chp18 统计学

基本上，统计学题型可以分成两大部分

1.未分组数据的题型（Ungrouped Data）

2.分组数据的题型（Grouped Data）

- 未分组数据的题型

例题： 现有数据

23	54	55	78	90	12	34	56	45	48
76	29	47	35	54	32	33	54	67	89

求

1. 算术平均数（集中趋势）
2. 众数（集中趋势）
3. 中位数（集中趋势）
4. 上四分位数（离中趋势）
5. 下四分位数（离中趋势）
6. 四分位距（离中趋势）
7. 四分位差（离中趋势）
8. 标准差（离中趋势）
9. 方差（离中趋势）
10. 平均差（离中趋势）
11. 变异系数

1. 所谓的算术平均数就是将所有的数据相加然后除以个数~也就是数据数量，代号为 $\bar{x}$

   $$\bar{x}=\frac{23+54+55+78+90+12+34+56+45+48+76+29+47+35+54+32+33+54+67+89}{20}$$

   $=50.55$

2. 众数就是出现最多次的那个数字（如果有两个数字出现次数一样多，那答案就有两个，以此类推）由小到大编排后会比较容易发现答案

12	23	29	32	33	34	35	45	47	48
54	54	54	55	56	67	76	78	89	90

   众数为 54

3. 中位数就是在最中间的数字（如果数据总数为奇数 21 个，那中位数就是第 $\frac{n+1}{2}=\frac{21+1}{2}=11$ 项）

   这题数据数量为偶数所以是第 $\frac{n}{2}=\frac{20}{2}=10$ 项与第 $\frac{n}{2}+1=\frac{20}{2}+1=11$ 项的平均数，那就是 48 与 54 之间的一半，所以中位数为 $\frac{48+54}{2}=51$

4. 上四分位数代号 Q_3，通常只要把数据线性排列就能够做到

12	23	29	32	33	34	35	45	47	48	54	54	54	55	56	67	76	78	89	90

   第 15 与第 16 项之间，所以 $Q_3=\frac{56+67}{2}=61.5$

5. 下四分位数代号是 Q_1，也是把数据线性排列就可以做到

   在第 5 与第 6 项之间，$Q_1=\frac{33+34}{2}=33.5$

6. 四分位距就是 Q_3-Q_1，也就是 $61.5-33.5=28$

7. 四分位差就是$Q.D.=\frac{Q_3-Q_1}{2}$，也就是$\frac{61.5-33.5}{2}=14$

8. 标准差公式为$s=\sqrt{\frac{\sum(x_i-\overline{x})^2}{n}}$ 或$s=\sqrt{\frac{\sum x_i^{\,2}}{n}-\overline{x}^2}$

x_i是数据本身，$\overline{x}$是平均数，n是数据总数，所以必须列表出来

x_i	$x_i^{\,2}$
12	144
23	529
29	841
32	1024
33	1089
34	1156
35	1225
45	2025
47	2209
48	2304
54	2916
54	2916
54	2916
55	3025
56	3136
67	4489
76	5776
78	6084
89	7921
90	8100
	59825

$$s=\sqrt{\frac{59825}{20}-50.55^2}\approx 20.88$$

9. 方差就是s^2，也就是标准差的平方$s^2=\left(\sqrt{\frac{59825}{20}-50.55^2}\right)^2=435.9475$

10. 平均差 $\frac{\sum|x_i-\bar{x}|}{n}$ 也就是每个数据去减平均数再全部总和起来再除以数据个数

x_i	$x_i-\bar{x}$	$\lvert x_i-\bar{x}\rvert$
12	12-50.55=-38.55	38.55
23	23-50.55=-27.55	27.55
29	29-50.55=-21.55	21.55
32	32-50.55=-18.55	18.55
33	33-50.55=-17.55	17.55
34	34-50.55=-16.55	16.55
35	35-50.55=-15.55	15.55
45	45-50.55=-5.55	5.55
47	47-50.55=-3.55	3.55
48	48-50.55=-2.55	2.55
54	54-50.55=3.45	3.45
54	54-50.55=3.45	3.45
54	54-50.55=3.45	3.45
55	55-50.55=4.45	4.45
56	56-50.55=5.45	5.45
67	67-50.55=16.45	16.45
76	76-50.55=25.45	25.45
78	78-50.55=27.45	27.45
89	89-50.55=38.45	38.45
90	90-50.55=39.45	39.45
		335

平均差=$\frac{335}{20}$=16.75

11. 变异系数代表符号V，$V=\frac{s}{\bar{x}}\times100\%$

所以$V=\frac{20.88}{50.55}\times100\%=41.3\%$

得到频数表的题型

组别	频数 f_i
<0	0
0-9	15
10-19	35
20-29	55
30-39	105
40-49	220
50-59	160
60-69	110
70-79	50
80-89	30
90-99	20
	800

求

1. 加权平均数（集中趋势）
2. 众数（集中趋势）
3. 中位数（集中趋势）
4. 上四分位数（离中趋势）
5. 下四分位数（离中趋势）
6. 四分位距（离中趋势）
7. 四分位差（离中趋势）
8. 第 7 十分位数 D_7（离中趋势）
9. 第 48 百分位数 P_{48}（离中趋势）
10. 标准差（离中趋势）
11. 方差（离中趋势）
12. 平均差（离中趋势）
13. 变异系数

1. 加权平均数

   得到频数表的题型在算平均数的时候没有办法把所有数据找出来~所以必须要利用乘的原理，但在那之前，一定要做到每次拿到频数表必做事项也就是把它改为组界，并把组中点做出来，代号为 x_i（感觉和之前数据的是一样的，对！我们就是要把组界的东西换成和数据一样的形式，就是把组中点当成那么多的数据的代表，把组中点当成数据）再把 x_i 和 f_i 乘起来，再加起来

组别	频数 f_i	组界	x_i	$x_i f_i$
0-9	15	-0.5 - 9.5	4.5	67.5
10-19	35	9.5 - 19.5	14.5	507.5
20-29	55	19.5 -29.5	24.5	1347.5
30-39	105	29.5 -39.5	34.5	3622.5
40-49	220	39.5 -49.5	44.5	9790
50-59	160	49.5 -59.5	54.5	8720
60-69	110	59.5 -69.5	64.5	7095
70-79	50	69.5 -79.5	74.5	3725
80-89	30	79.5 -89.5	84.5	2535
90-99	20	89.5 -99.5	94.5	1890
$\sum f_i =$	800		$\sum x_i f_i =$	39300

$$\overline{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{39300}{800} = 49.125$$

2. 众数，要先判断频数最多的哪一组，所以坐落在 220 那组也就是 40-49 那组

$$L + \left(\frac{d_1}{d_1 + d_2}\right)(U - L)$$

L是众数所在组的下组界

U-L是众数所在组的组距

d_1是众数所在组与前一组的频数之差

d_2是众数所在组与后一组的频数之差

所以众数为$39.5+\left(\frac{115}{115+60}\right)(49.5-39.5)=46.07$

3. 中位数公式为$M_e=L+\frac{\frac{n}{2}-f_s}{f}(U-L)$ 所谓的n其实就是$\sum f_i$ 就是频数总数，中位数只需要做多一个累积频数就可以做到答案了

组别	频数 f_i	组界	累积频数
0-9	15	-0.5 - 9.5	15
10-19	35	9.5 - 19.5	50
20-29	55	19.5 -29.5	105
30-39	105	29.5 -39.5	210
40-49	220	39.5 -49.5	430
50-59	160	49.5 -59.5	590
60-69	110	59.5 -69.5	700
70-79	50	69.5 -79.5	750
80-89	30	79.5 -89.5	780
90-99	20	89.5 -99.5	800
$\sum f_i=$	800		

先用$\frac{n}{2}=\frac{800}{2}=400$，并判断的出坐落在 39.5 -49.5 这一组别，公式里的 f_s 就是上一组别的累积频数也就是 210，f 就是这一组别的频数就是 220，全部带入公式得到

$$M_e=39.5+\frac{400-210}{220}(49.5-39.5)=48.14$$

4. 上四分位数道理相似，先用$\frac{3}{4}n=\frac{3}{4}\times 800=600$，并判断的出坐落在 59.5 -69.5 这一组别，公式里的 f_s 就是上一组别的累积频数也就是 590，f 就是这一组别的频数就是 110，全部带入公式得到

$$Q_3=59.5+\frac{600-590}{110}(69.5-59.5)=60.41$$

5. 下四分位数道理相似，先用$\frac{1}{4}n=\frac{1}{4}\times 800=200$，并判断的出坐落在 29.5 -39.5 这一组别，公式里的 f_s 就是上一组别的累积频数也就是 105，f 就是这一组别的频数就是 105，全部带入公式得到

$$Q_1=29.5+\frac{200-105}{105}(39.5-29.5)=38.55$$

6. 四分位距$60.41-38.55=21.86$

7. 四分位差就是 $\frac{60.41-38.55}{2}=10.93$

8. 第 7 十分位数道理相似，先用 $\frac{7}{10}n=\frac{7}{10}\times800=560$，并判断的出坐落在 49.5 -59.5 这一组别，公式里的 f_s 就是上一组别的累积频数也就是 430， f 就是这一组别的频数就是 160，全部带入公式得到

$$D_7=49.5+\frac{560-430}{160}(59.5-49.5)=57.63$$

9. 第 48 百分位数道理相似，先用 $\frac{48}{100}n=\frac{48}{100}\times800=384$，并判断的出坐落在 39.5 -49.5 这一组别，公式里的 f_s 就是上一组别的累积频数也就是 210， f 就是这一组别的频数就是 220，全部带入公式得到

$$P_{48}=39.5+\frac{384-210}{220}(49.5-39.5)=47.41$$

10. 标准差的公式给有分组的题目 $S=\sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}-\overline{x}^2}$ 或 $S=\sqrt{\frac{\sum (x_i-\overline{x})^2 f_i}{\sum f_i}}$

组别	频数 f_i	组界	x_i	x_i^2	$x_i^2 f_i$
0-9	15	-0.5 - 9.5	4.5	20.25	303.75
10-19	35	9.5 - 19.5	14.5	210.25	7358.75
20-29	55	19.5 -29.5	24.5	600.25	33013.75
30-39	105	29.5 -39.5	34.5	1190.25	124976.3
40-49	220	39.5 -49.5	44.5	1980.25	435655
50-59	160	49.5 -59.5	54.5	2970.25	475240
60-69	110	59.5 -69.5	64.5	4160.25	457627.5
70-79	50	69.5 -79.5	74.5	5550.25	277512.5
80-89	30	79.5 -89.5	84.5	7140.25	214207.5
90-99	20	89.5 -99.5	94.5	8930.25	178605
$\sum f_i=$	800			$\sum x_i^2 f_i=$	2204500

带入公式得 $s=\sqrt{\frac{2204500}{800}-49.125^2}\approx18.50$

11. 方差就是 $s^2=\left(\sqrt{\frac{2204500}{800}-49.125^2}\right)^2\approx342.36$

12. 平均差 $\frac{\sum|x_i-\overline{x}|f_i}{\sum f_i}$ 所以我们列表

组别	频数 f_i	组界	x_i	$x_i-\overline{x}$	$\lvert x_i-\overline{x}\rvert$	$\lvert x_i-\overline{x}\rvert f_i$
0-9	15	-0.5 - 9.5	4.5	-44.625	44.625	669.375
10-19	35	9.5 - 19.5	14.5	-34.625	34.625	1211.875
20-29	55	19.5 -29.5	24.5	-24.625	24.625	1354.375
30-39	105	29.5 -39.5	34.5	-14.625	14.625	1535.625
40-49	220	39.5 -49.5	44.5	-4.625	4.625	1017.5
50-59	160	49.5 -59.5	54.5	5.375	5.375	860
60-69	110	59.5 -69.5	64.5	15.375	15.375	1691.25
70-79	50	69.5 -79.5	74.5	25.375	25.375	1268.75
80-89	30	79.5 -89.5	84.5	35.375	35.375	1061.25
90-99	20	89.5 -99.5	94.5	45.375	45.375	907.5
$\sum f_i=$	800				$\sum\lvert x_i-\overline{x}\rvert=$	11577.5

众数 $\frac{11577.5}{800}=14.47$

13. 变异系数 $V=\frac{18.5}{49.125}\times100\%=37.66\%$