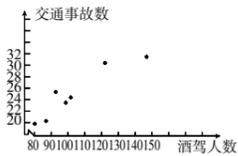


一、单项选择题

1.D 提示:对于 ABC,两个变量之间为确定性关系,即两个变量之间均为函数关系;对于 D,根据生活经验,天空中的云量和下雨是相关关系.故选 D.

2.A 提示:由表格中的数据,在直角坐标系中描出数据的散点图,如图所示,直观判断散点从左向右成带状分布,在一条直线附近,所以具有线性相关关系,且是正相关.故选 A.



(第 2 题图)

3.C 提示:①中,汽车的重量和汽车每消耗 1 升汽油所行驶的距离是负相关的关系;②中,平均日学习时间和平均学习成绩的关系是一个正相关;③中,某人每日吸烟量和其身体健康情况是负相关的关系;④中,圆的半径与面积是函数关系;⑤中,汽车的重量和每千米耗油量关系是一个正相关.所以②⑤中的两个变量属于线性正相关.故选 C.

4.A 提示:由题意可知,这组样本数据呈完全负相关,所以其相关系数 r=-1.故选 A.

5.B 提示:对于 A,数据组①的相关系数 r1=0,无相关性,故 A 错误;

对于 B、C、4 个数据组中,数据组②的相关系数 r2=-0.95,绝对值最大,其两变量线性相关性最强,故 B 正确,C 错误;

对于 D,由 r2>r1,得 D 错误.故选 B.

6.C 提示:因为 r1=0.837,r2=-0.957,所以变量 X 与 Y 之间呈正相关关系,变量 U 与 V 之间呈负相关关系,且 X 与 Y 之间的相关性弱于 U 与 V 之间的相关性.故选 C.

7.A 提示:对于 A,用相关系数 r 来衡量两个变量之间线性关系的强弱时,|r|越接近于 1,相关性越强,故 A 错误;对于 B,由相关系数 r 的含义知:当相关系数 r<0 时,表明两变量负相关,故 B 正确;

对于 C,由相关系数的含义知:当相关系数 r=0 时,表明两变量线性不相关,故 C 正确;

对于 D,|r|很接近于 1,故有较强的线性相关程度,故 D 正确.故选 A.

8.A 提示:由给出的四组数据的散点图可以看出,图 1 和图 3 是正相关,相关系数大于 0,

图 2 和图 4 是负相关,相关系数小于 0,图 1 和图 2 的点相对更加集中,所以相关性更强,所以 r1 接近于 1,r2 接近于-1,由此可得 r2<r1<0<r3<r1,故选 A.

二、多项选择题

9.AC 提示:通过对诗句的理解可判断选项 A 和 C 为相关关系,选项 B 不是相关关系,选项 D 中理想状态下自由下落的距离为 s=1/2gt^2,与速度无关,故选项 D 不是相关关系,故选 AC.

10.BC 提示:A 中的点无规律分布,范围很广,表明两个变量之间的相关程度很小;B、C 中的点分布在一条直线的附近,两个变量之间具有线性相关关系;D 中所有的点分布在一条曲线附近,所以不是线性相关关系.故选 BC.

11.AC 提示:对于 A,样本相关系数即为其标准化数据向量夹角的余弦值,故 A 正确;

对于 B、C,样本相关系数的取值范围是[-1,1],故 B 错误,C 正确;

对于 D,变量 x 与 y 的线性回归方程为 y=1.5x-2.15>0,则 x 与 y 正相关,故 D 错误.故选 AC.

12.BCD 提示:对于 A,2021 年 7 月的异地快递量为 572812.9 万件,同城快递量为 105191.1 万件,异地快递量小于同城快递量的 6 倍,故 A 错误;对于 B,因为 679556.6-599604.6 > 708642.6-679556.6,9 月异地快递增长率明显高于 10 月异地快递增长率,故 B 正

确;对于 C,由图可看出,除 2021 年 12 月异地快递量较 11 月略少,其余都有较明显增加,因此可以判断异地快递量与月份呈正相关关系,故 C 正确;

对于 D,由图可看出,同城和异地快递量最高都在 11 月份,故 D 正确.故选 BCD.

三、填空题

13.有 提示:物品大小的值由小变大时,销售价格也由小变大,因此,两个变量间有相关关系.

14.甲 提示:因为线性相关系数的绝对值越大,线性相关性越强,甲、乙、丙、丁四组数据的线性相关系数分别为-0.98,0.87,0.58,0.92,所以甲组数据的线性相关性最强.

15.-1 提示:因为过点(1,2),(2,0)的直线的斜率为 k=(2-0)/(1-2)=-2,所以过点(1,2),(2,0)的直线的方程为 y=-2(x-2),即 y=4-2x,经检验,点(4,-4),(-1,6)都在直线 y=4-2x 上,

所以 y 与 x 是函数关系,又因为 y=4-2x 单调递减,所以 y 与 x 的相关系数为-1.

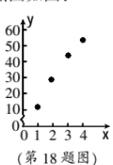
16.0.85 提示:r = (sum(x_i - x_bar)(y_i - y_bar) / (sum(x_i - x_bar)^2 * sum(y_i - y_bar)^2))^0.5 = 0.85

1.2 * sum(y_i - y_bar)^2 / (sqrt(2) * sum(y_i - y_bar)^2) = 1.2 / sqrt(2) approx 0.85

四、解答题

17.解:因为“Secchi 深度”越大说明湖水越清洁,湖水被水藻污染的程度越小,而叶绿素的总浓度越大湖水被水藻污染的程度越大,所以“Secchi 深度”和叶绿素的总浓度,这两个变量间是负相关.

18.解:作出散点图如图:



(第 18 题图)

由散点图可知,各点大致分布在一条直线附近,由此推断 x 与 y 线性相关.

由题中所给表格及参考数据得:

x_bar = 5/2, y_bar = 69/2, sum(x_i y_i) = 418, sqrt(sum(y_i - y_bar)^2) approx 32.6

sum(x_i^2) = 30, sum((x_i - x_bar)(y_i - y_bar)) = sum(x_i y_i - 4x_i y_bar) = 418 - 4 * 5/2 * 69/2 = 5/2

69/2 = 73, sqrt(sum((x_i - x_bar)^2)) = sqrt(sum(x_i^2 - 4x_i + 5)) = sqrt(30 - 4 * 5 + 5) = 5

sqrt(5) approx 2.24, 所以 r = (sum(x_i - x_bar)(y_i - y_bar) / (sqrt(sum(x_i - x_bar)^2) * sqrt(sum(y_i - y_bar)^2))) approx 0.9997

因为 y 与 x 的相关系数近似为 0.9997,可以推断该公司的年销量 y 与第 x 年呈正线性相关,且线性相关程度很强.

19.解:(1)由表格中的数据,可得 x_bar = (5+2+9+8+11)/5 = 7, y_bar = (2+3+10+5+15)/5 = 7

外卖甲的日接单量的方差 s_a^2 = ((5-7)^2 + (2-7)^2 + (9-7)^2 + (8-7)^2 + (11-7)^2) / 5 = 10

外卖乙的日接单量的方差 s_b^2 = ((2-7)^2 + (3-7)^2 + (10-7)^2 + (5-7)^2 + (15-7)^2) / 5 = 23.6

因为 x_bar = y_bar, s_a^2 < s_b^2,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)因为 r = (sum(x_i - x_bar)(y_i - y_bar) / (sqrt(sum(x_i - x_bar)^2) * sqrt(sum(y_i - y_bar)^2))) approx 0.857 > 0.75

又 sum((x_i - x_bar)(y_i - y_bar)) = 66, sqrt(sum((x_i - x_bar)^2) * sum((y_i - y_bar)^2)) approx 77

所以代入计算可得,相关系数 r approx 66/77 approx 0.857 > 0.75,所以可认为 y 与 x 之间有较强的线性相关关系.

20.解:(1)t_bar = (1+2+3+4+5+6+7)/7 = 4, sum(t_i - t_bar)^2 = 28, sum(t_i y_i - 7t_bar y_bar) = 42.1, sqrt(sum(y_i - 7y_bar)^2) = 8.1, sqrt(7) approx 2.65, 所以 r = (sum(t_i y_i - 7t_bar y_bar) / (sqrt(sum(t_i - t_bar)^2) * sqrt(sum(y_i - 7y_bar)^2))) approx 0.98

因为交易额 y 与 t 的相关系数近似为 0.98,说明交易额 y 与 t 具有很强的正线性相关,

从而可用线性回归模型拟合交易额 y 与 t 的关系.

(2)y_bar = 35/7 = 5(千万元), sum(t_i - 7t_bar)^2 = 28, sum(t_i y_i - 7t_bar y_bar) = 42.1, 则 b_hat = (sum(t_i y_i - 7t_bar y_bar) / sum(t_i - 7t_bar)^2) = 42.1 / 28 approx 1.5, a_hat = y_bar - b_hat = 5 - 1.5 * 4 = -1

所以 y 关于 t 的线性回归方程为 y=1.5t-1,将 t=8 代入线性回归方程得 y=1.5*8-1=11(千万元)=1.1(亿元),所以预测下一周的第一天的交易额为 1.1 亿元.

21.解:(1)由题意得, x_bar = 2019, y_bar = 1, 所以 r = (sum(x_i y_i - 5x_bar y_bar) / (sqrt(sum(x_i^2 - 5x_bar)^2) * sqrt(sum(y_i^2 - 5y_bar)^2))) approx 0.998 > 0.75, 所以 y 与 x 线性相关性很强.

(2)由题意知, b_hat = (sum(x_i y_i - 5x_bar y_bar) / sum(x_i^2 - 5x_bar^2)) = 3.6/10 = 0.36

a_hat = y_bar - b_hat x_bar = 1 - 2019 * 0.36 = -725.84, 所以 y 关于 x 的线性回归方程是 y=0.36x-725.84, 当 x=2023 时, y=0.36 * 2023 - 725.84 = 2.44, 所以该地区 2023 年足球特色学校有 244 个.

22.解:(1)设这棵树木平均一棵的根部横截面积为 x_bar, 平均一棵的材积量为 y_bar, 则根据题中数据得 x_bar = 0.6/10 = 0.06m^2, y_bar = 3.9/10 = 0.39m^3

(2)由题意知, r = (sum((x_i - x_bar)(y_i - y_bar)) / (sqrt(sum((x_i - x_bar)^2) * sum((y_i - y_bar)^2)))) approx 0.97

(3)设根部面积总和为 X, 总材积量为 Y, 则 X/Y = x_bar/y_bar, 故 Y = (0.39/0.06) * 186 = 1209m^3

第 23 期

一、单项选择题

1.C 2.C 提示:“是否吸烟”是分类变量,它的两个不同取值:吸烟和不吸烟.“是否患病”是分类变量,它的两个不同取值:患病和不患病.可知 A、B 都是一个分类变量所取的两个不同值.故选 C.

3.C 提示:由题意,得 c=120-73-25=22, a=74-22=52, b=73-52=21, 所以 a-b-c=52-21-22=9, 故选 C.

4.D 提示:根据分类变量 X 与 Y 的观察数据知, chi^2 = 3.174 > 2.706, 且 3.174 < 3.841, 所以变量 X 与 Y 不独立, 这个结论犯错误的概率不超过 10%. 故选 D.

5.B 提示:根据独立性检验的基本思想可知, 分类变量 X 与 Y 的随机变量 chi^2 的观测值越大, “X 与 Y 没有关系”的可信程度越小, 则“X 与 Y 有关系”的可信程度越大; chi^2 越小, “X 与 Y 有关系”的可信程度越小, “X 与 Y 没有关系”的可信程度越大, 故 ACD 错误, B 正确. 故选 B.

6.D 提示:根据题意, 填写列联表如下:

Contingency table with columns: 专业 A, 专业 B, 合计; rows: 女生, 男生, 合计. Values: 女生(12, 4, 16), 男生(38, 46, 84), 合计(50, 50, 100).

则 chi^2 = (100 * ((12 * 46 - 4 * 38)^2) / (16 * 84 * 50 * 50)) = 100 / 21 approx 4.762 > 4.762 > 3.841, 所以认为工科院校中“性别”与“专业”有关的把握有 95%, 故选 D.

7.A 提示:由表可得, chi^2 = (100 * (40 * 25 - 15 * 20)^2) / (55 * 45 * 60 * 40) approx 8.249 > 6.635, 有 99% 的把握认为阅读量多少与幸福感强弱有关, 故 A 正确, B 错误; 在幸福感强的人中随机抽取 1 人, 此人是阅读量多的人的概率约为 40 / (40 + 15) = 8/11 approx 0.72, 故 C 错误; 在阅读量大的人中随机抽取一

人, 此人是幸福感强的人的概率约为 40 / (40 + 20) approx 0.67, 故 D 错误. 故选 A.

8.A 提示:对于 A, 喜欢糯米类食物(粽子, 元宵)的有 50+30=80 人, 喜欢小麦类食物(饺子)的有 120 人, 故 A 错误;

对于 B, 根据题意得 2x2 列联表如下:

Contingency table with columns: 粽子、元宵, 饺子, 合计; rows: 男, 女, 合计. Values: 男(30, 70, 100), 女(50, 50, 100), 合计(80, 120, 200).

chi^2 = (200 * (30 * 50 - 50 * 70)^2) / (80 * 120 * 100 * 100) approx 8.33 > 6.635, 所以有 99% 的把握认为是是否喜欢糯米类食物(粽子, 元宵)与性别有关, 故 B 正确;

对于 C, 喜欢粽子的频率, 即概率为 50 / 200 = 1/4, 则 xi - B(n, 1/4), 故 C 正确;

对于 D, 根据题意得 2x2 列联表如下:

Contingency table with columns: 粽子, 元宵, 合计; rows: 男, 女, 合计. Values: 男(20, 10, 30), 女(30, 20, 50), 合计(50, 30, 80).

chi^2 = (80 * (20 * 20 - 30 * 10)^2) / (30 * 50 * 30 * 50) approx 0.356 < 2.706, 所以喜欢糯米类食物(粽子, 元宵)的人中, 是否喜欢粽子与性别没有明显关系, 故 D 正确. 故选 A.

二、多项选择题

9.BD 提示:由题意知, chi^2 = (100 * (35 * 25 - 25 * 15)^2) / (60 * 40 * 50 * 50) approx 4.17, 因为 chi^2 > 3.841, 所以有 95% 的把握认为“喜欢足球与性别相关”, 故 A 错误, B 正确;

因为 chi^2 < 6.635, 所以没有 99% 的把握认为“喜欢足球与性别有关”, 故 C 错误, D 正确. 故选 BD.

10.BD 提示:由题意, 把频率看作概率可得夜晚下雨的概率约为 (25+25)/100 = 1/2, 故 A 错误; 未出现“日落云里走”, 但夜晚下雨的概率约为 25/(25+45) = 5/14, 故 B 正确; 由 chi^2 = 19.05 > 6.635, 所以可知有 99% 的把握认为“‘日落云里走’是否出现”与“当晚是否下雨”有关, 故 D 正确, C 错误. 故选 BD.

11.CD 提示:由题意知, 被调查的男女生人数相同, 设男生的人数为 5m(m in N+), 则可列出 2x2 列联表:

Contingency table with columns: 男生, 女生, 合计; rows: 喜欢躺着看书, 不喜欢躺着看书, 合计. Values: 喜欢躺着看书(2m, 4m, 6m), 不喜欢躺着看书(3m, m, 4m), 合计(5m, 5m, 10m).

chi^2 = (10m * (2m * m - 4m * 3m)^2) / (6m * 4m * 5m * 5m) = 5m/3, 因为有 95% 的把握认为是是否喜欢躺着看书和性别有关, 所以 3.841 <= 5m/3 < 6.635, 解得 11.523 <= 5m < 19.905, 因为 m in N+, 所以 A、B 错误, C、D 正确. 故选 CD.

12.ABD 提示:依题意, 140/n = 7/10, 解得 n=60, 由 m/(m+80) = 3/5, 解得 m=120, 故 A 正确; chi^2 = (400 * ((140 * 80 - 120 * 60)^2) / (260 * 140 * 200 * 200)) = 400/91 approx 4.396 > 3.841, 则有 95% 的把握认为市民性别与喜欢冰雪运动有关系, 故 B 正确; 随机对一路人进行调查, 喜欢冰雪运动的频率为 (140+120)/400 = 65/100 = 65%, 则有 65% 的可能性对方喜欢冰雪运动, 故 C 错误; chi^2 = 4.396 < 6.635, 没有 99% 的把握认为市民性别与喜欢冰雪运动有关系, 故 D 正确. 故选 ABD.

三、填空题 13.女教授人数, 男教授人数, 女副教授人数, 男副教授人数 14.16:16 提示:由表中数据得 m=20-4=16, 所以经常打篮球的人数为 16+8=24, 故 n=40-24=16.

15.4.844; 能 提示:由表得 chi^2 = (50 * (13 * 20 - 7 * 10)^2) / (23 * 27 * 20 * 30) approx 4.844 > 3.841, 所以能在犯错误的概率不超过 0.05 的前提下认为主修统计专业与性别有关系.

N+, 可得 n=20, 因为 chi^2 approx 4.040 > 3.841, 所以, 有 95% 的把握认为该校学生对冬季奥运会项目的了解情况与性别有关.

19.解:(1)由题意, 可得以下 2x2 列联表:

Contingency table with columns: 患心脏病, 患其他病, 总计; rows: 秃顶, 不秃顶, 总计. Values: 秃顶(200, 150, 350), 不秃顶(400, 600, 1000), 总计(600, 750, 1350).

P1 = 200/350 = 4/7, P2 = 400/1000 = 2/5, 因为 P1 远大于 P2, 所以判断秃顶与患心脏病有关.

(2)由题意知, chi^2 = (1350 * ((200 * 600 - 150 * 400)^2) / (350 * 1000 * 600 * 750)) = 216/7 approx 30.86 > 6.635, 所以能够以 99% 的把握认为秃顶与患心脏病有关.

20.解:(1)由表中数据可得, x_bar = 1/5 * (1+2+3+4+5) = 3, y_bar = 1/5 * (10+11+14+25+20) = 16,

则 r = (sum((x_i - x_bar)(y_i - y_bar)) / (sqrt(sum((x_i - x_bar)^2) * sum((y_i - y_bar)^2)))) = (17 / sqrt((4+1+0+1+4) * (36+25+4+81+16))) = 17 / (9 * sqrt(5)) approx 0.84, 故相关系数 r 的大小约为 0.84, 因为 r 接近于 1, 所以收益 y 与总投资 x 的线性相关程度较强.

(2)由题意知, 2x2 列联表如下:

Contingency table with columns: 满意度, 满意, 不满意, 总计; rows: 性别, 男, 女, 总计. Values: 男(54, 18, 72), 女(36, 42, 78), 总计(90, 60, 150).

因为 chi^2 = (150 * ((54 * 42 - 18 * 36)^2) / (90 * 60 * 72 * 78)) approx 12.981 > 6.635, 所以有 99% 的把握认为消费者满意度与性别有关.

21.解:(1)由频率分布直方图可知, 10x * (0.006 + 0.008 + a + 0.026 + 0.042) = 1, 解得 a=0.018, 所以平均分的估计值为 0.08 * 55 + 0.26 * 65 + 0.42 * 75 + 0.18 * 85 + 0.06 * 95 = 73.8, 故受奖励的分数的估计值为 73.8 分.

(2)(i)由题意知, “良好”的人数为 10x * (0.018 + 0.006) * 100 = 24, 则列联表如下:

Contingency table with columns: 良好, 不良好, 合计; rows: 男, 女, 合计. Values: 男(8, 40, 48), 女(16, 36, 52), 合计(24, 76, 100).

(ii)由列联表得 chi^2 = (100 * ((8 * 36 - 16 * 40)^2) / (24 * 76 * 48 * 52)) approx 2.72 < 3.841, 所以没有 95% 的把握认为参赛学生的成绩是否良好与性别有关.

22.解:(1)根据题意可得性别与 A、B 组的 2x2 列联表为

Contingency table with columns: A, B, 合计; rows: 男, 女, 合计. Values: 男(18, 32, 50), 女(30, 20, 50), 合计(48, 52, 100).

所以 chi^2 = (100 * ((18 * 20 - 32 * 30)^2) / (48 * 52 * 50 * 50)) approx 5.769 > 3.841, 所以能有 95% 的把握认为 A 组用户与性别有关.

(2)因为女性用户中 A、B 组的人数比为 3:2, 所以从调查的女性用户中按分层随机抽样的方法选出 5 人, 其中 A 组抽 3 人, B 组抽 2 人, 所以再从这 5 人中抽 3 人共有 C3=10 种结果,

而这 3 人中有 2 人在 A 组包含 C2 * C1=6 种结果, 所以 3 人中有 2 人在 A 组的概率为 6/10 = 3/5.