

心理统计学第二讲

讲师 司马紫衣



高途学院APP下载



高途学院公众号





第四章 差异量数

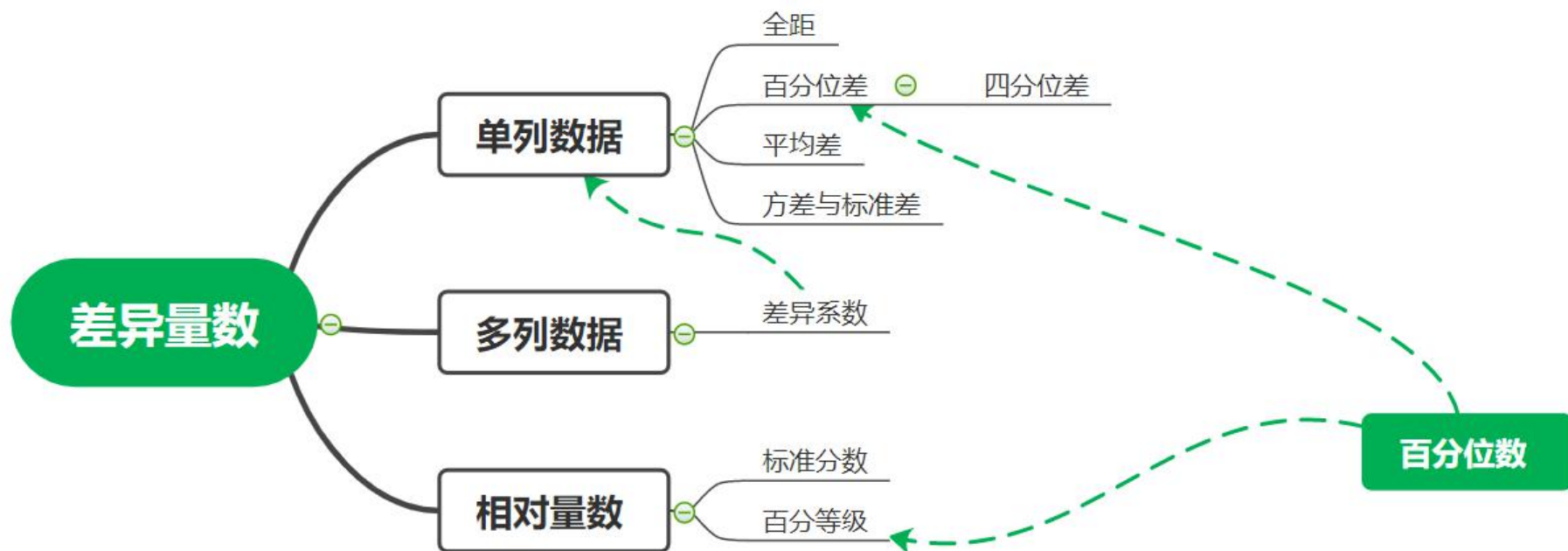


概览

- ◆ 全距与百分位差（单选）
- ◆ 平均差、方差和标准差（单选、简答）
- ◆ 差异系数（单选、名词解释）
- ◆ 标准分数（单选、简答）



本章结构





一、全距与百分位差

(一) 全距

$$R = X_{\max} - X_{\min}$$

特点：

不稳定、不可靠、不灵敏；

易受极端值的影响。



一、全距与百分位差

(二) 百分位差

1. 百分等级

百分等级是低于某测验分数的人数百分比。

计算：

(1) 未分组： $P_R = 100 - (100 \times R - 50) / N$ R为某人在测验中的排序。

(2) 已分组：
$$P_R = \frac{100}{N} \times [F_b + \frac{f(X - L_b)}{i}]$$

N是总数 i是组距

Lb是该分数所在组的精确下限

Fb是小于Lb的各组次数的和

f是百分位数所在组的次数



一、全距与百分位差

(二) 百分位差

2. 百分位数

百分位数是在数列中处于某一百分等级的原始数据。

$P_{50} = 70$ ，那么在70分以下就包括50%的数据。

计算：

(1) 未分组：先排序，找出第 $N \cdot P_R$ 个数据

(2) 已分组：
$$P_p = L_b + \frac{i}{f} \times \left(\frac{P}{100} \times N - F_b \right)$$



一、全距与百分位差

(二) 百分位差

3. 百分位差

百分位差是指某一百分位数与另一百分位数之间的差值。

例如： $P90 - P10$

优点：百分位差容易理解、易计算且较少受两极数值的影响。

缺点：不能反映出分布的中间数值的差异情况、稳定性较差。



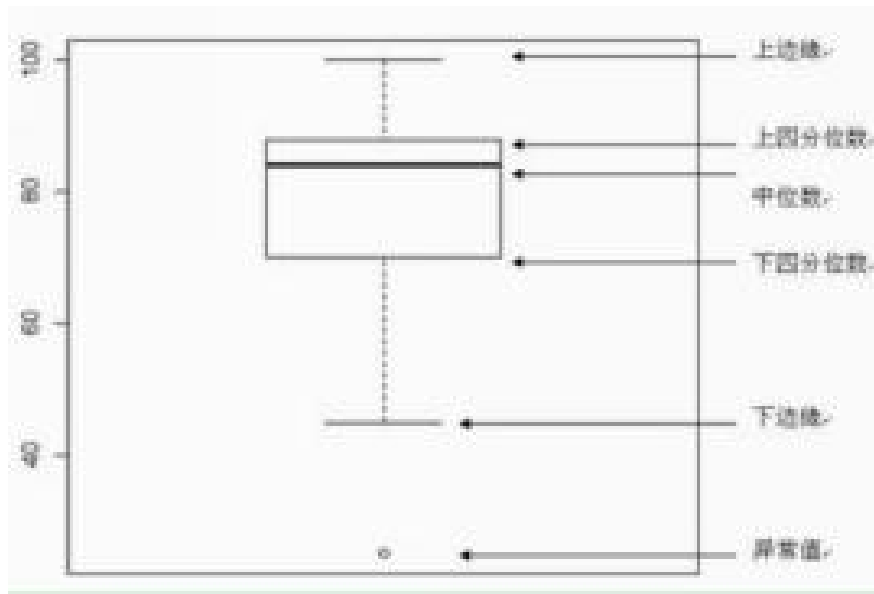
一、全距与百分位差

(二) 百分位差

4. 四分位差

它的值等于P25到P75距离的二分之一。

$$Q = \frac{Q_3 - Q_1}{2} = \frac{P_{75} - P_{25}}{2}$$





真题演练

现有一列数据，它们是4, 4, 5, 3, 5, 5, 2。这列数据的平均数、众数和全距依次是（ ）。(312, 2007)

- A. 4, 4, 2
- B. 4, 5, 3
- C. 5, 4, 4
- D. 5, 5, 1

甲班和乙班同时参加数学考试，甲班的 $Q3-Q1=38$ ，乙班的 $Q3-Q1=24$ ，则两个班分数的离散程度（ ）。(北师大, 2015)

- A. 甲 $>$ 乙
- B. 甲 $=$ 乙
- C. 甲 $<$ 乙
- D. 无法比较



真题演练

现有一列数据，它们是4, 4, 5, 3, 5, 5, 2。这列数据的平均数、众数和全距依次是（ ）。 (312, 2007)

- A. 4, 4, 2
- B. 4, 5, 3
- C. 5, 4, 4
- D. 5, 5, 1

答案：B

甲班和乙班同时参加数学考试，甲班的 $Q3-Q1=38$ ，乙班的 $Q3-Q1=24$ ，则两个班分数的离散程度（ ）。 (北师大, 2015)

- A. 甲 $>$ 乙
- B. 甲 $=$ 乙
- C. 甲 $<$ 乙
- D. 无法比较



真题演练

现有一列数据，它们是4, 4, 5, 3, 5, 5, 2。这列数据的平均数、众数和全距依次是（ ）。(312, 2007)

- A. 4, 4, 2
- B. 4, 5, 3
- C. 5, 4, 4
- D. 5, 5, 1

答案：B

甲班和乙班同时参加数学考试，甲班的 $Q3-Q1=38$ ，乙班的 $Q3-Q1=24$ ，则两个班分数的离散程度（ ）。(北师大, 2015)

- A. 甲 $>$ 乙
- B. 甲 $=$ 乙
- C. 甲 $<$ 乙
- D. 无法比较

答案：A



真题演练

下列关于百分等级的说法正确的是（ ）。 （多选） （西南，2018）

- A.百分等级是一种相对位置量数
- B.百分等级单位相等
- C.百分等级是相对于特定的被试团体而言的
- D.百分等级只具有顺序性，无法用它来说明不同被试之间分数差异的数量，也无法对同一被试在多项测验上的百分等级进行合成汇总。



真题演练

下列关于百分等级的说法正确的是（ ）。 （多选） （西南，2018）

A.百分等级是一种相对位置量数

B.百分等级单位相等

C.百分等级是相对于特定的被试团体而言的

D.百分等级只具有顺序性，无法用它来说明不同被试之间分数差异的数量，也无法对同一被试在多项测验上的百分等级进行合成汇总。

答案：ACD



真题演练

百分位分数 $P_{45}=65$ 表示（ ）。(312, 2018)

- A. 低于45分的人数占总人数的65%
- B. 高于45分的人数占总人数的65%
- C. 高于65分的人数占总人数的45%
- D. 低于65分的人数占总人数的45%



真题演练

百分位分数 $P_{45}=65$ 表示 ()。(312, 2018)

- A. 低于45分的人数占总人数的65%
- B. 高于45分的人数占总人数的65%
- C. 高于65分的人数占总人数的45%
- D. 低于65分的人数占总人数的45%

答案: D



真题演练

高途考研 | 学习成就美好

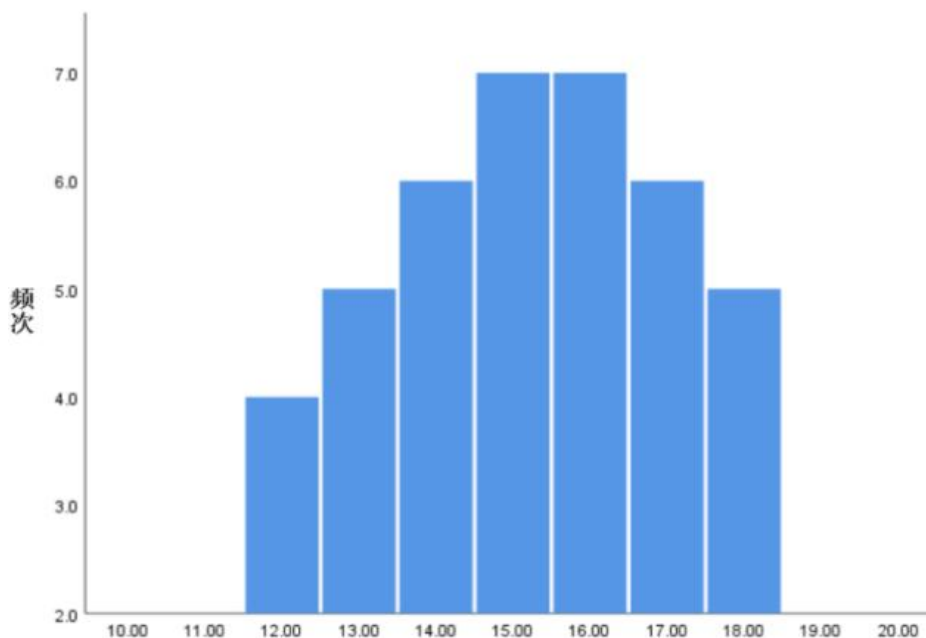
一组数据的大致分布情况如下图，若一个人的分数是14.5，那么其百分等级大概是（ ）。 （北航，2018）

A. 38

B. 16

C. 50

D. 34





真题演练

一组数据的大致分布情况如下图，若一个人的分数是14.5，那么其百分等级大概是（ ）。（北航，2018）

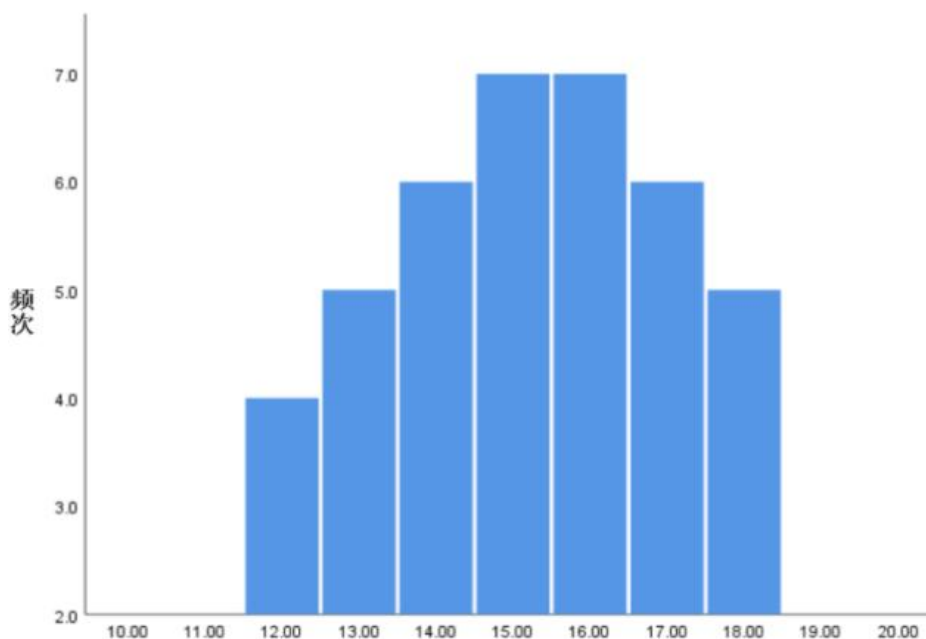
A. 38

B. 16

C. 50

D. 34

答案：A





二、平均差、方差和标准差

(一) 平均差

$$A.D. = \frac{\sum |X_i - \bar{X}|}{n}$$



二、平均差、方差和标准差

(二) 方差

$$s^2 = \frac{\sum x^2}{N} \quad x = X - \bar{X} \quad s^2 = \frac{\sum X^2}{N} - \left(\frac{\sum X}{N}\right)^2$$

特性:

可分解性

可加性

$$S_T^2 = \frac{\sum N_i S_i^2 + \sum N_i d_i^2}{\sum N_i} \quad d_i = \bar{X}_T - \bar{X}_i$$

S_T^2 是总方差

S_i^2 是各小组标准差

N_i 是各小组数据个数

\bar{X}_T 、 \bar{X}_i 分别是总平均数和各小组平均数



二、平均差、方差和标准差

(三) 标准差

$$s = \sqrt{s^2}$$

特性：

$$Y_i = X_i + C \quad s_Y = s_X$$

$$Y_i = X_i \times C \quad s_Y = |C| \times s_X$$

$$Y_i = X_i \times C + d \quad s_Y = |C| \times s_X$$

优点：反应灵敏、计算严密、计算简单、简明易解、适合于进一步用代数方法演算、较少受抽样变动影响；

缺点：但运算繁琐、易受极值影响、难理解。

切比雪夫定理，“正负三个标准差法则”。



习题演练

一组数据中每个数的值都是5，那么这组数据的标准差和方差分别是（ ）。 (312, 2009)

- | | |
|---------|----------|
| A. 0, 0 | B. 5, 25 |
| C. 0, 5 | D. 0, 25 |

有一组数据：2, 3, 4, 5, 6, 7, 8。该组数据的平均值和标准差分别是5和2。如果给这组数据的每个数都加上3，再乘以2，那么可以得到一组新数据。其平均数和标准差分别是（ ）。 (312, 2013)

- | | |
|----------|-----------|
| A. 8, 2 | B. 8, 5 |
| C. 16, 4 | D. 16, 10 |



习题演练

一组数据中每个数的值都是5，那么这组数据的标准差和方差分别是（ ）。 (312, 2009)

- | | |
|---------|----------|
| A. 0, 0 | B. 5, 25 |
| C. 0, 5 | D. 0, 25 |

答案：A

有一组数据：2, 3, 4, 5, 6, 7, 8。该组数据的平均值和标准差分别是5和2。如果给这组数据的每个数都加上3，再乘以2，那么可以得到一组新数据。其平均数和标准差分别是（ ）。 (312, 2013)

- | | |
|----------|-----------|
| A. 8, 2 | B. 8, 5 |
| C. 16, 4 | D. 16, 10 |



习题演练

一组数据中每个数的值都是5，那么这组数据的标准差和方差分别是（ ）。 (312, 2009)

- | | |
|---------|----------|
| A. 0, 0 | B. 5, 25 |
| C. 0, 5 | D. 0, 25 |

答案：A

有一组数据：2, 3, 4, 5, 6, 7, 8。该组数据的平均值和标准差分别是5和2。如果给这组数据的每个数都加上3，再乘以2，那么可以得到一组新数据。其平均数和标准差分别是（ ）。 (312, 2013)

- | | |
|----------|-----------|
| A. 8, 2 | B. 8, 5 |
| C. 16, 4 | D. 16, 10 |

答案：C



习题演练

下列几组数据中，标准差最大的一组是（ ）。(312, 2013)

A. -2, -1, 0, 1, 1, 2

B. 2, 6, 10, 14, 18

C. 5.756, 5.765, 5.890, 5.895, 5.923

D. 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00

描述数据离中趋势的统计量有（ ）。(多选) (312, 2007)

A. 方差

B. 标准差

C. 平均差

D. 四分位差



习题演练

下列几组数据中，标准差最大的一组是（ ）。(312, 2013)

A. -2, -1, 0, 1, 1, 2

B. 2, 6, 10, 14, 18

C. 5.756, 5.765, 5.890, 5.895, 5.923

D. 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00

答案：B

描述数据离中趋势的统计量有（ ）。(多选) (312, 2007)

A. 方差

B. 标准差

C. 平均差

D. 四分位差



习题演练

下列几组数据中，标准差最大的一组是（ ）。 (312, 2013)

A. -2, -1, 0, 1, 1, 2

B. 2, 6, 10, 14, 18

C. 5.756, 5.765, 5.890, 5.895, 5.923

D. 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00

答案：B

描述数据离中趋势的统计量有（ ）。 (多选) (312, 2007)

A. 方差

B. 标准差

C. 平均差

D. 四分位差

答案：ABCD



习题演练

下表是一次大规模考试中一道单项选择题的数据统计表。表中样本总人数为2320人，分成五组，每组464人。该选择题满分2分，正确选项为D。

选项	组一	组二	组三	组四	组五	选择比率
A	125	80	51	39	37	0.143
B	140	133	95	98	90	0.235
C	59	55	75	57	36	0.122
D	140	196	243	270	311	0.500

该题目得分的标准差为（ ）。 (312, 2016)

- A. 0.3 B. 0.5 C. 0.8 D. 1.0



习题演练

下表是一次大规模考试中一道单项选择题的数据统计表。表中样本总人数为2320人，分成五组，每组464人。该选择题满分2分，正确选项为D。

选项	组一	组二	组三	组四	组五	选择比率
A	125	80	51	39	37	0.143
B	140	133	95	98	90	0.235
C	59	55	75	57	36	0.122
D	140	196	243	270	311	0.500

该题目得分的标准差为（ ）。 (312, 2016)

- A. 0.3 B. 0.5 C. 0.8 D. 1.0

答案：D



三、差异系数

又称变异系数、相对标准差等，它是一种相对量数，为标准差对平均数的百分比。

$$CV = \frac{s}{\bar{X}} \times 100\%$$

适用条件：

- ① 同一团体不同观测值离散程度的比较；
- ② 对于水平相差较大，但进行的是同一种观测的各种团体的观测值的离散程度的比较。



习题演练

有三种年龄的分布：①0 岁、1 岁、2 岁；②0 个月、12 个月、24 个月；③0 天、365 天、730 天，关于这三组数据的离散程度，下面说法正确的是（ ）。(北师大, 2015)

- A. 第三组数据的离散程度最大
- B. 第三组数据的离散程度最小
- C. 三组数据的离散程度一样大
- D. 第一组数据的离散程度最小



习题演练

有三种年龄的分布：①0 岁、1 岁、2 岁；②0 个月、12 个月、24 个月；③0 天、365 天、730 天，关于这三组数据的离散程度，下面说法正确的是（ ）。 （北师大，2015）

- A. 第三组数据的离散程度最大
- B. 第三组数据的离散程度最小
- C. 三组数据的离散程度一样大
- D. 第一组数据的离散程度最小

答案：C



习题演练

某中学初一、初二的学生接受同一个测验。初一学生平均分为65分，标准差为5；初二学生平均分为80分，标准差为6。结论正确的是（ ）。

(312, 2011)

- A. 初一分数比初二分数离散程度大
- B. 初二分数比初一分数离散程度大
- C. 两个年级的分数离散程度无法比较
- D. 两个年级的分数离散程度一样大



习题演练

某中学初一、初二的学生接受同一个测验。初一学生平均分为65分，标准差为5；初二学生平均分为80分，标准差为6。结论正确的是（ ）。

(312, 2011)

- A. 初一分数比初二分数离散程度大
- B. 初二分数比初一分数离散程度大
- C. 两个年级的分数离散程度无法比较
- D. 两个年级的分数离散程度一样大

答案：A



四、标准分数

以标准差为单位表示一个原始分数在团体中所处位置的相对位置量数，也叫Z分数或基分数。

$$Z = \frac{X - \bar{X}}{s}$$

特性：

- (1) 无实际单位，以平均数为参照点，以标准差为单位。
- (2) 所有原始分数的Z分数之和为0，Z分数的平均数也为0。
所有原始分数的Z分数的标准差为1。
- (3) 原始分数转换为Z分数后，两者分布形状相同；若原始分数呈正态分布，则转换得到的所有Z分数均值为0，标准差为1的标准正态分布。



四、标准分数

优点：

可比性；可加性；明确性；稳定性。

缺点：

计算繁杂；有负值和零、有小数；在进行比较时须满足数据原始形态相同这一条件。

应用

- (1) 比较性质不同的观测值在各自数据分布中相对位置的高低。
- (2) 计算不同质的观测值的总和或平均值，以表示在团体中的相对位置。
- (3) 若标准分数中有小数、负数等不易被人接受的问题，可通过的线性公式将其转化成新的标准分数。



习题演练

一组服从正态分布的数据，平均数为50，标准差为5，则Z值为-2.58的原始数据是（ ）。 (312, 2008)

- A. 37.10
- B. 42.42
- C. 47.42
- D. 62.90

已知三年级学生历年的语文期末成绩平均分为80，标准差为5，数学期末成绩平均分为75，标准差为8，某学生的语文和数学期末成绩为82，请问他的哪科成绩相对更好一些（ ）。 (华中师大, 2016)

- A. 语文成绩
- B. 数学成绩
- C. 一样好
- D. 不确定



习题演练

一组服从正态分布的数据，平均数为50，标准差为5，则Z值为-2.58的原始数据是（ ）。(312, 2008)

- A. 37.10
- B. 42.42
- C. 47.42
- D. 62.90

答案：A

已知三年级学生历年的语文期末成绩平均分为80，标准差为5，数学期末成绩平均分为75，标准差为8，某学生的语文和数学期末成绩为82，请问他的哪科成绩相对更好一些（ ）。(华中师大，2016)

- A.语文成绩
- B.数学成绩
- C.一样好
- D.不确定



习题演练

一组服从正态分布的数据，平均数为50，标准差为5，则Z值为-2.58的原始数据是（ ）。(312, 2008)

- A. 37.10
- B. 42.42
- C. 47.42
- D. 62.90

答案：A

已知三年级学生历年的语文期末成绩平均分为80，标准差为5，数学期末成绩平均分为75，标准差为8，某学生的语文和数学期末成绩为82，请问他的哪科成绩相对更好一些（ ）。(华中师大，2016)

- A.语文成绩
- B.数学成绩
- C.一样好
- D.不确定

答案：B



第五章 相关量数

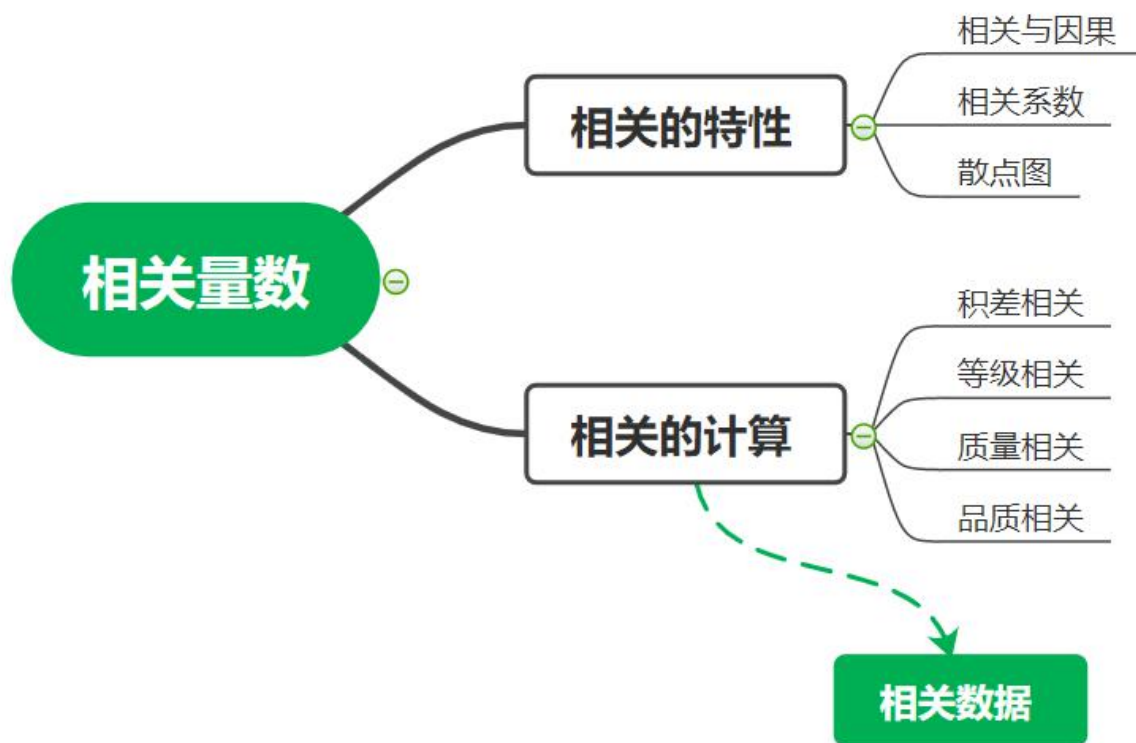


概览

- ◆ 相关的特性（选择、简答）
- ◆ 积差相关（选择、名词解释、简答）
- ◆ 等级相关（单选）
- ◆ 质量相关（单选、简答）
- ◆ 品质相关（单选）



本章结构





一、相关的特性

(一) 相关的概念

相关研究的是两变量之间的关系。变量之间的基本关系有三种类型：

1. 虚无关系：两变量相互之间无影响，无法从一个变量推导或者预测另一个变量。
2. 相关关系：一个变量变化时，另一个变量也在变化，或者一个变量的变化可以预测另一个变量的变化。
3. 因果关系：一个变量的变化引起或导致了另一个变量的变化。

相关关系和因果关系合称共变关系。

相关包含因果，相关不一定是因果。



一、相关的特性

根据因果哲学家穆勒的观点，因果关系需满足三个条件：

1. 共变关系
2. 时间顺序性
3. 排除其他可能的解释



一、相关的特性

(二) 相关系数

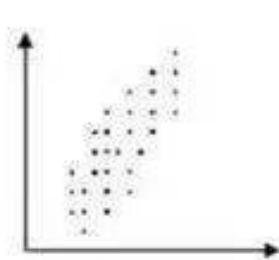
1. 取值范围是 $[-1, 1]$ 。
2. \pm 表示相关方向，绝对值表示相关的强弱程度。
3. r 值受到样本量和样本代表性的影响。
4. r 值不能进行加减乘除运算。
5. 相关是否显著，需要做检验。



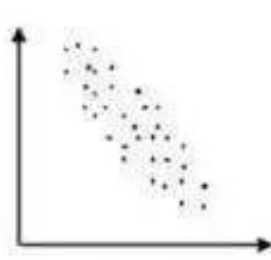
一、相关的特性

(三) 相关分类与散点图

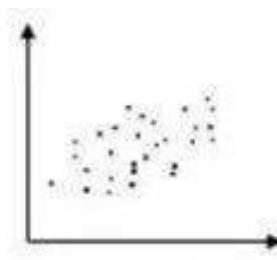
1. 正相关：两个变量变化的方向相同， $r > 0$
2. 负相关：两个变量变化的方向相反， $r < 0$
3. 零相关：两个变量的变化不存在关联， $r = 0$



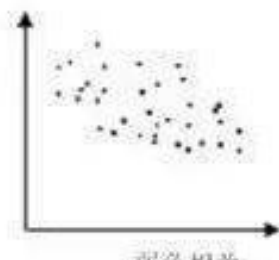
强正相关



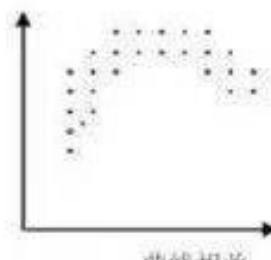
强负相关



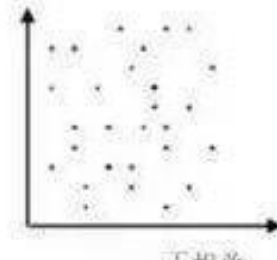
弱正相关



弱负相关



曲线相关



无相关



习题演练

如下哪个相关系数代表的相关关系最强 ()。(北航, 2016)

- A. -0.10 B. 0.46 C. 0.65 D. -0.90

散点图的形状为一条直线, 且两个变量方差均不为0, 它们之间的相关系数可能为 ()。(多选) (312, 2010)

- A. 1 B. 0.5 C. 0 D. -1



习题演练

如下哪个相关系数代表的相关关系最强 ()。(北航, 2016)

- A. -0.10 B. 0.46 C. 0.65 D. -0.90

答案: D

散点图的形状为一条直线, 且两个变量方差均不为0, 它们之间的相关系数可能为 ()。(多选) (312, 2010)

- A. 1 B. 0.5 C. 0 D. -1



习题演练

如下哪个相关系数代表的相关关系最强 ()。(北航, 2016)

- A. -0.10 B. 0.46 C. 0.65 D. -0.90

答案: D

散点图的形状为一条直线, 且两个变量方差均不为0, 它们之间的相关系数可能为 ()。(多选) (312, 2010)

- A. 1 B. 0.5 C. 0 D. -1

答案: AD



习题演练

一位研究者随机调查了50名城市居民为孩子购买课外读物的花费，另外还搜集了老师对这些孩子的总体评价，得到积差相关系数为0.53，下列推断中，正确的是（ ）。(312, 2010)

- A. 如果另外再随机调查50名乡镇居民，他们为孩子购买课外读物的花费与老师对其孩子总体评价之间的相关系数会非常接近0.53
- B. 用城市居民为孩子购买课外读物的花费预测老师对其孩子总体评价的准确率为53%
- C. 城市居民为孩子购买课外读物的花费决定老师对其子女的总体评价
- D. 城市居民为孩子购买课外读物的花费与老师对其孩子的总体评价之间存在中等程度的相关



习题演练

一位研究者随机调查了50名城市居民为孩子购买课外读物的花费，另外还搜集了老师对这些孩子的总体评价，得到积差相关系数为0.53，下列推断中，正确的是（ ）。(312, 2010)

- A. 如果另外再随机调查50名乡镇居民，他们为孩子购买课外读物的花费与老师对其孩子总体评价之间的相关系数会非常接近0.53
- B. 用城市居民为孩子购买课外读物的花费预测老师对其孩子总体评价的准确率为53%
- C. 城市居民为孩子购买课外读物的花费决定老师对其子女的总体评价
- D. 城市居民为孩子购买课外读物的花费与老师对其孩子的总体评价之间存在中等程度的相关

答案：D



二、积差相关

两个相关的变量是等距或者等比数据。

两列变量各自总体的分布都是正态的，至少接近正态。

计算线性相关。

$$r = \frac{\sum xy}{Ns_x s_y} = \frac{\sum xy}{\sqrt{\sum x^2 \times \sum y^2}}$$

$$r = \frac{1}{N} \sum Z_x Z_y$$



习题演练

当积差相关系数 $r=0$ 时，对两个变量之间关系的最佳解释是（ ）

(312, 2013)

- A. 相关程度很低
- B. 不存在任何相关
- C. 不存在线性相关关系
- D. 存在非线性相关关系

现有X、Y两列正态变量，其中 $S_x=4$, $S_y=3$, $\sum xy=24$, $n=10$ 。此两列变量的积差相关系数是（ ）

- A. 0.20
- B. 0.50
- C. 0.60
- D. 0.80



习题演练

当积差相关系数 $r=0$ 时，对两个变量之间关系的最佳解释是（ ）。

(312, 2013)

- A. 相关程度很低
- B. 不存在任何相关
- C. 不存在线性相关关系
- D. 存在非线性相关关系

答案：C

现有X、Y两列正态变量，其中 $S_x=4$, $S_y=3$, $\sum xy=24$, $n=10$ 。此两列变量的积差相关系数是（ ）。(312, 2019)

- A. 0.20
- B. 0.50
- C. 0.60
- D. 0.80



习题演练

当积差相关系数 $r=0$ 时，对两个变量之间关系的最佳解释是（ ）。

(312, 2013)

- A. 相关程度很低
- B. 不存在任何相关
- C. 不存在线性相关关系
- D. 存在非线性相关关系

答案：C

现有X、Y两列正态变量，其中 $S_x=4$, $S_y=3$, $\sum xy=24$, $n=10$ 。此两列变量的积差相关系数是（ ）。(312, 2019)

- A. 0.20
- B. 0.50
- C. 0.60
- D. 0.80

答案：A



习题演练

有300名考生，期中和期末考试成绩的相关系数为0.62，两次考试成绩平均分数均为80分。如果给所有考生的期中成绩各加8分，加分后两次成绩的相关系数是（ ）。 (312, 2016)

A. 0.52

B. 0.62

C. 0.70

D. 0.72



习题演练

有300名考生，期中和期末考试成绩的相关系数为0.62，两次考试成绩平均分数均为80分。如果给所有考生的期中成绩各加8分，加分后两次成绩的相关系数是（ ）。 (312, 2016)

A. 0.52

B. 0.62

C. 0.70

D. 0.72

答案：B



三、等级相关

使用的数据是顺序数据。

对总体分布不做要求。

1. 斯皮尔曼等级相关

$$r_R = 1 - \frac{6 \sum D^2}{N^3 - N} \quad D = R_X - R_Y$$

2. 肯德尔W系数：评价者直接使用等级评定

$$W = \frac{s}{\frac{K^2}{12}(N^3 - N)} \quad s = \sum R_i^2 - \frac{(\sum R_i)^2}{N}$$

3. 肯德尔U系数：评价者使用对偶比较法进行评定



四、质量相关

1. 点二列相关

一列为真正二分变量，另一列为等距或等比数据。

$$r_{pb} = \frac{\overline{X}_p - \overline{X}_q}{s_t} \cdot \sqrt{pq}$$

2. 二列相关

一列为人为二分变量，另一列为等距或等比数据。

$$r_b = \frac{\overline{X}_p - \overline{X}_q}{s_t} \cdot \frac{pq}{y} \qquad r_b = \frac{\overline{X}_p - \overline{X}_t}{s_t} \cdot \frac{p}{y}$$



五、品质相关

	A	非A
B	a	b
非B	c	d

1. 四分相关

两个变量都是人为二分变量。

$$r_{\Phi} = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \cos \left(\frac{180^\circ}{1 + \sqrt{\frac{ad}{bc}}} \right)$$

2. Φ 系数



习题演练

在人事测评中，计算面试官的评分一致性，采用（ ）。(北师大，2018)

- A. 皮尔逊的积差相关
- B. 斯皮尔曼的等级相关
- C. 肯德尔的和谐系数
- D. 费希尔的精确概率



习题演练

在人事测评中，计算面试官的评分一致性，采用（ ）。(北师大，2018)

- A. 皮尔逊的积差相关
- B. 斯皮尔曼的等级相关
- C. 肯德尔的和谐系数
- D. 费希尔的精确概率

答案：C



习题演练

一个测验由0、1计分的100道单选题构成，根据原始得分计算题目区分度的恰当方法是（ ）。(312, 2017)

- A. φ 相关
- B. 点二列相关
- C. 等级相关
- D. 四分相关

一项研究调查了不同性别的成年人对在公众场合吸烟的态度，结果如表所示。那么，性别与对待吸烟的态度之间的相关系数是（ ）。(312, 2008)

	赞同	反对
男	15	10
女	10	26

- A. 0.12
- B. 0.32
- C. 0.48
- D. 0.54



习题演练

一个测验由0、1计分的100道单选题构成，根据原始得分计算题目区分度的恰当方法是（ ）。(312, 2017)

- A. φ 相关
- B. 点二列相关
- C. 等级相关
- D. 四分相关

答案：B

一项研究调查了不同性别的成年人对在公众场合吸烟的态度，结果如表所示。那么，性别与对待吸烟的态度之间的相关系数是（ ）。(312, 2008)

	赞同	反对
男	15	10
女	10	26

- A. 0.12
- B. 0.32
- C. 0.48
- D. 0.54



习题演练

一个测验由0、1计分的100道单选题构成，根据原始得分计算题目区分度的恰当方法是（ ）。(312, 2017)

- A. φ 相关
- B. 点二列相关
- C. 等级相关
- D. 四分相关

答案：B

一项研究调查了不同性别的成年人对在公众场合吸烟的态度，结果如表所示。那么，性别与对待吸烟的态度之间的相关系数是（ ）。(312, 2008)

	赞同	反对
男	15	10
女	10	26

- A. 0.12
- B. 0.32
- C. 0.48
- D. 0.54

答案：B



第六章 概率分布

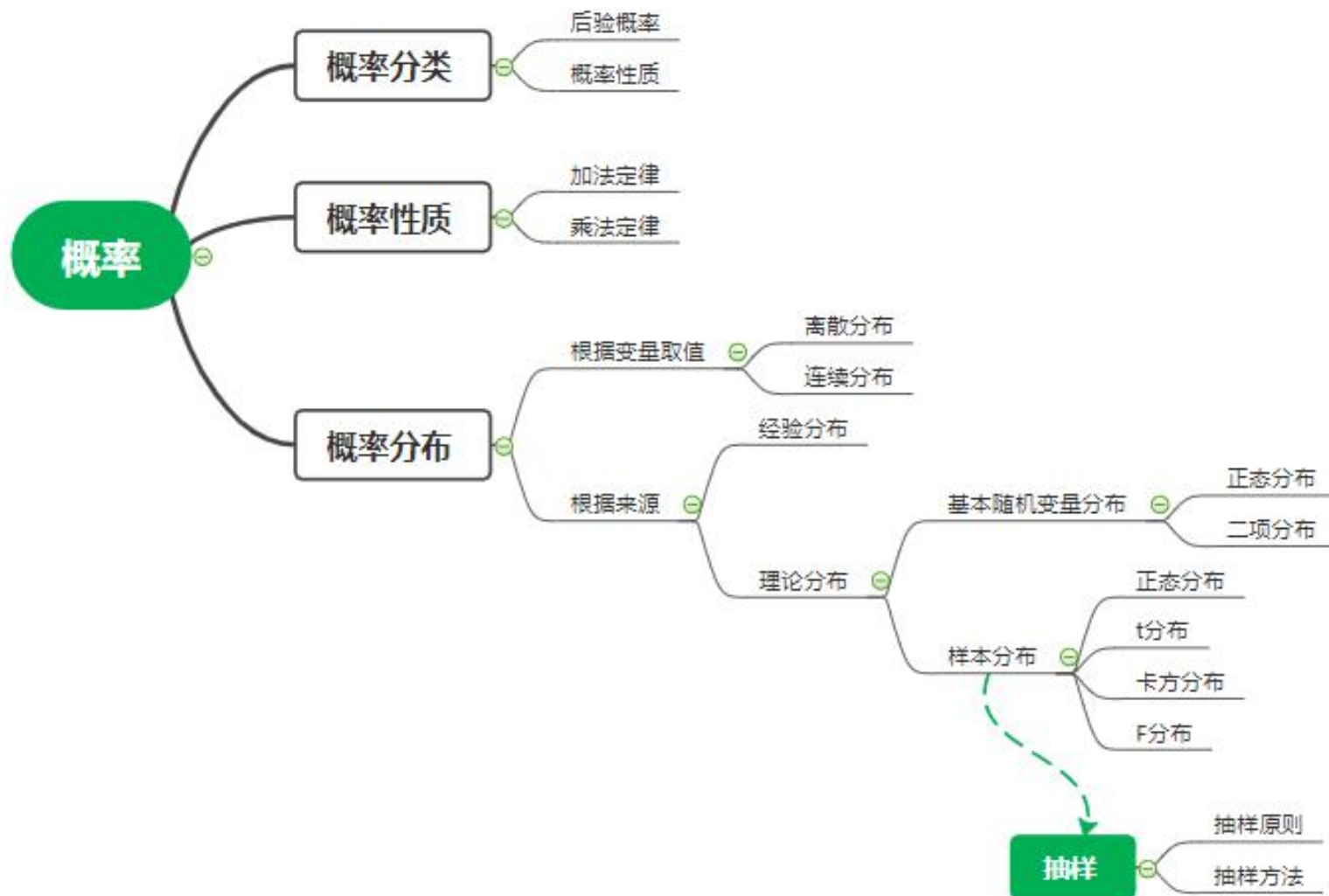


概览

- ◆ 概率（单选）
- ◆ 正态分布（选择、名词解释、简答）
- ◆ 二项分布（单选、名词解释、简答）
- ◆ 样本分布（选择、名词解释、简答）



本章结构





一、概率

(一) 定义

随机事件出现可能性大小的客观指标。

简单来说，就是变量的取值在事件总体中所占的概率。

1. 后验概率

在已经发生的无数次事件中，计算得到的某个事件（取值）的概率。

2. 先验概率

在事件发生之前通过理论方式计算所得某个事件（取值）的概率。



一、概率

(二) 基本性质

1. 公理系统

- (1) 任何事件的概率 ≥ 0
- (2) 100%发生的事件的概率=1
- (3) 100%不发生的事件的概率=0

公理 (2) 和 (3) 的逆定理不存在。

2. 加法定理

两个不相容事件A和B之和（取值分组/联合）的概率等于两个事件的概率（各自取值的概率）之和。

3. 乘法定律

两个独立事件同时出现的概率等于两个事件概率的乘积。



一、概率

(三) 分布类型

1. 离散分布和连续分布

离散分布：二项分布、泊松分布

连续分布：正态分布

2. 经验分布和理论分布

经验分布：样本数据的分布

理论分布

基本随机变量分布：正态分布、二项分布

抽样分布：t分布、卡方分布、F分布



习题演练

十次硬币投掷的顺序最可能是（H表示正面朝上，T表示背面朝上）（北师大，2017）

(1) HTHTHTHTHT

(2) HHTHTHTTTH

(3) TTTTTTTTTTTT

A. 顺序1最可能

B. 顺序2最可能

C. 顺序3最可能

D. 每个顺序概率相同



习题演练

十次硬币投掷的顺序最可能是（H表示正面朝上，T表示背面朝上）（北师大，2017）

(1) HTHTHTHTHT

(2) HHTHTHTTTH

(3) TTTTTTTTTTTT

A. 顺序1最可能

B. 顺序2最可能

C. 顺序3最可能

D. 每个顺序概率相同

答案：D



习题演练

两个四选一的选择题，一个考生全凭猜测，两题全对的概率是（ ）。

(北师大，2014)

A.0.025

B.0.0625

C.0.50

D.0.125



习题演练

两个四选一的选择题，一个考生全凭猜测，两题全对的概率是（ ）。

(北师大，2014)

A.0.025

B.0.0625

C.0.50

D.0.125

答案：B



二、正态分布

(一) 特征

1. 单峰、对称
2. 中央点最高，两侧逐渐下降，先向内弯，后向外弯，拐点在 ± 1 个标准差处，永不与横轴相交。
3. 曲线下面积为1
4. 一族分布
5. 标准差与概率面积有一定的关系

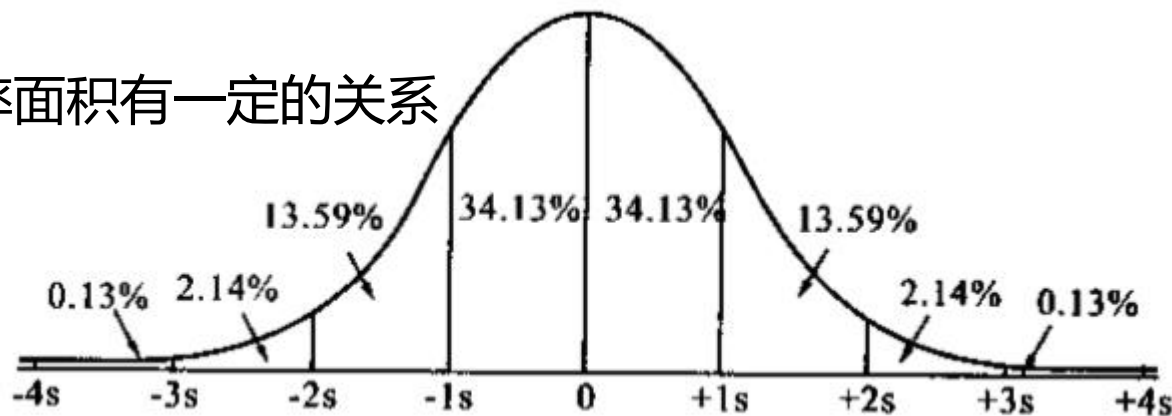


图 6-3 正态曲线下单位标准差面积比例

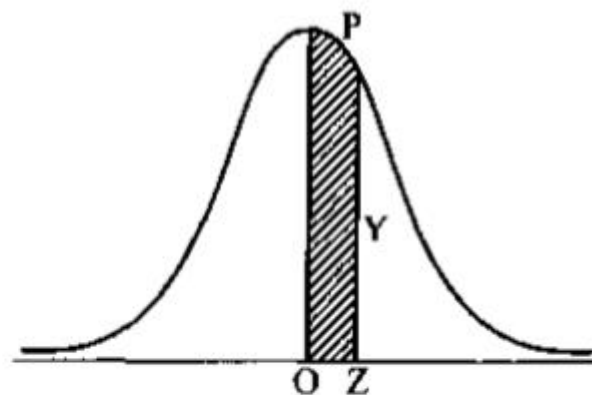


二、正态分布

(二) 查表

附录

附表1 正态分布表
(曲线下的面积与纵高)



Z	Y	P	Z	Y	P	Z	Y	P
0.00	0.39894	0.00000	0.30	0.38139	0.11791	0.60	0.33322	0.22575
0.01	0.39892	0.00399	0.31	0.38023	0.12172	0.61	0.33121	0.22907
0.02	0.39886	0.00798	0.32	0.37903	0.12552	0.62	0.32918	0.23237
0.03	0.39876	0.01197	0.33	0.37780	0.12930	0.63	0.32713	0.23565
0.04	0.39862	0.01595	0.34	0.37654	0.13307	0.64	0.32506	0.23891



二、正态分布

(三) 正态检验

1. 皮尔逊偏态量数法sk

$SK=0$ 为对称分布, $SK>0$ 为正偏态分布, $SK<0$ 为负偏态分布

2. 峰度 g_1 、偏度 g_2 检验法

偏度系数 g_1 : $g_1=0$ 为对称分布, $g_1>0$ 为正偏态分布, $g_1<0$ 为负偏态分布

峰度系数 g_2 : $g_2=0$ 为正态分布的峰度, $g_2>0$ 表明峰度比正态分布低阔, $g_2<0$ 表明峰度比正态分布高狭

3. 累加次数曲线法



二、正态分布

(四) 正态分布应用

1. 把顺序数据转换成等距数据

表 6-2 3 名教师对 100 名学生的评定结果

等级	评定结果(人数)		
	教师甲	教师乙	教师丙
A	5	10	20
B	25	20	25
C	40	40	35
D	25	20	15
E	5	10	5
总数	100	100	100

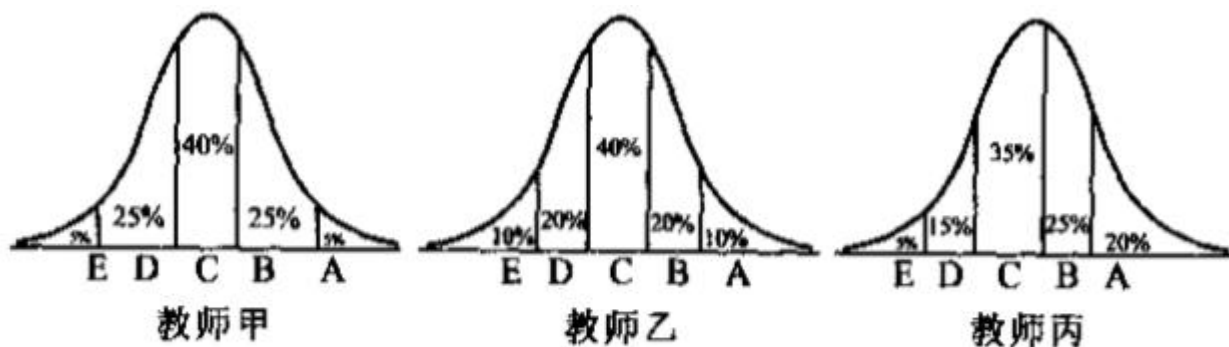


图 6-5 化等级评定为测量数据



二、正态分布

(四) 正态分布应用

2. 确定测验题目的难易度

表 6-5 难度分数的计算

测验题编号	通过率(%)	p 值	Z	Z + 5
1	99	0.49	-2.331	2.669
3	95	0.45	-1.645	3.355
5	85	0.35	-1.035	3.965
7	80	0.30	-0.84	4.160
9	70	0.20	-0.525	4.475
10	50	0	0	5.000
11	20	0.30	0.84	5.840
13	5	0.45	1.645	6.645
25	1	0.49	2.33	7.330



二、正态分布

(四) 正态分布应用

3. 在能力分组或者等级评定时确定人数

表 6-6 能力分为 5 组时各组人数的分布

分组	各组界限	比率 p	人数分布 ($p \times N$)
A	1.8σ 以上	0.0359	4
B	$0.6\sigma \sim 1.8\sigma$	0.2384	24
C	$-0.6\sigma \sim 0.6\sigma$	0.4514	44
D	$-1.8\sigma \sim -0.6\sigma$	0.2384	24
E	-1.8σ 以下	0.0359	4



二、正态分布

(四) 正态分布应用

4. 测验分数正态化

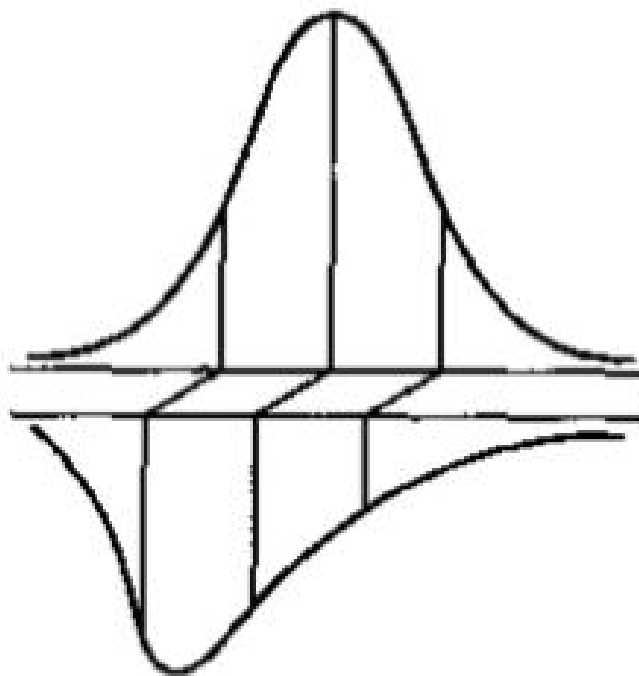


图 6-6 正态化示意图



习题演练

在标准正态曲线下，正、负三个标准差范围内的面积占总面积的比例是（ ）。 (312, 2009)

- A. 99.73% B. 99.00%
- C. 95.46% D. 95.00%

下列有关正态曲线的描述，不正确的是（ ）。 (312, 2014)

- A. 曲线以平均数为对称轴
- B. y 值越大， P 值越大
- C. 曲线与 x 轴之间区域面积为1
- D. 曲线呈现“中间高，两边低”的形状



习题演练

在标准正态曲线下，正、负三个标准差范围内的面积占总面积的比例是（ ）。 (312, 2009)

- A. 99.73% B. 99.00%
- C. 95.46% D. 95.00%

答案：A

下列有关正态曲线的描述，不正确的是（ ）。 (312, 2014)

- A. 曲线以平均数为对称轴
- B. y 值越大， P 值越大
- C. 曲线与 x 轴之间区域面积为1
- D. 曲线呈现“中间高，两边低”的形状



习题演练

在标准正态曲线下，正、负三个标准差范围内的面积占总面积的比例是（ ）。（312, 2009）

- A. 99.73% B. 99.00%
- C. 95.46% D. 95.00%

答案：A

下列有关正态曲线的描述，不正确的是（ ）。（312, 2014）

- A. 曲线以平均数为对称轴
- B. y 值越大， P 值越大
- C. 曲线与 x 轴之间区域面积为1
- D. 曲线呈现“中间高，两边低”的形状

答案：B



习题演练

培训机构甲让学生的数学成绩平均提高了30分，标准差为10；机构乙平均提高了40分，标准差为5。那么，更可能让学生数学成绩提高50分以上的培训机构是（ ）。(312, 2015)

- A. 机构甲，因为它有较大的标准差
- B. 两机构相同，概率都是0.0227
- C. 机构乙，因为它提高成绩均值较大
- D. 两机构相同，概率都是0.04



习题演练

培训机构甲让学生的数学成绩平均提高了30分，标准差为10；机构乙平均提高了40分，标准差为5。那么，更可能让学生数学成绩提高50分以上的培训机构是（ ）。(312, 2015)

- A. 机构甲，因为它有较大的标准差
- B. 两机构相同，概率都是0.0227
- C. 机构乙，因为它提高成绩均值较大
- D. 两机构相同，概率都是0.04

答案：B



习题演练

假定能力为正态分布，现取正、负各3个标准差，欲将1000名被试按照能力测验成绩等距划分为好、中、差三组。下列选项中与分配到各组的被试人数最接近的是（ ）。(312, 2018)

A. 199, 602, 199

B. 179, 642, 179

C. 159, 682, 159

D. 139, 722, 139



习题演练

假定能力为正态分布，现取正、负各3个标准差，欲将1000名被试按照能力测验成绩等距划分为好、中、差三组。下列选项中与分配到各组的被试人数最接近的是（ ）。(312, 2018)

A. 199, 602, 199

B. 179, 642, 179

C. 159, 682, 159

D. 139, 722, 139

答案：C



习题演练

某次高考分数呈正态分布，以此为基础可以（ ）。(多选) (312, 2009)

- A. 计算考生的标准分数
- B. 由P值，计算Z值
- C. 确定某一分数界限内的考生比例
- D. 知道计划录取人数后确定录取分数线



习题演练

某次高考分数呈正态分布，以此为基础可以（ ）。(多选) (312, 2009)

- A. 计算考生的标准分数
- B. 由P值，计算Z值
- C. 确定某一分数界限内的考生比例
- D. 知道计划录取人数后确定录取分数线

答案：ABCD

THANKS

- 期待下次相遇 -



高途学院APP下载



高途学院公众号

