

专题九 统计与成对数据的统计分析



考点38 随机抽样、用样本估计总体 (P3-30)

考点39 变量间的相关关系与统计案例 (P31-54)



考点38 随机抽样、用样本估计总体



经典3+2

1. [2023新课标 II 卷] 某学校为了解学生参加体育运动的情况，用比例分配的分层随机抽样方法作抽样调查，拟从初中部和高中部两层共抽取60名学生，已知该校初中部和高中部分别有400名和200名学生，则不同的抽样结果共有 (D)

A. $C_{400}^{45} \cdot C_{200}^{15}$ 种

B. $C_{400}^{20} \cdot C_{200}^{40}$ 种

C. $C_{400}^{30} \cdot C_{200}^{30}$ 种

D. $C_{400}^{40} \cdot C_{200}^{20}$ 种

【解析】 由题意，初中部和高中部学生人数的比值为 $\frac{400}{200} = \frac{2}{1}$ ，所以抽取的60名学生中初中部应有 $60 \times \frac{2}{3} = 40$ (人)，高中部应有 $60 \times \frac{1}{3} = 20$ (人)，所以不同的抽样结果共有 $C_{400}^{40} \cdot C_{200}^{20}$ 种，故选D.



2.[2024重庆质检]一组数据按从小到大的顺序排列为2, 4, m , 13, 16, 17, 若该组数据的中位数是极差的 $\frac{3}{5}$, 则该组数据的第40百分位数是(C)

A.4

B.45

C.5

D.9

【解析】由题意得该组数据的极差为 $17 - 2 = 15$, 故中位数是 $15 \times \frac{3}{5} = 9$.因为该组数据共6个, 所以中位数为 $\frac{m+13}{2} = 9$, 解得 $m = 5$.因为 $6 \times 40\% = 2.4$, 所以该组数据的第40百分位数为第3个数, 即该组数据的第40百分位数是5.故选C.



拓展变式

变条件[2024山东枣庄部分学校联考]一组数据从小到大排列,依次为3,5,6,7,8,9, x ,12,13,13,且该组数据的70%分位数不超过11,则 x 的取值范围是(**D**)

A.[9,12]

B.(9,11]

C.(9,10)

D.[9,10]

【解析】由题意知一共有10个数据,因为 $10 \times 70\% = 7$,所以70%分位数为 $\frac{x+12}{2}$,所以 $\begin{cases} \frac{x+12}{2} \leq 11, \\ 9 \leq x \leq 12, \end{cases}$ 解得 $9 \leq x \leq 10$.

故选D.



3.[2023山东潍坊统考]若一组样本数据 x_1, x_2, \dots, x_n 的平均数为10, 另一组样本数据 $2x_1 + 4, 2x_2 + 4, \dots, 2x_n + 4$ 的方差为8, 则两组样本数据合并为一组样本数据后的平均数和方差分别为(A)

A.17, 54

B.17, 48

C.15, 54

D.15, 48

【解析】设样本数据 x_1, x_2, \dots, x_n 的平均数为 \bar{x} , 则 $\bar{x} = 10$, 所以样本数据 $2x_1 + 4, 2x_2 + 4, \dots, 2x_n + 4$ 的平均数为 $\frac{2x_1+4+2x_2+4+\dots+2x_n+4}{n} = \frac{2(x_1+x_2+\dots+x_n)+4n}{n} = \frac{2 \times n \times \bar{x} + 4n}{n} = 2\bar{x} + 4 = 24$, 所以将两组样本数据合并为一组样本数据

后的平均数为 $\frac{1}{2n}(x_1 + x_2 + \dots + x_n + 2x_1 + 4 + 2x_2 + 4 + \dots + 2x_n + 4) = \frac{1}{2n}(10n + 24n) = 17$. 又样本数据

$2x_1 + 4, 2x_2 + 4, \dots, 2x_n + 4$ 的方差为 $\frac{1}{n}\{[2x_1 + 4 - (2\bar{x} + 4)]^2 + [2x_2 + 4 - (2\bar{x} + 4)]^2 + \dots +$

$[2x_n + 4 - (2\bar{x} + 4)]^2\} = \frac{4}{n}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2] = 8$, 所以

$\frac{1}{n}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2] = 2$, 即样本数据 x_1, x_2, \dots, x_n 的方差为2, 所以将两组样本数据合并

为一组样本数据后的方差为 $\frac{1}{2} \times [2 + (17 - 10)^2] + \frac{1}{2} \times [8 + (17 - 24)^2] = 54$, 故选A.



4.[2023四川成都八中摸底]某班统计一次数学测验成绩(单位:分)的平均数与方差,计算完毕后才发现有同学的成绩还未录入,只好重算一次.已知原平均数和原方差分别为 \bar{x}, s^2 ,新平均数和新方差分别为 \bar{x}_1, s_1^2 ,若此同学的成绩恰好为 \bar{x} ,则(C)

A. $\bar{x} = \bar{x}_1, s^2 = s_1^2$

B. $\bar{x} = \bar{x}_1, s^2 < s_1^2$

C. $\bar{x} = \bar{x}_1, s^2 > s_1^2$

D. $\bar{x} < \bar{x}_1, s^2 = s_1^2$

【解析】设这个班有 n 个同学,成绩分别是 $a_1, a_2, \dots, a_i, \dots, a_n$,第 i 个同学的成绩未录入,第一次计算时,总成绩是 $(n-1)\bar{x}$,方差 $s^2 = \frac{1}{n-1}[(a_1 - \bar{x})^2 + \dots + (a_{i-1} - \bar{x})^2 + (a_{i+1} - \bar{x})^2 + \dots + (a_n - \bar{x})^2]$,第二次计算时,新

平均数 $\bar{x}_1 = \frac{(n-1)\bar{x} + \bar{x}}{n} = \bar{x}$,新方差

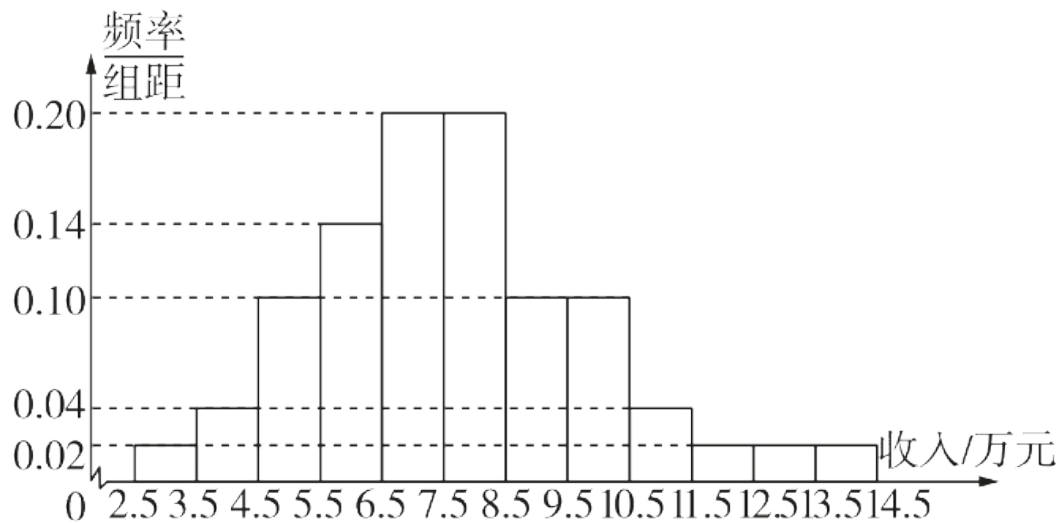
$$s_1^2 = \frac{1}{n}[(a_1 - \bar{x}_1)^2 + \dots + (a_{i-1} - \bar{x}_1)^2 + (a_{i+1} - \bar{x}_1)^2 + \dots + (a_n - \bar{x}_1)^2 + (\bar{x} - \bar{x}_1)^2] = \frac{n-1}{n}s^2, \text{故 } \bar{x} = \bar{x}_1, s^2 > s_1^2. \text{故}$$

选C.



5.[2021全国甲卷(理)]为了解某地农村经济情况,对该地农户家庭年收入进行抽样调查,将农户家庭年收入的调查数据整理得到如图所示的频率分布直方图:

根据此频率分布直方图,下面结论中不正确的是(C)



- A. 该地农户家庭年收入低于4.5万元的农户比率估计为6%
- B. 该地农户家庭年收入不低于10.5万元的农户比率估计为10%
- C. 估计该地农户家庭年收入的平均值不超过6.5万元
- D. 估计该地有一半以上的农户,其家庭年收入介于4.5万元至8.5万元之间



【解析】对于A, 根据频率分布直方图可知, 家庭年收入低于4.5万元的农户比率约为

$(0.02 + 0.04) \times 1 \times 100\% = 6\%$, 故A正确; 对于B, 根据频率分布直方图可知, 家庭年收入不低于10.5万元的农户

比率约为 $(0.04 + 0.02 + 0.02 + 0.02) \times 1 \times 100\% = 10\%$, 故B正确; 对于C, 根据频率分布直方图可知, 该地

农户家庭年收入的平均值约为

$3 \times 0.02 + 4 \times 0.04 + 5 \times 0.10 + 6 \times 0.14 + 7 \times 0.20 + 8 \times 0.20 + 9 \times 0.10 + 10 \times 0.10 + 11 \times 0.04 + 12 \times 0.02 +$

$13 \times 0.02 + 14 \times 0.02 = 7.68$

(万元), 故C错误; 对于D, 根据频率分布直方图可知, 家庭年收入介于4.5万元至8.5万元之间的农户比率约为

$(0.10 + 0.14 + 0.20 + 0.20) \times 1 \times 100\% = 64\% > 50\%$, 故D正确.



解题妙招

由于C难以计算，可以在判断选项A，B，D后，利用排除法得到**答案**为C.



6. [多选] [2024山东青岛调研] 一组样本数据11, 19, 15, 16, 19, 则这组数据的(**AB**)

A. 众数是19

B. 平均数是16

C. 中位数是15

D. 方差是44

【解析】 将这组数据从小到大排列为11, 15, 16, 19, 19, 所以众数为19, 中位数为16, 所以选项A正确, 选项C错误.

平均数 $\bar{x} = \frac{11+15+16+19+19}{5} = 16$, 方差 $s^2 = \frac{(11-16)^2+(15-16)^2+(16-16)^2+(19-16)^2+(19-16)^2}{5} = \frac{44}{5}$, 所以选项B正确, 选项

D错误. 综上, 选AB.



7. **[多选]** [2021新高考I卷]有一组样本数据 x_1, x_2, \dots, x_n , 由这组数据得到新样本数据 y_1, y_2, \dots, y_n , 其中 $y_i = x_i + c (i = 1, 2, \dots, n)$, c 为非零常数, 则(**CD**)

A. 两组样本数据的样本平均数相同

B. 两组样本数据的样本中位数相同

C. 两组样本数据的样本标准差相同

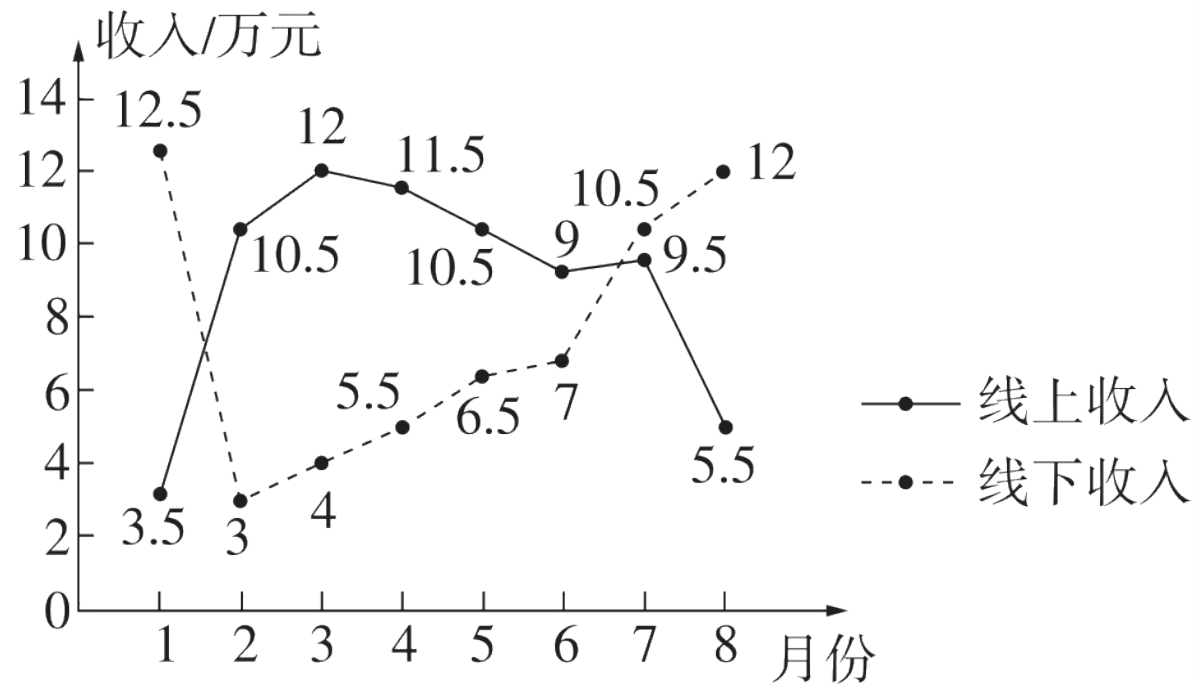
D. 两组样本数据的样本极差相同

【解析】 设样本数据 x_1, x_2, \dots, x_n 的平均数、中位数、标准差、极差分别为 \bar{x}, m, σ, t , 依题意得, 新样本数据 y_1, y_2, \dots, y_n 的平均数、中位数、标准差、极差分别为 $\bar{x} + c, m + c, \sigma, t$, 因为 $c \neq 0$, 所以C, D正确, 故选CD.



8. [多选] [2024天星原创]某商户收集并整理了其在2023年1月到8月线上和线下收入的数据，并绘制如图所示的折线图，则下列结论正确的是(**ACD**)

- A. 该商户这8个月中，月收入最高的是7月
- B. 该商户这8个月的线上总收入低于线下总收入
- C. 该商户这8个月中，线上、线下收入相差最小的是7月
- D. 该商户这8个月中，月收入不少于17万元的频率是 $\frac{1}{2}$

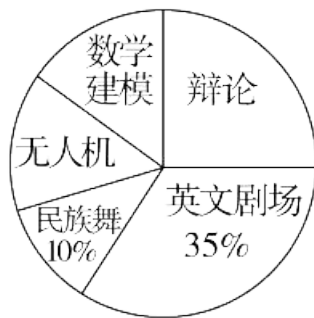
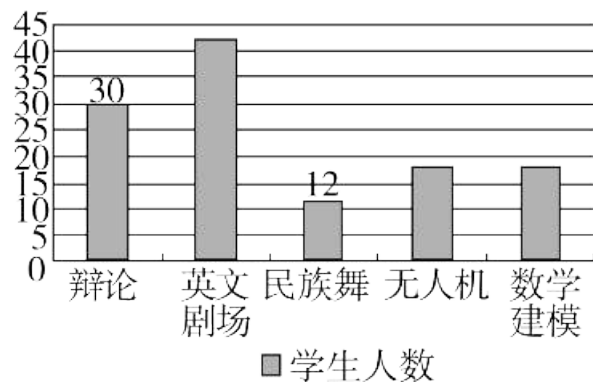




【解析】对于A: 该商户这8个月中, 月收入分别为16万元, 13.5万元, 16万元, 17万元, 17万元, 16万元, 20万元, 17.5万元, 月收入最高的是7月, A正确; (提醒: 月收入包含线上收入和线下收入) 对于B: 该商户这8个月的线上总收入为72万元, 线下总收入为61万元, B错误; 对于C: 根据折线图可看出该商户这8个月中, 线上、线下收入相差最小的是7月, C正确; 对于D: 该商户这8个月中, 月收入不少于17万元的有4个月, 故所求频率为 $\frac{1}{2}$, D正确.



9. [多选] [2023山东济南考试]某学校组建了辩论、英文剧场、民族舞、无人机和数学建模五个社团，高一学生全员参加，且每位学生只能参加一个社团.学校根据学生参加情况绘制如下统计图，已知参加无人机社团和参加数学建模社团的学生人数相等，下列说法正确的是(AC)



- A. 高一年级学生人数为120
- B. 参加无人机社团的学生人数为17
- C. 若按比例分层随机抽样从各社团抽取20人，则从无人机社团抽取的学生人数为3
- D. 若甲、乙、丙三人报名参加社团，则共有60种不同的报名方法



【解析】由题中统计图可知，参加民族舞社团的学生人数为12，占高一年级学生人数的10%，所以高一年级学生人数为 $12 \div 10\% = 120$ ，所以参加英文剧场社团的学生人数为 $120 \times 35\% = 42$ ，又参加辩论社团的学生人数为30，所以参加无人机社团的学生人数等于参加数学建模社团的学生人数等于 $(120 - 42 - 30 - 12) \div 2 = 18$ ，故A正确，B不正确.若按比例分层随机抽样从各社团抽取20人，则从无人机社团抽取的学生人数为 $20 \times \frac{18}{120} = 3$ ，C正确.若甲、乙、丙三人报名参加社团，则每人有5种选法，共有 $5^3 = 125$ （种）不同的报名方法，故D不正确.综上所述，选AC.



10. [多选] [2024安徽名校联考]某学校共有学生2 000人, 其中高一学生800人, 高二、高三学生各600人, 学校为了了解学生在暑假期间每天的读书时间(单位: 小时), 按照分层随机抽样的方法从全校学生中抽取100人, 其中高一学生、高二学生、高三学生每天读书时间的平均数分别为 $\bar{x}_1 = 2.7$, $\bar{x}_2 = 3.1$, $\bar{x}_3 = 3.3$, 每天读书时间的方差分别为 $s_1^2 = 1$, $s_2^2 = 2$, $s_3^2 = 3$, 则下列说法正确的是(**ACD**)

A. 从高一学生中抽取了40人

B. 抽取的高二学生每天的总读书时间是1 860小时

C. 被抽取的学生每天的读书时间的平均数为3小时

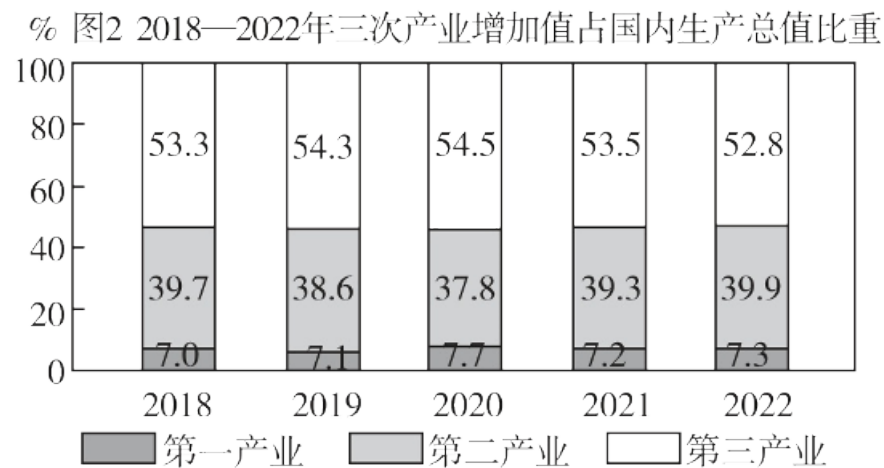
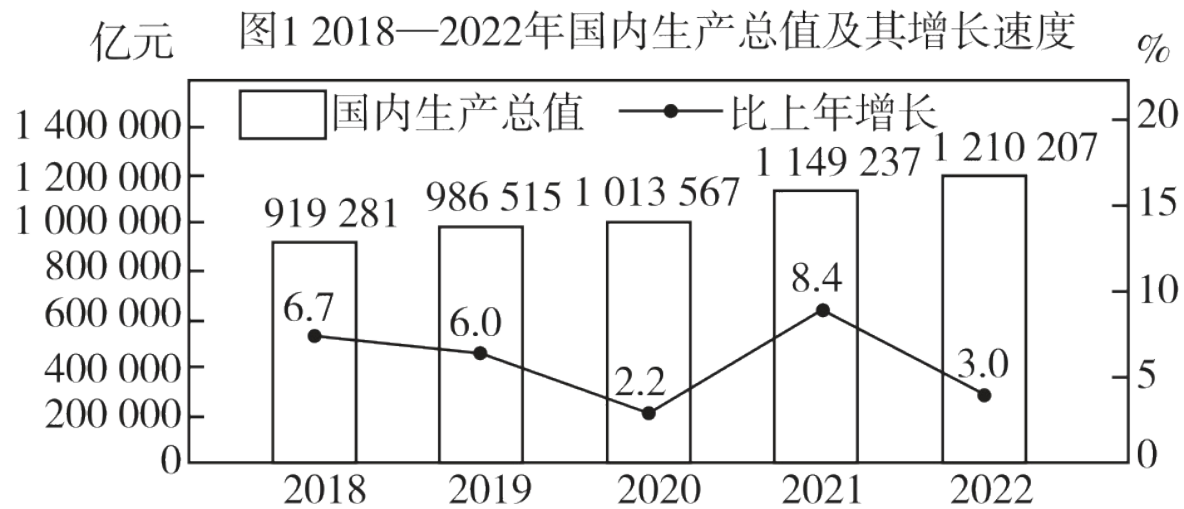
D. 估计该校全体学生每天的读书时间的方差为 $s^2 = 1.966$



【解析】对于A, 根据分层抽样, 分别从高一学生、高二学生、高三学生中抽取40人、30人、30人, 故A正确; 对于B, 抽取的高二学生每天的总读书时间是 $\bar{x}_2 \times 30 = 93$ (小时), 故B错误; 对于C, 被抽取的学生每天的读书时间的平均数为 $\frac{40}{100} \times 2.7 + \frac{30}{100} \times 3.1 + \frac{30}{100} \times 3.3 = 3$ (小时), 故C正确; 对于D, 被抽取的学生每天的读书时间的方差为 $\frac{40}{100} \times [1 + (2.7 - 3)^2] + \frac{30}{100} \times [2 + (3.1 - 3)^2] + \frac{30}{100} \times [3 + (3.3 - 3)^2] = 1.966$, 所以估计该校全体学生每天的读书时间的方差为 $s^2 = 1.966$, 故D正确. 故选ACD.



11. [多选] [2023四川宜宾诊断]国家统计局于2023年2月发布《中华人民共和国2022年国民经济和社会发展统计公报》，如图1是2018—2022年国内生产总值及其增长速度，图2是2018—2022年三次产业增加值占国内生产总值比重（三次产业包括第一产业、第二产业、第三产业）。根据图1，2，以下说法正确的是(**ABC**)



- A. 2018—2022年国内生产总值呈逐年增长的趋势
- B. 2020年与2022年国内生产总值的增长速度较上一年均有明显回落
- C. 2018—2022年第三产业增加值占国内生产总值比重的极差为1.7%
- D. 2020年第二产业增加值较2019年有所减少

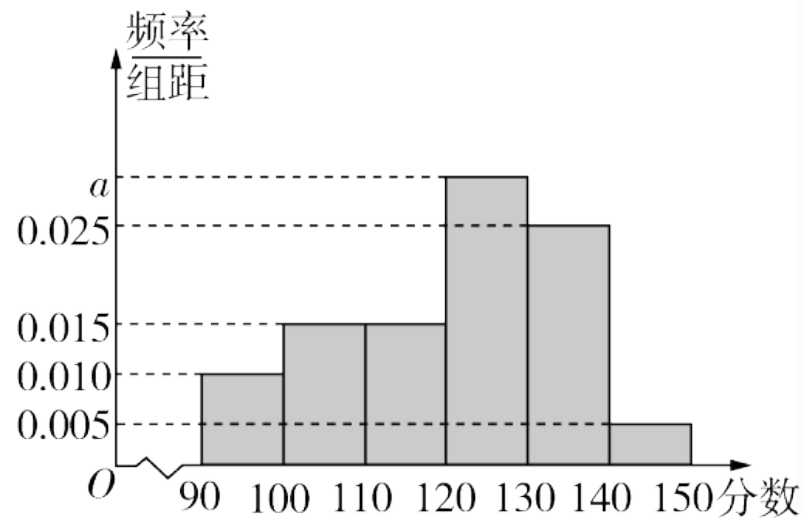


【解析】对于A：由题图1可以看出2018—2022年国内生产总值呈逐年增长的趋势，故A正确.对于B：由题图1可以看出折线在2020年与2022年时与上一年连线的斜率均小于0，故B正确.对于C：由题图2可以得出2018—2022年第三产业增加值占国内生产总值比重最大值为54.5%，最小值为52.8%，所以极差为 $54.5\% - 52.8\% = 1.7\%$ ，故C正确.对于D：结合题图1, 2可知，2019年第二产业的增加值为 $986\ 515 \times 38.6\% = 380\ 794.79$ （亿元），2020年第二产业的增加值为 $1\ 013\ 567 \times 37.8\% = 383\ 128.326$ （亿元），因为 $383\ 128.326 > 380\ 794.79$ ，所以2020年第二产业增加值较2019年有所增加，故D错误.故选ABC.



12. [多选] [2024天星原创]某校某年级为了测试该年级1 000名学生的数学水平,组织了一次数学测试,将这1 000名学生的成绩分成了六组,分别为 $[90,100)$, $[100,110)$, $[110,120)$, $[120,130)$, $[130,140)$, $[140,150]$,并整理得到如图所示的频率分布直方图,则下列结论正确的是(**ABD**)

- A.该年级学生成绩落在 $[120,130)$ 的人数为300
- B.该年级学生成绩的第一四分位数为110
- C.该年级学生成绩的众数为135分
- D.该年级学生成绩的平均分为121分





【解析】选项A:由频率分布直方图可得,该年级学生成绩落在 $[120,130)$ 的频率为

$1 - (0.01 + 0.015 + 0.015 + 0.025 + 0.005) \times 10 = 0.3$, (点拨:频率之和为1)则该年级学生成绩落在 $[120,130)$

的人数为 $1\ 000 \times 0.3 = 300$,故A正确.选项B:第一四分位数即25%分位数,由频率分布直方图可得,前两个区间的

频率之和恰为0.25,故B正确.选项C:由频率分布直方图可知,该年级学生成绩落在 $[120,130)$ 的频率最大,故该年

级学生成绩的众数为125分, (点拨:频率分布直方图中,最高小长方形的底边中点的横坐标表示众数)故C错误.

选项D:该年级学生成绩的平均分为

$95 \times 0.1 + 105 \times 0.15 + 115 \times 0.15 + 125 \times 0.3 + 135 \times 0.25 + 145 \times 0.05 = 121$ (分),故D正确.



13. [多选] [2023福建部分学校联考]某环保局对辖区内甲、乙、丙、丁四个地区的环境治理情况进行检查督导,若连续10天,每天空气质量指数(单位: $\mu\text{g}/\text{m}^3$) 不超过100,则认为该地区环境治理达标,否则认为该地区环境治理不达标.根据连续10天检查所得数据的数字特征推断,环境治理一定达标的地区是(**AD**)
- A.甲地区: 平均数为80, 方差为40
B.乙地区: 平均数为50, 众数为40
C.丙地区: 中位数为50, 极差为60
D.丁地区: 极差为10, 80%分位数为90



【解析】对于A选项, 设甲地区这10天中, 第*i*天的空气质量指数为 $x_i (i = 1, 2, \dots, 10)$, 则方差 $s^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x})^2$,

由 $\frac{1}{10} \sum_{i=1}^{10} (x_i - 80)^2 = 40$, 得 $\sum_{i=1}^{10} (x_i - 80)^2 = 400$, 如果这10天中有1天的空气质量指数超过100, 则必有

$\sum_{i=1}^{10} (x_i - 80)^2 > 400$, 与题意相矛盾, 所以甲地区这10天每天的空气质量指数都不超过100, 甲地区环境治理一定

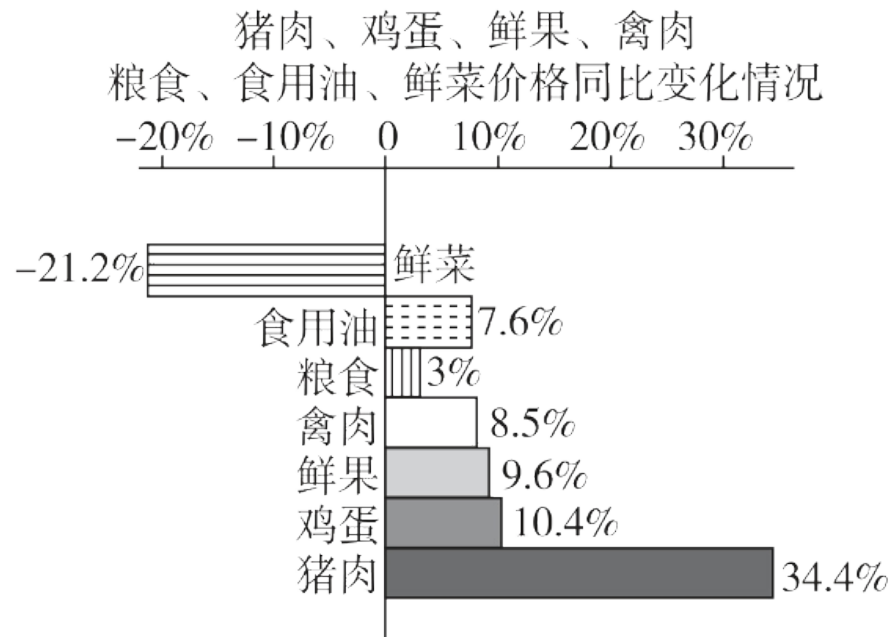
达标. 对于B选项, 若乙地区这10天的空气质量指数有8天为40, 1天为150, 1天为30, 满足平均数为50, 众数为40, 但该地区环境治理不达标, 所以乙地区环境治理不一定达标. 对于C选项, 若丙地区这10天的空气质量指数在第1天为110, 后面9天均为50, 满足中位数为50, 极差为60, 但该地区环境治理不达标, 所以丙地区环境治理不一定达标. 对于D选项, 如果丁地区这10天的空气质量指数的最大值超过100, 根据极差为10, 则最小值超过90, 这与80%分位数为90矛盾, 故最大值不超过100, 所以丁地区环境治理一定达标. 故选AD.



创新1+1

1. **新角度·创新统计图形式**[2024四川雅安部分学校联考]2023年4月, 国内猪肉、鸡蛋、鲜果、禽肉、粮食、食用油、鲜菜价格同比(与去年同期相比)的变化情况如图所示, 则下列说法正确的是(**D**)

- A. 猪肉、鸡蛋、鲜果、禽肉、粮食、食用油这6类食品中, 食用油价格同比涨幅最小
- B. 猪肉价格同比涨幅超过禽肉价格同比涨幅的5倍
- C. 去年4月鲜菜价格要比今年4月低
- D. 这7类食品价格同比涨幅的平均值超过7%





【解析】由图可知，粮食价格同比涨幅比食用油价格同比涨幅小，故A不正确；猪肉价格同比涨幅为34.4%，禽肉价格同比涨幅为8.5%， $34.4\% - 5 \times 8.5\% < 0$ ，故B不正确；鲜菜价格同比涨幅为-21.2%，说明去年4月鲜菜价格要比今年4月高，故C不正确；这7类食品价格同比涨幅的平均值为

$\frac{34.4\% + 10.4\% + 9.6\% + 8.5\% + 3\% + 7.6\% - 21.2\%}{7} \approx 7.47\% > 7\%$ ，故D正确。故选D。

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：
<https://d.book118.com/815334024022011311>