# ASSESSMENT

**Question 1:** What is the average landholding size with standard deviation from the sample data?

**Answer:** Total land holding (acre) from the Farmer details sheet has been used here.

| | |
|---|---|
| **Average** | **1.0605 acres** |
| **Standard Deviation** | **0.71644205 acres** |

**Question 2:** What is the total nitrogen application rate for the recent year for the Boro season from the data?

**Answer: Total nitrogen application rate = Total Nitrogen applied in Kg / Total Application Area**.
The total nitrogen applied is calculated using the N content level in the fertilizer applied.
(Note: Only the Application area is used here which is different from the Landholding area)

| | |
|---|---|
| **Total Nitrogen** | **872.5 kg** |
| **Total Application Area** | **21.88 acres** |
| **Nitrogen Application Rate** | **39.87659963 kg/ acre** |

**Question 3:** Do you see a problem with the data for the fertilizer application events across farmers and over the past year trends? Elaborate.

**Answer:** Data for the fertilizer application events across farmers and over the past year trends has been analyzed here.

| DAT | Year | Fertilizer Application Events | DAT | Fertilizer Application Events | DAT | Fertilizer Application Events |
|---|---|---|---|---|---|---|
| **-1** | **2011** | **80** | **21** | **40** | **42** | **40** |
| | **2016** | **86** | | **40** | | **52** |
| | **2020** | **88** | | **40** | | **52** |
| | **2021** | **88** | | **40** | | **52** |

For Days After Transplanting -1 & 42, the fertilizer application event has been an increasing trend whereas, for DAT 21, it has remained the same.

| Year | Days after Transplanting | | | Total Fertilizer Application Events | Total Quantity of Fertilizer (kg) | Total Grain Yield (kg/Acre) |
|---|---|---|---|---|---|---|
| | **-1** | **21** | **42** | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| **2011** | **80** | **40** | **40** | **160** | **4875** | **92700** |
| **2016** | **86** | **40** | **52** | **178** | **5445** | **92890** |
| **2020** | **88** | **40** | **52** | **180** | **5495** | **83740** |
| **2021** | **88** | **40** | **52** | **180** | **5520** | **81960** |

When the total fertilizer application events are considered for the entire year, we can see that, although **fertilizer application events have increased, the total grain yield is on a decreasing trend**. **That means either soil fertility has been deteriorating or increasing the fertilizer application does not affect the grain yield.**

Soil samples were collected for the set of farms surveyed. We have provided this in the same Excel file. From this file, answer the following questions.

**Question 4:** What are the average available nitrogen content levels in the data provided across the plots? Express this as mean, and standard deviation and write a note on the variation. Make decisions on the inclusions or exclusion of the data points and write a couple of lines on the overall quality of the data.

**Answer:** The soil samples from the farm is used here to calculate the average available nitrogen content levels. **The Sample from University data is not considered here**.

| | |
|---|---|
| **Average** | **63.78 mg/kg** |
| **Variance** | **1471.755611** |
| **Standard Deviation** | **38.36346713 mg/kg** |

**The variance of 1471.755611 is very high for the nitrogen data**.

| **Parameters** | **Nitrogen** |
|---|---|
| **Third Quartile** | **91.94** |
| **Maximum Value** | **140.02** |
| **Minimum Value** | **0.75** |
| **Median** | **65.935** |
| **First Quartile** | **34.72** |
| **Range** | **139.27** |

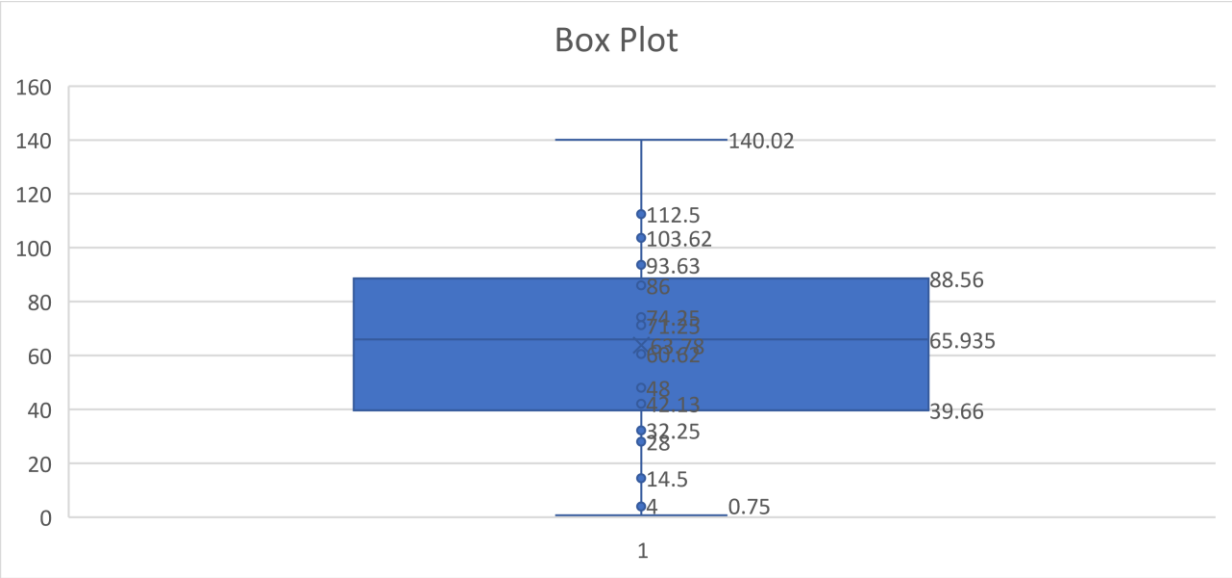
As we can see from the above table, the minimum & maximum values are far apart resulting in a range of 139.27.

**To check for outliers, Inter Quartile Range** is calculated and outlier limits have been calculated

| | | |
|---|---|---|
| **Inter Quartile Range** | **Q3-Q1** | **57.22** |
| **Outlier Limits** | **Q1 - IQR*1.5** | **-51.11** |
| | **Q3 + 1QR*1.5** | **177.77** |

Since the **maximum (140.02) & minimum (0.75) nitrogen values are well within the Outlier limits, we need not exclude any data. Hence all data can be included for analysis**.

**Box Plot** has been constructed to visualize,




**The overall quality of data:**

Though there is a large variation in nitrogen data, there exists no outlier. Hence data can be considered for further analysis (Also, nitrogen data is normally distributed). But, the phosphorus data have many **null values (15/20).** Hence, it might be difficult to make a decision using such data.

**In the Farmer details data**,

**WBB2022-019** has the total land holding as 1 acre whereas the area of the candidate plot is 5, this data is doubtful. This needs to be verified.

**Question 5:** We submitted blind replicate samples to the same lab to assess analytical errors. Calculate the mean and variance across these blind replicates and compare them with the average and variation from actual farm soil samples.

**Answer:** The **Sample from University** data has been considered as the blind replicate samples.

| | |
|---|---|
| **Average** | **98.35 mg/kg** |
| **Variance** | **633.14375** |

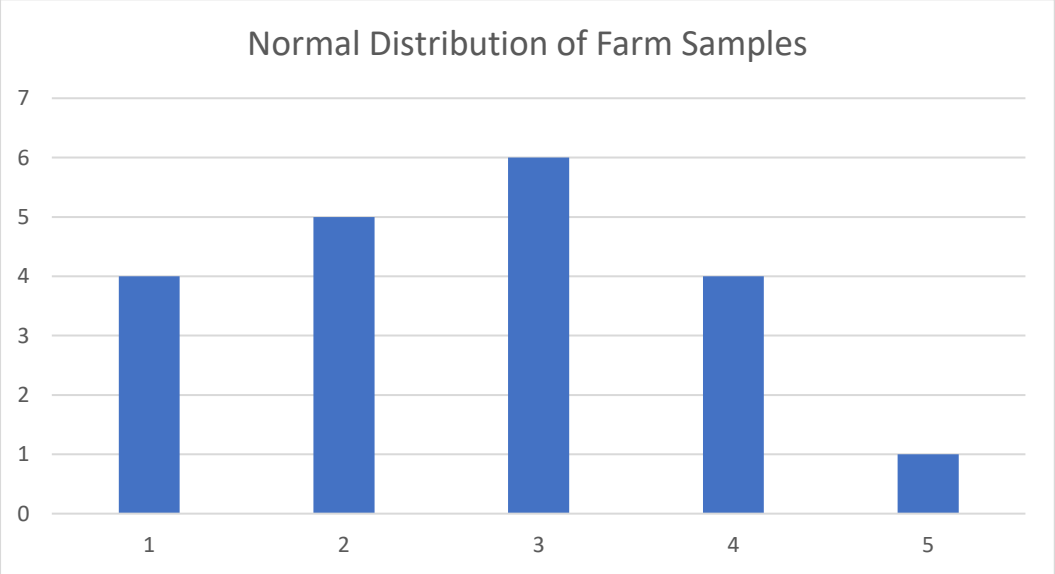
**Comparison of actual Farm Samples and blind Replicate Samples,**

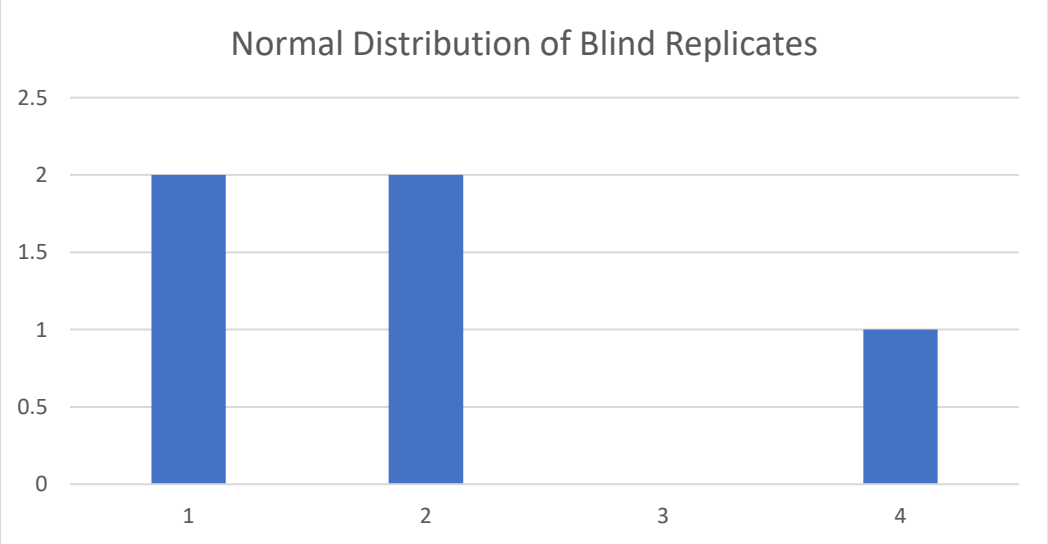
| Actual Farm Sample | | Blind Replicate Sample | |
|---|---|---|---|
| **Average** | **63.78 mg/kg** | **Average** | **98.35 mg/kg** |
| **Variance** | **1471.755611** | **Variance** | **633.14375** |

The above table shows that the values from the actual farm sample and blind replicate are **different.**

We can test this using the **t-Test (sample size < 30) and F-Test for mean and variance respectively**.

Samples ( farm and blind replicate) are from **2 independent populations**, we need to check whether the data are **Normally distributed** to apply the t-Test.




**( Note: Sample size of blind replicates is too small (5))**

**t- Test for mean comparison**

**Null Hypothesis**: The mean μ1 of farm samples is equal to the mean μ2 of blind samples

**Alternative Hypothesis:** μ1 is not equal to μ2

The test is conducted at a **95 % confidence level**.

t-Test: Two-Sample Assuming Unequal Variances

| | *Farm_Nitrogen* | *Blind_Nitrogen* |
|---|---|---|
| Mean | 63.78 | 98.35 |
| Variance | 1471.755611 | 633.14375 |
| Observations | 20 | 5 |
| | | |
| Hypothesized Mean Difference | 0 | |
| df | 9 | |
| t Stat | -2.443145961 | |
| P(T<=t) one-tail | 0.018586158 | |
| t Critical one-tail | 1.833112933 | |
| P(T<=t) two-tail | 0.037172317 | |
| t Critical two-tail | 2.262157163 | |

| | |
|---|---|
| α | 0.05 |

For 2 tailed test

| | |
|---|---|
| p-value | 0.037172317 |

**Since p-value < α, we can reject the Null Hypothesis and there exists a significant difference in their mean values**. Here, we have assumed that samples have unequal variances.

**We can test this using F-Test**

F-Test Two-Sample for Variances

| | *Farm_Nitrogen* | *Blind_Nitrogen* |
|---|---|---|
| Mean | 63.78 | 98.35 |
| Variance | 1471.755611 | 633.14375 |
| Observations | 20 | 5 |
| df | 19 | 4 |
| F | 2.324520475 | |
| P(F<=f) one-tail | 0.214952281 | |
| F Critical one-tail | 5.811359237 | |

| | |
|---|---|
| α | 0.05 |

For 2 tailed test

| | |
|---|---|
| p-value | 0.429904562 |

**F- Test for variance comparison**

**Null Hypothesis**: The variance σ1 of farm samples is equal to the variance σ2 of blind samples

**Alternative Hypothesis:** σ1 is not equal to σ2

**Since p-value > α, there exists no significant difference in the variances of both samples. Hence Null Hypothesis is true.**

This is **contradictory** to the assumption (unequal variances) made during the t-Test.

But, the F-Test is reliable only when the sample size is large. F-test does not yield accurate results when the sample sizes are small. We should consider the results of the t-Test ignoring the F-Test.

**Hence, from the t-Test, we can say that there exists a significant difference in the mean and variance values of farm and blind replicate samples**.

**Question 6:** Write a note on which of the following data parameters can be used to interpret the status of the soil in the study area: soil organic carbon, available nitrogen content, bulk density, and pH.

**Answer:**

**Soil organic carbon (SOC)** is the key element that determines soil quality, fertility, agricultural profitability, and atmospheric carbon dioxide ($CO_2$) fixation. The SOC affects the physio-chemical and biological properties of soil which simultaneously improves soil structure, water, and nutrient retention capacity.

**Available Nitrogen Content** is really important for plant growth (structure), plant food processing (metabolism), and the creation of chlorophyll. Nitrogen fertilization can significantly affect soil properties. Soil factors are closely associated with soil nutrient cycling and plant nutrient uptake and therefore affect productivity.

**Bulk density** is the weight of soil in a given volume. Solis with a bulk density higher than 1.6 g/cm3 tend to restrict root growth. Bulk density can be used to calculate soil properties per unit area. Bulk density gives a good indication of the suitability for root growth. It is generally desirable to have soil with a low bulk density (< 1.5 g/cm3)

**Soil pH** is a characteristic that describes the relative acidity or alkalinity of the soil. Soils are considered acidic if pH < 5, and very acidic if pH < 4. On the other hand, soils are considered alkaline if pH > 7.5, and very alkaline if pH > 8. The availability of some plant nutrients is greatly affected by soil pH. The “ideal” soil pH is close to neutral, and neutral soils are considered to fall within a range from a slightly acidic pH of 6.5 to a slightly alkaline pH of 7.5. It has been determined that most plant nutrients are optimally available to plants within this 6.5 to 7.5 pH range, plus this range of pH is generally very compatible to plant root growth.

Hence the above parameters can be used to interpret the status of the soil.

I have created a **Composite Soil Status Index (CSSI)** based on the above parameters.

**COMPOSITE SOIL STATUS INDEX**

**Mean Standardisation Method**: In this method, we normalize the value of each variable and then work out the average of the normalized values for all the variables. The average of the normalized values will be the composite index.

**CSSI = ( SOC / SOC ideal+ Nitogen/ Nitrogen ideal + BD / BD ideal + pH/ pH ideal) / 4**

**Healthy soils have CSSI approximately equal to 1.**

For our Farm samples,

| Parameters | Ideal Value | Average | Coefficient | Soil_Status_Index |
|---|---|---|---|---|
| **pH** | 5.5-7.5 | 6.5 | 0.153846154 | 1.22002219 |
| **Nitrogen** | | 63.78 | 0.015678896 | 1.285621337 |
| **Organic Carbon** | | 0.6315 | 1.583531275 | 1.079616217 |
| **Bulk Density** | 1.2-1.6 | 1.4 | 0.714285714 | 0.946359186 |
| | | | | 1.123578823 |
| | | | | 0.756801163 |
| | | | | 0.870610245 |
| | | | | 0.935160569 |
| | | | | 0.854724235 |
| | | | | 1.016365021 |
| | | | | 0.724182744 |
| | | | | 0.775233738 |
| | | | | 1.512026825 |
| | | | | 0.730887076 |
| | | | | 1.227056818 |
| | | | | 1.063010279 |
| | | | | 1.067287667 |
| | | | | 1.118811242 |
| | | | | 0.8662457 |
| | | | | 0.811014309 |
| | | | | 0.5 |

**Note**: The above method is the most simple method to compose the Index. More accurate is the one where the relationship between the variables is empirically determined (Like correlation, or regression). Then each variable is assigned weights according to the level of their contribution to the overall soil health). The CSSI should be considered for each soil type, and crop type.