第2课时　变量的相关性(2)



知识技能

1. 利用向量的方法研究变量的相关性．

2. 通过具体实例，理解线性相关关系中相关系数的性质．

思想方法

通过实例掌握利用相关系数的方法研究变量相关性的过程，体会利用相关系数分析变量相关性的统计思想．

核心素养

1. 在运用向量法探究相关性的过程中，发展逻辑推理素养．

2. 通过实例，体会相关性的应用，并进行相关计算，发展数学运算素养．



重点：相关系数的探求．

难点：相关系数的探求及性质．



问题导引

预习教材P141～145，思考下面的问题：

1. 用什么方法可以检验两个变量之间的线性相关程度？

提示　相关系数．

2. 你能举出一些实际生活中的例子，使它们的变量之间有明显的正相关、明显的负相关吗？你还能举出一些例子，使相关系数接近于0吗？

提示　广告费与销售额有明显的正相关；夏天的气温与热饮的销售量负相关；成年人的年龄与体重有接近于0的相关系数．

即时体验

1. 判断(正确的打“√”，错误的打“×”)：

(1) 相关系数*r*>0，表明两个变量线性相关性很强；(×)

(2) 相关系数*r*<0，表明两个变量无关；(×)

(3) 相关系数*r*越小，表明两个变量线性相关性越弱；(×)

(4) 两个变量的相关系数的绝对值越接近于1，它们的相关性越强．(√)

2. 对两个变量*x, y*进行线性相关检验，得到线性相关系数*r*1＝0.7859；对两个变量*u, v*进行线性相关检验，得到线性相关系数*r*2＝－0.9568.下列判断中正确的是(C)

A. 变量*x*与*y*正相关，变量*u*与*v*负相关，变量*x*与*y*的线性相关性较强

B. 变量*x*与*y*负相关，变量*u*与*v*正相关，变量*x*与*y*的线性相关性较强

C. 变量*x*与*y*正相关，变量*u*与*v*负相关，变量*u*与*v*的线性相关性较强

D. 变量*x*与*y*负相关，变量*u*与*v*正相关，变量*u*与*v*的线性相关性较强

提示　由线性相关系数*r*1＝0.7859＞0知*x*与*y*正相关，由线性相关系数*r*2＝－0.9568＜0知*u*与*v*负相关．又|*r*1|＜|*r*2|，所以变量*u*与*v*的线性相关性比*x*与*y*的线性相关性强．

3. 为了研究两个变量*x*与*y*之间的关系，分别选择了四个不同的模型进行拟合，它们的相关系数*r*如下表所示，其中拟合效果最好的模型是(A)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 模型 | 1 | 2 | 3 | 4 |
| *r* | 0.98 | 0.80 | 0.50 | 0.25 |

A. 模型1 B. 模型2

C. 模型3 D. 模型4



一、 问题情境

在上节课的学习中，年龄与脂肪含量的散点基本上在一条直线附近，我们可以粗略地估计这两个变量间有线性相关关系，但是它们线性相关的程度如何呢？又如何较为精确地刻画线性相关关系呢？



二、 数学建构

(一) 生成概念

下面我们用向量的方法来研究数据的相关性．

根据向量的数量积***a***·***b***＝|***a***||***b***|cos*θ*可知cos*θ*＝，其中*θ*为向量***a, b***的夹角．

问题1　对于向量***a***＝(*a*1, *a*2), ***b***＝(*b*1, *b*2), cos*θ*＝，当cos*θ*满足什么条件时，***a, b***趋于共线？

对于向量***a***＝(*a*1, *a*2), ***b***＝(*b*1, *b*2)，由cos*θ*＝知，|cos*θ*|越接近于1, ***a, b***的夹角*θ*就越接近于0或π，这时，向量***a, b***趋于共线．

对于两对数据(*x*1, *y*1), (*x*2, *y*2)，设点*A*1(*x*1, *y*1), *A*2(*x*2, *y*2)，线段*A*1*A*2的中点为*M*(， ).

构造向量***a, b, a***＝(*x*1－， *x*2－), ***b***＝(*y*1－， *y*2－)，当***a, b***共线时，存在非零实数*λ*，使得

***b***＝*λ****a***，

即

这说明向量与共线，即*A*1, *A*2, *M*三点共线(图1)．



图1

问题2　对于向量***a***＝(*a*1, *a*2, *a*3), ***b***＝(*b*1, *b*2, *b*3), cos*θ*＝，当cos*θ*满足什么条件时，***a, b***趋于共线？

对于向量***a***＝(*a*1, *a*2, *a*3), ***b***＝(*b*1, *b*2, *b*3)，由cos*θ*＝知，|cos*θ*|越大(越接近于1)，***a, b***的夹角*θ*就越接近于0或π，这时，向量***a, b***趋于共线．

对于三对数据(*x*1, *y*1), (*x*2, *y*2), (*x*3, *y*3)，设点*A*1(*x*1, *y*1), *A*2(*x*2, *y*2), *A*3(*x*3, *y*3)，取点*M*(， ).

构造向量***a, b, a***＝(*x*1－， *x*2－， *x*3－), ***b***＝(*y*1－， *y*2－， *y*3－)，并记〈***a, b***〉＝*θ*，则

*cos*θ＝.

当|cos*θ*|越大(越接近于1时)，***a, b***的夹角*θ*就越接近于0或π，这时，向量***a, b***趋于共线．当***a, b***共线时，存在非零实数*λ*，使得

***b***＝*λ****a***，

即

这说明，向量， ， 趋于共线，即*A*1, *A*2, *A*3, *M*四点接近于共线(图2)．



图2

问题3　一般地，对于向量***a***＝(*a*1, *a*2, …， *an*), ***b***＝(*b*1, *b*2, …， *bn*)，

*cos*θ＝

＝.

当cos*θ*满足什么条件时，***a, b***趋于共线？ [1]

|cos*θ*|越大(越接近于1)，***a, b***的夹角*θ*就越接近于0或π，这时，向量***a, b***趋于共线．

一般地，对于*n*对数据(*x*1, *y*1), (*x*2, *y*2), …， (*xn, yn*)，设点*A*1(*x*1, *y*1), *A*2(*x*2, *y*2), …， *An*(*xn, yn*)，取点*M*(， ).

构造向量***a, b, a***＝(*x*1－， *x*2－， …， *xn*－), ***b***＝(*y*1－， *y*2－， …， *yn*－)，并记〈***a, b***〉＝*θ*，则

*cos*θ＝.(\*)

当|cos*θ*|越大(越接近于1时)，***a, b***的夹角*θ*就越接近于0或π，这时，向量***a, b***趋于共线．当***a, b***共线时，存在非零实数*λ*，使得

***b***＝*λ****a***，

即

这说明，向量， ， …， 趋于共线，即*A*1, *A*2, …， *An, M*这(*n*＋1)点接近于共线(图3)．



图3

我们将(\*)式称为*n*对数据(*x*1, *y*1), (*x*2, *y*2), …， (*xn, yn*)的相关系数，记为*r*.

相关系数*r*可由下面的公式计算：

r＝

＝.

(二) 理解概念

相关系数*r*的性质：

(1) |*r*|≤1，即－1≤*r*≤1.

(2) *r*>0时，*y*与*x*呈正相关关系；*r*<0时，*y*与*x*呈负相关关系．

(3) |*r*|越接近于1, *y*与*x*的相关程度就越强；|*r*|越接近于0, *y*与*x*的相关程度就越弱．

(4) 当|*r*|＝1时，所有数据点都在一条直线上．

通常情况下，当|*r*|>0.5时，认为线性相关关系显著；当|*r*|<0.3时，认为几乎没有线性相关关系．

针对问题情境中的问题，用*x*表示年龄，*y*表示脂肪含量，利用上一课时提供的数据可得

*x*i＝23＋27＋…＋60＋61＝673，

*y*i＝9.5＋17.8＋…＋35.2＋34.6＝381.7，

*x*i*y*i＝23×9.5＋27×17.8＋…＋60×

 35.2＋61×34.6＝19403.2，

*x*＝232＋272＋…＋602＋612＝34181，

*y*＝9.52＋17.82＋…＋35.22＋34.62

＝11051.77，

所以

r＝

＝

≈0.967，

所以脂肪含量与年龄的相关程度很强．

三、 数学运用

例1[2]　(1) (多选)下列关于相关系数*r*的说法中正确的有(ABC)

A. *r*可以用来衡量两个变量之间的线性相关程度

B. |*r*|≤1，且|*r*|越接近0，相关程度越弱

C. |*r*|≤1，且|*r*|越接近1，相关程度越强

D. |*r*|≤1，且|*r*|越接近1，相关程度越弱

(2) 已知甲、乙、丙、丁四组数据变量间对应的线性相关系数分别为0.46, 0.79, －0.92, 0.85，则(C)

 甲组数据变量间的线性相关程度最强

 乙组数据变量间的线性相关程度最弱

 丙组数据变量间的线性相关程度最强

 丁组数据变量间的线性相关程度最强

(见学生用书课堂本P88)

[处理建议]　(1)根据相关关系的性质逐一辨析．(2)|*r*|越接近1，相关程度越强．

[规范板书]　解析　(2) 因为线性相关系数的绝对值越大，线性相关性越强，所以丙组数据的线性相关性最强．

[题后反思]　相关系数*r*的性质：(1)|*r*|≤1; (2)|*r*|越接近于1，两个变量的线性相关程度越强；(3)|*r*|越接近于0，两个变量的线性相关程度越弱．

　为了比较甲、乙、丙三组数据的线性相关性的强弱，小郑分别计算了甲、乙、丙三组数据的线性相关系数，其数值分别为0.939, 0.937, 0.948，则(D)

A. 甲组数据的线性相关性最强，乙组数据的线性相关性最弱

B. 乙组数据的线性相关性最强，丙组数据的线性相关性最弱

C. 丙组数据的线性相关性最强，甲组数据的线性相关性最弱

D. 丙组数据的线性相关性最强，乙组数据的线性相关性最弱

例2[3]　(1) 在一组样本数据为(*x*1, *y*1), (*x*2, *y*2), …， (*xn, yn*)(*n*≥2, *x*1, *x*2, …， *xn*不全相等)的散点图中，若所有样本点(*x*i*, y*i)(i＝1, 2, …， *n*)都在直线*y*＝－*x*＋2上，则这组样本数据的相关系数为(A)

A. －1 B. 1

C. － D.

(2) 已知变量*x, y*相对应的一组数据为(10, 1), (11.3, 2), (11.8, 3), (12.5, 4), (13, 5)，变量*x*′， *y*′相对应的一组数据为(10, 5), (11.3, 4), (11.8, 3), (12.5, 2), (13, 1)，用*r*1表示变量*x*与*y*之间的线性相关系数，用*r*2表示变量*x*′与*y*′之间的线性相关系数，则有(C)

A. *r*2<*r*1<0 B. 0<*r*2<*r*1

C. *r*2<0<*r*1 D. *r*1<*r*2<0

(见学生用书课堂本P89)

[处理建议]　(1) 由样本数据的所有样本点都在一条直线上，得出这组样本数据完全相关．

(2) 观察数据，分析正相关、负相关．由*r*1>0, *r*2<0比较*r*1, *r*2的大小．

[规范板书]　解析　(1) 根据直线方程是*y*＝－*x*＋2，得这两个变量负相关，故这组样本数据的样本相关系数为负值，且所有样本点(*x*i*, y*i)(i＝1, 2, …， *n*)都在直线上，则有|*r*|＝1，所以相关系数*r*＝－1.

(2) 从第一组数据看出*y*与*x*正相关，则*r*1>0；从第二组数据看出*y*与*x*负相关，则*r*2<0.故*r*2<0<*r*1.

[题后反思]　若所有样本点都在一条直线上，则*r*＝±1.此时，若为正相关，则*r*＝1；若为负相关，则*r*＝－1.

　在一组样本数据为(*x*1, *y*1), (*x*2, *y*2), …， (*xn, yn*)(*n*≥2, *x*1, *x*2, …， *xn*不全相等)的散点图中，若所有样本点(*x*i*, y*i)(i＝1, 2, …， *n*)都在直线*y*＝*x*＋1上，则这组样本数据的相关系数为(B)

A. －1 B. 1

C. － D.

[规范板书]　解析　这组样本数据的所有样本点(*x*i*, y*i)(i＝1, 2, …， *n*)都在直线*y*＝*x*＋1上，所以这组样本数据完全相关，即说明这组数据的样本完全正相关，其相关系数是1.

　某统计部门对四组数据进行统计分析后，获得如图所示的散点图，关于相关系数的比较正确的是(A)



(变式2)

A. *r*2<*r*4<0<*r*3<*r*1 B. *r*4<*r*2<0<*r*1<*r*3

C. *r*4<*r*2<0<*r*3<*r*1 D. *r*2<*r*4<0<*r*1<*r*3

[规范板书]　解析　根据散点图的特征，数据大致呈增长趋势的是正相关，大致呈递减趋势的是负相关；数据越集中在一条直线附近，说明相关性越强．

易知图1, 3为正相关，图2, 4为负相关，

故*r*1＞0, *r*3＞0, *r*2＜0, *r*4＜0.

又图1与图2中散点图更接近于一条直线，故*r*1＞*r*3, *r*2＜*r*4.

因此，*r*2＜*r*4＜0＜*r*3＜*r*1.

　(多选)为了对变量*x*与*y*的线性相关性进行检验，由样本点(*x*1, *y*1), (*x*2, *y*2), …， (*x*10, *y*10)求得两个变量的样本相关系数为*r*，下列说法中错误的有(ABD)

A. 若所有样本点都在直线*y*＝－2*x*＋1上，则*r*＝1

B. 若所有样本点都在直线*y*＝－2*x*＋1上，则*r*＝－2

C. 若|*r*|越大，则变量*x*与*y*的线性相关性越强

D. 若|*r*|越小，则变量*x*与*y*的线性相关性越强

[规范板书]　解析　若所有样本点都在直线*y*＝－2*x*＋1上，则直线斜率为负数，*r*＝－1, A, B均错误；若|*r*|越大，则变量*x*与*y*的线性相关性越强，C正确，D错误.

例3　基于移动互联技术的共享单车被称为“新四大发明”之一，短时间内就风靡全国，带给人们新的出行体验．某共享单车运营公司的市场研究人员为了了解公司的经营状况，对该公司最近六个月内的市场占有率进行了统计，结果如下表所示：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 月份 | 2020.8 | 2020.9 | 2020.10 | 2020.11 | 2020.12 | 2021.1 |
| 月份代码*x* | 1 | 2 | 3 | 4 | 5 | 6 |
| 市场占有率*y*/% | 11 | 13 | 16 | 15 | 20 | 21 |

(1) 请在给出的坐标纸中作出散点图；

(2) 计算相关系数；

(3) 将散点图与相关系数进行比照分析，并作出解释．[4]



(例3)

参考数据： (*x*i－)2＝17.5, (*x*i－)·(*y*i－)＝35, ≈36.5.

参考公式：相关系数

r＝.

(见学生用书课堂本P89)

[处理建议]　作出散点图，根据相关系数公式计算*r*，并作出判断．

[规范板书]　解　(1) 散点图如图所示．



(例3答图)

(2) ＝＝16, (*y*i－)2＝76，

所以r＝＝＝≈≈0.96,

从而两变量之间具有较强的线性相关关系．

(3) 从散点图以及相关系数这两个角度分析，都可以看出两个变量都具有较强的线性相关关系，两者的结论是一致的．

[题后反思]　在综合性问题中，对相关系数的理解可以从形和数两个角度出发．

一方面可以将图表数据转化成散点图，从图形角度直观感受线性相关，即为定性分析．在散点图中，样本点在某条直线附近越集中，两变量的线性相关关系越强；样本点在某条直线附近越分散，两变量的线性相关关系越弱．

另一方面可以利用公式计算得到相关系数的值，然后进行定量分析．|*r*|越大，两变量的线性相关关系越强；|*r*|越小，两变量的线性相关关系越弱．

这两个角度所获得的结论是一致的．

　互联网使我们的生活日益便捷，网络外卖也开始成为不少人日常生活中不可或缺的一部分，某市一调查机构针对该市市场占有率较高的甲、乙两家网络外卖企业(以下称外卖甲、外卖乙)的经营情况进行了调查，调查结果如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1日 | 2日 | 3日 | 4日 | 5日 |
| 外卖甲日接单*x*/百单 | 5 | 2 | 9 | 8 | 11 |
| 外卖乙日接单*y*/百单 | 2 | 3 | 10 | 5 | 15 |

(1) 试根据表格中这五天的日接单量情况，从统计的角度说明这两家外卖企业的经营状况；

(2) 据统计表明，*y*与*x*之间具有线性相关关系，请用相关系数*r*对*y*与*x*之间的相关性强弱进行判断．(若|*r*|>0.75，则可认为*y*与*x*有较强的线性相关关系，*r*的值精确到0.001)

参考数据： (*x*i－)(*y*i－)＝66, ≈77.

[处理建议]　通过分析甲、乙的均值与方差，比较二者之间的经营状况； 通过计算相关系数，分析*x*，*y*的线性相关关系．

[规范板书]　解　(1) ＝＝7, ＝＝7，

外卖甲的日接单量的方差*s*＝＝10，

外卖乙的日接单量的方差*s*＝＝23.6.

因为＝， *s*<*s*，即外卖甲平均日接单量与外卖乙平均日接单量相同，但外卖甲日接单量更集中一些，所以外卖甲比外卖乙经营状况更好．

(2) r＝≈≈0.857>0.75，

所以可认为*y*与*x*之间有较强的线性相关关系．

四、 课堂练习

1. 关于样本相关系数的说法中错误的是(C)

A. 可以用来判断成对样本数据相关的正负性

B. 可以是正的，也可以是负的

C. 样本相关系数越大，成对样本数据的线性相关程度越高

D. 取值范围是[－1, 1]

2. (多选)下列散点图与相关系数*r*一定不符合的有(ACD)



3. 在一组样本数据为(*x*1, *y*1), (*x*2, *y*2), …， (*xn, yn*)(*n*≥2, *x*1, *x*2, …， *xn*不全相等)的散点图中，若所有样本点(*x*i*, y*i)(i＝1, 2, …， *n*)都在直线*y*＝－*x*－3上，则这组样本数据的相关系数为(A)

A. －1 B. 1

C. － D.

4. 某项科研活动共进行了5次试验，其变量*x*与*y*的数据如下表所示：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 第1次 | 第2次 | 第3次 | 第4次 | 第5次 |
| *x* | 2 | 5 | 8 | 9 | 11 |
| *y* | 12 | 10 | 8 | 8 | 7 |

(1) 根据表中的数据，作出散点图；

(2) 计算*x*与*y*之间的相关系数，并刻画它们的相关程度．

参考公式：相关系数

r＝.

解　(1) 散点图如图所示：



(第4题)

(2) 由题意得＝7, ＝9，

 (*x*i－)(*y*i－)＝－28, (*x*i－)2＝50, (*y*i－)2＝16，

所以r＝＝

≈－0.99.

|*r*|≈0.99很接近1，说明*x, y*线性相关性很强．由于*r*<0，故其关系为负相关．

五、 课堂小结

1. 相关系数公式：

r＝

＝.

2. 相关系数*r*的性质：

(1) |*r*|≤1，即－1≤*r*≤1.

(2) *r*>0时，*y*与*x*呈正相关关系；*r*<0时，*y*与*x*呈负相关关系．

(3) |*r*|越接近于1, *y*与*x*的相关程度就越强；|*r*|越接近于0, *y*与*x*的相关程度就越弱.



[1] 从学生熟悉的向量背景入手，分别研究二维、三维、*n*维向量共线时，cos*θ*的性质，为研究*n*对数据的相关系数作准备．

[2] 理解相关系数的性质．

[3] 通过具体模型理解相关系数的概念．

[4] 运用相关系数解决实际问题．