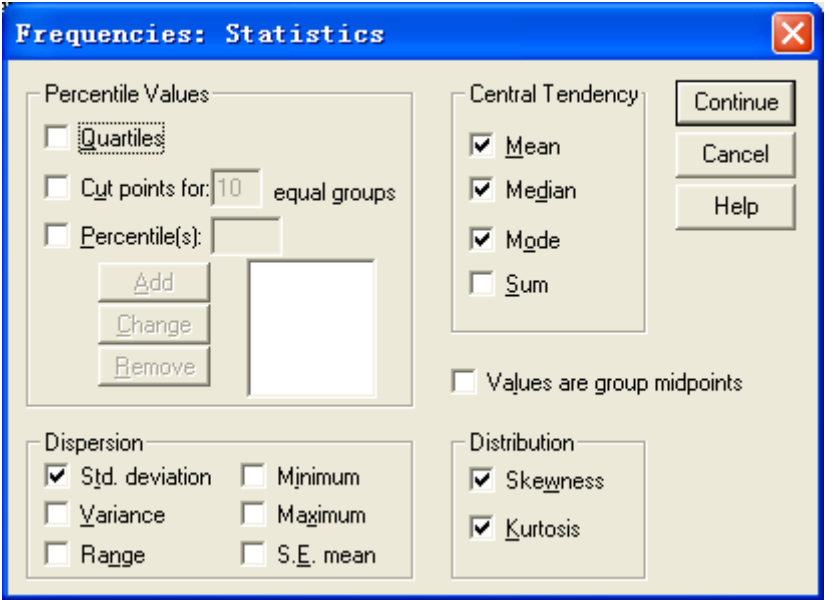
2) Analyze → Descriptive Statistics → Frequencies :

- Variable → 组中值
- Display frequency tables 选中



3) Statistic 选项中选择如下图:

- Central Tendency (集中趋势)
- Dispersion (分散参数):
- Distribution (分布参数):
- Values are group midpoint:

The image shows the 'Frequencies: Statistics' dialog box in SPSS. It is divided into several sections: 'Percentile Values' with checkboxes for 'Quartiles', 'Cut points for: 10 equal groups', and 'Percentile(s):' with 'Add', 'Change', and 'Remove' buttons; 'Central Tendency' with checked boxes for 'Mean', 'Median', and 'Mode', and an unchecked box for 'Sum'; 'Dispersion' with checked boxes for 'Std. deviation' and 'Range', and unchecked boxes for 'Variance', 'Minimum', 'Maximum', and 'S.E. mean'; and 'Distribution' with checked boxes for 'Skewness' and 'Kurtosis'. There are also buttons for 'Continue', 'Cancel', and 'Help' on the right, and an unchecked checkbox for 'Values are group midpoints' at the bottom right.

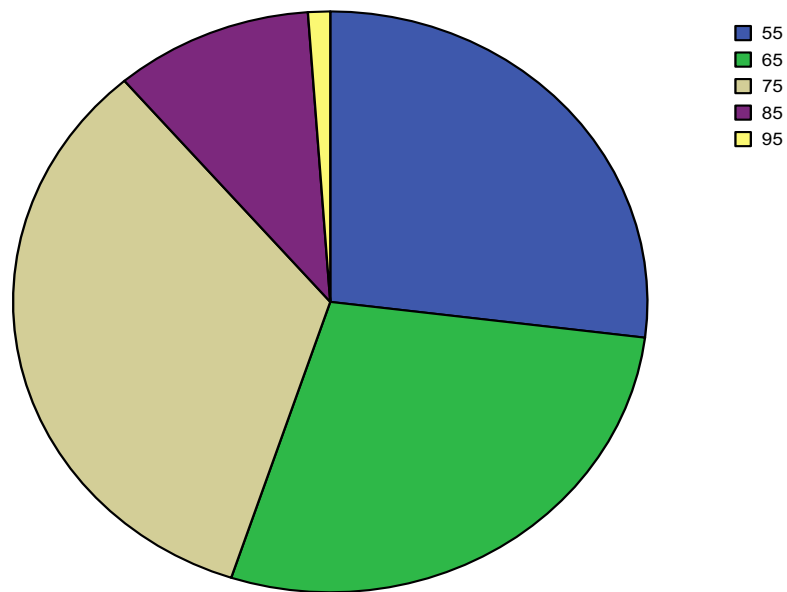
4) Chart 选择项中分别选择: 饼图 (并选择百分比) 和直方图以及分布曲线;

5) 点击 OK, 执行运算。Output 输出结果如下:

Statistics

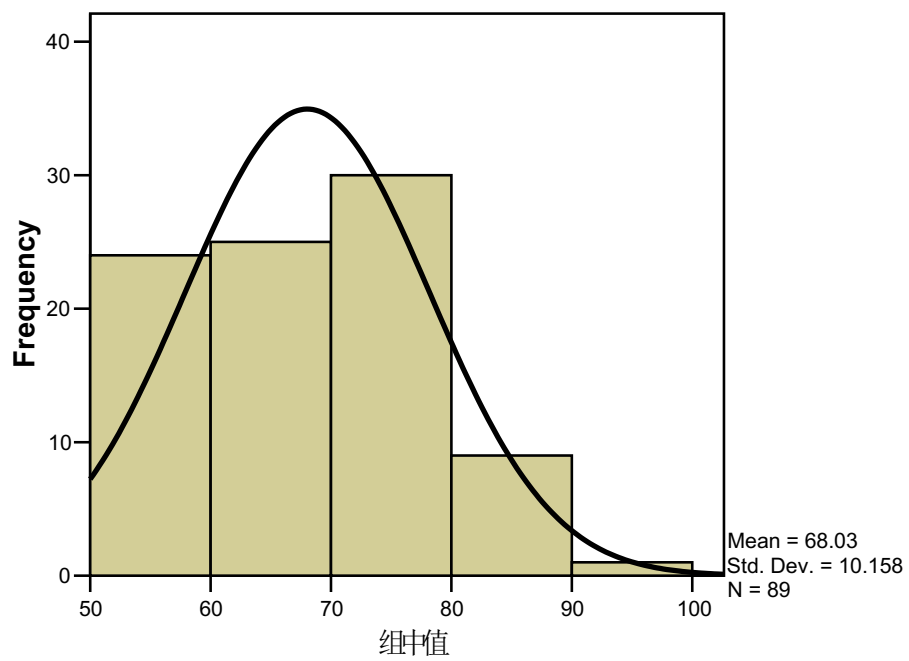
×éÖÐÖμ		
N	Valid	89
	Missing	0
Mean		68.03
Median		65.00
Mode		75
Std. Deviation		10.158
Skewness		.220
Std. Error of Skewness		.255
Kurtosis		-.775
Std. Error of Kurtosis		.506

组中值



Cases weighted by 频数

Histogram

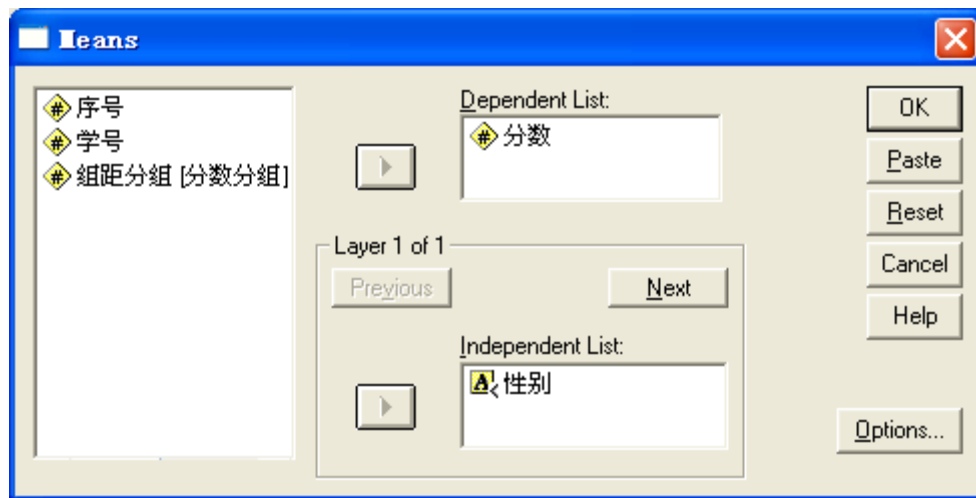


Cases weighted by 频数

3、男女同学平均分数比较

1) Analyze → Compare Means → means

- Dependent List → 分数
- Independent List → 性别



2) 其他选项保持默认，点击 OK 执行运算。

Report

分数

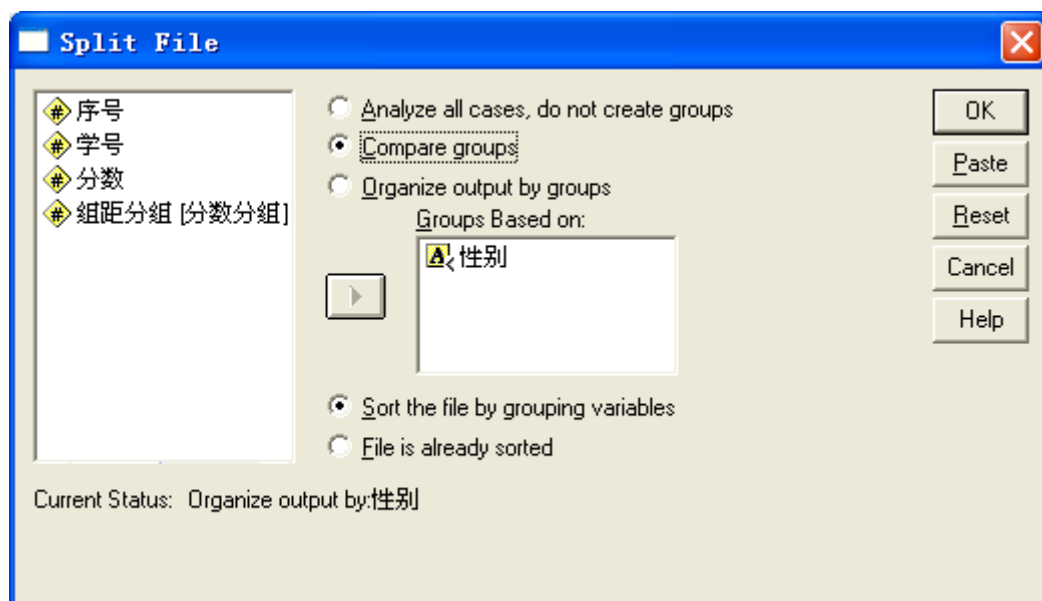
性别	Mean	N	Std. Deviation
男	65.86	49	13.954
女	65.35	40	11.504
Total	65.63	89	12.842

4、分数分组数列：男女同学分数比较

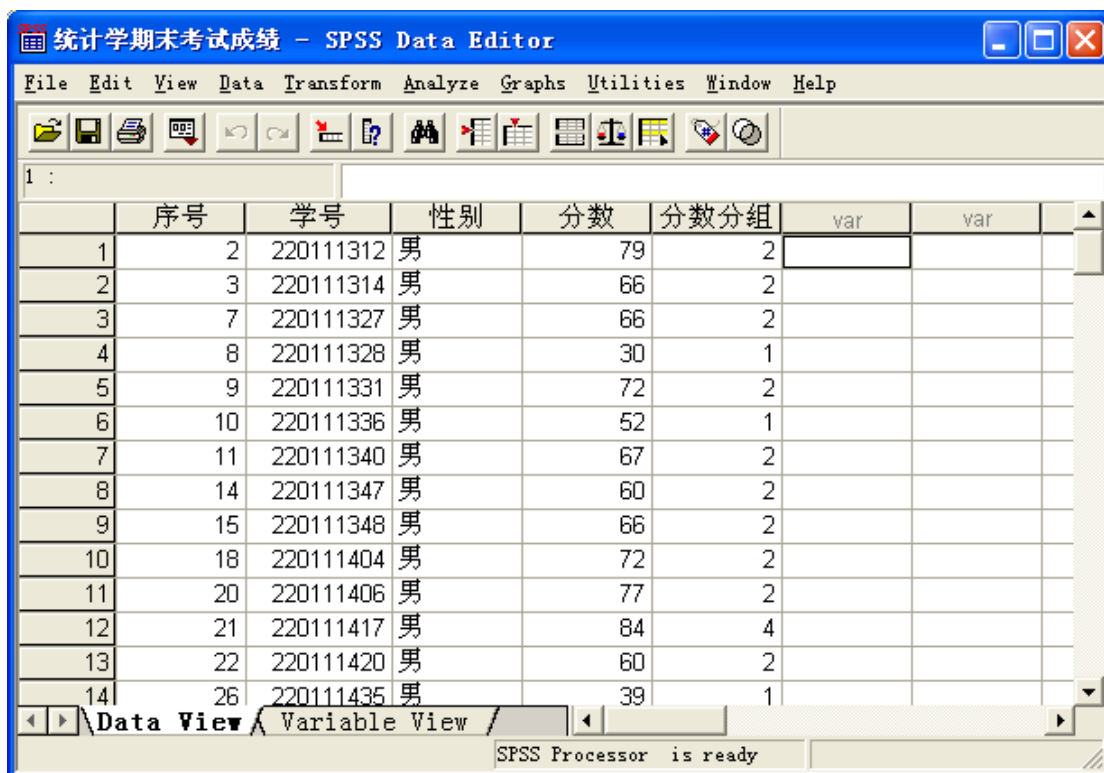
1) Data → Split File

2) 选中 Compare Groups

3) Group Base On→ 性别



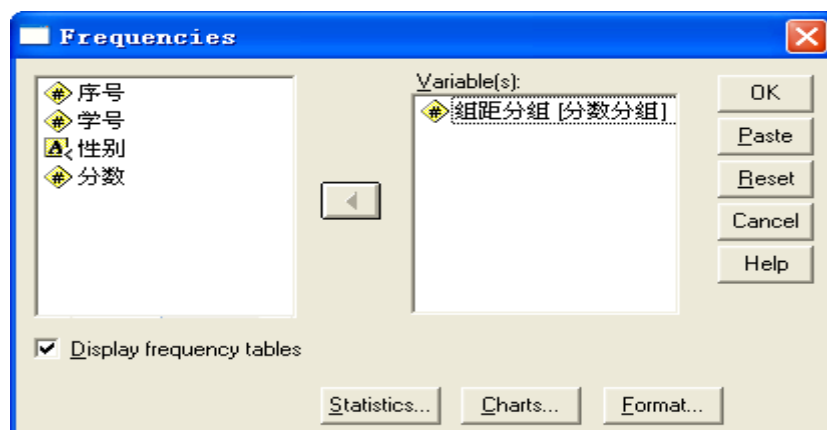
4) 点击 OK，结果在数据表中显示：



	序号	学号	性别	分数	分数分组	var	var
1	2	220111312	男	79	2		
2	3	220111314	男	66	2		
3	7	220111327	男	66	2		
4	8	220111328	男	30	1		
5	9	220111331	男	72	2		
6	10	220111336	男	52	1		
7	11	220111340	男	67	2		
8	14	220111347	男	60	2		
9	15	220111348	男	66	2		
10	18	220111404	男	72	2		
11	20	220111406	男	77	2		
12	21	220111417	男	84	4		
13	22	220111420	男	60	2		
14	26	220111435	男	39	1		

5) Analyze → Descriptive Statistics → Frequencies :

- Variable → 分数分组
- Display frequency tables 选中



6) Statistic 选项中选择：

- Central Tendency (集中趋势) : 全选
- Dispersion (分散参数): 全选
- Distribution (分布参数): 全选
- Values are group midpoint: 选中

Frequencies: Statistics

Percentile Values

☐ Quartiles

☐ Cut points for: 10 equal groups

☐ Percentile(s):

Add

Change

Remove

Central Tendency

☒ Mean

☒ Median

☒ Mode

☒ Sum

☒ Values are group midpoints

Dispersion

☒ Std. deviation

☒ Variance

☒ Range

☒ Minimum

☒ Maximum

☒ S.E. mean

Distribution

☒ Skewness

☒ Kurtosis

Continue

Cancel

Help

7) Chart 选择项中分别选择：直方图和分布曲线；

Frequencies: Charts

Chart Type

☐ None

☐ Bar charts

☐ Pie charts

☒ Histograms:

☒ With normal curve

Chart Values

☒ Frequencies

☐ Percentages

Continue

Cancel

Help

8) 点击 OK 执行运算。

×é¾à·Ö×é

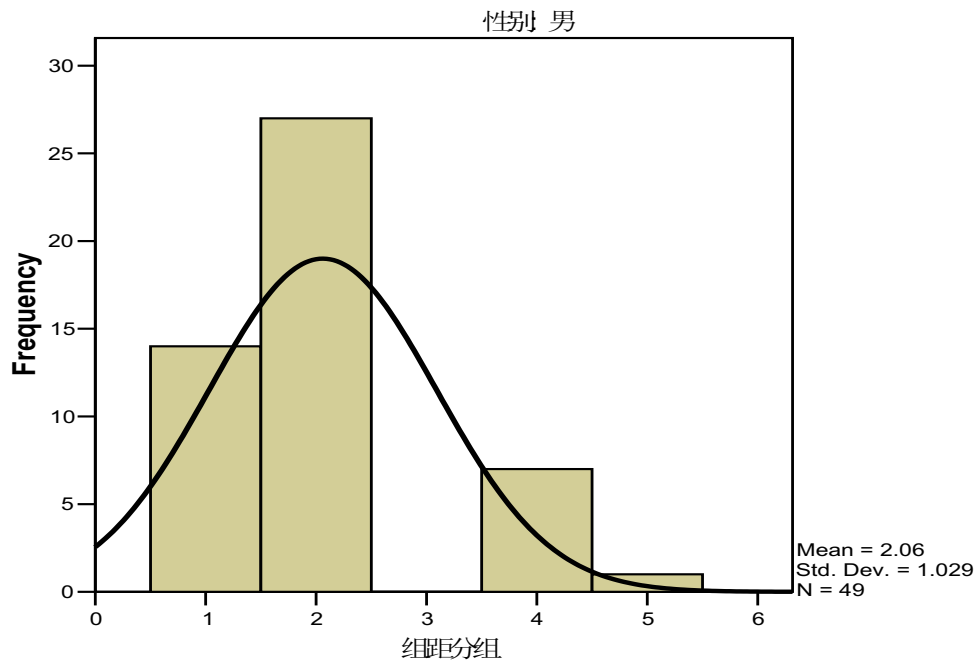
ĐÔ±ð			Frequency	Percent	Valid Percent	Cumulative Percent
ÄĐ	Valid	60ÖÖİÄ	14	28.6	28.6	28.6
		60-70	27	55.1	55.1	83.7
		80-90	7	14.3	14.3	98.0
		90ÖÖÊİ	1	2.0	2.0	100.0
		Total	49	100.0	100.0	
Ä®	Valid	60ÖÖİÄ	10	25.0	25.0	25.0
		60-70	28	70.0	70.0	95.0
		80-90	2	5.0	5.0	100.0
		Total	40	100.0	100.0	

Statistics

×é¾à·Ö×é

ÄÐ	N	Valid	49
		Missing	0
	Mean		2.06
	Std. Error of Mean		.147
	Median		2.00
	Mode		2
	Std. Deviation		1.029
	Variance		1.059
	Skewness		1.189
	Std. Error of Skewness		.340
	Kurtosis		.831
	Std. Error of Kurtosis		.668
	Range		4
	Minimum		1
	Maximum		5
	Sum		101
Ä®	N	Valid	40
		Missing	0
	Mean		1.85
	Std. Error of Mean		.105
	Median		2.00
	Mode		2
	Std. Deviation		.662
	Variance		.438
	Skewness		1.286
	Std. Error of Skewness		.374
	Kurtosis		4.217
	Std. Error of Kurtosis		.733
	Range		3
	Minimum		1
	Maximum		4
	Sum		74

组距分组



组距分组

