Film Data Analytics

Case – Why stretch values?

The ability to predict stretch values of stretch film is valuable for production with processing procedures. As our products are stretch film, stretch data are very important for improving the efficiency of production. In order to build a model to predict stretch values, processing factors that might affect stretch should be known.

What factors affect values of stretch?

At first, the processing factors which have possibilities to influence stretch values based on production experience are chosen to do the predictions.

Potential factors:

- Line speed
- Output
- Vacuum 1
- Vacuum 2
- Temperature 1
- Temperature 2

After analyzing correlations of these factors, we found out line speed and output have strong relationship and values in V2, T1, and T2 didn't change much. Therefore, we use output and V1 to be our predict factors first.

```
> summary(summer_lm)

Call:
lm(formula = stretch ~ output + V1, data = summer)

Residuals:
    Min      1Q  Median      3Q     Max
-29.764 -14.034  -7.381   6.768 168.499

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 640.89750  225.10206   2.847  0.00724 **
output       -0.07643    0.06597  -1.159  0.25428
V1           -0.36157    0.96279  -0.376  0.70946
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31.56 on 36 degrees of freedom
Multiple R-squared:  0.0459,    Adjusted R-squared:  -0.007107
F-statistic: 0.8659 on 2 and 36 DF,  p-value: 0.4292
```

Since p-value in this table is high that we cannot use this model to do prediction, we decided to remove output values and got the table below.

```
Call:
lm(formula = stretch ~ V1, data = summer)

Residuals:
    Min      1Q  Median      3Q     Max
-19.645  -8.501   1.199   7.543  25.265

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 446.3289    24.4506  18.254  < 2e-16 ***
V1           -1.8443     0.4769  -3.867  0.00049 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.2 on 33 degrees of freedom
Multiple R-squared:  0.3119,    Adjusted R-squared:  0.291
F-statistic: 14.96 on 1 and 33 DF,  p-value: 0.0004897
```

This summary shows low enough p value and the estimate of intercept is more reasonable. Than we use this model to predict the value of stretch percent base on the primary vacuum data.

How can we predict stretch values?

The model here uses vacuum 1 as selected factor. After developing the model, only two values show +/- 20 in residuals which means this model is good to use for prediction.