Dallin Romney
Design of Experiments

Multi-Factor ANOVA

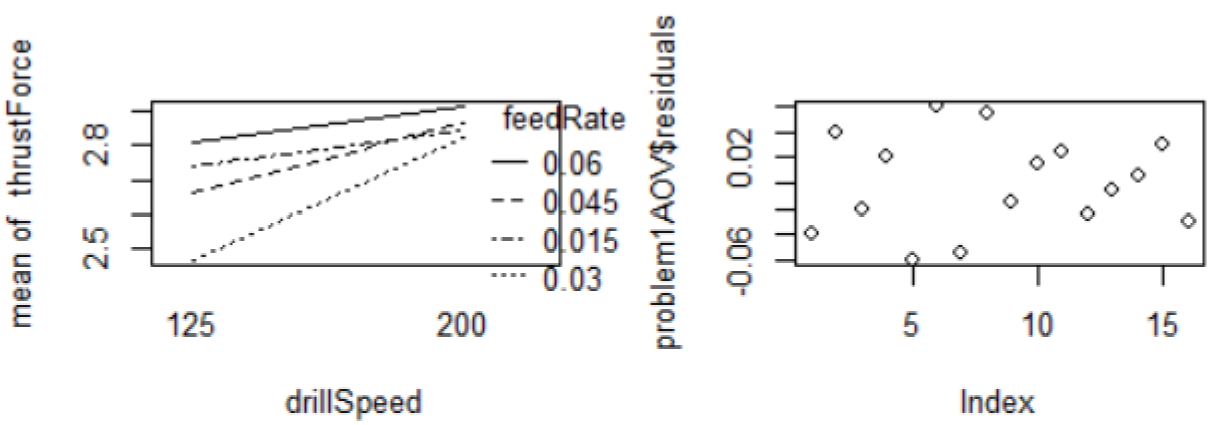
Part I: Multi-Factor ANOVA Problems from Open Stats textbook

5-7. A mechanical engineer is studying the thrust force developed by a drill press. He suspects that the drilling speed and the feed rate of the material are the most important factors. He selects four feed rates and uses a high and low drill speed chosen to represent the extreme operating conditions. He obtains the following results. Analyze the data and draw conclusions. Use $\alpha = 0.05$.

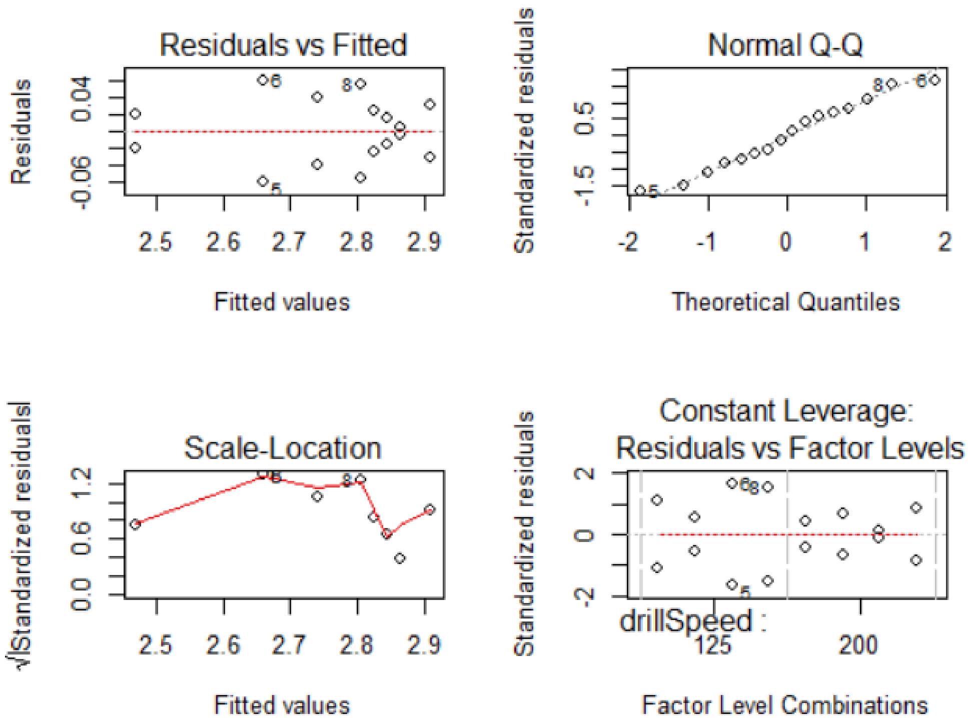
	Feed Rate			
Drill Speed	0.015	0.030	0.045	0.060
125	2.70	2.45	2.60	2.75
	2.78	2.49	2.72	2.86
200	2.83	2.85	2.86	2.94
	2.86	2.80	2.87	2.88

The interaction plot (top left) shows approximately parallel lines, denoting little interaction. The residual plot (top right) shows reasonably distributed and non-trendy ANOVA residuals. Likewise, the four residual plots (bottom) look normal; the data look reasonably homoscedastic and normal, with no outliers.

The ANOVA results show that there is a extremely significant correlation between drill speed and thrust force ($p = 6.6 \cdot 10^{-5}$). There is also a significant correlation between feed rate and thrust force, but less so ($p = 0.0026$). However, there is also significant interaction between the two ($p = 0.026$). Inspecting the data reveals that this can be interpreted as the following: drill speed has the most significant effect on thrust force, and the effect of feed rate on thrust force increases as drill speed increases.



```
> summary(problem1AOV)
                    Df  Sum Sq Mean Sq F value   Pr(>F)    
drillSpeed           1 0.14823 0.14823  57.010 6.61e-05 ***
feedRate             3 0.09250 0.03083  11.859  0.00258 ** 
drillSpeed:feedRate  3 0.04187 0.01396   5.369  0.02557 *  
Residuals            8 0.02080 0.00260                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

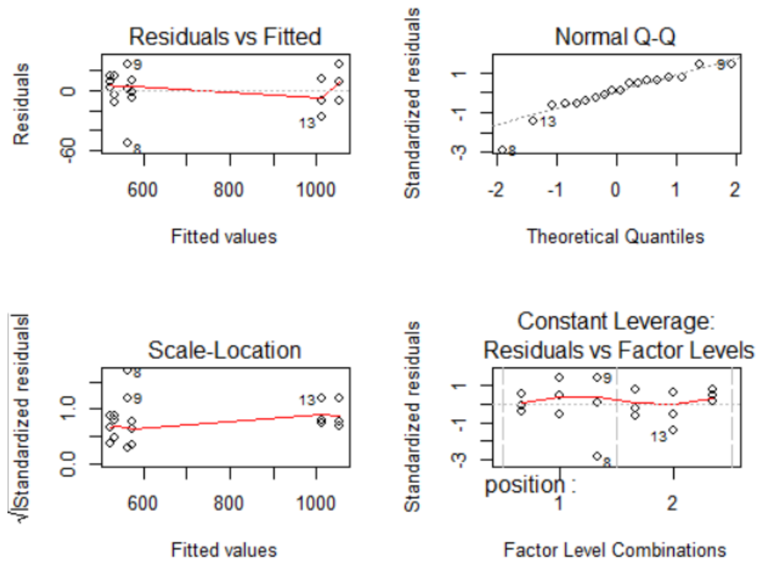


5-11. An experiment was conducted to determine if either firing temperature or furnace position affects the baked density of a carbon anode. The data are shown below:

	Temperature (°C)		
Position	800	825	850
1	570	1063	565
	565	1080	510
	583	1043	590
2	528	988	526
	547	1026	538
	521	1004	532

Suppose we assume that no interaction exists. Write down the statistical model. Conduct the analysis of variance and test hypotheses on the main effects. What conclusions can be drawn? Comment on the model's adequacy.

With no interaction, the linear model is Density = $\beta_0 + \beta_1 \cdot$Position + $\beta_2 \cdot$Temperature. The analysis reveals that there is a very significant effect of temperature on density ($p = 6.6 \cdot 10^{-5}$) and a fairly significant effect from position ($p = 0.00125$). However, this model is inadequate since the data is clearly nonlinear – the samples at 825°C have much higher densities than the other two temperatures, and 825 is between 800 and 850. Surprisingly, the data does appear to satisfy residual requirements fairly well.




```
              Df Sum Sq Mean Sq F value   Pr(>F)    
position       1   7160    7160    16.2  0.00125 ** 
temperature    2 945342  472671  1069.3 4.92e-16 ***
Residuals     14   6189     442                     
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

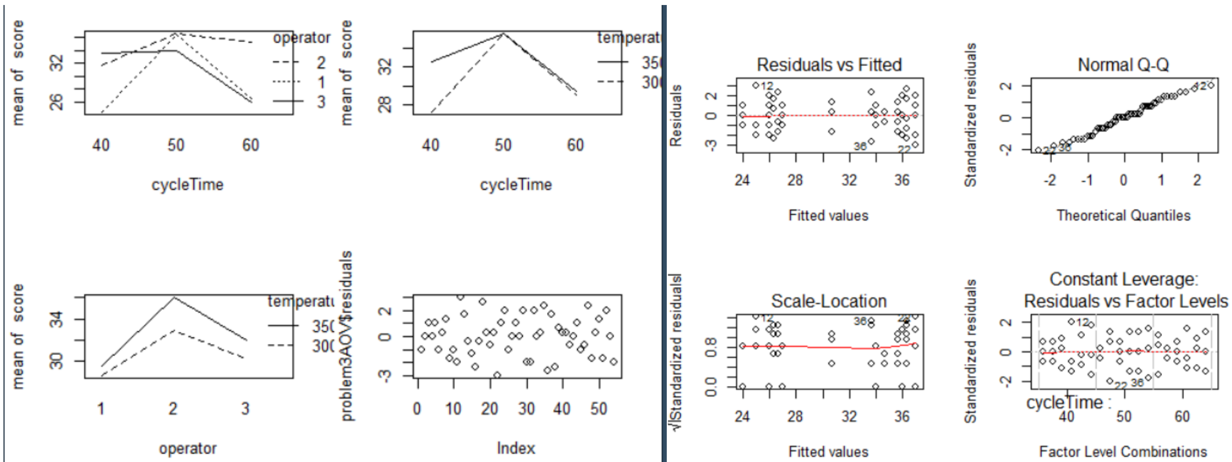
5-17. The quality control department of a fabric finishing plant is studying the effect of several factors on the dyeing of cotton–synthetic cloth used to manufacture men’s shirts. Three operators, three cycle times, and two temperatures were selected, and three small specimens of cloth were dyed under each set of conditions. The finished cloth was compared to a standard, and a numerical score was assigned. The results follow. Analyze the data and draw conclusions. Comment on the model’s adequacy.

	Temperature					
	300°			350°		
	Operator			Operator		
Cycle Time	1	2	3	1	2	3
	23	27	31	24	38	34
40	24	28	32	23	36	36
	25	26	29	28	35	39
	36	34	33	37	34	34
50	35	38	34	39	38	36
	36	39	35	35	36	31
	28	35	26	26	36	28
60	24	35	27	29	37	26
	27	34	25	25	34	24

The residuals of the model appear to satisfy the homoscedasticity and normal requirements of ANOVA. However, from the interaction plots, there appears to be significant interaction between cycle time and operator, as well as some interaction between cycle time and temperature. This is confirmed by the ANOVA analysis, which shows a very significant interaction effect between cycle time and operator (p = $1.8 \cdot 10^{-10}$), as well as a very significant interaction effect between cycle time and temperature (p = 0.0001) and some interaction between the 3 factors (p = 0.016).

All three factors have significant effects on the score (p = $8.1 \cdot 10^{-13}$ for cycle time, p = $7.4 \cdot 10^{-10}$ for operator, and p = 0.000393 for temperature), but that is probably mostly explained by interactions.

From the data, I would conclude that cycle time has the most significant effect on dyeing, and temperature has a small effect. Most likely, the model's inadequacies caused by interactions are caused by operators of different skill level or simply differences in operator methods; the operator also has a major impact on dyeing quality.




```
                             Df Sum Sq Mean Sq F value   Pr(>F)    
cycleTime                     2  436.0  218.00  66.508 8.14e-13 ***
temperature                   1   50.1   50.07  15.277 0.000393 ***
operator                      2  261.3  130.67  39.864 7.44e-10 ***
cycleTime:temperature         2   78.8   39.41  12.023 0.000100 ***
cycleTime:operator            4  355.7   88.92  27.127 1.98e-10 ***
temperature:operator          2   11.3    5.63   1.718 0.193895    
cycleTime:temperature:operator  4   46.2   11.55   3.523 0.015870 *  
Residuals                    36  118.0    3.28                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part II: Multi-Factor ANOVA Design Project

Suppose you want to determine whether the brand of laundry detergent and the water temperature affects the amount of dirt removed from your laundry. You buy two different brands of detergent ("Brand 1" and "Brand 2") and choose three different temperature levels ("cold", "warm", and "hot"). Then you divide your laundry randomly into 6×r piles of equal size and assign each r piles into the combination of ("Brand 1" and "Brand 2") and ("cold", "warm", and "hot").

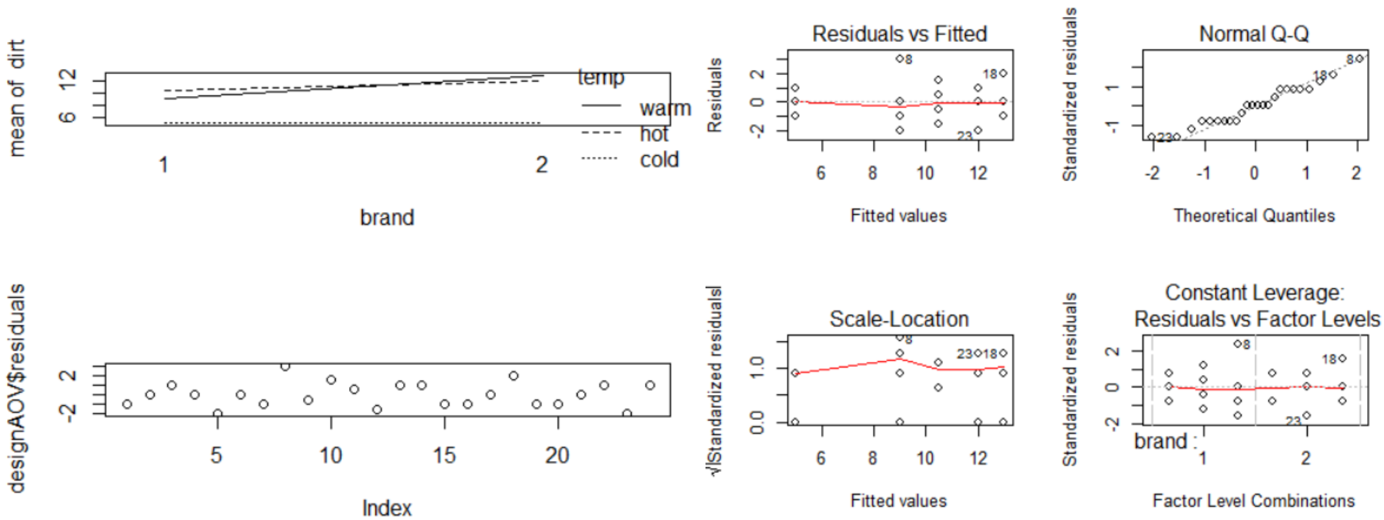
The experiment has two factors (Factor Detergent, Factor Temperature) at a = 2(Brand 1 and Brand 2) and b = 3(cold, warm and hot) levels. Thus there are ab = 3×2 = 6 different combinations of detergent and temperature. With each combination you wash r = 4 loads. r is called the number of replicates. This sums up to n = abr = 24 loads in total. The amounts Y_{ijk} of dirt removed when washing sub pile k (k = 1,2,3,4) with detergent i (i = 1,2) at temperature j (j = 1,2,3) are recorded below.

	Cold	Warm	Hot
Brand 1	4,5,6,5	7,9,8,12	10,12,11,9
Brand 2	6,6,4,4	13,15,12,12	12,13,10,13

Inspect your data and draw the appropriate conclusions. Include the appropriate ANOVA results and interaction plot. Comment on each brand of detergent and interactions that are seen.

The residual plot and fitted plots appear to satisfy mean zero, normal (approximately, a bit choppy), and homoscedasticity requirements for the model. From the top left plot (interactions), there does appear to be some interaction between brand and temperature: both brands work about the same at cold temperatures, but Brand 2 is actually most effective at warm rather than hot temperatures, while Brand 1 gets better with higher temperature.

Brand does appear to have a significant effect on cleanliness ($p = 0.0058$), but temperature is undoubtedly more significant ($p = 5.4 \cdot 10^{-8}$). In general, I would recommend focusing on temperature for each brand, and using Brand 2 over Brand 1 (not considering price).




```
           Df Sum Sq Mean Sq F value   Pr(>F)    
brand       1  20.17   20.17   9.811  0.00576 ** 
temp        2 200.33  100.17  48.730 5.44e-08 ***
brand:temp  2  16.33    8.17   3.973  0.03722 *  
Residuals  18  37.00    2.06                     
```