MAT 2377
Probability and Statistics for Engineers

Practice Set

Iraj Yadegari (uOttawa)

Winter 2020

University of Ottawa

This is based on course notes by Rafał Kulik, Patrick Boily, and textbook of MAT2377.

Q1. Two events each have probability 0.2 of occurring and are independent. The probability that neither occur is:

a)0.64 b)0.04 c)0.2 d)0.4 e)none of the preceding

Solution: since A and B are events, then

$$P(\text{neither}) = P((A \cup B)^c) = P(A^c \cap B^c).$$

Since A and B are independent, so are A^c and B^c (why?). Thus,

$$\begin{aligned} P(\text{neither}) &= P(A^c)P(B^c) \\ &= (1 - P(A))(1 - P(B)) \\ &= (1 - 0.2)(1 - 0.2) = 0.64. \end{aligned}$$

Q2. Two events each have probability 0.2 and are mutually exclusive. The probability that neither occurs is:

a)0.36 b)0.04 c)0.2 d)0.6 e)none of the preceding

Solution: since A and B are events, then

$$P(\text{neither}) = P((A \cup B)^c) = 1 - P(A \cup B).$$

But $P(A \cup B) = P(A) + P(B)$ since A and B are mutually exclusive, so

$$\begin{aligned} P(\text{neither}) &= 1 - P(A \cup B) = 1 - P(A) - P(B) \\ &= 1 - 0.2 - 0.2 = 0.6. \end{aligned}$$

Q3. A smoke-detector system consists of two parts A and B. If smoke occurs then the item A detects it with probability 0.95, the item B detects it with probability 0.98 whereas both of them detect it with probability 0.94. What is the probability that the smoke will not be detected?

a)0.01 b)0.99 c)0.04 d)0.96 e)none of the preceding

Solution: let A represent the event that part A detects smoke, and same for B. We have $P(A) = 0.95$, $P(B) = 0.98$, $P(A \cap B) = 0.94$. Then

$$\begin{aligned} P(\text{smoke not detected}) &= 1 - P(\text{smoke detected}) \\ &= 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B)) \\ &= 1 - (0.95 + 0.98 - 0.94) = 0.01. \end{aligned}$$

Q4. Three football players will attempt to kick a field goal. Let A_1, A_2, A_3 denote the events that the field goal is made by player $1, 2, 3$, respectively. Assume that A_1, A_2, A_3 are independent and $P(A_1) = 0.5$, $P(A_2) = 0.7$, $P(A_3) = 0.6$. Compute the probability that exactly one player is successful.

a)0.29 b)0.21 c)0.71 d)0.79 e)none of the preceding

Solution: we have

$$\begin{aligned}
P(\text{only player 1 succeeds}) &= P(A_1 \cap A_2^c \cap A_3^c) = P(A_1)P(A_2^c)P(A_3^c) \\
&= 0.5 \times 0.3 \times 0.4 = 0.06 \\
P(\text{only player 2 succeeds}) &= P(A_1^c \cap A_2 \cap A_3^c) = P(A_1^c)P(A_2)P(A_3^c) \\
&= 0.5 \times 0.7 \times 0.4 = 0.14 \\
P(\text{only player 3 succeeds}) &= P(A_1^c \cap A_2^c \cap A_3) = P(A_1^c)P(A_2^c)P(A_3) \\
&= 0.5 \times 0.3 \times 0.6 = 0.09
\end{aligned}$$

But these three events are mutually exclusive, so

$$P(\text{exactly one succeeds}) = P(1) \cup P(2) \cup P(3) = P(1) + P(2) + P(3) = 0.29.$$

Q5. In a group of 16 candidates for laboratory research positions, 7 are chemists and 9 are physicists. In how many ways can one choose a group of 5 candidates with 2 chemists and 3 physicists?

Solution: this is a two-stage procedure. There are $\binom{7}{2}$ ways of selecting 2 chemists among the 7 (the first stage), and $\binom{9}{3}$ ways of selecting 3 physicists (the second stage).

Thus, there are

$$\binom{7}{2}\binom{9}{3}=\frac{7!}{5!2!}\times\frac{9!}{6!3!}=\frac{7\cdot 6}{2}\times\frac{9\cdot 8\cdot 7}{3\cdot 2}=21\times 84=1764$$

ways of selecting a group of candidates with the required constraints.

Q6. There is a theorem of combinatorics that states that the number of permutations of n objects in which n_1 are alike of kind 1, n_2 are alike of kind 2, ..., and n_r are alike of kind r (that is, $n = n_1 + n_2 + \cdots + n_r$) is

$$\frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot n_r!}.$$

Find the number of different words that can be formed by rearranging the letters in the following words (include the given word in the count):

a)NORMAL b)HHTTTT c)ILLINI d)MISSISSIPPI

Solution:

- NORMAL: each letter is different, so $\frac{6!}{1!1!1!1!1!1!} = 6! = 6{\cdot}5{\cdot}4{\cdot}3{\cdot}2{\cdot}1 = 720$
  NORMAL, NROMAL, NOMRAL, etc.

- HHTTTT: $2\times$H and $4\times$T, so $\frac{6!}{2!4!} = \frac{6{\cdot}5}{2} = 15$
  HHTTTT, HTHTTT, HTTHTT, HTTTHT, etc.

- ILLINI: $3\times$I, $2\times$N, and $1\times$N, so $\frac{6!}{3!2!1!} = \frac{6{\cdot}5{\cdot}4}{2} = 60$
  ILLINI, ILILNI, ILLNII, etc.

- MISSISSIPPI: $4\times$I, $1\times$M, $4\times$S, and $2\times$P, so $\frac{11!}{4!1!4!2!} = \frac{39{,}916{,}800}{1152} = 34,650$.

Q7. A class consists of 490 engineering students and 510 science students. The students are divided according to their marks:

	Passed	Failed
Eng.	430	60
Sci.	410	100

If one person is selected randomly, the probability that it failed if it was an engineering student?

a)0.06 b)0.12 c)0.41 d)0.81 e)none of the preceding

Solution: there are 1000 students in total.

Let A and E represent the events that the student passed and that the student is an engineer, respectively. Then

$$P(A^c|E) = \frac{P(A^c \cap E)}{P(E)}) = \frac{60/100}{490/1000} = 0.12.$$

Q8. A company which produces a particular drug has two factories, A and B. 30% of the drug are made in factory A, 70% in factory B. Suppose that 95% of the drugs produced by factory A meet specifications while only 75% of the drugs produced by factory B meet specifications. If I buy a dose of the company's drug, what is the probability that it meets specifications?

a)0.81 b)0.95 c)0.75 d)0.7 e)none of the preceding

Solution: let M be the events that the drug meets specifications, A that it is produced by factory A, B that it is produced by factory B.

We have $P(A) = 0.7$, $P(B) = 0.3$, $P(M|A) = 0.95$, and $P(M|B) = 0.75$. According to the Law of Total Probability,

$$P(M) = P(M|A)P(A) + P(M|B)P(B) = 0.95 \cdot 0.7 + 0.75 \cdot 0.3 = 0.89.$$

Q9. A medical research team wished to evaluate a proposed screening test for Alzheimer's disease. The test was given to a random sample of 450 patients with Alzheimer's disease; in 436 cases the test result was positive. The test was also given to a random sample of 500 patients without the disease; only in 5 cases was the result was positive. It is known that in Canada 11.3% of the population aged $65+$ have Alzheimer's disease. Find the probability that a person has the disease given that their test was positive (choose the closest answer).

a)0.97 b)0.93 c)0.99 d)0.07 e)none of the preceding

Solution: let A and D be the events that the test is positive and that the person has the disease, respectively. From the statement of the problem, we have that

$$P(A|D) = \frac{436}{450}, P(A|D^c) = \frac{5}{500}, P(D) = 0.113.$$

According to Bayes Theorem,

$$\begin{aligned} P(D|A) &= \frac{P(A|D)P(D)}{P(A|D)P(D) + P(A|D^c)P(D^c)} \\ &= \frac{436/450 \cdot 0.113}{436/450 \cdot 0.113 + 5/500 \cdot (1 - 0.113)} = 0.925. \end{aligned}$$

Q10. Twelve items are independently sampled from a production line. If the probability that any given item is defective is 0.1, the probability of at most two defectives in the sample is closest to ...

a)0.38748 b)0.9872 c)0.7361 d)0.8891 e)none of the preceding

Solution: let $p = 0.1$ denote the probability that an item is defective. Then $1 - p = 0.9$ is the probability that an item is not defective. Let X denote the number of defective items in the sample.

The probability that none of the items is defective is

$$\begin{aligned} P(X = 0) &= P(\text{item } 1 \text{ is not defective}) \times \cdots \times P(\text{item } 12 \text{ is not defective}) \\ &= (1 - p)^{12} = 0.9^{12} \approx 0.2824 \end{aligned}$$

The probability that exactly one of the items is defective is

$$P(X = 1) = P(\text{only item } 1 \text{ is defective}) + \cdots + P(\text{only item } 12 \text{ is defective}).$$

The event that only item j is defective (for $j = 1, \ldots, 12$) occurs when item j is defective (with probability p) AND the remaining 11 items are not defective (each with probability $1-p$). Since the items are sampled independently,

$$P(\text{only item } j \text{ is defective}) = p(1-p)^{11} \approx 0.0314.$$

But there are $\binom{12}{1} = \frac{12!}{1!11!} = 12$ ways to chose which of the 12 items will be defective (sampling without replacement), so

$$\begin{aligned} P(X=1) &= \underbrace{p(1-p)^{11} + \cdots + p(1-p)^{11}}_{\binom{12}{1}=12 \text{ times}} \\ &= \binom{12}{1} p(1-p)^{11} \approx 12(0.0314) \approx 0.3766. \end{aligned}$$

The probability that exactly two of the items are defective is

$$P(X=2) = P(\text{only items } 1,2 \text{ are defective}) + \cdots$$
$$+ P(\text{only item } 11,12 \text{ are defective}).$$

The event that only items j,k are defective (for $j,k=1,\ldots,12,\ j\neq k$) occurs when items $j \neq k$ are defective (each with probability p) AND the remaining 10 items are not defective (each with probability $1-p$). Since the items are sampled independently,

$$P(\text{only items } j,k \text{ are defective}) = p^2(1-p)^{10} \approx 0.0031.$$

But there are $\binom{12}{2} = \frac{12!}{2!10!} = 66$ ways to chose which 2 of the items will be defective, so

$$
\begin{aligned}
P(X=2) &= \underbrace{p^2(1-p)^{10} + \cdots + p^2(1-p)^{10}}_{\binom{12}{2}=66 \text{ times}} \\
&= \binom{12}{2} p^2(1-p)^{10} \approx 66(0.0031) = 0.2301.
\end{aligned}
$$

The probability of at most two defective items in the sample is thus

$$
\begin{aligned}
P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\
&= \binom{12}{1} p^0(1-p)^{12} + \binom{12}{2} p^1(1-p)^{11} + \binom{12}{2} p^2(1-p)^{10} \\
&\approx 0.2824 + 0.3766 + 0.2071 = 0.8891
\end{aligned}
$$

Q11. A student can solve 6 problems from a list of 10. For an exam 8 questions are selected at random from the list. What is the probability that the student will solve exactly 5 problems?

a)0.98 b)0.02 c)0.28 d)0.53 e)none of the preceding

Solution: let X and $8-X$ be the number of questions on the exam that the student can and cannot solve, respectively.

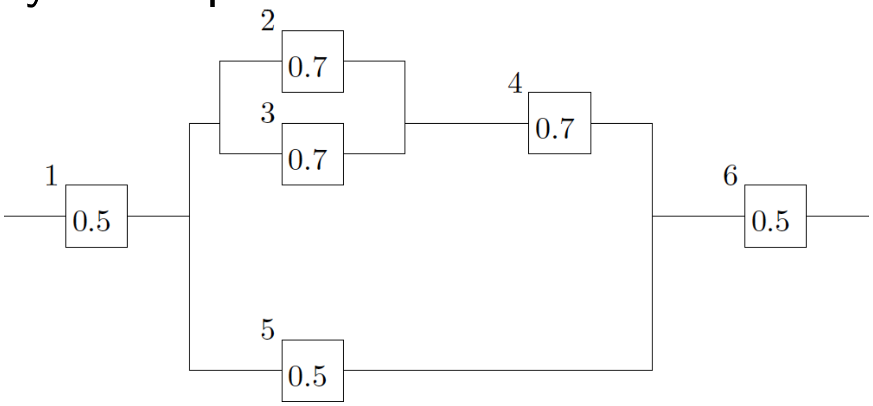
There are $\binom{6}{X}$ ways to randomly chose X exam questions that the student can solve, and $\binom{10-6}{8-X}=\binom{4}{8-X}$ ways to randomly chose $8-X$ questions that the student cannot solve: there are thus $\binom{6}{X}\binom{4}{8-X}$ ways to randomly chose X questions that the student can solve AND $8-X$ questions that the student cannot solve.

Since there are $\binom{10}{8}$ total ways to chose 8 exam questions randomly from the 10 problems,

$$P(X)=\frac{\binom{6}{X}\binom{4}{8-X}}{\binom{10}{8}}.$$

For $X=5$, $P(X=5)=\frac{\binom{6}{5}\binom{4}{3}}{\binom{10}{8}}=\frac{6\cdot 4}{45}\approx 0.53.$

Q12. Consider the following system with six components. We say that it is functional if there exists a path of functional components from left to right. The probability of each component functions is shown. Assume that the components function or fail independently. What is the probability that the system operates?



a) 0.1815 b) 0.8185 c) 0.6370 d) 0.2046 e) none of the preceding

Solution: let Box A consist of components $2,3,4,5$, Box B consist of components $2,3,4$, and Box C consist of components $2,3$. We will denote the probability that Box j operates by $P(j)$, $j\in\{A,B,C\}$.

We are interested in the probability $P(S)$ that the system operates. Because the components function or fail independently,

$$\begin{aligned}P(S)&=P(\text{component }1\text{ and Box }A\text{ and component }6\text{ operate})\\&=P(1\text{ operates})\times P(A)\times P(6\text{ operates})\\&=0.5\cdot P(A)\cdot 0.5=0.5^2P(A).\end{aligned}$$

There are two ways for Box A to operate: either component 5 operates (with probability 0.5) or Box B operates:

$$\begin{aligned}P(A) &= P(\text{component } 5 \text{ operates or Box } B \text{ operates})\\ &= P(5 \text{ operates}) + P(B)\\ &\qquad\qquad - P(\text{component } 5 \text{ operates and Box } B \text{ operates})\\ &= P(5 \text{ operates}) + P(B) - P(5 \text{ operates})P(B)\\ &= 0.5 + P(B) - 0.5P(B) = 0.5(1 + P(B)).\end{aligned}$$

Thus,

$$P(S) = 0.5^2 P(A) = 0.5^2 \cdot 0.5(1 + P(B) = 0.5^3(1 + P(B)).$$

In order for Box B to operate, we need both Box C and component 4 to operate, so that

$$\begin{aligned} P(B) &= P(\text{Box } C \text{ operates and component } 4 \text{ operates}) \\ &= P(4 \text{ operates})P(C) = 0.7P(C). \end{aligned}$$

Thus,

$$P(S) = 0.5^3(1 + P(B)) = 0.5^3(1 + 0.7P(C)).$$

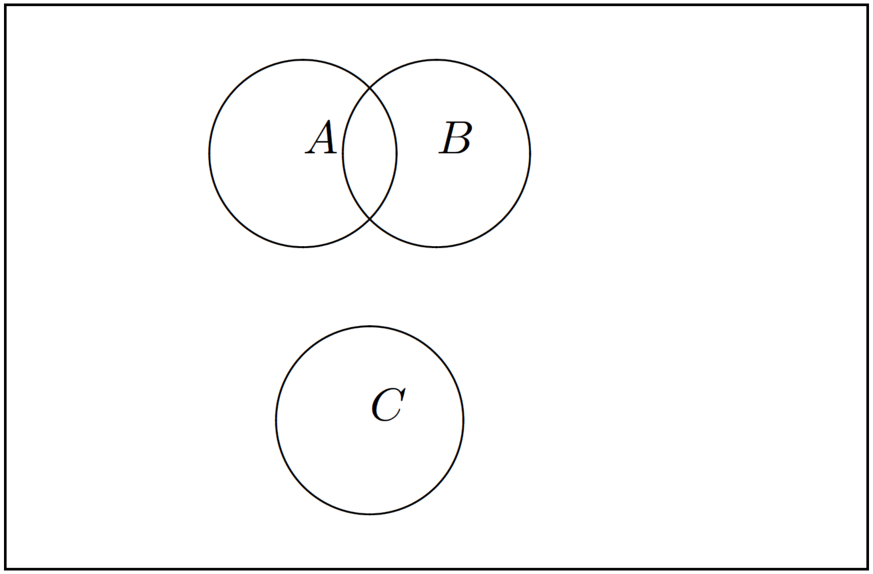
Finally, there are two ways for Box C to operate: either component 2 operates (with probability 0.7) or component 3 operates (also with probability 0.7):

$$\begin{aligned}
P(C) &= P(\text{component } 2 \text{ operates or component } 3 \text{ operates}) \\
&= P(2 \text{ operates}) + P(3 \text{ operates}) \\
&\qquad\qquad - P(\text{components } 2 \text{ operates and component } 3 \text{ operates}) \\
&= P(2 \text{ operates}) + P(3 \text{ operates}) - P(2 \text{ operates})P(3 \text{ operates}) \\
&= 0.7 + 0.7 - 0.7 \cdot 0.7 = 0.7(2 - 0.7) = 0.7 \cdot 1.3.
\end{aligned}$$

Thus,

$$\begin{aligned}
P(S) &= 0.5^3(1 + 0.7P(C)) = 0.5^3(1 + 0.7 \cdot 0.7 \cdot 1.3) \\
&= 0.5^3(1 + 0.7^2 \cdot 1.3) = 0.24046.
\end{aligned}$$

Q13. Three events are shown in the Venn diagram below.



Shade the region corresponding to the following events:

a)A^c

b)$(A \cap B) \cup (A \cap B^c)$

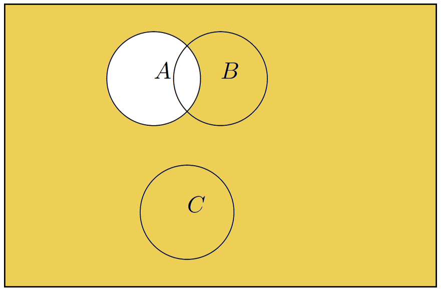
c)$(A \cap B) \cup C$

d)$(B \cup C)^c$

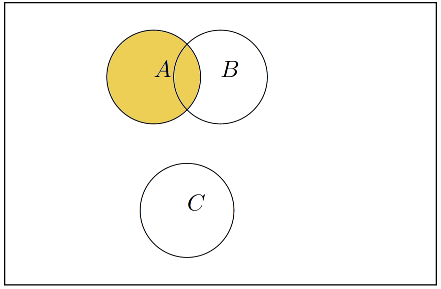
e)$(A \cap B)^c \cup C$

Solution:

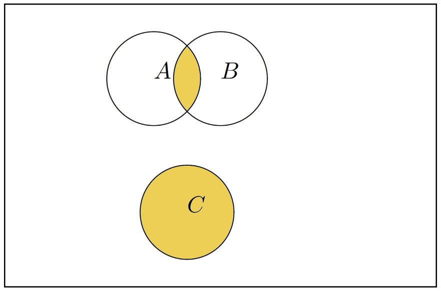
a) This is the set of all outcomes not in A:



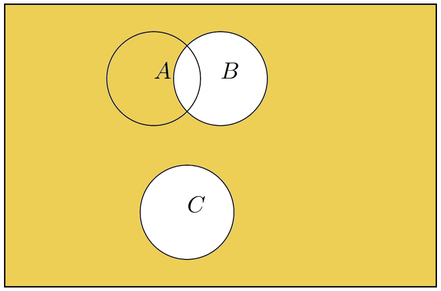
b) We can rewrite $(A \cap B) \cup (A \cap B^c) = A \cap (B \cup B^c) = A \cap \mathcal{S} = A$, so this is the set of all outcomes in A:



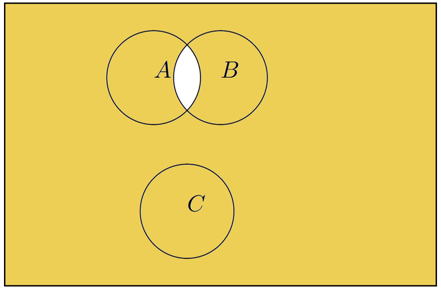
c) This is the set all outcomes either in C or in the intersection of A and B:



d) This is the set of all outcomes that are not in either B or C (or that are neither in B nor in C):



e) Since C and $A \cup B$ are mutually exclusive (disjoint), $C \subseteq (A \cap B)^c$ and $(A \cap B)^c \cup C = (A \cap B)^c$, so this is the set of all outcomes not in $A \cap B$:



Q14. Pieces of aluminum are classified according to the finishing of the surface and according to the finishing of edge. The results from 85 samples are summarized as follows:

	Edge	
Surface	excellent	good
excellent	60	5
good	16	4

Let A denote the event that a selected piece has ”excellent” surface, and let B denote the event that a selected piece has “excellent” edge. If samples are elected randomly, determine the following probabilities:

a)$P(A)$ b)$P(B)$ c)$P(A^c)$

d)$P(A \cap B)$ e)$P(A \cup B)$ f)$P(A^c \cup B)$

g) If the selected piece has excellent edge finishing, what is the probability that it has excellent surface finishing?

h) If the selected piece has good surface finishing, what is the probability that it has excellent edge finishing?

i) Are A and B independent?

Solution:

a) $P(A) = 65/85$

b) $P(B) = 76/85$

c) $P(A^c) = 1 - 65/85 = 20/85$

d) $P(A \cap B) = 60/85$

e) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 65/85 + 76/85 - 60/85 = 81/85$

f) $P(A^c \cup B) = P(A^c) + P(B) - P(A^c \cap B) = 20/85 + 76/85 - 16/85 = 80/85$

g) $P(A|B) = P(A \cap B)/P(B) = 60/76$

h) $P(B|A^c) = P(B \cap A^c)/P(A^c) = 16/20$

i) $P(A|B) \neq P(A)$ – events are not independent.

Q15. If $P(A) = 0.1$, $P(B) = 0.3$, $P(C) = 0.3$, and events A, B, C are mutually exclusive, determine the following probabilities:

a)$P(A \cup B \cup C)$ b)$P(A \cap B \cap C)$ c)$P(A \cap B)$

d)$P((A \cup B) \cap C)$ e)$P(A^c \cap B^c \cap C^c)$ f)$P[(A \cup B \cup C)^c]$

Solution:

a) $P(A \cup B \cup C) = P(A) + P(B) + P(C) = 0.7$, since the events are mutually exclusive.

b) $P(A \cap B \cap C) = 0$

c) $P(A \cap B) = 0$

d) $P((A \cup B) \cap C) = 0$

e) $P(A^c \cap B^c \cap C^c) = P[(A \cup B \cup C)^c] = 1 - P(A \cup B \cup C) = 0.3$ (draw Venn diagram)

f) $P[(A \cup B \cup C)^c] = 1 - P(A \cup B \cup C) = 0.3$

Q16. The probability that an electrical switch, which is kept in dryness, fails during the guarantee period, is 1%. If the switch is humid, the failure probability is 8%. Assume that 90% of switches are kept in dry conditions, whereas remaining 10% are kept in humid conditions.

a) What is the probability that the switch fails during the guarantee period?

b) If the switch failed during the guarantee period, what is the probability that it was kept in humid conditions?

Solution: Let F, H, D represent the event that the switch fails, that it is humid, and that it is dry, respectively. Note that H and D are mutually exclusive and exhaustive.

Given: $P(F|D) = 0.01$, $P(F|H) = 0.08$, $P(D) = 0.9$, $P(H) = 0.1$.

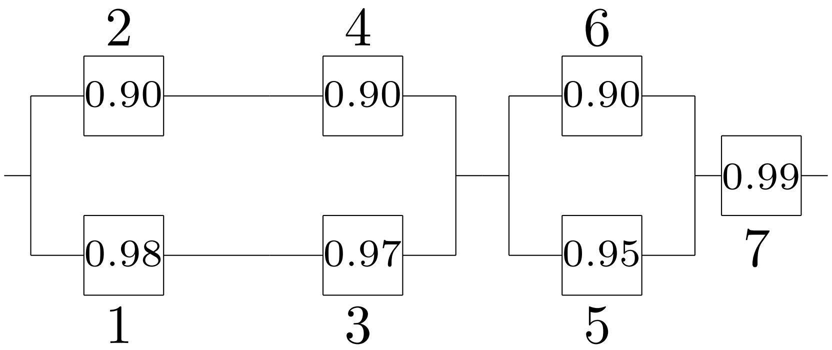
a) According to the Law of Total Probability,

$$\begin{aligned} P(F) &= P(F|D)P(D) + P(F|H)P(H) \\ &= 0.01 \cdot 0.9 + 0.08 \cdot 0.1 = 0.009 + 0.008 = 0.017. \end{aligned}$$

b) According to Bayes' Theorem,

$$P(H|F) = \frac{P(H \cap F)}{P(F)} = \frac{P(F|H)P(H)}{P(F)} = 0.4706.$$

Q17. The following system operates only if there is a path of functional device from left to the right. The probability that each device functions is as shown. What is the probability that the circuit operates? Assume independence.



Solution: let Box A consist of components $1, 2, 3, 4$; Box B of components $5, 6$; Box C of component 7. Since all components are independent,

$$P(\text{system works}) = P(A \text{ works})P(B \text{ works})P(7 \text{ works}).$$

Now, B is just a parallel system, so that it works if any of its two components work:

$$P(B \text{ works}) = 0.9 + 0.95 - 0.9 \cdot 0.95 = 1.85 - 0.855 = 0.995.$$

Furthermore, since all the components are independent of one another,

$$P(2 \text{ and } 4 \text{ work}) = P(2 \text{ works})P(4 \text{ works}) = 0.9 \cdot 0.9 = 0.81$$
$$P(1 \text{ and } 3 \text{ work}) = P(1 \text{ works})P(3 \text{ works}) = 0.98 \cdot 0.97 = 0.9506.$$

Now, A is a parallel system consisting of two independent sub-systems: $2, 4$ and $1, 3$, so that

$$\begin{aligned} P(A \text{ works}) &= P(2 \text{ and } 4 \text{ work}) + P(1 \text{ and } 3 \text{ work}) \\ &\qquad - P(2 \text{ and } 4 \text{ work})P(1 \text{ and } 3 \text{ work}) \\ &= 0.81 + 0.9506 - 0.81 \cdot 0.9506 = 0.9906. \end{aligned}$$

Thus

$$P(\text{system works}) = 0.9906 \cdot 0.995 \cdot 0.99 = 0.9758.$$

Q18. An inspector working for a manufacturing company has a 95% chance of correctly identifying defective items and 2% chance of incorrectly classifying a good item as defective. The company has evidence that 1% of the items it produces are nonconforming (defective).

1. What is the probability that an item selected for inspection is classified as defective?

2. If an item selected at random is classified as non defective, what is the probability that it is indeed good?

Solution: let A be the event that an item is classified as defective and D be the event that an item is defective; so that D^c is the event that an item is 'good'. What is known is that $P(D) = 0.01$, $P(A|D) = 0.95$, and $P(A|D^c) = 0.02$.

a) According to the the Law of Total Probability,

$$P(A) = P(A \cap D) + P(A \cap D^c) = P(A|D)P(D) + P(A|D^c)P(D^c) \approx 0.0293.$$

b) According to Bayes' Theorem,

$$P(D^c|A^c) = \frac{P(A^c|D^c)P(D^c)}{P(A^c)} = \frac{(1 - P(A|D^c))P(D^c)}{1 - P(A)} \approx 0.999.$$

Q19. Consider an ordinary 52-card North American playing deck (4 suits, 13 cards in each suit).

a) How many different $5-$card poker hands can be drawn from the deck?

b) How many different $13-$card bridge hands can be drawn from the deck?

c) What is the probability of an all-spade $5-$card poker hand?

d) What is the probability of a flush ($5-$cards from the same suit)?

e) What is the probability that a $5-$card poker hand contains exactly 3 Kings and 2 Queens?

f) What is the probability that a $5-$card poker hand contains exactly 2 Kings, 2 Queens, and 1 Jack?

Solution:

a) The number of possible $5-$card poker hands drawn from a deck of 52 playing cards is

$$ {}_{52}C_5 = \binom{52}{5} = \frac{52!}{5!47!} = 2,598,960. $$

b) The number of possible $13-$card poker hands drawn from a deck of 52 playing cards is

$$ {}_{52}C_{13} = \binom{52}{13} = \frac{52!}{13!39!} = 635,013,559,600. $$

c) The number of possible $5-$card hands that are all-spades is $N = \binom{13}{5}\binom{39}{0}$ because the 5 spades in the hand can be selected from the 13 spades in the deck in $\binom{13}{5}$ ways, after which zero non-spade can be selected in $\binom{39}{0} = 1$ way, and so $N = \binom{13}{5} \cdot 1 = \frac{13!}{5!8!} = 1287$. The probability of obtaining such a hand is thus

$$\frac{1287}{2,598,960} \approx 0.000495.$$

d) For any of the suits S, H, D, C, the probability of a $5-$card hand with all cards in the same suit has been computed above to be 0.000495. Thus,

$$\begin{aligned} P(\text{flush}) &= P(S \text{ flush}) + P(H \text{ flush}) + P(D \text{ flush}) + P(C \text{ flush}) \\ &\approx 4(0.000495) \approx 0.00198. \end{aligned}$$

e) Let A be the outcome which the hand consists of exactly 3 Ks and 2 Qs. We can select the 3 Ks in any one of $\binom{4}{3}$ ways and the 2 Qs in any one of $\binom{4}{2}$ ways, so that the number of such hands is

$$N_B = \binom{4}{3}\binom{4}{2} = \frac{4!}{1!3!} \cdot \frac{4!}{2!2!} = 4 \cdot (24/4) = 24,$$

and $P(B) = \frac{24}{2{,}598{,}960} \approx 0.0000092.$

f) Same idea, but this time

$$P(C) = \frac{\binom{4}{2}\binom{4}{2}\binom{4}{1}}{2,598,960} = \frac{144}{2,598,960} \approx 0.000055.$$

Q20. Students on a boat send messages back to shore by arranging seven coloured flags on a vertical flagpole.

a) If they have 4 orange flags and 3 blue flags, how many messages can they send?

b) If they have 7 flags of different colours, how many messages can they send?

c) If they have 3 purple flags, 2 red flags, and 4 yellow flags, how many messages can they send?

Solution:

a) The question is to find the number of distinguishable permutations of 4 Os and 3 Bs, which is to say

$$_7C_4 = \binom{7}{4} = \frac{7!}{4!3!} = 35.$$

b) If the flags were of different colours, we would be looking at the number of (distinguishable) permutations of 7 different objects, which is to say $7! = 5040$.

c) The number of distinguishable combinations of 3 Ps, 2 Rs, and 4 Ys is

$$\frac{9!}{3!4!2!} = 1260.$$

Q21. The Stanley Cup Finals in hockey or the NBA Finals in basketball continue until either the representative team form the Western Conference or from the Eastern Conference wins 4 games. How many different orders are possible ($WWEEEE$ means that the Eastern team won in 6 games) if the series goes

a)4 games? b)5 games? c)6 games? d)7 games?

Solution:

a) There are only 2 orders: $EEEE$ and $WWWW$.

b) Let's start with Finals won by the Eastern team, i.e.: 4 out of 5 of the wins were by E, with the 5th (and last) win by E. Note that this is equal to $\binom{3}{0} + \binom{3}{3}$.

In other words, only the order of the 4 first games matters, from which 3 must be wins by E (in order for the 5th game's win by E to end the Final): there are $\binom{4}{3} = 4$ ways of picking 3 wins by E in the first 4 games.

The same reasoning applies to Finals won by the Western team. There are thus 8 ways to end the Finals in 5 games. Note that this is equal to $\binom{4}{1} + \binom{4}{3}$.

c) A similar reasoning shows that there are $\binom{5}{3} = \frac{5\cdot4}{2} = 10$ ways for either the Eastern or the Western team to win in 6 games, so 20 ways in total. Note that this is equal to $\binom{5}{2} + \binom{5}{3}$.

d) A similar reasoning shows that there are $\binom{6}{3} = \frac{6\cdot5\cdot4}{6} = 20$ ways for either the Eastern or the Western team to win in 7 games, so 40 ways in total. Note that this is equal to $\binom{6}{3} + \binom{6}{3}$.

Q22. Consider an ordinary 52-card North American playing deck (4 suits, 13 cards in each suit), from which cards are drawn at random and without replacement, until 3 spades are drawn.

a) What is the probability that there are 2 spades in the first 5 draws?

b) What is the probability that a spade is drawn on the 6th draw given that there were 2 spades in the first 5 draws?

c) What is the probability that 6 cards need to be drawn in order to obtain 3 spades?

d) All the cards are placed back into the deck, and the deck is shuffled. 4 cards are then drawn from. What is the probability of having drawn a spade, a heart, a diamond, and a club, in that order?

Solution:

a) Let A be the event that there are 2 spades in the first 5 draws. Since the cards are drawn without replacement,

$$\begin{aligned} P(A) &= \frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} = \frac{13!}{2!11!}\cdot\frac{39!}{3!36!}\cdot\frac{5!47!}{52!} \\ &= \frac{13\cdot 12}{2}\cdot\frac{39\cdot 38\cdot 37}{6}\cdot\frac{120}{52\cdot 51\cdot 50\cdot 49\cdot 48} = \frac{1,026,492,480}{3,742,502,400}\approx 0.274. \end{aligned}$$

b) Let B be the event that the 6th draw was a spade. If A has already occurred, there are only 47 cards left for the 6th draw, out of which only 11 are spades. Thus $P(B|A)=\frac{11}{47}\approx 0.234.$

c) The event $A \cap B$ is the event that 6 cards need to be drawn in order to obtain 3 spades: first, there must be 2 spades in the first 5 draws (A), and once A has happened, there must be a spade on the 6th draw (B):

$$P(A \cap B) = P(B|A)P(A) \approx 0.234 \cdot 0.274 \approx 0.064.$$

d) If we assume that the first is a spade, the second is a heart, the third is a diamond, and the fourth is a club, then its probability is

$$P(S_1 \cap H_2 \cap D_3 \cap C_4) = P(S_1)P(H_2|S_1)P(D_3|S_1 \cap H_2)P(C_4|S_1 \cap H_2 \cap D_3)$$
$$= \frac{13}{52} \cdot \frac{13}{51} \cdot \frac{13}{50} \cdot \frac{13}{49} \approx 0.0044.$$

If we take all 4 cards at a time and order is not important, it is:

$$\frac{\binom{13}{1}\binom{13}{1}\binom{13}{1}\binom{13}{1}}{\binom{52}{4}} = \frac{13 \times 13 \times 13 \times 13}{(52 \times 51 \times 50 \times 49)/4!} \approx 4! \times 0.0044.$$

Q23. A student has 5 blue marbles and 4 white marbles in his left pocket, and 4 blue marbles and 5 white marbles in his right pocket. If they transfer one marble at random from their left pocket to his right pocket, what is the probability of them then drawing a blue marble from their right pocket?

Solution: for notation, let BL, BR, and WL denote drawing a blue marble from the left pocket, a blue marble from the right pocket, and a white marble from the left pocket, respectively. By the Law of Total Probability,

$$\begin{aligned} P(BR) &= P(BL \cap BR) + P(WL \cap BR) \\ &= P(BL)P(BR|BL) + P(WL)P(BR|WL) \\ &= \frac{5}{9} \cdot \frac{5}{10} + \frac{4}{9} \cdot \frac{4}{10} = \frac{41}{90} \approx 0.456. \end{aligned}$$

Q24. An insurance company sells a number of different policies; among these, 60% are for cars, 40% are for homes, and 20% are for both. Let A_1, A_2, A_3, A_4 represent people with only a car policy, only a home policy, both, or neither, respectively. Let B represent the event that a policyholder renews at least one of the car or home policies.

a) Compute $P(A_1)$, $P(A_2)$, $P(A_3)$, and $P(A_4)$.

b) From past data, we know that $P(B|A_1) = 0.6$, $P(B|A_2) = 0.7$, $P(B|A_3) = 0.8$. Given that a client selected at random has a car or a home policy, what is the probability that they will renew one of these policies?

Solution:

a) $P(A_1) = 0.4$, $P(A_2) = 0.2$, $P(A_3) = 0.2$. Because the events are mutually excl. & exhaustive, $P(A_4) = 1 - P(A_1) - P(A_2) - P(A_3) = 0.2$.

b) The event $A_1 \cup A_2 \cup A_3$ represent those clients who have a car policy, a home policy, or both. Thus,

$$\begin{aligned} P(B|A_1 \cup A_2 \cup A_3) &= \frac{P(B \cap (A_1 \cup A_2 \cup A_3))}{P(A_1 \cup A_2 \cup A_3)} \\ &= \frac{P((B \cap A_1) \cup (B \cap A_2) \cup (B \cap A_3))}{P(A_1 \cup A_2 \cup A_3)}. \end{aligned}$$

But A_1, A_2, A_3 are mutually exclusive, so that $B \cap A_1, B \cap A_2, B \cap A_3$ are also mutually exclusive, and

$$\begin{aligned}
P(B|A_1 \cup A_2 \cup A_3) &= \frac{P(B \cap A_1) + P(B \cap A_2) + P(B \cap A_3)}{P(A_1) + P(A_2) + P(A_3)} \\
&= \frac{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)}{P(A_1) + P(A_2) + P(A_3)} \\
&= \frac{0.6 \cdot 0.4 + 0.7 \cdot 0.2 + 0.8 \cdot 0.2}{0.4 + 0.2 + 0.2} = \frac{0.54}{0.80} = 0.675.
\end{aligned}$$

Q25. An urn contains four balls numbered 1 through 4. The balls are selected one at a time, without replacement. A match occurs if ball m is the mth ball selected. Let the event A_i denote a match on the ith draw, $i = 1, 2, 3, 4$.

a) Compute $P(A_i)$, $i = 1, 2, 3, 4$.

b) Compute $P(A_i \cap A_j)$, $i, j = 1, 2, 3, 4$, $i \neq j$.

c) Compute $P(A_i \cap A_j \cap A_k)$, $i, j, k = 1, 2, 3, 4$, $i \neq j, i \neq k, j \neq k$.

d) What is the probability of at least 1 match?

Solution: (difficult!)

a) $P(A_i) = \frac{3!}{4!}$.

b) $P(A_i \cap A_j) = \frac{2!}{4!}$.

c) $P(A_i \cap A_j \cap A_k) = \frac{1!}{4!}$.

d) $P(A_1 \cup A_2 \cup A_3 \cup A_4) = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!}$.

Q26. The probability that a company's workforce has at least one accident in a given month is $(0.01)k$, where k is the number of days in the month. Assume that the number of accidents is independent from month to month. If the company's year starts on January 1, what is the probability that the first accident occurs in April?

Solution: for any month X, let X represent the event that an accident takes place during month X. If the monthly probabilities are independent of one another, then

$$\begin{aligned} P(J^c \cap F^c \cap M^c \cap A) &= P(J^c)P(F^c)P(M^c)P(A) \\ &= (1-31(0.01)) \cdot (1-28(0.01)) \cdot (1-31(0.01)) \cdot 30(0.01) \\ &= 0.69 \cdot 0.72 \cdot 0.69 \cdot 0.30 \approx 0.103. \end{aligned}$$

Q27. A Pap smear is a screening procedure used to detect cervical cancer. Let T^- and T^+ represent the events that the test is negative and positive, respectively, and let C represent the event that the person tested has cancer.

- Among patients that do not have the disease, the test reports a 'positive' with probability 0.19;
- Among patients that have the disease, the test reports a 'negative' with prob 0.16.

In North America, the rate of incidence for this cancer is roughly 8 out of 100,000 women. Based on these numbers, do you think that the Pap smear is an effective procedure? What factors influence your conclusion?

Solution: from the statement of the problem, we have

$$P(T^-|C) = 0.16, \quad P(T^+|C) = 0.84, \quad P(T^+|C^c) = 0.19,$$
$$P(T^-|C^c) = 0.81, \quad P(C) = 0.00008, \quad P(C^c) = 0.99992.$$

According to Bayes' Theorem and the Law of Total Probability,

$$\begin{aligned} P(C|T^+) &= \frac{P(T^+|C)P(C)}{P(T^+)} = \frac{P(T^+|C)P(C)}{P(T^+|C)P(C) + P(T^+|C^c)P(C^c)} \\ &= \frac{0.84 \cdot 0.00008}{0.84 \cdot 0.00008 + 0.19 \cdot 0.99992} \approx 0.0000354. \end{aligned}$$

For every million positive Pap smears, only 354 represent true cases of cervical cancer. The procedure is ineffective because the cancer rate is small, and because the error rates of the procedure are relatively high.

Q28. Of three different fair dice, one each is given to Elowyn, Llewellyn, and Gwynneth. They each roll the die they received.

Let $E = \{\text{Elowyn rolls a } 1 \text{ or a } 2\}$, $LL = \{\text{Llewellyn rolls a } 3 \text{ or a } 4\}$, and $G = \{\text{Gwynneth rolls a } 5 \text{ or a } 6\}$ be 3 events of interest.

a) What are the probabilities of each of E, LL, and G occurring?

b) What are the probabilities of any two of E, LL, and G occurring simultaneously?

c) What is the probability of all three of the events occurring simultaneously?

d) What is the probability of at least one of E, LL, or G occurring?

Solution:

a) $P(E) = P(LL) = P(G) = 1/3.$

b) $P(E \cap LL) = P(LL \cap G) = P(G \cap E) = 1/9.$

c) $P(E \cap LL \cap G) = 1/27.$

d) Using De Morgan's Law and assuming that the events are independent,

$$\begin{aligned} P(E \cup LL \cup G) &= 1 - P((E \cup LL \cup G)^c) = 1 - P(E^c \cap LL^c \cap G^c) \\ &= 1 - P(E^c)P(LL^c)P(G^c) \\ &= 1 - (1 - P(E))(1 - P(LL))(1 - P(G)) \\ &= 1 - (1 - 1/3)^3 = 1 - 8/27 \approx 0.704. \end{aligned}$$

Q29. Over the course of two baseball seasons, player A obtained 126 hits in 500 at-bats in Season 1, and 90 hits in 300 at-bats in Season 2; player B, on the other hand, obtained 75 hits in 300 at-bats in Season 1, and 145 hits in 500 at-bats in Season 2. A player's batting average is the number of hits they obtain divided by the number of at-bats.

a) Which player has the best batting average in Season 1? In Season 2?

b) Which player has the best batting average over the 2-year period?

c) What is happening here?

Solution:

a) In season 1, player A's batting average is $\frac{126}{500} = 0.252$; player B's is $\frac{75}{300} = 0.250$. In season 2, player A's batting average is $\frac{90}{300} = 0.300$; player B's is $\frac{145}{500} = 0.290$. In both seasons, player A has a stronger batting average.

b) Over the 2 seasons, player A's batting average is $\frac{126+90}{500+300} = 0.270$, while player B's batting average is $\frac{75+145}{300+500} = 0.275$; player B has a stronger battin average.

c) Simpson's paradox!

Q30. A stranger comes to you and shows you what appears to be a normal coin, with two distinct sides: Heads (H) and Tails (T). They flip the coin 4 times and record the following sequence of tosses: $HHHH$.

a) What is the probability of obtaining this specific sequence of tosses? What assumptions do you make along the way in order to compute the probability? What is the probability that the next toss will be a T.

b) The stranger offers you a bet: they will toss the coin another time; if the toss is T, they give you $100\$$, but if it is H, you give them $10\$$. Would you accept the bet (if you are not morally opposed to gambling)?

c) Now the stranger tosses the coin 60 times and records $60 \times H$ in a row: $H \cdots H$. They offer you the same bet. Do you accept it?

d) What if they offered $1000\$$ instead? $1,000,000\$$?

Solution:

a) $(1/2)^4 = 0.0625$; independence of tosses and fairness of coin; $1/2 = 0.5$.

b) The bet is to your advantage: half the time you will receive $100\$$, half the time you will lose $10\$$; on average, you receive more than you lose if you take the bet.

c) You know what? Even though it's theoretically possible for a fair coin to be tossed independently 60 times and to come up H 60 times in a row, the probability of this happening is so vanishingly small at $(1/2)^{60} = 8.7 \times 10^{-19}$ that I don't actually believe that my assumptions were warranted: either the coin is not fair, or the tosses are not independent. At any rate, I smell a rat and I do not take the bet.

Q31. The sample space of a random experiment is $\{a, b, c, d, e, f\}$ and each outcome is equally likely. A random variable is defined as follows

outcome	a	b	c	d	e	f
X	0	0	1.5	1.5	2	3

Determine the probability mass function of X. Determine the following probabilities:

a)$P(X = 1.5)$ b)$P(0.5 < X < 2.7)$ c)$P(X > 3)$

d)$P(0 \leq X < 2)$ e)$P(X = 0 \text{ or } 2)$

Solution: the probability mass function is

$$P(X=0)=P(\{a,b\})=\tfrac{1}{6}+\tfrac{1}{6}=\tfrac{1}{3},\ P(X=1.5)=P(\{c,d\})=\tfrac{1}{3},$$
$$P(X=2)=P(\{e\})=\tfrac{1}{6},\ P(X=3)=P(\{f\})=\tfrac{1}{6}.$$

a) $P(X=1.5)=\frac{2}{6}=\frac{1}{3}$

b) $P(0.5<X<2.7)=P(X=1.5)+P(X=2)=\frac{3}{6}=\frac{1}{2}$

c) $P(X>3)=0$

d) $P(0\leq X<2)=P(X=0)+P(X=1.5)=\frac{4}{6}=\frac{2}{3}$

e) $P(X=0\text{ or }X=2)=\frac{3}{6}=\frac{1}{2}$

Q32. Determine the mean and the variance of the random variable defined in **Q1**.

Solution: we have

$$\begin{aligned}
\mathrm{E}[X] &= \sum_i X_i P(X = X_i) \\
&= 0 \cdot P(X=0) + 1.5 \cdot P(X=1.5) + 2 \cdot P(X=2) + 3 \cdot P(X=3) \\
&= 0 \cdot \tfrac{1}{3} + 1.5 \cdot \tfrac{1}{3} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} = \tfrac{4}{3} \approx 1.33 \\
\mathrm{Var}[X] &= \sum_i (X_i - \mathrm{E}[X])^2 P(X = X_i) \\
&= (0 - \tfrac{4}{3})^2 P(0) + (1.5 - \tfrac{4}{3})^2 P(1.5) + (2 - \tfrac{4}{3})^2 P(2) + (3 - \tfrac{4}{3})^2 P(3) \\
&= (\tfrac{4}{3})^2 \cdot \tfrac{1}{3} + (\tfrac{1}{6})^2 \cdot \tfrac{1}{3} + (\tfrac{2}{3})^2 \cdot \tfrac{1}{6} + (\tfrac{5}{3})^2 \cdot \tfrac{1}{6} = 41/36 \approx 1.39.
\end{aligned}$$

Q33. We say that X has **uniform distribution** on a set of values $\{X_1, \ldots, X_k\}$ if

$$P(X = X_i) = \frac{1}{k}, \qquad i = 1, \ldots, k.$$

The thickness measurements of a coating process are **uniformly distributed** with values 0.15, 0.16, 0.17, 0.18, 0.19. Determine the mean and variance of the thickness measurements.

Is this result compatible with a uniform distribution?

Solution: First, note that for $X_i \in \{0.15, 0.16, 0.17, 0.18, 0.18.0.19\}$ we have $P(X_i) = P(X = X_i) = 1/5$. The mean is

$$\mathrm{E}[X] = \sum_{i=1}^{5} X_i P(X_i) = \frac{1}{5}(0.15 + 0.16 + 0.17 + 0.18 + 0.19) = 0.17.$$

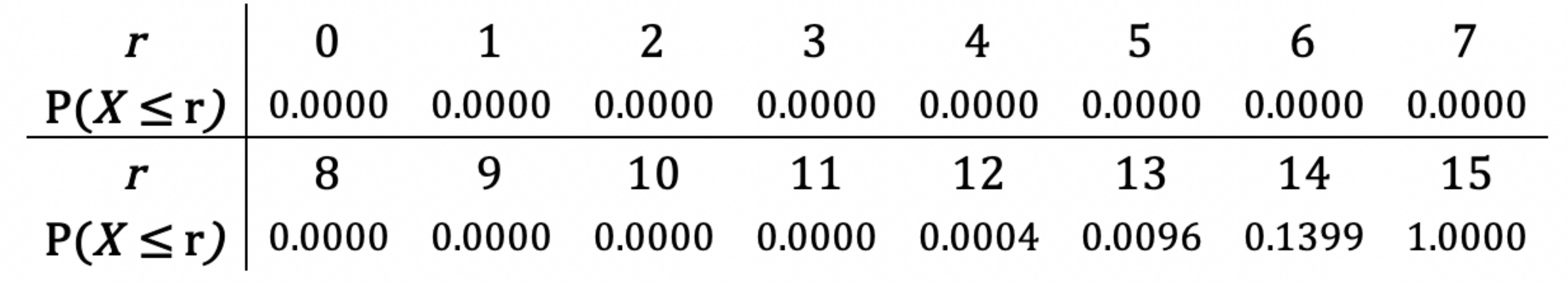
In that case, the variance is

$$\begin{aligned}\mathrm{Var}[X] &= \sum_{i=1}^{5} (X_i - \mathrm{E}[X])^2 P(X_i) \\ &= \frac{1}{5}\left((0.15 - 0.17)^2 + \cdots + (0.19 - 0.17)^2\right) \\ &= \frac{1}{5}\left(0.02^2 + 0.01^2 + 0^2 + 0.01^2 + 0.02^2\right) = 0.0002\end{aligned}$$

Q34. Samples of rejuvenated mitochondria are mutated in 1% of cases. Suppose 15 samples are studied and that they can be considered to be independent (from a mutation standpoint). Determine the following probabilities:

a) no samples are mutated;
b) at most one sample is mutated, and
c) more than half the samples are mutated.

Use the following CDF table for the $\mathcal{B}(n,p)$, with $n = 15$ and $p = 0.99$:



r	0	1	2	3	4	5	6	7
$P(X \le r)$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
r	8	9	10	11	12	13	14	15
$P(X \le r)$	0.0000	0.0000	0.0000	0.0000	0.0004	0.0096	0.1399	1.0000

Solution: let X be the number of non-mutated samples; then X has binomial distribution with $n = 15$ and $p = 0.99$ (no mutation is considered a success; note that if we chose mutation as a trial success, then $p = 0.01$ and we cannot use the Table):

a) 0 mutated sample $=$ 15 non-mutated samples, thus we need to evaluate $P(X = 15) = P(X \leq 15) - P(X \leq 14) = 1 - 0.1399 \approx 0.8601;$

b) at most 1 mutated sample $=$ at least 14 non-mutated samples, thus we need to evaluate

$$P(X \geq 14) = 1 - P(X < 14) = 1 - P(X \leq 13) = 1 - 0.0096 = 0.9904;$$

c) more than half the samples are mutated $=$ fewer than (or exactly) half the samples are non-mutated; $P(X \leq 7.5) = P(X \leq 7) = 0.0000.$

In R, the cumulative binomial distribution function for r or fewer successes among n trials (each with probability p) is given by the function `pbinom(r,size=n,prob=p)`. The probability for exactly r successes is `dbinom(r,size=n,prob=p)` (if the order is preserved, you can remove the strings `size=` and `prob=`).

a)

```
> pbinom(15,15,0.99)-pbinom(14,15,0.99) = 0.8600584
> dbinom(15,15,0.99) = 0.8600584
```

b)

```
> pbinom(15,15,0.99)-pbinom(13,15,0.99) = 0.9903702
> dbinom(14,15,0.99) + dbinom(15,15,0.99) = 0.9903702
```

c)

```
> pbinom(7.5,15,0.99) = 6.045248e-13
> pbinom(7,15,0.99) = 6.045248e-13
```

Q35. Samples of 20 parts from a metal punching process are selected every hour. Typically, 1% of the parts require re-work. Let X denote the number of parts in the sample that require re-work. A process problem is suspected if X exceeds its mean by more than three standard deviations.

a) What is the probability that there is a process problem?

b) If the re-work percentage increases to 4%, what is the probability that X exceeds 1?

c) If the re-work percentage increases to 4%, what is the probability that X exceeds 1 in at least one of the next five sampling hours?

Solution:

a) We have $X \sim \mathcal{B}(n,p)$, $n = 20$, $p = 0.01$ (where a trial success = a part requires re-work). By definition of the binomial distribution, $\mathrm{E}[X] = np = 0.2$, $\mathrm{Var}[X] = np(1-p) = 0.2 \times 0.99 = 0.198$, and $\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]} \approx 0.44$.

What we want to compute is $P(X > \mathrm{E}[X] + 3 \cdot \mathrm{SD}[X])$. Thus,

$$\begin{aligned} P(X > 0.2 + 3 \cdot 0.44) &= P(X > 1.535) = P(X \geq 2) \\ &= 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - \binom{20}{0} 0.01^0 \cdot 0.99^{20} - \binom{20}{1} 0.01^1 \cdot 0.99^{19} \\ &(= 1 - \texttt{pbinom}(1, 20, 0.01)) \approx 0.017. \end{aligned}$$

Alternatively, you may work with $Y \sim \mathcal{B}(n, 0.99)$, where Y is the $\#$ of samples which do not require re-work.

b) We have the same set-up, but with $p = 0.04$. We are still interested in

$$P(X > 1) = P(X \geq 2) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1))$$

$$= 1 - \binom{20}{0} 0.04^0 \cdot 0.96^{20} - \binom{20}{1} 0.04^1 \cdot 0.96^{19}$$

$$(= \mathtt{1 - pbinom(1, 20, 0.04)}) \approx 0.19.$$

c) Now, we have 5 hourly samples, and each of them consists of 20 items. let W be the number of hourly samples where the number of items which require re-work is larger than 1 (where a trial success $=$ hourly sample has more than 1 defective item, i.e. $X > 1$).

We have $W \sim \mathcal{B}(5, p_0)$, where $p_0 = P(X > 1) = 0.19$ is the probability of a trial success. Thus, we need to evaluate

$$P(W \geq 1) = 1 - P(W = 0) = 1 - \binom{5}{0} 0.19^0 \cdot 0.81^5 \approx 0.651.$$

This example highlights the procedure for problems of this nature: we identify the appropriate distribution model; we evaluate its parameters; we identify the appropriate probability to evaluate, and we compute its value.

Q36. In a clinical study, volunteers are tested for a gene that has been found to increase the risk for a particular disease. The probability that the person carries a gene is 0.1.

a) What is the probability that 4 or more people will have to be tested in order to detect 1 person with the gene?

b) How many people are expected to be tested in order to detect 1 person with the gene?

c) How many people are expected to be tested before 2 with the gene are detected?

Solution: if X is the number of tests before the 1^{st} success (gene detection), then X has geometric distribution with $p = 0.1$

a) In this case, we want to evaluate

$$P(X \geq 4) = \sum_{k=4}^{\infty} (1-p)^{k-1} p = p \frac{(1-p)^3}{1-(1-p)} = (1-p)^3 = 0.729.$$

b) By definition, $\mathrm{E}[X] = 1/p = 1/0.1 = 10.$

c) We can think of this procedure as splitting the patients into 2 groups, randomly, and testing each of the groups one by one. It takes on average 10 tests before the gene is detected in either one of the groups, so 20 for the two combined groups.

Q37. The number of failures of a testing instrument from contaminated particles on the product is a Poisson random variable with a mean of 0.02 failure per hour.

a) What is the probability that the instrument does not fail in an $8-$hour shift?

b) What is the probability of at least 1 failure in a $24-$hour day?

Solution:

a) Let X be the number of failures per $8-$hour shift. The failure rate per 8 hours is $8 \times 0.02 = 0.16$.

If X is Poisson random variable with $\lambda = 0.16$, we want to evaluate $P(X = 0) = \exp(-0.16) = \texttt{ppois}(0, 0.16) \approx 0.85$.

b) In this case, Y is the number of failures over a $24-$hour day. The failure rate over a full day is $24 \times 0.02 = 0.48$.

If Y is Poisson with $\lambda = 0.48$, we want to evaluate

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = 1 - \exp(-0.48) \\ &= \texttt{1} - \texttt{ppois}(0, 0.48) \approx 0.38. \end{aligned}$$

Q38. Use `R` to generate a sample from a binomial distribution and from a Poisson distribution (select parameters as you wish).

Use `R` to compute the sample means and sample variances. Compare these values to population means and population variances.

Solution: samples from the desired distributions are generated by the functions `rbinom(n,size,prob)` and `rpois(n,lambda)`.

We simulate `n=1000` values for each sample. **WARNING:** in the course, n is used to represent the number of trials; in `R`, n is used to represent the sample size, and the number of trials is represented by the parameter `size`. It's not what I would have chosen.

For the binomial distribution X, we use `size=20` and `prob=0.2`; for the Poisson distribution Y, we use `lambda=8.5`.

The true underlying means and variances are

$$\mathrm{E}[X] = 20(0.2) = 4, \quad \mathrm{Var}[X] = 20(0.2)(0.8) = 3.2$$

$$\mathrm{E}[Y] = \mathrm{Var}[Y] = 8.5.$$

The samples and estimated values agree with the theoretical ones:

```
X=rbinom(1000,20,0.2);
mean(X); var(X);
[1] 4.022 [1] 3.0745901
```

```
Y=rpois(1000,8.5);
mean(Y); var(Y);
[1] 8.388 [1] 9.090547
```

Agreement is a broad term to use. What does it mean, in this context? We'll revisit this at a later stage.

Q39. A container of 100 light bulbs contains 5 bad bulbs. We draw 10 bulbs without replacement. Find the probability of drawing at least 1 defective bulb.

a)0.4164 b)0.584 c)0.1 d)0.9 e)none of the preceding

Solution: although this sounds like it could be a binomial experiment, it is in fact just a classical probability question; we are not working with independent trials because the bulbs are sampled WITHOUT replacement).

Let X be the number of defective bulbs. We want to evaluate

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0).$$

But

$$P(X = 0) = \frac{\binom{95}{10}\binom{5}{0}}{\binom{100}{10}} = 0.584,$$

so $P(X \geq 1) = 1 - 0.584 = 0.416.$

Q40. Let X be a disrete random variable with range $\{0, 1, 2\}$ and probability mass function (p.m.f.) given by $f(0) = 0.5$, $f(1) = 0.3$, and $f(2) = 0.2$. The expected value and variance of X are, respectively,

a)0.7, 0.61 b)0.7, 1.1 c)0.5, 0.61 d)0.5, 1.1 e)none of the preceding

Solution: by definition,

$$\mathrm{E}[X] = \sum_i X_i f(X_i) = 0 \cdot 1.5 + 1 \cdot 0.3 + 2 \cdot 0.2 = 0.7$$

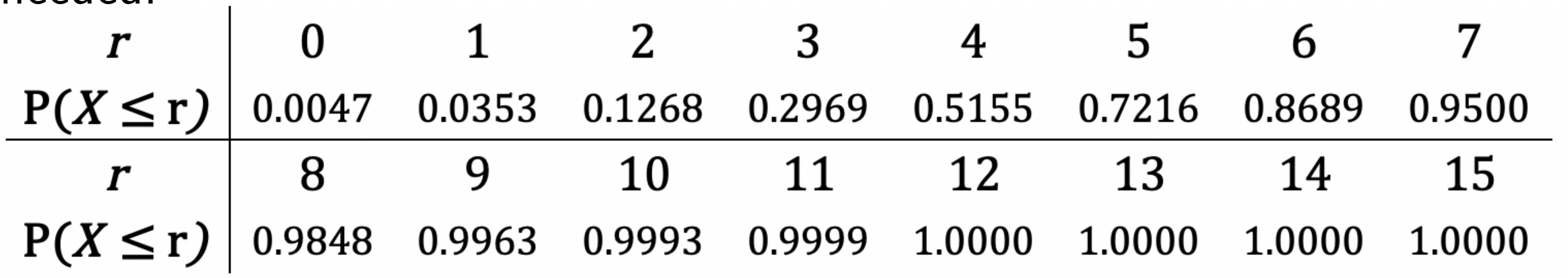
and

$$\begin{aligned}\mathrm{Var}[X] &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \sum_i X_i^2 f(X_i) - (\mathrm{E}[X])^2 \\ &= 0^2 \cdot 1.5 + 1^2 \cdot 0.3 + 2^2 \cdot 0.2 - 0.7^2 = 0.61.\end{aligned}$$

Q41. A factory employs several thousand workers, of whom 30% are not from an English-speaking background. If 15 members of the union executive committee were chosen from the workers at random, evaluate the probability that exactly 3 members of the committee are not from an English-speaking background.

a)0.17 b)0.83 c)0.98 d)0.51 e)none of the preceding

Use the following CDF table for the $\mathcal{B}(n,p)$, with $n = 15$ and $p = 0.30$ if needed:



r	0	1	2	3	4	5	6	7
$P(X \leq r)$	0.0047	0.0353	0.1268	0.2969	0.5155	0.7216	0.8689	0.9500
r	8	9	10	11	12	13	14	15
$P(X \leq r)$	0.9848	0.9963	0.9993	0.9999	1.0000	1.0000	1.0000	1.0000

Solution: let X be the number of executive committee members not from an English-speaking background. Because the number of employees is large, and the number of committee members is relatively small, we can treat the act of selecting committee members as (approximately) independent trials. Then $X \sim \mathcal{B}(15, 0.3)$ and we want to evaluate

$$P(X = 3) = \binom{15}{3} 0.3^3 \cdot 0.7^{12} \approx 0.17.$$

Alternatively, we could use the table and compute:

$$P(X = 3) = P(X \leq 3) - P(X \leq 2) = 0.2969 - 0.1268 = 0.1701.$$

Q42. Assuming the context of **Q11**, what is the probability that a majority of the committee members do not come from an English-speaking background?

Solution: let X be as in **Q11**. We want to evaluate

$$P(X > 7.5) = P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.9500 = 0.0500.$$

Q43. In a video game, a player is confronted with a series of opponents and has an 80% probability of defeating each one. Success with any opponent (that is, defeating the opponent) is independent of previous encounters. The player continues until defeated. What is the probability that the player encounters at least three opponents?

a)0.8 b)0.64 c)0.5 d)0.36 e)none of the preceding

You may need to use the following formula:

$$\sum_{x=k}^{\infty} q^x = \frac{q^k}{1-q}, \quad 0 < q < 1.$$

Solution: let X be the number of encountered opponents (including the last one, which is a loss by the player); X follows a geometric distribution, where a trial success is a loss against an opponent.

The probability of success is $p = 0.2$. Since X is geometric, and since the player faces at least 1 encounter,

$$P(X = x) = (1 - p)^{x-1}p, \quad x = 1, 2, \ldots .$$

We want to evaluate

$$P(X \geq 3) = \sum_{x=3}^{\infty}(1 - p)^{x-1}p = p\frac{(1 - p)^2}{p} = (1 - p)^2 = 0.64.$$

Q44. Assuming the context of **Q13**, how many encounters is the player expected to have?

a) 5 b) 4 c) 8 d) 10 e) none of the preceding

Solution: let X be as in **Q13**. According to the expectation formula for a geometric distribution, $\mathrm{E}[X] = 1/p = 1/0.2 = 5$.

Q45. From past experience it is known that 3% of accounts in a large accounting company are in error. The probability that exactly 5 accounts are audited before an account in error is found, is:

a)0.242 b)0.011 c)0.030 d)0.026 e)none of the preceding

Solution: let X be the number of accounts that need to be audited before an account in error is found. Then X follows a geometric distribution with a probability of success $p = 0.03$. We want to evaluate

$$\begin{aligned} P(X = 5) &= P(\text{First } 4 \text{ are not in error})P(5\text{th is in error}) \\ &= 0.97^4 \cdot 0.03 \approx 0.026. \end{aligned}$$

Q46. A receptionist receives on average 2 phone calls per minute. Assume that the number of calls can be modeled using a Poisson random variable. What is the probability that he does not receive a call within a $3-$minute interval?

a)e^{-2} b)$e^{-1/2}$ c)e^{-6} d)e^{-1} e)none of the preceding

Solution: let X be the number of calls received over a $3-$minute interval. The call rate per $3-$minute interval is $3 \times 2 = 6$.

If X is a Poisson random variable with $\lambda = 6$, we want to evaluate

$$P(X = 0) = \exp(-6) = e^{-6} \approx 0.0025$$

.

Q47. Consider a random variable X with probability density function (p.d.f.) given by

$$f(x) = \begin{cases} 0 & \text{if } x \leq -1 \\ 0.75(1-x^2) & \text{if } -1 \leq x < 1 \\ 0 & \text{if } x \geq 1 \end{cases}$$

What is the expected value and the standard deviation of X?

a) $0,\ 3$ b) $0,\ 0.447$ c) $1,\ 0.2$ d) $1,\ 3$ e) none of the preceding

Solution: the expected value of X is given by

$$\begin{aligned}\mathrm{E}[X] &= \int_{-\infty}^{\infty} xf(x)\,dx = 0.75\int_{-1}^{1} x(1-x^2)\,dx = 0.75\int_{-1}^{1}(x-x^3)\,dx \\ &= 0.75\left[\frac{x^2}{2}-\frac{x^4}{4}\right]_{-1}^{1} = 0.75\left[\left(\frac{1}{2}-\frac{1}{4}\right)-\left(\frac{1}{2}-\frac{1}{4}\right)\right] = 0\end{aligned}$$

The standard deviation is

$$\begin{aligned}\mathrm{SD}[X] &= \sqrt{\int_{-\infty}^{\infty}(x-E[X])^2 f(x)\,dx} = \sqrt{0.75\int_{-1}^{1}(x^2-x^4)\,dx} \\ &= \sqrt{0.75\left[\frac{x^3}{3}-\frac{x^5}{5}\right]_{-1}^{1}} = \sqrt{2(0.75)\left[\frac{1}{3}-\frac{1}{5}\right]} = \sqrt{0.2} \approx 0.447.\end{aligned}$$

Q48. A random variable X has a cumulative distribution function (c.d.f.)

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x/2 & \text{if } 0 < x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

What is the mean value of X?

a) 1 b) 2 c) 0 d) 0.5 e) none of the preceding

Solution: the corresponding p.d.f. is

$$f(x) = F'(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1/2 & \text{if } 0 < x < 2 \\ 0 & \text{if } x \geq 2 \end{cases}$$

Therefore,

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\, dx = 0.5 \int_0^2 x\, dx = 0.5 \left[\frac{x^2}{2}\right]_0^2 = 0.5 \left[\frac{2^2}{2} - \frac{0^2}{2}\right] = 1.$$

Q49. Let X be a random variable with p.d.f. given by $f(x) = \frac{3}{2}x^2$ for $-1 \leq x \leq 1$, and $f(x) = 0$ otherwise. Find $P(X^2 \leq 0.25)$.

a)0.250 b)0.125 c)0.500 d)0.061 e)none of the preceding

Solution: the corresponding distribution function is

$$F(x) = \int_{-1}^{x} f(t)\,dt = \begin{cases} 0 & \text{if } x \leq -1 \\ \frac{x^3}{2} + \frac{1}{2} & \text{if } -1 \leq x \leq 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

(You should verify that $F(x)$ is continuous.) In that case,

$$\begin{aligned} P(X^2 \leq 0.25) &= P(-0.5 \leq X \leq 0.5) = F(0.5) - F(-0.5) \\ &= \left(\frac{0.5^3}{2} + \frac{1}{2}\right) - \left(\frac{(-0.5)^3}{2} + \frac{1}{2}\right) = 0.125. \end{aligned}$$

Q50. In the inspection of tin plate produced by a continuous electrolytic process, 0.2 imperfections are spotted per minute, on average. Find the probability of spotting at least 2 imperfections in 5 minutes. Assume that we can model the occurrences of imperfections as a Poisson process.

a)0.736 b)0.264 c)0.632 d)0.368 e)none of the preceding

Solution: let X be the number of imperfections found in 5 minutes. The imperfection spotting rate per 5 minutes is $0.2 \times 5 = 1$.

If X is a Poisson random variable with $\lambda = 1$, we want to evaluate

$$\begin{aligned} P(X \geq 2) &= 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1)) \\ &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - \exp(-\lambda) - \lambda \exp(-\lambda) \\ &= 1 - 2\exp(-1) \approx 0.264. \end{aligned}$$

Q51. If $X \sim \mathcal{N}(0,4)$, the value of $P(|X| \geq 2.2)$ is (using the normal table):

a)0.2321 b)0.8438 c)0.2527 d)0.2713 e)0.7286 f) none of the preceding

Solution: let $Z = \frac{X-0}{\sqrt{4}}$. Then $Z \sim \mathcal{N}(0,1)$ and we have

$$\begin{aligned} P(|X| \geq 2.2) &= 1 - P(|X| \leq 2.2) = 1 - P(-2.2 \leq X \leq 2.2) \\ &= 1 - P\left(\frac{-2.2-0}{\sqrt{4}} \leq \frac{X-0}{\sqrt{4}} \leq \frac{2.2-0}{\sqrt{4}}\right) \\ &= 1 - P(-1.1 \leq Z \leq 1.1) \\ &= 1 - (\Phi(1.1) - \Phi(-1.1)) \\ &= 1 - (\texttt{pnorm}(1.1,0,1) - \texttt{pnorm}(-1.1,0,1)) \approx 0.2713. \end{aligned}$$

This could have been computed directly as

$$1 - P(-2.2 \leq X \leq 2.2) = 1 - (\texttt{pnorm}(2.2,0,2) - \texttt{pnorm}(-2.2,0,2)).$$

Q52. If $X \sim \mathcal{N}(10, 1)$, the value of k such that $P(X \leq k) = 0.701944$ is closest to

a)0.59 b)0.30 c)0.53 d)10.53 e)10.30 f)10.59

Solution: let $Z = \frac{X-10}{\sqrt{1}} = X - 10$. Then $Z \sim \mathcal{N}(0,1)$ and we have

$$P(X \leq k) = P\left(\frac{X-10}{1} \leq \frac{k-10}{1}\right) = P(Z \leq k-10) = 0.701944.$$

According to the table (or `R`),

$$k - 10 = \Phi^{-1}(0.701944) = \texttt{qnorm}(0.701944, 0, 1) \approx 0.53,$$

whence $k \approx 10.53$.

Q53. The time it takes a supercomputer to perform a task is normally distributed with mean 10 milliseconds and standard deviation 4 milliseconds. What is the probability that it takes more than 18.2 milliseconds to perform the task? (use the normal table or `R`).

a)0.9798 b)0.8456 c)0.0202 d)0.2236 e)0.5456 f) none of the preceding

Solution: let $X \sim \mathcal{N}(10, 4^2)$ and $Z = \frac{X-10}{4}$. Then $Z \sim \mathcal{N}(0,1)$ and

$$\begin{aligned} P(X \geq 18.2) &= 1 - P(X \leq 18.2) = 1 - P\left(\frac{X-10}{4} \leq \frac{18.2-10}{4}\right) \\ &= 1 - P(Z \leq 2.05) \approx 0.0202. \end{aligned}$$

Q54. Roll a fair $4-$sided die twice, and let X equal the larger of the two outcomes if they are different and the common value if they are the same. Find the p.m.f. and the c.d.f. of X.

Solution: The outcome space for this experiment is

$$\mathcal{S} = \{(d_1, d_2) : d_1 = 1, 2, 3, 4; d_2 = 1, 2, 3, 4\}.$$

It is assumed that each of these 16 outcomes has equal probability $1/16$.

By definition, $X(d_1, d_2) = \max\{d_1, d_2\}$ for any $(d_1, d_2) \in \mathcal{S}$. Then

$$P(X = 1) = P[(1, 1)] = \tfrac{1}{16},\ P(X = 2) = P[\{((1, 2), (2, 1), (2, 2)\}] = \tfrac{3}{16},$$
$$P(X = 3) = P[\{(1, 3), (2, 3), (3, 1), (3, 2), (3, 3)\}] = \tfrac{5}{16}$$
$$P(X = 4) = P[\{(1, 4), (2, 4), (3, 4), (4, 1), (4, 2), (4, 3), (4, 4)\}] = \tfrac{7}{16},$$

and the p.d.f. can be simply re-written as

$f(x) = P(X = x) = \frac{1}{16}(2x-1),$ for $x = 1, 2, 3, 4,$ and $f(x) = 0$ otherwise.

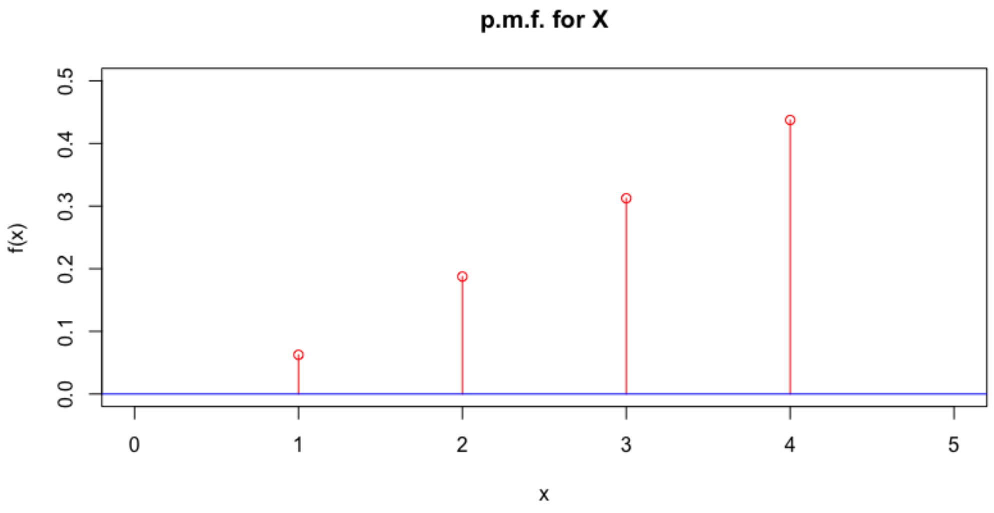
We can verify that

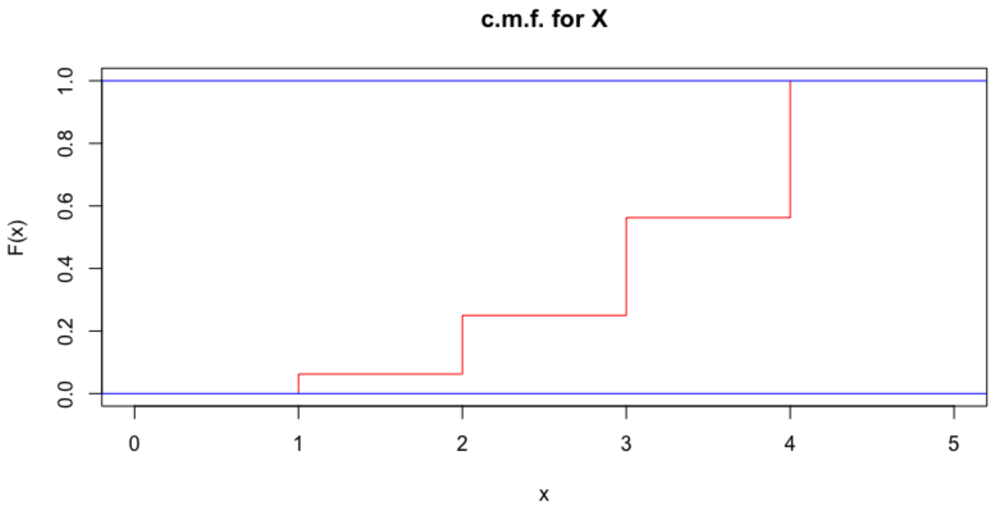
$$\begin{aligned}\sum_{x=1}^{4} f(x) &= \frac{1}{16}\sum_{x=1}^{4}(2x-1) = \frac{2}{16}\sum_{x=1}^{4} x - \frac{1}{16}\sum_{x=1}^{4} 1 \\ &= \frac{2}{16}\cdot\frac{4(5)}{2} - \frac{1}{16}\cdot 4 = \frac{5}{4} - \frac{1}{4} = 1.\end{aligned}$$

Since $2x > 1$ for $x = 1, 2, 3, 4$, $f(x) \geq 0$ for all x and f is indeed a p.m.f.

The c.d.f. is

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{\lfloor x \rfloor}{16} & \text{if } 1 \leq x < 4 \\ 1 & \text{if } x \geq 4 \end{cases}$$


p.m.f. for X
f(x)
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5
x


c.m.f. for X
F(x)
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
x

The graph on the preceding slide was produced with the following (not commented) R code:

```
X <- 1:4
P <- c(1/16,3/16,5/16,7/16)
require(graphics)
par(mfrow=c(2,1))
plot(X,P,type="h",col=2,main="PMF",xlim=c(0,5),ylim=c(0,0.5),xlab="x", ylab="f(x)")
points(X,P,col=2)
abline(h=0,col=4)
F <- cumsum(P)
plot(c(1,X),c(0,F),type="s",main="CMF",xlim=c(0,5),ylim=c(0,1),col=2,xlab="x",ylab="F(x)")
abline(h=0:1,col=4)
```

Q55. Compute the mean and the variance of X as defined in **Q24**, as well as $\mathrm{E}[X(5-X)]$.

Solution: let X be as in question **Q24**. Then

$$\mathrm{E}[X]=\sum_{x=1}^{4}xP(X=x)=1\cdot\frac{1}{16}+2\cdot\frac{3}{16}+3\cdot\frac{5}{16}+4\cdot\frac{7}{16}=\frac{25}{8}=3.125$$

$$\begin{aligned}\mathrm{Var}[X]&=\mathrm{E}[X^2]-\mathrm{E}^2[X]=\sum_{x=1}^{4}x^2P(X=x)-\left(\frac{25}{8}\right)^2\\&=\left(1^2\cdot\tfrac{1}{16}+2^2\cdot\tfrac{3}{16}+3^2\cdot\tfrac{5}{16}+4^2\cdot\tfrac{7}{16}\right)-\tfrac{625}{64}=\tfrac{233}{16}-\tfrac{625}{64}=\tfrac{307}{64}\approx 4.797,\end{aligned}$$

whereas $\mathrm{E}[X(5-X)]=\mathrm{E}[5X-X^2]=5\mathrm{E}[X]-\mathrm{E}[X^2]$. The first and second moments were computed above: $\mathrm{E}[X]=\frac{25}{8}$ and $\mathrm{E}[X^2]=\frac{233}{16}$, so

$$\mathrm{E}[X(5-X)]=5\cdot\frac{25}{8}-\frac{233}{16}=\frac{17}{16}\approx 1.0625.$$

Q56. In 80% of cases when a basketball player attempts a free throw, they are successful. Assume that each of the free throw attempts are independent. Let X be the minimum number of attempts in order to succeed 10 times. Find the p.m.f. of X and the probability that $X = 12$.

Solution: the player requires at least $x = 10$ throws to be successful on 10 separate attempts.

There must be 10 successes in total (with probability 0.8^{10}), and $x - 10$ failures (with probability 0.2^{10-x}).

Furthermore, the player has to be successful on the xth throw (otherwise x would not be the minimum number of attempts in order to succeed 10 times); within the first $x - 1$ throws, exactly 9 must be successful, and there are ${}_{x-1}C_9 = \binom{x-1}{9}$ ways for these to be ordered.

Thus, the p.m.f. of X is

$$f(x) = P(X = x) = \binom{x-1}{9}(0.8)^{10}(0.2)^{x-10}, \quad x = 10, 11, 12, ...$$

and $P(X = 12) = f(12) = \binom{12}{9}(0.80)^{10}(0.20)^2 \approx 0.2362.$

Q57. Let X be the minimum number of independent trials (each with probability of success p) that are needed to observe r successes. The p.m.f. of X is

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-1}, \quad x = r, r+1, \ldots$$

The mean and variance of X are $\mathrm{E}[X] = \frac{r}{p}$ and $\mathrm{Var}[X] = \frac{r(1-p)}{p^2}$. Compute the mean minimum number of independent free throw attempts required to observe 10 successful free throws if the probability of success at the free thrown line is 80%. What about the standard deviation of X?

Solution: let X be as in **Q26**. We have

$$\mathrm{E}[X] = \frac{r}{p} = \frac{10}{0.80} = 12.5 \text{ and } \mathrm{Var}[X] = \frac{r(1-p)}{p^2} = \frac{10(0.20)}{0.80^2} \approx 3.125,$$

from which we conclude that $\mathrm{SD}[X] \approx \sqrt{3.125} \approx 1.768.$

Q58. If $n \geq 20$ and $p \leq 0.05$, it can be shown that the binomial distribution with n trials and an independent probability of success p can be approximated by a Poisson distribution with parameter $\lambda = np$. This is called the **Poisson approximation**:

$$\frac{(np)^x e^{-np}}{x!} \approx \binom{n}{x} p^x (1-p)^{n-x}.$$

A manufacturer of light bulbs knows that 2% of its bulbs are defective. What is the probability that a box of 100 bulbs contains exactly at most 3 defective bulbs? Use the Poisson approximation to estimate the probability.

Solution: let X be the number of defective bulbs in the box of 100. Since $X \sim \mathcal{B}(100, 0.02)$, we have

$$P(X = x) = \binom{100}{x}(0.02)^x(0.98)^{100-x}$$

and

$$P(X \leq 3) = \sum_{x=0}^{3}\binom{100}{x}(0.02)^x(0.98)^{100-x}.$$

But $n = 100 \geq 20$ and $p = 0.02 \leq 0.05$ and the Poisson approximation applies. If $\lambda = np = 2$, then

$$\frac{2^x e^{-2}}{x!} \approx \binom{100}{x}(0.02)^x(0.98)^{100-x},$$

so

$$
\begin{aligned}
P(X \leq 3) &\approx \sum_{x=0}^{3} \frac{2^x e^{-2}}{x!} \\
&= e^{-2}\left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!}\right) \approx 0.857.
\end{aligned}
$$

Q59. Consider a discrete random variable X which has a uniform distribution over the first positive m integers, i.e.

$$f(x) = P(X = x) = \frac{1}{m}, \quad x = 1, \dots, m,$$

and $f(x) = 0$ otherwise. Compute the mean and the variance of X. For what values of m is $\mathrm{E}[X] > \mathrm{Var}[X]$?

Solution: we have

$$\mathrm{E}[X] = \sum_{x=1}^{m} x f(x) = \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m}\sum_{x=1}^{m} x = \frac{1}{m} \cdot \frac{m(m+1)}{2} = \frac{m+1}{2},$$

$$\mathrm{Var}[X] = E[X^2] - E^2[X] = \sum_{x=1}^{m} x^2 f(x) - \frac{(m+1)^2}{4} = \frac{1}{m}\sum_{x=1}^{m} x^2 - \frac{(m+1)^2}{4}$$

$$= \frac{1}{m} \cdot \frac{m(m+1)(2m+1)}{6} - \frac{(m+1)^2}{4} = \frac{m^2-1}{12}.$$

The mean is greater than the variance when

$$\frac{m+1}{2} > \frac{m^2-1}{12} \leftrightarrow 6(m+1) > m^2 - 1 \leftrightarrow m^2 - 6m - 7 < 0 \leftrightarrow 1 \leq m < 7.$$

Q60. Let X be a random variable. What is the value of b (where b is not a function of X) which minimizes $\mathrm{E}[(X-b)^2]$?

Solution: write

$$\begin{aligned} g(b) &= \mathrm{E}[(X-b)^2] = \mathrm{E}[X^2 - 2bX + b^2] = \mathrm{E}[X^2] - 2b\mathrm{E}[X] + \mathrm{E}[b^2] \\ &= \mathrm{E}[X^2] - 2b\mathrm{E}[X] + b^2. \end{aligned}$$

To find the minimum of $g(b)$ with respect to b, set $g'(b) = 0$ and solve for b:

$$\begin{aligned} g'(b) &= -2\mathrm{E}[X] = 2b = 0 \\ b &= \mathrm{E}[X]. \end{aligned}$$

Since $g''(b) = 2 > 0$, $\mathrm{E}[X]$ is the value of b that minimizes $\mathrm{E}[(X-b)^2]$.

Q61. An experiment consists in selecting a bowl, and then drawing a ball from that bowl. Bowl B_1 contains two red balls and four white balls; bowl B_2 contains one red ball and two white balls; and bowl B_3 contains five red balls and four white balls. The probabilities for selecting the bowls are not uniform: $P(B_1) = 1/3$, $P(B_2) = 1/6$, and $P(B_3) = 1/2$, respectively.

a) What is the probability of drawing a red ball $P(R)$?

b) If the experiment is conducted and a red ball is drawn, what is the probability that the ball was drawn from bowl B_1? B_2? B_3?

Solution:

a) Since B_1, B_2 and B_3 are mutually exclusive, we can apply the Law of Total Probability to obtain

$$\begin{aligned} P(R) &= P(R|B_1)P(B_1) + P(R|B_2)P(B_2) + P(R|B_3)P(B_3) \\ &= P(R|B_1)(1/3) + P(R|B_2)(1/6) + P(R|B_3)(1/2). \end{aligned}$$

But from the problem statement, we have $P(R|B_1) = \frac{2}{6}$, $P(R|B_2) = \frac{1}{3}$, $P(R|B_3) = \frac{5}{9}$, so

$$P(R) = (2/6)(1/3) + (1/3)(1/6) + (5/9)(1/2) = 4/9 \approx 0.44.$$

b) We are looking for the conditional probabilities $P(R|B_1)$, $P(R|B_2)$, and $P(R|B_3)$. According to Bayes' Theorem,

$$P(R|B_1) = \frac{P(R|B_1)P(B_1)}{P(R)} = \frac{(2/6)(1/3)}{4/9} = \frac{2}{8} = \frac{1}{4}$$

$$P(B_2|R) = \frac{(1/3)(1/6)}{4/9} = \frac{1}{8}$$

$$P(B_3|R) = \frac{(5/9)(1/2)}{4/9} = \frac{5}{8}.$$

Once the red ball has been picked, the probability concerning B_3 seems more favourable because B_3 has a larger percentage of red balls than do B_1 and B_2.

Q62. The time to reaction to a visual signal follows a normal distribution with mean 0.5 seconds and standard deviation 0.035 seconds.

a) What is the probability that time to react exceeds 1 second?

b) What is the probability that time to react is between 0.4 and 0.5 seconds?

c) What is the time to reaction that is exceeded with probability of 0.9?

Solution: let X be the time to reaction. Then $X \sim \mathcal{N}(0.5, (0.035)^2)$.

a) We need to evaluate

$$P(X > 1) = 1 - \Phi\left(\tfrac{1-0.5}{0.035}\right) = 1 - \Phi(14.29) \approx 0.$$

b) We need to evaluate

$$P(0.4 < X < 0.5) = \Phi\left(\tfrac{0.5-0.5}{0.035}\right) - \Phi\left(\tfrac{0.4-0.5}{0.035}\right) = \Phi(0) - \Phi(-2.86) \approx 0.4979.$$

c) We want to find x such that $0.90 = P(X > x) = 1 - \Phi\left(\tfrac{x-0.5}{0.035}\right)$. Thus, $\Phi\left(\tfrac{x-0.5}{0.035}\right) = 0.10 \;\Rightarrow\; \tfrac{x-0.5}{0.035} = -1.28 \;\Rightarrow\; x = 0.4552.$

Q63. Suppose that the random variable X has the following cumulative distribution function:

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ x^3, & 0 \leq x \leq 1 \\ 1, & x \geq 1. \end{cases}$$

a) Compute $P(X > 0.5)$.

b) Compute $P(0.2 < X < 0.8)$.

c) Find the probability density function of X.

d) Find $\mathrm{E}[X]$ and $\mathrm{Var}[X]$.

Solution:

a) $P(X > 0.5) = 1 - F(0.5) = 1 - 0.5^3 = 0.875.$

b) $P(0.2 < X < 0.8) = F(0.8) - F(0.2) = (0.8)^3 - (0.2)^3 = 0.504.$

c) $f_X(x) = F'(x) = 3x^2, \quad 0 < x < 1.$

d)

$$\mu_X = \mathrm{E}[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 3\, x^3\, dx = \frac{3}{4}x^4 \Big|_0^1 = \frac{3}{4}$$

$$\mathrm{E}[X^2] = \int_{-\infty}^{\infty} x^2\, f(x)\, dx = \int_0^1 3\, x^4\, dx = \frac{3}{5}x^5 \Big|_0^1 = \frac{3}{5}.$$

Thus, $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mu_X^2 = (3/5) - (3/4)^2 = 3/80 = 0.0375.$

Q64. Assume that arrivals of small aircrafts at an airport can be modeled by a Poisson random variable with an average of 1 aircraft per hour.

a) What is the probability that more than 3 aircrafts arrive within an hour?

b) Consider 15 consecutive and disjoint $1-$hour intervals. What is the probability that in none of these intervals we have more than 3 aircraft arrivals?

c) What is the probability that exactly 3 aircrafts arrive within 2 hours?

Solution:

a) Let X be the number of aircrafts that arrive at the airport within one hour. Thus, $X \sim \mathcal{P}(\lambda)$, with $\lambda = 1$, and

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) \\ &= 1 - \left[e^{-1}\frac{1^0}{0!} + e^{-1}\frac{1^1}{1!} + e^{-1}\frac{1^2}{2!} + e^{-1}\frac{1^3}{3!} \right] \approx 0.01899. \end{aligned}$$

We can also compute this probability with $\texttt{R}$:

$$\texttt{1} - \texttt{ppois}(\texttt{3}, \texttt{1}) \approx 0.01899.$$

b) Let Y be the number of $1-$hour intervals with more than 3 arrivals. If the arrivals are independent, we can view Y as a binomial experiment with $p = 0.01899$. Thus, $Y \sim B(15, 0.01899)$ and

$$\begin{aligned} P(Y=0) &= \binom{15}{0}(0.01899)^0(1-0.01899)^{15} = \texttt{dbinom}(0,15,0.01899) \\ &\approx 0.7501. \end{aligned}$$

c) Let W be the number of arrivals within 2 hours. Thus, $W \sim \mathcal{P}(\lambda^*)$ with $\lambda^* = 2$ and

$$P(W=3) = e^{-2} \cdot \frac{2^3}{3!} = \texttt{dpois}(3,2) \approx 0.1804.$$

Q65. Refer to the situation described in **Q3**.

a) What is the length of the interval such that the probability of having no arrival within this interval is 0.1?

b) What is the probability that one has to wait at least 3 hours for the arrival of 3 aircrafts?

c) What is the mean and variance of the waiting time for 3 aircrafts?

Solution:

a) Let T be the time between two consecutive arrivals (in hours). Thus, $T \sim \text{Exp}(\lambda)$, for $\lambda = 1$. We want to find t such that

$$0.1 = P(\text{no arrivals in } [0, t]) = P(T > t) = e^{-\lambda\, t}.$$

Thus, $t \approx -\ln(0.1)/\lambda = 2.3026$ hours.

b) Let S be the number of arrivals within 3 hours. Thus, $S \sim \mathcal{P}(\lambda^{**})$, with $\lambda^{**} = 3$.

The probability that one has to wait at least 3 hours for the arrival of 3 aircrafts is the probability that at most 2 aircrafts arrive within 3 hours.

Thus, we compute

$$P(S \leq 2) = P(S=0) + P(S=1) + P(S=2) = e^{-3}\left[\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!}\right]$$
$$= \texttt{ppois}(2,3) \approx 0.4232.$$

c) Let T be the waiting time (in hours) for 3 arrivals. Thus, $T \sim \Gamma(\lambda, r)$ with $\lambda = 1$ and $r = 3$. We compute

$$\mathrm{E}[T] = \frac{r}{\lambda} = 3 \quad \text{and} \quad \mathrm{Var}[T] = \frac{r}{\lambda^2} = 3.$$

Q66. Assume that X is normally distributed with mean 10 and standard deviation 3. In each case, find the value x such that:

a) $P(X > x) = 0.5$

b) $P(X > x) = 0.95$

c) $P(x < X < 10) = 0.2$

d) $P(-x < X - 10 < x) = 0.95$

e) $P(-x < X - 10 < x) = 0.99$

Solution: if $X \sim \mathcal{N}(10, 3^2)$, then $Z = \frac{X-10}{3} \sim \mathcal{N}(0,1)$. Using the table, we find:

a) $0.5 = P(X > x) = 1 - \Phi(\frac{x-10}{3}) \Rightarrow \Phi(\frac{x-10}{3}) = 0.5 \Rightarrow \frac{x-10}{3} = 0 \Rightarrow x = 10.$

b) $0.95 = P(X > x) = 1 - \Phi(\frac{x-10}{3}) \Rightarrow \Phi(\frac{x-10}{3}) = 0.05 \Rightarrow \frac{x-10}{3} \approx -1.64 \Rightarrow x \approx 5.08.$

c) $0.2 = P(x < X < 10) = \Phi(\frac{10-10}{3}) - \Phi(\frac{x-10}{3}) \Rightarrow \Phi(\frac{x-10}{3}) = \Phi(0) - 0.2 = 0.3 \Rightarrow \frac{x-10}{3} \approx -0.52 \Rightarrow x = 8.44.$

d) $0.95 = P(-x < X - 10 < x) = \Phi(x/3) - \Phi(-x/3) = \Phi(x/3) - [1 - \Phi(x/3)] \Rightarrow \Phi(x/3) = (0.95+1)/2 = 0.975 \Rightarrow x/3 \approx 1.96 \Rightarrow x \approx 5.88.$

e) $0.99 = P(-x < X - 10 < x) = \Phi(x/3) - \Phi(-x/3) = \Phi(x/3) - [1 - \Phi(x/3)] \Rightarrow \Phi(x/3) = (0.99+1)/2 = 0.995 \Rightarrow x/3 \approx 2.58 \Rightarrow x \approx 7.74.$

Q67. Let $X \sim$ Exp(λ) with mean 10. What is $P(X > 30|X > 10)$ equal to?

a)$1 - \exp(-2)$ b)$\exp(-2)$ c)$\exp(-3)$

d)$1/10$ e)$\exp(-200)$ f) none of the preceding

Solution: we have $10 = \mathrm{E}[X] = \frac{1}{\lambda}$, so $\lambda = \frac{1}{10}$. From the memory-less property of the exponential distribution,

$$P(X > 30 | X > 10) = P(X > 20) = \exp\left(-\lambda \cdot 20\right) = \exp(-2) \approx 0.1353.$$

Q68. Let X denote a number of failures of a particular machine within a month. Its probability mass function is given by

x	0	1	2	3	4	5
$P(X = x)$	0.17	0.23	0.19	0.13	0.08	0.2

The probability that there are fewer than 3 failures within a month, and the expected number of failures within a month are, respectively,

a)$0.28; 2.50$ b)$0.72; 2.32$ c)$0.59; 2.32$

d)$0.80; 2.50$ e)none of the preceding

Solution:

Q69. A company's warranty document states that the probability that a new swimming pool requires some repairs within the first year is 20%. What is the probability, that the sixth sold pool is the first one which requires some repairs within the first year?

a)0.6068 b)0.3932 c)0.9345 d)0.0655 e)none of the preceding

Solution:

Q70. In a group of ten students, each student has a probability of 0.7 of passing the exam. What is the probability that exactly 7 of them will pass an exam?

a)0.9829 b)0.2668 c)0.0480 d)0.9520 e)none of the preceding

Solution: Let X be the number of students who pass the exam. We assume $X \sim \mathcal{B}(10, 0.7)$. Then

$$P(X = 7) = \binom{10}{7}(0.7)^7(0.3)^3 = \frac{10!}{7!3!}(0.7)^7(0.3)^3 \approx 0.2668.$$

Q71. Two companies A and B consider making an offer for road construction. The company A makes the submission. The probability that B submits the proposal is $1/3$. If B does not submit the proposal, the probability that A gets the job is $3/5$. If B submits the proposal, the probability that A gets the job is $1/3$. What is the probability that A will get the job?

a)0.6667 b)0.5111 c)0.7500 d)0.3333 e)none of the preceding

Solution:

Q72. In a box of 50 fuses there are 8 defective ones. We choose 5 fuses randomly (without replacement). What is the probability that all 5 fuses are not defective?

a)0.4015 b)0.84 c)0.3725 d)0.4275 e)none of the preceding

Solution:

Q73. Consider a random variable X with the following probability density function:

$$f(x) = \begin{cases} 0 & \text{if } x \leq -1 \\ \frac{3}{4}(1-x^2) & \text{if } -1 < x < 1 \\ 0 & \text{if } x \geq 1 \end{cases}$$

The value of $P(X \leq 0.5)$ is

a)$11/32$ b)$27/32$ c)$16/32$ d)1 e)none of the preceding

Solution: We need to compute:

$$P(X \leq 0.5) = \int_{-1}^{0.5} \frac{3}{4}(1 - x^2)\,dx = \left[\frac{3}{4}x\right]_{-1}^{0.5} - \left[\frac{1}{4}x^3\right]_{-1}^{0.5} = 27/32.$$

Q74. A receptionist receives on average 2 phone calls per minute. If the number of calls follows a Poisson process, what is the probability that the waiting time for call will be greater than 1 minute?

a)$e^{-1/15}$ b)$e^{-1/30}$ c)e^{-2} d)e^{-1} e)none of the preceding

Solution: We have a Poisson process with $\lambda = 2$. The waiting time in a Poisson process is exponentially distributed. Let W be an exponential random variable with parameter $\lambda = 2$. Then

$$P(W > 1) = 1 - P(W \leq 1) = 1 - (1 - \exp(-2 \cdot 1)) = \exp(-2).$$

Q75. A company manufactures hockey pucks. It is known that their weight is normally distributed with mean 1 and standard deviation 0.05. The pucks used by the NHL must weigh between 0.9 and 1.1. What is the probability that a randomly chosen puck can be used by NHL?

a) 1 b) 0.9545 c) 0.4560 d) 0.9772 e) none of the preceding

Solution: We want to evaluate

$$
\begin{aligned}
P(0.9 < X < 1.1) &= P\left(\frac{0.9 - 1.0}{0.05} < Z < \frac{1.1 - 1.0}{0.05}\right) \\
&= \Phi(2) - \Phi(-2) = 0.977250 - 0.022750 = 0.9545.
\end{aligned}
$$

Q76. Consider the following dataset:

12 14 6 10 1 20 4 8

The median and the first quartile of the dataset are, respectively:

a)$9,\ 5$ b)$5.5,\ 6$ c)$10,\ 5$ d)$5,\ 10$ e)none of the preceding

Solution:

Q77. Let X denote a number of failures of a particular machine within a month. Its probability mass function is given by

x	0	1	2	3	4	5
$P(X=x)$	0.17	0.23	0.19	0.13	0.08	0.2

- The probability that there are less than 3 failures within a month, and
- the expected number of failures within a month

are, respectively:

a)0.28; 2.50 b)0.72; 2.32 c)0.59; 2.32 d)0.80; 2.50 e)none of the preceding

Solution:

Q78. Consider the following R output:

```
> pbinom(15,100,0.25)                > pbinom(16,100,0.25)
[1] 0.01108327                       [1] 0.02111062
> pbinom(17,100,0.25)                > pbinom(30,100,0.25)
[1] 0.03762626                       [1] 0.8962128
> pbinom(31,100,0.25)                > pbinom(32,100,0.25)
[1] 0.9306511                        [1] 0.9554037
```

Let X be a binomial random variable with $n = 100$ and $p = 0.25$. Using the R output above, calculate $P(16 \leq X \leq 31)$.

a)0.9196 b)0.9095 c)0.9348 d)0.9443 e)none of the preceding

Solution:

Q79. Suppose that samples of size $n = 25$ are selected at random from a normal population with mean 100 and standard deviation 10. What is the probability that sample mean falls in the interval

$$(\mu_{\overline{X}} - 1.8\sigma_{\overline{X}}, \mu_{\overline{X}} + 1.0\sigma_{\overline{X}})?$$

Solution: Recall that $\overline{X} = \mu_X = 25$, $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{5} = 2$. (Note, however, that this information is completely irrelevant).

Instead, we use the fact that $(\overline{X} - \mu_{\overline{X}})/\sigma_{\overline{X}} \sim \mathcal{N}(0,1)$, so that

$$\begin{aligned} P\left(\mu_{\overline{X}} - 1.8\sigma_{\overline{X}} < \overline{X} < \mu_{\overline{X}} + 1.0\sigma_{\overline{X}}\right) &= P\left(-1.8 < \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} < 1.0\right) \\ &= \Phi(1.0) - \Phi(-1.8) \\ &= 0.8413 - 0.0359 = 0.8054. \end{aligned}$$

Q80. The compressive strength of concrete is normally distributed with mean $\mu = 2500$ and standard deviation $\sigma = 50$. A random sample of size 5 is taken. What is the standard error of the sample mean?

Solution: by definition,

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{5}} = 22.3607.$$

Q81. Suppose that $X_1 \sim \mathcal{N}(3,4)$ and $X_2 \sim \mathcal{N}N(3,45)$. Given that X_1 and X_2 are independent random variables, what is a good approximation to $P(X_1 + X_2 > 9.5)$?

a)0.3085 b)0.6915 c)0.5279 d)0.4271 e)none of the preceding

Solution: since X_1 and X_2 are independent,

$$X_1 + X_2 \sim \mathcal{N}(3+3, 4+45) = \mathcal{N}(6, 49).$$

If $Y = X_1 + X_2$, then

$$\begin{aligned} P(X_1 + X_2 > 9.5) = P(Y > 9.5) &= P\left(\frac{Y-6}{7} > \frac{9.5-6}{7}\right) \\ &= P(Z > 0.5) \approx 1 - 0.6915 = 0.3085 \end{aligned}$$

Q82. The amount of time that a customer spends waiting at an airport check-in counter is a random variable with mean $\mu = 8.2$ minutes and standard deviation $\sigma = 1.5$ minutes. Suppose that a random sample of $n = 49$ customers is taken. Compute the approximate probability that the average waiting time for these customers is:

a)Less than 10 min. b)Between 5 and 10 min. c)Less than 6 min.

Solution: let $\overline{X}$ be the mean waiting time for 49 clients. Then according to the Central Limit Theorem, $\overline{X} \sim \mathcal{N}\left(8.2, \frac{1.5^2}{49}\right)$.

a) $P(\overline{X} < 10) = \Phi\left(\frac{10-8.2}{1.5/\sqrt{49}}\right) = \Phi(8.4) \approx 1.$

b) $P(5 < \overline{X} < 10) = \Phi\left(\frac{10-8.2}{1.5/\sqrt{49}}\right) - \Phi\left(\frac{5-8.2}{1.5/\sqrt{49}}\right) = \Phi(8.4) - \Phi(-14.93) \approx 1 - 0 = 1.$

c) $P(\overline{X} < 6) = \Phi\left(\frac{6-8.2}{1.5/\sqrt{49}}\right) = \Phi(-10.26) \approx 0.$

Q83. A random sample of size $n_1 = 16$ is selected from a normal population with a mean of 75 and standard deviation of 8. A second random sample of size $n_2 = 9$ is taken independently from another normal population with mean 70 and standard deviation of 12. Let $\overline{X}_1$ and $\overline{X}_2$ be the two sample means. Find

a) The probability that $\overline{X}_1 - \overline{X}_2$ exceeds 4.

b) The probability that $3.5 < \overline{X}_1 - \overline{X}_2 < 5.5$.

Solution: we have $\overline{X}_1 - \overline{X}_2 \sim \mathcal{N}(75-70, 8^2/16 + 12^2/9) = \mathcal{N}(5, 20)$.

a) Thus,

$$P(\overline{X}_1 - \overline{X}_2 > 4) = P\left(Z > \frac{4-5}{\sqrt{20}}\right) = 1 - \Phi\left(\frac{4-5}{\sqrt{20}}\right)$$
$$= 1 - \Phi(-0.22) = 0.5871.$$

b) Furthermore,

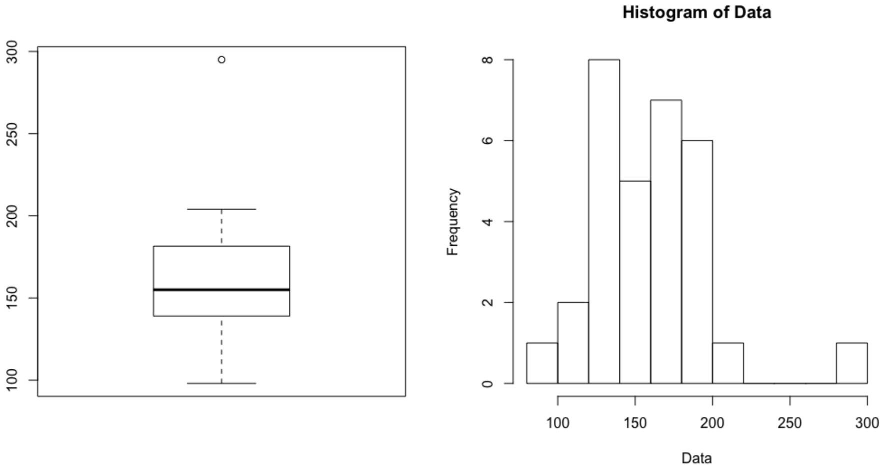
$$P(3.5 \leq \overline{X}_1 - \overline{X}_2 \leq 5.5) = \Phi\left(\frac{5.5-5}{\sqrt{20}}\right) - \Phi\left(\frac{3.5-5}{\sqrt{20}}\right)$$
$$= \Phi(0.11) - \Phi(-0.34) = 0.5438 - 0.3669 = 0.1769$$

Q84. Discuss the normality of the following dataset:

```
170,295,200,165,140,190,195,142,138,148,110,140,103,176,125,
126,204,196,98,123,124,152,177,168,175,186,140,147,174,155,195
```

Solution: the following piece of R code will produce a boxplot and a histogram for the data:

```
Data=c(170,295,200,165,140,190,195,142,138,148,110,140,103,
176,125,126,204,196,98,123,124,152,177,168,175,186,140,147,174,155,195);
par(mfrow=c(1,2)); boxplot(Data);hist(Data,breaks=10)
```



The observations do not seem to be symmetric; there is an outlier; the data does not seem to be normal.

Q85. Using R, illustrate the central limit theorem by generating $M = 300$ samples of size $n = 30$ from:

- a normal random variable with mean 10 and variance 0.75;
- a binomial random variable with 3 trials and probability of success 0.3.

Repeat the same procedure for samples of size $n = 200$. What do you observe?

Hint: In each case, assess the normality using a histogram and a QQ plot.

Solution:

Q86. Suppose that the weight in pounds of a North American adult can be represented by a normal random variable with mean 150 lbs and variance 900 lbs^2. An elevator containing a sign "Maximum 12 people" can safely carry 2000 lbs. The probability that 12 North American adults will not overload the elevator is closest to

a)0.9729 b)0.4501 c)0.0271 d)0.0001 e)1.3 f) none of the preceding

Solution: let S be the total weight of 12 adults. Then

$$S \sim \mathcal{N}(12 \times 150, 12 \times 900) = \mathcal{N}(1800, 10800),$$

and the probability that they will not overload the elevator is

$$\begin{aligned} P(S < 2000) &= P\left(\frac{S - 1800}{\sqrt{10800}} < \frac{2000 - 1800}{\sqrt{10800}}\right) \\ &= P(Z < 1.924) = \texttt{pnorm(1.924501, 0, 1)} \approx 0.9729. \end{aligned}$$

Q87. Let $X_1, \cdots, X_{50}$ be an independent random sample from a Poisson distribution with mean 1. Set $Y = X_1 + \cdots + X_{50}$. The approximate probability $P(48 \leq Y \leq 52)$ is closest to:

a)0.6368 b)0.4534 c)0.2227 d)0.9988 e)0.5000 f) none of the preceding

Solution: for each Poisson variable X_i, we have $\mathrm{E}[X_i] = \mathrm{Var}[X_i] = 1$. Thus $\mathrm{E}[Y] = \mathrm{Var}[Y] = 50$, and, according to the Central Limit Theorem, we have

$$\begin{aligned} P(48 \leq Y \leq 52) &= P\left(\frac{48-50}{\sqrt{50}} \leq \frac{Y-50}{\sqrt{50}} \leq \frac{52-50}{\sqrt{50}}\right) \\ &\approx P(-0.2828 < Z < 0.2828) \\ &= \Phi(0.2828) - \Phi(-0.2828) \approx 0.2227. \end{aligned}$$

Q88. A new type of electronic flash for cameras will last an average of 5000 hours with a standard deviation of 500 hours. A quality control engineer intends to select a random sample of 100 of these flashes and use them until they fail. What is the probability that the mean life time of the sample of 100 flashes will be less than 4928 hours?

a)0.0749 b)0.9251 c)0.0002 d)0.4532 e)none of the preceding

Solution: we have $\mu = 5000$ and $\sigma = 500$. Let $\overline{X}$ be the mean life time of $n = 100$ flashes. Then $\mu_{\overline{X}} = 5000$ and $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{10}$. We standardize to obtain

$$\begin{aligned} P(\overline{X} < 4298) &= P\left(\frac{\overline{X} - 5000}{500/10} < \frac{4928 - 5000}{500/10}\right) \\ &\approx P(Z < -1.44) = \Phi(-1.44) \approx 0.0749. \end{aligned}$$

Q89. A manufacturer of fluoride toothpaste regularly measures the concentration of of fluoride in the toothpaste to make sure that it is within the specifications of $0.85 - 1.10$ mg/g. The table on the next page lists 100 such measurements. Build a relative frequency histogram of the data (a histogram with area $= 1$).

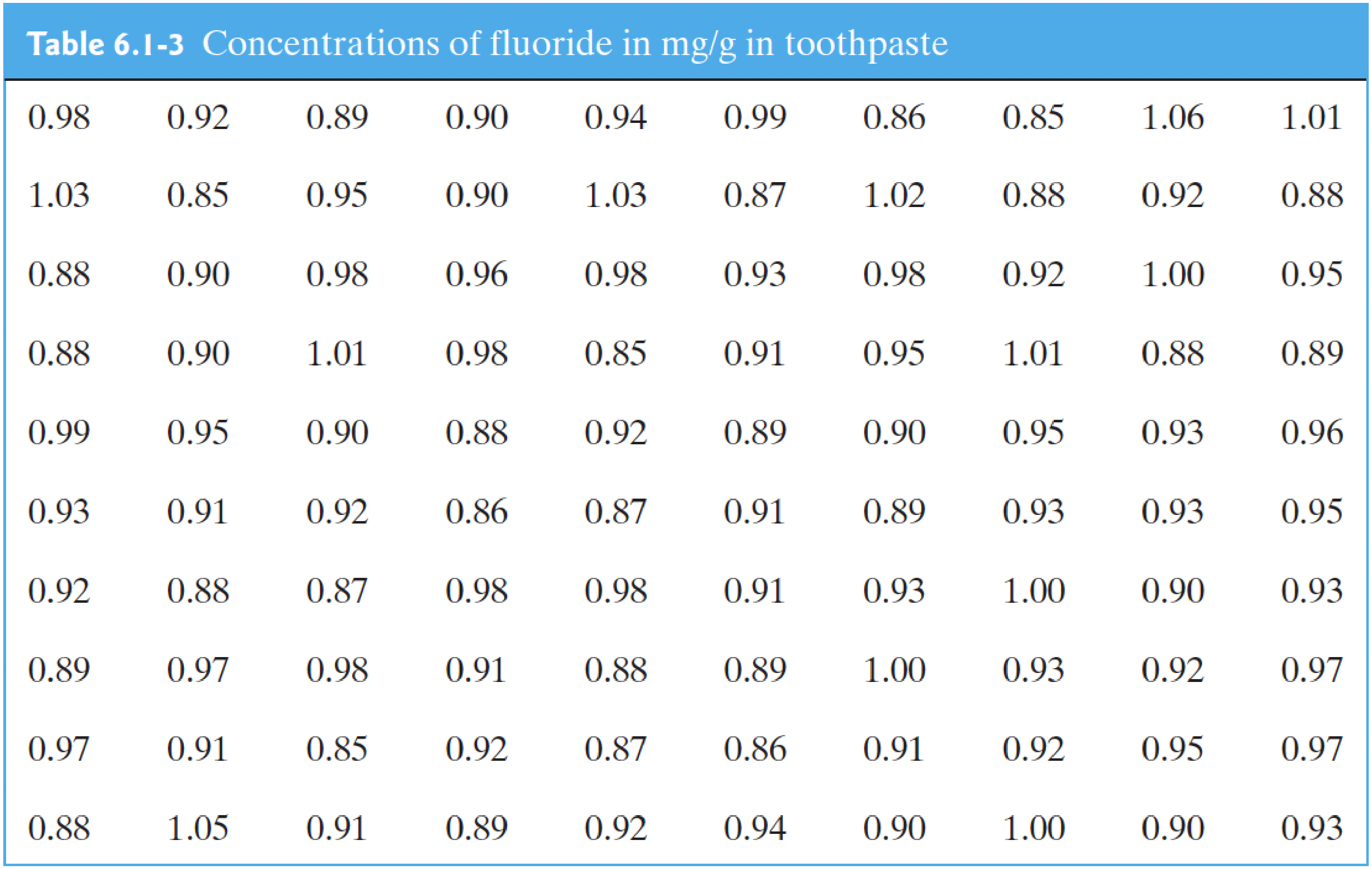


Table 6.1-3 Concentrations of fluoride in mg/g in toothpaste

0.98	0.92	0.89	0.90	0.94	0.99	0.86	0.85	1.06	1.01
1.03	0.85	0.95	0.90	1.03	0.87	1.02	0.88	0.92	0.88
0.88	0.90	0.98	0.96	0.98	0.93	0.98	0.92	1.00	0.95
0.88	0.90	1.01	0.98	0.85	0.91	0.95	1.01	0.88	0.89
0.99	0.95	0.90	0.88	0.92	0.89	0.90	0.95	0.93	0.96
0.93	0.91	0.92	0.86	0.87	0.91	0.89	0.93	0.93	0.95
0.92	0.88	0.87	0.98	0.98	0.91	0.93	1.00	0.90	0.93
0.89	0.97	0.98	0.91	0.88	0.89	1.00	0.93	0.92	0.97
0.97	0.91	0.85	0.92	0.87	0.86	0.91	0.92	0.95	0.97
0.88	1.05	0.91	0.89	0.92	0.94	0.90	1.00	0.90	0.93

Solution: the minimum of these measurements is 0.85 and the maximum is 1.06. The range is $1.06 - 0.85 = 0.21$. We use $k = 8$ classes of length 0.03. Note that $8(0.03) = 0.24 > 0.21$. We start at 0.835 and end at 1.075. These boundaries are the same distance below the minimum and above the maximum. In the table on the next slide, we give the values of the height of each of the histogram's rectangle (the total area has to add up to 1.) The heights are given by

$$h(x) = \frac{f_i}{(0.03)(100)} = \frac{f_i}{3},$$

where f_i is the frequency of concentrations in class i. The plot of the histogram is shown on the subsequent slide.

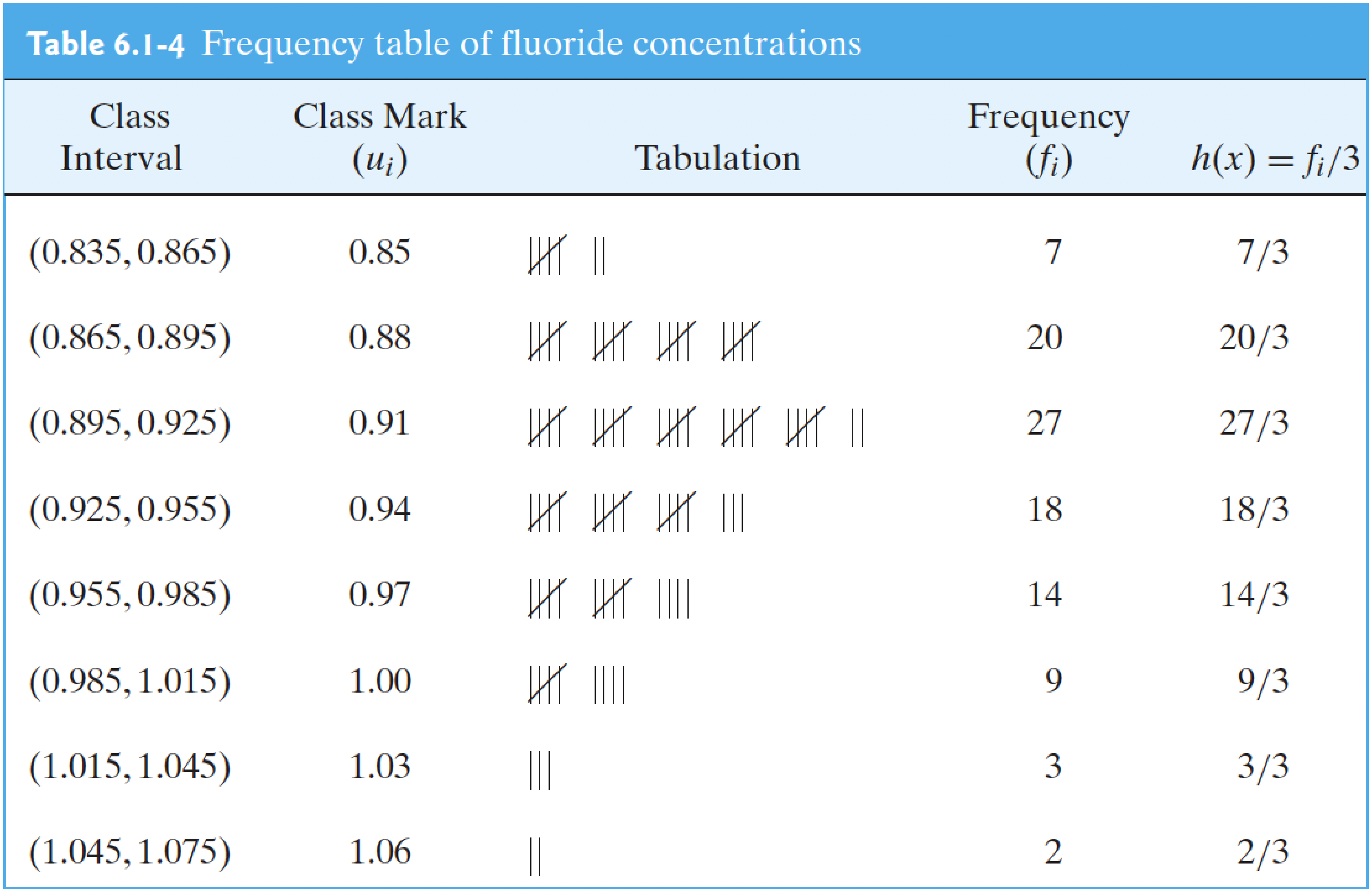
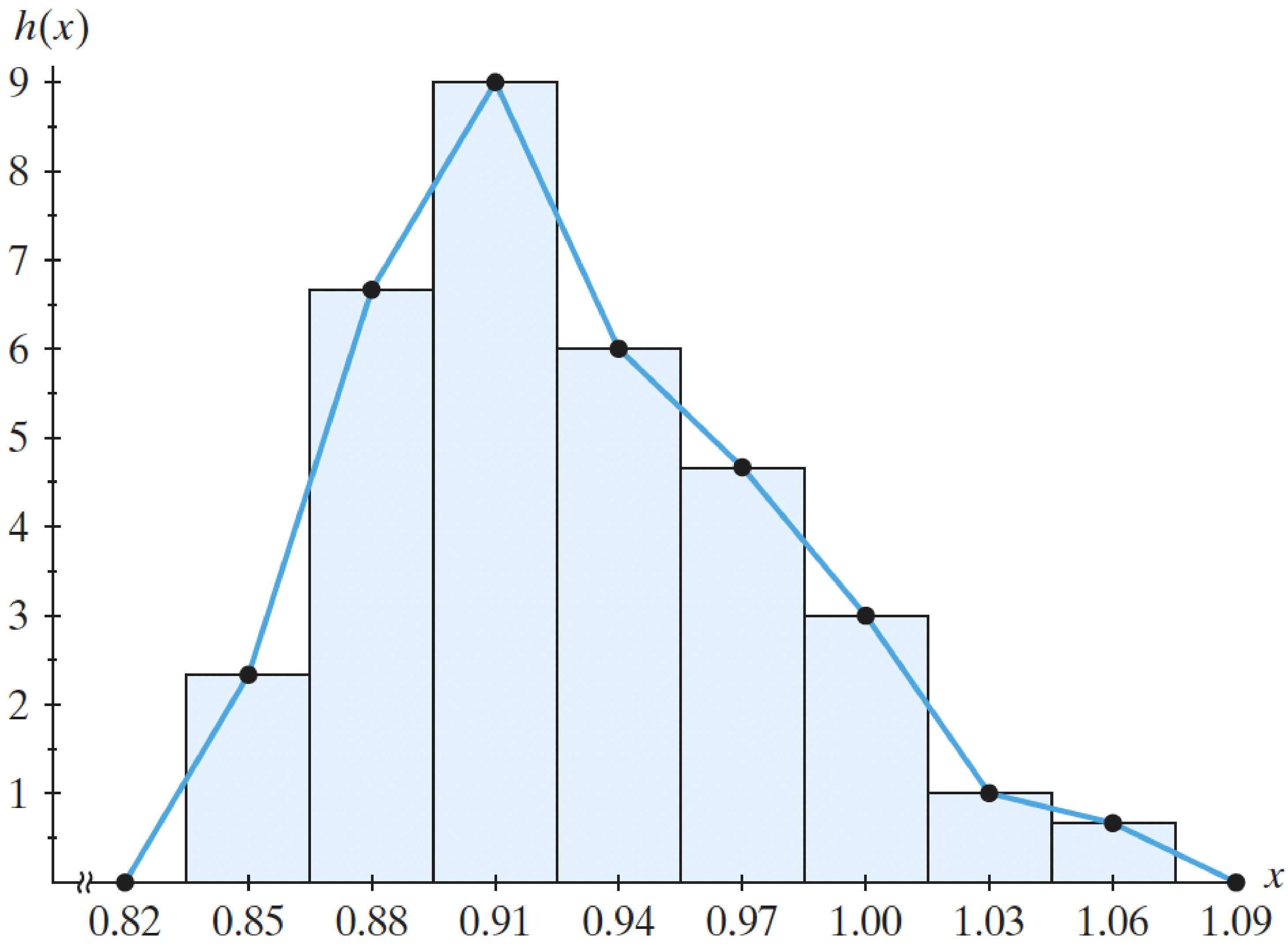


Table 6.1-4 Frequency table of fluoride concentrations

Class Interval	Class Mark (u_i)	Tabulation	Frequency (f_i)	$h(x) = f_i/3$
$(0.835, 0.865)$	0.85	𝍸 𝍷𝍷	7	$7/3$
$(0.865, 0.895)$	0.88	𝍸 𝍸 𝍸 𝍸	20	$20/3$
$(0.895, 0.925)$	0.91	𝍸 𝍸 𝍸 𝍸 𝍸 𝍷𝍷	27	$27/3$
$(0.925, 0.955)$	0.94	𝍸 𝍸 𝍸 𝍷𝍷𝍷	18	$18/3$
$(0.955, 0.985)$	0.97	𝍸 𝍸 𝍷𝍷𝍷𝍷	14	$14/3$
$(0.985, 1.015)$	1.00	𝍸 𝍷𝍷𝍷𝍷	9	$9/3$
$(1.015, 1.045)$	1.03	𝍷𝍷𝍷	3	$3/3$
$(1.045, 1.075)$	1.06	𝍷𝍷	2	$2/3$


$h(x)$
9
8
7
6
5
4
3
2
1
0.82
0.85
0.88
0.91
0.94
0.97
1.00
1.03
1.06
1.09
x

Q90. Use the data from **Q89**.

a) Compute the data's mean $\overline{x}$ and it's standard deviation s_x (use a computer program, for goodness' sake!)

b) Using the frequency table of fluoride concentrations (Table 6.1-4), you can also approximate the mean and variance. Let u_i be the **class mark** for each of the histogram's 8 classes (the midpoint along the rectangles' widths), n be the total number of observations, and k be the number of classes. Then

$$\overline{u} = \frac{1}{n}\sum_{i=1}^{k} f_i u_i \quad \text{and} \quad s_u^2 = \frac{1}{n-1} f_i (u_i - \overline{u})^2.$$

Compute $\overline{u}$ and s_u. How do they compare with $\overline{x}$ and s_x?

Solution:

a) The following `R` code will do the trick:

```
> CCs_Fl<-c(0.98,0.92,0.89,0.90,0.94,0.99,0.86,0.85,1.06,1.01,
      1.03,0.85,0.95,0.90,1.03,0.87,1.02,0.88,0.92,0.88,
      0.88,0.90,0.98,0.96,0.98,0.93,0.98,0.92,1.00,0.95,
      0.88,0.90,1.01,0.98,0.85,0.91,0.95,1.01,0.88,0.89,
      0.99,0.95,0.90,0.88,0.92,0.89,0.90,0.95,0.93,0.96,
      0.93,0.91,0.92,0.86,0.87,0.91,0.89,0.93,0.93,0.95,
      0.92,0.88,0.87,0.98,0.98,0.91,0.93,1.00,0.90,0.93,
      0.89,0.97,0.98,0.91,0.88,0.89,1.00,0.93,0.92,0.97,
      0.97,0.91,0.85,0.92,0.87,0.86,0.91,0.92,0.95,0.97,
      0.88,1.05,0.91,0.89,0.92,0.94,0.90,1.00,0.90,0.93)
> mean(CCs_Fl)
[1] 0.9293
> sd(CCs_Fl)
[1] 0.0489538
```

b) Using the relative frequencies f_i and the class marks u_i from the frequency tables, we get

$$\overline{u} = \frac{1}{100}\sum_{i=1}^{8} f_i u_i = \frac{92.83}{100} = 0.9283,$$

$$s_u^2 = \frac{1}{100-1}\sum_{i=1}^{8} f_i(u_i - \overline{u})^2 = \frac{1}{100-1}\left(\sum_{i=1}^{8} f_i u_i^2 - \frac{1}{100}\left(\sum_{i=1}^{8} f_i u_i\right)^2\right)$$

$$= \frac{0.237411}{99} \approx 0.002398,$$

so $s_u = \sqrt{0.002398} = 0.04897.$

Q91. Use the data from **Q89**.

a) Provide a the $5-$number summary of the data $(q_0, q_1, q_2, q_3, q_4)$, as well as the interquartile range IQR.

b) Display the $5-$number summary as a boxplot chart.

Solution:

a) To make it easy to compute the quartiles, we provide an **ordered stem-and-leaf diagram** of the concentrations on the next slide. There were 100 observations, so

$$Q_0 = 0.85,\ Q_1 = 0.89,\ Q_2 = 0.92,\ Q_3 = 0.97,\ Q_4 = 1.06.$$

The IQR, meanwhile, is $Q_3 - Q_1 = 0.08$.

b) There are no outliers since

$$Q_1 - 1.5 \times \text{IQR} = 0.85 - 1.5(0.08) < Q_0, \text{ and}$$
$$Q_3 + 1.5 \times \text{IQR} = 0.97 + 1.5(0.08) > Q_4.$$

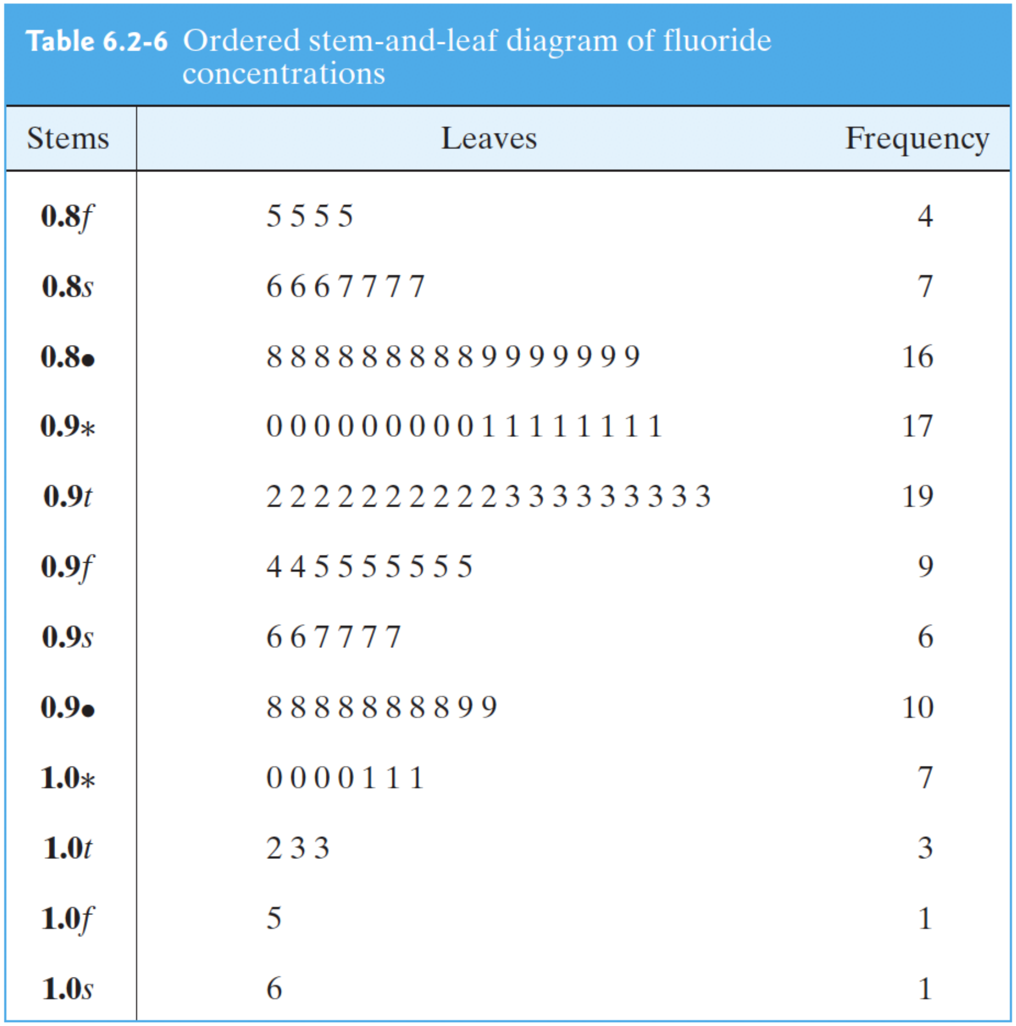
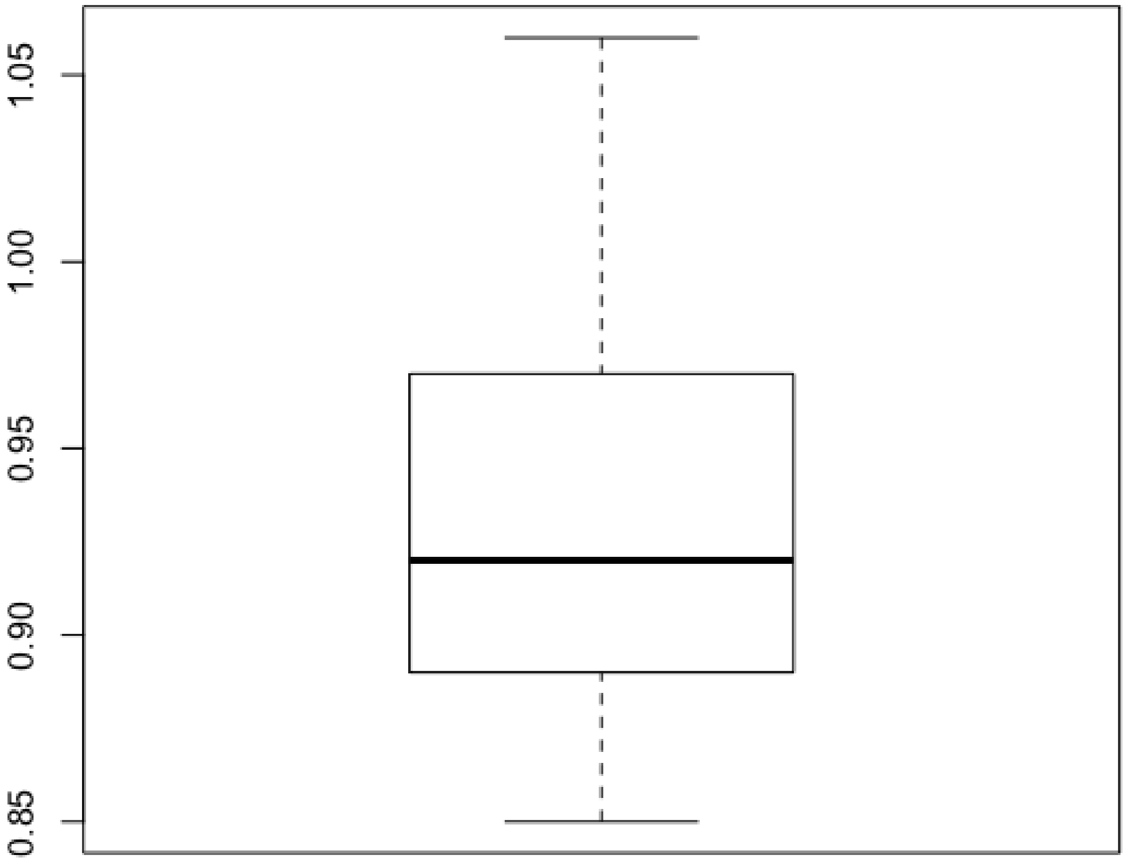


Table 6.2-6 Ordered stem-and-leaf diagram of fluoride concentrations

Stems	Leaves	Frequency
0.8*f*	5 5 5 5	4
0.8*s*	6 6 6 7 7 7 7	7
0.8•	8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9	16
0.9$*$	0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1	17
0.9*t*	2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3	19
0.9*f*	4 4 5 5 5 5 5 5 5	9
0.9*s*	6 6 7 7 7 7	6
0.9•	8 8 8 8 8 8 8 8 9 9	10
1.0$*$	0 0 0 0 1 1 1	7
1.0*t*	2 3 3	3
1.0*f*	5	1
1.0*s*	6	1


0.85
0.90
0.95
1.00
1.05

Q92. Use the data from **Q89**. Compute the **midrange** $\frac{1}{2}(Q_0 + Q_4)$, the **trimean** $\frac{1}{4}(Q_1 + 2Q_2 + Q_3)$, and the **range** $Q_4 - Q_0$ for the fluoride data.

Solution: The midrange is $\frac{1}{2}(Q_0 + Q_4) = \frac{1}{2}(0.85 + 1.06) = 0.955$.

The trimean is $\frac{1}{4}(Q_1 + 2Q_2 + Q_3) = \frac{1}{4}(0.89 + 2(0.92) + 0.97) = 0.925$.

The range is $Q_4 - Q_0 = 1.06 - 0.85 = 0.21$.

Q93. A new cure has been developed for a certain type of cement that should change its mean compressive strength. It is known that the standard deviation of the compressive strength is 130 kg/cm^2 and that we may assume that it follows a normal distribution. 9 chunks of cement have been tested and the observed sample mean is $\overline{X} = 4970$. Find the 95% confidence interval for the mean of the compressive strength.

a)$[4858.37, 5081.63]$ b)$[4885.07, 5054.93]$ c)$[4858.37, 5054.93]$

d)$[4944.52, 4995.48]$ e)none of the preceding

Solution: for a normal population with known standard deviation σ, the 95% C.I. is

$$\overline{X} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 4970 \pm 1.96\left(\frac{130}{\sqrt{9}}\right) \approx [4885.07, 5054.93].$$

Q94. Consider the same set-up as in **Q93**, but now 100 chunks of cement have been tested and the observed sample mean is $\overline{X} = 4970$. Find the 95% confidence interval for the mean of the compressive strength.

a)$[4858.37, 5081.63]$ b)$[4885.07, 5054.93]$ c)$[4858.37, 5054.93]$

d)$[4944.52, 4995.48]$ e)none of the preceding

Solution: for a normal population with known standard deviation σ, the $95\% C.I.$ is

$$\overline{X} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 4970 \pm 1.96\left(\frac{130}{\sqrt{100}}\right) \approx [4944.52, 4995.48].$$

Compare with the the results of the preceding question, and observe the effect of the bigger sample size.

Q95. Consider the same set-up as in **Q93**, but now we do not know the standard deviation of the normal distribution. 9 chunks of cement have been tested, and the measurements are

$$5001, 4945, 5008, 5018, 4991, 4990, 4968, 5020, 5003.$$

Find the 95% confidence interval for the mean of the compressive strength.

a)$[4858.37, 5081.63]$ b)$[4885.07, 5054.93]$ c)$[4858.37, 5054.93]$

d)$[4944.52, 4995.48]$ e)none of the preceding

Solution: we have $\overline{X} \approx 4993.78$ and $s \approx 24.18$. For a normal population with unknown standard deviation σ, the 95% C.I. is

$$\overline{X} \pm t_{0.025,8}\frac{s}{\sqrt{n}} = 4993.78 \pm 2.306\left(\frac{24.18}{\sqrt{9}}\right) \approx [4975.19, 5012.37].$$

Q96. A steel bar is measured with a device which a known precision of $\sigma = 0.5$mm. Suppose we want to estimate the mean measurement with an error of at most 0.2mm at a level of significance $\alpha = 0.05$. What sample size is required? Assume normality?

a)25 b)24 c)6 d)7 e)none of the preceding

Solution: if we assume normality, the error is

$$E = \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \quad \Longrightarrow n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2.$$

Since $E = 0.2$, $\alpha = 0.05$, and $\sigma = 0.5$, then

$$n = \left(\frac{z_{0.025}\sigma}{E}\right)^2 = \left(\frac{1.96 \times 0.5}{0.2}\right)^2 = 24.01,$$

so $n \geq 25$.

Q97. In a random sample of 1000 houses in the city, it is found that 228 are heated by oil. Find a 99% C.I. for the proportion of homes in the city that are heated by oil.

a)$[0.202, 0.254]$ b)$[0.197, 0.259]$ c)$[0.194, 0.262]$

d)$[0.185, 0.247]$ e)none of the preceding

Solution: according to the formula for the confidence interval of a proportion, we have

$$\hat{p} \pm z_{0.005}\sqrt{\frac{\hat{p}(1-\hat{p}}{n}} = \frac{228}{1000} \pm 2.576\sqrt{\frac{228/1000(1-228/1000)}{1000}}$$
$$\approx [0.194, 0.262].$$

Q98. Past experience indicates that the breaking strength of yarn used in manufacturing drapery material is normally distributed and that $\sigma = 2$ psi. A random sample of 15 specimens is tested and the average breaking strength is found to be $\overline{x} = 97.5$ psi.

a) Find a 95% confidence interval on the true mean breaking strength.

b) Find a 99% confidence interval on the true mean breaking strength.

Solution: in both instances we have a normal population with known σ.

a) A 95% confidence interval is provided by

$$\overline{x} \pm z_{.025}\frac{\sigma}{\sqrt{n}} = 97.5 \pm 1.96 \left(\frac{2}{\sqrt{15}}\right) = 97.5 \pm 1.012140 \approx [96.49, 98.51].$$

b) A 99% confidence interval is provided by

$$\overline{x} \pm z_{.005}\frac{\sigma}{\sqrt{n}} = 97.5 \pm 2.576 \left(\frac{2}{\sqrt{15}}\right) = 97.5 \pm 1.330241 \approx [96.17, 98.83].$$

Q99. The diameter holes for a cable harness follow a normal distribution with $\sigma = 0.01$ inch. For a sample of size 10, the average diameter is 1.5045 inches.

a) Find a 99% confidence interval on the mean hole diameter.

b) Repeat this for $n = 100$.

Solution: we have a normal population with known σ.

a) A 99% confidence interval for $n = 10$ is provided by

$$\overline{x} \pm z_{.005}\frac{\sigma}{\sqrt{n}} = 1.5045 \pm 1.96\left(\frac{0.01}{\sqrt{10}}\right) = 1.5045 \pm 0.008146027$$
$$\approx [1.496354, 1.512646].$$

b) A 99% confidence interval for $n = 100$ is provided by

$$\overline{x} \pm z_{.005}\frac{\sigma}{\sqrt{n}} = 1.5045 \pm 1.96\left(\frac{0.01}{\sqrt{100}}\right) = 1.5045 \pm 0.002576$$
$$\approx [1.501924, 1.507076].$$

Q100. A journal article describes the effect of delamination on the natural frequency of beams made from composite laminates. The observations are as follows:

$$230.66, 233.05, 232.58, 229.48, 232.58, 235.22.$$

Assuming that the population is normal, find a 95% confidence interval on the mean natural frequency.

Solution: we have $n = 6$, $\overline{x} = 232.2617$, $s = 1.993935$. Since we do not know the true standard deviation, a 95% confidence interval is provided by

$$\overline{x} \pm t_{0.025,5}\frac{s}{\sqrt{n}} = 232.2617 \pm 2.571\,\left(\frac{1.993935}{\sqrt{6}}\right) = [230.169, 234.355].$$

Q101. A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of $\mu = 12$ kilograms with standard deviation of $\sigma = 0.5$ kilograms.

a) What should be the sample size so that with probability 0.95 we will estimate the mean thread elongation with error at most 0.15 kg?

b) What should be the sample size so that with probability 0.95 we will estimate the mean thread elongation with error at most 0.05 kg?

Solution:

a) We must have

$$n \geq \left(\frac{z_{.025}\,\sigma}{E}\right)^2 = \left(\frac{(1.96)(0.5)}{0.15}\right)^2 = 42.68.$$

b) We must have

$$n \geq \left(\frac{z_{.025}\,\sigma}{E}\right)^2 = \left(\frac{(1.96)(0.5)}{0.05}\right)^2 = 384.16.$$

Q102. The brightness of television picture tube can be evaluated by measuring the amount of current required to achieve a particular brightness level. An engineer thinks that one has to use 300 microamps of current to achieve the required brightness level. A sample of size $n = 20$ has been taken to verify the engineer's hypotheses.

a) Formulate the null and the alternative hypotheses. Use a two-sided test alternative.

b) For the sample of size $n = 20$ we obtain $\overline{x} = 319.2$ and $s = 18.6$. Test the hypotheses from part a) with $\alpha = 5\%$ by computing a critical region. Calculate the p-value.

c) Use the data from part b) to construct a 95% confidence interval for the mean required current.

Solution:

a) We want to verify $\mu = 300$, thus we test $H_0 : \mu = 300$ against $H_1 : \mu \neq 300$.

b) The observed value of the test statistic is

$$t_0 = \frac{\overline{x} - 300}{s/\sqrt{n}} = \frac{319.2 - 300}{18.6/\sqrt{20}} = 4.61.$$

Critical (Rejection) Region: We reject H_0 if $|t_0| > t_{0.025,19} = 2.093$.

Conclusion: Since $|t_0| = 4.61$, we reject H_0. On the level of significance 5%, we conclude that the mean is not equal 300.

p-**value approach:**

$$2\,P(\overline{X} > 319.2) = 2 \times P\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}} > \frac{319.2-300}{18.6/\sqrt{20}}\right) = 2 \times P(t_{19} > 4.61).$$

From the table you can find out that $P(T > 4.61) < 0.0005$ so that p-value$< 2(0.0005) = 0.001$, with the same conclusion.

c) Assuming that population is normal, the 95% C.I. for the mean is

$$\overline{x} \pm t_{0.025,19}\frac{s}{\sqrt{n}} = 319.2 \pm (2.093)\frac{18.6}{\sqrt{20}} = [310.495, 327.905].$$

In particular, since 300 is not in the C.I., we reject H_0.

Q103. We say that a particular production process is **stable** if it produces at most 2% defective items. Let p be the true proportion of defective items.

a) We sample $n = 200$ items at random and consider hypotheses testing about p. Formulate null and alternative hypotheses.

b) What is your conclusion of the above test, if one observes 3 defective items out of 200? Note: you have to choose an appropriate level α.

Solution:

a) $H_0 : p = 0.02$, $H_1 : p \neq 0.02$

b) p-**value approach:** let X be the number of defective items.

$$2 \times P(X \leq 3) = 2 \times P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{3-4}{\sqrt{200(0.02)(0.98)}}\right)$$
$$= 2P(Z < -0.51).$$

Note that we evaluate $P(X \leq 3)$ since the observed percentage (1.5%) is lower than the claimed defective rate. We do not reject H_0 at α since the $p-$value is

$$2P(Z < -0.51) = 0.6101.$$

Q104. Ten engineers' knowledge of basic statistical concepts was measured on a scale of $0 - 100$, before and after a short course in statistical quality control. The result are as follows:

Engineer	1	2	3	4	5	6	7	8	9	10
Before X_{1i}	43	82	77	39	51	66	55	61	79	43
After X_{2i}	51	84	74	48	53	61	59	75	82	53

Let μ_1 and μ_2 be the mean mean score before and after the course. Perform the test $H_0 : \mu_1 = \mu_2$ against $H_A : \mu_1 < \mu_2$. Use $\alpha = 0.05$.

Solution: The differences $D_i = X_{1i} - X_{2i}$ are:

Engineer	1	2	3	4	5	6	7	8	9	10
Before X_{1i}	43	82	77	39	51	66	55	61	79	43
After X_{2i}	51	84	74	48	53	61	59	75	82	53
Difference D_i	-8	-2	3	-9	-2	5	-4	-14	-3	-10

so that $\overline{D} = -4.4$, $S_D = 5.91$. We compute the p-value as

$$P(\overline{D} \le -4.4) = P\left(\frac{\overline{D}}{S_D/\sqrt{n}} \le \frac{-4.4}{\sqrt{31.21/10}}\right) = P(t_9 \le -2.35)$$

$$= P(t_9 > 2.35) \in (0.02, 0.05).$$

Since the p-value is smaller than 0.05, we reject H_0 at $\alpha = 0.05$ in favour of H_1, i.e. the average score improves after the course.

105. A company is currently using titanium alloy rods it purchases from supplier A. A new supplier (supplier B) approaches the company and offers the same quality (at least according to supplier B's claim) rods at a lower price. The company is certainly interested in the offer. At the same time, the company wants to make sure that the safety of their product is not compromised. The company randomly selects ten rods from each of the lots shipped by suppliers A and B and measures the yield strengths of the selected rods. The observed sample mean and sample standard deviation are 651 MPa and 2 MPa for supplier's A rods, respectively, and the same parameters are 657 MPa and 3 MPa for supplier B's rods. Perform the test $H_0 : \mu_A = \mu_B$ against $\mu_A \neq \mu_B$. Use $\alpha = 0.05$. Assume that the variances are equal but unknown.

Solution: This is a two-sample test: $H_0 : \mu_A = \mu_B$, $H_1 : \mu_A \neq \mu_B$. We have $\overline{x}_1 = 651$, $\overline{x}_2 = 657$, $s_1 = 2$, $s_2 = 3$. The observed difference in means is $\overline{x}_1 - \overline{x}_2 = -6$. The test statistic is

$$T_0 = \frac{\overline{x}_1 - \overline{x}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2},$$

where S_p^2 is the **pooled variance** which is computed as follows:

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = 6.5.$$

We compute the p-value as

$$2P(\overline{x}_1 - \overline{x}_2 \leq -6) = 2P\left(\frac{\overline{x}_1 - \overline{x}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \leq -5.26\right) = 2P(t_{18} < -5.26)$$
$$= 2P(t_{18} > 5.26) < 2(0.0005) = 0.001.$$

This is smaller than $\alpha = 0.05$, thus we reject H_0 in favour of H_1 at level $\alpha = 0.05$.

106. The deflection temperature under load for two different types of plastic pipe is being investigated. Two random samples of 15 pipe specimens are tested, and the deflection temperatures observed are as follows:

Type 1: 206, 188, 205, 187, 194, 193, 207, 185, 189, 213, 192, 210, 194, 178, 205.

Type 2: 177, 197, 206, 201, 180, 176, 185, 200, 197, 192, 198, 188, 189, 203, 192.

Does the data support the claim that the deflection temperature under load for type 1 pipes exceeds that of type 2? Calculate the p-value, using $\alpha = 0.05$, and state your conclusion.

Solution: for this $2-$sample test, we test $H_0 : \mu_1 = \mu_2$ vs. $H_0 : \mu_1 > \mu_2$. We have $\overline{x}_1 = 196.4$, $\overline{x}_2 = 192.0667$, $s_1^2 = 109.8286$, $s_2^2 = 89.06667$, $n_1 = n_2 = 15$, and

$$s_p^2 = \frac{(15-1)109.8286 + (15-1)89.06667}{15+15-2} = 99.44762.$$

We are in Case 2 (σ_1^2, σ_2^2 unknown, small samples), so the test statistic is

$$T_0 = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{1/n_1 + 1/n_2}} \sim t(n_1 + n_2 - 2).$$

The observed value of the test statistic is

$$t_0 = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}} = 1.19.$$

From the $t-$table we get $t_{0.05}(28) = 1.701$, so that $t_0 < t_{0.05}(28)$, meaning that we cannot reject H_0; there is no evidence that the deflection temperature under load for type 1 pipes exceeds that for type 2 pipes. The p-value is

$$P(t(28) > 1.19) \in (0.1, 0.25),$$

since $p(t(28) > 1.313) = 0.1$ and $P(t(28) > 0.683) = 0.25$.

Q107. It is claimed that 15% of a certain population is left-handed, but a researcher doubts this claim. They decide to randomly sample 200 people and use the anticipated small number to provide evidence against the claim of 15%. Suppose 22 of the 200 are left-handed. Compute the $p-$value associated with the hypothesis (assuming a binomial distribution), and provide an interpretation.

Solution: we assume that the binomial distribution is appropriate. Let X denote the (random, i.e. before observing) number of left-handed people in the sample, and let p denote the true proportion of left-handed people in the population. We can set up the formal hypothesis test as follows:

- Model: $X \sim \mathcal{B}(200, p)$, where p is the true proportion.
- $H_0: p = 0.15$ (claim), against $H_1: p < 0.15$ (suspicion: $12/200 = 11\%$)
- Evidence against H_0: small values of X. Observed value: 22.
- $p-$value: $P(X \leq 22)$ under $X \sim \mathcal{B}(200, 0.15)$ (i.e. when H_0 true). But $P(X \leq 22) = \texttt{pbinom}(22, 200, 0.15) \approx 0.0645$. The "small-ish" $p-$value provides some evidence against the claim of 15%.

Q108. A child psychologist believes that nursery school attendance improves children's social perceptiveness (SP). They use 8 pairs of twins, randomly choosing one to attend nursery school and the other to stay at home, and then obtains scores for all 16. In 6 of the 8 pairs, the twin attending nursery school scored better on the SP test. Compute the $p-$value associated with the hypothesis (assuming a binomial distribution), and provide an interpretation.

Solution:

- Model $X \sim \mathcal{B}(8,p)$, where X is $\#$ of pairs in the sample where the twin attending nursery school scored better, and p is the true probability that a twin attending nursery school scores better
- H_0: "Attending nursery school has no effect on SP", $H_0 : p = 0.5$, against H_1: "Attending nursery school improves SP". $H_1 : p > 0.5$
- If H_0 is true, $X \sim \mathcal{B}(8,0.5)$; if H_1 true, X would tend to take **larger values** than it would under H_0. Thus larger values of X provide more evidence against H_0 (in the direction of H_1).
- The $p-$value under H_0 is $P\left(X \geq 6\right), \quad X \sim \mathcal{B}(8,0.5)$. The $p-$value is

$$P\left(X \geq 6\right) = 1 - P(X \leq 5) = \texttt{pbinom}(6,8,0.5) = 0.1445.$$

Interpretation: if there was no real effect, we would see 6 or more improvements out of 8 around 14% of the time, just by chance (which is fairly large, all things considered). The data does not provide compelling evidence against H_0, the null hypothesis (no effect). Consequently, the researcher cannot convince us that attending nursery school improves SP.

Q109. It is claimed that the breaking strength of yarn used in manufacturing drapery material is normally distributed with mean 97 and $\sigma = 2$ psi. A random sample of nine specimens is tested and the average breaking strength is found to be $\overline{X} = 98$ psi. Formulate a test for this situation. Should it be $1-$sided or $2-$sided? What value of α should you use? What conclusion does the test and the sample yield?

Solution:

Q110. A civil engineer is analyzing the compressive strength of concrete. It is claimed that its mean is 80 and variance is known to be 2. A random sample of size 60 yields the sample mean 59. Formulate a test for this situation. Should it be $1-$sided or $2-$sided? What value of α should you use? What conclusion does the test and the sample yield?

Solution:

Q111. The sugar content of the syrup in canned peaches is claimed to be normally distributed with mean 10 and variance 2. A random sample of $n = 10$ cans yields a sample mean 11. Another random sample of $n = 10$ cans yields a sample mean 9. Formulate a test for this situation. Should it be $1-$sided or $2-$sided? What value of α should you use? What conclusion does the test and the sample yield?

Solution:

Q112. A certain power supply is stated to provide a constant voltage output of 10kV. Ten measurements are taken and yield the sample mean of 11kV. Formulate a test for this situation. Should it be $1-$sided or $2-$sided? What value of α should you use? What conclusion does the test and the sample yield?

Solution:

Q113. The mean water temperature downstream from a power water plant cooling tower discharge pipe should be no more than 100F. Past experience has indicated that that the standard deviation is 2F. The water temperature is measured on nine randomly chosen days, and the average temperature is found to be 98F. Formulate a test for this situation. Should it be $1-$sided or $2-$sided? What value of α should you use? What conclusion does the test and the sample yield?

Solution:

Q114. We are interested in the mean burning rate of a solid propellant used to power aircrew escape systems. We want to determine whether or not the mean burning rate is 50 cm/second. A sample of 10 specimens is tested and we observe $\overline{X} = 48.5$. Assume normality with $\sigma = 2.5$.

Solution: we test for $H_0 : \mu = 50$ against $H_1 : \mu \neq 50$.

The $p-$value is

$$2\times \min\left(P\left(Z \geq \frac{48.5-\mu_0}{\sigma/\sqrt{n}}\right), P\left(Z \leq \frac{48.5-\mu_0}{\sigma/\sqrt{n}}\right)\right)$$
$$= 2 \times P(Z \leq -2.4) \approx 2 \times 0.0082 = 0.0164.$$

We do not reject H_0 for $\alpha = 0.01$, but we reject it for $\alpha = 0.05$.

Q115. Ten individuals have participated in a diet modification program to stimulate weight loss. Their weight both before and after participation in the program is shown below:

Before	$195, 213, 247, 201, 187, 210, 215, 246, 294, 310$
After	$187, 195, 221, 190, 175, 197, 199, 221, 278, 285$

Is there evidence to support the claim that this particular diet-modification program is effective in producing mean weight reduction? Use $\alpha = 0.05$. Compute the associated $p-$value.

Solution: this is a paired $t-$test, not a $2-$sample test. We compute the after-before difference:

D_i	$-8, -18, -26, -11, -12, -13, -16, -25, -16, -25$

We test for $H_0 : \mu_D = 0$ against $H_0 : \mu_D < 0$. We have $\overline{d} = -17$, $s_D^2 = 41.11$. The test statistic T_0 follows a $t(10-1)$ distribution under H_0. The observed value of the test statistic is

$$t_0 = \frac{\overline{d}}{s_D/\sqrt{10}} = -8.38.$$

The associated $p-$value is $P(t(9) < -8.38) = P(t(9) > 8.38) < 0.0005$, and so there is enough evidence to reject the hypothesis (at $\alpha = 0.05$) that the diet does not reduce weight.

Q116. We want to test the hypothesis that the average content of containers of a particular lubricant equals 10L against the two-sided alternative. The contents of a random sample of 10 containers are

$$\begin{array}{ccccc} 10.2 & 9.7 & 10.1 & 10.3 & 10.1 \\ 9.8 & 9.9 & 10.4 & 10.3 & 9.5 \end{array}$$

Find the $p-$value of this two-sided test. Assume that the distribution of contents is normal. Note that $\sum_{i=1}^{10} x_i^2 = 1006.79$, if x_i represent the measurements.

a)$0.05 < p < 0.10$ b)$0.10 < p < 0.20$ c)$0.25 < p < 0.40$

d)$0.50 < p < 0.80$ e)none of the preceding

Solution: we test for $H_0 : \mu = 10$ against $H_1 : \mu > 10$. We have $\overline{x} = 10.03$ and $s^2 = 0.08678$. The observed value of the test statistic is

$$t_0 = \frac{\overline{x} - 10}{s/\sqrt{n}} = \frac{10.03 - 10}{\sqrt{0.08678}/\sqrt{10}} = 0.322.$$

There are 10 observations, so we use $\nu = 10 - 1 = 9$ degrees of freedom of for the $2-$sided test (see appropriate table).

The $p-$value is thus $P(T(9) > .322)$, which falls between 0.25 and 0.40 since $P(T(9) > 0.703) = 0.25$ and $P(T(9) > 0.261) = 0.40$. Multiply by two to get the answer.

Q117. An engineer measures the weight of $n = 25$ pieces of steel, which follows a normal distribution with variance 16. The average weight for the sample is $\overline{X} = 6$. They want to test for $H_0 : \mu = 5$ against $H_1 : \mu > 5$. What is the $p-$value for the test?

a)0.05000 b)0.10565 c)0.89435 d)1.0000 e)none of the preceding

Solution: under H_0, we have

$$\begin{aligned}P(\overline{X} > 6|\mu = 5) &= P\left(Z > \frac{6-5}{4/5}\right) = P(Z > 1.25)\\ &= 1 - 0.89435 = 0.10565.\end{aligned}$$

Q118. The thickness of a plastic film (in mm) on a substrate material is thought to be influenced by the temperature at which the coating is applied. A completely randomized experiment is carried out. 11 substrates are coated at 125F, resulting in a sample mean coating thickness of $\overline{x}_1 = 103.5$ and a sample standard deviation of $s_1 = 10.2$. Another 11 substrates are coated at 150F, for which $\overline{x}_2 = 99.7$ and $s_2 = 11.7$ are observed. We want to test equality of means against the two-sided alternative. The value of the appropriate test statistics and the decision are $($for $\alpha = 0.05)$:

a)0.81; Reject H_0.

b)0.81; Do not reject H_0.

c)1.81; Reject H_0.

d)1.81; Do not reject H_0.

e)none of the preceding

Note: assume that population variances are unknown but equal.

Solution: this is two-sample test with small samples and unknown variances, so we need the pooled variance

$$s_p^2 = \frac{(11-1)10.2^2 + (11-1)11.7^2}{11+11-2} = 120.465,$$

or $s_p = 10.97$. The observed value of the test statistic is

$$t_0 = \frac{\overline{x}_1 - \bar{x}_2}{s_p\sqrt{1/n_1 + 1/n_2}} = \frac{103.5 - 99.7}{10.97\sqrt{1/101 + 1/11}} = 0.81.$$

Since $t_{0.05/2}(11+11-2) = 2.086 > t_0$, we don't reject H_0.

Q119. The following output was produced with `t.test` command in `R`.

```
One Sample t-test
data:  x
t = 2.0128, df = 99, p-value = 0.02342
alternative hypothesis: true mean is greater than 0
```

Based on this output, which statement is correct?

a) If the type I error is 0.05, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu > 0$;

b) If the type I error is 0.05, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu \neq 0$;

c) If the type I error is 0.01, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu > 0$;

d) If the type I error is 0.01, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu < 0$;

e) Type I error is 0.02342.

Solution:

Q120. A pharmaceutical company claims that a drug decreases a blood pressure. A physician doubts this claim. They test 10 patients and records results before and after the drug treatment:

```
> Before=c(140,135,122,150,126,138,141,155,128,130)
> After=c(135,136,120,148,122,136,140,153,120,128)
```

At the `R` command prompt, they type:

```
> test.t(Before,After,alternative="greater")
    data:  Before and After
    t = 0.5499, p-value = 0.2946
    alternative hypothesis: true difference in means is greater than 0
    sample estimates: mean of x mean of y
        136.5     133.8
```

Their assistant claims that the command should instead be:

```
> test.t(Before,After,paired=TRUE,alternative="greater")
```

```
data: Before and After t = 3.4825, df = 9, p-value = 0.003456
alternative hypothesis: true difference in means is greater than 0
sample estimates: mean of the differences
    2.7
```

Which answer is best?

a) The assistant uses the correct command. There is not enough evidence to justify that the new drug decreases blood pressure;
b) The assistant uses the correct command. There is enough evidence to justify that the new drug decreases blood pressure for any reasonable choice of α;
c) The physician uses the correct command. There is not enough evidence to justify that the new drug decreases blood pressure;
d) The physician uses the correct command. There is enough evidence to justify that the new drug decreases blood pressure for any reasonable choice of α;
e) Nobody is correct, $t-$tests should not be used here.

Solution: the correct answer is b).

Q121. A company claims that the mean deflection of a piece of steel which is 10ft long is equal to 0.012ft. A buyer suspects that it is bigger than 0.012ft. The following data x_i has been collected:

```
0.0132 0.0138 0.0108 0.0126 0.0136 0.0112 0.0124 0.0116 0.0127 0.0131
```

Assuming normality and that $\sum_{i=1}^{10} x_i^2 = 0.0016$, what are the $p-$value for the appropriate one-sided test and the corresponding decision?

a)$p \in (0.05, 0.1)$ and reject H_0 at $\alpha = 0.05$.

b)$p \in (0.05, 0.1)$ and do not reject H_0 at $\alpha = 0.05$.

c)$p \in (0.1, 0.25)$ and reject H_0 at $\alpha = 0.05$.

d)$p \in (0.1, 0.25)$ and do not reject H_0 at $\alpha = 0.05$.

e)none of the preceding

Solution: we test for $H_0 : \mu = 0.012$ against $H_1 : \mu > 0.012$. As the variance of the underlying population is unknown, we will be using the one-sided $t-$test. The estimated sample variance is

$$S^2 = \frac{1}{10-1}\left(\sum_{i=1}^{10} x_i^2 - \frac{1}{10}\left(\sum_{i=1}^{10} x_i\right)^2\right) = 0.00000102.$$

The observed mean is $\overline{x} = 0.0125$. We calculate the corresponding $p-$value:

$$P(\overline{X} > \overline{x}) = P\left(\frac{\overline{X}-\mu_0}{S/\sqrt{n}} > \frac{\overline{x}-\mu_0}{S/\sqrt{n}}\right) = P\left(\frac{\overline{X}-\mu_0}{S/\sqrt{n}} > \frac{0.0125-0.012}{\sqrt{0.00000102/10}}\right)$$

$$= P(t(10-1) > 1.5638) = P(t(9) > 1.5638) \in (0.05, 0.1)$$

and we do not reject H_0 at $\alpha = 0.05$.

Q122. In an effort to compare the durability of two different types of sandpaper, 10 pieces of type A sandpaper were subjected to treatment by a machine which measures abrasive wear; 11 pieces of type B sandpaper were subjected to the same treatment. We have the following observations:

```
xA 27 26 24 29 30 26 27 23 28 27
xB 24 23 22 27 24 21 24 25 24 23 20
```

Note that $\sum x_{A,i} = 267$, $\sum x_{B,i} = 257$, $\sum x_{A,i}^2 = 7169$, $\sum x_{B,i}^2 = 6041$. Assuming normality and equality of variances in abrasive wear for A and B, we want to test for equality of mean abrasive wear for A and B. The appropriate $p-$value is

a)$p < 0.01$

b)$p > 0.2$

c)$p \in (0.01, 0.05)$

d)$p \in (0.1, 0.2)$

e)$p \in (0.05, 0.1)$

f) none of the preceding

Solution: this is a two sample test. We test for $H_0 : \mu_A = \mu_B$ against $H_1 : \mu_A \neq \mu_B$. We compute $s_A^2 = 4.45$, $s_B^2 = 3.65$, $s_p^2 = 4.03$, $\overline{x}_A = 26.71$, $\overline{x}_B = 23.26$. The $p-$value is

$$\begin{aligned} 2P(\overline{X}_A - \overline{X}_B > \overline{x}_A - \overline{x}_B) &= 2P\left(\frac{\overline{X}_A - \overline{X}_B}{S_p\sqrt{1/n_A + 1/n_B}} > \frac{3.34}{\sqrt{4.03}\sqrt{1/10+1/11}}\right) \\ &= 2P(t(10+11-2) > 3.8037) \\ &= 2P(t(19) > 3.8037) < 0.01, \end{aligned}$$

because $P(t(19) > 3.8037) < 0.005$.

Q123. The following output was produced with `t.test` command in `R`.

```
One Sample t-test
data:  x
t = 32.9198, df = 999, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
```

Based on this output, which statement is correct?

a) If the type I error is 0.05, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu > 0$;

b) If the type I error is 0.05, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu \neq 0$;

c) If the type I error is 0.01, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu > 0$;

d) If the type I error is 0.01, then we reject $H_0 : \mu = 0$ in favour of $H_1 : \mu < 0$;

e) None of the preceding.

Solution: the correct answer is a).

Q124. Consider a sample $\{X_1, \ldots, X_{10}\}$ from a normal population $X_i \sim \mathcal{N}(4, 9)$. Denote by $\overline{X}$ and S^2 the sample mean and the sample variance, respectively. Find c such that

$$P\left(\frac{\overline{X} - 4}{S/\sqrt{10}} \leq c\right) = 0.99$$

a)1.833 b)2.326 c)1.645 d)2.821 e)none of the preceding

Solution: an equivalent statement is to find c such that

$$P\left(\frac{\overline{X}-4}{S/\sqrt{10}} \geq c\right) = 0.01.$$

We know that $\frac{\overline{X}-4}{S/\sqrt{10}} \sim T(10-1)$. From the table, we have

$$P(t(9) > 2.821) = 0.01,$$

thus $c = 2.821$.

Q125. Consider the following dataset:

```
2.6  3.7  0.8  9.6  5.8 -0.8  0.7 0.6
4.8  1.2  3.3  5.0  3.7  0.1 -3.1 0.3
```

The median and the interquartile range of the sample are, respectively:

a)$2.4, 3.3$ b)$1.9, 3.8$ c)$1.9, 1.8$ d)$2.9, 12.2$ e)none of the preceding

Solution: the correct answer is a).

Q126. An article in *Computers and Electrical Engineering* considered the speed-up of cellular neural networks (CNN) for a parallel general-purpose computing architecture. Various speed-ups are observed:

```
3.77  3.35  4.21  4.03  4.03  4.63
4.63  4.13  4.39  4.84  4.26  4.60
```

Assume that the population is normally distributed. The 99% C.I. for the mean speed-up is:

a)$[4.155, 4.323]$ b)$[3.863, 4.615]$ c)$[4.040, 4.438]$

d)$[3.77, 4.60]$ e)none of the preceding

Solution: let's do this in `R`:

```
> x=c(3.77,3.35,4.21,4.03,4.03,4.63,4.63,4.13,4.39,4.84,4.26,4.60);
> alpha=0.01
> n=length(x)
> mean(x)-qt(1-alpha/2,n-1)*sd(x)/sqrt(n)
    3.863531
> mean(x)+qt(1-alpha/2,n-1)*sd(x)/sqrt(n)
    4.614802
```

The function `qt(beta,nu)` finds the β quantile of the Student $t-$distributions with ν degrees of freedom; it is equivalent to the `qnorm()` function.

Q127. An engineer measures the weight of $n = 25$ pieces of steel, which follows a normal distribution with variance 16. The average observed weight for the sample is $\overline{x} = 6$. The two-sided 95% C.I. for the mean μ is:

a)$[-0.272, 12.272]$ b)$[4.432, 7.568]$ c)$[3.250, 8.750]$

d)$[4.120, 7.522]$ e)none of the preceding

Solution: since the weights follow a normal distribution with known variance, the C.I. is

$$\overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 6 \pm 1.96\frac{4}{5} = [4.432, 7.568].$$

Q128. Assume that random variables $\{X_1, \ldots, X_8\}$ follow a normal distribution with mean 2 and variance 24. Independently, assume that random variables $\{Y_1, \ldots, X_{16}\}$ follow a normal distribution with mean 1 and variance 16. Let $\overline{X}$ and $\overline{Y}$ be the corresponding sample means. Then $P(\overline{X} + \overline{Y} > 4)$ is:

a)0.7721 b)0.30855 c)0.69165 d)0.9883 e)none of the preceding

Solution: since the X_i and Y_j are independent,

$$\overline{X} + \overline{Y} \sim \mathcal{N}\left(2+1, \frac{24}{8} + \frac{16}{16}\right) = \mathcal{N}(3,4).$$

Thus

$$P(\overline{X} + \overline{Y} > 4) = P\left(\frac{\overline{X} + \overline{Y} - 3}{\sqrt{4}} > \frac{4-3}{\sqrt{4}}\right) = P(Z > 0.5)$$
$$= 1 - P(Z < 0.5) = 1 - \Phi(0.5) \approx 1 - 0.6915 = 0.3085.$$

Q129. A medical team wants to test whether a particular drug decreases diastolic blood pressure. Nine people have been tested. The team measured blood pressure before (X) and after (Y) applying the drug. The corresponding means were $\overline{X} = 91$, $\overline{Y} = 87$. The sample variance of the differences was $S_D^2 = 25$. The $p-$value for the appropriate one-sided test is between:

a)0 and 0.025 b)0.025 and 0.05 c)0.05 and 0.1

d)0.1 and 0.25 e)0.25 and 1 f) none of the preceding

Solution: this is a one-sided paired $t-$test, $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X > \mu_Y$. The observed difference of the means is $\overline{d} = 4$. The associated $p-$value is is

$$\begin{aligned} P(\overline{D} \geq \overline{d}) &= P(\overline{D} \geq 4) = P\left(\frac{\overline{D}}{S_D/\sqrt{n}} \geq \frac{4}{5/3}\right) \\ &= P(t(n-1) > 2.4) = P(t(8) > 2.4) < 0.025, \end{aligned}$$

since $P(t(8) > 2.4) \in (0.01, 0.025)$ according to the table.

Q130. A researcher studies a difference between two programming languages. Twelve experts familiar with both languages were asked to write a code for a particular function using both languages and the time for writing those codes was registered. The observations are as follows.

Expert	01	02	03	04	05	06	07	08	09	10	11	12
Lang 1	17	16	21	14	18	24	16	14	21	23	13	18
Lang 2	18	14	19	11	23	21	10	13	19	24	15	29

Construct a 95% C.I. for the mean difference between the first and the second language. Do we have any evidence that one of the languages is preferable to the other (i.e. the average time to write a function is shorter)?

Solution:

../Images/WechatIMG1.pdf

Q131. For a set of 12 pairs of observations on (x_i, y_i) from an experiment, the following summary for x and y is obtained:

$$\sum_{i=1}^{12} x_i = 25,\ \sum_{i=1}^{12} y_i = 432,\ \sum_{i=1}^{12} x_i^2 = 59,\ \sum_{i=1}^{12} x_i y_i = 880.5,\ \sum_{i=1}^{12} y_i^2 = 15648.$$

The estimated value of y at $x = 5$ from the least squares regression line is:

a)27.78 b)47.77 c)41.87 d)55.97 e)none of the preceding

Solution: assuming the linear regression model is warranted, the estimated value at $x = 5$ is given by

$$\hat{y}(5) = b_0 + b_1(5).$$

We have

$$\begin{aligned}
\overline{x} &= \frac{1}{12}\sum_{i=1}^{12} x_i = \frac{25}{12}, \quad \overline{y} = \frac{1}{12}\sum_{i=1}^{12} y_i = 36 \\
S_{xx} &= \sum_{i=1}^{12} x_i^2 - 12\overline{x}^2 = \frac{83}{12}, \quad S_{xy} = \sum_{i=1}^{12} x_i y_i - 12\overline{xy} = -19.5, \\
b_1 &= -\frac{19.5}{83/12} = -2.82, \quad b_0 = 36 - (-2.82)(25/12) = 41.87, \\
\hat{y}(5) &= 41.87 - 2.82(5) = 27.78.
\end{aligned}$$

Q132. Assuming that the simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ is appropriate for $n = 14$ observations, the estimated regression line is computed to be

$$\hat{y} = 0.66490 + 0.83075x.$$

Given that $S_{yy} = 4.1289$ and $S_{xy} = 4.49094$, compute the estimated standard error for the slope.

a)0.3176 b)0.0783 c)0.0855 d)0.0073 e)none of the preceding

Solution: the estimated standard error for the slope is

$$\mathrm{se}(b_1) = \sqrt{\hat{\sigma}^2/S_{xx}}.$$

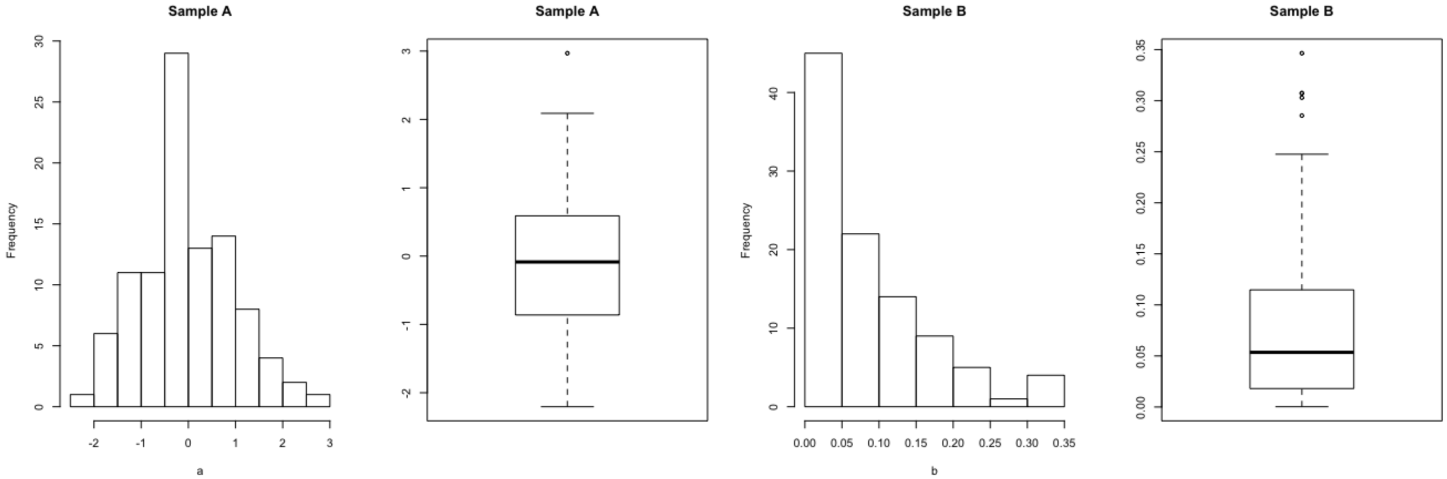
But

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} \quad \text{and} \quad S_{xx} = \frac{S_{xy}}{b_1}$$

so that

$$\mathrm{se}(b_1) = \sqrt{\frac{b_1(S_{yy} - b_1 S_{xy})}{(n-2)S_{xy}}} = \sqrt{\frac{0.83075(4.1289 - 0.83075 \cdot 4.49094)}{(14-2)4.49094}}$$
$$= 0.07833.$$

Q136. The following charts show a histogram and a boxplot for two samples, A and B. Based on these charts, we may conclude that



a)only A arises from a normal population

b)only B arises from a normal population

c)both A and B arise from a normal population

Solution: it is reasonable to expect that A arises from a normal population, but the skew and asymmetric distribution for B means it does not come from a normal distribution.

Q137. We have a dataset with $n = 25$ pairs of observations (x_i, y_i), and

$$\sum_{i=1}^{n} x_i = 325.000,\ \sum_{i=1}^{n} y_i = 658.972,$$

$$\sum_{i=1}^{n} x_i^2 = 5525.000,\ \sum_{i=1}^{n} x_i y_i = 11153.588,\ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Note that $t_{0.05/2}(23) = 2.069$. The point estimate for the slope of the regression line is

a)1.99 b)-1.99 c)0.49 d)0.59 e)none of the preceding

Solution: we have $b_1 = \frac{S_{xy}}{S_{xx}}$. Note that

$$\overline{x} = \frac{1}{25}\sum_{i=1}^{25} x_i = 13,\ \overline{y} = \frac{1}{25}\sum_{i=1}^{25} y_i = 26.359,\ S_{xx} = \sum_{i=1}^{25} x_i^2 - 25\overline{x}^2 = 1300$$

$$S_{xy} = \sum_{i=1}^{25} x_i y_i - 25\overline{xy} = 2586.952,\ S_{yy} = \sum_{i=1}^{25} y_i^2 - 25\overline{y}^2 = 5261.613,$$

$$b_1 = \frac{2586.952}{1300} = 1.99, \quad b_0 = 26.359 - (1.990)(13) = 0.49.$$

$\hat{y}(30) = 0.49 + 1.99(30) = 60.19.$

Q138. We have a dataset with $n = 25$ pairs of observations (x_i, y_i), and

$$\sum_{i=1}^{n} x_i = 325.000,\ \sum_{i=1}^{n} y_i = 658.972,$$

$$\sum_{i=1}^{n} x_i^2 = 5525.000,\ \sum_{i=1}^{n} x_i y_i = 11153.588,\ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Note that $t_{0.05/2}(23) = 2.069$. The point estimate for the intercept of the regression line is

a)1.99 b)-1.99 c)0.49 d)0.59 e)none of the preceding

Solution: see answer to **Q137**.

Q139. We have a dataset with $n = 25$ pairs of observations (x_i, y_i), and

$$\sum_{i=1}^{n} x_i = 325.000,\ \sum_{i=1}^{n} y_i = 658.972,$$

$$\sum_{i=1}^{n} x_i^2 = 5525.000,\ \sum_{i=1}^{n} x_i y_i = 11153.588,\ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Note that $t_{0.05/2}(23) = 2.069$. What is the prediction of y for $x = 30$?

a)60.19 b)16.67 c)30 d)30.54 e)none of the preceding

Solution: see answer to **Q137**.

Q140. We have a dataset with $n = 25$ pairs of observations (x_i, y_i), and

$$\sum_{i=1}^{n} x_i = 325.000, \ \sum_{i=1}^{n} y_i = 658.972,$$

$$\sum_{i=1}^{n} x_i^2 = 5525.000, \ \sum_{i=1}^{n} x_i y_i = 11153.588, \ \sum_{i=1}^{n} y_i^2 = 22631.377.$$

Note that $t_{0.05/2}(23) = 2.069$. Is the linear regression significant?

Solution: we are testing for $H_0 : \beta_1 = 0$ against $H_0 : \beta_1 \neq 0$. Let's use $\alpha = 0.05$. If we reject H_0 in favour of H_1, then the evidence suggests that there is a linear relationship between X and Y.

Under H_0, the test statistic $T_0 = \frac{b_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} \sim t_{0.05/2}(23)$. Using the results from the previous questions, we have

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{5261.613 - 1.99 \cdot 2586.952}{23} = 4.94;$$

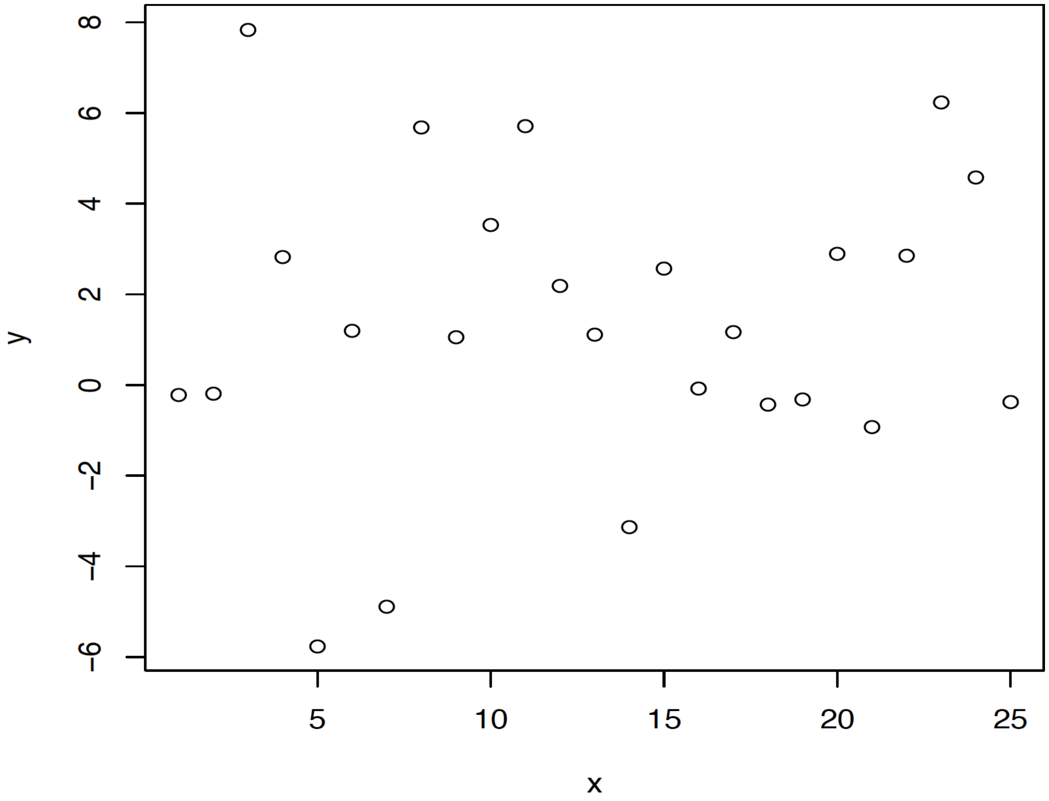
the observed statistic is thus

$$t_0 = \frac{1.99}{\sqrt{4.94/1300}} = 32.27 < t_{0.05/2}(23) = 2.069.$$

Thus we reject H_0 in favour of significance of regression.

Q141. For the following data the correlation coefficient is most likely to be

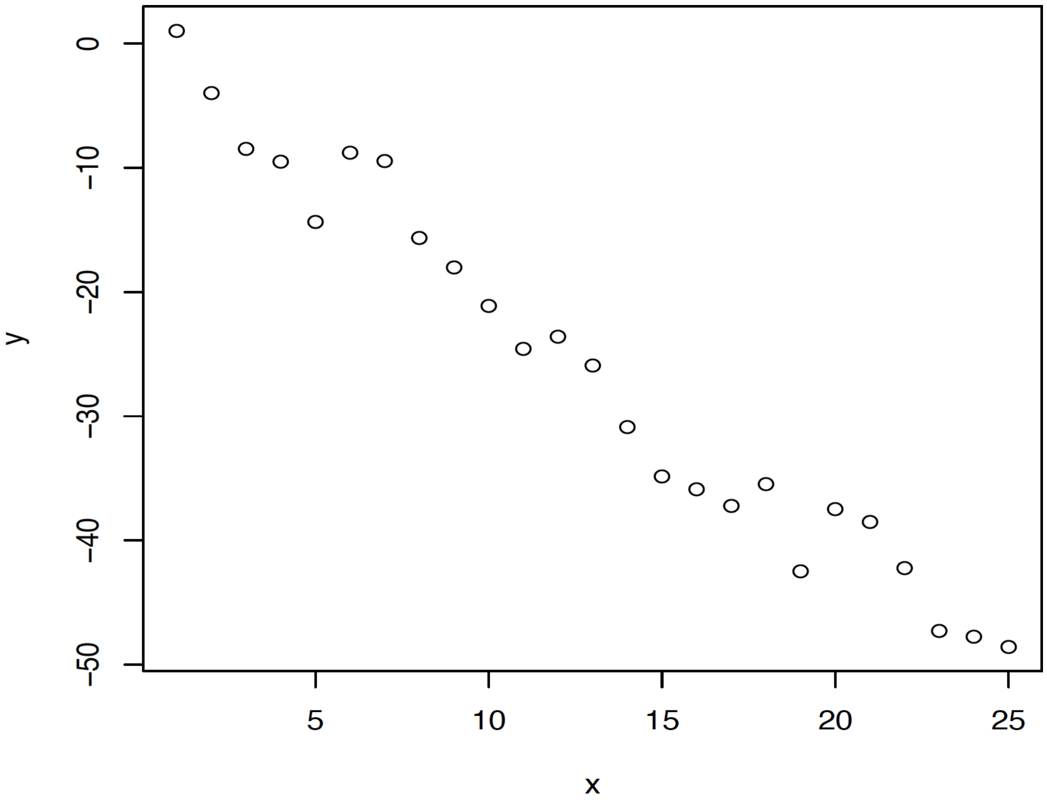
a)0.01 b)0.98 c)-0.5 d)-0.98



Solution: the scatterplot shows no real structure or relationship between x and y. The most likely answer si $\rho = 0.01$.

Q142. For the following data the correlation coefficient is most likely to be

a)0.01 b)0.98 c)-0.5 d)-0.98



Solution: the scatter plot shows a clear anti-correlated pattern between x and y – when x increases, y decreases and vice-versa. The most likely value is $\rho = -0.98$.

Q143. A company employs 10 part-time drivers for its fleet of trucks. Its manager wants to find a relationship between number of km driven (X) and number of working days (Y) in a typical week. The drivers are hired to drive half-day shifts, so that 3.5 stands for 7 half-day shifts.

The manager wants to use the linear regression model $Y = \beta_0 + \beta_1 x + \epsilon$ on the following data:

	1	2	3	4	5	6	7	8	9	10
x	825	215	1070	550	480	920	1350	325	670	1215
y	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

Note that $\sum x_i^2 = 7104300$, $\sum y_i^2 = 99.75$, and $\sum x_i y_i = 26370$. What is the fitted regression line?

Solution: we have

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 = 1297860$$

and

$$S_{xy} = 4653,$$

so that

$$b_1 = S_{xy}/S_{xx} = 0.0036,$$

and

$$b_0 = \sum_{i=1}^{n} y_i/n - b_1 \sum_{i=1}^{n} x_i = 0.1181;$$

hence the fitted line is $\hat{y} = 0.1181 + 0.0036x$.

Q144. Using the data from question **Q143**, what value is the correlation coefficient of x and y closest to?

a)0.437 b)0.949 c)0.113 d)1.123 e)none of the preceding

Solution: as in question **Q143**, we have $S_{xx} = 12978600$ and $S_{xy} = 4653$. Furthermore, we have

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2 = 18.525,$$

so that the correlation coefficient is

$$\rho_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{4653}{\sqrt{18.525 \cdot 1297860}} \approx 0.949$$

Q145. We want to test significance of regression, i.e. $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$. The value of the appropriate statistic and the decision for $\alpha = 0.05$ is:

a)8.55; do not reject H_0

b)2.31; reject H_0

c)8.55; reject H_0

d)2.31; do not reject H_0

e)none of the preceding

Solution: the estimated variance is

$$\hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{1.8434}{8} = 0.23.$$

Consequently, the test statistic is

$$t_0 = \frac{b_1}{\sqrt{\hat{\sigma}^2/S_{xx}}} = \frac{0.0036}{\sqrt{0.23/1297860}} = 8.551701.$$

Since $t_{0.05/2}(n-2) = t_{0.025}(8) = 2.306$, we reject H_0.

Q146. Regression methods were used to analyze the data from a study investigating the relationship between roadway surface temperature in F (x) and pavement defection (y). Summary quantities were $n = 20$,

$$\sum y_i = 12.75,\ \sum y_i^2 = 8.86,\ \sum x_i = 1478\ \sum x_i^2 = 143,215.8\ \sum x_i y_i = 1083.67.$$

a)Calculate the least squares estimates of the slope and intercept. Estimate σ^2.

b)Use the equation of the fitted line to predict what pavement deflection would be observed when the surface temperature is 90F.

c)Give a point estimate of the mean pavement deflection when the surface is 85F.

d)What change in mean pavement deflection would be expected for a 1F change in surface temperature?

Solution:

a) We have

$$b_1 = \frac{S_{xy}}{S_{xx}}, \quad b_0 = \overline{y} - b_1\overline{x}, \quad \hat{\sigma}^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2},$$

where

$$\begin{aligned} S_{xy} &= \sum x_i y_i - \tfrac{1}{n}(\sum x_i)(\sum y_i) = 141.445 \\ S_{xx} &= \sum x_i^2 - \tfrac{1}{n}(\sum x_i)^2 = 33991.6 \\ S_{yy} &= \sum y_i^2 - \tfrac{1}{n}(\sum y_i)^2 = 0.731875, \end{aligned}$$

so that $b_1 = 0.00416$, $b_0 = 0.32999$, and $\hat{\sigma}^2 = 0.00797$

b) $\hat{y}(90) = b_0 + b_1 \cdot 90 = 0.70$

c) The question can be rephrased as “use the equation of the fitted line to predict what pavement deflection would be observed when the surface temperature is 85F”, i.e. $\hat{y}(85) = b_0 + b_1 \cdot 85 = 0.68$.

d) That is simply the slope: $b_1 = 0.00416$

Q147. Consider the data from **Q146**.

a) Test for significance of regression using $\alpha = 0.05$. Find the p-value for this test. What conclusion can you draw?

b) Estimate the standard errors of the slope and intercept.

Solution:

a) We test for $H_0 : \beta_1 = 0$, against $H_1 : \beta_1 \neq 0$. The test statistic is $T_0 = \frac{b_1 - 0}{\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}} \sim t(n-2)$. Its observed value is $t_0 = \frac{b_1 - 0}{\sqrt{\hat{\sigma}^2 / S_{xx}}} = 8.6$. The p-value (using $t(18)$ table) is $2P(t_{18} > 8.6) < 0.001$, and so we reject H_0 in favour of a linear relationship between x and y.

b) The standard errors are

$$\mathrm{se}(b_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}, \quad \mathrm{se}(b_0) = \sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}}\right]}.$$

So, $\mathrm{se}(b_1) = 0.00048$, $\mathrm{se}(b_0) = 0.04098$.

Q148. Solve this question using `R`.

a) Generate a sample `x` of size $n = 100$ from a normal distribution;

b) Define `y=1+2*x+rnorm(100)`;

c) Plot scatter plot;

d) Find the estimators of the regression parameters and add the line to the scatter plot;

f) Compute the correlation coefficient

g) Plot the residuals;

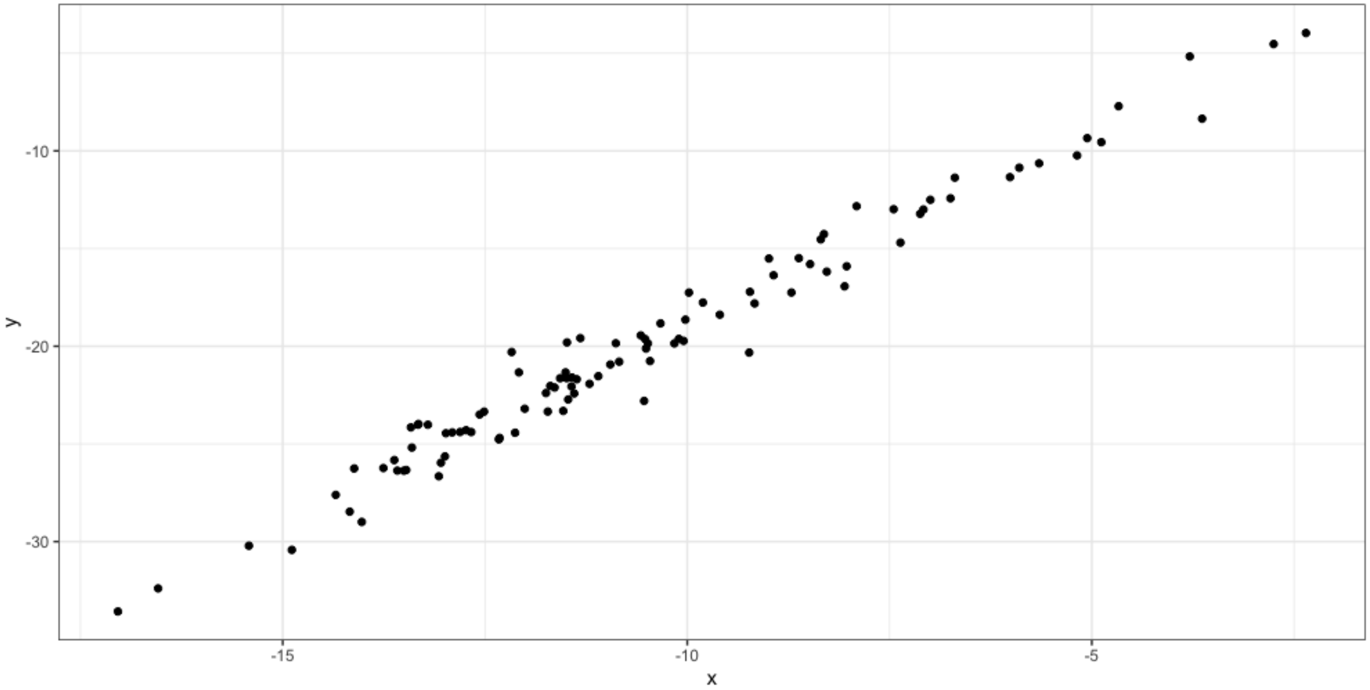
h) Comment on your results.

Solution: The following code will do the trick.

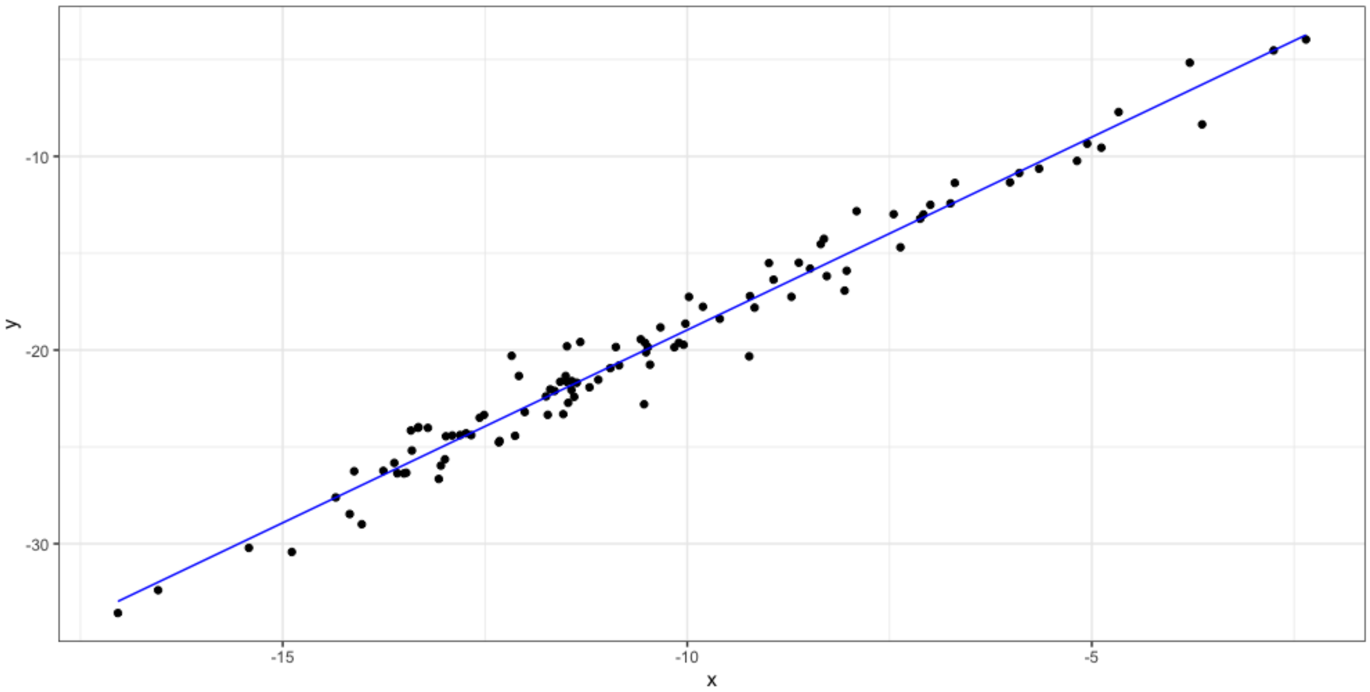
```
> library(ggplot2) ## required for plotting
> set.seed(1234) ## so we all get the same results

# a), b), c)
> x = rnorm(100, mean = -10,sd=3)
> y = 1 + 2*x + rnorm(100)
> data.Q148 = data.frame(x,y)
> ggplot(data.Q148) + geom_point(aes(x=x, y=y)) +
    theme_bw()

# d)
> model <- lm(y ~ x, data=data.Q148)
> summary(model)
    Call:
    lm(formula = y ~ x, data = data.Q148)
    Residuals:
         Min       1Q   Median       3Q      Max
    -2.88626 -0.61401  0.00236  0.58645  2.98774

    Coefficients:
                Estimate Std. Error t value Pr(>|t|)
    (Intercept)  0.95020    0.37674   2.522   0.0133 *
    x            1.99131    0.03459  57.566   <2e-16 ***
    ---
    Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Residual standard error: 1.037 on 98 degrees of freedom
    Multiple R-squared:  0.9713,Adjusted R-squared:  0.971
    F-statistic:  3314 on 1 and 98 DF,  p-value: < 2.2e-16

> ggplot(model) + geom_point(aes(x=x, y=y)) +
    geom_line(aes(x=x, y=.fitted), color="blue" ) + theme_bw()
```

```
# e)
> Sxy=sum((x-mean(x))*(y-mean(y)))
> Sxx=sum((x-mean(x))^2)
> Syy=sum((y-mean(y))^2)
> rho=Sxy/(sqrt(Sxx*Syy))
> rho
    0.9855334

# f)
> ggplot(model) + geom_point(aes(x=x, y=y)) +   ### plotting residuals
    geom_line(aes(x=x, y=.fitted), color="blue" ) +
    geom_linerange(aes(x=x, ymin=.fitted, ymax=y), color="red") +
    theme_bw()

> ggplot(model) +
    geom_point(aes(x=.fitted, y=.resid)) + theme_bw()
```

The corresponding plots are shown in the following slides. You may get different results if you use a different normal distribution to generate x.



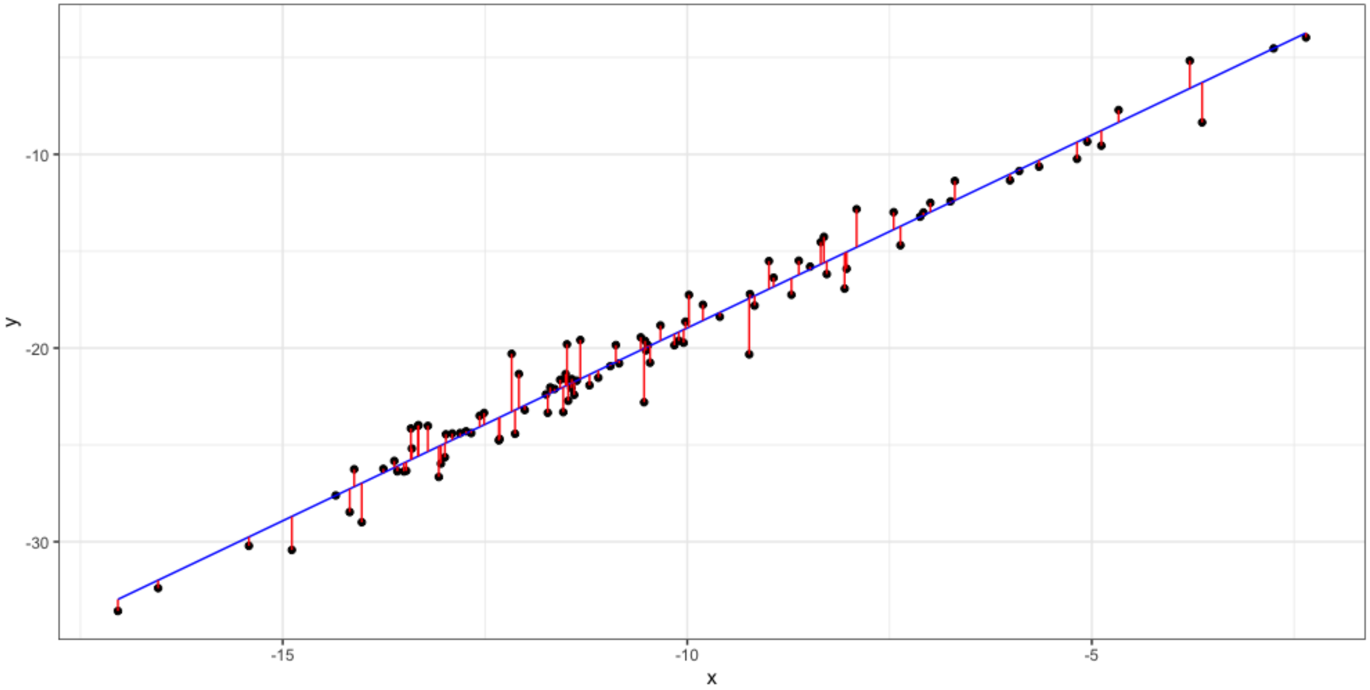
scatterplot



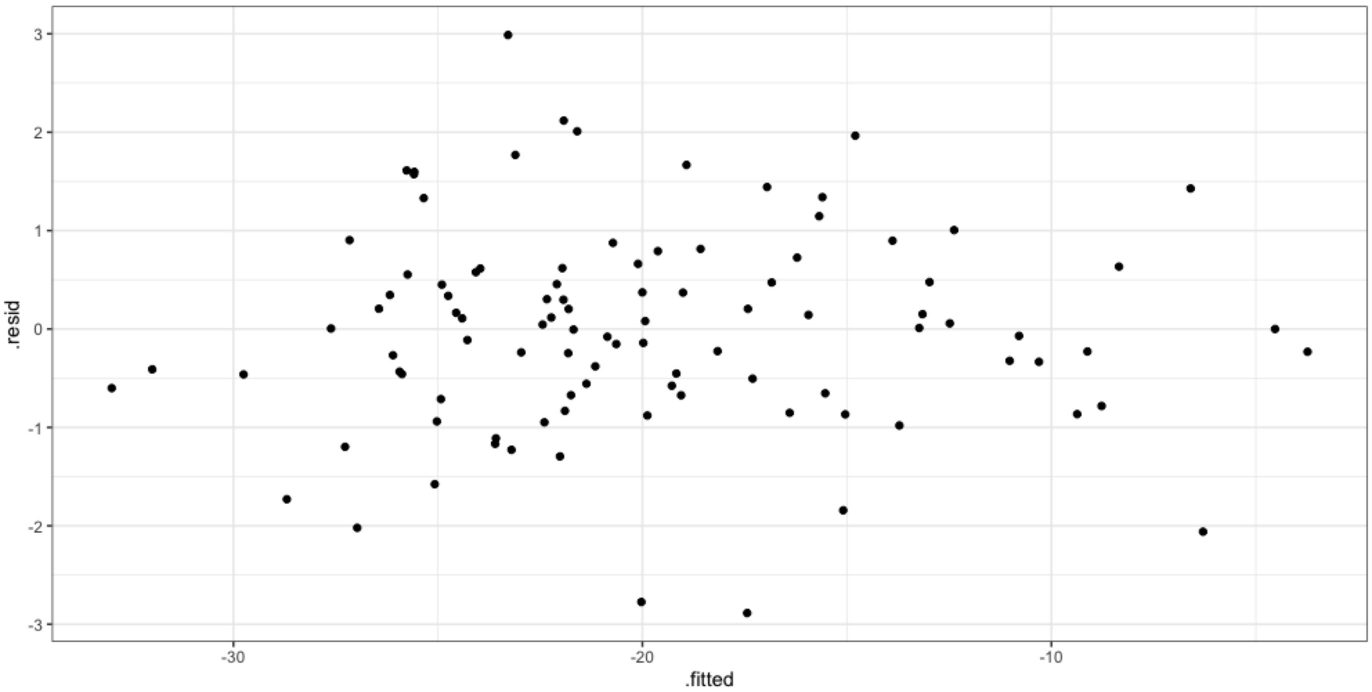
line of best fit: $\hat{y} = 0.95020 + 1.99131x$

quite close to the true relationship

coefficient of correlation: $\rho = 0.986$



residuals against fitted
they are quite small



residuals against fitted (rotated)

no specific structure, seems like the linear model is warranted