ASSIGNMENT 2 :-

PROBLEM 3.6:

Consider the partial datasets extracted from body fat prediction dataset on kaggle.  
Only body fat column is extracted. The dataset is partitioned into two sets, Xla and Xny . Suppose that first 100 instances belong to LA and the remaining 152 instances belong to NYC. Let the subscripts ‘la’ and ‘ny’ indicate two cities. Two datasets are exclusive:  
n = nla +nny and X = Xla ∪Xny .

Following parameters for each city are already computed:

A maths table with a cartoon character and numbers

AI-generated content may be incorrect.

Find the parameter values in real time without accessing the original dataset.

a). Find the mean value, μ(X).

-> μ(X) = naμ(Xa) + nbμ(Xb)  
 na + nb

μ(X) = 100 × 18.191 + 152 × 19.7822  
252

= 19.1508

b) Find the variance value, σ2(X).  
σ²(X) = na(σ²(Xa) + μ(Xa)²) + nb(σ²(Xb) + μ(Xb)²)   
 na + nb − μ(X)²

σ^2(X) =

100 × (80.763 + 18.191²) + 152 × (61.513 + 19.7822²)  
 252 -– 19.151²

= 69.7579

c). Find E(X²).   
E(X²) = σ²(X) + μ²(X)

= 69.7579 + 19.1508²

= 436.51

≈ 436.5108

d). Find the Pearson’s moment coefficient of skewness, η₃(X).  
ηₚ(X) = Mₚ(X)  
 σp(X)

M₃(Xₗₐ) = η₃( Xₗₐ) (√σ²( Xₗₐ)³

M₃( Xₗₐ) = 0.2055 × (√80.763)³

= 149.1525 ≈ 149.1882

M₃(ₙᵧ) = 0.1641 × (√61.513)³

= 79.1697 ≈ 79.1709

M3(X) = nₗₐ(M₃(Xₗₐ) − B1(Xₗₐ)) + nₙᵧ (M3(Xₙᵧ) − B1(Xₙᵧ ))  
 nₗₐ + nₙᵧ + B₁(X)

B1(X) = −μ3(X) − 3μ(X)σ²(X)

B1(X) = −19.15083 − 3 × 19.1508 × 69.7579 ≈ −11031

B1(Xla) = −18.1913 − 3 × 18.191 × 80.763 ≈ −10427

B1(Xny) = −19.78223 − 3 × 19.7822 × 61.513 ≈ −11392

M3(X) =

(100 × (149.1525 + 10427) + 152 × (79.1697 + 11392))

(100 + 152 − 11031)

= 85.0041 ≈ 84.7608

η3(X) = 85.0041

√69.7579³

= 0.1459 ≈ 0.1455

e) Find E(X³).

E(X³) = M₃(X) + μ₃(X) + 3μ(X)σ²(X)

E(X³) = 85.0041 + 19.15083 + 3 × 19.1508 × 69.7579

E(X³) = 11116

f) Find the fourth standardized moment coefficient of kurtosis, η₄(X).

ηₚ(X) = Mₚ(X)

σₚ(X)

M₄(Xla) = η₄(Xla) × σ⁴(Xla)

M₄(Xla) = 2.0279 × 80.763² = 13227 ≈ 13228

M₄(Xny) = 3.2783 × 61.513² = 12405 ≈ 12404

M₄(X) =

(nla × (M₄(Xla) − 4B₁(Xla)) + nny × (M₄(Xny) − 4B₁(Xny)))

(nla + nny) + 4B₁(X) (Equation 3.68)

pBk(X) = ∑ (p choose k) E(X^(p−i)) (−μ(X))ⁱ where 0 ≤ k ≤ p

4B₁(X) = −4E(X³)μ(X) + 6E(X²)μ²(X) − 3μ⁴(X)

4B₁(X) = −4 × 11116 × 19.1508 + 6 × 436.51 × 19.1508² − 3 × 19.1508⁴ = −294500

E(X³ₗₐ) = 79.1697 + 18.1913 + 3 × 18.191 × 80.763 = 10576

E(X²ₗₐ) = σ²(Xla) + μ²(Xla) = 80.763 + 18.191²

= 411.6755

4B₁(Xla) = −4 × 10576 × 18.191 + 6 × 411.6755 × 18.191² − 3 × 18.191⁴

= −280690

E(X³ₙᵧ) = 79.1709 + 19.78223 + 3 × 19.7822 × 61.513 = 11471

E(X²ₙᵧ) = σ²(Xₙᵧ) + μ²(Xₙᵧ) = 61.513 + 19.7822² = 452.8484 ≈ 452.8497

4B₁(Xₙᵧ) = −4 × 11471 × 19.7822 + 6 × 452.8484 × 19.7822² − 3 × 19.7822⁴ = −303820

M₄(X) = (100 × (13227 + 280690) + 152 × (12405 + 303820))

(100 + 152 − 294500)

= 12873 ≈ 12891

η₄(X) = 12873 / 69.7579²

= 2.6454 ≈ 2.6491

g) Find E(X⁴).

E(X⁴) = M₄(X) − 4B₁(X)

E(X⁴) = 12873 + 303820

E(X⁴) = 316693

h) Validate your answers in a) ∼ g) using the full dataset and discuss the differences.

| Parameter | Real-Time Value | By Definition |

|------------|----------------|--------------|

| μ(X) | 19.1508 | 19.1508 |

| σ²(X) | 69.7579 | 69.7579 |

| E(X²) | 436.51 | 436.5108 |

| M₃(X) | 85.0041 | 84.7608 |

| η₃(X) | 0.1459 | 0.1455 |

| E(X³) | 11116 | 11116 |

| M₄(X) | 12873 | 12891 |

| η₄(X) | 2.6454 | 2.6491 |

| E(X⁴) | 316693 | ? |

For questions j) ∼ m), suppose that a person (xₗₐ,₁₀₀ = 22.2) in LA moved to NYC.

j) Find the following mean values:

μ(Xₗₐ − {22.2}) and μ(Xₙᵧ ∪ {22.2}).

Let Xₙᵧ₊ = Xₙᵧ ∪ {22.2} and Xₙᵧ₋ = Xₙᵧ − {22.2} for simplicity.

μ(X₁∼ₙ) = ((n − 1)μ(X₁∼ₙ₋₁) + xₙ)

n

μ(Xₙᵧ₊) = (152 × 19.7822 + 22.2)

153

μ(Xₙᵧ₊) = 19.798

μ(X₁∼ₙ₋₁) = (nμ(X₁∼ₙ) − xₙ) / (n − 1)

μ(Xₗₐ₋) = (100 × 19.1508 − 22.2) / 99

μ(Xₗₐ₋) = 18.150

k) Find the following variance values:

σ²(Xₗₐ − {22.2}) and σ²(Xₙᵧ ∪ {22.2}).

σ²(X₁∼ₙ) = ((n − 1)(σ²(X₁∼ₙ₋₁) + μ²(X₁∼ₙ₋₁)) + xₙ²)

n − μ²(X₁∼ₙ)

σ²(Xₙᵧ₊) = (152 × (61.513 + 19.7822²) + 22.2²)

153 − 19.798²

σ²(Xₙᵧ₊) = 61.149 ≈ 61.1487

σ²(X₁∼ₙ₋₁) = (n(σ²(X₁∼ₙ) + μ²(X₁∼ₙ)) − xₙ²) / (n − 1) − μ²(X₁∼ₙ₋₁)

σ²(Xₗₐ₋) = (100 × (80.763 + 18.191²) − 22.2²) / 99 − 18.1505²

σ²(Xₗₐ₋) = 81.41

l) Find the following Pearson’s moment coefficients of skewness:

η₃(Xₗₐ − {22.2}) and η₃(Xₙᵧ ∪ {22.2}).

To find η₃(Xₙᵧ ∪ {22.2}):

M₃(X₁∼ₙ) =

((n − 1)(M₃(X₁∼ₙ₋₁) − B₁(X₁∼ₙ₋₁)) + xₙ³)

n + B₁(X₁∼ₙ)

M₃(Xₙᵧ₊) = (152(M₃(Xₙᵧ) − B₁(Xₙᵧ)) + 22.2³)

153 + B₁(Xₙᵧ₊)

B₁(Xₙᵧ₊) = −μ₃(Xₙᵧ₊) − 3μ(Xₙᵧ₊)σ²(Xₙᵧ₊)

= −19.798³ − 3 × 19.798 × 61.149

= −11392

M₃(Xₙᵧ₊) = (152 × (79.1697 + 11392) + 22.2³)

153 − 11392

= 75.705

η₃(Xₙᵧ₊) = M₃(Xₙᵧ₊) / σ³(Xₙᵧ₊)

= 75.705 / √(61.149³)

= 0.1583 ≈ 0.1586

To find η₃(Xₗₐ − {22.2}):

M₃(X₁∼ₙ₋₁) =

(n(M₃(X₁∼ₙ) − B₁(X₁∼ₙ)) − xₙ³)

(n − 1) + B₁(X₁∼ₙ₋₁)

M₃(Xₗₐ₋) = (100(M₃(Xₗₐ) − B₁(Xₗₐ)) − 22.2³) / 99 + B₁(Xₗₐ₋)

B₁(Xₗₐ₋) = −μ₃(Xₗₐ₋) − 3μ(Xₗₐ₋)σ²(Xₗₐ₋)

= −18.1505³ − 3 × 18.1505 × 81.415

= −10413

M₃(Xₗₐ₋) = (100 × (149.1525 + 10427) − 22.2³) / 99 − 10413

= 159.47

η₃(Xₗₐ₋) = M₃(Xₗₐ₋)

σ³(Xₗₐ₋)

= 159.47 / √(81.415³)

= 0.2171 ≈ 0.2177

m) Find the following fourth standardized moment coefficients of kurtosis:

η₄(Xₗₐ − {22.2}) and η₄(Xₙᵧ ∪ {22.2}).

To find η₄(Xₙᵧ₊),

Mₚ(X₁∼ₙ) = ((n − 1)(Mₚ(X₁∼ₙ₋₁) − pB₁(X₁∼ₙ₋₁)) + xₙᵖ) / n + pB₁(X₁∼ₙ) (Equation 3.67)

M₄(Xₙᵧ₊) = (152 × (M₄(Xₙᵧ) − 4B₁(Xₙᵧ)) + 22.2⁴) / 153 + 4B₁(Xₙᵧ₊)

4B₁(Xₙᵧ₊) = −4E(X³ₙᵧ₊)μ(X) + 6E(X²ₙᵧ₊)μ²(Xₙᵧ₊) − 3μ⁴(Xₙᵧ₊)

E(X³ₙᵧ₊) = M₃(Xₙᵧ₊) + μ³(Xₙᵧ₊) + 3μ(Xₙᵧ₊)σ²(Xₙᵧ₊)

= 75.705 + 19.798³ + 3 × 19.798 × 61.149

≈ **11467.63**

E(X²ₙᵧ₊) = σ²(Xₙᵧ₊) + μ²(Xₙᵧ₊)

= 61.149 + 19.798²

= 453.11

n) Validate your answers in j) ∼ m) using the full dataset and discuss the differences.

* For NYC dataset (nₙᵧ = 152):

μ(Xₙᵧ) = 19.7822

σ²(Xₙᵧ) = 61.513

η₃(Xₙᵧ) = 0.1641

η₄(Xₙᵧ) = 3.2783

M₃(Xₙᵧ) = 79.1697

M₄(Xₙᵧ) = 12405

4B₁(Xₙᵧ) = −303820

* For updated dataset after adding 22.2 (n = 252)

μ(X ∪ {22.2}) = 19.798

σ²(X ∪ {22.2}) = 61.149

η₃(X ∪ {22.2}) = 0.1583

η₄(X ∪ {22.2}) = 2.0064

M₃(X ∪ {22.2}) = 75.705

M₄(X ∪ {22.2}) = 7502.16

4B₁(X ∪ {22.2}) = -303436.57

* For LA dataset (nₗₐ = 100)

μ(Xₗₐ) = 18.191

σ²(Xₗₐ) = 80.763

η₃(Xₗₐ) = 0.2055

η₄(Xₗₐ) = 2.0279

M₃(Xₗₐ) = 149.1525

M₄(Xₗₐ) = 13227

4B₁(Xₗₐ) = −280690

* For the full dataset (n = 252)

μ(X) = 19.1508

σ²(X) = 69.7579

η₃(X) = 0.1459

η₄(X) = -6.0038

M₃(X) = 85.0041

M₄(X) = -22449.37

4B₁(X) = -303436.57

PROBLEM 3.7:

In order to investigate the inheritance of traits in pea plants, Sir Francis Galton conducted a scientific experiment. Galton’s pea data table, which is originally from [6, p.226] contains a list of frequencies of daughter seeds of various sizes according to the size of their parent seeds. Parent pea seeds of various diameter, 15 ∼ 21. The unit of the diameter of seed is 0.01 inch. Each parent seed has 100 filial seeds and their diameter frequency distribution is given in the following table.

A table with numbers and a few black text

AI-generated content may be incorrect.

Let X15 be the set of daughter seeds who parent seed’s diameter is 15. Let Xx,15∼18 is a subset of daughter seeds whose diameters range from 15 to 18 and their parent seed diameter is x. Let Xx,<15 is a subset of daughter seeds whose diameter is less than 15 and they are worthless as a commodity because they are too small and their exact diameter value is unknown. Let F15,17 = 11 denote the frequency of case where parent and daughter seed diameters are 15 and 17. Let F15,15∼18 = ⟨14, 9, 11, 14⟩ be the subset frequency.

A) Find the median values of each Xₓ for x ∈ {15, · · · , 21}.

wμi(X15, P15) = 15

because ∑¹⁵ᵢ₌₋∞ P (X15, i) =

46 + 14  
100 = 0.6 ≥ 0.5

And ∑²¹ᵢ₌₁₅ =

14 + 9 + 11 + 14 + 4 + 2 + 0  
 100 = 0.54 ≥ 0.5

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Parent seeds diameter x | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| wμi(Xx, Px) | 15 | 16 | 16 | 16 | 15 | 17 | 17 |

b). Find the median values of each Xx,15∼21 for x ∈ {15, · · · , 21}.  
For example, μi(X15,15∼21) = 17

wμi(X15,15∼21, P15,15∼21) = 17

because ∑¹⁸ᵢ₌₁₅ P (X15,i) =

14 + 9 + 11  
 54 = 0.62963 ≥ 0.5

AND ∑²¹ᵢ₌₁₈ P (X15,i) =

14 + 9 + 11  
 54 = 0.54 ≥ 0.5

wμi(X16,15∼21, P16,15∼21) = 16.5

because ∑¹⁶ᵢ₌₁₅ P (X16,i) =

15 + 18  
 66 = 0.5

And ∑²¹ᵢ₌₁₇ P (X16,i) =

16 + 13 + 3 + 1 + 0  
 66 = 0.5 = 0.5

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Parent seeds diameter x | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| wμi(Xx,15∼21, Px,15∼21) | 17 | 16.5 | 17 | 17 | 17 | 17 | 18 |

c) Find the mean values of each Xx,15∼21 for x ∈ {15, · · · , 21}.  
  
μw(X15,15∼21, P15,15∼21) =

14 × 15 + 9 × 16 + · · · + 0 × 21  
 54

= 16.833

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parent seeds diameter x | 16 | 17 | 18 | 19 | 20 | 21 |
| μw(Xx,15∼21, Px,15∼21) | 16.606 | 16.667 | 16.955 | 16.954 | 17.403 | 17.603 |

d). Find the variance values of each Xx,15∼21 for x ∈ {15, · · · , 21}.  
σ²w(X15,15∼21, P15,15∼21) =  
14 × (15 − 16.833)² + 9 × (16 − 16.833)² + · · · + 0 × (21 − 16.833)²  
 54

= 2.0648

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parent seeds diameter x | 16 | 17 | 18 | 19 | 20 | 21 |
| μw(Xx,15∼21, Px,15∼21) | 1.5418 | 1.7143 | 1.801 | 2.4748 | 2.2145 | 2.2138 |

e) Find the MAD around median values of each

Xx,15∼21 for x ∈ {15, · · · , 21}.

Wδᵢ(X15,15∼21, P15,15∼21) =

14 × |15 − 17| + 9 × |16 − 17| + · · · + 0 × |21 − 17|  
 54

= 1.2037

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parent seeds diameter x | 16 | 17 | 18 | 19 | 20 | 21 |
| μw(Xx,15∼21, Px,15∼21) | 1.0606 | 1.0952 | 1.0758 | 1.3077 | 1.2338 | 1.1923 |

f) Find the Groeneveld & Meeden’s coefficient values of each Xx,15∼21 for x ∈ {15, · · · , 21}.

For example, κₘᵢδᵢ (X15,15∼21) = −12.408.

κₘᵢδᵢ (X15,15∼21, P15,15∼21) =

μw(X15,15∼21, P15,15∼21) − wμi(X15,15∼21, P15,15∼21)  
 wδi(X15,15∼21, P15,15∼21)

= 16.833 − 17  
 1.2037

= −0.13874

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parent seeds diameter x | 16 | 17 | 18 | 19 | 20 | 21 |
| μw(Xx,15∼21, Px,15∼21) | 0.2 | −0.3044 | −0.0423 | -0.0353 | 0.3263 | -0.3333 |

g) Suppose μ(X₂₀) = 16.33 and μ(X₂₁) = 16.5. Find the mean value, μ(X₂₀ ∪ X₂₁)?  
  
μ(X₂₀ ∪ X₂₁) = 100 × μ(X₂₀) + 100 × μ(X₂₁)  
 100 + 100

= 1633 + 1650  
 200

= 16.415

h) Suppose μ(X₂₀) = 16.33, μ(X₂₁) = 16.5, σ²(X₂₀) = 5.8811, and σ²(X₂₁) = 6.41. Find the variance value, σ²(X₂₀∪ X₂₁)?

σ²(X₂₀∪ X₂₁)  
= 100 × (σ²(X₂₀) + μ(X₂₀)²) + 100 × (σ²(X₂₁) + μ(X₂₁)²)  
 100 + 100 − μ(X₂₁ ∪ X₂₁)²

= 100 × (5.8811 + 16.332) + 100 × (6.41 + 16.52)  
 100 + 100 − 16.4152

= 6.1528

I) Suppose μ(X₁₅) = 14.9 and μ(X15,15∼21) = 16.8333. Find the mean value, μ(X15,<15)?

μ(X15) = μ(X15,<15) ∪ μ(X15,15∼21),

μ(X15) = 46 × μ(X15,<15) + 54 × μ(X15,15∼21)  
 100

14.9 = 46 × μ(X15,<15) + 54 × 16.8333  
 100

μ(X15,<15) = 14.9 × 100 − 54 × 16.8333  
 46

= 12.63

j) Suppose μ(X15) = 14.9, μ(X15,15∼21) = 16.8333, σ²(X15) = 6.23, and σ2(X15,15∼21) = 2.0648. Find the variance value, σ²(X15,<15)?

σ²(X15) + μ(X15)² =

46(σ²(X15,<15) + μ(X15,<15)2) + 54(σ²(X15,15∼21) +μ(X15,15∼21)²)  
 100

6.23 + 14.92 = 46 × (σ2(X15,<15) + 12.632) + 54 × (2.0648 + 16.83332)  
 100

σ²(X₁₅, <₁₅) = 100 × (6.23 + 14.92) − 54 × (2.0648 + 16.8333²)  
 46 − 12.63²

= 1.59