Karimunda kalimu 11-03-2025

Seminar on publishing scientific papers and preparation of scientific research projects

Statistical analysis with Python

1. Download the provided Excel file
2. Calculate the mean and SD for each treatment.
3. Test statistical differences among the treatments using ANOVA and post hoc tests.
4. Prepare bar graphs with whiskers using mean +- SD
5. The letters above the bars show which treatments are statistically different from the others.

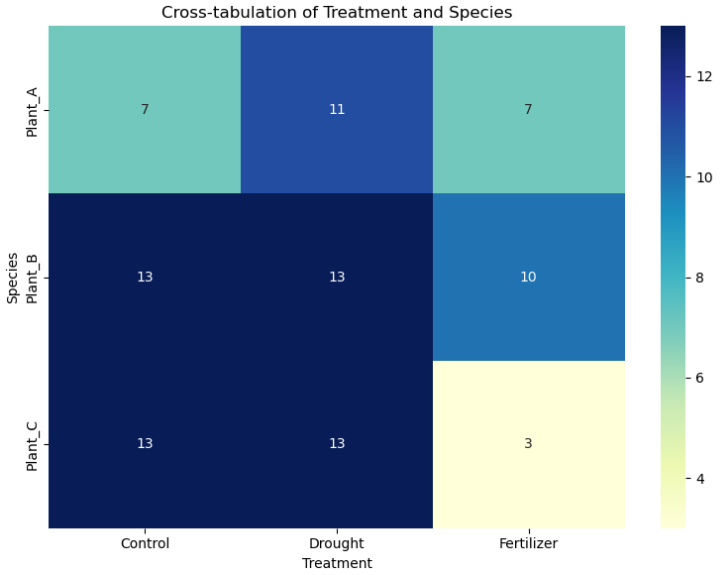
Answers:

1. **After downloading the file**, I performed some operations that helped me understand the file better. The file consists of two categorical variables, "Species and Treatment," and two numerical variables, "Height (cm) and chlorophyll." Moreover, 90 rows and 4 columns make the shape of the file. Furthermore, it has 3 Species, namely Plant_C, Plant_A, and Plant_B), and 3 Treatments, namely (Control, Drought, and Fertilizer).

   I used crosstab function to check if the data is balanced or unbalanced

Treatment	Control	Drought	Fertilizer
Species			
Plant_A	7	11	7
Plant_B	13	13	10
Plant_C	13	13	3

   From the table above and graph below, I have realized that the data is unbalanced, because sample sizes differ across groups.



2. **Mean and SD for each treatment**

	Height (cm)		Chlorophyll	
	mean	std	mean	std
Treatment				
Control	24.78	2.86	40.41	5.25
Drought	20.29	2.84	30.86	4.70
Fertilizer	30.43	3.22	48.02	6.74

3. **Statistical differences among the treatments using ANOVA and post hoc tests**

```
ANOVA for Height (cm): F_onewayResult(statistic=78.67155184350746, pvalue=3.098016635113847e-20)
ANOVA for Chlorophyll: F_onewayResult(statistic=69.89898069863159, pvalue=7.921083930193082e-19)
```

Analysis of variance revealed significant differences in plant height ($F=78.67$, $p<0.05$) and chlorophyll content ($F=69.89$, $p<0.05$) across the treatments.

For the Post Hoc Test, I chose Turkey HSD

```
Tukey HSD for Height:
  Multiple Comparison of Means - Tukey HSD, FWER=0.05
=====================================================
 group1   group2   meandiff p-adj lower   upper  reject
-----------------------------------------------------
Control    Drought  -4.4856   0.0 -6.159 -2.8122   True
Control Fertilizer   5.6496   0.0 3.6691  7.6301   True
Drought Fertilizer  10.1352   0.0 8.1956 12.0749   True
-----------------------------------------------------

Tukey HSD for Chlorophyll:
   Multiple Comparison of Means - Tukey HSD, FWER=0.05
=========================================================
 group1   group2   meandiff p-adj  lower    upper  reject
---------------------------------------------------------
Control    Drought  -9.5471   0.0 -12.6325 -6.4617   True
Control Fertilizer   7.6127   0.0   3.9612 11.2643   True
Drought Fertilizer  17.1599   0.0  13.5835 20.7362   True
---------------------------------------------------------
```

<u>Analysis of the mean Height and Cholorophyll Results</u>

The mean height of the Drought group is **4.4856 units lower** than the Control group.

The mean height of the Fertilizer group is **5.6496 units higher** than the Control group.

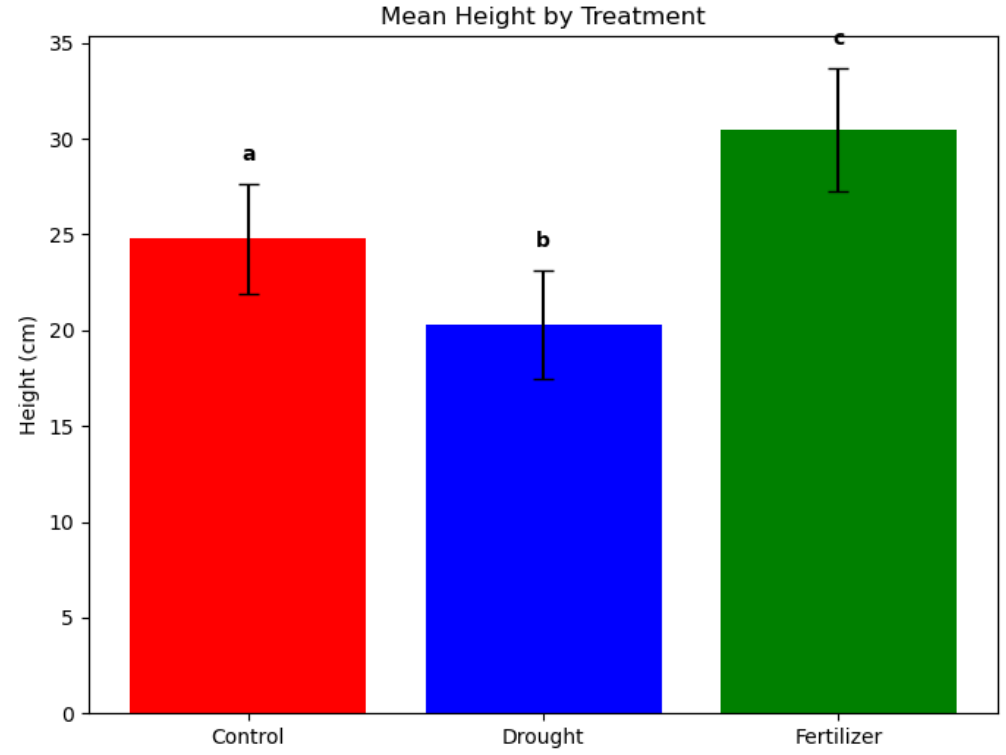
The mean height of the Fertilizer group is **10.1352 units higher** than the Drought group.

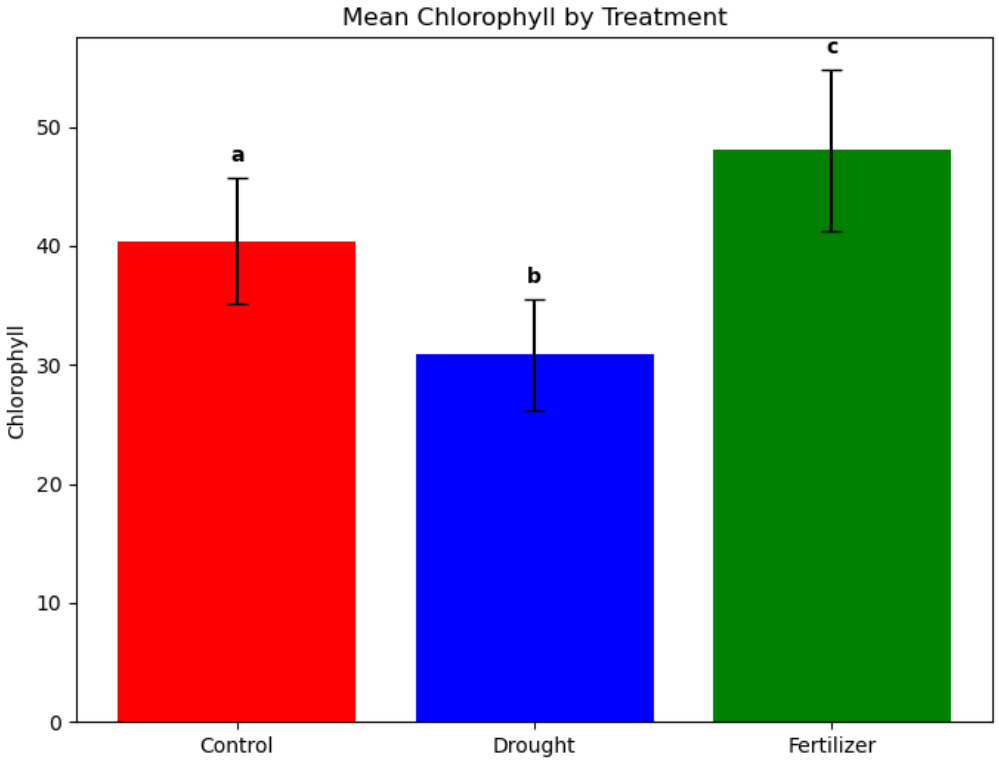
The mean chlorophyll content of the Drought group is **9.5471 units lower** than the Control group.

The mean chlorophyll content of the Drought group is **9.5471 units lower** than the Control group.

The mean chlorophyll content of the Fertilizer group is **17.1599 units higher** than the Drought group.

4. Bar graphs with whiskers using mean +- SD





Letters (a, b, c) represent statistical groupings where the Fertilizer treatment group (c) demonstrates significantly greater mean values than the Control group (a), which in turn exhibited higher values than the Drought group (b).