Progressive Education Society's

Modern College of Arts, Science & Commerce

(Autonomous) Shivajinagar, Pune - 411005

STATISTICS PROJECT

on

Analysis of major crops- Sugarcane and Wheat produced in INDIA

A project report on

Analysis of major crop-Sugarcane and Wheat produced in INDIA

Submitted by

Name	Roll. No
Samruddhi Joshi	2100017
Sakshi Sarde	2100087
Madhura Khopade	2100229
Sakshi Murvade	2100272
Sakshi Gailwar	2102184

CERTIFICATE

A Project Report entitled,

"Analysis of major crops- Sugarcane and Wheat produced in INDIA".

Is approved as partial fulfilment of requirement of the University of Pune for the award of degree of Bachelor of Science (Statistics).

Project In-Charge **Dr. Manisha Sane**

(Head of Department)

Internal Examiner **External Examiner**

Date:- **Date:-**

INDEX

ACKNOWLEDGEMENT

We would like to express OUR deepest gratitude to all those who have contributed to the completion of this statistical graduate project.

First and foremost, we would like to express our special thanks of gratitude to Department of Statistics, Modern College of Arts, Commerce & Science, Shivajinagar for giving us the opportunity to prepare a project. This project truly helped us to apply statistics in real life.

We would like to sincerely thanks Dr. M. M. Sane Ma'am, Head of department (Modern College of Arts, Commerce & Science, Shivajinagar, Pune- 05) for giving us an opportunity to work on this project.

We are immensely thankful to our guide Mrs. Arti Ma'am for her valuable inputs & continued motivation.

We would also like to acknowledge the support of our classmates and parents for their support encouragement and understanding throughout the completion of the project.

MOTIVATION

Agriculture stands as the cornerstone of human civilization, providing sustenance, livelihoods, and economic stability to communities worldwide. In recent years, the agricultural sector has encountered unprecedented challenges, including climate change, resource scarcity, population growth, and evolving consumer demands.

Motivation behind this project is to use innovative approaches that leverage data-driven insights to optimize agricultural practices, enhance productivity, and ensure food security for future generations.

By examining the relation between agricultural productivity and environmental factors, we aspire to identify strategies to increase production of crops as needed for the growing population of India.

COLLECTION OF DATA

The Data used in the analysis in this project is collected from **https://www.kaggle.com/.**

This data was primarily sourced and integrated from following sites:

1.https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics

2. https://www.fao.org/faostat/en/#data

3. https://data.gov.in/catalog/rainfall-india

4. https://environicsindia.in/

5. https://www.imdpune.gov.in/library/public/e-book110.pdf

INTRODUCTION

Agriculture remains the backbone of many economies, providing livelihoods for a significant portion of the global population and supplying essential food and raw materials. However, the agricultural landscape is constantly evolving, influenced by various natural as well as technological factors such as rainfall, area, use of fertilisers and pesticides. As a result, it is imperative to examine agricultural trends systematically to adapt to changing circumstances and optimize productivity.

This project aims to use statistical techniques to analyse factors affecting production of some of the major crops produced in INDIA to enhance productivity and provide food security for future generations.

In this project we have tried to analyse how production may vary state to state resulting from various factors such as area used for farming, rainfall, use of fertilisers and pesticides.

OBJECTIVE

To analyse trends in agricultural production: This includes examining changes in crop yields, area used for farming, and overall agricultural output over different states.

To assess resource utilisation: This involves evaluating the utilization of land, fertilizers, pesticides for the production of the crops.

DATASET

1. Crop year: The year when the crop was grown.

2. Season: The specific cropping season (e.g., Kharif, Rabi, Whole Year).

3. State: The Indian state where the crop was cultivated.

4. Area: The total land area (in hectares) under cultivation for the specific crop.

5. Production: The quantity of crop production (in metric tons).

6. Annual Rainfall: The annual rainfall received in the crop-growing region (in mm).

7. Fertilizer: The total amount of fertilizer used for the crop (in kilograms).

8. Pesticide: The total amount of pesticide used for the crop (in kilograms).

9. Yield: The calculated crop yield (production per unit area).

Crop	Crop Year	Season	State	Area	Production	Annual Rainfall	Fertilizer	Pesticide	Yield
Sugarcane	1999	Whole Year	Maharashtra	590100	53140400	1166.5	62627313	159327	86.55576923
Sugarcane	2000	Whole Year	Maharashtra	595300	49568700	1057.8	58440601	154778	77.65407407
Sugarcane	2001	Whole Year	Maharashtra	578200	45139800	1035.8	59051566	150332	69.57740741
Sugarcane	2002	Whole Year	Maharashtra	573100	42617000	967.4	54255377	143275	65.24666667
Sugarcane	2003	Whole Year	Maharashtra	442500	25668400	1037.4	43798650	106200	56.65814815
Sugarcane	2004	Whole Year	Maharashtra	326900	23913700	1052	35416346	68649	64.51
Sugarcane	2005	Whole Year	Maharashtra	500700	38813700	1387.2	60043944	105147	66.29923077
Sugarcane	2006	Whole Year	Maharashtra	848800	66277400	1439.4	108400248	186736	70.67444444
Sugarcane	2007	Whole Year	Maharashtra	109280	88437200	1257.2	145779520	174848	74.23888889
Sugarcane	2008	Whole Year	Maharashtra	768400	60648300	1052.8	109911936	69156	65.53407407
Sugarcane	2009	Whole Year	Maharashtra	755900	64159300	1002.4	117784338	128503	76.29222222
Sugarcane	2010	Whole Year	Maharashtra	964500	85691500	1389	160213095	231480	73.83961538
Sugarcane	2011	Whole Year	Maharashtra	102200	89456100	1174.09	171205440	337260	76.54074074
Sugarcane	2012	Whole Year	Maharashtra	938100	75335000	1003.4	141465480	290811	73.4448
Sugarcane	2013	Whole Year	Maharashtra	937100	83954000	1409.8	135401579	253017	79.05115385
Sugarcane	2014	Whole Year	Maharashtra	102950	91538100	1001.6	155413320	339735	71.18857143
Sugarcane	2015	Whole Year	Maharashtra	987100	69235400	875.7	155872961	325743	59.97133333
Sugarcane	2016	Whole Year	Maharashtra	633269	54236984	1151.1	97048474.25	221644.15	69.409
Sugarcane	2017	Whole Year	Maharashtra	902035	82539200	1129.5	142016390.4	342773.3	66.33833333
Sugarcane	2018	Whole Year	Maharashtra	116283	89770489	1387.4	188611999.2	406992.6	65.7037931
Sugarcane	2019	Whole Year	Maharashtra	822407	69312919	1555.8	141256626.3	304290.59	69.62148148

Crop	Crop Year	Season	State	Area	Production	Annual Rainfall	Fertilizer	Pesticide	Yield
Sugarcane	1999	Whole Year	Karnataka	37133	37397876	1203.8	39409465	100259.6	102.486
Sugarcane	2000	Whole Year	Karnataka	417141	42923496	1213.3	40950732	108456.7	98.80038
Sugarcane	2001	Whole Year	Karnataka	406950	33016618	1002.9	41561804	105807	84.00269
Sugarcane	2002	Whole Year	Karnataka	382719	32485308	860.8	36232008	95679.75	83.89231
Sugarcane	2003	Whole Year	Karnataka	243341	16015440	867.6	24085892	58401.84	72.1276
Sugarcane	2004	Whole Year	Karnataka	308857	28999269	1266.7	29393921	95745.67	91.74737
Sugarcane	2005	Whole Year	Karnataka	221462	19647650	1316.2	26557723	46507.02	90.63773
Sugarcane	2006	Whole Year	Karnataka	269440	23641948	1146.3	34410182	59276.8	90.2876
Sugarcane	2007	Whole Year	Karnataka	307084	26028309	1323.3	40965006	49133.44	83.41654
Sugarcane	2008	Whole Year	Karnataka	281100	24266201	1164.1	40208544	25299	85.54
Sugarcane	2009	Whole Year	Karnataka	336521	32285023	1321	52436702	57208.57	94.28792
Sugarcane	2010	Whole Year	Karnataka	436798	42276713	1317.7	72556516	104831.5	97.33138
Sugarcane	2011	Whole Year	Karnataka	431405	39470478	1178.986	72268966	142363.7	89.37346
Sugarcane	2012	Whole Year	Karnataka	425393	35510714	956.2	64149264	131871.8	83.3069
Sugarcane	2013	Whole Year	Karnataka	420022	38042382	1235.6	60688979	113405.9	87.8824
Sugarcane	2014	Whole Year	Karnataka	480222	43838057	1238.5	72494313	158473.3	92.69038
Sugarcane	2015	Whole Year	Karnataka	449586	36314129	1024.9	70994125	148363.4	85.546

Sugarcane	2016	Whole Year	Karnataka	397098	27378389	1147.2	60855269	138984.3	75.5836
Sugarcane	2017	Whole Year	Karnataka	400000	37460951	1111.7	62976000	152000	91.89957
Sugarcane	2018	Whole Year	Karnataka	495932	42323211	1419.1	80440170	173576.2	83.26862
Sugarcane	2019	Whole Year	Karnataka	428540	36034028	1422.3	73606030	158559.8	81.45552

Crop	Crop Year	Season	State	Area	Production	Annual Rainfall	Fertilizer	Pesticide	Yield
Sugarcane	1999	Whole Year	Goa	1335	68440	3879.3	141683.55	360.45	51.27
Sugarcane	2000	Whole Year	Goa	1250	72750	3886.9	122712.5	325	58.2
Sugarcane	2001	Whole Year	Goa	1205	70565	2530.6	123066.65	313.3	58.56
Sugarcane	2002	Whole Year	Goa	1188	65470	2516.3	112467.96	297	55.11
Sugarcane	2003	Whole Year	Goa	1210	57607	3011.6	119765.8	290.4	47.61
Sugarcane	2004	Whole Year	Goa	1210	60583	2518	131091.4	254.1	52.75
Sugarcane	2005	Whole Year	Goa	1110	55867	3392.8	133111.2	233.1	53.16
Sugarcane	2006	Whole Year	Goa	1120	58279	2928.8	143035.2	246.4	55.455
Sugarcane	2007	Whole Year	Goa	1034	56027	3674.7	137935.6	165.44	57.615
Sugarcane	2008	Whole Year	Goa	1021	49253	3271.6	146043.84	91.89	49.52
Sugarcane	2009	Whole Year	Goa	893	52343	3334.9	139147.26	151.81	50.95
Sugarcane	2010	Whole Year	Goa	921	49109	4193.3	152987.31	221.04	48.715
Sugarcane	2011	Whole Year	Goa	915	46584	3340.39	153280.8	301.95	47.28
Sugarcane	2012	Whole Year	Goa	854	46055	3048.9	128783.2	264.74	47.765
Sugarcane	2013	Whole Year	Goa	872	47669	3642.6	125995.28	235.44	48.06
Sugarcane	2014	Whole Year	Goa	808	49215	3491.2	121975.68	266.64	54.35
Sugarcane	2015	Whole Year	Goa	1139	58077	2587.1	179859.49	375.87	53.265
Sugarcane	2016	Whole Year	Goa	897	40222	3266.9	137465.25	313.95	41.525
Sugarcane	2017	Whole Year	Goa	910	47160	2772.5	143270.4	345.8	45.41
Sugarcane	2018	Whole Year	Goa	888	35296	4406	144033.6	310.8	34.38
Sugarcane	2019	Whole Year	Goa	811	53708	4489.5	139297.36	300.07	70.115

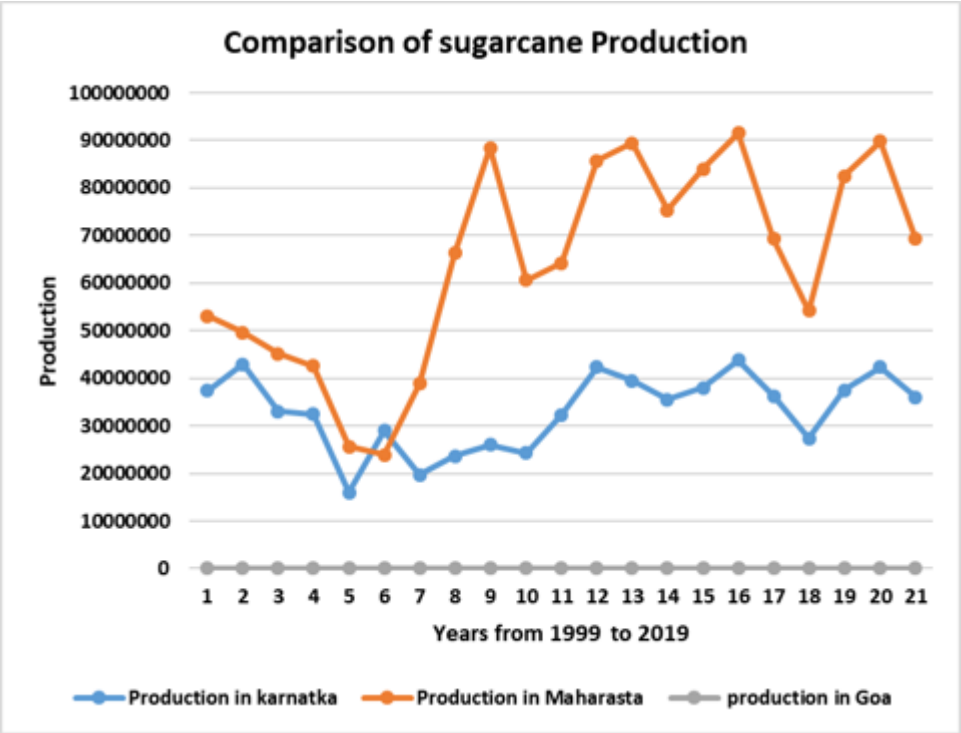
Crop	Crop Year	Season	State	Area	Production	Annual Rainfall	Fertilizer	Pesticide	Yield
Wheat	2000	Rabi	Maharashtra	754200	947600	1057.8	74039814	196092	1.197
Wheat	2001	Rabi	Maharashtra	776000	1077400	1035.8	79252880	201760	1.313793
Wheat	2002	Rabi	Maharashtra	759800	983900	967.4	71930266	189950	1.234333
Wheat	2003	Rabi	Maharashtra	664700	778200	1037.4	65792006	159528	1.171667
Wheat	2004	Rabi	Maharashtra	755700	1016300	1052	81872538	158697	1.199667
Wheat	2005	Rabi	Maharashtra	932800	1300700	1387.2	1.12E+08	195888	1.329333
Wheat	2006	Rabi	Maharashtra	1230800	1871000	1439.4	1.57E+08	270776	1.451667
Wheat	2007	Rabi	Maharashtra	1252800	2371100	1257.2	1.67E+08	200448	1.782333
Wheat	2008	Rabi	Maharashtra	1021800	1516400	1052.8	1.46E+08	91962	1.431667
Wheat	2009	Rabi	Maharashtra	1080900	1740200	1002.4	1.68E+08	183753	1.53
Wheat	2010	Rabi	Maharashtra	1306600	2300800	1389	2.17E+08	313584	1.709333
Wheat	2011	Rabi	Maharashtra	878100	1498800	1174.095	1.47E+08	289773	1.646667
Wheat	2012	Rabi	Maharashtra	785000	1198700	1003.4	1.18E+08	243350	1.461034
Wheat	2013	Rabi	Maharashtra	1028200	1480000	1409.8	1.49E+08	277614	1.462414
Wheat	2014	Rabi	Maharashtra	1067300	1307700	1001.6	1.61E+08	352209	1.201379
Wheat	2015	Rabi	Maharashtra	910854	981351	875.7	1.44E+08	300581.8	1.027813
Wheat	2016	Rabi	Maharashtra	1272006	2214012	1151.1	1.95E+08	445202.1	1.734839
Wheat	2017	Rabi	Maharashtra	1137732	1884842	1129.5	1.79E+08	432338	1.505484
Wheat	2018	Rabi	Maharashtra	834423.3	1249407	1387.4	1.35E+08	292048.1	1.388667
Wheat	2019	Rabi	Maharashtra	1056965	1793440	1555.8	1.82E+08	391077.1	1.564

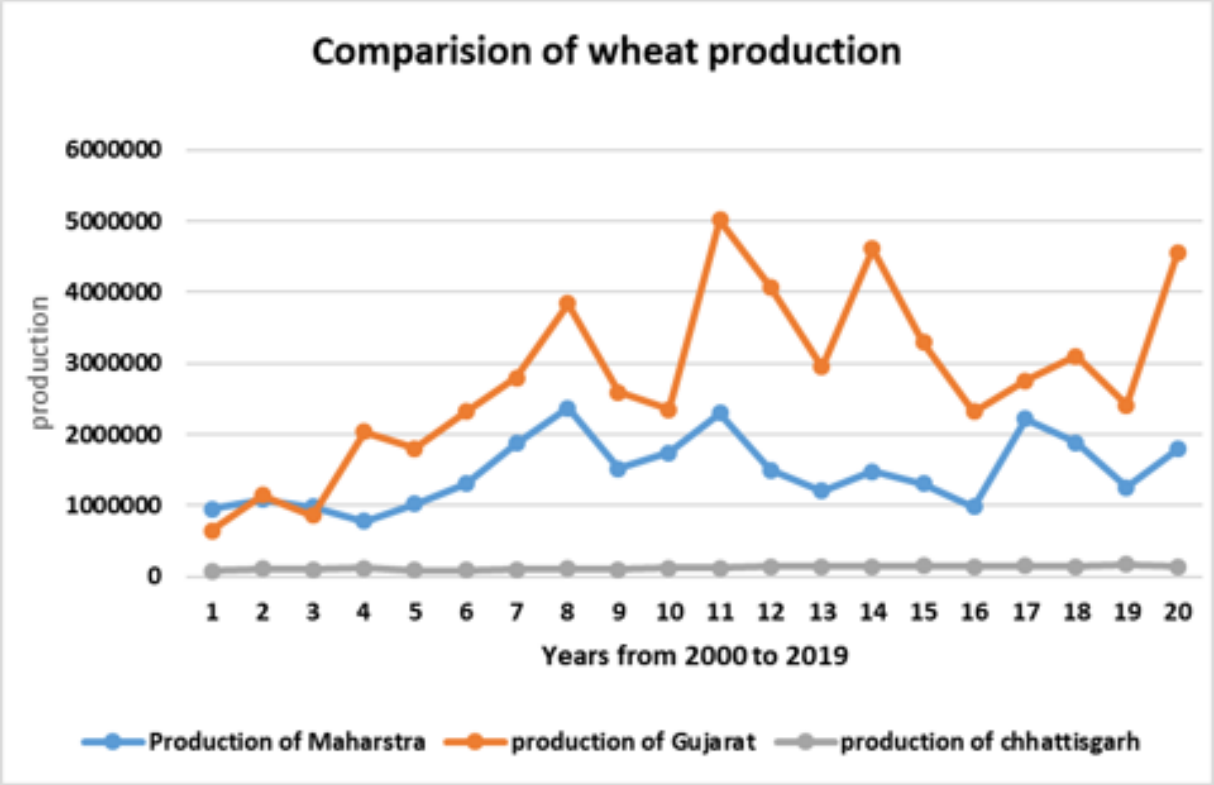
Crop	Crop Year	Season	State	Area	Production	Annual Rainfall	Fertilizer	Pesticide	Yield
Wheat	2000	Rabi	Chhattisgarh	77844	79488	896.4	7641945.48	20239.44	1.214375
Wheat	2001	Rabi	Chhattisgarh	97861	103480	1550	9994543.93	25443.86	1.195
Wheat	2002	Rabi	Chhattisgarh	93846	98592	1001.4	8884400.82	23461.5	1.185
Wheat	2003	Rabi	Chhattisgarh	106109	113138	1643.7	10502668.82	25466.16	1.20875
Wheat	2004	Rabi	Chhattisgarh	99245	85215	1122.8	10752203.3	20841.45	0.963125
Wheat	2005	Rabi	Chhattisgarh	97074	85191	1292.7	11641114.08	20385.54	1.0125
Wheat	2006	Rabi	Chhattisgarh	93213	93952	1317.2	11904232.23	20506.86	1.06
Wheat	2007	Rabi	Chhattisgarh	95020	104548	1246.2	12675668	15203.2	1.299444444
Wheat	2008	Rabi	Chhattisgarh	94834	97424	1117.9	13565055.36	8535.06	1.15
Wheat	2009	Rabi	Chhattisgarh	109163	118923	956.9	17009778.66	18557.71	1.301764706
Wheat	2010	Rabi	Chhattisgarh	103701	121745	1283.1	17225773.11	24888.24	1.395882353
Wheat	2011	Rabi	Chhattisgarh	104825	135121	1302.7	17560284	34592.25	1.512
Wheat	2012	Rabi	Chhattisgarh	102230	143226	1366.8	15416284	31691.3	1.6104
Wheat	2013	Rabi	Chhattisgarh	105033	140756	1418.3	15176218.17	28358.91	1.5332
Wheat	2014	Rabi	Chhattisgarh	103195	153321	1274.7	15578317.2	34054.35	1.745384615
Wheat	2015	Rabi	Chhattisgarh	105785	142329	1136	16704509.35	34909.05	1.522
Wheat	2016	Rabi	Chhattisgarh	102114	151078	1298.9	15648970.5	35739.9	1.660769231
Wheat	2017	Rabi	Chhattisgarh	102195	141642	1124.5	16089580.8	38834.1	1.497692308
Wheat	2018	Rabi	Chhattisgarh	99925	166979	1515.5	16207835	34973.75	1.792692308
Wheat	2019	Rabi	Chhattisgarh	116274	135888	1420.3	19971222.24	43021.38	1.336538462

Crop	Crop Year	Season	State	Area	Production	Annual Rainfall	Fertilizer	Pesticide	Yield
Wheat	2000	Rabi	Gujarat	286100	649000	379.7	28086437	74386	2.346842105
Wheat	2001	Rabi	Gujarat	470100	1144700	666.5	48011313	122226	2.474736842
Wheat	2002	Rabi	Gujarat	435600	856600	458.4	41238252	108900	2.185789474
Wheat	2003	Rabi	Gujarat	759500	2036500	902.5	75175310	182280	2.68875
Wheat	2004	Rabi	Gujarat	727400	1805500	719.8	78806516	152754	2.6312
Wheat	2005	Rabi	Gujarat	858800	2319200	942.4	102987296	180348	2.5792
Wheat	2006	Rabi	Gujarat	1072300	2791500	1056.2	136943433	235906	2.5696
Wheat	2007	Rabi	Gujarat	1273900	3837700	1021.6	169938260	203824	2.9304
Wheat	2008	Rabi	Gujarat	1091400	2592600	746.1	156113856	98226	2.349615385
Wheat	2009	Rabi	Gujarat	878000	2351300	618	136809960	149260	2.617692308
Wheat	2010	Rabi	Gujarat	1588600	5013400	1107.5	263882346	381264	3.0388
Wheat	2011	Rabi	Gujarat	1350600	4072100	890.5	226252512	445698	3.01
Wheat	2012	Rabi	Gujarat	1023500	2944000	460.6	154343800	317285	2.9216
Wheat	2013	Rabi	Gujarat	1442288	4607993	1006.5	208396193.1	389417.76	3.072
Wheat	2014	Rabi	Gujarat	1171449	3293487	605.6	176841941	386578.17	2.8178125
Wheat	2015	Rabi	Gujarat	857696	2315849	584.3	135438775.4	283039.68	2.6759375
Wheat	2016	Rabi	Gujarat	999370	2757755	710.5	153153452.5	349779.5	2.7428125
Wheat	2017	Rabi	Gujarat	1058778	3101037	814.8	166694008.3	402335.64	2.926666667
Wheat	2018	Rabi	Gujarat	797156	2407132	1125.4	129298703.2	279004.6	3.0109375

Wheat	2019	Rabi	Gujarat	1393487	4553681	1067.8	239345327.1	515590.19	3.1434375

Graphical representation:





Correlation of various factors affecting production of sugarcane in Maharashtra

Correlation between Area and production

Correlation coefficient between Area and production : 0.967043

This correlation coefficient of 0.967043 indicates a strong positive correlation between area and production of sugarcane and suggests that increasing the area of land allocated for sugarcane cultivation could lead to a significant increase in sugarcane production.

Correlation between Annual Rainfall and production

Correlation coefficient between Annual rainfall and production : **0.306006154**

A correlation coefficient of 0.306 indicates a positive but relatively weak correlation between annual rainfall and production. This suggests that while there is some tendency for production to increase with higher rainfall, other factors likely play a significant role in determining production levels.

Correlation between Fertilisers and production

Correlation coefficient between Fertilisers and production : **0.944137894**

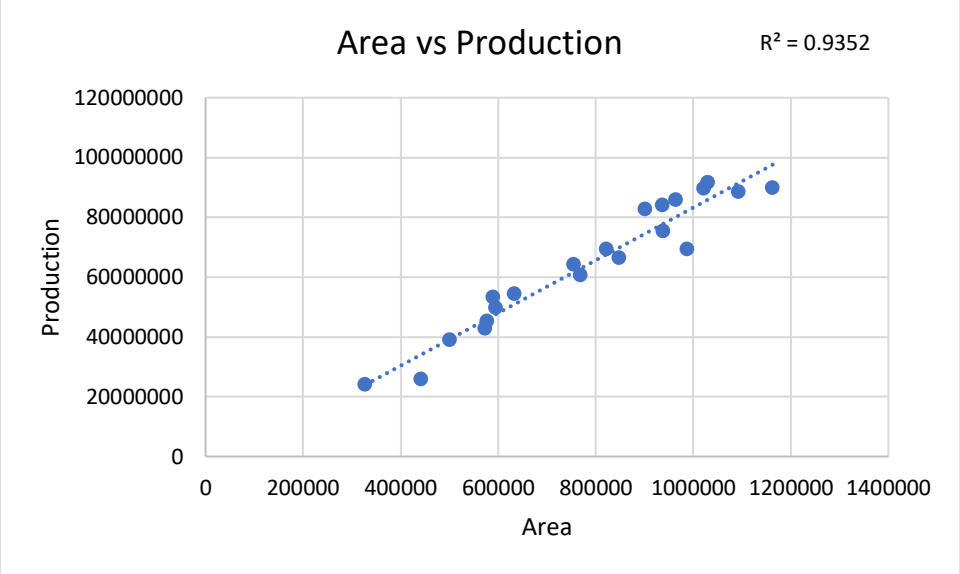
The correlation coefficient of 0.944137894 indicates a strong positive relationship between fertilizers and sugarcane production, suggesting that increased fertilizer usage is associated with higher sugarcane yields.

Correlation between Pesticides used and production

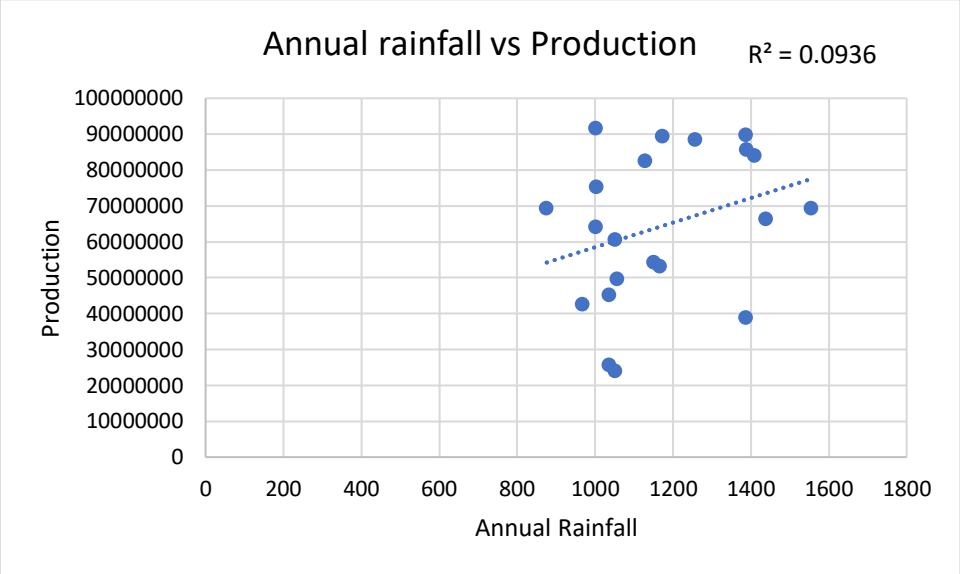
Correlation coefficient between Annual rainfall and production : **0.776000285**

A correlation coefficient of 0.776 is relatively high, indicating that changes in pesticide usage are strongly associated with changes in sugarcane production.

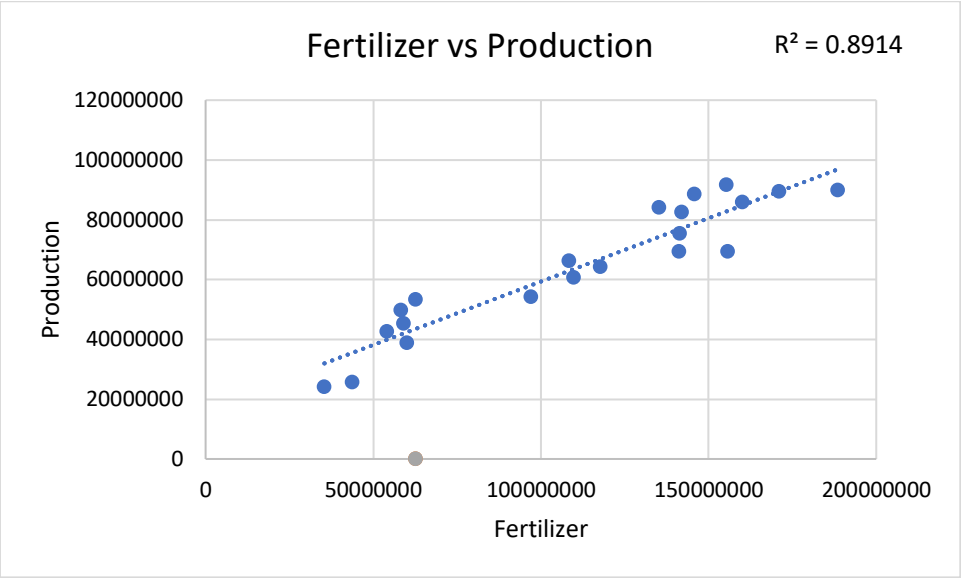
Regression model on sugarcane produced vs. factors affecting it's production in Maharashtra



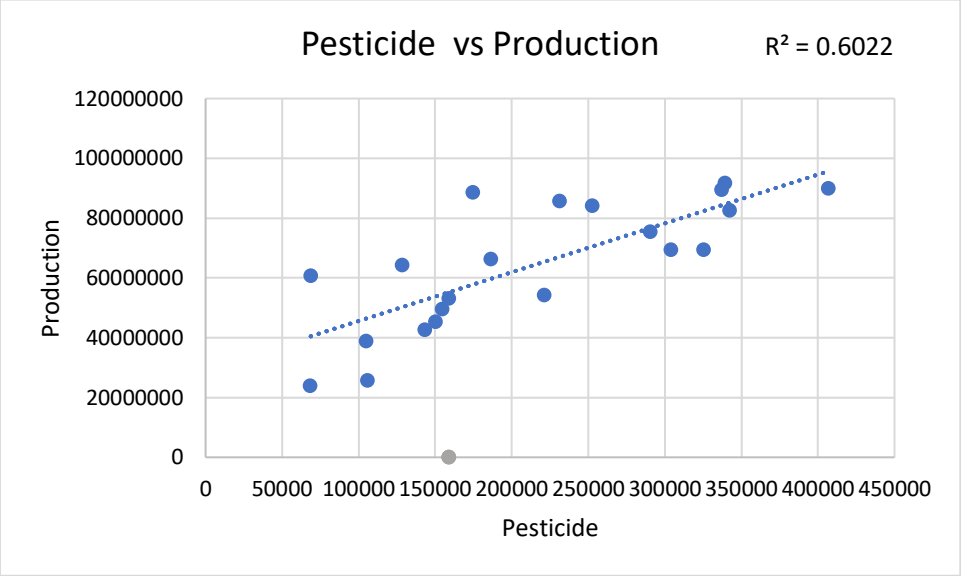
An R-squared value of 0.9352 indicates that approximately 93.52% of the variability in sugarcane production can be explained by the variability in the area of sugarcane cultivation. Area under sugarcane cultivation can be a very good predictor of sugarcane production.



Since the R-squared value is low, it indicates that there is a weak linear relationship between annual rainfall and sugarcane production. This suggests that only about 9.36% of the variability in sugarcane production can be explained by annual rainfall.



The high value R-square indicates that there is a strong relationship between the amount of fertilizers used and sugarcane production. It suggests that changes in the amount of fertilizers used are associated with large changes in sugarcane production.



The R-squared value of 0.6022 suggests that approximately 60.22% of the variation in sugarcane production can be explained by the variation in pesticides used. This suggests that pesticide usage is an important factor in explaining the differences in sugarcane production levels.

Correlation of various factors affecting production of wheat in Maharashtra

Correlation between Area and production

Correlation coefficient between Area and production : **0.928736**

The correlation coefficient of 0.928736 indicates a very strong positive relationship between the area of land used for wheat cultivation and the wheat production.While correlation does not imply causation, in this case, the strong positive correlation suggests that the area of land dedicated to wheat cultivation likely contributes significantly to wheat production.

Correlation between Annual Rainfall and production

Correlation coefficient between Annual rainfall and production : 0.465756

The correlation coefficient of 0.465756 indicates a moderate positive relationship between rainfall and wheat production. This suggests that as rainfall increases, wheat production tends to increase as well, and vice versa. However, the correlation is not perfect indicating that other factors may also influence wheat production. This aligns with the expectation that adequate rainfall is generally beneficial for crop growth.

Correlation between Fertilisers and production

Correlation coefficient between Fertilisers and production : 0.874769

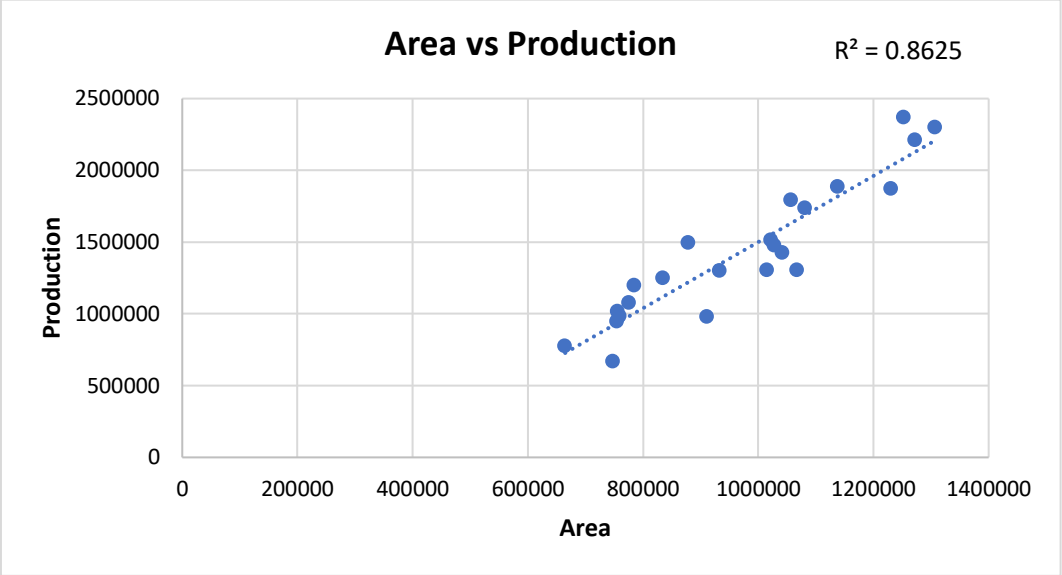
The correlation coefficient of 0.874769 suggests a strong positive linear relationship between fertilizer usage and wheat production. As the amount of fertilizer used increases, wheat production tends to increase as well.

Correlation between Pesticides used and production

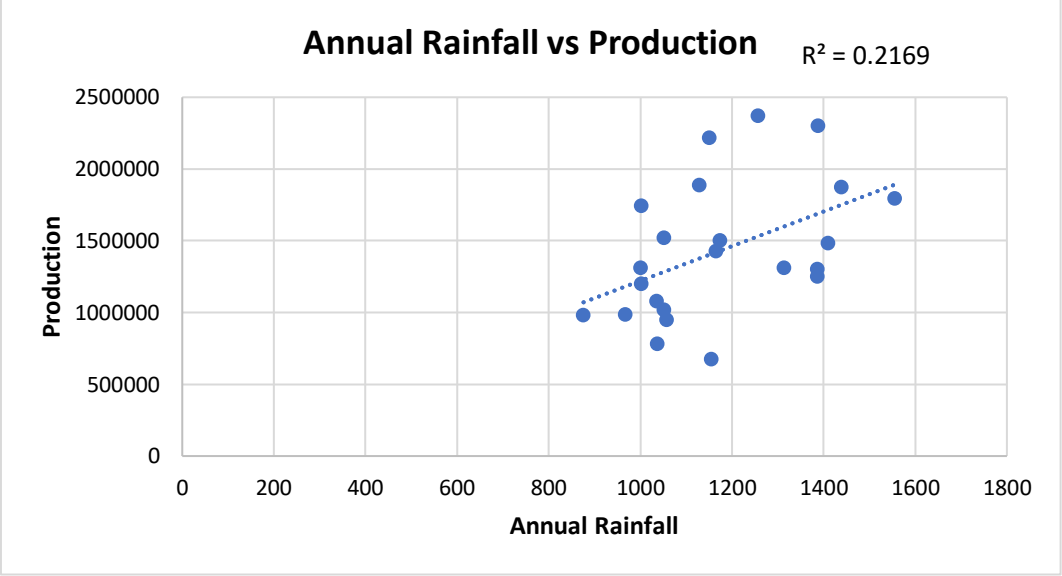
Correlation coefficient between Annual rainfall and production: **0.44734**

The correlation coefficient of 0.44734 indicates a moderate strength of association between pesticides used and wheat production. While the correlation is not perfect, it is still capable to increase production by avoiding the effects of pest.

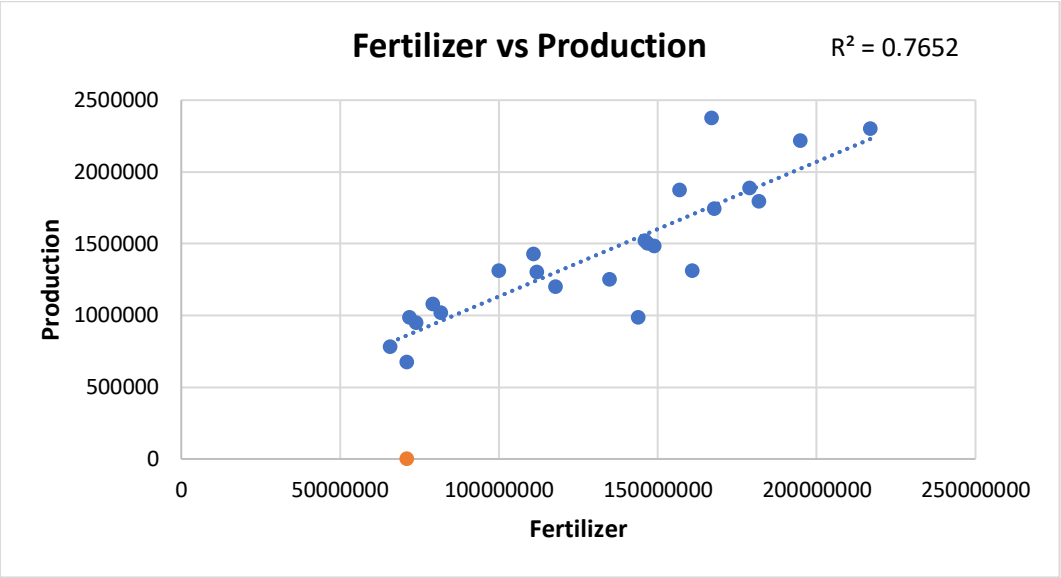
Regression model on wheat produced vs. factors affecting its production in Maharashtra



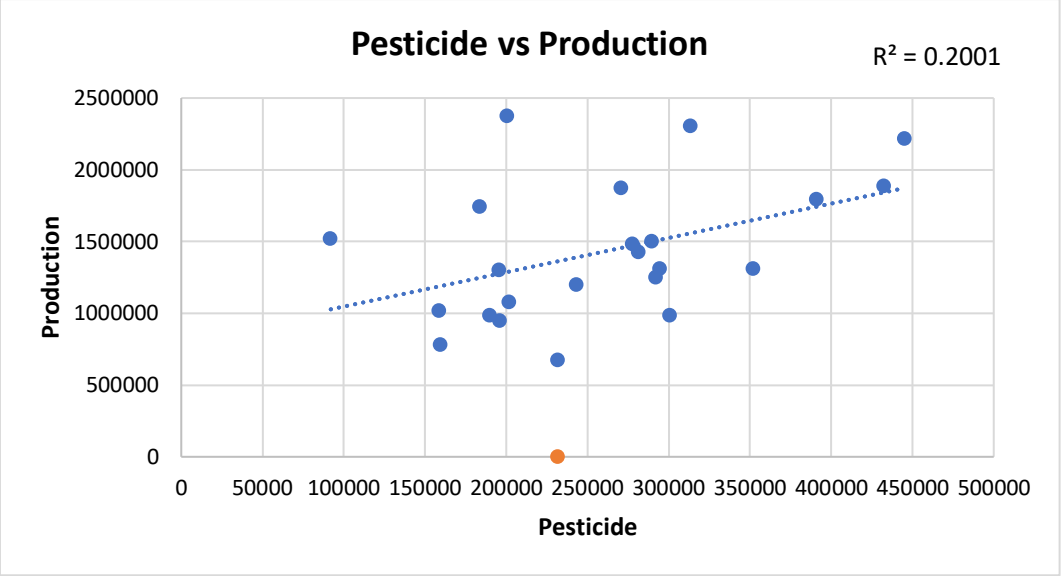
An R-squared value of 0.8625 indicates that 86.25% of the variability in wheat production can be explained by the variability in the area of wheat cultivation. Changes in the area of cultivation are strongly associated with changes in wheat production.



An R-squared value of 0.2169 indicates that 21.69% of the variability in wheat production is explained by annual rainfall.



About 76.52% of the variability in wheat production can be attributed to the variation in the amount of fertilizers used. This means that a large proportion of the fluctuations in wheat yield across time periods can be explained by changes in fertilizer application.



The correlation coefficient of 0.2001 suggests a weak positive relationship between the amount of pesticides used and the production of wheat. This means that as the amount of pesticides used increases, there is a tendency for wheat production to increase, but the relationship is not very strong.

Comparison of production of sugarcane between Maharashtra and Goa and factors affecting it.

We performed Wilk-Shapiro test on the data given above and checked for normality for every variable. The results and interpretations are given below:

Shapiro-Wilk normality test

data: yield1

W = 0.9811, p-value = 0.9401

Shapiro-Wilk normality test

data: yield2

W = 0.95217, p-value = 0.3741

Shapiro-Wilk normality test

data: area1

W = 0.96025, p-value = 0.5211

Shapiro-Wilk normality test

data: area2

W = 0.91199, p-value = 0.06013

Shapiro-Wilk normality test

data: production1

W = 0.93494, p-value = 0.1729

Shapiro-Wilk normality test

data: production2

W = 0.97198, p-value = 0.7764

Shapiro-Wilk normality test

data: rainfall1

W = 0.91496, p-value = 0.06886

Shapiro-Wilk normality test

data: rainfall2

W = 0.95018, p-value = 0.3433

Shapiro-Wilk normality test

data: fertilizer1

W = 0.91947, p-value = 0.08468

Shapiro-Wilk normality test

data: pesticide1

W = 0.94209, p-value = 0.2396

Shapiro-Wilk normality test

data: pesticide2

W = 0.94063, p-value = 0.2243

Here we can see that for every variable, the p value is greater than the significance level 0.05. Hence, we can conclude that all the variables approximately follow Normal distribution. Also, we are making comparisons of different variables between two states and so, they are independent of each other. That means, production of one state does not affect the production of another. Thus, we can use t-test.

We are using t-test to compare the means of different variables regarding sugarcane production in Maharashtra and Goa. The reason to choose these states for testing is because as we live in Maharashtra, it felt appropriate to make the analysis on it. While the reason to choose Goa is that, it is one of the lowest sugarcane producing states.

1.Comparison between yields of sugarcane of the states Maharashtra and Goa.

We want to test,

- **Null Hypothesis (H0)**: The true difference in means between the yield of two states is equal to zero.
- **Alternative Hypothesis (H1)**: The true difference in means between the yield of two states is not equal to zero.

We have used T-test for testing of hypothesis in r software.

Let yield1= yield of sugarcane Maharashtra.

Let yield2= yield of sugarcane Goa.

Unit of measurement= Production per unit area.

```
> t.test(yield1,yield2)
```

Output:

```
        Welch Two Sample t-test

data:  yield1 and yield2

t = 8.7555, df = 39.924, p-value = 7.837e-11

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 14.55104 23.28560

sample estimates:

mean of x mean of y

 70.39761  51.47929
```

Interpretation:

- **t-value**: The t-value is 8.7555. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is positive, it suggests that the mean of yield1 is significantly larger than the mean of yield2.
- **p-value**: The p-value is 7.837e-11, which is extremely small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of yields of sugarcane of the states Maharashtra and Goa. We can also conclude that the yield of sugarcane obtained from Maharashtra is much more than that of Goa.

We will now compare the available regressors of both the states to see which regressor contributes the most to the difference between yields of the states.

2.Comparison between areas of sugarcane production of Maharashtra and Goa.

We want to test,

- **Null Hypothesis (H0)**: The true difference in means between the areas two states is equal to zero.

- **Alternative Hypothesis (H1)**: The true difference in means between the areas two states is not equal to zero.

Let area1=area of cultivation of sugarcane of Maharashtra.

Let area2= area of cultivation sugarcane of Goa.

Unit of measurement= Hectares.

Input:

```
> t.test(area1,area2)
```

Output:

```
        Welch Two Sample t-test

data:  area1 and area2
t = 15.405, df = 20, p-value = 1.472e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 677259.2 889403.3
sample estimates:
 mean of x  mean of y
784359.381   1028.143
```

Interpretation:

t-value: The t-value is 15.405. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is positive, it suggests that the mean of area1 is significantly larger than the mean of area2.

p-value: The p-value is 1.472e-12, which is extremely small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of areas of cultivation of sugarcane of the states

Maharashtra and Goa. We can also conclude that the area of cultivation of sugarcane obtained from Maharashtra is much more than that of Goa.

3.Comparison between production of sugarcane of Maharashtra and Goa.

We want to test,

- **Null Hypothesis (H0)**: The true difference in means between the production of two states is equal to zero.
- **Alternative Hypothesis (H1)**: The true difference in means between the production of two states is not equal to zero.

Let production1= production of sugarcane of Maharashtra.

Let production2= production of sugarcane of Goa.

Unit of measurement= Metric tons.

Input:

```
> t.test(production1,production2)
```

Output:

```
        Welch Two Sample t-test

data:  production1 and production2
t = 13.851, df = 20, p-value = 1.036e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 54536282 73874510
sample estimates:
mean of x mean of y
 64259695     54299
```

Interpretation:

- **t-value**: The t-value is 13.851. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is positive, it suggests that the mean of production1 is significantly larger than the mean of production2.

- **p-value**: The p-value is 1.036e-11, which is extremely small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:
We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of productions of cultivation of sugarcane of the states Maharashtra and Goa. We can also conclude that the production of cultivation of sugarcane obtained from Maharashtra is much more than that of Goa.

4.Comparison between annual rainfall between Maharashtra and Goa.

We want to test,

- **Null Hypothesis (H0)**: The true difference in means between the annual rainfall of two states is equal to zero.
- **Alternative Hypothesis (H1)**: The true difference in means between the annual rainfall of two states is not equal to zero.

Let rainfall1= annual rainfall measured in sugarcane growing regions in Maharashtra.

Let rainfall2= annual rainfall measured sugarcane growing regions in Goa.

Unit of measurement= Millimeters.

Input:

```
> t.test(rainfall1,rainfall2)
```

Output:

```
        Welch Two Sample t-test

data:  rainfall1 and rainfall2
t = -15.757, df = 23.913, p-value = 3.955e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2458.621 -1889.055
sample estimates:
mean of x mean of y
 1168.252  3342.090
```

Interpretation:

- **t-value**: The t-value is -15.757. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is negative, it suggests that the mean of rainfall1 is significantly smaller than the mean of rainfall2.
- **p-value**: The p-value is 3.955e-14, which is extremely small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of productions of cultivation of sugarcane of the states Maharashtra and Goa. We can also conclude that the annual rainfall recorded for sugarcane crop in Goa is much more than that in Maharashtra.

5.Comparison between use of fertilizer for sugarcane of Maharashtra and Goa.

We want to test,

- **Null Hypothesis (H0)**: The true difference in means between the annual rainfall of two states is equal to zero.
- **Alternative Hypothesis (H1)**: The true difference in means between the annual rainfall of two states is not equal to zero.

Let fertilizer1= Amount of fertilizer needed for cultivation of sugarcane in Maharashtra.

Let fertilizer2= Amount of fertilizer needed for cultivation of sugarcane in Goa.

Unit of measurement= Kilograms.

Input:

> t.test(fertilizer1,fertilizer2)

Output:

Welch Two Sample t-test

data: fertilizer1 and fertilizer2

t = 5.3767, df = 25.653, p-value = 1.301e-05

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

36657652 82082690

sample estimates:

mean of x mean of y

111619772 52249601

Interpretation:

- **t-value**: The t-value is 5.3767. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples.
- **p-value**: The p-value is 1.301e-05, which is very small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of usage of fertilizers of sugarcane of the states Maharashtra and Goa. We can also conclude that the usage of fertilizers of sugarcane obtained from Maharashtra is much more than that of Goa.

6.Comparison between use of pesticides for sugarcane of Maharashtra and Goa.

We want to test,

- **Null Hypothesis (H0)**: The true difference in means between the use of pesticides of two states is equal to zero.
- **Alternative Hypothesis (H1)**: The true difference in means between the use of pesticides of two states is not equal to zero.

Let pesticide1 = Amount of pesticide used for cultivation of sugarcane in Maharashtra.

Let pesticide2= Amount of pesticide used for cultivation of sugarcane in Goa.

Unit of measurement= Kilograms.

Input:

> t.test(pesticide1,pesticide2)

Output:

Welch Two Sample t-test

data: pesticide1 and pesticide2

t = 9.7131, df = 20, p-value = 5.15e-09

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

168080.3 260018.0

sample estimates:

mean of x mean of y

214318.935 269.771

Interpretation:

- **t-value**: The t-value is 9.7131. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples.
- **p-value**: The p-value is 5.15e-09, which is very small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of usage of pesticides of sugarcane of the states Maharashtra and Goa. We can also conclude that the usage of pesticides of sugarcane obtained from Maharashtra is much more than that of Goa.

Final conclusion:

The above analysis of sugarcane production of the states Maharashtra and Goa shows that:

Even though the annual rainfall recorded for the sugarcane cultivating areas in Goa is much larger than that of Maharashtra, the other available factors, i.e. area, production, usage of fertilizers and pesticides are much larger in Maharashtra. This leads to huge sugarcane yield in Maharashtra.

Comparison of production of sugarcane between Maharashtra and Karnataka and factors affecting it.

We are using t-test to compare the means of different variables regarding sugarcane production in Maharashtra and Karnataka. The reason to choose these states for testing is because as we live in Maharashtra, it felt appropriate to make the analysis on it. While the reason to choose Karnataka is that, it is one of the major sugarcane producing states.

1.Comparison between yields of sugarcane of the states Maharashtra and Karnataka.

We want to test,

Null Hypothesis (H0): The true difference in means between the yield of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the yield of two states is not equal to zero.

We have used T-test for testing of hypothesis in r software.

Let yield1= yield of sugarcane Maharashtra.

Let yield2= yield of sugarcane Karnataka.

Unit of measurement= Production per unit area.

Input:

```
> t.test(yield1,yield2)
```

Output:

```
        Welch Two Sample t-test

data:  yield1 and yield2
t = -7.9833, df = 39.808, p-value = 8.615e-10
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -21.91394 -13.05884
sample estimates:
mean of x mean of y
 70.39761  87.88400
```

Interpretation:

- t-value: The t-value is approximately -7.9833. This indicates how many standard errors the difference between the means is away from zero. Since it's negative, it suggests that the mean of yield2 is significantly larger than the mean of yeild1.
- p-value: The p-value is approximately 8.615e-10, which is very close to zero. This extremely small p-value suggests strong evidence against the null hypothesis, indicating that there is a significant difference in means between yield1 and yield2.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of yields of sugarcane of the states Maharashtra and Karnataka. We can also conclude that the yield of sugarcane obtained from Karnataka is much more than that of Maharashtra.

We will now compare the available regressors of both the states to see which regressor contributes the most to the difference between yields of the states.

2.Comparison between areas of sugarcane production of Maharashtra and Karnataka.

We want to test,

Null Hypothesis (H0): The true difference in means between the areas two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the areas two states is not equal to zero.

Let area1=area of cultivation of sugarcane of Maharashtra.

Let area2= area of cultivation sugarcane of Karnataka.

Unit of measurement= Hectares.

Input:

```
> t.test(area1,area2)
```

Output:

```
        Welch Two Sample t-test

data:  area1 and area2
t = 7.6043, df = 24.4, p-value = 6.864e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 297103.8 518191.8
sample estimates:
```

mean of x mean of y

784359.4 376711.6

Interpretion:

- t-value: The t-value is 7.6043. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is positive, it suggests that the mean of area1 is significantly larger than the mean of area2.
- p-value: The p-value is 6.864e-08, which is extremely small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of areas of cultivation of sugarcane of the states Maharashtra and Karnataka. We can also conclude that the area of cultivation of sugarcane obtained from Maharashtra is much more than that of Karnataka.

3.Comparison between production of sugarcane of Maharashtra and Karnataka.

We want to test,

Null Hypothesis (H0): The true difference in means between the production of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the production of two states is not equal to zero.

Let production1= production of sugarcane of Maharashtra.

Let production2= production of sugarcane of Karnataka.

Unit of measurement= Metric tons.

Input:

```
> t.test(production1,production2)
```

Output:

```
        Welch Two Sample t-test

data:  production1 and production2

t = 6.2978, df = 25.433, p-value = 1.267e-06

alternative hypothesis: true difference in means is not equal to 0
```

95 percent confidence interval:

20970272 41324719

sample estimates:

mean of x mean of y

64259695 33112200

Interpretation:

- t-value: The t-value is 6.2978. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is positive, it suggests that the mean of production1 is significantly larger than the mean of production2.
- p-value: The p-value is 1.267e-06, which is extremely small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of productions of cultivation of sugarcane of the states Maharashtra and Karnataka. We can also conclude that the production of cultivation of sugarcane obtained from Maharashtra is much more than that of Karnataka.

4.Comparison between annual rainfall between Maharashtra and Karnataka.

We want to test,

Null Hypothesis (H0): The true difference in means between the annual rainfall of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the annual rainfall of two states is not equal to zero.

Let rainfall1= annual rainfall measured in sugarcane growing regions in Maharashtra.

Let rainfall2= annual rainfall measured sugarcane growing regions in Karnataka.

Unit of measurement= Millimeters.

Input:

> t.test(rainfall1,rainfall2)

Output:

Welch Two Sample t-test

data: rainfall1 and rainfall2

t = -0.17955, df = 39.015, p-value = 0.8584

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-119.6661 100.1527

sample estimates:

mean of x mean of y

1168.252 1178.009

Interpretation:

- t-value: The t-value is -0.17955. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is close to zero and negative, it suggests that there is not a significant difference between the mean rainfall of rainfall1 and rainfall2.
- p-value: The p-value is 0.8584, which is relatively large. This suggests that there is not enough evidence to reject the null hypothesis. Specifically, it indicates that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is relatively high. Therefore, we fail to reject the null hypothesis.

Conclusion:

We fail to reject the null hypothesis and cannot conclude that there is a statistically significant difference in means between the annual rainfall of Maharashtra and Karnataka. The mean rainfall of Maharashtra and Karnataka are not significantly different from each other.

5.Comparison between use of fertilizer for sugarcane of Maharashtra and Karnataka.

We want to test,

Null Hypothesis (H0): The true difference in means between the annual rainfall of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the annual rainfall of two states is not equal to zero.

Let fertilizer1= Amount of fertilizer needed for cultivation of sugarcane in Maharashtra.

Let fertilizer2= Amount of fertilizer needed for cultivation of sugarcane in Karnataka.

Unit of measurement= Kilograms.

Input:

> t.test(fertilizer1,fertilizer2)

Output:

Welch Two Sample t-test

data: fertilizer1 and fertilizer2

t = 5.3767, df = 25.653, p-value = 1.301e-05

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

36657652 82082690

sample estimates:

mean of x mean of y

111619772 52249601

Interpretation:

- t-value: The t-value is 5.3767. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is positive, it suggests that the mean of fertilizer1 is significantly larger than the mean of fertilizer2.
- p-value: The p-value is 1.301e-05, which is extremely small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of usage of fertilizers of sugarcane of the states Maharashtra and Karnataka. We can also conclude that the usage of fertilizers of sugarcane obtained from Maharashtra is much more than that of Karnataka.

6.Comparison between use of pesticides for sugarcane of Maharashtra and Karnataka.

We want to test,

Null Hypothesis (H0): The true difference in means between the use of pesticides of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the use of pesticides of two states is not equal to zero.

Let pesticide1 = Amount of pesticide used for cultivation of sugarcane in Maharashtra.

Let pesticide2= Amount of pesticide used for cultivation of sugarcane in Karnataka.

Unit of measurement= Kilograms.

Input:

```
> t.test(pesticide1,pesticide2)
```

Output:

```
        Welch Two Sample t-test

data:  pesticide1 and pesticide2
t = 4.5211, df = 27.113, p-value = 0.0001093
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 59216.55 157592.25
sample estimates:
mean of x mean of y
 214318.9  105914.5
```

Interpretation:

- t-value: The t-value is 4.5211. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples. Since the t-value is positive, it suggests that the mean of pesticide1 is significantly larger than the mean of pesticide2.
- p-value: The p-value is 0.0001093, which is very small. This indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of usage of pesticides of sugarcane of the states Maharashtra and Karnataka. We can also conclude that the usage of pesticides of sugarcane obtained from Maharashtra is much more than that of Karnataka.

Final conclusion:

The above analysis of sugarcane production of the states Maharashtra and Karnataka shows that:

Maharashtra is a greater producer of sugarcane. We can see that the area of cultivation, usage of pesticides and fertilizers is much larger for Maharashtra than Karnataka. Rainfall does not contribute to the difference seen in the production of sugarcane In two states.

Comparison of production of wheat between Maharashtra and Chhattisgarh and factors affecting it.

We performed Wilk-Shapiro test on the data given above and checked for normality for every variable. The results and interpretations are given below:

Shapiro-Wilk normality test

data: yield1

W = 0.96989, p-value = 0.7525

Shapiro-Wilk normality test

data: yield1

W = 0.96989, p-value = 0.7525

Shapiro-Wilk normality test

data: area1

W = 0.93943, p-value = 0.2339

Shapiro-Wilk normality test

data: area2

W = 0.92139, p-value = 0.1054

Shapiro-Wilk normality test

data: production1

W = 0.93882, p-value = 0.2277

Shapiro-Wilk normality test

data: production2

W = 0.94592, p-value = 0.3093

Shapiro-Wilk normality test

data: rainfall1

W = 0.90531, p-value = 0.05192

Shapiro-Wilk normality test

data: rainfall2

W = 0.94707, p-value = 0.3248

Shapiro-Wilk normality test

data: fertilizer1

W = 0.94104, p-value = 0.2508

Shapiro-Wilk normality test

data: fertilizer2

W = 0.95506, p-value = 0.4504

Shapiro-Wilk normality test

data: pesticide1

W = 0.95341, p-value = 0.421

Shapiro-Wilk normality test

data: pesticide2

W = 0.96965, p-value = 0.7475

Here we can see that for every variable, the p value is greater than the significance level 0.05. Hence, we can conclude that all the variables approximately follow Normal

distribution. Also, we are making comparisons of different variables between two states and so, they are independent of each other. That means, production of one state does not affect the production of another. Thus, we can use t-test.

We are using t-test to compare the means of different variables regarding wheat production in Maharashtra and Chhattisgarh. The reason to choose these states for testing is because as we live in Maharashtra, it felt appropriate to make the analysis on it. While the reason to choose Chhattisgarh is that, it is one of the minor wheat producing states.

1.Comparison between yields of wheat of the states Maharashtra and Chhattisgarh.

We want to test,

Null Hypothesis (H0): The true difference in means between the yield of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the yield of two states is not equal to zero.

We have used T-test for testing of hypothesis in r software.

Let yield1= yield of wheat Maharashtra.

Let yield2= yield of wheat Chhattisgarh.

Unit of measurement= Production per unit area.

Input:

```
> t.test(yield1,yield2)
```

Output:

```
        Welch Two Sample t-test

data:  yield1 and yield2

t = -17.272, df = 35.686, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -1.474529 -1.164545

sample estimates:

mean of x mean of y

 1.417155  2.736692
```

Interpretation:

- t-value: The t-value is -17.272. This value indicates the magnitude and direction of the difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as < 2.2e-16, which is very close to zero. This extremely low p-value indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is extremely low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of yields of wheat of the states Maharashtra and Chhattisgarh. We can also conclude that the yield of wheat obtained from Chhattisgarh is much more than that of Maharashtra.

We will now compare the available regressors of both the states to see which regressor contributes the most to the difference between yields of the states.

2.Comparison between areas of wheat production of Maharashtra and Chhattisgarh.

We want to test,

Null Hypothesis (H0): The true difference in means between the areas two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the areas two states is not equal to zero.

Let area1=area of cultivation of wheat of Maharashtra.

Let area2= area of cultivation wheat of Chhattisgarh.

Unit of measurement= Hectares.

Input:

```
> t.test(area1,area2)
```

Output:

```
        Welch Two Sample t-test

data:  area1 and area2

t = 19.692, df = 19.058, p-value = 3.968e-14

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
```

781890.0 967829.5

sample estimates:

mean of x mean of y

975334.0 100474.3

Interpretation:

- t-value: The t-value is 19.692. This value indicates the magnitude and direction of the difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 3.968e-14, which is extremely close to zero. This very low p-value indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is extremely low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of area of cultivation of wheat of the states Maharashtra and Chhattisgarh. We can also conclude that the area of cultivation of wheat obtained from Maharashtra is much more than that of Chattisgarh.

3.Comparison between production of wheat of Maharashtra and Chhattisgarh.

We want to test,

Null Hypothesis (H0): The true difference in means between the production of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the production of two states is not equal to zero.

Let production1= production of wheat of Maharashtra.

Let production2= production of wheat of Chhattisgarh.

Unit of measurement= Metric tons.

Input:

> t.test(production1,production2)

Output:

Welch Two Sample t-test

data: production1 and production2

t = 12.671, df = 19.114, p-value = 9.498e-11

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

1131259 1578723

sample estimates:mean of x mean of y

1475592.6 120601.8

Interpretation:

- t-value: The t-value is 12.671. This value indicates the magnitude and direction of the difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 9.498e-11, which is extremely close to zero. This very low p-value indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is extremely low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of production of wheat of the states Maharashtra and Chhattisgarh. We can also conclude that the production of wheat obtained from Maharashtra is much more than that of Chattisgarh.

4.Comparison between annual rainfall between Maharashtra and Chhattisgarh.

We want to test,

Null Hypothesis (H0): The true difference in means between the annual rainfall of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the annual rainfall of two states is not equal to zero.

Let rainfall1= annual rainfall measured in wheat growing regions in Maharashtra.

Let rainfall2= annual rainfall measured wheat growing regions in Chhattisgarh.

Unit of measurement= Millimeters.

Input:

> t.test(rainfall1,rainfall2)

Output:

Welch Two Sample t-test

data: rainfall1 and rainfall2

t = 5.5162, df = 36.836, p-value = 2.877e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

236.6686 511.5409

sample estimates:

mean of x mean of y

1168.340 794.235

Interpretation:

- t-value: The t-value is 5.5162. This value indicates the magnitude and direction of the difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 2.877e-06, which is extremely small. This very low p-value indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is extremely low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of annual rainfall recorded in areas of cultivation of wheat of the states Maharashtra and Chhattisgarh. We can also conclude that the annual rainfall recorded in areas of cultivation of wheat obtained from Maharashtra is much more than that of Chhattisgarh.

5.Comparison between use of fertilizer for wheat of Maharashtra and Chhattisgarh.

We want to test,

Null Hypothesis (H0): The true difference in means between the annual rainfall of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the annual rainfall of two states is not equal to zero.

Let fertilizer1= Amount of fertilizer needed for cultivation of wheat in Maharashtra.

Let fertilizer2= Amount of fertilizer needed for cultivation of wheat in Chhattisgarh.

Unit of measurement= Kilograms.

Input:

> t.test(fertilizer1,fertilizer2)

Output:

Welch Two Sample t-test

data: fertilizer1 and fertilizer2

t = 12.414, df = 19.214, p-value = 1.261e-10

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

102682223 144291467

sample estimates:

mean of x mean of y

137494375 14007530

Interpretation:

- t-value: The t-value is 12.414. This value indicates the magnitude and direction of the difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 1.261e-10, which is extremely small. This very low p-value indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is extremely low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of fertilizer used in cultivation of wheat of the states Maharashtra and Chhattisgarh. We can also conclude that fertilizer used in cultivation of wheat obtained from Maharashtra is much more than that of Chhattisgarh.

6.Comparison between use of pesticides for wheat of Maharashtra and Chhattisgarh.

We want to test,

Null Hypothesis (H0): The true difference in means between the use of pesticides of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the use of pesticides of two states is not equal to zero.

Let pesticide1 = Amount of pesticide used for cultivation of wheat in Maharashtra.

Let pesticide2= Amount of pesticide used for cultivation of wheat in Chhattisgarh.

Unit of measurement= Kilograms.

Input:

```
> t.test(pesticide1,pesticide2)
```

Output:

```
        Welch Two Sample t-test

data:  pesticide1 and pesticide2
t = 10.898, df = 19.327, p-value = 1.072e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 187773.5 276919.2
sample estimates:
mean of x mean of y
 259331.6   26985.2
```

Interpretation:

- t-value: The t-value is 10.898. This value indicates the magnitude and direction of the difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 1.072e-09, which is extremely small. This very low p-value indicates strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is extremely low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of pesticides used in cultivation of wheat of the states Maharashtra and Chhattisgarh. We can also conclude that pesticides used in cultivation of wheat obtained from Maharashtra is much more than that of Chhattisgarh.

Final conclusion:

The above analysis of wheat production of the states Maharashtra and Chhattisgarh shows that: The production of wheat in Maharashtra is much larger than in Chhattisgarh. Here, the total area that cultivates wheat, annual rainfall recorded in wheat growing areas, usage of fertilizers and pesticides all of these variables are higher in Maharashtra than Chhattisgarh.

Hence, naturally, the production of wheat is larger in Maharashtra.

Comparison of production of wheat between Maharashtra and Gujarat and factors affecting it.

We performed Wilk-Shapiro test on the data given above and checked for normality for every variable. The results and interpretations are given below:

Shapiro-Wilk normality test

data: yield1

W = 0.96989, p-value = 0.7525

Shapiro-Wilk normality test

data: yield1

W = 0.96989, p-value = 0.7525

Shapiro-Wilk normality test

data: area1

W = 0.93943, p-value = 0.2339

Shapiro-Wilk normality test

data: area2

W = 0.97872, p-value = 0.9165

Shapiro-Wilk normality test

data: production1

W = 0.93882, p-value = 0.2277

Shapiro-Wilk normality test

data: production2

W = 0.96727, p-value = 0.6965

Shapiro-Wilk normality test

data: rainfall1

W = 0.90531, p-value = 0.05192

Shapiro-Wilk normality test

data: rainfall2

W = 0.94707, p-value = 0.3248

Shapiro-Wilk normality test

data: fertilizer1

W = 0.94104, p-value = 0.2508

Shapiro-Wilk normality test

data: fertilizer2

W = 0.96927, p-value = 0.7394

Shapiro-Wilk normality test

data: pesticide1

W = 0.95341, p-value = 0.4219

Shapiro-Wilk normality test

data: pesticide2

W = 0.9487, p-value = 0.3478

Here we can see that for every variable, the p value is greater than the significance level 0.05. Hence, we can conclude that all the variables approximately follow Normal distribution. Also, we are making comparisons of different variables between two states and so, they are independent of each other. That means, production of one state does not affect the production of another. Thus, we can use t-test.

We are using t-test to compare the means of different variables regarding wheat production in Maharashtra and Gujarat. The reason to choose these states for testing is because as we live in Maharashtra, it felt appropriate to make the analysis on it. While the reason to choose Gujarat is that, it is one of the major wheat producing states.

1.Comparison between yields of wheat of the states Maharashtra and Gujarat.

We want to test,

Null Hypothesis (H0): The true difference in means between the yield of two states is equal to zero

Alternative Hypothesis (H1): The true difference in means between the yield of two states is not equal to zero

Let area1=area of cultivation of wheat of Maharashtra.

Let area2= area of cultivation wheat of Gujarat.

Unit of measurement= Hectares.

Input:

```
> t.test(area1,area2)
```

Output:

Welch Two Sample t-test

data: area1 and area2

t = -0.016465, df = 30.313, p-value = 0.987

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-183373.0 180438.6

sample estimates:

mean of x mean of y

975334.0 976801.2

Interpretation:

- t-value: The t-value is -0.016465. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples.
- p-value :The p-value is reported as 0.987. This value is relatively high, indicating weak evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is quite high. Therefore, we fail to reject the null hypothesis

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of yields of wheat of the states Maharashtra and Gujarat. We can also conclude that the yield of wheat obtained from Gujarat is much more than that of Maharashtra.

We will now compare the available regressors of both the states to see which regressor contributes the most to the difference between yields of the states.

2.Comparison between areas of wheat production of Maharashtra and Gujarat.

We want to test,

Null Hypothesis (H0): The true difference in means between the areas two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the areas two states is not equal to zero.

Let area1=area of cultivation of wheat of Maharashtra.

Let area2= area of cultivation wheat of Gujarat.

Unit of measurement= Hectares.

Input:

```
> t.test(area1,area2)
```

Output:

```
        Welch Two Sample t-test

data:  area1 and area2

t = -0.016465, df = 30.313, p-value = 0.987

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -183373.0  180438.6

sample estimates:

mean of x mean of y

 975334.0  976801.2
```

Interpretation:

- t-value: The t-value is -0.016465. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 0.987. This value is relatively high, indicating weak evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is quite high. Therefore, we fail to reject the null hypothesis.

Conclusion:

We fail to reject the null hypothesis and conclude that there is no significance difference between the means of areas that cultivate wheat in Maharashtra and Gujarat.

3.Comparison between production of wheat of Maharashtra and Gujarat.

We want to test,

Null Hypothesis (H0): The true difference in means between the production of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the production of two states is not equal to zero.

Let production1= production of wheat of Maharashtra.

Let production2= production of wheat of Gujarat.

Unit of measurement= Metric tons.

Input:

```
> t.test(production1,production2)
```

Output:

```
        Welch Two Sample t-test

data:  production1 and production2

t = -4.456, df = 24.77, p-value = 0.0001555

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -1896695.1  -697223.1

sample estimates:

mean of x mean of y

  1475593   2772552
```

Interpretation:

- t-value: The t-value is -4.456. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 0.0001555. This value is very low, indicating strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of production of wheat of the states Maharashtra

and Gujarat. We can also conclude that the production of wheat obtained from Gujarat is much more than that of Maharashtra.

4.Comparison between annual rainfall between Maharashtra and Gujarat.

We want to test,

Null Hypothesis (H0): The true difference in means between the annual rainfall of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the annual rainfall of two states is not equal to zero.

Let rainfall1= annual rainfall measured in wheat growing regions in Maharashtra.

Let rainfall2= annual rainfall measured wheat growing regions in Gujarat.

Unit of measurement= Millimeters.

Input:

```
> t.test(rainfall1,rainfall2)
```

Output:

```
        Welch Two Sample t-test

data:  rainfall1 and rainfall2

t = 5.5162, df = 36.836, p-value = 2.877e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 236.6686 511.5409

sample estimates:

mean of x mean of y

 1168.340   794.235
```

Interpretation:

- t-value: Th e t-value is 5.5162. This value indicates the magnitude of difference between the means of the two samples relative to the variability within the samples.

- p-value: The p-value is reported as 2.877e-06 (or approximately 0.000002877). This value is very low, indicating strong evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is very low. Therefore, we reject the null hypothesis.

Conclusion:

We do not have enough evidence to accept the null hypothesis and conclude that there is a significant difference in the means of productions of cultivation of wheat of the states Maharashtra and Gujarat. We can also conclude that the annual rainfall recorded for wheat crop in Maharashtra is much more than that in Gujarat.

5.Comparison between use of fertilizer for wheat of Maharashtra and Gujarat.

We want to test,

Null Hypothesis (H0): The true difference in means between the annual rainfall of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the annual rainfall of two states is not equal to zero.

Let fertilizer1= Amount of fertilizer needed for cultivation of wheat in Maharashtra.

Let fertilizer2= Amount of fertilizer needed for cultivation of wheat in Gujarat.

Unit of measurement= Kilograms.

Input:

```
> t.test(fertilizer1,fertilizer2)
```

Output:

```
Welch Two Sample t-test

data:  fertilizer1 and fertilizer2

t = -0.22048, df = 33.447, p-value = 0.8268

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -39803418  32016400
```

sample estimates:

mean of x mean of y

137494375 141387885

Interpretation:

- t-value: The t-value is -0.22048. This value indicates the magnitude and direction of the difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 0.8268. This value is relatively high, indicating weak evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is relatively high. Therefore, we fail to reject the null hypothesis.

Conclusion:

We fail to reject the null hypothesis and conclude that there is no significance difference between the means of fertilizers used to that cultivate wheat in Maharashtra and Gujarat.

6.Comparison between use of pesticides for wheat of Maharashtra and Gujarat.

We want to test,

Null Hypothesis (H0): The true difference in means between the use of pesticides of two states is equal to zero.

Alternative Hypothesis (H1): The true difference in means between the use of pesticides of two states is not equal to zero.

Let pesticide1 = Amount of pesticide used for cultivation of wheat in Maharashtra.

Let pesticide2= Amount of pesticide used for cultivation of wheat in Gujarat.

Unit of measurement= Kilograms.

Input:

> t.test(pesticide1,pesticide2)

Output:

Welch Two Sample t-test

data: pesticide1 and pesticide2

t = -0.099194, df = 34.756, p-value = 0.9216

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -76728.52 69581.37

sample estimates:

mean of x mean of y

 259331.6 262905.1

Interpretation:

- t-value: The t-value is -0.099194. This value indicates the magnitude and direction of the difference between the means of the two samples relative to the variability within the samples.
- p-value: The p-value is reported as 0.9216. This value is relatively high, indicating weak evidence against the null hypothesis. Specifically, it suggests that the probability of observing such extreme results (or more extreme) under the assumption that there is no difference in means between the two populations is relatively high. Therefore, we fail to reject the null hypothesis.

Conclusion:

We fail to reject the null hypothesis and conclude that there is no significance difference between the means of pesticides used to that cultivate wheat in Maharashtra and Gujarat.

Final conclusion:

The above analysis of wheat production of the states Maharashtra and Gujarat shows that:

The yield of wheat produced in Gujarat is much larger than that of Maharashtra. We can see from the above analysis that the production of wheat in Gujarat is higher than Maharashtra. However, the annual rainfall recorded in Maharashtra is higher than Gujarat. Area of cultivation, usage of fertilizers and pesticides do not contribute to the difference in yields between the two states.

Conclusion:

By comparing production of two major crops produced in India in different states we identified the the variable that can have big or small impact on the production of these crops.

Used Tools:

Following tolls were used during the analysys:

1) MS - Excel
2) MS - Word
3) R – studio

References:

https://www.statista.com/statistics/1365756/india-wheat-production-by-leading-state/

https://www.hindustantimes.com/cities/chandigarh-news/grain-gain-punjab-beats-mp-regains-top-wheat-producing-state-tag-101622230045277.html