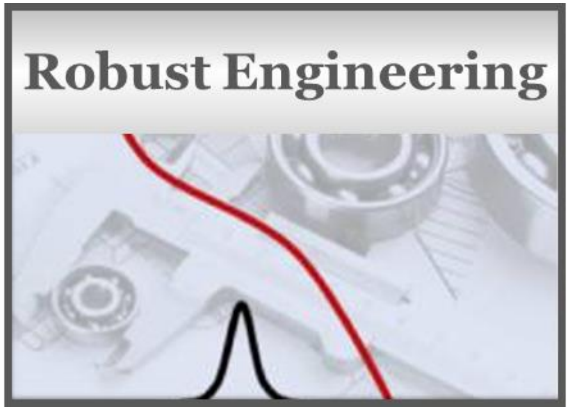
UNIVERSITY of BRADFORD

Statistiques Appliquées Formation Industrielle



Session 10: Analysing A Multi-Level Experiment

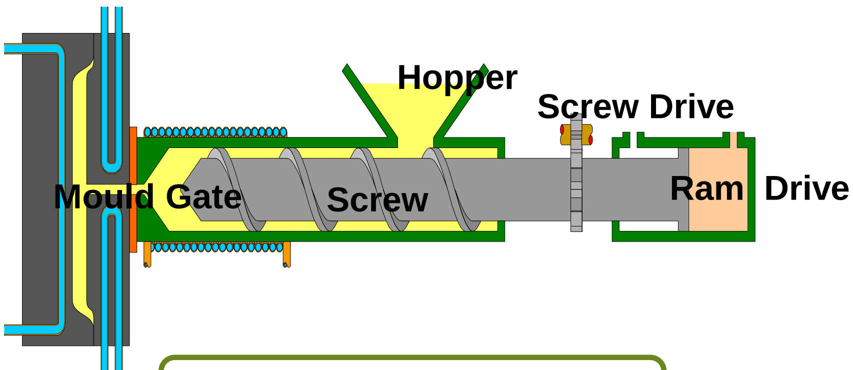
In This Session We Will…

- Use an example to illustrate the design and analysis of a three-level experiment.
- Compare two possible designs, using various measures of design quality.
- Examine a plot of the response surface and compare it with the corresponding plot from a two-level experiment.
- Look at the regression output (ANOVA table, R^2, residual plots and PRESS RMSE).
- Explain that PRESS RMSE is often misleading in small experiments.
- Discuss the use of t tests to remove terms from the equation
  - and suggest that if t tests are used, the threshold p-value should be relatively high (e.g. 15%).

Continuing The Bumper Bracket Moulding Story

- To limit the size of the follow-up experiment, it was decided to include only 4 factors.
- Based on the results of the screening experiment, factor x_4 (Holding time) was held constant; the remaining factors had the same –1/+1 levels as before, with a third level half-way between.

Factor	Name	Units	–1	0	+1
x_1	Melt temp	°C	230	250	270
x_2	Screw speed	rpm	50	175	300
x_3	Holding pressure	MPa	50	150	250
x_5	Injection rate	mm/sec	50	175	300



e.g. we could use 150

The 'middle' levels don't have to be exactly halfway

Coded –0.2

Continuing The Bumper Bracket Moulding Story

- The experimenters considered the following designs:
  - a face-centred Central Composite with 2 centre runs (26 runs in total).
  - a D-optimal design, also with 26 runs.
- For each design they looked at;
  - (relative) standard errors of the quadratic coefficients,
  - (relative) prediction standard deviation,
  - orthogonality measures and leverages.

Quadratic coefficients always have the highest SEs

Comparing The Designs

Relative SEs

Coefficient	CC design	D-optimal
b_{11}	0.62	0.57
b_{22}	0.62	0.56
b_{33}	0.62	0.51
b_{55}	0.62	0.57

The D-optimal design gives lower SEs

PSD

	CC design	D-optimal
Maximum PSD	0.88	0.93

The CC design does a little better on this criterion

Their properties are similar, so the experimenters chose the standard CC design

Leverages

	CC design	D-optimal
Highest leverage	0.66	0.77

The CC design does better on this criterion

Orthogonality

	CC design	D-optimal
Highest VIF	1.09	1.03

Both of these are acceptable

The Follow-up Experiment

Melt temp	Screw speed	Holding pressure	Injection rate	Tensile strength (MPa)
–1	–1	–1	–1	38.62
+1	–1	–1	–1	37.32
–1	+1	–1	–1	38.62
+1	+1	–1	–1	37.54
–1	–1	+1	–1	41.35
+1	–1	+1	–1	40.30
–1	+1	+1	–1	41.36
+1	+1	+1	–1	40.82
–1	–1	–1	+1	40.62
+1	–1	–1	+1	41.32
–1	+1	–1	+1	40.61
+1	+1	–1	+1	42.23
–1	–1	+1	+1	42.45
+1	–1	+1	+1	42.84
–1	+1	+1	+1	42.66
+1	+1	+1	+1	42.83
–1	0	0	0	41.31
+1	0	0	0	41.31
0	–1	0	0	41.58
0	+1	0	0	41.85
0	0	–1	0	40.12
0	0	+1	0	41.83
0	0	0	–1	40.16
0	0	0	+1	42.07
0	0	0	0	41.20
0	0	0	0	41.44

The fitted response surface

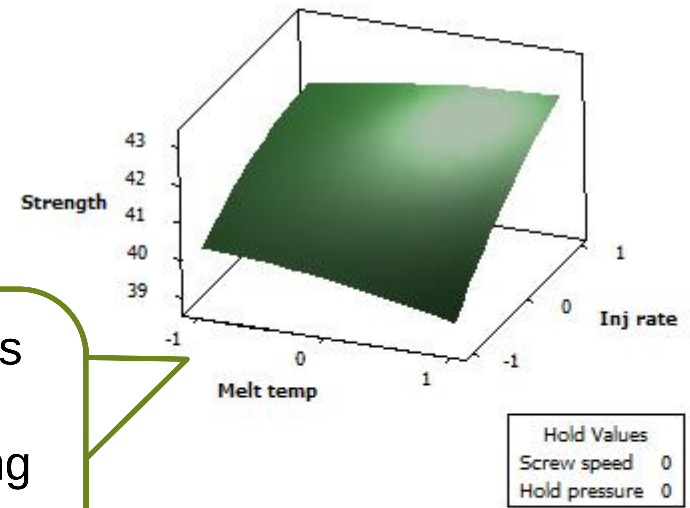
b_0	41.42
b_1	-0.06
b_2	0.12
b_3	1.08
b_5	1.20
b_{11}	-0.14
b_{22}	0.26
b_{33}	-0.48
b_{55}	-0.34
b_{12}	0.09
b_{13}	-0.06
b_{15}	0.43
b_{23}	-0.02
b_{25}	0.02
b_{35}	-0.36

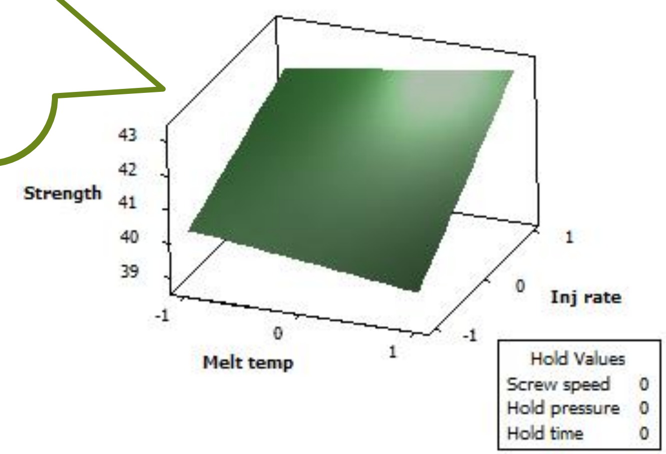
Linear Terms

Quadratic Terms

Interaction Terms

Compare this with the corresponding plot from the screening experiment





ANOVA and R^2

Source of variation	Sum of squares	Degrees of freedom
Regression (RegSS)	54.848	14
Residual (ResSS)	0.955	11
Total (SSy)	55.805	25

terms less the intercept

$n - \# \, terms$

$n - 1$

$$R^2 = \frac{\text{RegSS}}{\text{SSy}} = \frac{54.848}{55.805} = 0.983$$

This is the simplest form of ANOVA when applying multiple regression

Colab OLS produces a much more elaborate table

As explained in the tutorial, the expanded table is of limited value

Using The Residuals To Check Our Assumptions

- We fitted the response surface by least squares.
- This method is most effective when the random variation in the data:
  - has a roughly Gaussian distribution,
  - has a constant standard deviation σ that does not change as the factor levels are changed.

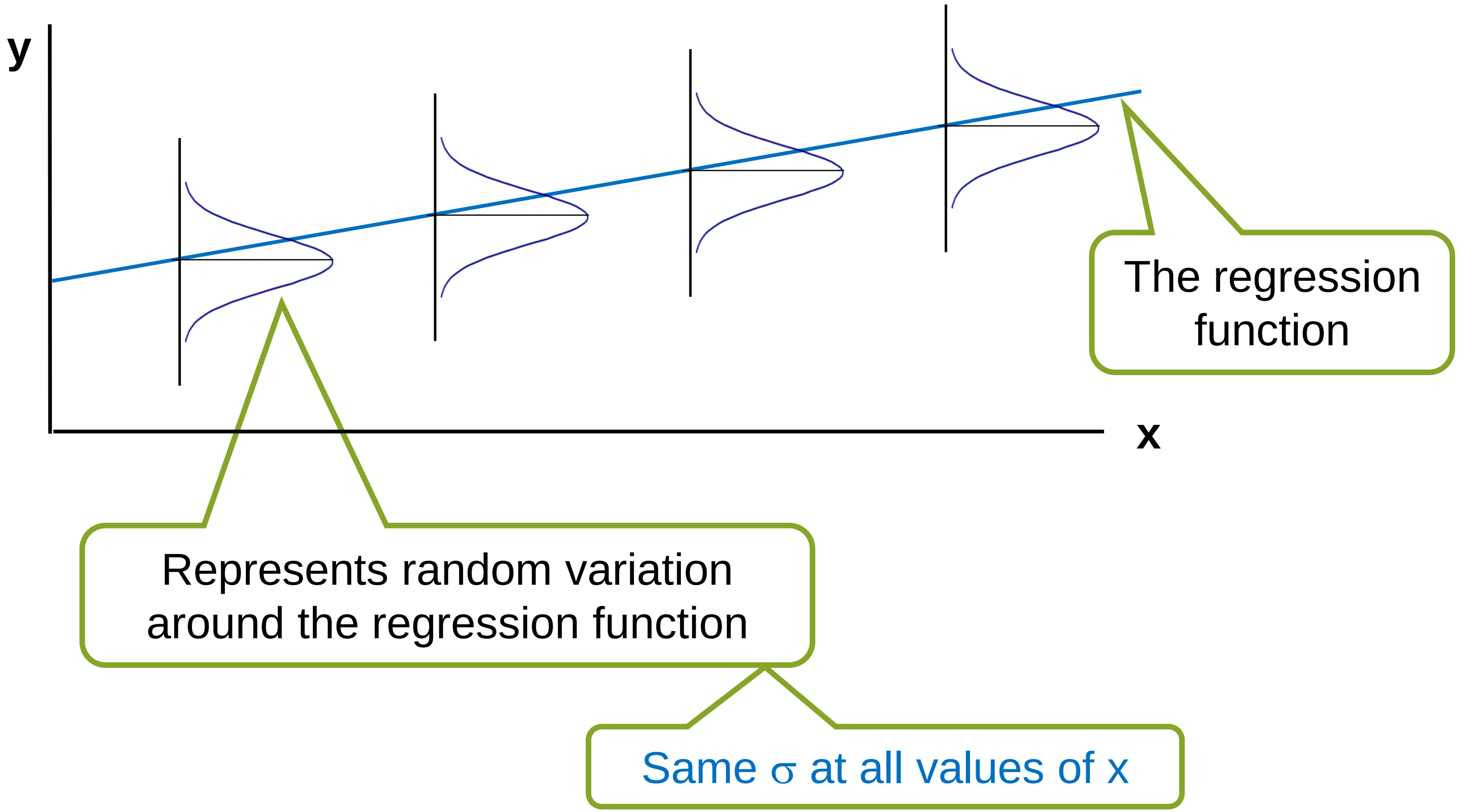
The constant σ assumption

- These assumptions can be checked by plotting the scaled residuals in various ways.

Scaled residuals are dimensionless

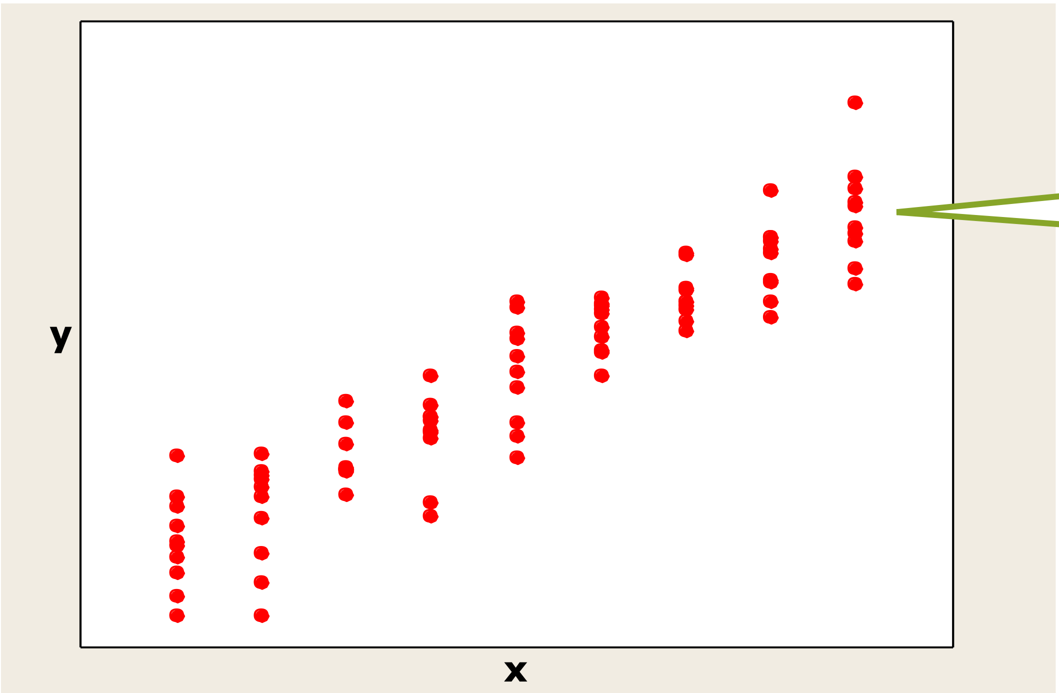
Also referred to as deleted residuals

The Constant 'σ' Assumption



Checking The Constant σ Assumption

- When $\mu(x)$ is linear, a large set of data from this model might look like this:

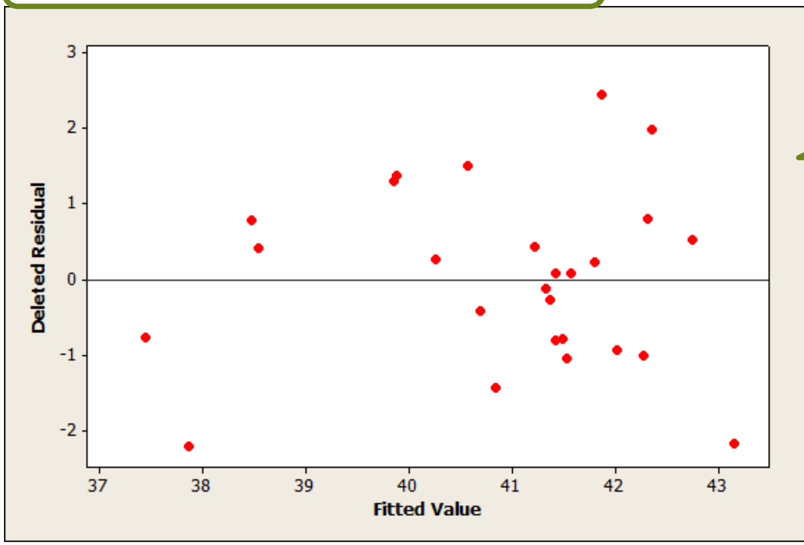


Variation is roughly constant across the x-values

In smaller data sets it is better to plot residuals

Plots of The Deleted Residuals

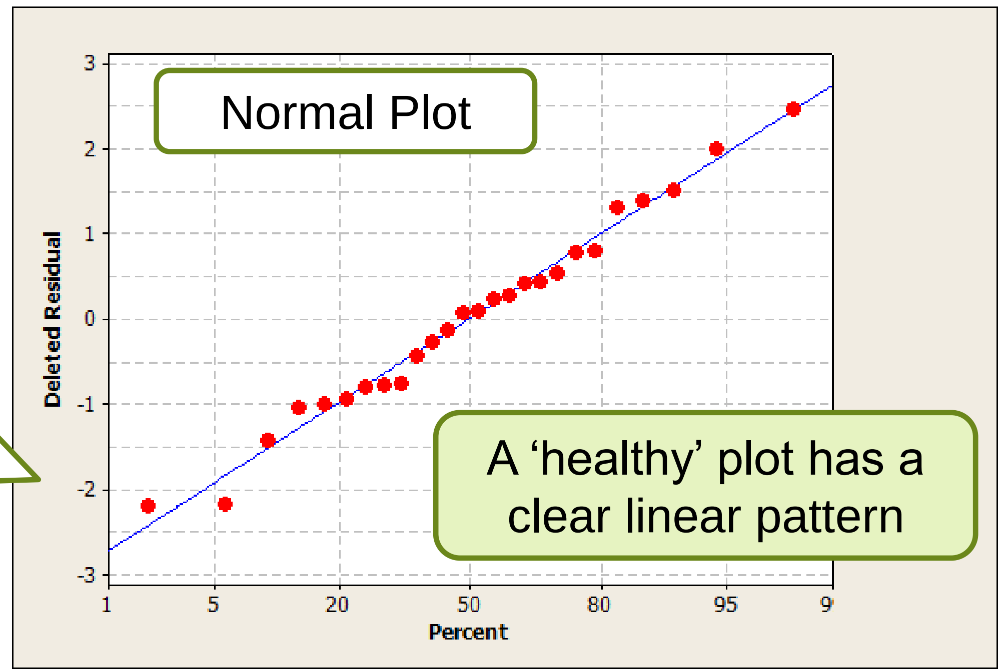
Plot against fitted values



A 'healthy' plot has a random scatter above and below zero, with constant amplitude

This plot is not ideal ('hole' at top left) but doesn't show any trends

Normal Plot



A 'healthy' plot has a clear linear pattern

This plot is used to check the assumption that random part of the response has normal distribution

PRESS RMSE

- PRESS = sum of squares of leave-one-out prediction errors = 6.858.

$$PRESS\ RMSE = \sqrt{\frac{6.858}{26}} = 0.51\ MPa$$

- Caution: in small experiments, where , $\frac{\#\text{ terms fitted}}{\#\text{ runs}} > \frac{1}{2}$, PRESS RMSE is likely to give an unduly pessimistic idea of future prediction errors

In the example, $\frac{\#\text{ terms fitted}}{\#\text{ runs}} = \frac{15}{26}$

…. so we must use PRESS RMSE cautiously

A ‘golden rule’ for a good multi-level experiment: aim to have
$\#\,runs > 2 \times (\#\,terms\ you\ want\ to\ fit)$

Removing Terms From The Response Surface Equation

- In this example we have enough degrees of freedom to apply a t test to each regression coefficient.
  - if the coefficient is not statistically significant because the p-value is too large, we could remove this term from the equation.
- This appears to be a sensible idea because if an effect is really zero (e.g. the quadratic effect of screw speed) the regression coefficient is a random number.
  - we would rather not include random numbers in our predictions!
- Removing terms also increases ResDF.
  - and a simpler equation may be easier to interpret and compare with other equations.

Removing Terms From The Equation (Cont.)

- Unfortunately, statistical theory shows that in some scenarios this procedure may increase the prediction errors (because of the chance that real effects are not detected – Type II error).
- However, using a high threshold for the p-values (e.g. 0.15 or 15%) limits the potential increase in prediction error.
- Note: we may choose to eliminate some terms based on engineering arguments.
  - e.g. a positive quadratic coefficient may not make engineering sense.

t tests For The Injection Moulding Data

Coefficient	Value	Standard error	t-ratio	p-value
b_1	-0.061	0.069	-0.87	0.40
b_2	0.118	0.069	1.70	0.12
b_3	1.080	0.069	15.55	0.00
b_5	1.197	0.069	17.23	0.00
b_{11}	-0.143	0.184	-0.78	0.45
b_{22}	0.262	0.184	1.42	0.18
b_{33}	-0.478	0.184	-2.60	0.02
b_{55}	-0.338	0.184	-1.84	0.09
b_{12}	0.089	0.074	1.21	0.25
b_{13}	-0.061	0.074	-0.82	0.43
b_{15}	0.428	0.074	5.81	0.00
b_{23}	-0.024	0.074	-0.33	0.75
b_{25}	0.022	0.074	0.30	0.77
b_{35}	-0.358	0.074	-4.86	0.00

b_{25} has the largest p-value

As we saw in the sintering example, if b_{25} is removed, the other p-values will change

…. but not by much, because the CC design is 'nearly orthogonal'

Removing Terms One By One (‘Backward Elimination’)

b_1	0.40	0.38	0.36	0.36	0.35	0.36	0.37
b_2	0.12	0.10	0.09	0.09	0.08	0.09	0.09
b_3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{11}	0.45	0.43	0.42				
b_{22}	0.18	0.16	0.15	0.19	0.19	0.20	
b_{33}	0.02	0.02	0.02	0.01	0.01	0.01	0.01
b_{55}	0.09	0.08	0.07	0.04	0.03	0.04	0.08
b_{12}	0.25	0.23	0.21	0.21	0.20		
b_{13}	0.43	0.41	0.39	0.39			
b_{15}	0.00	0.00	0.00	0.00	0.00	0.00	0.00
b_{23}	0.75	0.74					
b_{25}	0.77						
b_{35}	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Reading left to right, we eliminate the term with the biggest p-value in the column

After 6 steps, all p-values are < 0.15 except for b_1

…. and we wouldn’t usually eliminate a linear term

This process can be automated (often called stepwise regression)

In This Session We Have…

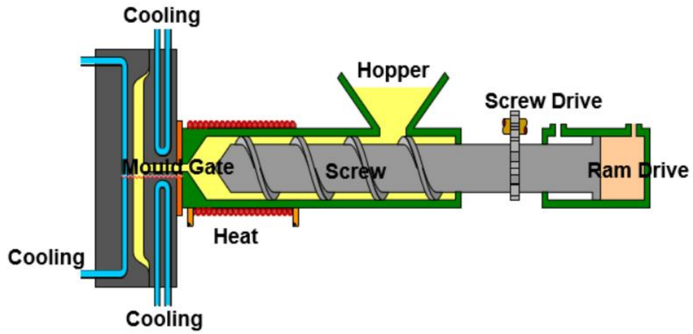
- Compared two possible designs, using various measures of design quality.
- Examined a plot of the response surface and compared it with the corresponding plot from a two-level experiment.
- Looked at the regression output (ANOVA table, R^2, residual plots and PRESS RMSE).
- Explained that PRESS RMSE is often misleading in small experiments.
- Discussed the use of t tests to remove terms from the equation
    - and suggested that if t tests are used, the threshold p-value should be relatively high (e.g. 15%).

Session 10: Analysing A Multi Level Experiment

Tutorial and Exercise

Tutorial

- **Session TS10: Analysing a Multi-Level Experiment**
- **Objective:**
  - Develop Python skills to analyse a multi-level designs based on a 4 factor face-centred CC design.
  - This tutorial is based on the Lecture notes in TS10 Analysing a multi-level experiment- see lecture slides for details.
- **Engineering Scenario: Continuing the bumper bracket moulding story**
  - After carrying out the screening experiment for the bumper bracket study, we would now like to design a follow-up experiment. The follow-up experiment is based on the significant factors identified during the screening-experiment.



Tutorial

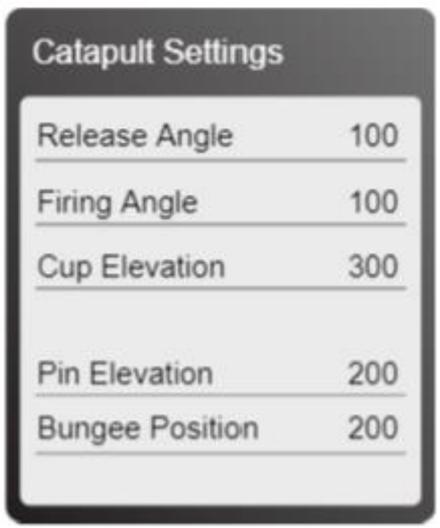
- **Session TS10: Analysing a Multi-Level Experiment**
- **Python Environment**

  A self-guided tutorial has been created as a Colab notebook with pre-designed Python code and notes. For this tutorial, follow the instructions in the notes, upload data files and run the code. No modification of code is required. Interpret the results in accordance with the Technical session.
- **Tutorial Tasks**
  1. Construct a response surface design for 4 factors.
  2. Analyse the data.
  3. Check the residual diagnostics.
  4. Apply the t-test to the regression coefficents.
  5. Plot the response surface.

Exercise

- **Sessions TS08+09 & 10: Three Level Experiment**
- **Objective**
- To design, analyse and discuss the results of a follow up three level experiment for the Virtual Catapult.

  https://sigmazone.com/catapult/



Exercise

- **Sessions TS08+09 & 10: Three Level Experiment**

- **Python Environment**

  The exercise has been created as a Colab notebook with notes. Follow the instructions in the notes, and create your own code using tutorials 08 to 11 as a guide. Interpret the results in accordance with the Technical Sessions.

- **Objectives:**
  - To plan, run and analyse a three-level experiment on the catapult, using **three** factors selected on the basis of your screening results
  - To make predictions for factor combinations that you have not already tested
  - To test your predictions by firing the catapult.

- **Guidelines**

  Amongst other things, you need to decide:
  - Which design to use; which three factors to use; which levels to use for each factor; how many runs to make.
  - If you decide to run a CC design you need to decide how many runs to make at the centre point. If you choose a D-optimal design you should consider whether to add one or more runs at the centre.