## 14.4.3　用频率直方图估计总体分布

## 14.4.4　百分位数

学习目标　1.用频率直方图估计总体分布.2.结合实例，能用样本估计百分位数.3.理解百分位数的统计含义．

知识点　百分位数

1．*k*百分位数的定义：

一般地，一组数据的*k*百分位数是这样一个值*pk*，它使得这组数据中至少有*k*%的数据小于或等于*pk*，且至少有(100－*k*)%的数据大于或等于*pk*.

2．计算一组*n*个数据的大样本的*k*百分位数的一般步骤如下：

第1步　将所有数值按从小到大的顺序排列；

第2步　计算*k*·；

第3步　如果结果为整数，那么*k*百分位数位于第*k*·位和下一位数之间，通常取这两个位置上数值的平均数为*k*百分位数；

第4步　如果*k*·不是整数，那么将其向上取整(即其整数部分加上1)，在该位置上的数值即为*k*百分位数．

3．四分位数

25,50,75这三个百分位数把一组由小到大排列后的数据分成四等份，因此称为四分位数，其中25百分位数也称为下四分位数，75百分位数也称为上四分位数．

1．若一组样本数据各不相等，则其75百分位数大于25百分位数．(　√　)

2．若一组样本数据的10百分位数是23，则在这组数据中有10%的数据大于23.(　×　)

3．若一组样本数据的24百分位数是24，则在这组数据中至少有76%的数据大于或等于24.(　√　)

4．某次数学测试成绩的70百分位数是85分，则至少有70%的同学测试成绩小于或等于85分．(　√　)

一、用频率直方图估计总体分布

例1　为了了解高一年级学生的体能情况，某校抽取部分学生进行一分钟跳绳次数测试，将所得数据整理后，画出频率直方图(如图所示)，图中从左到右各小矩形的面积之比为2∶4∶17∶15∶9∶3，第二小组的频数为12.

(1)第二小组的频率是多少？样本容量是多少？

(2)若次数在110以上(含110次)为达标，则该校高一年级全体学生的达标率约是多少？

解　(1)频率直方图是以面积的形式来反映数据落在各小组内的频率大小的，

因此第二小组的频率为＝0.08.

因为第二小组的频率＝，

所以样本容量＝＝＝150.

(2)由频率直方图可知该校高一年级全体学生的达标率约为×100%＝88%.

反思感悟　频率直方图反映了样本在各个范围内取值的可能性，由抽样的代表性利用了样本在某一范围内的频率，可近似地估计在这一范围内的可能性．

跟踪训练1　为检查某工厂所生产的8万台电风扇的质量，随机抽取20台，其无故障连续使用时限(单位：h)统计如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 分组 | 频数 | 频率 | 频率/组距 |
| [180,200) | 1 | 0.05 | 0.002 5 |
| [200,220) | 1 | 0.05 | 0.002 5 |
| [220,240) | 2 | 0.10 | 0.005 0 |
| [240,260) | 3 | 0.15 | 0.007 5 |
| [260,280) | 4 | 0.20 | 0.010 0 |
| [280,300) | 6 | 0.30 | 0.015 0 |
| [300,320) | 2 | 0.10 | 0.005 0 |
| [320,340] | 1 | 0.05 | 0.002 5 |
| 合计 | 20 | 1 | 0.050 0 |

(1)作出频率直方图；

(2)估计8万台电风扇中无故障连续使用时限不低于280 h的有多少台；

(3)假设同一组中的数据用该组区间的中点值代替，估计这8万台电风扇的平均无故障连续使用时限．

解　(1)频率直方图如图所示．

(2)由题意得8×(0.30＋0.10＋0.05)＝3.6，所以估计8万台电风扇中无故障连续使用时限不低于280 h的有3.6万台．

(3)由频率直方图得＝190×0.05＋210×0.05＋230×0.10＋250×0.15＋270×0.20＋290×0.30＋310×0.10＋330×0.05＝269(h)．

故估计这8万台电风扇的平均无故障连续使用时限为269 h.

二、百分位数的计算

例2　从某公司生产的产品中，任意抽取12件，得到它们的质量(单位：kg)如下：

7．9,9.0,8.9,8.6,8.4,8.5,8.5,8.5,9.9,7.8,8.3,8.0，

分别求出这组数据的25百分位数，75百分位数，95百分位数．

解　将所有数据从小到大排列，得7．8,7.9,8.0,8.3,8.4,8.5,8.5,8.5,8.6,8.9,9.0,9.9，

因为共有12个数据，

所以25×＝3,75×＝9,95×＝11.4，

则25百分位数是＝8.15，

75百分位数是＝8.75，

95百分位数是第12个数据为9.9.

反思感悟　计算一组*n*个数据的*p*百分位数的一般步骤：

(1)排列：按照从小到大排列原始数据；

(2)算*i*：计算*i*＝*n*×*p*%；

(3)定数：若*i*不是整数，大于*i*的最小整数为*j*，则*p*百分位数为第*j*项数据；若*i*是整数，则*p*百分位数为第*i*项与第(*i*＋1)项数据的平均数．

跟踪训练2　某歌手电视大奖赛中，七位评委对某选手打出如下分数：7.9,8.1,8.4,8.5,8.5,8.7,9.9，则其50百分位数为\_\_\_\_\_\_\_\_．

答案　8.5

解析　∵50×＝3.5，

∴其50百分位数是第4个数据为8.5.

三、百分位数的综合应用

例3　教育厅为了了解和掌握高考考生的实际答卷情况，随机地取出了100名考生的数学成绩(单位：分)，将数据分成了11组，制成了如图所示的频率分布表：

|  |  |  |
| --- | --- | --- |
| 分组 | 频数 | 频率 |
| [80,85) | 1 | 0.01 |
| [85,90) | 2 | 0.02 |
| [90,95) | 4 | 0.04 |
| [95,100) | 14 | 0.14 |
| [100,105) | 24 | 0.24 |
| [105,110) | 15 | 0.15 |
| [110,115) | 12 | 0.12 |
| [115,120) | 9 | 0.09 |
| [120,125) | 11 | 0.11 |
| [125,130) | 6 | 0.06 |
| [130,135] | 2 | 0.02 |
| 合计 | 100 | 1 |

(1)求样本数据的60,80百分位数；

(2)估计高考考生的数学成绩的90百分位数．

解　从频率分布表得，前六组的频率之和为0.01＋0.02＋0.04＋0.14＋0.24＋0.15＝0.60，

前七组的频率之和为0.60＋0.12＝0.72，

前八组的频率之和为0.72＋0.09＝0.81，

前九组的频率之和为0.81＋0.11＝0.92.

(1)由前六组的频率之和为0.60，得样本数据的60百分位数为110，样本数据的80百分位数一定在第八组[115,120)内，由115＋5×≈119.4，估计样本数据的80百分位数约为119.4.

(2)由前八组的频率之和为0.81，前九组的频率之和为0.92，知90百分位数一定在第九组[120,125)内，由120＋5×≈124.1，估计高考考生的数学成绩的90百分位数为124.1.

反思感悟　由频率直方图求百分位数的方法

(1)要注意频率直方图中小矩形的面积，就是数据落在该组的频率．

(2)一般采用方程的思想，设出*k*百分位数，根据其意义列出方程并求解即可．

跟踪训练3　为了了解一片经济林的生长情况，随机抽测了其中60株树木的底部周长(单位：cm)，所得数据均在区间[80,130]上，其频率直方图如图所示，你能估计一下60株树木的50百分位数和75百分位数吗？

解　由题意知分别落在各区间上的频数

在[80,90)上为60×0.15＝9，

在[90,100)上为60×0.25＝15，

在[100,110)上为60×0.3＝18，

在[110,120)上为60×0.2＝12，

在[120,130]上为60×0.1＝6.

从以上数据可知50百分位数一定落在区间[100,110)上，

由100＋10×＝100＋≈103.3；

75百分位数一定落在区间[110,120)上，

由110＋10×＝110＋＝112.5.

综上可知，50百分位数和75百分位数分别估计为103.3 cm，112.5 cm.

1．(多选)已知100个数据的75百分位数是9.3，则下列说法不正确的是(　　)

A．这100个数据中一定有75个数小于或等于9.3

B．把这100个数据从小到大排列后，9.3是第75个数据

C．把这100个数据从小到大排列后，9.3是第75个数据和第76个数据的平均数

D．把这100个数据从小到大排列后，9.3是第75个数据和第74个数据的平均数

答案　ABD

解析　因为75×＝75为整数，所以第75个数据和76个数据的平均数为75百分位数，是9.3，则C正确，其他选项均不对，故选ABD.

2．数据7.0,8.4,8.4,8.4,8.6,8.7,9.0,9.1的30百分位数为(　　)

A．8.4 B．8.5 C．8.6 D．8.3

答案　A

解析　因为30×＝2.4，故30百分位数是第三个数据8.4.

3．某地区对当地3 000户家庭的2020年所得年收入情况进行调查统计，年收入的频率直方图如图所示，数据(单位：千元)的分组依次为[20,40)，[40,60)，[60,80)，[80,100]，则年收入不超过6万元的家庭大约有(　　)

A．900户 B．600户

C．300户 D．150户

答案　A

解析　由频率直方图得，年收入不超过6万元的家庭的频率为(0.005＋0.010)×20＝0.3，所以年收入不超过6万元的家庭大约有0.3×3 000＝900(户)．

4．下列一组数据的25百分位数是(　　)

2．1,3.0,3.2,3.8,3.4,4.0,4.2,4.4,5.3,5.6

A．3.2 B．3.0

C．4.4 D．2.5

答案　A

解析　把该组数据按照由小到大排列，可得：2．1,3.0,3.2,3.4,3.8,4.0,4.2,4.4,5.3,5.6，

由25×＝2.5，不是整数，则第3个数据3.2是25百分位数．

5．一组样本数据的频率直方图如图所示，试估计此样本数据的50百分位数为\_\_\_\_\_\_\_\_．

答案

解析　样本数据低于10的比例为(0.08＋0.02)×4＝0.40，样本数据低于14的比例为0.40＋0.09×4＝0.76，所以此样本数据的50百分位数在[10,14)内，估计此样本数据的50百分位数为10＋×4＝.

1．知识清单：

(1)*k*百分位数．

(2)四分位数．

2．方法归纳：数据分析、数形结合．

3．常见误区：求*k*百分位数时，忽略将数据从小到大排列．

1．以下数据为参加数学竞赛决赛的15人的成绩(单位：分)：78,70,72,86,88,79,80,81,94,84,56,

98,83,90,91，则这15人成绩的80百分位数是(　　)

A．90 B．90.5 C．91 D．91.5

答案　B

解析　把成绩按从小到大的顺序排列为56,70,72,78,79,80,81,83,84,86,88,90,91,94,98，

因为80×＝12，所以这15人成绩的80百分位数是＝90.5.

2．如图是某工厂对一批新产品长度(单位：mm)检测结果的频率直方图．估计这批产品的平均数与中位数分别为(　　)

A．22.5,20 B．22.5,22.75

C．22.75,22.5 D．22.75,25

答案　C

解析　由题意，这批产品的平均数为＝5×(0.02×12.5＋0.04×17.5＋0.08×22.5＋0.03×27.5＋0.03×32.5)＝22.75，

其中位数为*x*0＝20＋＝22.5.故选C.

3．某厂10名工人在一小时内生产零件的个数分别是15,17,14,10,15,17,17,16,14,12，设该组数据的平均数为*a*,50百分位数为*b*，则有(　　)

A．*a*＝13.7，*b*＝15.5 B．*a*＝14，*b*＝15

C．*a*＝12，*b*＝15.5 D．*a*＝14.7，*b*＝15

答案　D

解析　把该组数据按从小到大的顺序排列为10,12,14,14,15,15,16,17,17,17，其平均数*a*＝×(10＋12＋14＋14＋15＋15＋16＋17＋17＋17)＝14.7，因为50×＝5，所以这10名工人一小时内生产零件的50百分位数为*b*＝＝15.

4．某棉纺厂为了了解一批棉花的质量，从中随机抽取了100根棉花纤维的长度(棉花纤维的长度是棉花质量的重要指标)，所得数据都在区间[5,40]中，其频率直方图如图所示，估计棉花纤维的长度的样本数据的80百分位数是(　　)

A．29 mm B．29.5 mm

C．30 mm D．30.5 mm

答案　A

解析　棉花纤维的长度在30 mm以下的比例为(0.01＋0.01＋0.04＋0.06＋0.05)×5＝0.85＝85%，

在25 mm以下的比例为85%－25%＝60%，

因此，80百分位数一定位于[25,30)内，

由25＋5×＝29，

可以估计棉花纤维的长度的样本数据的80百分位数是29 mm.

5．已知甲、乙两组数据(已按从小到大的顺序排列)：

甲组：27,28,39,40，*m*,50；

乙组：24，*n*,34,43,48,52.

若这两组数据的30百分位数、80百分位数分别相等，则等于(　　)

A. B.

C. D.

答案　A

解析　因为×6＝1.8，×6＝4.8，所以30百分位数为*n*＝28,80百分位数为*m*＝48，所以＝＝.

6．1,2,3,4,5,6,7,8,9,10的25百分位数为\_\_\_\_\_\_，75百分位数为\_\_\_\_\_\_\_\_，90百分位数为\_\_\_\_\_\_\_\_．

答案　3　8　9.5

解析　因为数据个数为10,10×＝2.5,10×＝7.5，10×＝9.

所以该组数据的25百分位数为*x*3＝3,75百分位数为*x*8＝8,90百分位数为＝＝9.5.

7．某学校组织学生参加数学测试，成绩的频率直方图如图，数据的分组依次为[20,40)，[40,60)，[60,80)，[80,100]，则60分为成绩的\_\_\_\_\_\_\_\_百分位数．

答案　30

解析　因为分数位于[20,40)，[40,60)的频率之和为(0.005＋0.01)×20＝0.3，所以60分为成绩的30百分位数．

8．某市要对两千多名出租车司机的年龄进行调查，现从中随机抽出100名司机，已知抽到的司机年龄都在[20,45]岁之间，根据调查结果得出司机的年龄情况残缺的频率直方图如图所示，利用这个残缺的频率直方图估计该市出租车司机年龄的中位数大约是\_\_\_\_\_\_\_\_．(保留一位小数点)

答案　33.6

解析　在频率直方图中，所有矩形面积之和为1，

所以，数据位于[25,30)的频率为1－(0.01＋0.07＋0.06＋0.02)×5＝0.2，

前两个矩形的面积之和为0.01×5＋0.2＝0.25，

前三个矩形的面积之和为0.25＋0.07×5＝0.6，

所以，中位数位于区间[30,35)，设中位数为*a*，

则有0.25＋(*a*－30)×0.07＝0.5，解得*a*≈33.6(岁).

9．求下列数据的四分位数．

13,15,12,27,22,24,28,30,31,18,19,20.

解　把12个数据按从小到大的顺序排列可得：12,13,15,18,19,20,22,24,27,28,30,31，

计算25×＝3,50×＝6,75×＝9，

所以数据的25百分位数为＝16.5，50百分位数为＝21，75百分位数为＝27.5.

10．某网络营销部门随机抽查了某市200名网友在2021年11月11日的网购金额，所得数据如下表：

|  |  |  |
| --- | --- | --- |
| 网购金额(单位：千元) | 人数 | 频率 |
| [0,1) | 16 | 0.08 |
| [1,2) | 24 | 0.12 |
| [2,3) | *x* | *p* |
| [3,4) | *y* | *q* |
| [4,5) | 16 | 0.08 |
| [5,6] | 14 | 0.07 |
| 合计 | 200 | 1.00 |

已知网购金额低于3千元与不低于3千元的人数比恰为3∶2.

(1)试确定*x*，*y*，*p*，*q*的值，并补全频率直方图；

(2)估计网购金额的25百分位数(结果保留3位有效数字)．

解　(1)根据题意有

解得

所以*p*＝0.4，*q*＝0.25.

补全频率直方图如图所示．

(2)由(1)可知，网购金额低于2千元的频率为0.08＋0.12＝0.2，

网购金额低于3千元的频率为0.2＋0.4＝0.6，

所以网购金额的25百分位数在[2,3)内，

则网购金额的25百分位数估计为2＋×1＝2．125≈2.13(千元)．

11．港珠澳大桥于2018年10月24日正式通车，它是中国境内一座连接香港、珠海和澳门的桥隧工程，桥隧全长55千米，桥面为双向六车道高速公路，大桥通行限速100 km/h. 现对大桥某路段上汽车行驶速度进行抽样调查，画出频率直方图(如图)．根据直方图估计在此路段上汽车行驶速度的众数和行驶速度超过90 km/h的频率分别为(　　)

A．85,0.25 B．90,0.35

C．87.5,0.25 D．87.5,0.35

答案　D

解析　由频率直方图估计在此路段上汽车行驶速度的众数为＝87.5，

由频率直方图估计在此路段上汽车行驶速度超过90 km/h的频率为(0.05＋0.02)×5＝0.35，

∴由频率直方图估计在此路段上汽车行驶速度超过90 km/h的频率为0.35.

12．已知30个数据的60百分位数是8.2，这30个数据从小到大排列后第18个数据是7.8，则第19个数据是\_\_\_\_\_\_\_\_．

答案　8.6

解析　由60×＝18，设第19个数据为*x*，则＝8.2，解得*x*＝8.6，即第19个数据是8.6.

13．如图是某市2021年4月1日至4月7日每天最高、最低气温的折线统计图，这7天的日最高气温的10百分位数为\_\_\_\_\_\_，日最低气温的80百分位数为\_\_\_\_\_\_．

答案　24 ℃　16 ℃

解析　由折线图可知，把日最高气温按照从小到大排序，得24,24.5,24.5,25,26,26,27，因为共有7个数据，所以7×＝0.7，不是整数，所以这7天日最高气温的10百分位数是第1个数据，为24 ℃.把日最低气温按照从小到大排序，得12,12,13,14,15,16,17，因为共有7个数据，所以7×＝5.6，不是整数，所以这7天日最低气温的80百分位数是第6个数据，为16 ℃.

14．某年级120名学生在一次百米测试中，成绩全部介于13秒与18秒之间，将测试结果分成5组：[13,14)，[14,15)，[15,16)，[16,17)，[17,18]，得到如图所示的频率直方图，如果从左到右的5个小矩形的面积之比为1∶3∶7∶6∶3，那么成绩的70百分位数约为\_\_\_\_\_\_\_\_秒．

答案　16.5

解析　设成绩的70百分位数为*x*，因为＝0.55，＝0.85，所以*x*∈[16,17)，

所以0.55＋(*x*－16)×＝0.70，解得*x*＝16.5.

15．数据3.2,3.4,3.8,4.2,4.3,4.5，*x*,6.6的65百分位数是4.5，则实数*x*的取值范围是(　　)

A．[4.5，＋∞) B．[4.5,6.6)

C．(4.5，＋∞) D．(4.5,6.6]

答案　A

解析　因为65×＝5.2，所以这组数据的65百分位数是第6个数据4.5，则*x*≥4.5，故选A.

16．某市为了鼓励居民节约用电，实行“阶梯式”电价，将该市每户居民的月用电量划分为三档，月用电量不超过200千瓦时的部分按0.5元/千瓦时收费，超过200千瓦时但不超过400千瓦时的部分按0.8元/千瓦时收费，超过400千瓦时的部分按1.0元/千瓦时收费．

(1)求某户居民用电费用*y*(单位：元)关于月用电量*x*(单位：千瓦时)的函数解析式；

(2)为了了解居民的用电情况，通过抽样获得了今年1月份100户居民每户的用电量，统计分析后得到如图所示的频率直方图．若这100户居民中，今年1月份用电费用低于260元的占80%，求*a*，*b*的值；

(3)根据(2)中求得的数据计算用电量的75百分位数．

解　(1)当0≤*x*≤200时，*y*＝0.5*x*；

当200＜*x*≤400时，

*y*＝0.5×200＋0.8×(*x*－200)＝0.8*x*－60；

当*x*＞400时，

*y*＝0.5×200＋0.8×200＋1.0×(*x*－400)＝*x*－140.

所以*y*与*x*之间的函数解析式为*y*＝

(2)由(1)可知，当*y*＝260时，*x*＝400，即用电量低于400千瓦时的占80%，

结合频率直方图可知

解得

(3)设75百分位数为*m*，

因为用电量低于300千瓦时的所占比例为(0.001＋0.002＋0.003)×100＝60%，

用电量低于400千瓦时的占80%，

所以75百分位数*m*在[300,400)内，

所以0.6＋(*m*－300)×0.002＝0.75，

解得*m*＝375，即用电量的75百分位数为375千瓦时．