EE2211 Tutorial 3

Dr Feng LIN

Q1

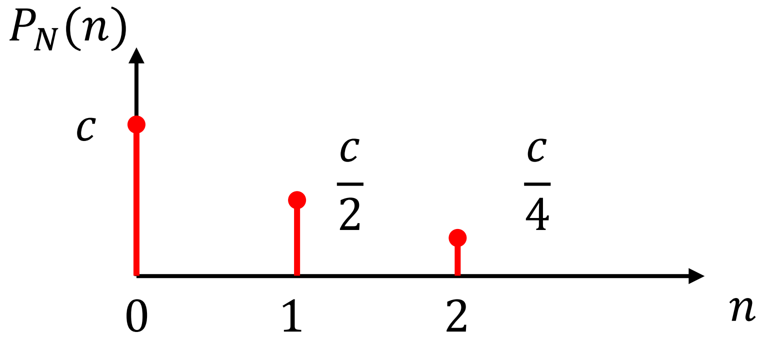
The random variable N has probability mass function (PMF)

$$P_N(n) = \begin{cases} c\left(\frac{1}{2}\right)^n, & n = 0,1,2 \\ 0, & \text{otherwise} \end{cases}$$

(a) What is the value of the constant *c*?

(b) What is $Pr[N \leq 1]$?

Q1



(a) We wish to find the value of c that makes the PMF sum up to one.

$$\sum_{n=0}^{2} P_N(n) = c + \frac{c}{2} + \frac{c}{4} = 1 \quad \text{implying } c = \frac{4}{7}$$

(b) The probability that $N \leq 1$ is

$$Pr[N \leq 1] = Pr[N = 0] + Pr[N = 1] = 4/7 + 2/7 = 6/7\ .$$

Q2

The random variable X has probability density function (PDF)

$$f_X = \begin{cases} cx, & 0 \leq x \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

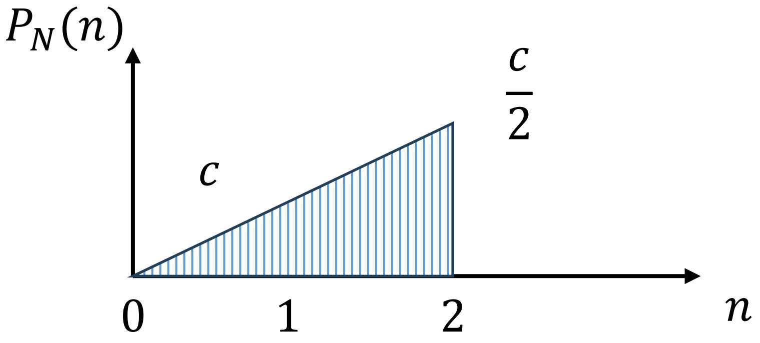
Use the PDF to find

(a) The constant c,

(b) $Pr[0 \leq X \leq 1]$

(c) $Pr\left[-\frac{1}{2} \leq X \leq \frac{1}{2}\right]$

Q2



(a) From the above PDF we can determine the value of c by integrating the PDF and setting it equal to 1.

$$\int_0^2 cx\,dx = c\left[\frac{x^2}{2}\right]_0^2 = c\left[\frac{4}{2}\right] = 2c = 1$$

Therefore $c = 1/2$.

(b) $Pr[0 \leq X \leq 1] = \int_0^1 x/2\,dx = 1/4$

(c) $Pr[-1/2 \leq X \leq 1/2] = \int_0^{1/2} x/2\,dx = 1/16$

Q3

Let A = {resistor is within 50Ω of the nominal value}. The probability that a resistor is from machine B is $\Pr[B] = 0.3$. The probability that a resistor is acceptable, i.e., within 50Ω of the nominal value, is $\Pr[A] = 0.78$. Given that a resistor is from machine B, the conditional probability that it is acceptable is $\Pr[A|B] = 0.6$. What is the probability that an acceptable resistor comes from machine B?

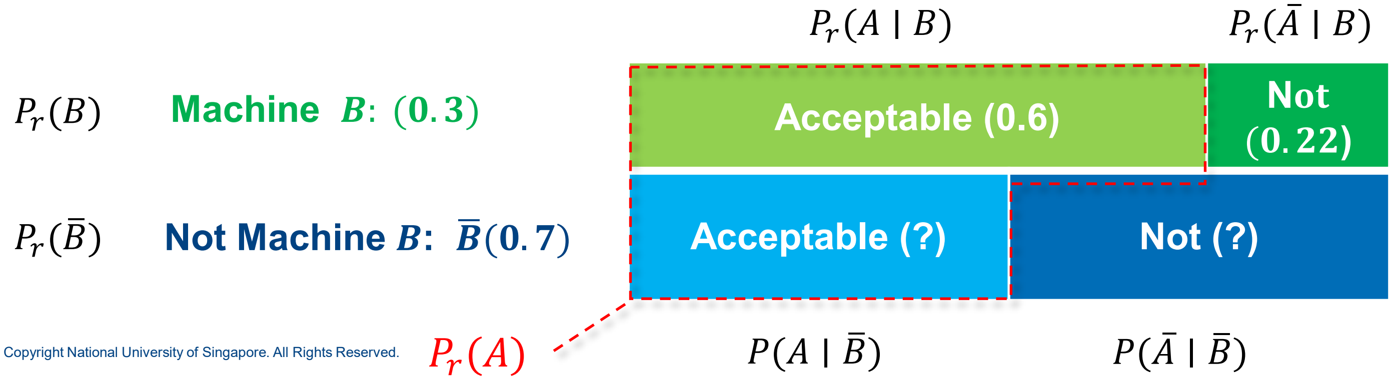
Q3

Step 1: Identify the events and given probability

- Let A be the event that a resistor is acceptable (within 50Ω of the nominal value)
- Let B be the event that a resistor is from machine B

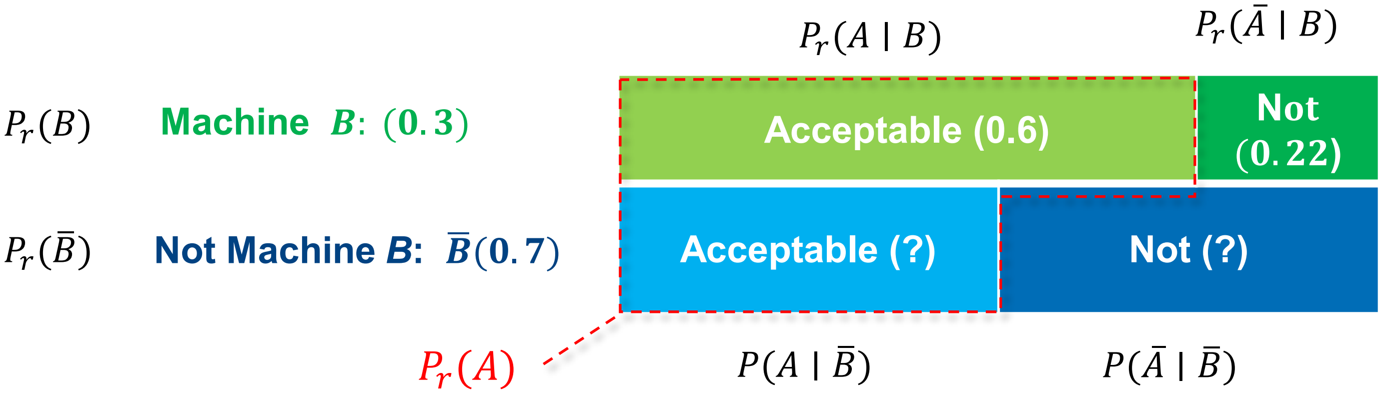
Then we have

$$P_r[B] = 0.3 \qquad P_r[A] = 0.78 \qquad P_r[A|B] = 0.6$$



Q3

$$P_r[B] = 0.3 \qquad P_r[A] = 0.78 \qquad P_r[A|B] = 0.6$$



Step2: Use Bayes' theorem

$$P_r(B|A) = \frac{P_r(A|B)P_r(B)}{P_r(A)} = \frac{0.6 \times 0.3}{0.78} \approx 0.23$$

Q4

Consider tossing a fair six-sided die. There are only six outcomes possible, Ω = {1, 2, 3, 4, 5, 6}. Suppose we toss two dice and assume that each throw is independent.

(a) What is the probability that the sum of the dice equals seven?

i. List out all pairs of possible outcomes together with their sums from the two throws. (hint: enumerate all the items in range(1,7))

ii. Collect all of the (a, b) pairs that sum to each of the possible values from two to twelve (including the sum equals seven). (hint: use dictionary from collections import default dict to collect all of the (a, b) pairs that sum to each of the possible values from two to twelve)

(b) What is the probability that half the product of three dice will exceed their sum?

Q4 (a)

Step1: Determine the sample space

Each die has 6 faces, so when we rolling two dices, the total outcomes is

$$6 \times 6 = 36$$

Step2: List all pairs that sum to 7

$(1,6)\quad(2,5)\quad(3,4)\quad(4,3)\quad(5,2)\quad(2,5)\quad(6,1)$

There are 6 pairs that satisfy this condition.

Step 3: Calculate probability

$$P_r(sum = 7) = \frac{6}{36} = \frac{1}{6}$$

Q4

(a) What is the probability that the sum of the dice equals seven?

i. The first thing to do is characterize the measurable function for this as X : (a, b) → (a + b). Next, we associate all of the (a, b) pairs with their sum. A Python dictionary can be created like this:

```
#(i)
d={(i,j):i+j for i in range(1,7) for j in range(1,7)}

print('d=')
for key, value in d.items():
    print(f"{key}: {value}")
```

Note: Another useful data type built into Python is the dictionary (see Mapping Types — dict). Dictionaries are sometimes found in other languages as "associative memories" or "associative arrays". Unlike sequences, which are indexed by a range of numbers, dictionaries are indexed by keys, which can be any immutable type; strings and numbers can always be keys.
https://docs.python.org/3/tutorial/datastructures.html#dictionaries
On an abstract level, it consists of a key with an associated value.

Q4

```python
#(i)
d={(i,j):i+j for i in range(1,7) for j in range(1,7)}

print('d=')
for key, value in d.items():
    print(f"{key}: {value}")
```

```
d=
(1, 1): 2
(1, 2): 3
(1, 3): 4
(1, 4): 5
(1, 5): 6
(1, 6): 7
(2, 1): 3
(2, 2): 4
(2, 3): 5
(2, 4): 6
(2, 5): 7
(2, 6): 8
(3, 1): 4
(3, 2): 5
(3, 3): 6
(3, 4): 7
(3, 5): 8
(3, 6): 9
(4, 1): 5
(4, 2): 6
(4, 3): 7
(4, 4): 8
(4, 5): 9
(4, 6): 10
(5, 1): 6
(5, 2): 7
(5, 3): 8
(5, 4): 9
(5, 5): 10
(5, 6): 11
(6, 1): 7
(6, 2): 8
(6, 3): 9
(6, 4): 10
(6, 5): 11
(6, 6): 12
```

Q4

(ii) The next step is to collect all of the (a, b) pairs that sum to each of the possible values from two to twelve. This step can be considered as inverting the dictionary mapping.

```
#(ii) collect all of the (a, b) pairs that sum to each of the possible values from two to twelve
from collections import defaultdict
dinv = defaultdict(list)
for i,j in d.items(): dinv[j].append(i)

print('dinv=')
for key, value in dinv.items():
    print(f"{key}: {value}")
```

```
dinv=
2: [(1, 1)]
3: [(1, 2), (2, 1)]
4: [(1, 3), (2, 2), (3, 1)]
5: [(1, 4), (2, 3), (3, 2), (4, 1)]
6: [(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)]
7: [(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)]
8: [(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)]
9: [(3, 6), (4, 5), (5, 4), (6, 3)]
10: [(4, 6), (5, 5), (6, 4)]
11: [(5, 6), (6, 5)]
12: [(6, 6)]
```

For example, dinv[7] contains the following list of pairs that sum to seven, [(1, 6), (2, 5), (5, 2),(6, 1), (4, 3), (3, 4)].

Q4

The next step is to compute the probability measured for each of these items. Using the independence assumption, this means we have to compute the sum of the products of the individual item probabilities in dinv. Because we know that each outcome is equally likely, the probability of every term in the sum equals 1/36.

```
# Compute the probability measured for each of these items
# including the sum equals seven
X={i:len(j)/36. for i,j in dinv.items() }

print('X=')
for key, value in X.items():
    print(f"{key}: {value}")
```

```
X=
2: 0.027777777777777776
3: 0.05555555555555555
4: 0.08333333333333333
5: 0.1111111111111111
6: 0.1388888888888889
7: 0.16666666666666666
8: 0.1388888888888889
9: 0.1111111111111111
10: 0.08333333333333333
11: 0.05555555555555555
12: 0.027777777777777776
```

Q4

(b) What is the probability that half the product of three dice will exceed their sum?

Using the same method above, we create the first mapping as follows:

```
d={(i,j,k):((i*j*k)/2>i+j+k) for i in range(1,7)
                              for j in range(1,7)
                                  for k in range(1,7)}

# print('d=')
# for key, value in d.items():
#     print(f"{key}: {value}")
```

```
d=
(1, 1, 1): False
(1, 1, 2): False
(1, 1, 3): False
(1, 1, 4): False
(1, 1, 5): False
(1, 1, 6): False
(1, 2, 1): False
(1, 2, 2): False
(1, 2, 3): False
(1, 2, 4): False
(1, 2, 5): False
(1, 2, 6): False
(1, 3, 1): False
(1, 3, 2): False
(1, 3, 3): False
(1, 3, 4): False
(1, 3, 5): False
(1, 3, 6): False
(1, 4, 1): False
(1, 4, 2): False
(1, 4, 3): False
(1, 4, 4): False
(1, 4, 5): False
(1, 4, 6): True
(1, 5, 1): False
(1, 5, 2): False
(1, 5, 3): False
(1, 5, 4): False
(1, 5, 5): True
(1, 5, 6): True
```

Q4

The keys of this dictionary are the triples and the values are the logical values of whether or not half the product of three dice exceeds their sum. Now, we do the inverse mapping to collect the corresponding lists.

```python
dinv = defaultdict(list)
for i,j in d.items(): dinv[j].append(i)

print('dinv=')
for key, value in dinv.items():
    print(f"{key}: {value}")
```

```
dinv=
False: [(1, 1, 1), (1, 1, 2), (1, 1, 3), (1, 1, 4), (1, 1, 5), (1, 1, 6), (1, 2, 1), (1, 2, 2), (1, 2, 3), (1, 2, 4), (1, 2, 5), (1, 2, 6), (1, 3, 1), (1, 3, 2), (1, 3, 3), (1, 3, 4), (1, 3, 5), (1, 3, 6), (1, 4, 1), (
1, 4, 2), (1, 4, 3), (1, 4, 4), (1, 4, 5), (1, 5, 1), (1, 5, 2), (1, 5, 3), (1, 5, 4), (1, 6, 1), (1, 6, 2), (1, 6, 3), (2, 1, 1), (2, 1, 2), (2, 1, 3), (2, 1, 4), (2, 1, 5), (2, 1, 6), (2, 2, 1), (2, 2, 2), (2, 2, 3),
 (2, 2, 4), (2, 3, 1), (2, 3, 2), (2, 4, 1), (2, 4, 2), (2, 5, 1), (2, 6, 1), (3, 1, 1), (3, 1, 2), (3, 1, 3), (3, 1, 4), (3, 1, 5), (3, 1, 6), (3, 2, 1), (3, 2, 2), (3, 3, 1), (3, 4, 1), (3, 5, 1), (3, 6, 1), (4, 1, 1
), (4, 1, 2), (4, 1, 3), (4, 1, 4), (4, 1, 5), (4, 2, 1), (4, 2, 2), (4, 3, 1), (4, 4, 1), (4, 5, 1), (5, 1, 1), (5, 1, 2), (5, 1, 3), (5, 1, 4), (5, 2, 1), (5, 3, 1), (5, 4, 1), (6, 1, 1), (6, 1, 2), (6, 1, 3), (6, 2,
 1), (6, 3, 1)]
True: [(1, 4, 6), (1, 5, 5), (1, 5, 6), (1, 6, 4), (1, 6, 5), (1, 6, 6), (2, 2, 5), (2, 2, 6), (2, 3, 3), (2, 3, 4), (2, 3, 5), (2, 3, 6), (2, 4, 3), (2, 4, 4), (2, 4, 5), (2, 4, 6), (2, 5, 2), (2, 5, 3), (2, 5, 4), (2
, 5, 5), (2, 5, 6), (2, 6, 2), (2, 6, 3), (2, 6, 4), (2, 6, 5), (2, 6, 6), (3, 2, 3), (3, 2, 4), (3, 2, 5), (3, 2, 6), (3, 3, 2), (3, 3, 3), (3, 3, 4), (3, 3, 5), (3, 3, 6), (3, 4, 2), (3, 4, 3), (3, 4, 4), (3, 4, 5),
(3, 4, 6), (3, 5, 2), (3, 5, 3), (3, 5, 4), (3, 5, 5), (3, 5, 6), (3, 6, 2), (3, 6, 3), (3, 6, 4), (3, 6, 5), (3, 6, 6), (4, 1, 6), (4, 2, 3), (4, 2, 4), (4, 2, 5), (4, 2, 6), (4, 3, 2), (4, 3, 3), (4, 3, 4), (4, 3, 5)
, (4, 3, 6), (4, 4, 2), (4, 4, 3), (4, 4, 4), (4, 4, 5), (4, 4, 6), (4, 5, 2), (4, 5, 3), (4, 5, 4), (4, 5, 5), (4, 5, 6), (4, 6, 1), (4, 6, 2), (4, 6, 3), (4, 6, 4), (4, 6, 5), (4, 6, 6), (5, 1, 5), (5, 1, 6), (5, 2,
2), (5, 2, 3), (5, 2, 4), (5, 2, 5), (5, 2, 6), (5, 3, 2), (5, 3, 3), (5, 3, 4), (5, 3, 5), (5, 3, 6), (5, 4, 2), (5, 4, 3), (5, 4, 4), (5, 4, 5), (5, 4, 6), (5, 5, 1), (5, 5, 2), (5, 5, 3), (5, 5, 4), (5, 5, 5), (5, 5
, 6), (5, 6, 1), (5, 6, 2), (5, 6, 3), (5, 6, 4), (5, 6, 5), (5, 6, 6), (6, 1, 4), (6, 1, 5), (6, 1, 6), (6, 2, 2), (6, 2, 3), (6, 2, 4), (6, 2, 5), (6, 2, 6), (6, 3, 2), (6, 3, 3), (6, 3, 4), (6, 3, 5), (6, 3, 6), (6,
 4, 1), (6, 4, 2), (6, 4, 3), (6, 4, 4), (6, 4, 5), (6, 4, 6), (6, 5, 1), (6, 5, 2), (6, 5, 3), (6, 5, 4), (6, 5, 5), (6, 5, 6), (6, 6, 1), (6, 6, 2), (6, 6, 3), (6, 6, 4), (6, 6, 5), (6, 6, 6)]
```

Q4

Note that dinv contains only two keys, True and False. Again, because the dice are independent, the probability of any triple is $1/(6^3)$. Finally, we collect this for each outcome as in the following,

```
X={i:len(j)/6.0**3 for i,j in dinv.items() }

print('X=')
for key, value in X.items():
    print(f"{key}: {value}")
```

```
X=
False: 0.37037037037037035
True: 0.6296296296296297
```

Q5

Assuming a normal (Gaussian) distribution with mean 30 Ω and standard deviation of 1.8 Ω, determine the probability that a resistor coming off the production line will be within the range of 28 Ω to 33 Ω. (Hint: use stats.norm.cdf function from scipy import stats)

Q5

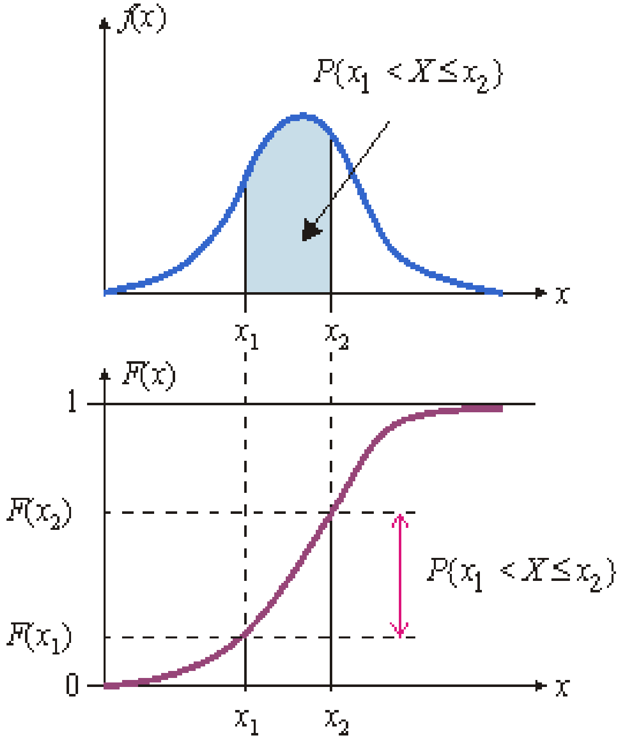
Answer:

Wiki: The cumulative distribution function of a real-valued random variable X is the function given by

$$F_X(x) = Pr(X \leq x)$$

where the right-hand side represents the probability that the random variable X takes on a value less than or equal to x. The probability that X lies in the semi-closed interval $(a, b]$, where $a < b$, is therefore

$$Pr(a < X \leq b) = F_X(b) - F_X(a).$$



Q5

```python
from scipy import stats

# define constants
mu = 30 # mean = 30Ω
sigma = 1.8 # standard deviation = 1.8Ω
x1 = 28 # lower bound = 28Ω
x2 = 33 # upper bound = 33Ω
```

- Import scipy.stats that contains statistical functions, including probability distributions.
- Define mean 30Ω and standard deviation of 1.8Ω for a normal (Gaussian) distribution.
- Set the range between 28Ω and 33Ω

Q5

stats.norm.cdf(x1, mu, sigma): This calculates the cumulative distribution function (CDF) of the normal distribution at x1 (28Ω). The CDF gives the probability that a normally distributed random variable is less than or equal to x1.

```
## calculate probabilities
# probability from Z=0 to lower bound
p_lower = stats.norm.cdf(x1,mu,sigma)

# probability from Z=0 to upper bound
p_upper = stats.norm.cdf(x2,mu,sigma)

# probability of the interval
Prob = (p_upper) - (p_lower)
```

Prob: The probability that the random variable falls within the interval [28Ω, 33Ω]. It is calculated by subtracting the probability of being less than 28Ω (p_lower) from the probability of being less than 33Ω (p_upper).

Q5

```
• # print the results print('\n')
• print(f'Normal distribution: mean = {mu}, std dev = {sigma} \n')
• print(f'Probability of occurring between {x1} and {x2}: ')
• print(f'--> inside interval Pin = {round(Prob*100,1)}%')
• print(f'--> outside interval Pout = {round((1-Prob)*100,1)}% \n')
• print('\n')
```

```
Normal distribution: mean = 30, std dev = 1.8

Probability of occurring between 28 and 33:
--> inside interval Pin = 81.9%
--> outside interval Pout = 18.1%
```

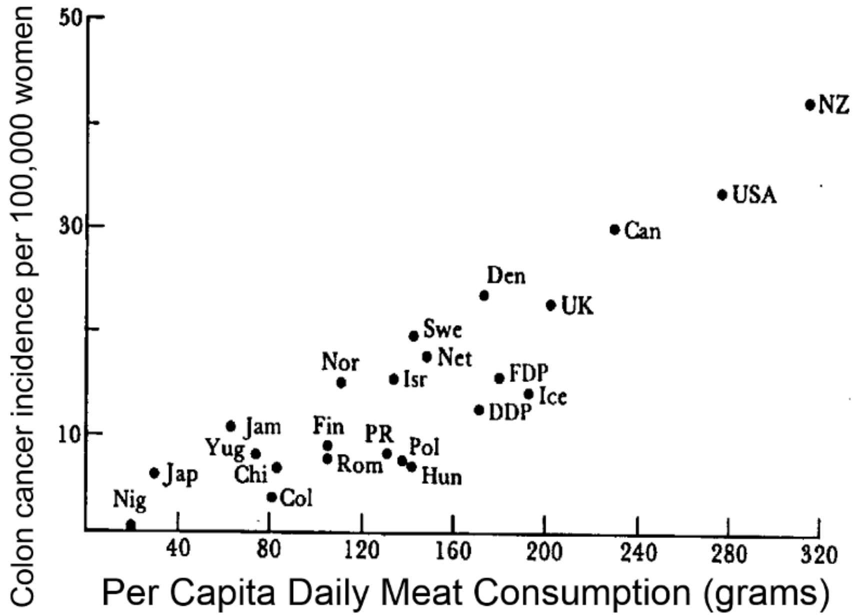
Q6

For each of the following graphs,

(i) State what you think the evidence is trying to suggest.

(ii) Give a reason why you agree or disagree with what the evidence is suggesting.

(iii) Identify whether the variable of the y-axis and the variable of the x-axis are correlated and/or causal?



http://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module1BDescriptiveStudies_and_Statistics/PH717-Module1B-DescriptiveStudies_and_Statistics6.html

Q6

Discussion

(i) Colon cancer is correlated to the amount of daily meat consumption.

(ii) There is a clear linear trend; countries with the lowest meat consumption have the lowest rates of colon cancer, and the colon cancer rate among these countries progressively increases as meat consumption increases.

(iii) Probably causal.

Q7

If A and B are correlated, but they're actually caused by C, which of the following statements are correct?

a) A and C are correlated

b) B and C are correlated

c) A causes B to happen

d) A causes C to happen

Q7

Ans:

a) A and C are correlated (Yes, A and C are correlated because A is caused by C)
b) B and C are correlated (Yes, B and C are correlated because B is caused by C)
c) A causes B to happen (No, A and B share the confounding factor C, but A and B don’t have causal relationship)
d) A causes C to happen (No, C causes A to happen, however, we are not sure if A causes C to happen).

Q8

We toss a coin and observe which side is facing up. Which of the following statements represent valid probability assignments for observing head P[‘H’] and tail P[‘T’]?

a) P[‘H’]=0.2, P[‘T’]=0.9

b) P[‘H’]=0.0, P[‘T’]=1.0

c) P[‘H’]=-0.1, P[‘T’]=1.1

d) P[‘H’]=P[‘T’]=0.5

Q8

Ans:

a) P[‘H’]=0.2, P[‘T’]=0.9 (Invalid because they sum to 1.1, which doesn’t conform to the Axioms of probability)

b) P[‘H’]=0.0, P[‘T’]=1.0 (Valid because it conforms to the Axioms of probability)

c) P[‘H’]=-0.1, P[‘T’]=1.1 (Invalid because probability cannot have negative value)

d) P[‘H’]=P[‘T’]=0.5 (Valid because it conforms to the Axioms of probability)

Q9

Two vectors $a = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$ are linearly dependent?

Q9

- Ans: False (because a is not a multiply of b.)

Q10

The rank of the matrix $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ is ______

Answer:

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 3 \\ 0 & -2 \end{bmatrix}$$

The rank is 2

Q10

The rank of the matrix $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$ is ______

Method 2

The determinant of the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is calculated as:

$$\det(A) = ad - bc = 1 \times 4 - 2 \times 3 = -2 \neq 0$$

If the determinant is non-zero, the matrix is full rank, which means the rank of the matrix is 2.

Q11

The rank of the matrix $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$ is ________

Q11

Perform Gaussian Elimination to obtain Row Echelon Form

step 1: make the first entry in first column as 1

The first now already has a leading 1 in the first column, so

- To eliminate 4 below the 1, we substact 4 times ~~Row~~ the firs row from the second row

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix}$$

Row2 = Row2 − 4×Row1

- Eliminate the 7 below the 1

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix}$$

Row3 = Row3 − 7×Row1

step 2: make the second entry in the second column a Leading 1

Row2 = $-\frac{1}{3}$ × Row2

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -6 & -12 \end{bmatrix}$$

step 3. Elimate the entry below the leading 1 in the second column.

Row3 = Row3 + 6 × Row2

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -6 & -12 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Rank is 2

THANK YOU