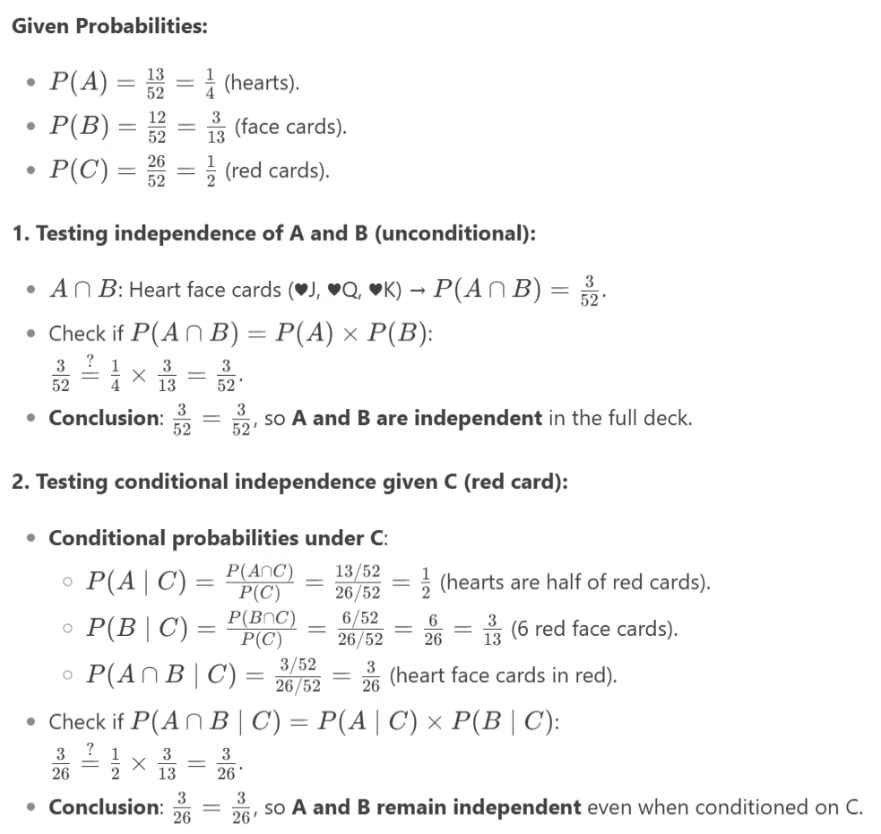
**Question 1**

1. (0.99×0.01)+(0.02×0.99)=0.0297.
2. 0.99×0.01 / 0.0297 = 0.33 (33.3%)
3. 0.98×0.99 / (1 - 0.0297) = 0.9997 (99.97%)

**Question 2**

**Question 3**

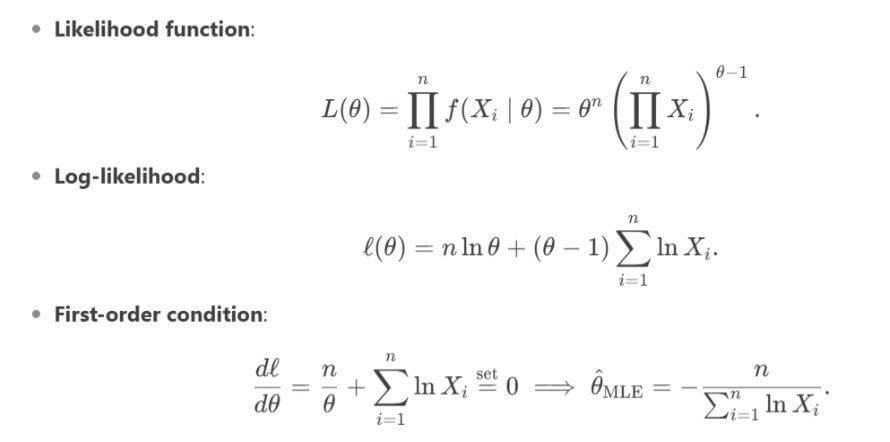


**Question 4**

1. Let be the fraction of left-handed individuals, so the fraction of right-handed individuals is . The overall proportion of athletes.is given by:

1. Given implies . Hence the ratio of left-handed to right-handed individuals is
2. If there are twice as many right-handed as left-handed individuals, let the number of left-handed individuals be and right-handed be then . Substitute into the expression for 0.8167.

**Question 5**

****

**Question 6**

(a) Are the events and independent? Why?

Let denote the event that all three of them have the same birthday. Then P() = P() = . At the same time P() \* P() = (. So since P() = P() \* P(), and are independent events

(b) Are the events and independent? Why?

Just like in part (a), P() = P() = and P() \* P() = (. So as P() = P() \* P(), events and are independent.

(c) Are the events and independent? Why?

P() = P() = .However, P()\* P()\* P() = (.So as P() is not equal to P()\* P()\* P(), events and are not independent.

**Question 7**

1. For each x=1,2,3…6, total outcomes=36. Then,

P(*X*=x) = [#outcomes where min(x1, x2) = x] / 36.

There are two cases when min(x1, x2) = x:

(1) both dice equal to x: 1 outcome

(2) one die is x, the other is greater than x: (x,k) and (k,x) for k = x+1, …, 6: in total 2(6-x) outcomes.

Thus, P(*X*=x) = [1 + 2(6-x)] / 36 = (13-2x) / 36.

1. E[*X*] = 91/36 = 2.527. Var[*X*] = 2555/1296 = 1.971.

**Question 8**

The table below shows some data from the early days of a clothing company. Each row shows the sales for a year, and the amount spent on advertising in that year.

|  |  |  |
| --- | --- | --- |
| Year | Advertising (Million Dollars) | Sales (Million Dollars) |
| 1 | 22 | 650 |
| 2 | 31 | 855 |
| 3 | 33 | 1064 |
| 4 | 44 | 1191 |
| 5 | 51 | 1420 |

1. Use linear regression to calculate the relationship between the advertising and sales (slop, and intercept).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | y | x- | y- | (x-)\*( y-) |
| 22 | 650 | -14.2 | -386 | 5481.2 |
| 31 | 855 | -5.2 | -181 | 941.2 |
| 33 | 1064 | -3.2 | 28 | -89.6 |
| 44 | 1191 | 7.8 | 155 | 1209 |
| 51 | 1420 | 14.8 | 384 | 5683.2 |

,

= 25.49 and .

Then the linear regression equation is y = 25.49x + 113.2, which describes the relationship between advertising (x) and sales (y).

1. Calculate the correlation coefficient between the amounts of advertising and sales.

and .

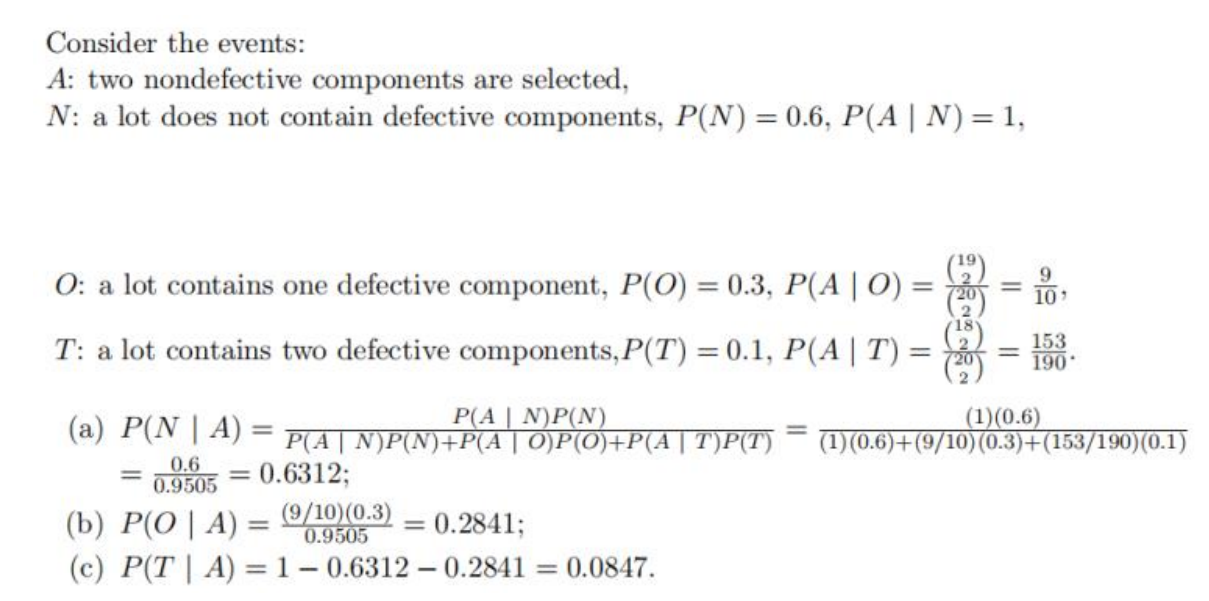
The correlation coefficient = 0.976.

1. If there is no advertising, what is the expected sales? If the amount of advertising is 58 million dollars in the next year, please predict the sales.

If there is no advertising, x=0, the expected sales y= 25.49\*0 + 113.2 = 113.2,

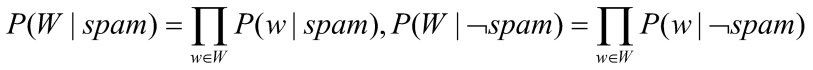
if the amount of advertising x=58, then the sales will be y =25.49\*58 + 113.2 = 1591.74.

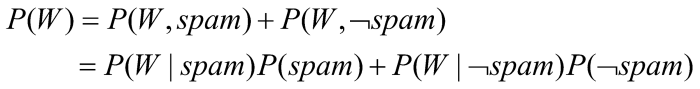
**Question 9**

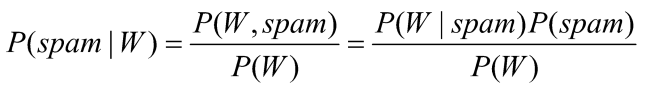
****

**Question 10**









Email I:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hypothesis** | **Prior odds** | **Likelihood ratio** | **Posterior odds** | **P(spam | email)** |
| spam | 0.8 | 0.000163 | 0.000130 | 0.101751 |
| ¬spam | 0.2 | 0.005753 | 0.001151 |

Email II:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hypothesis** | **Prior odds** | **Likelihood ratio** | **Posterior odds** | **P(spam | email)** |
| spam | 0.8 | 0.296208 | 0.236966 | 0.980170 |
| ¬spam | 0.2 | 0.023970 | 0.004794 |

# Input the email with your keyboard #

email = input('Please enter the test Email: ') # gets the test email

email = email.lower() # convert string to lowercase

print(email)

# Input the spam detection keyword #

word1 = input('Please enter the first key word: ') # gets the first keyword

word1 = word1.lower() # convert the first key word to lowercase

word2 = input('Please enter the second key word: ') # gets the second keyword

word2 = word2.lower() # convert the second keyword to lowercase

word3 = input('Please enter the third key word: ') # gets the third keyword

word3 = word3.lower() # convert the third keyword to lowercase

word4 = input('Please enter the fourth key word: ') # gets the fourth keyword

word4 = word4.lower() # convert the fourth keyword to lowercase

word5 = input('Please enter the fifth key word: ') # gets the fifth keyword

word5 = word5.lower() # convert the fifth keyword to lowercase

# Detect whether the key word occured in your email #

n\_word1 = email.count(word1) # count the number of the first keyword in the email

n\_word2 = email.count(word2) # count the number of the second keyword in the email

n\_word3 = email.count(word3) # count the number of the third keyword in the email

n\_word4 = email.count(word4) # count the number of the fourth keyword in the email

n\_word5 = email.count(word5) # count the number of the fifth keyword in the email

# The prior odds of spam is 80% #

p\_spam = 1.0 # initialize the likelihood ratio of the spam email

p\_no\_spam = 1.0 # initialize the likelihood ratio of the no spam email

prior\_odds\_spam = 0.8 # prior odds of spam

# Calculate the Likelihood ratio #

if n\_word1 != 0: # the first keyword occurred in the email

p\_spam = p\_spam \* 0.88

p\_no\_spam = p\_no\_spam \* 0.47

if n\_word2 != 0: # the second keyword occurred in the email

p\_spam = p\_spam \* 0.66

p\_no\_spam = p\_no\_spam \* 0.10

if n\_word3 != 0: # the third keyword occurred in the email

p\_spam = p\_spam \* 0.11

p\_no\_spam = p\_no\_spam \* 0.60

if n\_word4 != 0: # the fourth keyword occurred in the email

p\_spam = p\_spam \* 0.51

p\_no\_spam = p\_no\_spam \* 0.51

if n\_word5 != 0: # the fifth keyword occurred in the email

p\_spam = p\_spam \* 0.005

p\_no\_spam = p\_no\_spam \* 0.40

# Calculate the Posterior odds #

posterior\_odds\_spam = p\_spam \* prior\_odds\_spam # posterior odds of spam

posterior\_odds\_no\_spam = p\_no\_spam \* (1-prior\_odds\_spam) # posterior odds of no spam

p\_isSpam = posterior\_odds\_spam / (posterior\_odds\_spam+posterior\_odds\_no\_spam) # probability of spam P(spam | email)

print("p\_spam: %.6f, p\_no\_spam: %.6f, posterior\_odds\_spam: %.6f, posterior\_odds\_no\_spam: %.6f, p\_isSpam: %.6f"%(p\_spam, p\_no\_spam, posterior\_odds\_spam, posterior\_odds\_no\_spam, p\_isSpam))