040100025

概率与统计

Prob. & Stat. 课程组 最后编辑

日期：2022年4月6日

内容

1 第二章

2.1

1. 四所大学--1、2、3、4正在参加一个假日篮球赛。在第一轮中，1号将与2号比赛，3号将与4号比赛。然后，两个赢家将为冠军而战，两个输家也将出战。一个可能的结果可以用1324表示（在第一轮比赛中，1击败2，3击败4，然后1击败3，2击败4）。

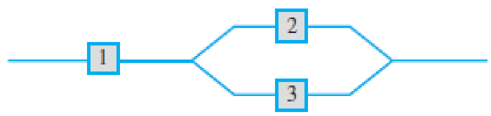
    a. 列出S的所有结果。

    b. 让A表示1赢得比赛的事件。列出A中的结果。

    c. 让B表示2进入冠军赛的事件。列出B中的结果。

    d. $A \cup B$和$A \cap B$中的结果是什么？A的结果是什么？′

2. 如附图所示，三个元件连接起来形成一个系统。由于2-3子系统中的组件是并联的，如果两个单独的组件中至少有一个发挥作用，该子系统就会发挥作用。为了使整个系统发挥作用，部件1必须发挥作用，2-3子系统也必须发挥作用。实验包括确定每个组件的状况[S（成功）代表功能正常的组件，F（失败）代表功能不正常的组件]。



    a. 在事件A中包含哪些结果，即三个组成部分中正好有两个发挥作用？

    b. 事件B中包含哪些结果，即至少有两个组件发挥作用？

    c. 哪些结果包含在系统功能的事件C中？

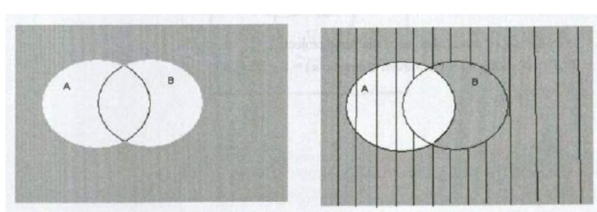
    d. 在C中列出结果′，$A \cup C$，$A \cap C$，$B \cup C$和$B \cap C$。

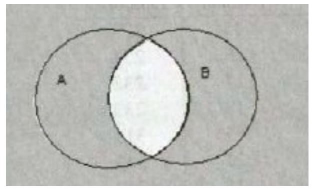
3. 使用维恩图来验证任何事件A和B的以下两种关系（这些被称为德摩根定律）：

    a. $(a \cup b)' = a' \cap b'$

    b. $(a \cap b)' = a' \cup b'$

    [提示：在每个部分，画一个与左边相对应的图，另一个与右边相对应的图。]





2.2

4. 让A表示下一次向统计软件顾问请求帮助的事件与SPSS软件包有关，让B表示下一次请求帮助SAS的事件。假设$P(A)=0.30$，$P(B)=0.50$。
   a. 为什么不是$P(A)+P(B)=1$的情况？
   b. 计算$P(A')$。
   c. 计算$P(A \cup B)$。
   d. 计算$P(A' \cap B')$。

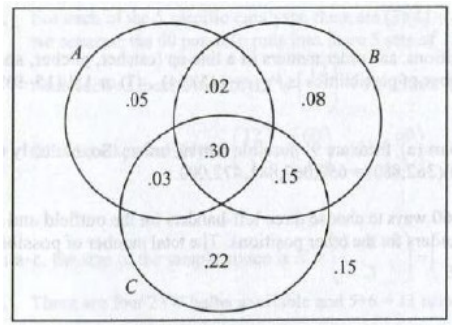
5. 人类对印刷电路板上的焊点进行目测可能是非常主观的。部分问题源于众多类型的焊接缺陷（例如，焊盘不湿润、膝部可见、空洞），甚至一个焊点拥有一个或多个这些缺陷的程度。因此，即使是训练有素的检查员也会对某一特定接头的处置产生分歧。在一批10,000个接头中，检查员A发现了724个被判定为有缺陷的接头，检查员B发现了751个这样的接头，其中1159个接头被至少一个检查员判定为有缺陷。假设这10,000个接头中的一个被随机抽取。
   a. 被选中的接头被两个检查员中的任何一个判定为缺陷的概率是多少？
   b. 被选中的接头被检查员B判断为有缺陷，而被检查员A判断为没有缺陷的概率是多少？

6. 某一类型的新车最受欢迎的三个选项是内置GPS（A），天窗（B），和自动变速器（C）。如果40%的购买者要求A，55%要求B，70%要求C，63%要求A或B，77%要求A或C，80%要求B或C，85%要求A或B或C，请确定以下事件的概率。

   [提示："A或B"是指两个选项中至少有一个被要求的事件；试着画一个维恩图，并给所有区域贴上标签。]
   a. 下一个购买者将要求在三个选项中至少有一个。
   b. 下一个购买者将不选择这三个选项。
   c. 下一个购买者将只要求自动变速器，而不是其他两个选项中的任何一个。
   d. 下一个购买者将正好选择这三个选项中的一个。

2.3

7. 作曲家贝多芬写了9部交响曲、5部钢琴协奏曲（钢琴和管弦乐队的音乐）和32部钢琴奏鸣曲（钢琴独奏的音乐）。
   a. 有多少种方法可以先演奏贝多芬交响曲，然后再演奏贝多芬钢琴协奏曲呢？



b. 一家广播电台的经理决定在每个连续的晚上（每周7天）播放一首贝多芬交响曲，然后是贝多芬钢琴协奏曲，接着是贝多芬钢琴奏鸣曲。在完全相同的节目必须重复之前，这个政策可以持续多少年？

8. 一家生产工厂有10名白班工人，8名倒班工人，6名晚班工人。一位质量控制顾问将从这些工人中选择5人进行深入访谈。假设选择的方式是，任何一组5名工人被选中的机会与其他任何一组相同（从24人中抽出5人，不作替换）。

   a. 有多少次选择的结果是所有5名工人都来自白班？所有5名被选中的工人都来自白班的概率是多少？
   b. 所有5名被选中的工人来自同一班次的概率是多少？
   c. 所选工人中至少有两个不同班次的代表，其概率是多少？
   d. 在工人的样本中，至少有一个班次没有代表的概率是多少？

9. 一个供应室的箱子里有15个紧凑型荧光灯泡，其中5个额定功率为13瓦，6个额定功率为18瓦，还有4个额定功率为23瓦。假设这些灯泡中有三个是随机选择的。

   a. 所选灯泡中正好有两个额定功率为23瓦的概率是多少？
   b. 所有三个灯泡都有相同等级的概率是多少？
   c. 每种类型的灯泡被选中的概率是多少？
   d. 如果逐个选择灯泡，直到得到一个23瓦的灯泡，那么至少需要检查6个灯泡的概率是多少？

10. 在五张牌的扑克游戏中，顺子由五张相邻面值的牌组成（例如，梅花9，红桃10，红桃J，黑桃Q和梅花K）。假设A可以是大牌也可以是小牌，如果你发了一手五张牌，那么大牌10是顺子的概率是多少？它是顺子的概率是多少？同花顺的概率是多少（所有的牌都是一样的）？

2.4

11. 某一国家的人口由三个民族组成。每个人都属于四个主要血型之一。所附的*联合概率表*给出了不同族群-血型组合中的个人比例。

		血型			
		O	**A**	**B**	**AB**
	1	.082	.106	.008	.004
族裔群体	**2**	.135	.141	.018	.006
	3	.215	.200	.065	.020

假设从人口中随机选出一个个体，用A={被选中的A类型}、B={被选中的B类型}和C={被选中的第3种族群体}定义事件。

a. 计算$P(A)$，$P(C)$，和$P(A\cap C)$。

b. 计算$P(A\ C)$和$P(C\ A)$，并在上下文中解释这些概率分别代表什么。

c. 如果被选中的人没有B型血，那么他或她来自1族的概率是多少？

12. 根据2013年7月31日cnn.com上的一篇帖子，在一个咬到花生的孩子死亡后，*《儿科》*杂志2010年的一项研究发现，美国18岁以下的儿童中有8%至少有一种食物过敏。在有食物过敏的人中，大约39%有严重的反应史。

    a. 如果随机选择一个18岁以下的儿童，他或她至少有一种食物过敏并有严重反应史的概率是多少？

    b. 还有报告说，事实上有30%的过敏者对多种食物过敏。如果随机选择一个18岁以下的孩子，他或她对多种食物过敏的概率是多少？

13. 鹿蜱可以是莱姆病或人类粒细胞性埃希氏病（HGE）的携带者。根据最近的一项研究，假设某一地点的所有蜱虫中有16%携带莱姆病，10%携带HGE，而至少携带其中一种疾病的蜱虫中有10%实际上同时携带这两种疾病。如果随机选择的一只蜱虫被发现携带HGE，那么所选蜱虫也是莱姆病携带者的概率是多少？

14. 如果$P(B|A) > P(B)$，表明$P(B'|A) < P(B')$。

15. 某种类型的部件以10个为一个批次运给供应商。假设所有这些批次中的50%不包含有缺陷的部件，30%包含一个有缺陷的部件，20%包含两个有缺陷的部件。从一个批次中随机选择两个部件进行测试。在下列各条件下，该批产品中出现0、1、2个缺陷部件的概率是多少？

    a. 两个被测试的部件都没有缺陷。

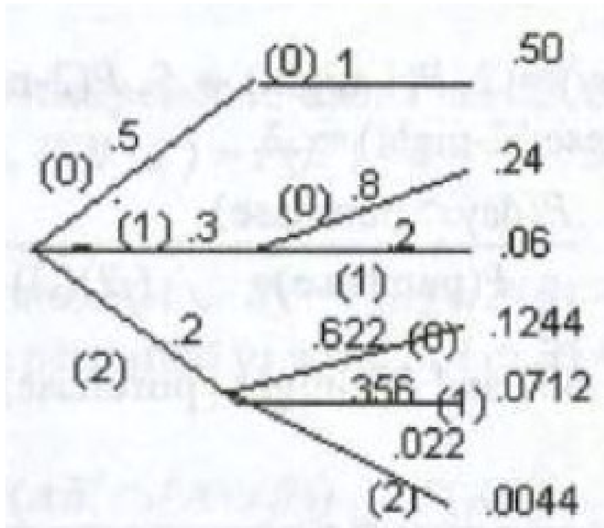
    b. 两个被测试的部件中，有一个是有缺陷的。[提示：为三种不同类型的批次画一个有三个第一代分支的树状图。]

2.5

16. 一个飞机接缝需要25个铆钉。如果这些铆钉中的任何一个有缺陷，该缝就必须返工。假设铆钉的缺陷是相互独立的，每个铆钉都有相同的概率。

    a. 如果所有接缝中有15%需要返工，那么铆钉有缺陷的概率是多少？

    b. 一个有缺陷的铆钉的概率应该有多小，才能确保所有接缝中只有10%需要返工？



17. 到达分销商的部件由两个不同的检查员进行缺陷检查（每个部件由两个检查员检查）。第一个检查员检测出所有存在缺陷的90%，第二个检查员也是如此。至少有一位检查员没有发现所有缺陷部件中20%的缺陷。发生以下情况的概率是多少？

    a. 一个有缺陷的部件只会被第一个检查员发现？恰恰是由两个检查员中的一个发现？

    b. 一批产品中所有三个有缺陷的部件都能逃脱两个检查员的检测（假设对不同部件的检查是相互独立的）？

18. 假设在一只狐狸的左耳和右耳都贴上相同的标签。然后将狐狸放生一段时间。考虑两个事件C_1 = 左耳标签丢失，C_2 = 右耳标签丢失。设$\vartheta=P(C_1)=P(C_2)$，并假设C_1和C_2是独立事件。假设C 和C 是独立的事件，请推导出正好丢失一个标签的概率的表达式（涉及ϑ），因为最多丢失一个标签（"红狐的耳标丢失"，*J. Wildlife Mgmt.*）

补充练习

19. 一颗卫星计划从佛罗里达州的卡纳维拉尔角发射，另一颗卫星计划在加利福尼亚的范登堡空军基地发射。让A表示范登堡的发射如期进行的事件，让B代表卡纳维拉尔角的发射如期进行的事件。如果A和B是独立事件，$P(A)>P(B)$，$P(A\cup B)=0.626$，$P(A\cap B)=0.144$，请确定$P(A)$和$P(B)$的值。

20. 个人A有一个由五个亲密朋友组成的圈子（B、C、D、E和F）。A从圈外听到了某个谣言，并邀请这五个朋友参加一个聚会来传播这个谣言。首先，A从这五个人中随机选择一个，并把这个谣言告诉被选中的人。然后，这个人从剩下的四个人中随机选择一个，重复这个谣言。继续下去，从那些还没有听过谣言的人中选出一个新的人，由刚刚听过谣言的人告诉他们，直到每个人都被告知。

    a. 谣言按照B、C、D、E、F的顺序重复出现的概率是多少？

    b. F是聚会上第三个被告知该谣言的人的概率是多少？

    c. F是最后一个听到这个传言的人的概率是多少？

d. 如果在每个阶段，目前 "拥有 "谣言的人不知道谁已经听到了谣言，并从所有五个可能的人中随机选择下一个接受者，那么在聚会上讲了十遍之后，F仍然没有听到谣言的概率是多少？

21. 不考虑2月29日生日的可能性，假设一个随机选择的人同样可能在其他365天中的任何一天出生。

a. 如果随机选择十个人，所有的人都有不同的生日，其概率是多少？至少有两个人的生日是相同的？

b. 在(a)部分中，用k代替10，那么最小的k是什么，它至少有50％的机会使两个或两个以上的人有相同的生日？

c. 如果随机选择十个人，那么至少有两个人的生日是相同的，或者至少有两个人的社会安全号码的最后三位数是相同的，这种概率是多少？[注："研究巧合的方法 "一文（F. Mosteller和P. Diaconis, *J. Amer. Stat. Assoc.*, 1989: 853-861）讨论了这种类型的问题。]

22. 一个盒子里有以下四张纸条，每张纸条的尺寸完全相同：（1）中奖1；（2）中奖2；（3）中奖3；（4）中奖1、2和3。将随机抽取一张纸条。设A_1 = 中奖1，A_2 = 中奖2，A_3 = 中奖3。证明A_1 和A_2 是独立的，A_1 和A_3 是独立的，A_2 和A_3 也是独立的（这就是成对的独立性）。然而，表明$P(A_1 A_2 A_3) = P(A_1)P(A_2)P(A_3)$，所以这三个事件不是相互独立的。[注：$A_1$ =抽签1或4。]

∩ ∩ ⁄ -

2 第三章

3.1

1. **7.**对于这里定义的每个随机变量，请描述该变量的可能值集，并说明该变量是否是离散的。
   a. X = 随机选择的标准鸡蛋盒中未破损的鸡蛋数量
   b. Y = 某一课程的班级名单上，在开课第一天缺席的学生人数
   c. U = 一个球手在击球前必须挥杆的次数
   d. X = 随机选择的响尾蛇的长度
   e. Z = 随机选择的amazon.com购物的销售税百分比
   f. Y = 随机选择的土壤样本的pH值
   g. X = 随机选择的网球拍的张力（psi）。
   h. X = 三个网球运动员必须旋转球拍以获得UUU或DDD以外的东西的总次数（以决定哪两个人接下来比赛）。

2. **9.**一个名叫克劳狄斯的人位于附图中的0点（图1）。使用适当的随机化设备（如四面体骰子，一个有四个面的骰子），Claudius首先移动到四个地点之一B_1, B_2, B_3, B_4。一旦到了这些地点之一，另一个随机化设备被用来决定Claudius接下来是回到0还是访问另外两个相邻点之一。然后这个过程继续进行；在每次移动之后，通过抛出适当的骰子或硬币来决定向（新）相邻点之一的另一次移动。
   a. 让X=克劳狄斯在首次返回0之前所走的棋步数。X的可能值是多少？X是离散的还是连续的？
   b. 如果也允许沿着连接0到A_1、A_2、A_3 和A_4 的对角线路径移动，请回答（a）部分的问题。

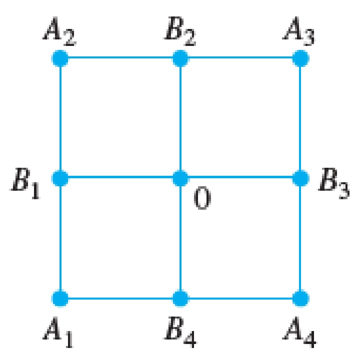


图1：问题9

3.2

3. **17.**一个新电池的电压可能是可接受的（A）或不可接受的（U）。某支手电筒需要两节电池，所以电池将被独立选择和测试，直到有两节可接受的电池为止。

发现。假设90%的电池都有可接受的电压。让Y表示必须测试的电池的数量。

a. 什么是$p(2)$，即$P(Y=2)$？

b. 什么是$p(3)$？[提示：有两种不同的结果会导致$Y=3$]。

c. 为了使$Y=5$，所选的第五个电池必须是什么？列出四个结果，其中$Y=5$，然后确定$p(5)$。

d. 利用你在(a)-(c)部分答案中的模式，得到$p(y)$的一般公式。

4. **23(22).**某银行在纽约市的一家分行有6台自动取款机。让X代表一天中某一时刻使用的机器的数量。X的cdf如下：

$$F(x) = \begin{cases} 0 & x < 0 \\ .06 & 0 \le x < 1 \\ .19 & 1 \le x < 2 \\ .39 & 2 \le x < 3 \\ .67 & 3 \le x < 4 \\ .92 & 4 \le x < 5 \\ .97 & 5 \le x < 6 \\ 1 & 6 \le x \end{cases}$$

直接从cdf中计算出以下概率：

a. $p(2)$，即$P(X=2)$

b. $P(X > 3)$

c. $P(2 \le X \le 5)$

d. $P(2 < X < 5)$

5. **25(24).**在例3.12中，让Y=实验结束前出生的女孩的数量。在$p=P(B)$和$1p=P(G)$的情况下，Y的pmf是多少？[提示：首先列出Y的可能值，从最小的开始，然后继续，直到你看到一个一般的公式。]

3.3

6. **29(28).**例子中给出了所购U盘中内存X(GB)的pmf。
3.13为

x	1	2	4	8	16
$p(x)$	.05	.10	.35	.40	.10

计算如下：

a. $E(X)$

b. $V(X)$直接来自定义

c. X的标准偏差

d. $V(X)$使用快捷的公式

7. **33(32).** Let X be a Bemoulli rv with pmf as in Example 3.18.

a. 计算$E\,X^2$。

b. 证明$V(X)=p(1-p)$。

c. 计算$E\,X^9$。

8. **37(36).**一份工作的n个候选人的排名是1，2，3，......。让X=随机选择的候选人的等级，所以X有pmf

$$p(x) = \begin{cases} 1/n & x = 1, 2, 3, \ldots, n \\ 0 & \text{否则} \end{cases}$$

(这被称为离散均匀分布)。使用捷径公式计算$E(X)$和$V(X)$。[提示：前n个正整数之和为$n(n+1)/2$，而它们的平方之和为$n(n+1)(2n+1)/6$。]

9. **41(40).**利用表达式（3.13）中的定义，证明$V(aX+b)=a^2\cdot\sigma_X^2$。[提示：用$h(X) = aX + b$ $E[h(X)] = a\mu + b$ 其中$\mu = E(X)$]。

3.4

10. **63(57).**

a. 证明$b(x; n, 1 - p) = b(n - x, n, p)$。

b. 证明$B(x; n, 1 - p) = 1 - B(n - x - 1; n, p)$。[提示：最多$x$个S相当于至少（$n - x$）F's]。

c. (a)和(b)部分暗示了在附录表A.1中包括大于0.5的p值的必要性是什么？

11. **67(61).**请参考练习44中给出的切比雪夫不等式。计算$P(|X - \mu| \geq k\sigma)$为$k = 2$和$k=3$时，$X \sim$ Bin(20, .5)，并与相应的上界进行比较。对X重复一次$\sim$ Bin(20, .75).

3.5

12. **73(67).**20对参加桥牌比赛的人被列为种子选手1, . . ., 20.在比赛的第一部分，这20人被随机分为10对东西向和10对南北向。

a. 前10对组合中，有x个组合最终向东向西比赛的概率是多少？

b. 所有前五名的对子最后打出相同方向的概率是多少？

c. 如果有$2n$对组合，X的pmf=前n对组合中最终向东向西比赛的人数是多少？什么是$E(X)$和$V(X)$？

13. **77(71).**三个兄弟和他们的妻子决定生孩子，直到每个家庭有两个女孩子。X=兄弟俩所生男孩子总数的pmf是多少？什么是$E(X)$，它与每个兄弟所生男孩子的预期数量相比如何？

3.6

14. **91(85).**假设树木按照二维泊松过程在森林中分布，参数α（每英亩的预期树木数）等于80。

a. 在某一四分之一英亩的土地上，最多有16棵树的概率是多少？

b. 如果森林占地85,000英亩，预计森林中的树木数量是多少？

c. 假设你在森林中选择一个点，构建一个半径为0.1英里的圆。让X=该圆形区域内树木的数量。X的pmf是多少？[提示：1平方英里=640英亩]。

15. **93(87).**

a. 在泊松过程中，在时间区间（0，t）和区间（t，$t+\Delta t$）中必须发生什么，才能使整个区间（0，$t+\Delta t$）中没有事件发生？利用这一点和假设1-3，写出$P_0(t+\Delta t)$和$P_0(t)$之间的关系。

b. 用(a)部分的结果写出差值$P_0(t+\Delta t)$ $P_0(t)$的表达式。然后除以Δt，让Δt为0，得到一个涉及(d/dt)$P_0(t)$的方程，即$P_0(t)$相对于t的导数。

c. 验证$P_0(t) = e^{-\alpha t}$ 满足(b)部分的方程式。

d. 可以用类似于(a)和(b)部分的方式证明，$P_k(t)$ s必须满足微分方程组的要求

$$\frac{d}{dt}P_k(t) = \alpha P_{k-1}(t) - \alpha P_k(t)$$

$$k = 1, 2, 3, \ldots$$

验证$P_k(t) = e^{-\alpha t}(\alpha t)^k$ /k！满足该系统。(这实际上是唯一的解。)

补充练习

16. **99(93).**一个*k-out-of-n*系统是指当且仅当系统中的*n*个单独组件中至少有*k*个组件发挥作用时，该系统才会发挥作用。如果各个部件独立运作，每个部件的概率为0.9，那么5个部件中的3个部件运作的概率是多少？

17. **105(99).**一台发电机组的购买者要求在接受该机组之前连续成功启动*c*次。假设单个启动的结果是相互独立的。让*p*表示任何特定启动成功的概率。所关注的随机变量是*X*=在接受之前必须进行的创业次数。如果p=.9，P（X 8 ）是多少？[提示：对于x 5 ，用在小值x 3、x 4、......的pmf值来 "递归 "表达$p(x)$。，2.](这个问题是由 "启动示范测试的评估 "一文提出的，J.质量技术，1983：103-106）。

18. **113(107).**一项关于某种疾病存在的测试，给出假阳性读数的概率为0.20（表明一个人患有这种疾病，但事实并非如此），给出假阴性结果的概率为.10个假阴性的结果。假设有十个人接受测试，其中五个人患有该疾病，五个人没有。设*X*=结果为阳性的读数。

a. *X*有二项分布吗？解释一下你的推理。

b. 十个测试结果中正好有三个是阳性的概率是多少？

19. **117(111).**一个计算机磁盘存储设备有十个同心轨道，从最外面到最里面编号为1, 2, ...10，从最外面到最里面，还有一个访问臂。让p_i = 任何特定的数据请求将使存取臂到达轨道i（$i = 1, \ldots, 10$）的概率。假设在连续的寻求中访问的轨道是独立的。让*X*=访问臂在两个连续请求中经过的轨道的数量（不包括访问臂刚刚离开的轨道，所以可能的*X*值是x=0，1，......, 9).计算*X*的pmf。 [提示：P（机械臂现在在轨道i上且$X=j$）=P（X=jlarm nowon i）p_i。在条件概率写成p_1，.. ., p_{10}，根据总概率定律，通过对i进行求和，就可以得到所需的概率。]

20. **119(113).**利用以下事实

$$\sum_{\text{所}} (x-\mu)^2\,p(x) \ge \sum_{x:|x-\mu|=k\sigma} (x-\mu)^2\,p(x)$$

来证明练习44中给出的切比雪夫的不等式。

3 第四章

4.1

1. **5.**一位大学教授从不在一小时内结束他的讲座，而且总是在一小时后的2分钟内结束他的讲座。设X=从一小时结束到讲座结束之间的时间，假设X的pdf为

$$f(x) = \begin{cases} kx^2 & 0 \le x \le 2 \\ 0 & 否则 \end{cases}$$

a. 找出k的值并画出相应的密度曲线。[提示："K "的图形下的总面积是多少？
$f(x)$为1]。

b. 讲座在一小时结束后1分钟内结束的概率是多少？

c. 讲座超过一小时后继续进行60至90秒的概率是多少？

d. 讲座在一小时结束后继续进行至少90秒的概率是多少？

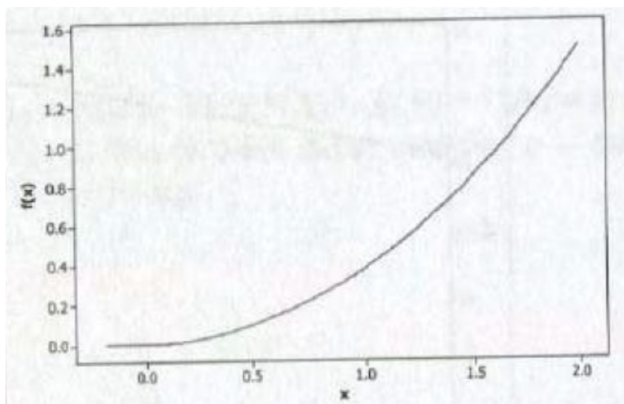


图2：问题5

4.2

2. **15.**让X表示放置在一个1 英尺^3 的包装容器中的物品所占据的空间量。X的pdf值为

$$f(x) = \begin{cases} 90x^8(1-x) & 0 < x < 1 \\ 0 & 否则 \end{cases}$$

a. 绘制PDF的图形。然后得到X的cdf并绘制成图。

b. 什么是P（$X \le .5$）[即F（.5）]？

c. 使用(a)中的cdf，什么是$P(.25 < X \le .5)$？P（$.25 \le X \le .5$）是多少？

d. 分布中的第75个百分点是多少？

e. 计算$E(X)$和σ_X。

f. X离其平均值超过1个标准差的概率是多少？

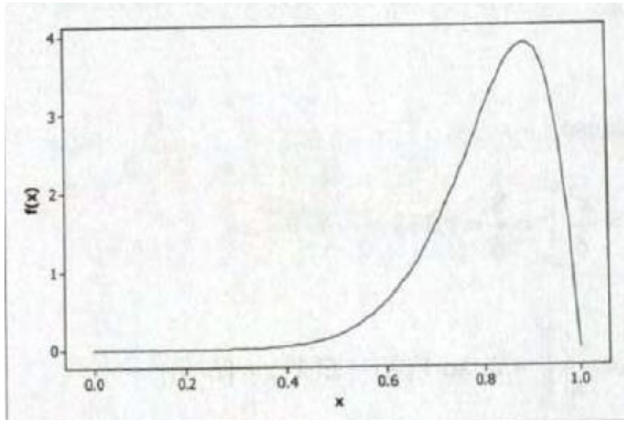


图3：问题15：pdf

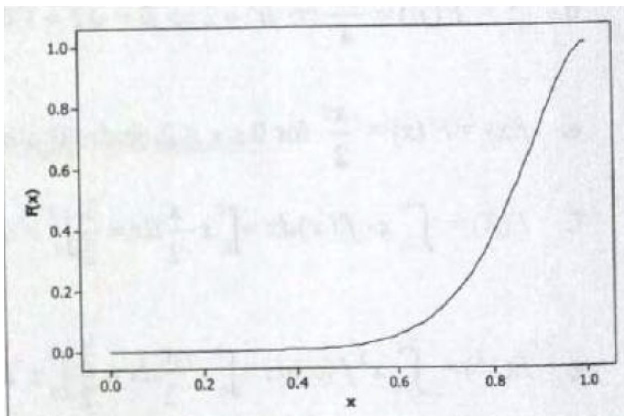


图4：问题15：cdf

3. **21.**一位生态学家希望标出一个半径为10米的圆形取样区域。然而，所产生的区域的半径实际上是一个随机变量R，其pdf

$$f(r) = \begin{cases} \frac{3}{4}\left[1 - (10 - r)^2\right] & 9 \leq r \leq 11 \\ 0 & \text{否则} \end{cases}$$

   所得圆形区域的预期面积是多少？

4.3

4. **31(29).**对于以下的α值，确定z_α：
   a. $\alpha = .0055$
   b. $\alpha = .09$
   c. $\alpha = .663$
5. **51(45).**切比雪夫不等式，（见第三章练习44），对连续分布和离散分布都有效。它指出，对于任何满足$k \geq 1$的数字，$P(|X - \mu| \geq k\sigma) \leq 1/k^2$（见第三章练习44的解释）。在$k=1$、2、3的正态分布情况下，获得这个概率，并与上界进行比较。

6. **57(51).**

   a. 证明如果X具有参数为μ和σ的正态分布，那么$Y=aX+b$（X的线性函数）也具有正态分布。Y的分布参数[即E（Y）和V（Y）]是什么？[提示：将Y的cdf，P（$Y y$ ）写成涉及X的pdf的积分 ，然后相对于y进行微分，得到Y的pdf。] $\leq$

   b. 如果当以°C测量时，温度是正态分布，平均数为115，标准差为2，那么以°F测量的温度的分布可以说是什么？

4.4

7. **59(58).**设X=两个连续到达当地银行开车窗口的时间。如果X具有$\lambda=1$的指数分布（这与$\alpha=1$的标准伽马分布相同），请计算如下：

   a. 两次连续到达之间的预期时间

   b. 连续到达的时间间隔的标准偏差

   c. $P(X \leq 4)$

   d. $P(2 \leq X \leq 5)$

8. **69(62).** 一个系统由五个相同的部件串联而成，如图所示（图5）：如

图5： 问题69

只要有一个部件失效，整个系统就会失效。假设每个组件的寿命是指数分布的，$\lambda=0.01$，并且组件的故障是相互独立的。定义事件A_i = {其中一个组件至少持续t小时，$i = 1, . . ., 5$，所以A_i s是独立的事件。让X=系统失效的时间--即五个组件中最短（最小）的寿命。}

   a. 事件{X≥t}等同于什么事件，涉及$A_1, . . ., A_5$？

   b. 利用A_i 's的独立性，计算P（$X t$）。然后得到$F(t)\leq P(X t)$和X的pdf。X的分布类型是什么？

   c. 假设有n个组件，每个组件都有参数为λ的指数寿命，那么X的分布类型是什么？

9. **71(65).**

   a. 事件$X^2 \leq y$等同于涉及X本身的什么事件？}

   b. 如果X具有标准正态分布，使用(a)部分来写出等于$P X^2 y$的积分。然后对y进行微分，得到X^2 [N（0，1）变量的平方]的pdf。最后，证明X^2 具有$v = 1$ df的齐次分布[见(4.10)]。[提示：使用下面的方法身份] ≤。

$$\frac{d}{\text{染料}}\left(\int_{a(y)}^{b(y)} f(x)dx\right) = f[b(y)] - b'(y) - f[a(y)] - a'(y)$$

4.5

10. **75(69).** 让X具有Weibull分布，其pdf来自表达式（4.11）。验证一下，$\mu = \beta\Gamma(1+1/\alpha)$。[提示：在$E(X)$的积分中，使变量$y = (x/\beta)^{\alpha}$，所以$x = \beta y^{1/\alpha}$。]

11. **85(77).**让X有一个标准的β密度，参数为α和β。
    a. 验证本节中给出的$E(X)$的公式。
    b. 计算$E[(1-X)^m]$。如果X代表一种物质中由某种成分组成的比例，那么不由这种成分组成的预期比例是多少？

补充练习

12. **99(89).**一根两端都被夹住的12英寸的钢筋要承受越来越大的应力，直到它断裂。假设Y=发生断裂时与左端的距离。假设Y有pdf

$$f(y) = \begin{cases} \frac{1}{24}y\left(1-\frac{y}{12}\right) & 0 \le y \le 12 \\ 0 & \text{否则} \end{cases}$$

计算如下：
    a. Y的cdf，并画出它的图形。
    b. $P(Y \le 4)$, $P(Y > 6)$, and $P(4 \le Y \le 6)$
    c. $E(Y)$, $E(Y^2)$，和 $V(Y)$
    d. 断裂点出现在离预期断裂点2英寸以上的概率。
    e. 断裂发生时，较短段的预期长度。

13. **101(91).**某项任务的完成时间X的cdf $F(x)$由以下公式给出

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^3}{3} & 0 \le x < 1 \\ 1-\frac{1}{2}\left(\frac{7}{3}-x\right)\left(\frac{7}{4}-\frac{3}{4}x\right) & 1 \le x \le \frac{7}{3} \\ 1 & x > \frac{7}{3} \end{cases}$$

    a. 得到pdf $f(x)$并画出其图形。
    b. 计算P（$.5 \le X \le 2$）。
    c. 计算$E(X)$。

14. **107(97).**让X表示发生某种化学反应的温度。假设X有pdf

$$f(x) = \begin{cases} \frac{1}{9}(4-x^2) & -1 \le x \le 2 \\ 0 & \text{否则} \end{cases}$$

    a. 画出$f(x)$的图形。
    b. 确定cdf并画出它的草图。
    c. 0是发生反应的中位温度吗？如果不是，中位温度是比0小还是大？
    d. 假设该反应在十个不同的实验室中各独立进行一次，每个实验室中反应时间的pdf值如实给出。设Y=10个实验室中温度超过1的数量。(给出任何参数的名称和值）。

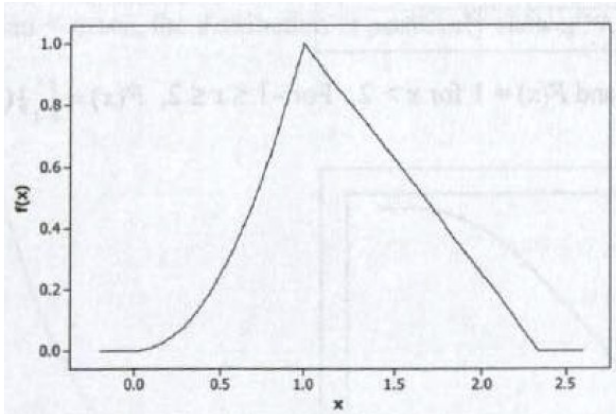


图6：问题101

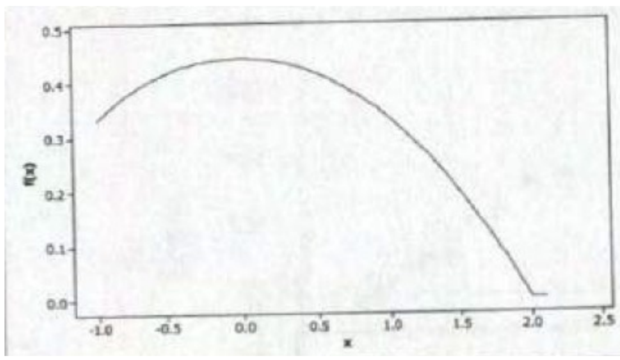


图7：问题107

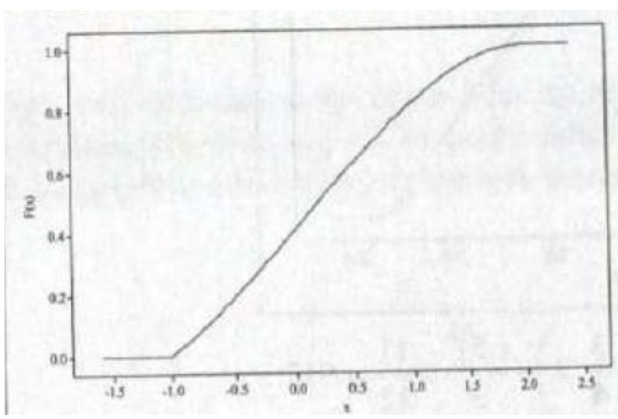


图8：问题107

15. **109(99).**文章"通过对腐蚀剖面的统计分析来预测腐蚀"（腐蚀科学，1985年：305-315）建议，在涉及碳锰钢暴露于酸化海水的实验中，最深坑的深度X的cdf如下。

$$F(x, \alpha, \beta) = e^{-e^{(x-\alpha/\beta)}} \; -\infty < x < \infty$$

作者提出了α=150和β=90的数值。假设这就是正确的模型。

a. 最深的坑的深度最多为150的概率是多少？至多300？在150和300之间？

b. 在90%的此类实验中，最大坑的深度将低于什么值？

c. X的密度函数是什么？

d. 密度函数可以显示为单模态（一个单峰）。这个峰值出现在测量轴上的哪个值之上？(这个值是模式。) e. 可以证明，$E(X)$ ≈ .5772β+α。对于给定的α和β的值，平均数是多少，它与中位数和模式的比较如何？画出密度函数的图形。[注：这被称为最大极值分布。]

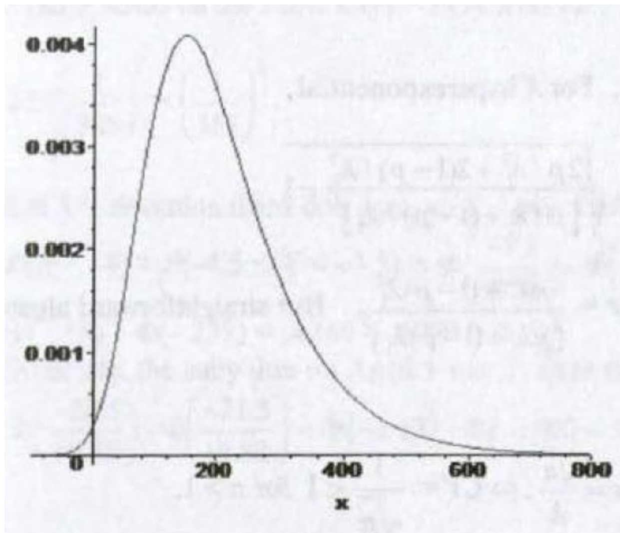


图9：问题109

16. **111(101).**连续分布的模式是使$f(x)$最大化的值x^*。
    a. 参数为μ和σ的正态分布的模式是什么？
    b. 具有参数A和B的均匀分布是否具有单一模式？为什么或为什么不呢？
    c. 什么是参数为λ的指数分布的模式？(画个图）。
    d. 如果X有一个参数为α和β的伽马分布，且$\alpha>1$，请找出模式。[提示：ln[$f(x)$]。如果$f(x)$是最大化的，那么取ln[$f(x)$]的导数可能更简单。]
    e. 有v个自由度的卡方分布的模式是什么？

17. **115(105).**让I_i是一个晶体管的输入电流，I_0是输出电流。那么电流增益与ln（I/I_{0i}）成正比。假设比例常数为1（相当于选择一个特定的测量单位），那么电流增益=X=ln（I/I_{0i}）。假设X是正态分布，μ=1，σ=0.05。
    a. I/I_{0i}的比例是什么类型的分布？
    b. 输出电流是输入电流的两倍以上的概率是多少？
    c. 输出与输入电流之比的预期值和方差是多少？

18. **117(107).**让Z有一个标准的正态分布，并通过$Y=\sigma Z+\mu$定义一个新的Rv Y，表明Y有一个参数为μ和σ的正态分布。利用这一点找到Y的cdf，然后对y进行微分] 。 ≤ ≤

19. **119(109).**在练习117和118中，以及其他许多情况下，人们有X的pdf $f(x)$，希望知道$y = h(X)$的pdf。假设$h()$是一个可逆函数，所以$y=h(x)$可以通过解x得到$x=k(y)$。那么可以证明，Y的pdf是

$$g(y) = f[k(y)] - |k'(y)|$$

   a. 如果X具有A=0和B=1的均匀分布，求出Y=-ln(X)的pdf。
   b. 用这个结果做习题117。
   c. 用这个结果做练习118(b)。

20. **123(111).**让U在区间[0，1]上有一个均匀分布。那么具有这种分布的观察值可以从计算机的随机数发生器中得到。让X=-（$1/\lambda$）ln（1-U）。
   a. 证明X具有参数为λ的指数分布 [提示：X的ODF为$F(x)$ = $P(X \simeq x)$; $X \le x$等同于$U \le$?]
   b. 你将如何使用(a)部分和随机数发生器从参数为λ=10的指数分布中获得观察值？

4 第五章

5.1

1. **3.**某一市场同时有一条快速结账线和一条超快速结账线。让X_1 表示一天中某一时刻在快速结账处排队的顾客人数，让X_2 表示同一时刻在特快结账处排队的顾客人数。假设X_1 和X_2 的联合pmf如附表所示。

	x_2 0	1	2	3
0	.08	.07	.04	.00
1	.06	.15	.05	.04
$x\ 2_1$	.05	.04	.10	.06
3	.00	.03	.04	.07
4	.00	.01	.05	.06

.

   a. $P(X_1 = 1, X_2 = 1)$是多少，即每条线上正好有一个顾客的概率？
   b. $P(X_1 = X_2)$是多少，即两条线的顾客人数相同的概率？
   c. 让A表示一条线上的顾客比另一条线上的顾客至少多两个的事件。用X_1 和X_2 来表示A，并计算这一事件的概率。
   d. 两条线上的顾客总数正好是四个的概率是多少？至少四个？

2. **5.**在一家百货公司等待礼品包装服务的顾客人数是一个rv X，可能的值是0，1，2，3，4和相应的概率.1，.2，.3，.25，.15。一个随机选择的顾客将有1、2或3个包裹需要包装，概率分别为.6、.3和.1。让Y=排队等候的顾客需要包装的包裹总数（假设一个顾客提交的包裹数量与其他顾客提交的数量无关）。

   a. 确定P（X=3，Y=3），即P（3，3）。
   b. 确定P（4，11）。
   c. 哪些结果包含在系统功能的事件C中？
   d. 在C中列出结果′，$A \cup C$，$A \cap C$，$B \cup C$和$B \cap C$。

3. **9.**某一类型车辆的每个前轮胎应该被填充到26psi的压力。假设每个轮胎的实际气压对右轮胎来说是一个随机变量X，对左轮胎来说是一个随机变量Y，其共同的pdf为

$$f(x, y) = \begin{cases} K(x^2 + y^2) & 20 \le x \le 30,\ 20 \le y \le 30 \\ 0 & \text{否则} \end{cases}$$

   a. K的值是多少？
   b. 两个轮胎都没装满的概率是多少？
   c. 两个轮胎之间的气压差最多为2psi的概率是多少？
   d. 单独确定右轮胎中的（边际）气压分布。
   e. X和Y是独立的RV吗？

4. **19.** 练习9中给出了左右前轮胎的联合压力pdf。

a. 确定在X=x的情况下Y的条件性pdf和在*X的情况*下X的条件性pdf。
Y=y。
b. 如果发现右轮胎的压力为22psi，那么左轮胎的压力至少为25psi的概率是多少？将此与P（Y≥25）进行比较。
c. 如果发现右轮胎的压力是22psi，那么左轮胎的预期压力是多少，这个轮胎的压力标准差是多少？

5.2

5. **23 在**练习3中，在快速收银台排队的顾客人数和在超级快速收银台排队的顾客人数之差为$X_1 - X_2$。计算预期差。

6. **31**
   a. 计算练习9中X和Y的协方差。
   b. 计算这个X和Y的相关系数ρ。

7. **35**
   a. 使用期望值的规则来证明$\mathrm{Cov}(aX + b, cY + d) = ac\,\mathrm{Cov}(X, Y)$。
   b. 使用(a)部分以及方差和标准差的规则来表明，当a和c具有相同的符号时，$\mathrm{Corr}(aX+b, cY+d)=\mathrm{Corr}(X, Y)$。
   c. 如果a和c的符号相反会怎样？

5.3

8. **37** 某品牌的洗碗机肥皂以三种规格出售：25盎司、40盎司和65盎司。20%的购买者选择25盎司的包装，50%选择40盎司的包装，其余30%选择65盎司的包装。让X_1 和X_2 表示由两个独立的购买者选择的包装尺寸。
   a. Determine the sampling distribution of $\bar{X}$, calculate $E(\bar{X})$, and compare to μ.
   b. 确定样本方差S的抽样分布^2，计算$E\ S^2$，并与之比较。
   σ^2。

9. **43** 假设某台机器分配的液体量是均匀分布的，下限A=8盎司，上限B=10盎司。请描述你将如何进行模拟实验，以比较样本量n=5、10、20和30时的（样本）第四分布的抽样分布。

5.4

10. **51** 随机选择的抵押贷款申请人填写某种表格所花的时间具有正态分布，其平均值为10分钟，标准差为2分钟。如果某天有5个人填写表格，另一天有6个人填写表格，那么每天的样本平均时间最多为11分钟的概率是多少？

11. **55(56)** 某城市在任何特定工作日开出的停车罚单数量具有泊松分布，参数为μ=50。
    a. 计算在某一天发出35到70张票的近似概率。[提示：当λ很大时，泊松rv近似于正态分布]。

b. 计算在为期5天的一周内发出的门票总数在225和275之间的大致概率。

c. 使用软件获得(a)和(b)中的确切概率，并与它们的近似值进行比较。

12. **57** 假设某大学的学生在某一项目上花费的时间X（以小时为单位）的分布是参数为α=50和β=2的gamma。因为α很大，可以证明X具有近似的正态分布。利用这一事实，计算一个随机选择的学生在该项目上最多花费125小时的近似概率。

5.5

13. **63** 参照练习三，a.

a. 计算X_1 = 快速结账的顾客数量和X_2 =之间的协方差。
在superexpress结账的客户数量。

b. 计算$V(X_1 + X_2)$。这与$V(X_1) + V(X_2)$相比如何？

14. **65** 假设当某种化合物的pH值为5.00时，随机选择的初学化学的学生测得的pH值是一个随机变量，平均值为5.00，标准差为0.2。将一大批该化合物进行细分，在上午的实验中给每个学生一个样本，在下午的实验中给每个学生一个样本。设$\bar{X}$=上午的学生测定的平均pH值，$\bar{Y}$ =下午的学生测定的平均pH值。

a. 如果pH值是一个正态变量，每个实验室有25名学生，请计算$P(-.1 \le \bar{X}-\bar{Y} \le .1)$。
[提示：$\bar{X}-\bar{Y}$是正态变量的线性组合，所以是正态分布。计算
$\mu_{\bar{X}-\bar{Y}}$ 和 $\sigma_{\bar{X}-\bar{Y}}$。

b. 如果每个实验室有36名学生，但pH值的测定不被假定为正常，请计算（大约）$P(-.1 \le \bar{X}-\bar{Y} \le .1)$。

15. **67** 一根PVC管要插在另一根管子里面。第一块的长度是正态分布，平均值为20英寸，标准差为0.5英寸。第二块的长度是一个正态分布，平均值和标准差分别为15英寸和0.4英寸。重叠量是正态分布，平均值为1英寸，标准偏差为0.1英寸。假设长度和重叠量是相互独立的，那么插入后的总长度在34.5英寸和35英寸之间的概率是多少？

补充练习

16. **79** 假设对于某个人来说，早餐的热量摄入是一个随机变量，预期值为500，标准差为50；午餐的热量摄入是随机的，预期值为900，标准差为100；晚餐的热量摄入是一个随机变量，预期值为2000，标准差为180。假设不同餐次的摄入量是相互独立的，那么在接下来的一年（365天）中，每天的平均卡路里摄入量最多是3500的概率是多少？
?[提示：让X_i，Y_i，Z_i表示第i天的三种卡路里摄入量。那么总的摄入量是由以下公式得出的
$\sigma(x_i + y_i + z_i)$。]

17. **81** 我们已经看到，如果$E(X_1) = E(X_2) = \cdots = E(X_n) = \mu$，那么$E(X_1 + \cdots + X_n) = n\mu$。在某些应用中，所考虑的$X_i$'s的数量不是一个固定的数字$n$，而是一个rv N。例如，让N=某一天进入维修车间的部件数量，让X_i表示第i个部件的维修车间时间。那么总的维修时间是$X_1 + X_2 + \cdots + X_N$，是一个随机数的随机变量之和。当N与X_i's无关时，可以证明：

$$E(X_1 + \cdots + X_N) = E(N) \cdot \mu$$

a. 如果在某一天预计送来的部件数量为10个，随机提交的部件的预计维修时间为40分钟，那么在任何一天提交的部件的预计总维修时间是多少？

b. 假设某种类型的部件按照泊松过程来维修，其速率为每小时5个。每个部件的预期缺陷数为3.5。在4小时的时间里，提交维修的部件的缺陷总数的预期值是多少？请务必指出你的答案是如何从刚才的一般结果中得出的。

18. **89(87)**

a. 使用线性组合方差的一般公式来写出$V(aX+Y)$的表达式。然后让$a = \sigma_Y/\sigma_X$，并说明$\rho \geq -1$。[提示：方差总是≥0，$\mathrm{Cov}(X, Y) = \sigma_X \cdot \sigma_Y \cdot \rho$。]

b. 通过考虑$V(aX-Y)$，得出结论：$\rho \leq 1$。

c. 利用$V(W)=0$，只有当W是常数时，才能说明$\rho=1$，只有当$Y=aX+b$时。

19. **93(91)** 从某一地区随机抽取一个岩石标本，在两个不同的时间段内进行称重。让W表示实际重量，X_1和X_2表示两次测量的重量。那么$X_1 = W + E_1$，$X_2 = W + E_2$，其中E_1和E_2是两个测量误差。假设E_i's相互独立，也独立于W，$V(E_1) = V(E_2) = \sigma^2$。

a. 用ρ表示，两个测量的权重X_1和X_2之间的相关系数^E，以σ^2_W，实际重量的方差，和σ^2_X，测量重量的方差。

b. 当$\sigma_W = 1$ kg和$\sigma_E = .01$ kg时计算ρ。

5 第六章

6.1

1. 例1.2中介绍了某一类型的混凝土梁的抗弯强度（MPa）的附带数据。

   5.9 7.2 7.3 6.3 8.1 6.8 7.0 7.6 6.8 6.5 7.0 6. 37.9 9.0
   8.2 8.7 7.8 9.7 7.4 7.7 9.7 7.8 7.7 11.6 11.3 11.8 10.7

   a. 计算所有以这种方式制造的梁的概念群体的强度平均值的点估计，并说明你使用的估计器。[提示：$\Sigma x_i = 219.8$。]
   b. 计算将所有此类梁中最弱的50%与最强的50%分开的强度值的点估计值，并说明你使用的估计器。
   c. Calculate and interpret a point estimate of the population standard deviation σ. Which estimator did you use? [Hint: $\Sigma x_i^2 = 1860.94$.
   d. 计算所有这些梁的比例的点估计，其抗弯强度超过了10MPa。[提示：如果一个观测值超过了10，就认为是 "成功"。]
   e. 计算人口变异系数σ/μ的点估计值，并说明你使用的估计器。

2. 考虑以下关于低粘度涂料涂层厚度的观察样本（"实现制造过程的目标值：一个案例研究，" J. of Quality Technology，1992: 22-26）：

   .83 .88 .88 1.04 1.09 1.12 1.29 1.31
   1.48 1.49 1.59 1.62 1.65 1.71 1.76 1.83

   假设涂层厚度的分布是正态的（正态概率图强烈支持这一假设）。

   a. 计算涂层厚度平均值的点估计值，并说明你使用的估计器。
   b. 计算涂层厚度分布的中位数的点估计值，并说明你使用的估计器。
   c. 计算一个点估计值，该值将厚度分布中所有数值中最大的10%与剩余的90%分开，并说明你使用的估计器。[提示：用μ和σ表示你要估计的东西。]
   d. 估计P（X<1.5），即所有厚度值小于1.5的比例。[提示：如果你知道μ和σ的值，你可以计算出这个概率。这些值是不可用的，但可以估计。]
   e. 你在(b)部分使用的估计器的估计标准误差是多少？

3. 作为一个例子，在这种情况下，可以合理地使用几个不同的统计数据来计算一个点估计，考虑一个有N张发票的群体。与每张发票相关的是其 "账面价值"，即该发票的记录金额。让T表示总账面价值，一个已知的数额。这些账面价值中有些是错误的。审计工作将通过随机选择n张发票并确定每张发票的审计（正确）值来进行。假设抽样得出以下结果（以美元计）。

	1	2	3	4	5
账面价值	300	720	526	200	127
经审计的价值	300	520	526	200	157
误差	0	200	0	0	- 30

让

$\bar{Y}$ = 样本平均账面价值

$\bar{X}$= 样本平均审计值

$\bar{D}$= 样本平均误差

提出三种不同的统计方法来估计总的审计值（即正确值）--一种只涉及N和$\bar{X}$，另一种涉及T、N和$\bar{D}$，最后一种涉及T和$\bar{X}/\bar{Y}$。如果N=5000，T=1，761，300，请计算出三个相应的点估计。("审计中的统计模型和分析 "一文，《统计科学》，1989年：2-33，讨论了这些估计器的特性)。

4. Using a long rod that has length μ, you are going to lay out a square plot in which the length of each side is μ. Thus the area of the plot will be μ^2. However, you do not know the value of μ, so you decide to make n independent measurements $X_1, X_2, \ldots, X_n$ of the length. Assume that each X_i has mean μ (unbiased measurements) and variance σ^2.

   a. Show that $\bar{X}^2$ is not an unbiased estimator for μ^2. [Hint: For any rv Y, $E(Y^2) = V(Y) + [E(Y)]^2$. Apply this with $Y = \bar{X}$.]

   b. 对于μ来说，k的什么值是无偏的估计器$\bar{X}^2 - kS^2$？[提示：计算$E(\bar{X}^2 - kS^2)$。

5. 在n_1个随机选择的男性吸烟者中，X_1吸食过滤嘴香烟，而在n_2个随机选择的女性吸烟者中，X_2吸食过滤嘴香烟。让p_1和p_2分别表示随机选择的男性和女性抽过滤嘴香烟的概率。

   a. 证明$(X_1/n_1)-(X_2/n_2)$是p_1-p_2的无偏估计。[提示：$E(X_i) = n_i p_i$ for i = 1, 2.]

   b. (a)部分中估计者的标准误差是多少？

   c. 你将如何使用观测值x_1和x_2来估计你的估计器的标准误差？

   d. 如果$n_1 = n_2 = 200$, $x_1 = 127$, and $x_2 = 176$, 使用(a)部分的估计器来获得估计值 $p_1 - p_2$。

   e. 使用(c)部分的结果和(d)部分的数据来估计估计器的标准误差。

6. 考虑一个随机样本$X_1, \ldots, X_n$，来自于pdf

$$f(x;\vartheta) = .5(1+\vartheta x) \quad -1 \le x \le 1$$

其中$-1 \le \vartheta \le 1$（这种分布出现在粒子物理学中）。证明$\hat{\theta}=3\bar{X}$是一个无偏见的[提示：首先确定$\mu=E(X)=E(\bar{X})$]。

7. 让$X_1, X_2, \ldots, X_n$代表来自雷利分布的随机样本，其pdf为

$$f(x;\vartheta) = \frac{x}{\vartheta}e^{-x^2/(2\theta)} \quad x > 0$$

   a. It can be shown that $E(X^2) = 2\vartheta$. Use this fact to construct an unbiased estimator of ϑ based on ΣX_i^2（并使用期望值的规则来表明它是无偏的）。

   b. 从以下n = 10个关于涡轮叶片在特定条件下的振动应力的观察中估计ϑ：

16.88 10.23 4.59 6.66 13.68

14.23 19.87 9.40 6.51 10.95

8. 假设一种植物在1年内的真实平均生长量μ与第二种植物的生长量相同，但第一种植物的生长量方差为σ^2，而第二种植物的方差为$4\sigma^2$。让$X_1, \ldots, X_m$是第一种类型的m个独立增长观测值[所以$E(X_i) = \mu$, $V(X_i) = \sigma^2$]，并让$Y_1, \ldots, Y_n$是第二种类型的n个独立的增长观测值 $E(Y_i) = \mu$, $V(Y_i) = 4\sigma^2$。

a. 证明估计器$\hat{\mu}=\delta\bar{X}+(1-\delta)\bar{Y}$对$\mu$是无偏的（对于$0<\delta<1$，估计器是一个两个单独样本的加权平均数）。

b. 对于固定的m和n，计算$V(\hat{\mu})$，然后找到最小化$V(\hat{\mu})$的δ值。[提示：将$V(\hat{\mu})$与δ相差。]

6.2

9. 让X表示一个随机选择的学生花在某项能力测试上的分配时间比例。假设X的pdf是

$$f(x;\vartheta)\quad \begin{cases}(\vartheta+1)x^{\theta} & 0\le x\le 1\\ 0 & \text{否则}\end{cases}$$

其中 $-1<\vartheta$.对10名学生进行随机抽样，得到的数据是$x_1=.92, x_2=.79, x_3=.90, x_4=.65, x_5=.86$，$x_6=.47$，$x_7=.73$，$x_8=.97$，$x_9=.94$，$x_{10}=.77$。

a. 使用矩量法来获得ϑ的估计值，然后计算这个数据的估计值。

b. 获得ϑ的最大似然估计值，然后计算给定数据的估计值。

10. 10个测试点焊的每个点的剪切强度被确定，得出以下数据（psi）：

392 376 401 367 389 362 409 415 358 375

a. 假设剪切强度是正态分布，用最大似然法估计真实的平均剪切强度和剪切强度的标准差。

b. 再次假设是正态分布，请估计95%的焊缝的强度值将低于这个值。[提示：就μ和σ而言，第95个百分位数是多少？现在使用不变性原则]。

c. 假设我们决定检查另一个试验点焊缝。让X=焊缝的抗剪强度。使用给定的数据来获得$P(X\le400)$的mle。[提示：$P(X\le400)=\Phi((400-\mu)/\sigma)$。

11. 让$X_1,\ldots,X_n$是参数为α和β的伽马分布的随机样本。

a. 推导出方程，其解产生α和β的最大似然估计值，你认为它们可以明确解决吗？

b. 证明$\mu=\alpha\beta$的mle是$\hat{\mu}=\bar{X}$。

12. 让$X_1,X_2,\ldots,X_n$代表来自雷利分布的一个随机样本，其密度函数在练习15中给出。确定

a. ϑ的最大似然估计器，然后计算该练习中给出的振动应力数据的估计值。这个估计值与练习15中建议的无偏估计值是否相同？

b. 振动应力分布的中值的mle。[提示：首先用ϑ表示中位数]。

13. 考虑一个随机样本$X_1,X_2,\ldots,X_n$，来自移位指数pdf的随机样本

$$f(x;\lambda,\vartheta)=\begin{cases}\lambda e^{-\lambda(x-\theta)} & x\ge\vartheta\\ 0 & \text{否则}\end{cases}$$

ϑ=0给出了之前考虑的指数分布的pdf（在零的右边有正的密度）。例4.5中出现了一个移位的指数分布的例子，其中感兴趣的变量是交通流中的时间间隔，ϑ=0.5是可能的最小时间间隔。

a. 获得ϑ和λ的最大似然估计值。

b. 如果n=10个时间头程观察，得出的数值为3.11，0.64，2.55，2.20，5.44、3.42，10.39，8.93，17.82，和1.30，计算ϑ和λ的估计值。

补充练习

14. 如果对于任何一个$\epsilon>0$，$P(\hat{\theta})$是一致的，则称估计器$\hat{\theta}$是一致的。ϑ $\rightarrow \epsilon$) $\rightarrow \infty$ 0，因为n . 也就是说，$\hat{\theta}$是如果随着样本量的增加，$\hat{\theta}$离ϵ越来越远的可能性越来越小，那么就会保持一致。当$\sigma^2 < \infty$时，用Chebyshev的方法证明$\bar{X}$是μ的一个一致估计值。第3章练习43中的不等式。[提示：该不等式可以改写成以下形式

$$P(|Y - \mu_Y| \geq \epsilon) \leq \sigma_Y^2/\epsilon^2$$

现在将Y与$\bar{X}$相提并论]。

15. a. Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on $[0, \vartheta]$. Then the mle of ϑ is $\hat{\vartheta} = Y = \max(X_i)$. Use the fact that $Y \leq y$ iff each $X_i \leq y$ to derive the cdf of Y. Then show that the pdf of $Y = \max(X_i)$ is

$$f_Y(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & 0 \leq y \leq \vartheta \\ 否则为0 & \end{cases}$$

b.用(a)部分的结果表明，mle是有偏的，但$(n+1)\max(X_i)/n$是无偏的。

16. n个试样中的每一个都要在同一个秤上称量两次。让X_i和Y_i表示第i个试样的两个观测重量。假设X_i和Y_i彼此独立，各自为正态分布，均值为μ_i（试样i的真实重量），方差为σ^2。

a. 证明σ^2的最大似然估计是$\hat{\sigma}^2 = \Sigma(X_i - Y_i)^2/(4n)$。[提示：如果$\bar{z} = (z_1 + z_2)/2$，那么$\Sigma(z_i - \bar{z})^2 = (z_1 - z_2)^2/2$。]

b. mle $\hat{\sigma}^2$是σ的无偏估计值吗^2？找到σ的无偏估计值^2。[提示：对于任何rv Z，$E Z^2 = V(Z) + [E(Z)]^2$。将此应用于$Z = X_i - Y_i$。]

6 第七章

7.1

1. 假设随机抽取了50瓶某品牌的止咳糖浆，并测定了每瓶糖浆的酒精含量。让μ表示该品牌所有瓶子的平均酒精含量。假设得出的95%置信区间是（7.8，9.4）。
   a. 从这个相同的样本中计算出的90%的置信区间会比给定的区间更窄还是更宽？解释一下你的推理。
   b. 考虑以下陈述：μ在7.8和9.4之间有95%的可能性。这个说法是否正确？为什么或为什么不正确？
   c. 请考虑以下说法：我们可以高度肯定，95%的这种咳嗽糖浆的瓶子的酒精含量都在7.8和9.4之间。这种说法是否正确？为什么或为什么不正确？
   d. 考虑以下陈述：如果选择一个大小为50的样本，然后计算相应的95%区间的过程重复100次，所得到的区间中有95个将包括μ。为什么或为什么不正确？

2. 假设从任何特定煤层采集的煤样的氦气孔率（百分比）为正态分布，真实标准偏差为0.75。
   a. 如果某个接缝的20个试样的平均孔隙率为4.85，计算该接缝的真实平均孔隙率的95%CI。
   b. 根据16个样品平均孔隙率为4.56的试样，计算出另一个接缝的真实平均孔隙率的98%CI。
   c. 如果要使95%的区间宽度为0.40，需要多大的样本量？
   d. 在99%的置信度下，估计真实的平均孔隙率在0.2以内，需要多大的样本量？

3. 对某一类型的n=15个热泵进行随机抽样，得出以下关于使用寿命的观察结果（以年为单位）：

   2.0 1.3 6.0 1.9 5.1 .4 1.0 5.3

   15.7 .7 4.8 .9 12.2 5.3 .6

   a. 假设寿命分布是指数型的，使用与例题平行的参数 7.5，以获得预期（真实平均）寿命的95%CI。
   b. 为了达到99%的置信度，(a)部分的区间应该如何改变？
   c. 寿命分布的标准偏差的95%CI是什么？[提示：什么是指数型随机变量的标准差？]

4. 考虑一个统计顾问将为不同客户获得的下一个1000个μ的95%CI。假设这些区间所依据的数据集是相互独立选择的。在这1000个区间中，你希望有多少个区间能捕捉到相应的μ值？在这些区间中，有940到960个区间含有相应的μ值的概率是多少？[提示：让Y=1000个区间中包含μ*的*数量，Y是哪种随机变量？]

7.2

5. 燃气烹饪、厨房通风和暴露于燃烧产物 "一文（《室内空气》，2006年：65-73）报告说，对50个有燃气烹饪设备的厨房样本进行了为期一周的监测，样本平均CO_2 水平（ppm）为654.16，样本标准偏差为164.43。

a. 计算并解释所有家庭中真实的平均CO_2 水平的95%（双侧）置信区间，并从中选择样本。

b. 假设调查人员在收集数据之前对*s*的值做了一个粗略的猜测，即175。在95%的置信度下，要获得50ppm的区间宽度，需要多大的样本量？

6. 皮尤宗教和公共生活论坛在2009年12月9日报告说，在对2003名美国成年人的调查中，25%的人说他们相信占星术。

a. 计算并解释所有相信占星术的成年美国人的比例在99%的置信水平下的置信区间。

b. 无论$\hat{p}$的值是多少，99%CI的宽度最多为0.05，需要多少样本量？

7. 一个大学区的校长曾经上过概率和统计学的课程，他认为任何一天的教师缺 勤人数都是参数为μ的泊松分布，使用附带的50天的缺勤数据来获得μ的大样本 CI。

$$Z = \frac{\bar{X} - \mu}{\sqrt{\mu/n}}$$

具有近似的标准正态分布。现在像推导p的区间一样进行，做一个概率声明（概率为1 α），并解决由此产生的μ的不等式（见（7.10）后面的论证）。

缺勤次数	0	1	2	3	4	5	6	7	8	9	10
频率	1	4	8	10	8	7	5	3	2	1	1

8. 假设$X_1, \ldots, X_n$ 构成一个来自指数分布的随机样本，其平均值未知 μ.描述一种构建具有特定置信系数的μ的置信区间的方法 γ（$0 < \gamma < 1$）。提示：确定常数c_1 和c_2 ，使$P(c_1 < (1/\mu)\sum_{i=1}^{n} X_i < c_2) = \gamma$。

7.3

9. 随机抽取n=8个某种类型的E-玻璃纤维测试样本，得出的样本平均界面剪切屈服应力为30.2，样本标准偏差为3.1（"On Interfacial Failure in Notched Unidirectional Glass/Epoxy Composites," J. of Composite Materials, 1985: 276-286）。假设界面剪切屈服应力是正态分布的，计算真实平均应力的95%CI（就像所引用文章的作者那样）。

10. 一个由14个特定类型的接头试样组成的样本的平均比例极限应力为8.48兆帕，样本标准偏差为0.79兆帕（"Pegged Timber Connections中轴承强度系数的特征"，J. of Structural Engr., 1997: 326-332）。

a. 计算并解释所有这些关节的真实平均比例极限应力的95%下限置信度。如果有的话，你对比例极限应力的分布做了什么假设？

b. 计算并解释这种类型的单个关节的比例极限应力的95%预测下限。

11. 26名近海石油工人参加了模拟逃生演习，得出了完成逃生的时间（秒）的附带数据（"Oxygen Consumption and Ventilation During Escape from an Offshore Platform," Ergonomics, 1997: 281-292）：

389	356	359	363	375	424	325	394	402
373	373	370	364	366	364	325	339	393
392	369	374	359	356	403	334	397	

a. 计算人口平均逃逸时间的置信度上限，置信度为 95%.

b. 使用95%的预测水平，计算一个额外工人逃逸时间的预测上界。这个界限与(a)部分的信心界限相比如何？

c. 假设有两个额外的工人被选来参加模拟逃生演习。用 X_{27} 和 X_{28} 来表示他们的逃生时间，让 $\bar{X}_{new}$ 表示这两个值的平均值。修改单个X值的PI公式，得到$\bar{X}_{new}$的PI，并根据给定的逃生数据计算出95%的双侧区间。

7.4

12. 对LNG船安全壳中使用的n=9个脉冲电源气体金属电弧焊的样本，确定了横向膨胀量（密尔）。结果样本的标准偏差为s=2.81密耳。假设是正态的，求出σ的95%CI^2 和σ的95%CI。

13. 电线放电加工（WEDM）是一种用于制造导电硬质金属部件的工艺。它使用一个连续移动的金属丝作为电极。线材电极上的涂层可使线材电极芯冷却并提供更好的切割性能。文章 "用于线切割加工的高性能线电极--回顾"（J. of Engr. Manuf., 2012: 1757-1773）给出了以下关于总涂层厚度的样本观察（单位：μm
)的八个线状电极用于线切割加工：

21 16 29 35 42 24 24 25

计算涂层厚度分布的标准偏差的99%CI。无论分布的性质如何，这个区间是否有效？解释一下。

补充练习

14. 假设某类部件的寿命分布是指数型的，参数为λ，因此预期寿命为$\mu=1/\lambda$。挑选出n个这样的组件样本，并将每个组件投入运行。假设组件的寿命是独立的。让Y_1 表示第一次故障发生的时间，Y_2 表示第二次故障发生的时间，以此类推，所以$T_r = Y_1 + + Y_r + (n\ r)Y_r$ 是终止时的总累积寿命。然后可以证明，$2\lambda T_r$ 具有$2r$ df的卡方分布。利用这一事实，为真正的平均寿命$1/\lambda$制定一个100(1 α)%的CI公式。作为一个例子，假设有20个部件被测试，r=10。那么如果前十个故障时间是11，15，29，33，35，40，47，55，58，和72。从数据中计算出一个95%的CI。

15. 让$X_1, X_2, \ldots, X_n$ 是连续概率分布中的随机样本，其中位数为$\tilde{\mu}$（因此，$P(X_i \le \tilde{\mu}) = P(X_i \ge \tilde{\mu}) = .5$）。

a. 表明

$$P(\min(X_i) < \tilde{\mu} < \max(X_i)) = 1 - \left(\frac{1}{2}\right)^{n-1}$$

所以(min (x), max (x))是μ的100(1-α)%置信区间，$\alpha=^{1}$ $^{n-1}$。[提示：$\frac{}{2}$
事件{min $(X_i) \le \mu \le$ max (X_i)}的补充是{max $(X_i) \le \mu$}∪{min $(X_i) \ge \mu$}，但是
max $(X_i) \le \tilde{\mu}$ iff $X_i \le \tilde{\mu}$ for all i.]

b. 对6个正常男婴中的每一个，在婴儿食用不含异亮氨酸的饮食时，测定氨基酸丙氨酸的量（mg/100 mL），得出以下数据：

2.84 3.54 2.80 1.44 2.94 2.70

计算这种饮食方式下婴儿丙氨酸真实中位数的97%CI（《婴儿必需氨基酸的需求》，Amer.J. of Nutrition, 1964：322-330).

c. Let $x_{(2)}$ denote the second smallest of the x_i's and $x_{(n-1)}$ denote the second largest of the x_i's. What is the confidence level of the interval $x_{(2)}, x_{(n-1)}$ for $\tilde{\mu}$?

16. 让$X_1, X_2, \ldots, X_n$ 是区间[0, ϑ]上均匀分布的一个随机样本，因此，

$$f(x) = \begin{cases} \frac{1}{\vartheta} & 0 \le x \le \vartheta \\ 0 & \text{otherwise} \end{cases}$$

那么如果$Y = \max(X_i)$，可以证明Rv $U = Y/\vartheta$有密度函数

$$f_U(u) = \begin{cases} nu^{n-1} & 0 \le u \le 1 \\ 0 & \text{否则} \end{cases}$$

a. 使用$f_U(u)$来验证：

$$P\left((\alpha/2)^{1/n} < \frac{Y}{\vartheta} \le (1-\alpha/2)^{1/n}\right) = 1-\alpha$$

并以此推导出ϑ的100(1-α)%CI。

b. 验证$P(\alpha^{1/n} \le Y/\vartheta \le 1) = 1-\alpha$，并根据这个概率声明得出$\vartheta$的100(1-$\alpha$)%CI。

c. 之前得出的两个区间哪个更短？如果我的早班车等待时间是均匀分布的，观察到的等待时间是$x_1 = 4.2$, $x_2 = 3.5$, $x_3 = 1.7$, $x_4 = 1.2$, and $x_5 = 2.4$, 用这两个区间中较短的一个来推导ϑ的95%CI。

17. 假设0≤γ≤α，那么当n很大时，μ的100(1-α)%CI为

$$\left(\bar{x} - z_\gamma \frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha-\gamma} \frac{s}{\sqrt{n}}\right)$$

该选择$\gamma = \alpha/2$得到的是第二节中得出的通常的区间。$\gamma \ne \alpha/2$中得出的通常区间；如果，这个区间就不是关于$\bar{x}$的对称性。这个区间的宽度是$w = s(z_\gamma + z_{\alpha-\gamma})/\sqrt{n}$。证明$w$是最小化的为选择$\gamma=\alpha/2$，所以对称区间是最短的。[提示：(a) 根据定义$z_\alpha$，$\Phi(z_\alpha) = 1-\alpha$, so that $z_\alpha = \Phi^{-1}(1-\alpha)$; (b) the relationship between the derivative of a function $y = f(x)$，反函数$x = f^{-1}(y)$为$(d/dy)f^{-1}(y) = 1/f'(x)$。

18. a.使用例7.5的结果，得到指数分布的参数λ的95%的置信度下限，并根据例中给出的数据计算该下限。

b.如果寿命X具有指数分布，那么寿命超过t的概率为P（$X>t$）$=e^{-\lambda t}$。使用（a）部分的结果，可以得到故障时间超过100分钟的概率的95%的置信度下限。

19. 让$X_1, X_2, \ldots, X_n$ 是独立和相同分布的随机变量，$X_1 \sim$ Exponential(λ)。证明$2\lambda \sum_{i=1}^{n} X_i$ 有一个自由度为$2n$的卡方分布，即$2\lambda \sum_{i=1}^{n} X_i \sim \chi^2(2n)$。

7 第八章

8.1

1. 让μ表示真正的平均放射性水平（皮库里/升）。5pCi/L这个值被认为是安全水和不安全水的分界线。你建议测试以下两个选项中的哪一个：
   (a) $H_0: \mu = 5$与$H_a: \mu > 5$相比；或
   (b) $H_0: \mu = 5$与$H_a: \mu < 5$？
2. 在同意为某一类型的高压注油海底电力电缆购买大量聚乙烯护套的订单之前，一家公司希望看到确凿的证据，证明护套厚度的真实标准偏差小于0.05毫米。应该对哪些假设进行测试，以及为什么？在这种情况下，什么是I型和II型错误？
3. 两家不同的公司申请在某一地区提供有线电视服务。让p表示在所有潜在用户中，倾向于第一家公司而不是第二家公司的比例。考虑在25人的随机样本基础上测试$H_0: p = .5$与$H_a: p = .5$。让测试统计量X代表样本中支持第一家公司的人数，X代表X的观察值。
   a. 在这个问题情境中，描述I型和II型错误。
   b. 假设X=6。哪些X的值至少与H_0一样矛盾？
   c. 当H_0为真时，测试统计量X的概率分布是什么？用它来计算
   当x=6时，P-值。
   d. 如果当P-值为0.044时要拒绝H_0，请计算当P=0.4时发生II型错误的概率，当P=0.3时再次计算，当P=0.6和P=0.7时也计算。[提示：P-值>.044等同于涉及x的什么不等式（见例8.4）？
   e. 使用（d）的测试程序，如果25个被询问者中有6个赞成公司1，你会得出什么结论？
4. 要通过对一个10公斤的试样称重25次来检查衡器的校准。假设不同称量的结果是相互独立的，并且每次试验的重量是正态分布，σ=0.200 kg。让μ表示秤上真正的平均重量读数。
   a. 应该检验哪些假设？
   b. 以样本平均数本身作为检验统计量，当$\bar{x}$=9.85时，P值是多少，在显著性水平0.01时你会得出什么结论？
   c. 对于α=0.01的测试，当事实上μ=10.1时，重新校准被判断为不必要的概率是多少？当μ=9.8时？

8.2

5. 让μ表示对某一刺激的真实平均反应时间。对于H的Z检验$_0$：μ=5与
   $H_a: \mu > 5$，请确定以下每个z检验统计量值的P值。
   a.1.42 B . 90 C . 1.96 D . 2.48 E .-.11
6. 回答例8.7中轮胎问题的下列问题。
   a. 如果$\bar{x}$=30，960，采用水平α=0.01的检验，那么决定是什么？
   b. 如果使用0.01水平测试，β(30, 500)是多少？
   c. 如果使用0.01的水平检验，同时要求β(30, 500)=0.05，需要多大的样本量n？

d. 如果$\bar{x}$=30，960，可以拒绝H_0 的最小α是多少（基于n=16）？

7. 测定了某品牌氢化植物油的16个样品中每个样品的熔点，结果是$\bar{x}$=94.32。假设熔点的分布是正态的，σ=1.20。

   a. 使用双尾水平.01检验$H_0 : \mu = 95$与$H_a : \mu \neq 95$。
   b. 如果使用0.01水平的检验，当μ=94时，β(94)是什么，即出现II型错误的概率？
   c. 当α=.01时，需要多大的n值才能保证β(94)=.1？

8. 某一类型的铝质水泥中SiO_2 的理想百分比是5.5。为了测试某一生产设施的真实平均百分比是否为5.5，对16个独立获得的样品进行分析。假设样品中SiO_2 的百分比为正态分布，$\sigma = .3$，$\bar{x} = 5.25$。

   a. 这是否确凿地表明，真正的平均百分比与5.5不同？
   b. 如果真实的平均百分比是μ=5.6，并且使用基于n=16的水平α=.01检验，那么检测到这种偏离H 的概率是多少$_0$ ？
   c. 要满足α=.01和β(5.6)=.01，需要多大的n值？

9. 证明对于任何$\Delta > 0$，当群体分布为正态分布且σ已知时，双尾检验满足$\beta(\mu_0 - \Delta) = \beta(\mu_0 + \Delta)$，因此$\beta(\mu')$ 关于μ_0 是对称的。

10. 对于一个固定的替代值μ' ，表明$\beta(\mu')$ $\rightarrow$ 0为n $\rightarrow \infty$ 对于单尾或双尾的z 在已知σ的正态人口分布的情况下进行检验。

8.3

11. 某一类型的球轴承的真正平均直径应该是*0.5*英寸。将进行单样本t检验，以了解情况是否属实。在下列情况中，什么结论是合适的？
    a. $n = 13, t = 1.6, \alpha = .05$
    b. $n = 13, t = -1.6, \alpha = .05$
    c. $n = 25, t = -2.6, \alpha = .01$
    d. n=25，t=-3.9

12. 用来在道路上划线的涂料必须能反射足够的光线，以便在夜间清晰可见。让μ表示正在考虑的一种新型涂料的真实平均反射仪读数。$H_0 : \mu = 20$与$H_a : \mu > 20$的测试将基于正常人群分布中大小为n的随机样本。在以下每种情况下，什么结论是合适的？
    a. $n = 15, t = 3.2, \alpha = .05$
    b. $n = 9, t = 1.8, \alpha = .01$
    c. $n = 24, t = -.2$

13. 文章 "铁路轨道生命周期成本的不确定性估计"（J. of Rail and Rapid Transit, 2009）提出了以下数据：某条铁路线的弯曲轨道上的高轨断裂的修复时间（min）。

    159 120 480 149 270 547 340 43 228 202 240 218

    数据的正态概率图显示了一个合理的线性模式，所以维修时间的群体分布至少是近似正态的，这是可信的。样本的平均值和标准差分别为249.7和145.1。

a. 是否有令人信服的证据可以断定真正的平均维修时间超过200分钟？用0.05的显著性水平对假设进行检验。

b. 使用σ=150，当真正的平均维修时间实际为300分钟时，(a)中使用的测试的第二类错误概率是多少？也就是说，β（300）是多少？

14. 重新考虑练习1.53中首次介绍的大盘成长型共同基金的费用率（%）的附带样本数据。

0.52 1.06 1.26 2.17 1.55 0.99 1.10 1.07 1.81 2.05
0.91 0.79 1.39 0.62 1.52 1.02 1.10 1.78 1.01 1.15

正态概率图显示了一个合理的线性模式。

a. 是否有令人信服的证据证明人口平均费用率超过1%？用0.01的显著性水平对相关假设进行检验。

b. 请参考(a)项，在上下文中描述第一类和第二类错误，并说明你在得出结论时可能犯的错误。获得数据的来源报告说，在所有762个这样的基金的群体中，μ=1.33。那么你在得出结论时是否真的犯了错误？

c. 假设σ=0.5，请确定并解释(a)中对实际值的检验力。
(b)中所述的μ。

8.4

15. 考虑使用Z检验来检验H_0 ：$P = .6$。确定以下每种情况下的P值。
   a. $H_a : p > .6, z = 1.47$
   b. $H_a : p < .6, z = -2.70$
   c. $H_a : p \neq .6, z = -2.70$
   d. $H_a : p < .6, z = .25$

16. 肥胖者的一个常见特征是他们的身体质量指数至少为30[BMI=体重/（身高）^{2}，其中身高为米，体重为公斤]。文章 "肥胖对石油化工工人的疾病缺勤和生产力的影响"（Annals of Epidemiology, 2008: 8-14）报告说，在一个女工样本中，262人的体重指数低于25，159人的体重指数至少为25但低于30，120人的体重指数超过30。是否有令人信服的证据可以得出结论，抽样人群中超过20%的人是肥胖的？

a. 提出并检验适当的假设，其显著性水平为0.05。

b. 在这种情况下，请解释什么是I型和II型错误。

c. 当肥胖者的实际比例为25%时，不得出超过20%的人口是肥胖者的结论的概率是多少？

17. 一家航空公司制定了一个行政人员俱乐部的计划，其前提是5%的现有客户有资格成为会员。对500名顾客进行随机抽样，发现有40人符合条件。

a. 利用这些数据，在0.01的水平上检验公司的前提是正确的这一零假设，而另一零假设则是不正确的。

b. 当使用(a)部分的测试时，当事实上10%的现有客户符合条件时，公司的前提将被判断为正确的概率是多少？

补充练习

18. 当$X_1, X_2, ..., X_n$是独立的泊松变量，每一个都有参数μ，并且n很大，样本的平均值$\bar{X}$近似于正态分布，那$\mu=E(\bar{X})$，$V(\bar{X})=\mu/n$。这意味着

$$Z = \frac{\bar{X} - \mu}{\sqrt{\mu/n}}$$

具有近似的标准正态分布。对于测试$H_0: \mu = \mu_0$，我们可以用μ代替μ_0在Z的方程中，得到一个检验统计量。这个统计量实际上优于分母为$S/\sqrt{n}$的大样本统计量（当X_i's是泊松时），因为它是明确针对泊松假设。如果某位统计学家在某一时期收到的咨询请求的数量为每周5天的工作时间具有泊松分布，在36周内的咨询请求总数为160，这是否表明每周请求的真实平均数超过4.0？用$\alpha=0.02$进行检验。

19. 当群体分布是正态的，且n很大时，样本标准差S近似于正态分布，$E(S)\approx\sigma$和$V(S)\approx\sigma^2/(2n)$。我们已经知道，在这种情况下，对于任何n，$\bar{X}$是正态的$E(\bar{X})=\mu$，$V(\bar{X})=\sigma^2/n$。

    a. 假设基础分布是正态的，那么第99个百分位数$\vartheta=\mu+2.33\sigma$的近似无偏估计值是多少？

    b. 当X_i's是正常的，可以证明$\bar{X}$和S是独立的Rv（一个测量的位置，而其他措施则是扩散）。用它来计算(a)部分的估计器$\hat{\theta}$的$V(\hat{\theta})$和$\sigma_{\hat{\theta}}$。估计的标准误差$\hat{\sigma}_{\hat{\theta}}$是多少？

    c. 写出一个用于检验$H_0: \vartheta = \vartheta_0$，当$H_0$为真时，具有近似标准正态分布的检验统计量。如果某个地区的土壤pH值呈正态分布，64个土壤样本产生$\bar{x}=6.33$，$s=.16$，这是否为得出所有可能的样本中至少有99%的pH值小于6.75的结论提供了有力的证据？用$\alpha=0.01$进行检验。

20. 让$X_1, X_2, ..., X_n$是参数为λ的指数分布的随机样本。那么可以证明$2\lambda\Sigma X_i$具有$\nu=2n$的奇偶分布（首先证明$2\lambda X_i$具有$\nu=2$的奇偶分布）。

    a. 利用这一事实得到检验$H_0: \mu = \mu_0$。然后解释当替代假设为$H_a: \mu < \mu_0$时，你将如何确定P值。[提示：$E(X_i) = \mu = 1/\lambda$，所以$\mu = \mu_0$相当于$\lambda = 1/\mu_0$。]

    b. 假设对十个相同的部件进行测试，每个部件的失效时间呈指数分布。结果故障时间为

    95 16 11 3 42 71 225 64 87 123

    使用(a)部分的测试程序来决定数据是否强烈地表明真实的平均寿命小于先前声称的75。[提示：参考表A.7]。

8 第九章

9.1

1. **5.** Persons having Reynaud's syndrome are apt to suffer a sudden impairment of blood circulation in fingers and toes. In an experiment to study the extent of this impairment, each subject immersed a forefinger in water and the resulting heat output cal/cm^2/min was measured. For m = 10 subjects with the syndrome, the average heat output was $\bar{x}$ = .64, and for n = 10 nonsufferers, the average output was 2.05. Let μ_1 and μ_2 denote the true average heat outputs for the two types of subjects. Assume that the two distributions of heat output are normal with σ_1 = .2 and σ_2 = .4.

   a. 考虑测试$H_0 : \mu_1\ \mu_2 = 1.0$与$H_2 : \mu_1\ \mu_2 < 1.0$，水平为0.01。用文字描述H_2 的内容，然后进行检验。

   b. 当μ_1 和μ_2 之间的实际差异为μ_1 -μ_2 =时，出现II型错误的概率是多少？
   -1.2？

   c. 假设 $m = n$，当 $\mu_1 - \mu_2 = -1.2$ 时，需要多少样本量才能确保$\beta = .1$？

2. **11.**对152名20-30岁的男性危险废物工人和86名女性工人的样本测定了血液中的铅含量，结果男性的平均标准误差为5.5 ± 0.3，女性为3.8 ± 0.2（"Temporal Changes in Blood Leads of Hazardous Waste Workers in New Jersey, 1984-1987, " Environ.监测和评估，1993年：99-107）。以提供可靠性和精确性信息的方式，计算出男性和女性工人真实平均血铅水平之间的差异估计。

9.2

3. **19.**假设μ_1 和μ_2 是某一类型的汽车在50mph时的真实平均停车距离，配备了有两种不同类型的制动系统。使用显著性水平为0.01的双样本t检验来检验 $H_0 : \mu_1 - \mu_2 = -10$与$H_a : \mu_1 - \mu_2 < -10$的数据：$m$=6，6，$\bar{x}$ = 115.7, s_1 = 5.03, n = $\bar{y}$=129.3、和s_2 = 5.38。

4. **23.**粘合衬的使用频率越来越高，以支持外层织物并改善各种服装的形状和悬垂性。文章 "Compatibility of Outer and Fusible Interlining Fabrics in Tailored Garments"（Textile Res. J.，1997: 137-142）给出了高质量（H）织物和劣质（P）织物试样在100gm/cm时的延展性（%）的附带数据。

H	1.2	.9	.7	1.0	1.7	1.7	1.1	.9	1.7
	1.9	1.3	2.1	1.6	1.8	1.4	1.3	1.9	1.6
	.8	2.0	1.7	1.6	2.3	2.0			
P	1.6	1.5	1.1	2.1	1.5	1.3	1.0	2.6	

   a. 构建正态概率图，以验证两个样本都是从正态人口分布中选出的合理性。

   b. 构建一个比较性的图表。它是否表明高质量织物标本的真正平均延伸性与劣质标本的延伸性之间存在差异？

   c. 高质量样本的平均数和标准差分别为1.508和0.444，而劣质样本的平均数和标准差为1.588和0.530。使用双样本t检验来决定两类织物的真实平均延伸性是否不同。

高质量面料的正态概率图（图10） 劣质面料的正态概率图

（图11）

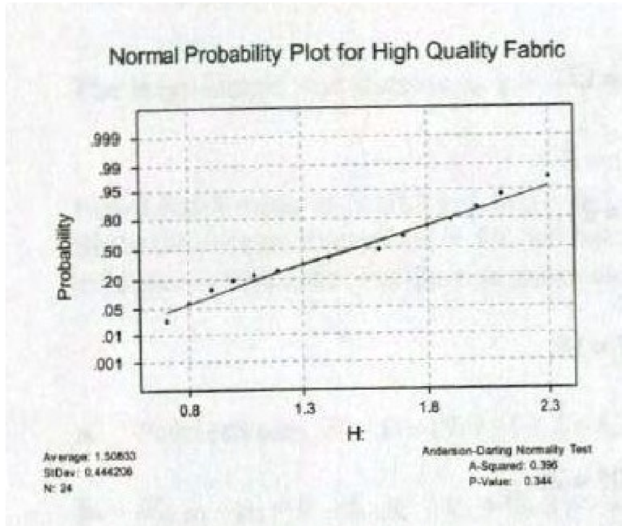


图10： 问题23

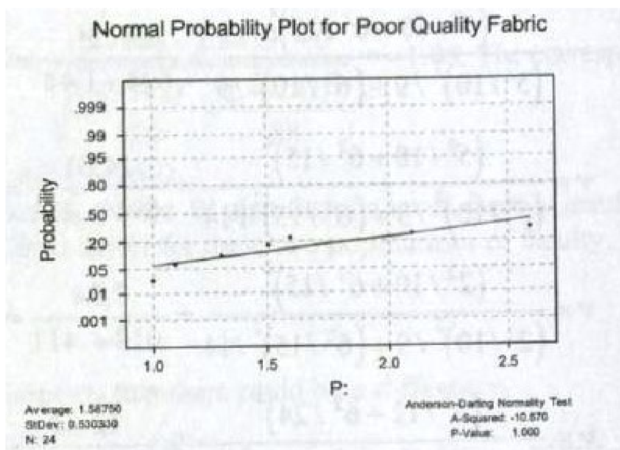


图11： 问题23

高质量和低质量织物的比较箱形图（图12）

5. **29.**饮料罐和塑料瓶的强度"（J. of Testing and Evaluation, 1993: 129- 131）"内部气体压力对压缩的影响 "一文中包含了一个装满草莓饮料的12盎司铝罐样本和另一个装满可乐的样本的压缩强度（磅）数据，这些数据是否表明可乐的额外碳酸化导致了更高的平均压缩强度？将你的答案建立在*P*值的基础上。你的分析需要哪些假设？

饮料	样本数量	样本平均值	样本SD 草
莓饮料	15	540	21
可乐	15	554	15

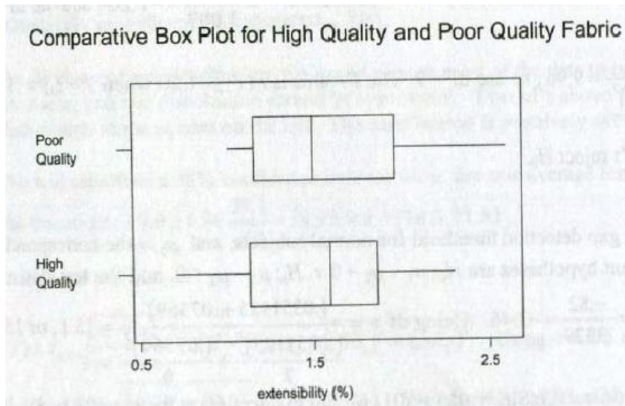


图12： 问题23

9.3

6. **39.**科学家和工程师经常希望比较两种不同的技术来测量或确定一个变量的值。在这种情况下，兴趣集中在测试测量的平均差异是否为零。文章"评价氘稀释技术与测定母乳摄入量的测试称量程序"（Amer. J. of Clinical Nutr., 1983: 996-1003）报告了14个随机的婴儿所摄入的奶量数据。选定的婴儿。

	婴儿				
	1	2	3	4	5
DD方法	1509	1418	1561	1556	2169
钨的方法	1498	1254	1336	1565	2000
差异	11	164	225	-9	169

	婴儿				
	6	7	8	9	1 0
DD方法	1760	1098	1198	1479	1281
钨的方法	1318	1410	1129	1342	1124
差异	442	-312	69	137	157

	婴儿			
	11	12	13	1 4
DD方法	1414	1954	2174	2058
钨的方法	1468	1604	1722	1518
差异	-54	350	452	540

a. 差异的人口分布是正常的，这是否合理？

b. 两种方法测得的摄入量的真正平均差值是否是零以外的？确定检验的*P*值，并在显著性水平*0.*05时用它来得出结论。

7. **43.**库欣氏病的特点是由于肾上腺或垂体功能障碍导致的肌肉无力。为了提供有效的治疗，尽早发现儿童库欣氏病是很重要的。在《经鼻腔微腺体切除术治疗儿童和青少年库欣氏病》（New Emgl. J. of Med., 1984: 889）一文中给出了15名患儿的发病年龄和诊断年龄（月）。以下是发病年龄和诊断年龄之间的差异值：

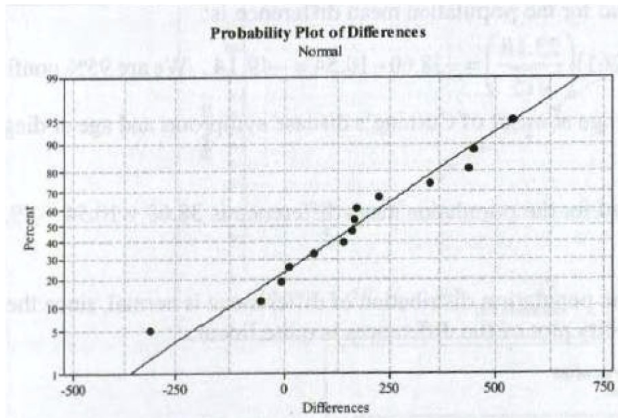


图13：问题39

-24 -12 -55 -15 -30 -60 -14 -21
-48 -12 -25 -53 -61 -69 -80

a. 附带的正态概率图（图14）是否对差异的群体分布的近似正态性产生了强烈的怀疑？

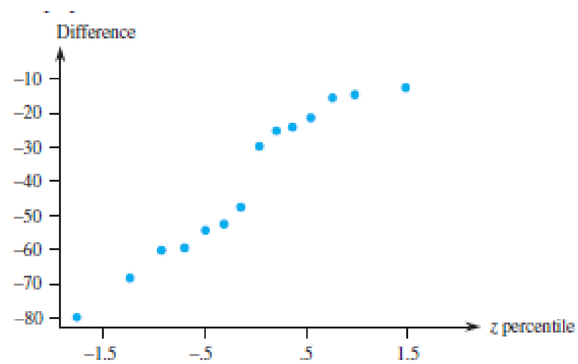


图14： 问题43

b. 计算群体平均差异的95%的置信度下限，并解释所得的界限。

c. 假设（诊断时的年龄）-（发病时的年龄）的差异已经计算出来了。相应的人群平均差异的95%的置信度上限是多少？

9.4

8. **55(53).**在医学调查中，比值$\vartheta = p / p_{12}$ 往往比差值p_1 p_2 更有意义（例如，给予治疗1的个体是给予治疗2的个体的多少倍？ ）。让$\hat{\theta} = \hat{p}_1 / \hat{p}_2$ 。当m和n都很大时，统计量$\ln(\hat{\theta})$具有近似的正态分布，具有近似的平均值$\ln(\vartheta)$和近似的标准差$[(m-x)/(mx)+(n-y)/(ny)]^{1/2}$ 。
a. 利用这些事实，可以得到用于估计$\ln(\vartheta)$的大样本95%CI公式，然后得到ϑ的CI
本身。

b. 回到例1.3的心脏病发作数据，在95%的置信度下计算出ϑ的合理值区间。这个区间对阿司匹林的疗效有什么提示？

9. **57(55).**两种不同类型的合金，A和B，已经被用来制造将用于某种工程应用的小型拉伸连接的实验试样。每个试样的极限强度(ksi)被确定，其结果在随附的频率分布中被总结出来。

	A	B
26_ < 30	6	4
30_ < 34	12	9
34_ < 38	15	19
38_ < 42	7	10
	$m = 40$	$m = 42$

计算所有合金A和B试样的真实比例之差的95%CI。

其极限强度至少为34ksi。

9.5

10. **59(57).**获得或计算以下数量：

a. $F_{.05,5,8}$

b. $F_{.05,8,5}$

c. F95,5,8

d. $F_{.95,8,5}$

e. F分布的第99个百分点，$v_1 = 10$，$v_2 = 12$

f. F分布的第一百分位数，$v_1 = 10$，$v_2 = 12$

g. P（$F \le 6.16$）为$v_1 = 6$，$v_2 = 4$

h. $P(.177 \le F \le 4.74)$ 为$v_1 = 10, v_2 = 5$

11. **65(63).** 文 章 "Progment of Compressive Properties of Failed Concrete Cylinders with Poly- mer Impregnation"(J. of Testing and Evaluation, 1977: 333-337)报告了两种不同的聚合物用于修复失效混凝土的裂缝时的压缩模量(psi 10⁶×)的以下数据。

环氧树脂	1.75 2.12 2.05 1.97
MMA预聚物	1.77 1.59 1.70 1.69

首先使用文中建议的方法获得一般的置信区间公式，从而获得变异率的90%CI。

补充练习

12. **67(65).**随之而来的关于12 10 8英寸箱子的压缩强度（磅）的总结数据出现在×《使×用固定和浮动测试压板的单壁瓦楞运输容器的压缩》（J. Testing and Evaluation, 1992: 318-320）一文中。作者说："使用固定压板和浮动压板方法的压缩强度之间的差异被发现与正常的压缩强度相比是很小的。

相同的盒子之间的压缩强度的变化"。你同意吗？你的分析是以任何假设为前提的吗？

方法	样本量	样本平均值	样本SD 固
定的	10	807	27
漂浮	10	757	41

13. **69(67).**如果在问卷中加入某种激励措施，是否会影响问卷的回复率？在一个实验中，110份没有激励措施的问卷有75份被退回，而98份包括抽奖机会的问卷则有66份回复（"慈善机构，没有；抽奖，没有；现金，有"，《民意季刊》，1996：542-562）。这些数据是否表明，包括激励措施会增加回应的可能性？说明并检验相关的假设，其显著性水平为0.10。

14. **71(69).** The article "Quantitative MRI and Electrophysiology of Preoperative Carpal Tunnel Syndrome in a Female Population" (Ergonomics, 1997: 642-649) reported that (-473.13, 1691.9) was a large-sample 95% confidence interval for the difference between true average thenar muscle volume mm^3 for sufferers of carpal tunnel syndrome and true average volume for nonsufferers. Calculate and interpret a 90% confidence interval for this difference.

15. **89(85).**假设$H_0 : \mu_1 \mu_2 = 0$与$H_\alpha : \mu_1 \mu_2 > 0$的0级检验，假设$\sigma_1 = \sigma_2 = 10$和两个分布的正态性，使用相同的样本量（$m = n$）。评估当$\mu_1 \mu_2 = 1$和$n = 25$、100、2500和10,000时发生II型错误的概率。你能想到在实际问题中，$\mu_1 \mu_2 = 1$的差异没有什么实际意义吗？在这样的问题中，n=10,000的样本量是否是可取的？

16. **95(91).**参照练习94，为$\mu_1 \mu_2$ 制定一个大样本的置信区间公式。以95%的置信度计算所给数据的区间。