INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

CONTROL OF OVERTIME HOURS IN A PAPER MILL BY FIXING ITS LIMITS

VISHNU M
23M0025

PROJECT REPORT
May 2024

CONTROL OF OVERTIME HOURS IN A PAPER MILL BY FIXING ITS LIMITS

Objective:

There are two plants at two different locations. For the sake of data security we call them plant A and plant B.

OT is the time an employee is asked to work beyond scheduled office hours and for this OT the person is paid additional amount which is proportional to the hours of OT.

The data at our hand is department-wise monthly summary of OT for four years for both the plants along with information on a few explanatory variables like absenteeism and downtime. It is expected that higher the absenteeism and downtime i.e. the hours the paper machine remains idle either due unexpected failure (resulting in unplanned downtime) or planned downtime (for repair and maintenance).
It will be obvious from the data that the OT/head or person is much higher at plant B than that at A.
The justifications given for higher OT at plant B are higher absenteeism and downtime.
The question is that **do these two variables fully explain the difference between the two. If not, what should be the permitted level of OT at plant B to bring both the plants at the same level? Hence we have to fix a limit of OT hours for plant B.**

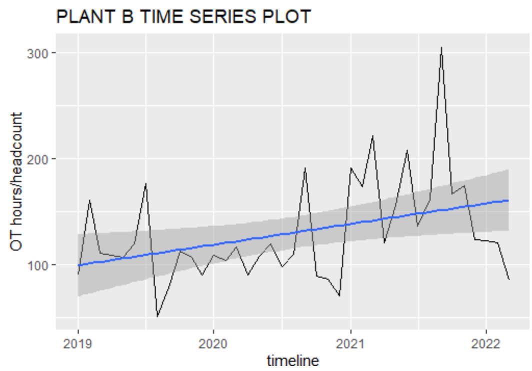
The above question is just one of the many questions that needed to be answered for **reducing OT at plant B**.

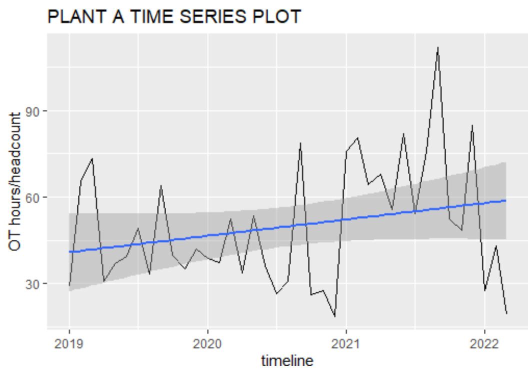
DATA CLEANING AND SORTING:

Department wise monthly data for four years is given. The data for both the plants is brought in the same format using MS Excel by following the steps given below:

1) Deleting the data of those departments which are not common to both the plants.
2) Clubbing similar departments together such as clubbing three electrical to get one department.
   NOTE: The data corresponding to columns like downtime and total number of man working days can be computed by direct summing. But for columns like absentees percentage, direct summing up should not be done. Absent % should then be computed separately from the no.of man absent days and total man working days.

TREND ANALYSIS:





COMMENTS: 1) Overall range of OT hours per headcount for plant B is much higher than that of plant A.

2) Both plots have a slight increasing trend.
3) There seems to be some complicated sort of seasonality.

ANALYSIS OF SEAONAL VARIATION:

OT hours for PLANT B:

	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
2019	88.56046	114.67286	123.10746	109.72839
2020	102.06309	132.41237	97.48305	97.17118
2021	145.29202	210.35708	167.76648	187.25191
2022	108.98108			

OT hours for PLANT A:

	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
2019	33.6314	47.4287	48.8837	47.6438
2020	36.7429	46.0429	29.7410	35.3766
2021	66.4081	80.5993	66.2299	55.0225
2022	29.9448			

Observation: OT hours in second quarter is significantly higher for both the plants.
Possible Explanation: The two plants are Dubai based. During Ramadan time (April) there is a norm in Dubai to have 6 hours of work per person instead of 8. So these 2 hours of lost working time is compensated in the form of OT hours.

SETTING UP A MODEL FOR PLANT B:

Y: response is OT hours/headcount corresponding to a department for a month.

Explanatory variables:

X1= planned downtime hours in that dept for that month

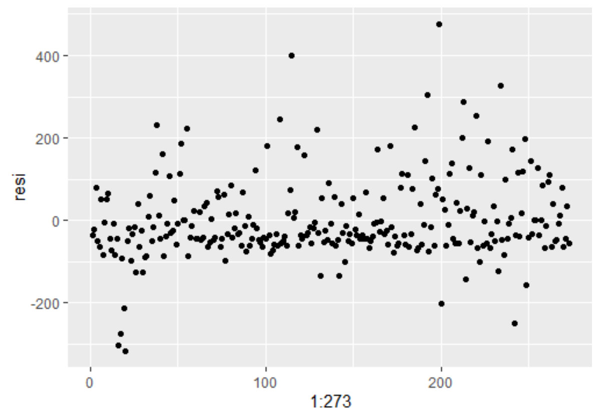
X2= unplanned downtime hours in that dept for that month

X3=percent of absent in a dept in a month.

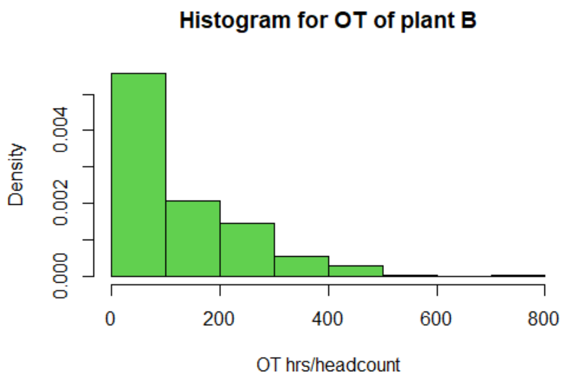
Calculated by—(no.of man days of absenteeism)*100/ (total no.of working man days)

On plotting the residual plot we observe that it is not at all random.

Also, there exists seasonality and some increasing trend hence years and months should also be included as explanatory variables.



Now we also plot the histogram of OT hours and observe that it is highly skewed.



So we make a log transformation (none of the responses are zero).Hence we fit the log linear model of response on the above mentioned variables.

For months variable, dummy variable created taking April as baseline.

Observation:

Significant variables: planned downtime, unplanned downtime, absent percentage, year.

R^2 value: 0.35

SETTING UP A MODEL FOR PLANT A:

Y: response is OT hours/headcount corresponding to a department for a month.

Explanatory variables:

X1= planned downtime hours in that dept for that month

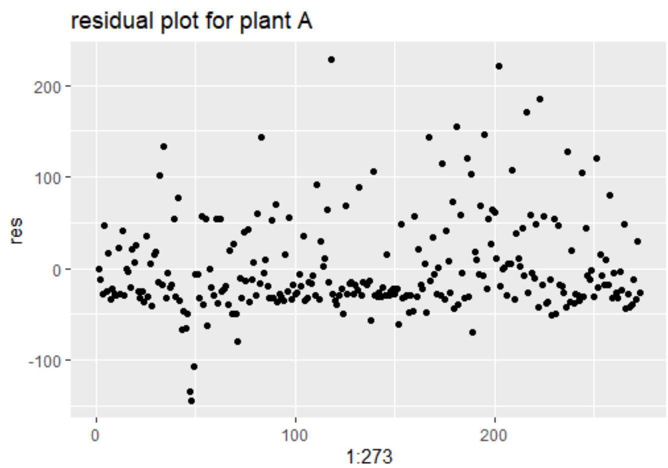
X2= unplanned downtime hours in that dept for that month

X3=percent of absent in a dept in a month.

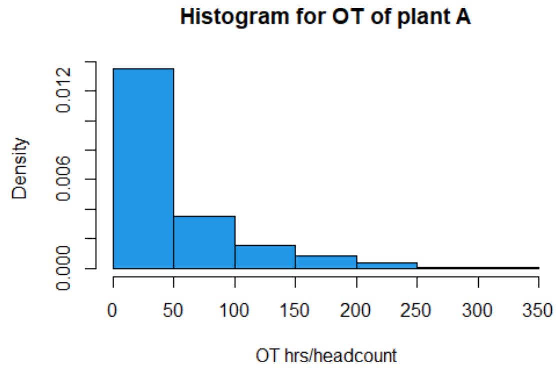
Calculated by—(no.of man days of absenteeism)*100/ (total no.of working man days)

On plotting the residual plot we observe that it is not at all random.

Also, there exists seasonality and some increasing trend hence years and months should also be included as explanatory variables.



In similar way as in plant B we fit a log linear model to plant A since histogram of OT hours/headcount is highly positively skewed here as well.



Now we make the following observations:

NOTE: In this case 7 of the response values(out of 273) are zero so we remove the data corresponding to these 7 values and then fit the log linear model

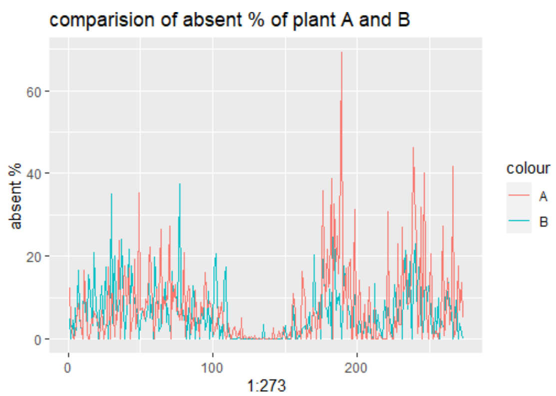
1) For months variable, dummy variable created taking April as baseline.

Significant variables: planned downtime, unplanned downtime, year, month of may.

R^2 value: 0.1975.

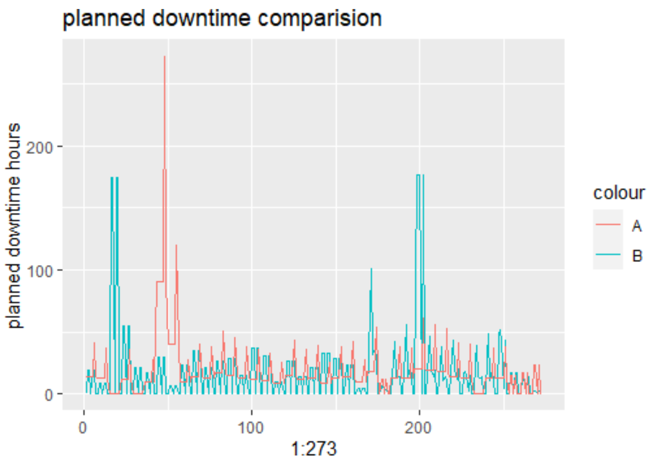
INDIVIDUAL VARIABLE ANALYSIS:

1) Absent percentage:



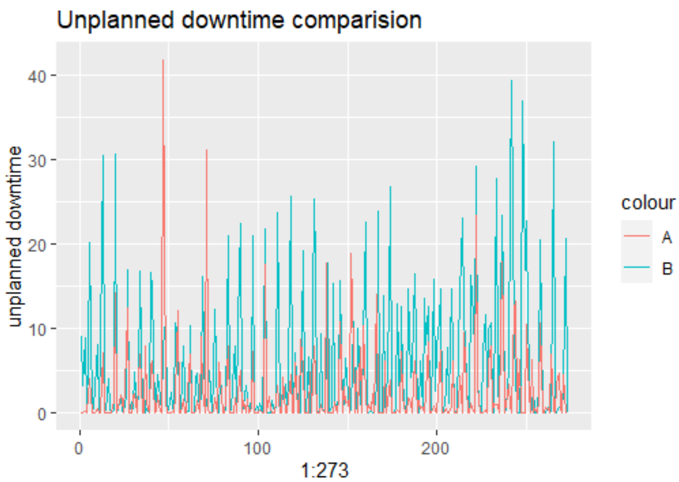
Observation: From the graph it is clear that plant A has higher absenteeism on an average. But OT hours/head count is higher for plant B. Hence higher OT hours/headcount in plant B is not due to higher absenteeism.

2) Planned downtime:



Observation: Except at certain points, on an average planned downtime of both the plants is more or less same.

3) Unplanned downtime:



Observation: Unplanned downtime for plant B seems to be higher.

Now the question is that if the entire difference in OT hours is due to difference in unplanned downtime.

[*Footnote: Interpretation of log-linear model—*
*log(Y)=α+β*x+€*

- *Here for 1 unit increase in X there should be an expected increase of log(Y) by β units.*
- *Each 1-unit increase in X multiplies the expected value of Y by e^ β.]*

1) Final regression equation for plant B—
log(Y)=b_o+ $b_1 * planned\ downtime$+$b_2 * unplanned\ downtime$+$b_3 * absent\ \%$+ month effect+ year effect
where b_1=1.129e-02 b_2=6.005e-02 b_3=3.009e-02
2) Difference in average unplanned downtime hours between plant A and plant B is 3.1078 hours/month.

3) Thus the expected difference in Y due to difference in unplanned downtime hours is 3.1078*$e^{6.005\mathrm{e}-02}$ = 3.3 hours/month .(For 1 unit increase in unplanned downtime there is $e^{6.005\mathrm{e}-02}$ units increase in Y)
4) But the average observed difference in OT hours is 80.178 hours/month.
5) Hence only 4.1158% of observed difference is explained by higher unplanned downtime.

Second question was that, what target of OT should plant B set to come to the same level as plant A.

1) Difference in absentees percentage between plant A and B is 1.04%, plant A having higher absentee percentage.
2) Difference in average unplanned downtime hours between plant A and plant B is 3.1078 hours/month.
3) Difference in average unplanned downtime hours between plant A and plant B is 0.259 hours/month.

Expected difference in OT hours is thus calculated as—

-1.04*$e^{3.009\mathrm{e}-02}$+3.1078*$e^{6.005\mathrm{e}-02}$ + 0.259*$e^{1.129\mathrm{e}-02}$ = 2.49 hours/month.

Hence plant B should fix their limit of OT hours as (49.42+2.49) hours/month= 51.91 hours/month.

Here mean OT hours/headcount for plant A is 49.42 hours/month.