

Verification of Cause

1 Question

The objective of this investigation is to verify that our main suspect is our dominant cause. Last investigation, we found evidence that x12 is the dominant cause therefore we perform our investigation on x12.

2 Plan

To verify x12 is our dominant cause we will perform a verification experiment. The observed range of x12 is $[-7.5, 19.6]$ and we will choose the number of replication of each factor combination to be 6 which means the number of runs will be 12. The total cost of this investigation will be \$492.

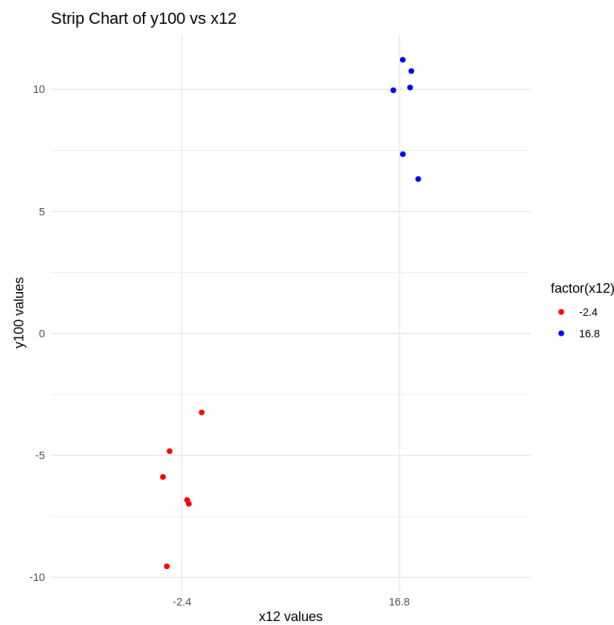
3 Data

Below is a view of the dataset

| daycount | shift | partnum | y100 | x12 |
|----------|-------|---------|------|------|
| 30 | 3 | 42721 | -7.0 | -2.4 |
| 30 | 3 | 42722 | 6.4 | 16.8 |
| 30 | 3 | 42723 | -9.6 | -2.4 |
| 30 | 3 | 42724 | 10.8 | 16.8 |
| 30 | 3 | 42725 | -5.8 | -2.4 |
| 30 | 3 | 42726 | 10.0 | 16.8 |

4 Analysis

4.1 Scatterplots



The chart indicates a potential correlation between y100 and x12, as evidenced by a marked shift in the distribution of y100 values in response to the changes in x12. Specifically, y100 values transition from negative to positive as x12 increases from -2.4 to 16.8, suggesting a positive association between these two variables.

4.2 Linear Model

```
Call:
lm(formula = y100 ~ 0 + x12, data = verification_data)

Residuals:
    Min       1Q   Median       3Q      Max
-8.174 -4.624 -2.978  0.018  1.218

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
x12    0.59417    0.09787   6.071 8.07e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.068 on 11 degrees of freedom
Multiple R-squared:  0.7701,    Adjusted R-squared:  0.7492
F-statistic: 36.86 on 1 and 11 DF,  p-value: 8.065e-05
```

Figure 1: Linear Model

The linear regression analysis demonstrates a significant positive relationship between 'y100' and 'x12'. With every unit increase in 'x12', 'y100' increases by an estimated 0.59417 units. The model explains about 77% of the variance in 'y100', indicating a good fit. Statistical tests confirm the relationship's significance, with the predictor 'x12' substantially contributing to the model, as evidenced by a low p-value and a high F-statistic.

5 Conclusion

The scatter plot and linear model both indicate a significant relationship between y_{100} and x_{12} , implying that x_{12} is indeed a dominant cause of variation.