Some Experiences in Modern Experimental Design

Marcus B. Perry

The University of Alabama
Dept. of Information Systems, Statistics, and Management Science

Introduction

- In today's Industry 4.0, industrial processes are increasing in complexity, presenting significant challenges to the industrial experimenter
- Many experimental design and analysis challenges today stem from data that are non-normal and/or correlated
- Non-normal and correlated experimental data is quite common in modern industry, and is often analyzed in practice by either:
  1. Use a classical design and (if needed) apply a data transformation to the response, i.e., transform to normality, then proceed with an analysis of the transformed response using well-known normal theory models, or
  2. Design and analyze the experiment under the framework of the generalized linear model (GLM)
     - Should involve optimal design for non-linear models

Data Transformations to Normality?

Advantages

- Data transformations are practitioner friendly and easy to implement
- Computationally inexpensive
- Assuming the data transformation is adequate, well-known classical designs such as 2-level factorials and their fractions are still "good" designs for analyzing the transformed response.
- Use of data transformations permits one to use well-known normal theory-based model fitting and inference methods with the transformed response, to include small-sample inference.
- Use of well-known model selection and model diagnostics strategies can be used with the transformed response

Data Transformations to Normality?

Disadvantages

- No guarantee that a single transformation will simultaneously induce all three of the desirable properties of normality, constant variance, and linearity
  - Easy to see if one considers analysis of discrete data where errors are, e.g., Poisson:
    - $\sqrt{Y}$ approximately stabilizes variance
    - $Y^{2/3}$ does better for approximating normality
    - $\log_e(Y)$ produces additivity of the systematic effects
- Transformation may not be defined at the boundaries of the sample space
- At times a transformation can result in nonsensical values, e.g., see Meyers and Montgomery (1997).

Data Transformations to Normality?

If final model is used for prediction purposes:

- Predicted values in original units are subject to *retransformation bias*

** Easy to see if one considers Jensen's inequality, i.e., $E[h(Z)] \geq h(E[Z])$, where Z is a random variable and $h(\cdot)$ is a convex function

Example (Square-Root Transformation)

- Suppose $\sqrt{Y} = Z \sim N(\mu, \sigma^2)$ denotes a transformation from the original units Y to the transformed units Z, so that $Y = h(Z) = Z^2$, then

$$E[Y] = E[h(Z)] = E[Z^2] = \mu^2 + \sigma^2 > h(E[Z]) = h(\mu) = \mu^2$$

- Retransformation bias is then given by

$$E[Y] - h(E[Z]) = Var[Z] = \sigma^2$$

- This problem and solutions are discussed in Neyman and Scott (1960), Land (1974), Miller (1984), Montgomery and Peck (1992), Perry and Walker (2015), Perry (2018a), and Perry (2018b)

Generalized Linear Model (GLM)

Advantages

- GLMs represent a transformation on the population mean, not the data $\Rightarrow$ response variable does not change
- Naturally resolves the issues outlined above posed by data transformations
- GLMs permit significantly more modeling flexibility by allowing a separate modeling of linearity and variance relationships
- Added flexibility allows the analyst to focus more on selecting an appropriate model as opposed to finding manipulations to make the data fit a restricted class of models
- Extensions of GLMs to GEEs and GLMMs allow for correlated observations

Generalized Linear Model (GLM)

Disadvantages

- Determining which design points to run is much more complicated under the GLM framework due to the design dependence problem, e.g., see Khuri et al. (2006) and Woods et al. (2017) $\Rightarrow$ Classical designs are not generally the "best" designs
    - Significant amount of work in optimal design for nonlinear models, particularly for Binomial, Poisson and Gamma models
    - Work in optimal design under the GLM framework when the experiment involves randomization restrictions is less known, one exception is Atkinson and Woods (2015).
- Model fitting/testing/diagnostics generally becomes more challenging, particularly in the case where observations are correlated
- Statistical inference relies on asymptotic results $\Rightarrow$ problematic for model editing when experimental runs are expensive or time consuming

Data Transformation or GLM - Which to use?

- Unless the response is binary or low-level counts, I highly recommend the use of data transformations.

Electron Microscopy Experiments

Electron Microscopy Experiments

- Electron microscopy experiments analyzed in Perry *et al.* (2015)
- Machining of copper bars
- Experimental factors include two machining parameters
  1. Rake angle (factor A)
  2. Cutting speed (factor B)
- Response: Grain Size
- Goal of experiments: Build model to predict material properties from the distribution of the nano-grains on the machined surface

Sampling Process

Electron Microscopy Experiments - Nonstandard Experimental Data

1. For any given bar stock, the sampled grain sizes are NOT a simple random sample $\Rightarrow$ randomization restrictions!
   - 2-stage sampling process $\Rightarrow$ nested error structure
     i. Random sampling at bar-stock level
     ii. Random sampling at chip level
2. Grain size is not normally distributed

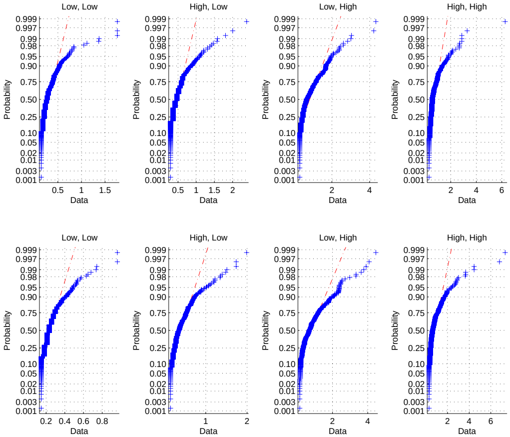
Sampling Process

Electron Microscopy Experiments - Experimental Design

- Materials scientists chose to run two replicates of a 2^2 factorial design
  - Total of 8 experimental runs (or 8 bar-stock)
- For the i^{th} experimental run ($i = 1, 2, ..., 8$):
  1. Only one chip was randomly selected from the available population of chips
  2. A vector of grain size observations of length n_i was recorded from the corresponding machined chip's surface

Electron Microscopy Experimental Data

Normal Probability Plots of Grain Size



Alternatives for retrospective analysis of experimental data

Alternatives for retrospective analysis of the microscopy experimental data:

1. **Standard ordinary least squares (OLS) analysis on 8-point averages**

Electron Microscopy Data - Analysis of 8-Point Averages

- Let $\bar{Y}$=average grain size; A=Rake Angle; B=Cutting Speed
- Postulated model: $\bar{Y}_i = \theta_0 + \theta_1 A + \theta_2 B + \theta_{12} AB + \epsilon_i$
- $\epsilon_i' s \sim iid\ N(0, \sigma^2)$

* Similar model can be fit using $2 \log_e S$ as response to identify significant dispersion effects

OLS Analyses

Design

Std	Run	Factor 1 A:Rake Angle	Factor 2 B:Cutting Speed	Response 1 R1
1	5	-1	-1	0.329074
2	1	-1	-1	0.260349
3	6	1	-1	0.529659
4	8	1	-1	0.430093
5	2	-1	1	0.882943
6	7	-1	1	0.904229
7	3	1	1	0.758901
8	4	1	1	0.901006

Analysis

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	0.4503	1	0.4503	48.26	0.0004	significant
B-Cutting Speed	0.4503	1	0.4503	48.26	0.0004	
Residual	0.0560	6	0.0093			
Lack of Fit	0.0383	2	0.0192	4.35	0.0993	not significant
Pure Error	0.0176	4	0.0044			
Cor Total	0.5062	7				

Alternatives for retrospective analysis of experimental data

Alternatives for retrospective analysis of the microscopy experimental data:

1. Standard ordinary least squares (OLS) analysis on 8-point averages
2. **Data transformation to normality, then perform analysis on transformed response using the linear mixed model (LMM)**

Analysis of Transformed Data using LMM

- Consider the Box-Cox transformations

$$Y(\lambda) = \begin{cases} \frac{Y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log_e(Y) & \lambda = 0 \end{cases}$$

  where Y is the original response variable, λ is the transformation parameter, and $Y(\lambda)$ is the transformed response variable.
- Postulated model:

$$y_{ij}(\lambda) = \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB + \delta_i + \epsilon_{ij}$$

  for $i = 1, ..., 8$ and $j = 1, ..., 64$
- Also, β_0, β_1, β_2, and β_{12} denote the fixed-effect model components and $\delta_i \sim N(0, \sigma_\delta^2)$ and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, where $Cov(\delta_i, \epsilon_{ij}) = 0$ for all i and j, denote the random model components.

Analysis of Transformed Data using LMM

Fixed Effects [Type III]

Response 1: R1
Transform: Inverse Sqrt
Constant: 0

REML (REstricted Maximum Likelihood) analysis
Kenward-Roger p-values

Source	Term df	Error df	F-value	p-value	
Whole-plot	3	4.00	11.65	0.0191	significant
a-a	1	4.00	3.61	0.1301	
b-b	1	4.00	29.85	0.0055	
ab	1	4.00	1.50	0.2882	

Coefficients in Terms of Coded Factors

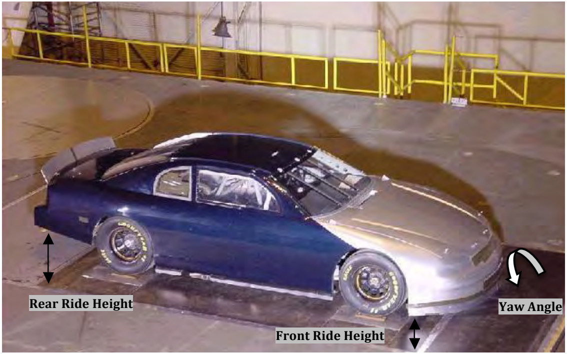
Source	Coefficient Estimate	Standard Error	VIF
Intercept	1.52	0.0508	
a-a	-0.0966	0.0508	1.0000
b-b	-0.2778	0.0508	1.0000
ab	0.0622	0.0508	1.0000

Variance Components

Source	Variance	Standard Error	95% CI Low	95% CI High
Group	0.0182	0.0146	-0.0105	0.0468
Residual	0.1590	0.0100	0.1411	0.1806
Total	0.1772			

NASCAR Wind Tunnel Experiments

NASCAR Wind Tunnel Experiments



Factor	Type	Low level	High level
Front ride height (x_1)	HTC	−0.5 in	+0.5 in
Rear ride height (x_2)	HTC	−1.0 in	+1.0 in
Yaw angle (x_3)	ETC	−3°	+1°
Grill tape (x_4)	ETC	0%	100%

NASCAR Wind Tunnel Experiments

Grill Tape (-1)

Grill Tape (+1)

NASCAR Wind Tunnel Experiments

- Resulting design is a *split-plot* design with 8 whole plots and 4 subplots per whole plot.
- Let's design and analyze the experiments in *Design Expert*.

wp	FRH	RRH	Yaw	Grill
1	-1	-1	-1	-1
	-1	-1	1	-1
	-1	-1	-1	1
	-1	-1	1	1
2	1	-1	-1	-1
	1	-1	1	-1
	1	-1	-1	1
	1	-1	1	1
3	-1	1	-1	-1
	-1	1	1	-1
	-1	1	-1	1
	-1	1	1	1
4	1	1	-1	-1
	1	1	1	-1
	1	1	-1	1
	1	1	1	1

wp	FRH	RRH	Yaw	Grill
5	-1	-1	-1	-1
	-1	-1	1	-1
	-1	-1	-1	1
	-1	-1	1	1
6	1	-1	-1	-1
	1	-1	1	-1
	1	-1	-1	1
	1	-1	1	1
7	-1	1	-1	-1
	-1	1	1	-1
	-1	1	-1	1
	-1	1	1	1
8	1	1	-1	-1
	1	1	1	-1
	1	1	-1	1
	1	1	1	1

NASCAR Wind Tunnel Experiments - Analysis

Fixed Effects [Type III]

Response 1: R1
Transform: Natural Log
Constant: 0

REML (REstricted Maximum Likelihood) analysis
Kenward-Roger p-values

Source	Term df	Error df	F-value	p-value	
Whole-plot	3	25.00	489.07	< 0.0001	significant
a-FRH	1	25.00	836.66	< 0.0001	
b-RRH	1	25.00	595.04	< 0.0001	
ab	1	25.00	35.51	< 0.0001	
Subplot	3	25.00	532.28	< 0.0001	significant
C-Yaw	1	25.00	141.05	< 0.0001	
D-Grill	1	25.00	1440.00	< 0.0001	
aC	1	25.00	15.80	0.0005	

Coefficients in Terms of Coded Factors

Source	Coefficient Estimate	Standard Error	VIF
Intercept	-0.1140	0.0023	
Whole-plot Terms:			
a-FRH	-0.0676	0.0023	1.0000
b-RRH	0.0570	0.0023	1.0000
ab	0.0139	0.0023	1.0000
Subplot Terms:			
C-Yaw	-0.0277	0.0023	1.0000
D-Grill	0.0886	0.0023	1.0000
aC	0.0093	0.0023	1.0000

Variance Components

Source	Variance	Standard Error	95% CI Low	95% CI High
Group	0.0000	0.0000	0.0000	0.0000
Residual	0.0002	0.0000	0.0001	0.0003
Total	0.0002			

Consequences of analyzing SPD as a CRD?

- If variability between whole plots exists, the standard errors of the factor effects are incorrect when analyzing a SPD as a CRD.
  - Inactive whole plot effects can be falsely deemed as active
  - Active subplot effects can be falsely deemed as inactive

Consequences of analyzing SPD as a CRD?

- Consider a 2^4 split-plot design with 2 HTC and 2 ETC factors.
- Simulated 3 replicates $\Rightarrow$ 12 whole-plots and 48 subplots
- Variance components: $\sigma_\delta^2 = 3.0$ and $\sigma_\epsilon^2 = 1.5$

WP	a	b	C	D
1	-1	-1	-1	-1
	-1	-1	1	-1
	-1	-1	-1	1
	-1	-1	1	1
2	1	-1	-1	-1
	1	-1	1	-1
	1	-1	-1	1
	1	-1	1	1
3	-1	1	-1	-1
	-1	1	1	-1
	-1	1	-1	1
	-1	1	1	1
4	1	1	-1	-1
	1	1	1	-1
	1	1	-1	1
	1	1	1	1

Consequences of analyzing SPD as a CRD?

Type of Effect	Coef	True Value	SPD P-Value	CRD P-Value
Whole Plot Effects	Int	100	0.0000	0.0000
	a	0	0.5004	0.2253
	b	0	0.5494	0.2818
	ab	0	0.2956	0.0581
Subplot Effects	C	1	0.0003	0.0019
	D	0	0.8259	0.8569
	aC	-1	0.0006	0.0034
	aD	0	0.6584	0.7167
	bC	0	0.3473	0.4392
	bD	0.5	0.0952	0.1664
	CD	1	0.0000	0.0000

True Model:

$E(y) = 100 + C - AC + 0.5(BD) + CD$

Consequences of analyzing SPD as a CRD?

- Bottom line: If between whole-plot variability exists, i.e., $\sigma_\delta^2 > 0$, then analyzing a split-plot design as a completely randomized design results in use of incorrect standard errors for computing the test statistics.

Coef	Estimate	SPD s.e.	CRD s.e.
Int	100.44	0.4158	0.2380
a	-0.2934	0.4158	0.2380
b	0.2599	0.4158	0.2380
ab	0.4653	0.4158	0.2380
C	0.7939	0.1948	0.2380
D	0.0432	0.1948	0.2380
aC	-0.7453	0.1948	0.2380
aD	-0.0870	0.1948	0.2380
bC	0.1861	0.1948	0.2380
bD	0.3359	0.1948	0.2380
CD	1.1892	0.1948	0.2380

Final Remarks

- Often times experiments are too difficult or expensive to run in a completely randomized fashion, leading to a split-plot treatment structure.
- Designs executed with randomization restrictions often require nonstandard analyses techniques.
  - If one ignores the correlation between observations due to the spit-plot nature of the design, one can easily be misled as to which factor effects may or may not be active.
- Fortunately, *Design Expert* has the capabilities to design and analyze split-plot designs, and integrates the use of data transformations throughout the SPD analysis phase.
  - Very practitioner-friendly!
  - Industry 4.0 readiness!

References

Meyers, R. H. and Montgomery, D. C. (1997). A Tutorial on Generalized Linear Models, *Journal of Quality Technology*, 29(3), 274-291.

Neyman, J. and Scott, E. L. (1960). "Correction for Bias Introduced by a Transformation of Variables". Annals of Mathematical Statistics 31, 643–655.

Land, C. E. (1974). "Confidence Interval Estimation for Means After Data Transformation to Normality". Journal of the American Statistical Association 69, pp. 795–802 (Correction, 71 p. 255).

Miller, D. M. (1984). "Reducing Transformation Bias in Curve Fitting". The American Statistician 38, pp. 124–126.

Montgomery, D. C. and Peck, E. A. (1992). Introduction to Linear Regression Analysis, 2nd ed. John Wiley & Sons, Inc., New York, NY.

References

- Perry, M. B. (2018), Approximate Prediction Intervals in the Original Units When Fitting Models Using Data Transformations, *Quality and Reliability Engineering International*, 34(1), 53-65.
- Perry, M. B. (2018), Prediction Intervals for the Original Response When Using Manly's Exponential Transformations, *Quality Engineering*, 30(2), pp. 195-211.
- Perry, M.B. and Walker, M.L. (2015). A Prediction Interval Estimator for the Original Response when using Box-Cox Transformations, *Journal of Quality Technology*, 47(3), 278-297.
- Khuri, A. I., Mukherjee, B. Sinha, B. K. and Ghosh, M. (2006). Design Issues for Generalized Linear Models: A Review, *Statistical Science*, 21(3), 376-399.
- Atkinson, A. C. and Woods, D. C. (2015). Designs for Generalized Linear Models. Preprint.
- Woods, D.C., Overstall, A. M., Adamou, M. and Waite, T.W. (2017). Bayesian Design of Experiments for Generalized Linear Models and Dimensional Analysis with Industrial and Scientific Application. *Quality Engineering*, 29(1), 91-103.

Thanks for Listening!