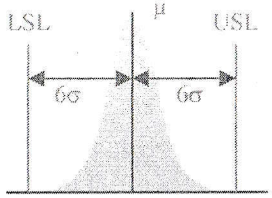
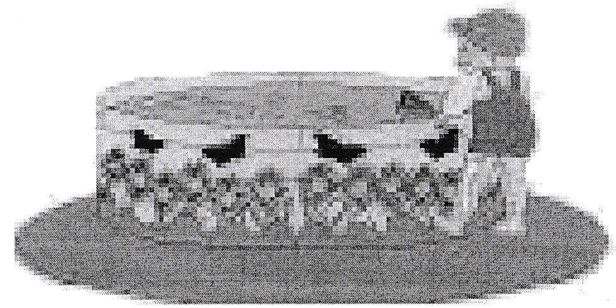
Pooling

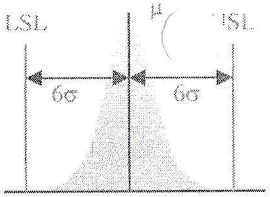


The Tech Group

Module Objectives

- What is pooling?
- Why use pooling?
- When to use pooling?
- Pooling concepts
  - How to start
  - When to stop
- Risks of pooling
- Examples

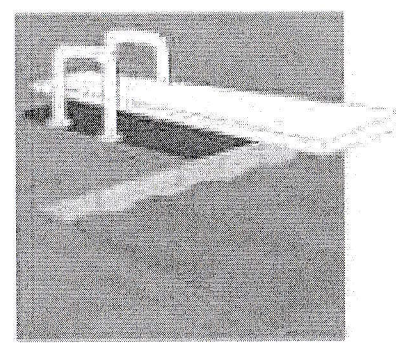


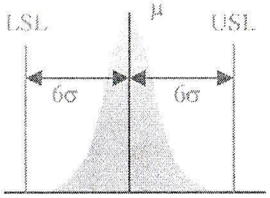


Pooling

- **What Is Pooling**
  - Pooling is the act of removing factors or interactions that are not significant, in order to gain degrees of freedom to estimate experimental error.

- **Why Use Pooling**
  - Pooling of the mean squares can help facilitate the identification of significant effects.
  - Allows you to reduce the model and provide a more powerful search for main effects.
  - Minimize terms in a predictive equation

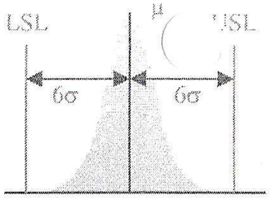




The Tech Group

When To Use Pooling

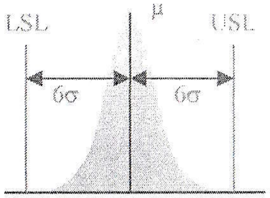
- **Use Pooling…**
  - When the mean square error has fewer than 6 degrees of freedom.
  - If the F statistic for the factor being pooled is not significant at a large alpha value (.25)
- **Do Not Use Pooling…**
  - When the mean square error has more than 6 degrees of freedom.
  - If the F statistic for the factor being pooled is significant at a large alpha value (.25)



Pooling Concepts

- **Eliminate Effects**
  - Based on use of other tools

- **Pool Less Significant Effects Into Error**
  - Recalculate the mean square and F statistic ratios before testing the main effects. This would be equivalent to eliminating the interactions line from the model in an ANOVA analysis.

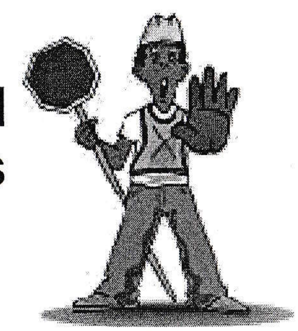


The Tech Group

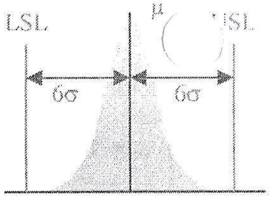
Pooling Concepts

- **Where To Start**
  - Pool the terms with the smallest effect
  - As additional terms are pooled , the error should not increase significantly if the terms are in fact insignificant.

- **When To Stop**
  - It is possible to pool to much. If this happens the noise will increase significantly and will make other significant terms appear insignificant.



  - When R square adjusted level begins to fall significantly (Fitted Line Plot)



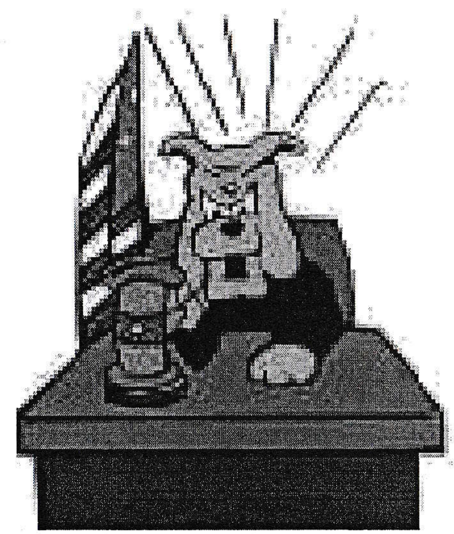
The Tech Group

Pooling Rules

1) Begin with smallest effects (insignificant high order interactions)

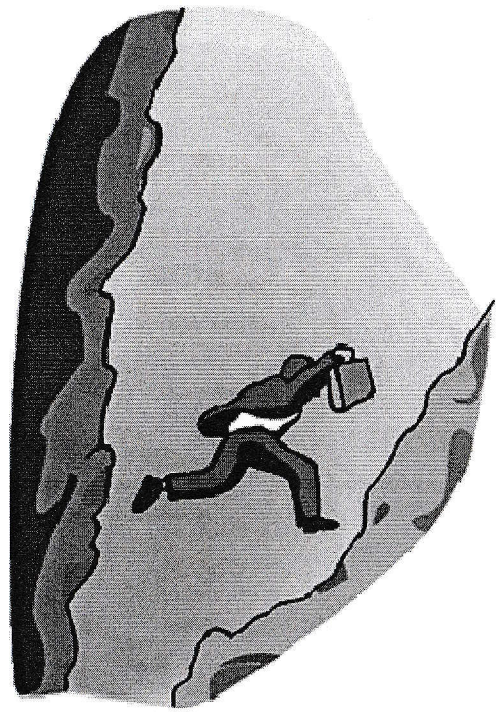
2) Do not pool a main effect without removing all interactions.

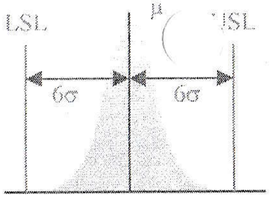
3) If there is a significant interaction do not pool the main effects associated with it.



Risks of Pooling

- There is risk that one may pool the mean square for a factor that is *really* significant with error. This mistake would show as an artificial inflation of the new residual mean square error and make other significant effects harder to detect.





The Tech Group

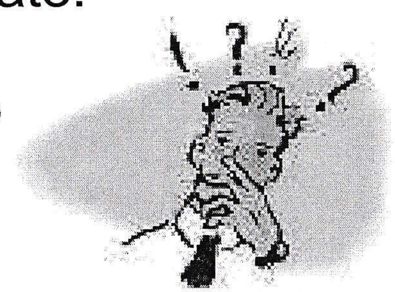
Real World Experience

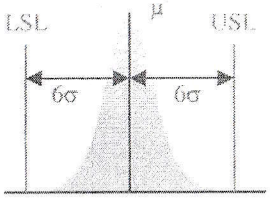
Frequently, resources are not available to run a replicate of an experiment. For example, we would like to run a 24 design (16 runs / treatments) which would be replicated (32 runs / treatments), but the budget will only allow us to run 16 total runs.

What can we do in this situation?

- We can drop a factor we suspect is not likely to be significant, thus ending up with a 2^3 design. A 2^3 design is a 8 run full factorial, therefore we could replicate.
- We could run a fractional factorial 2^{4-1} and replicate.
- We could keep the 2^4 design and eliminate the replicate.

However, what if we could only afford 8 runs?



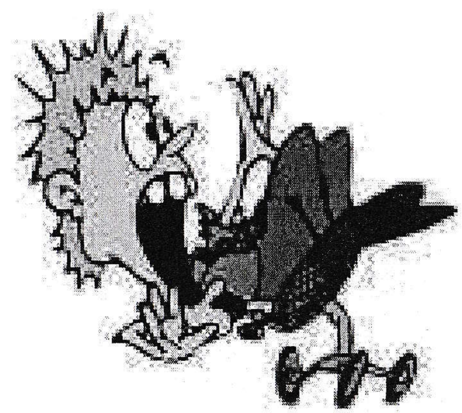


Downfalls of Un-replicated Designs

The Tech Group

- We have no estimate of random chance variation (experimental error)

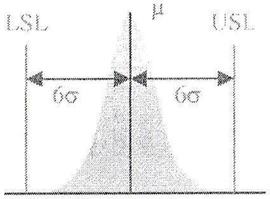
- We can't …
  - Use ANOVA to analyze
  - Compare random variation (within) to the variation between group (factor) means.



Tips For Analyzing Un-replicated Designs

The Tech Group

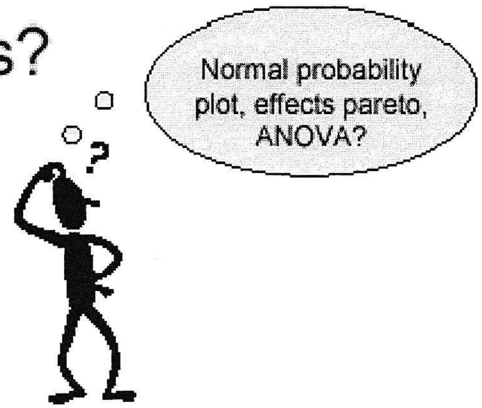
- Normal Probability Plot of the Effects
  - A graphical technique that is somewhat subjective.
  - Minitab uses Lenth's method to label significant effects when no error term exists. For replicated designs Minitab uses ANOVA (default a = 0.10).
- Pareto Chart of the Effects
  - Another graphical technique that is subjective.
  - Minitab uses Lenth's method to generate the significance line when no error term exists. For replicated designs Minitab uses ANOVA (default a = 0.10).
- ANOVA
  - Assume that certain high-order interactions are negligible and combine their mean squares to estimate error. This is an appeal to the sparsity of effects principle; that is, most systems are dominated by some of the main effects and low-order interactions and most high-order interactions are insignificant.
  - The risk with this method is that occasionally, real high-order interactions exist.



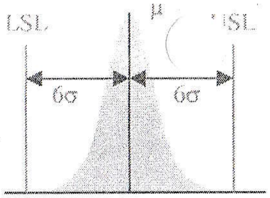
Pooling Concepts

The Tech Group

- Suppose that I could eliminate some effects based on use of the other tools?



- Could I selectively pool the less significant effects into error so that I could use ANOVA to improve my estimates of significance?

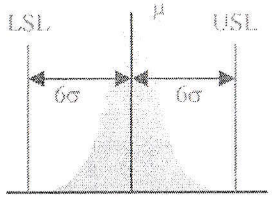


Pooling Example-1

Using MINITAB

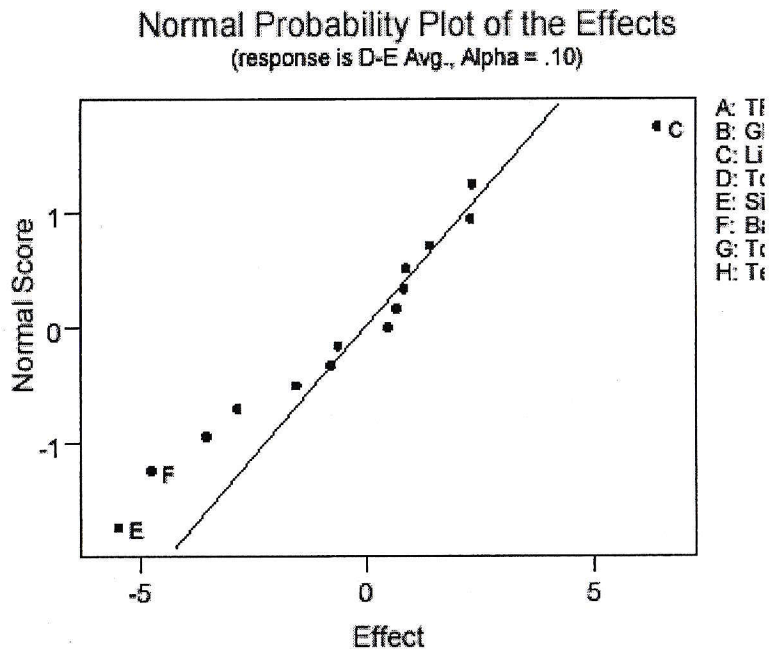
- To improve PC-4 release consistency, a DOE was planned to determine the key factors effecting silicone coverage and aged release based on the silicone coater set-up. Because of the number of factors (8) that needed to be considered, a full factorial design would have required 256 runs. The cost and coater time required was prohibitive.
- Based on a balance of cost, coater time, and information required, the decision was made to go with a 2^{8-4}_{IV} design that would allow us to identify the main effects (ignoring three way interactions) and identify possible two way interactions for further work. The experiment was not replicated because of the cost and time required.

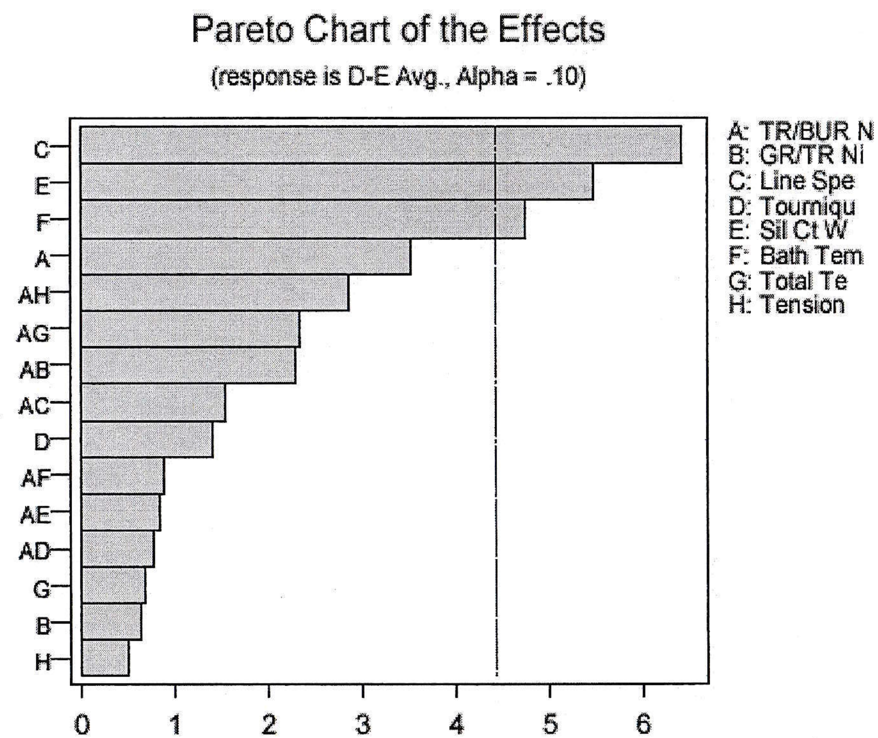
The effects pareto and normal probability plot for the initial analysis are on the following page.



The Tech Group

Pooling Example-1





Pooling Example-1

Fractional Factorial Fit

Estimated Effects and Coefficients for D-E (coded units)

Term	Effect	Coef
Constant		24.310
TR/BUR N	-3.527	-1.764
GR/TR Ni	-0.623	-0.311
Line Spe	6.420	3.210
Tourniqu	1.395	0.697
Sil Ct W	-5.487	-2.744
Bath Tem	-4.747	-2.374
Total Te	0.675	0.338
Tension	0.485	0.242
TR/BUR N*GR/TR Ni	2.290	1.145
TR/BUR N*Line Spe	-1.532	-0.766
TR/BUR N*Tourniqu	-0.763	-0.381
TR/BUR N*Sil Ct W	0.830	0.415
TR/BUR N*Bath Tem	0.875	0.438
TR/BUR N*Total Te	2.327	1.164
TR/BUR N*Tension	-2.857	-1.429

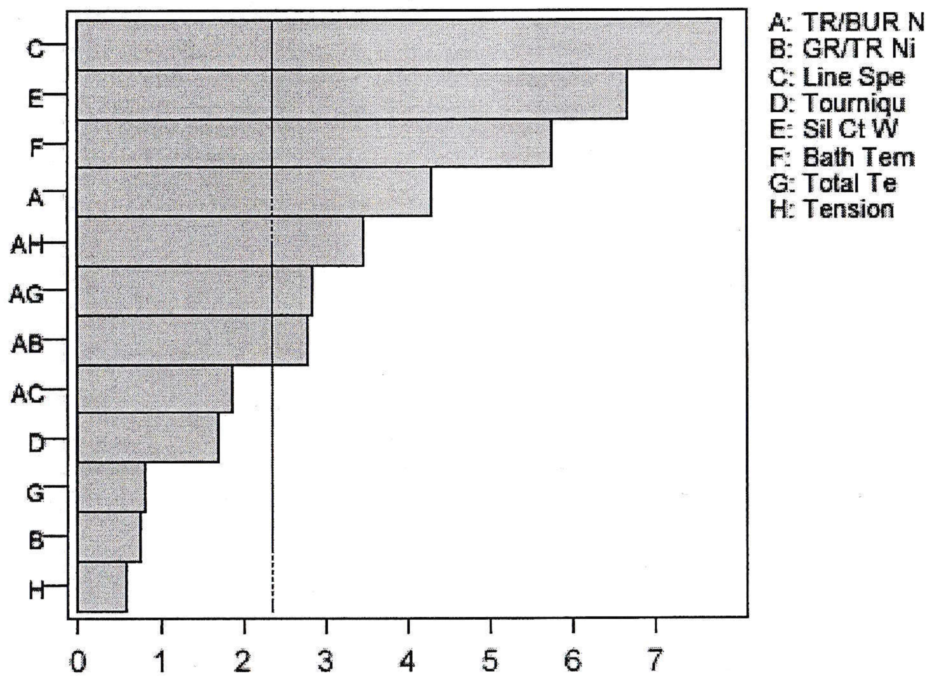
Analysis of Variance for D-E (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	8	437.34	437.34	54.67	*	*
2-Way Interactions	7	92.84	92.84	13.26	*	*
Residual Error	0	0.00	0.00	0.00		
Total	15	530.19				

Pooling Example-1

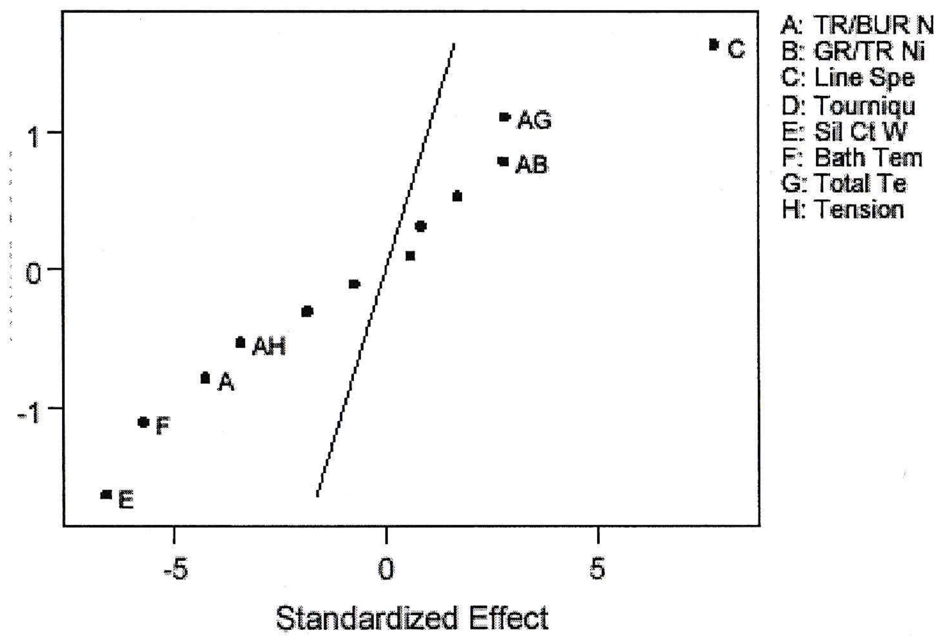
Pareto Chart of the Standardized Effects

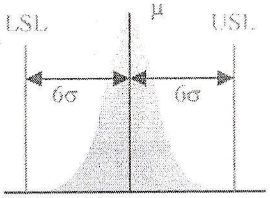
(response is D-E Avg., Alpha = .10)



Normal Probability Plot of the Standardized Effects

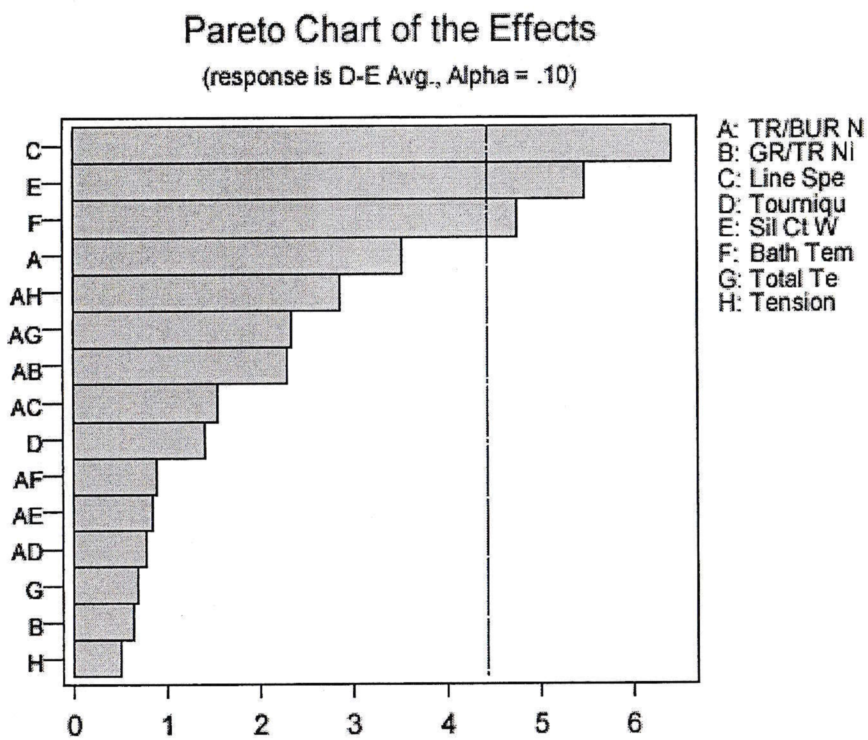
(response is D-E Avg., Alpha = .10)



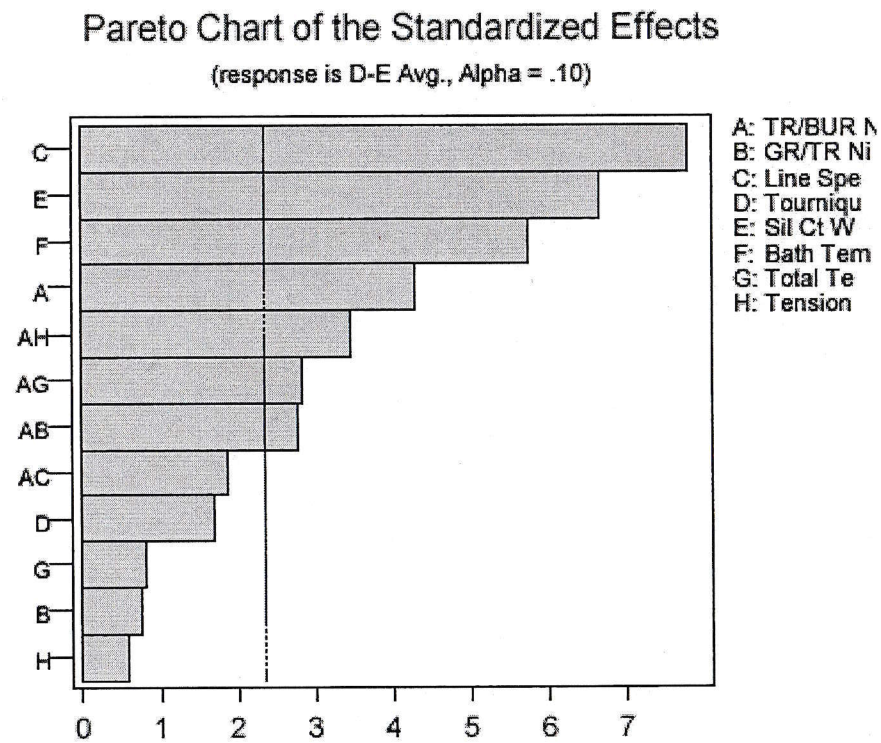


The Tech Group

Pooling Example-1



Before Pooling



After Pooling

Pooling Example-1

Fractional Factorial Fit: D-E Avg. versus TR/BUR Nip P, GR/TR Nip Im, ...

Estimated Effects and Coefficients for D-E (coded units)

Term	Effect	Coef	SE Coef	T	P
Constant		24.310	0.4119	59.02	0.000
TR/BUR N	-3.527	-1.764	0.4119	-4.28	0.023
GR/TR Ni	-0.623	-0.311	0.4119	-0.76	0.505
Line Spe	6.420	3.210	0.4119	7.79	0.004
Tourniqu	1.395	0.697	0.4119	1.69	0.189
Sil Ct W	-5.487	-2.744	0.4119	-6.66	0.007
Bath Tem	-4.747	-2.374	0.4119	-5.76	0.010
Total Te	0.675	0.338	0.4119	0.82	0.473
Tension	0.485	0.242	0.4119	0.59	0.597
TR/BUR N*GR/TR Ni	2.290	1.145	0.4119	2.78	0.069
TR/BUR N*Line Spe	-1.532	-0.766	0.4119	-1.86	0.160
TR/BUR N*Total Te	2.327	1.164	0.4119	2.83	0.066
TR/BUR N*Tension	-2.857	-1.429	0.4119	-3.47	0.040

Analysis of Variance for D-E (coded units)

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	8	437.342	437.342	54.668	20.14	0.016
2-Way Interactions	4	84.701	84.701	21.175	7.80	0.062
Residual Error	3	8.144	8.144	2.715		
Total	15	530.186				

NOTE: Minitab displays T values instead of F values.

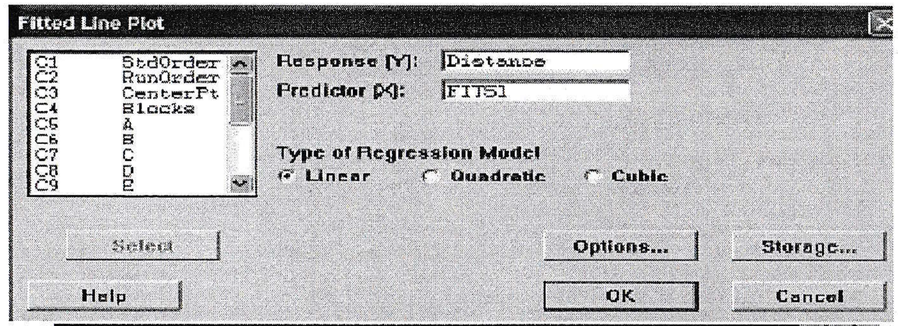
They are equivalent in this case (The square of a T random variable with ν degrees of freedom is an F random variable with 1 numerator and ν denominator degrees of freedom).

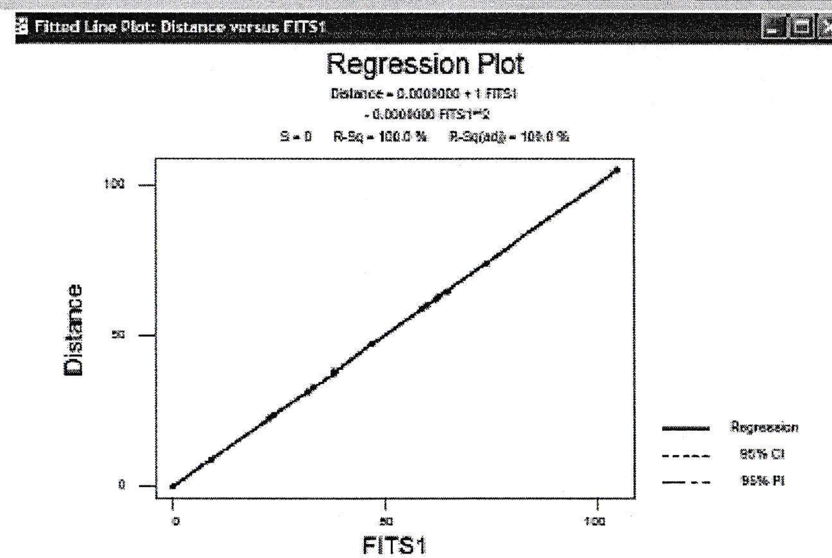
Pooling Example-2

The Tech Group

Creating A Regression Plot and R Squared Adjusted

- Stat > Regression> Fitted Line Plot
- Select Distance in Y
- Select FITS1 in X
- Check Linear



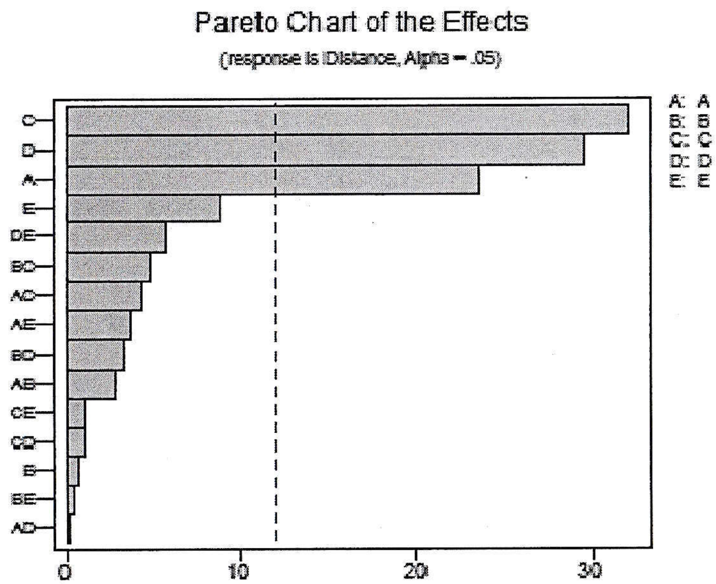


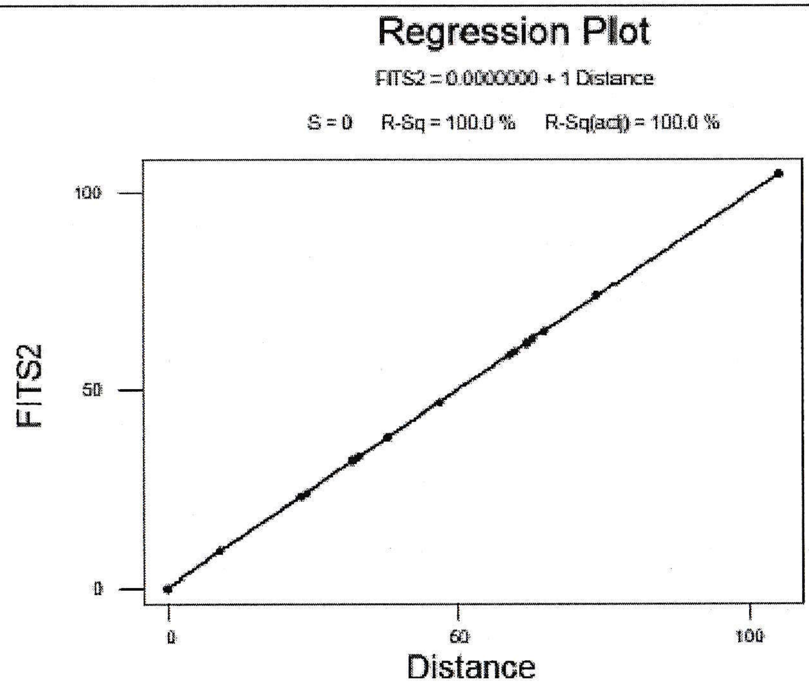
- R^2 Adjusted - % of data that is described by the model
  - Between 0 and 100%
  - 100% - all data can be explained by the model

Pooling Example-2

The Tech Group

Deleted 3,4,5 way interactions – R2 Adjusted 100%

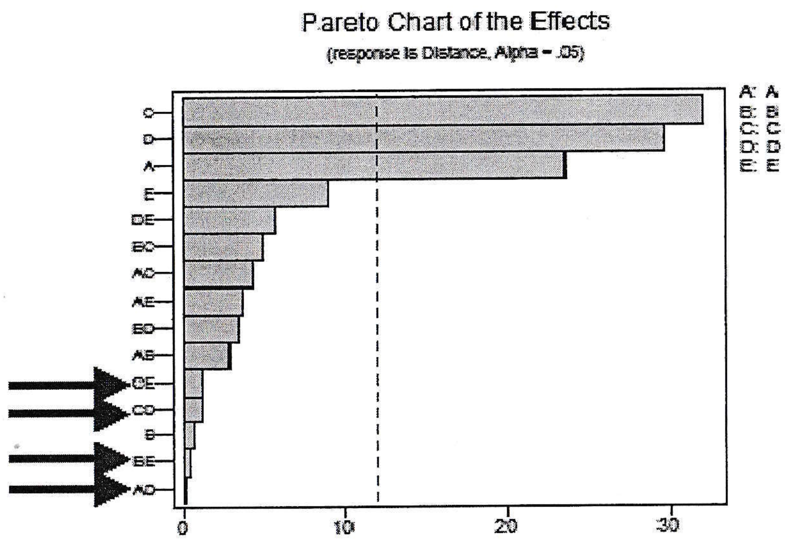


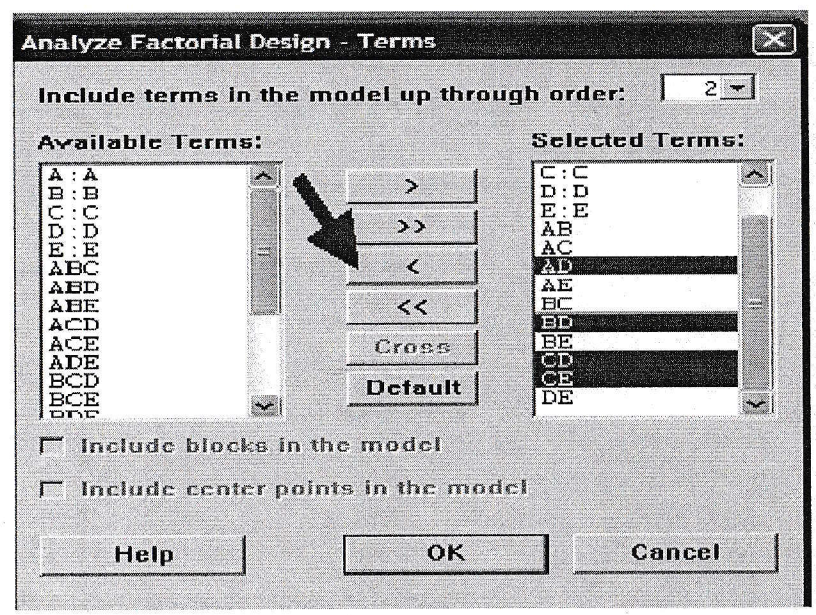


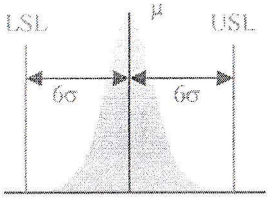
Pooling Example-2

Pooling in Minitab

Stat > DOE > Factorial >Analyze Factorial Design > Terms
Select insignificant terms
Select single back arrow $\boxed{<}$
The insignificant terms will be removed from the DOE model
Ok, Ok





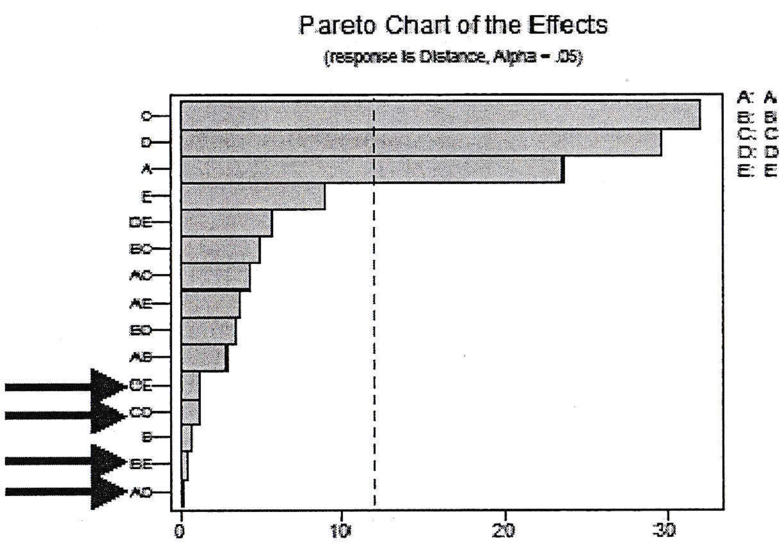


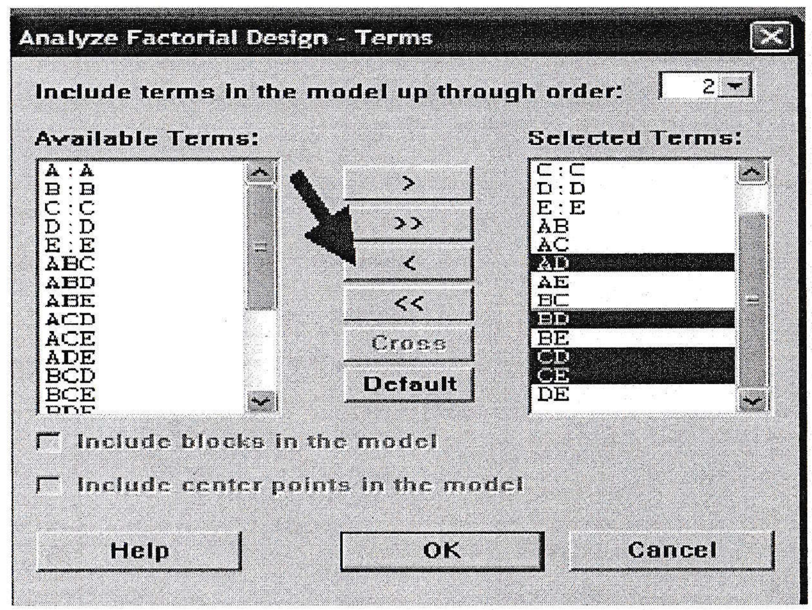
Pooling Example-2

The Tech Group

Pooling in Minitab

Stat > DOE > Factorial >Analyze Factorial Design > Terms
Select insignificant terms
Select single back arrow <
The insignificant terms will be removed from the DOE model
Ok, Ok

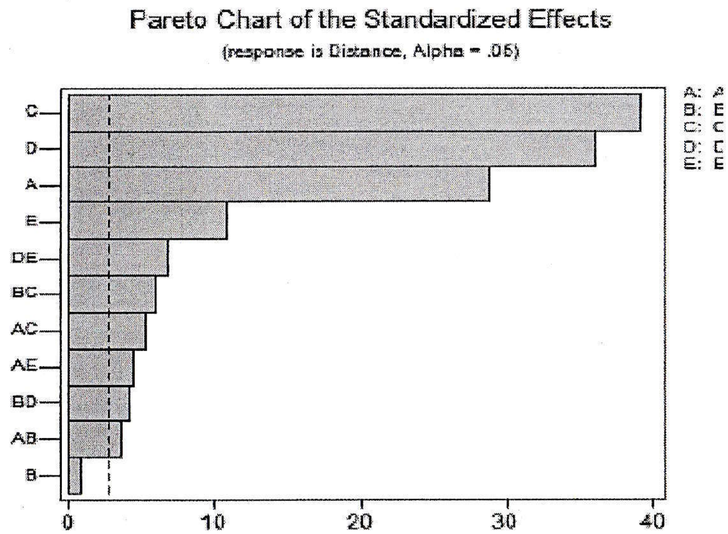




Pooling Example-2

The Tech Group

Pooled Out AD, BE, CD, CE – R squared adjusted 99.9%



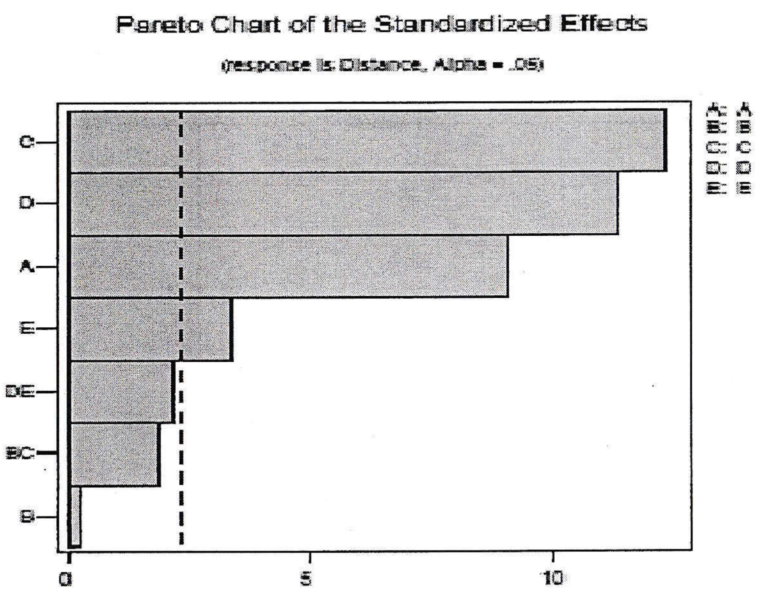
All remaining terms are significant
except B, the BC interaction is significant, there for the B term cannot be pooled

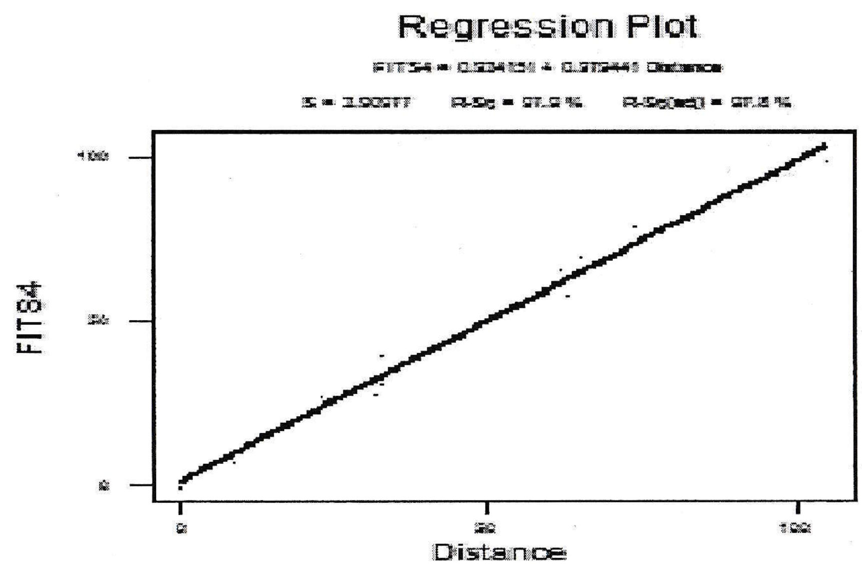
Would you stop pooling here?

Pooling Example-2

The Tech Group

Pooled Out AB, BD, AE, AC – R squared adjusted 97.5%





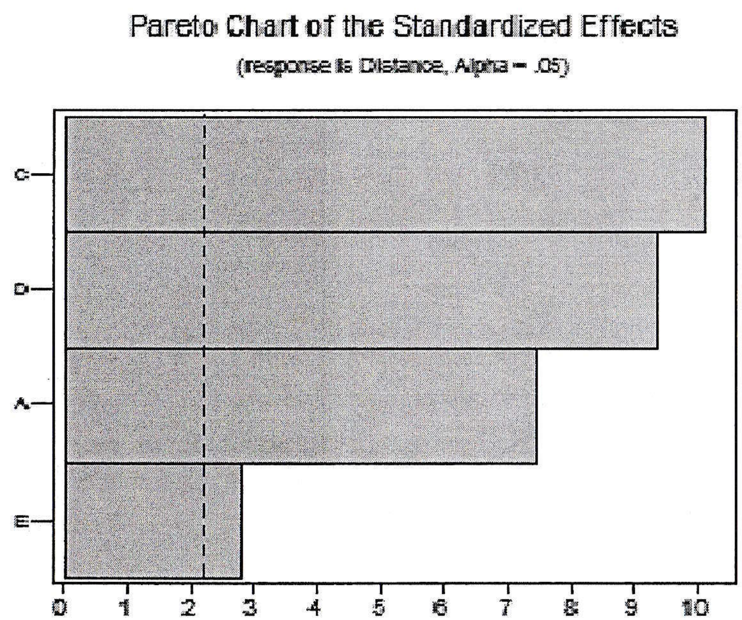
DE and BC are now insignificant when they were significant – Possible Type II error (Beta error)

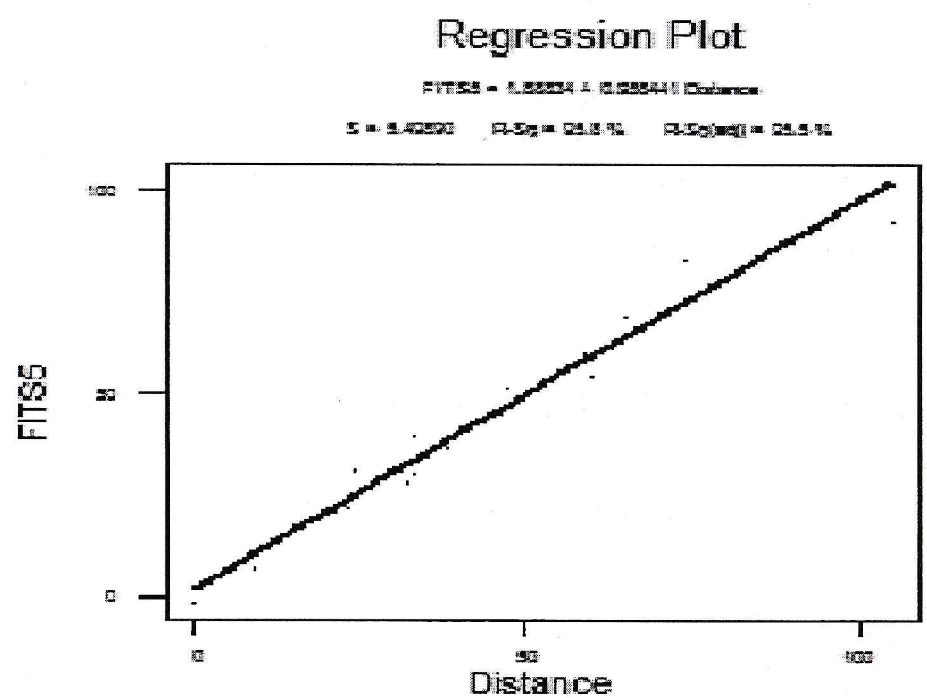
From the effects plot, does the effect of DE and BC have a practical effect on the response?

Pooling Example-2

The Tech Group

Pooled Out B, DE, BC – R squared adjusted 95.5%

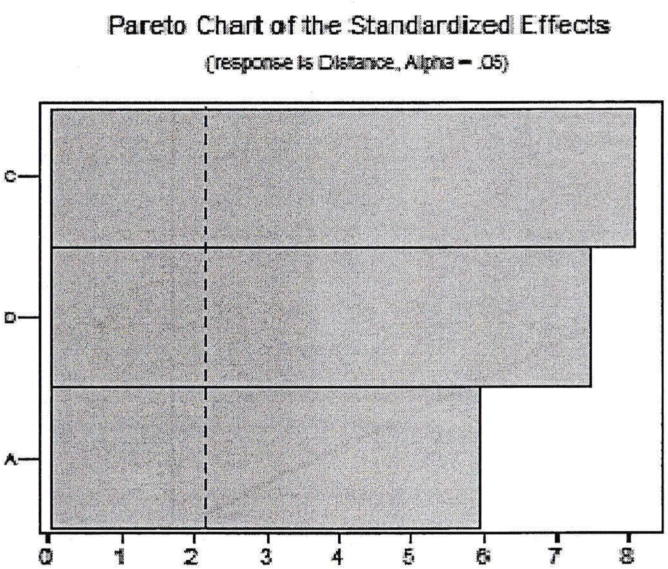


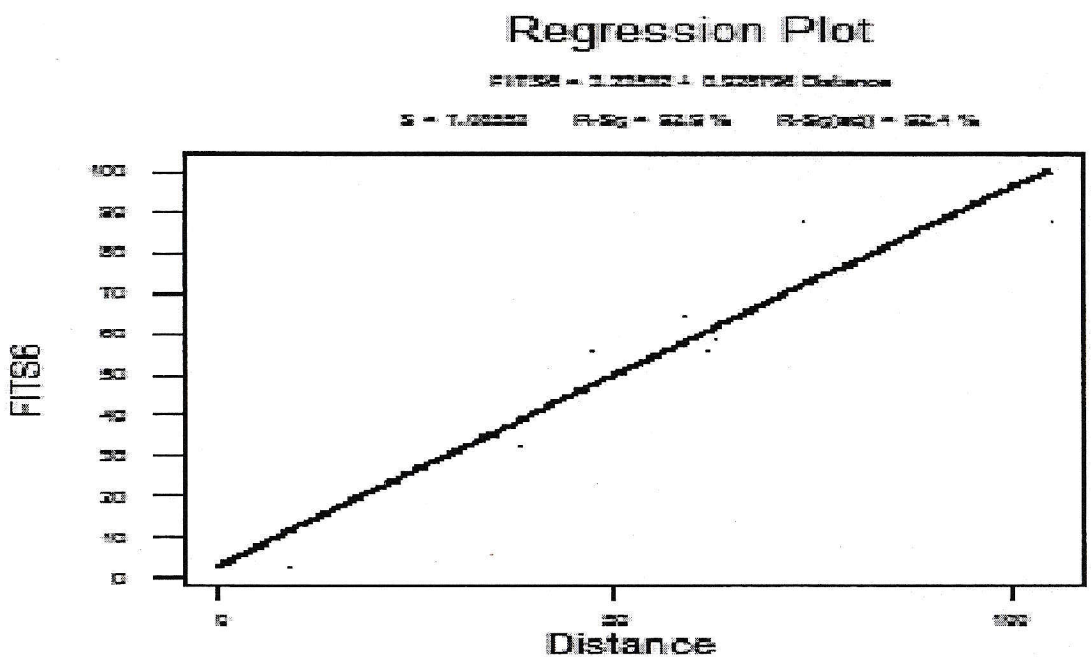


Pooling Example-2

The Tech Group

Pooled Out E – R squared adjusted 92.4%

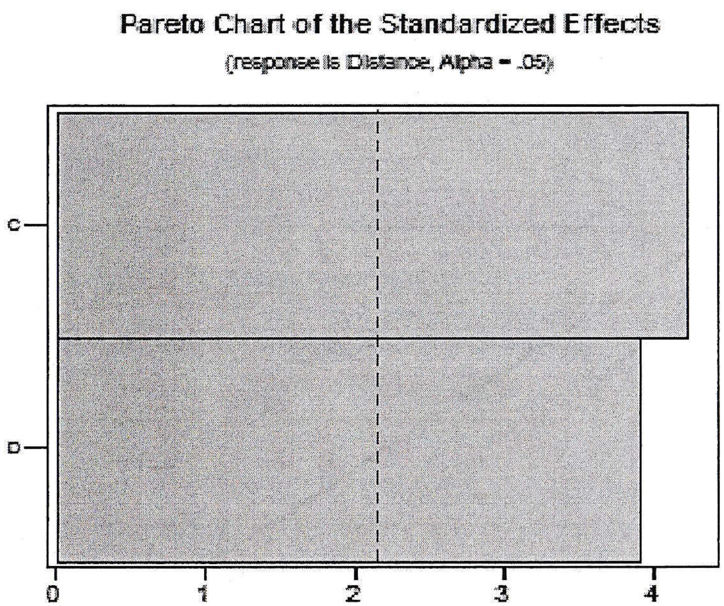


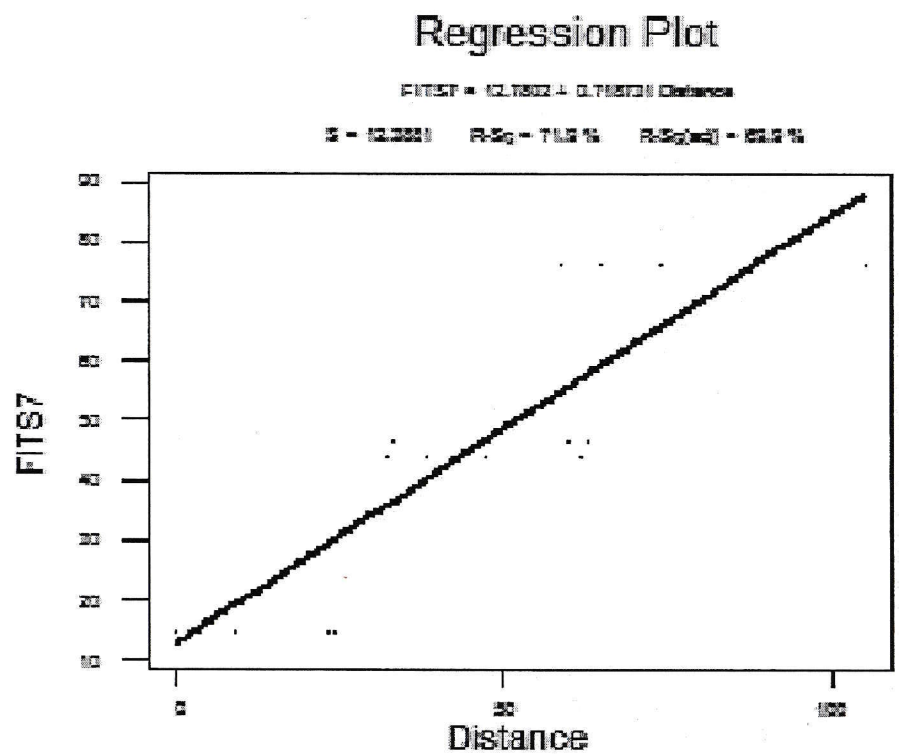


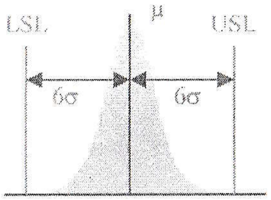
Pooling Example-2

The Tech Group

Pooled Out A – R squared adjusted 69.0%







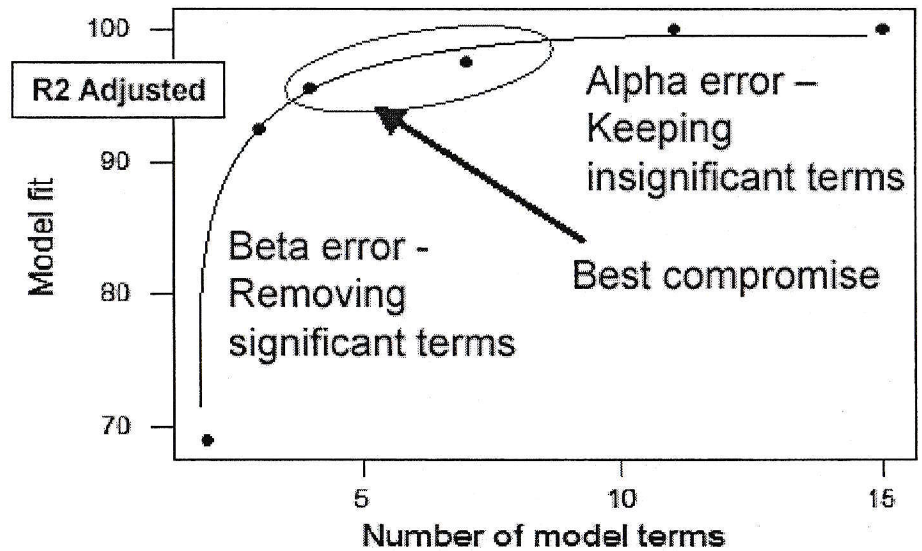
Pooling Example-2

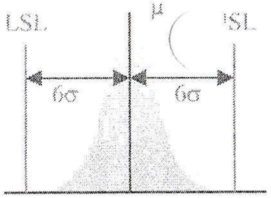
The Tech Group

Sparcity of Effects Principle

Alpha error – results in action taken when not necessary causing cost and schedule impact

Beta error – results in significant terms being ignored causing customer impact and rework





The Tech Group

Summary

- Pooling the mean square of insignificant factors and interactions provides degrees of freedom to estimate experimental error.
- Pool when the mean square error has fewer than 6 degrees of freedom.
- Pool only if F statistic for factor or interaction is **<u>NOT</u>** significant at a large alpha value (.25)
- Start with terms with smallest effects and stop when R^2 falls significantly.