Abhishek Saseendran: 01FB16ECS018

Akash Bhat: 01FB16ECS038

HYPOTHESIS TESTING

IDS ASSIGNMENT-2

THIS ASSIGNMENT IS BEING MADE WITH THE PURPOSE OF CONDUCTING A HYPOTHESIS TEST ON THE SAMPLE DATA AND COME TO A CONCLUSION

PROBLEM

WE HAVE CONDUCTED A HYPOTHESIS TESTING ON THE PARAMETER -MEAN RICE YEILD OF THE TWO SAMPLE POPULATIONS AND CONCLUDED WHICH ONE IS HIGHER.

THE PARAMETER: MEAN PRODUCTION OF RICE IN KOPPAL AND MEAN PRODUCTION OF RICE IN DHARWAD

(Values referred from previous assignment)

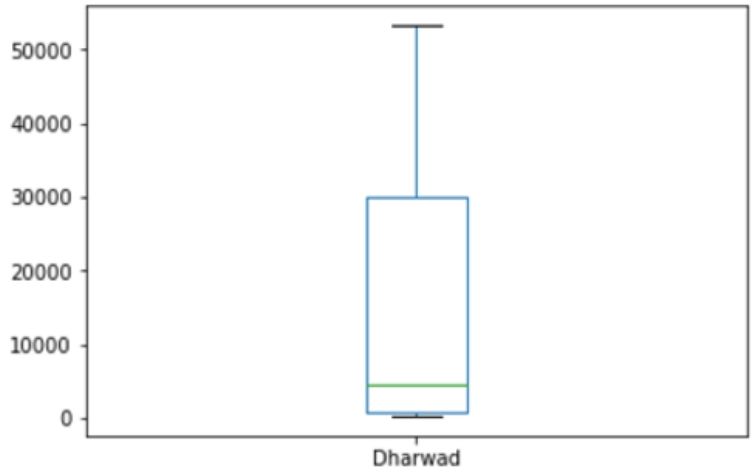
SAMPLE DATA

(from previous)

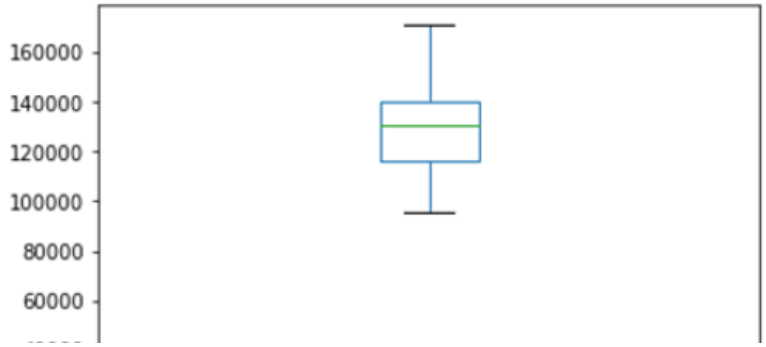
Production of Rice(in kg)			
Year	Season	DHARWAD	KOPPAL
1997	Kharif	103406	119948
1997	Rabi	30	113461
1997	Summer	6181	131751
1998	Kharif	33695	128590
1998	Rabi	3	124532
1998	Summer	2032	129035
1999	Kharif	37985	135034
1999	Summer	2594	117413
2000	Kharif	53225	154793
2000	Summer	2701	13
2001	Kharif	8323	13687
2001	Summer	1008	153179
2002	Kharif	7557	26966
2002	Summer	656	125369.02
2003	Kharif	6476	114229
2003	Summer	878	134083
2004	Kharif	15484.88	135541
2004	Summer	899	141720

2005	Kharif	39091	98424
2005	Summer	905	145069
2006	Kharif	32376	136130
2006	Summer	397	149430.06
2007	Kharif	44801	114654
2007	Summer	791	114920.38
2008	Kharif	27647	130501
2008	Summer	408	170981
2009	Kharif	30669.81	152810
2009	Summer	846	161960
2010	Kharif	33658	115410
2010	Summer	883	131451
2011	Kharif	25790	133102
2011	Summer	663	134099
2012	Kharif	24009	128944
2012	Summer	305	145559
2013	Kharif	25080	95208

```
No. of years:(Sample size)  17.0
----Dharwad Rice Production----
Min production : 301.0  kg
Max production :  53225.0  kg
Mean production: 14561.755588235294  kg
Median production: 4588.5  kg
Quartiles : [   301.        854.       4588.5     29914.1075]
```




```
----Koppal Rice Production----
Min production : 13687.0  kg
Max production :  170981.0  kg
Mean production: 125234.80764705883  kg
Median production: 130976.0  kg
Quartiles : [  13687.    115910.75  130976.    140322.5 ]
```



STATING THE HYPOTHESES

Null Hypothesis:

H_o: $\mu_x - \mu_y >= 0$

Alternative Hypothesis:

H_a: $\mu_x - \mu_y < 0$

μ_x =Mean production of rice in Dharwad

μ_y =Mean production of rice in Koppal

We assume the production of Koppal is lesser than that of Dharwad

Parameters used for the hypothesis testing:

We have the two Data Sets X and Y corresponding to Dharwad and Koppal respectively.

Mean production in Dharwad(μ_x) : 14561.755 kg

Mean production in Koppal(μ_y) : 125234.80 kg

Standard deviation(SD_X): 16504.18 kg

Standard deviation(SD_Y): 31426.88 kg

Number of samples in the population(N_X): 17

Number of samples in the population(N_Y): 17

Let us assume the test is being conducted under 95% confidence

The difference $\mu_x - \mu_y$ follows normal distribution.

The Null distribution is given by:

$\sim(\mu, \sigma^2)$

$\sim (\mu_x - \mu_y, SD_x^2/n_x + SD_y^2/n_y)$

$\sim (0 , 16504.18^2/17 + 31426.88^2/17)$

Also from the data,

We can calculate the t statistic as the sample size, $n < 30$

Degree of freedom $=n-1=17-1=16$

$t= (\mu_x - \mu_y - 0)/(\sigma /sqrt(n))$

$t= (\mu_x - \mu_y - 0)/(sqrt(16504.18^2/17 + 31426.88^2/17)/sqrt(17))$

t=-3.1178 , we take mod of the value, 3.1178

And corresponding P value is close to 0.001, which is less than 0.05 and approaches 0.

Conclusion:

Hence, we can reject the NULL Hypothesis,

and assume Alternate Hypothesis is true.

Hence we conclude that $\mu_x - \mu_y < 0$

That is, the mean production of rice is greater in Koppal than in Dharwad.