清华大学统计学辅修课程

Design and Analysis of Experiments

Lecture 7 –
Experiment Data Analysis via Regression

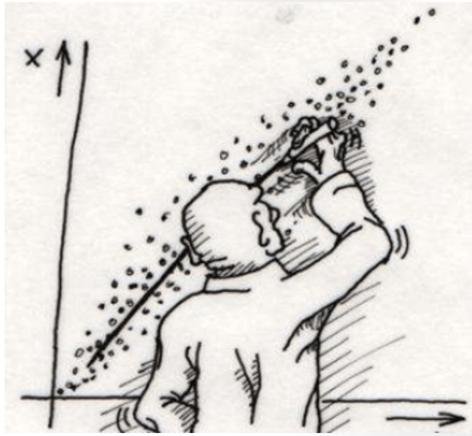
周在莹

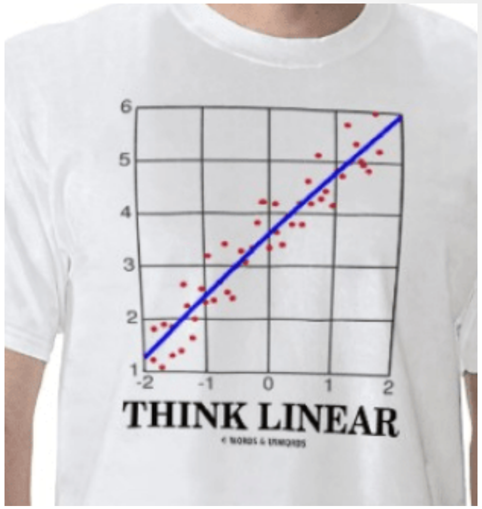
清华大学统计学研究中心

http://www.stat.tsinghua.edu.cn

Outline

- Experiments with a single quantitative factor
  - Regression model vs. fixed effects model
  - Model with non-linear terms
- One Quality Factor & One Quantitative Factor
  - Regression on each level of one factor

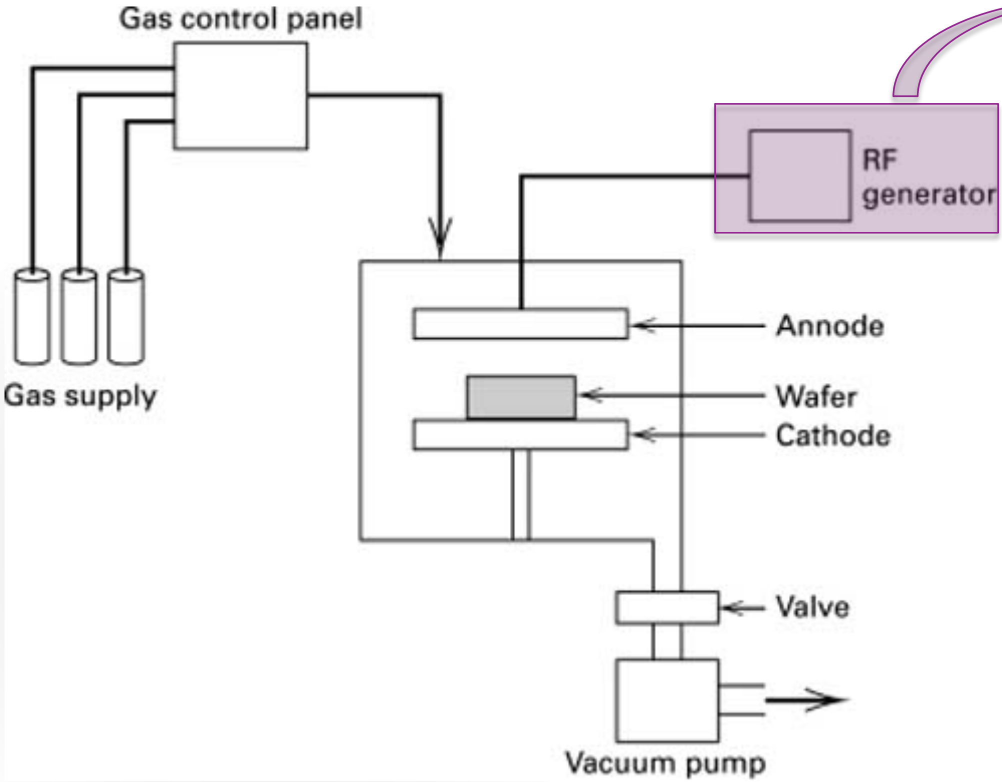




Experiments with a Single Quantitative Factor

- **Elements:**
  - Quantitative factor A with a levels
  - Response Y
- **Goals:**
  - Evaluate the impact of factor A to response Y
  - Find the best choice of A
- **Analysis strategies:**
  - Treat a levels of A as nominal variables (i.e., fixed effects)
  - Fit a regression model of A & Y

Single Factor Experiment: An Example



A single-wafer plasma etching tool

Test Sequence	Excel Random Number (Sorted)	Power
1	12417	200
2	18369	220
3	21238	220
4	24621	160
5	29337	160
6	32318	180
7	36481	200
8	40062	160
9	43289	180
10	49271	200
11	49813	220
12	52286	220
13	57102	160
14	63548	160
15	67710	220
16	71834	180
17	77216	180
18	84675	180
19	89323	200
20	94037	200

Data from the Experiment

Power (W)	Observations 1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

Etch Rate Data (in Å/min) from the Plasma Etching Experiment)

Fixed Effects Model (FEM)

- Statistical model

$$y_{ij} = \underbrace{\boxed{\mu}}_{\text{an overall mean}} + \overbrace{\boxed{\tau_i}}^{\text{Effect of the } i\text{-th factor}} + \underbrace{\boxed{\varepsilon_{ij}}}_{\text{Experimental error} \sim N(0,\sigma^2)}, \qquad i = 1, 2, \cdots, a, j = 1, 2, \cdots, n$$

$$\sum_{i=1}^{a} \tau_i = 0$$

- Equivalent regression model

$$y = X\beta + \varepsilon, \beta = \begin{pmatrix} \mu \\ \tau_1 \\ \cdots \\ \tau_a \end{pmatrix},$$

$$X = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 1
\end{pmatrix}$$

Linear Regression Model for the Example

Power (W)	Observations 1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

- Regression model with a different coding

$$y = X\beta + \varepsilon, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \end{pmatrix}$$

- Statistical model

Linear effect of factor A

$$y_{ij} = \boxed{\beta_0} + \boxed{\beta_1} x_i + \boxed{\varepsilon_{ij}}, \qquad i = 1,2,\cdots,a, j = 1,2,\cdots,n$$

Grand mean

Experimental error ~ $N(0,\sigma^2)$

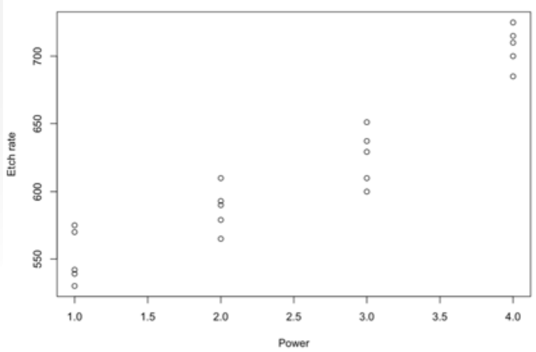
The data can be divided into 5 groups with same structure

3 levels are good enough to study a linear effect

- In the matrix form

$$y = X\beta + \varepsilon$$

$$X = \begin{pmatrix} 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \end{pmatrix}$$



Linear effect of factor A

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}, \begin{cases} i = 1, 2, \ldots, a \\ j = 1, 2, \ldots, n \end{cases}$$

Grand mean

Experimental error ~ $N(0,\sigma^2)$

Take a close look at group 1

$$X = \begin{pmatrix} 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 160 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 180 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 200 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \\ 1 & 220 \end{pmatrix}$$

Power (W)	Observations					Totals	Averages
	1	2	3	4	5		
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

Transform the Data for Better Performance

Power (W)	Observations				
	1	2	3	4	5
160	575	542	530	539	570
180	565	593	590	579	610
200	600	651	610	637	629
220	725	700	715	685	710

x	$x-\bar{x}$	$\frac{x-\bar{x}}{\Delta}$	
180	-20	-1	$-\frac{1}{\sqrt{2}}$
200	0	0	0
220	$+20$	$+1$	$+\frac{1}{\sqrt{2}}$

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}$$

$$X = \begin{pmatrix} 1 & 180 \\ 1 & 200 \\ 1 & 220 \end{pmatrix}$$

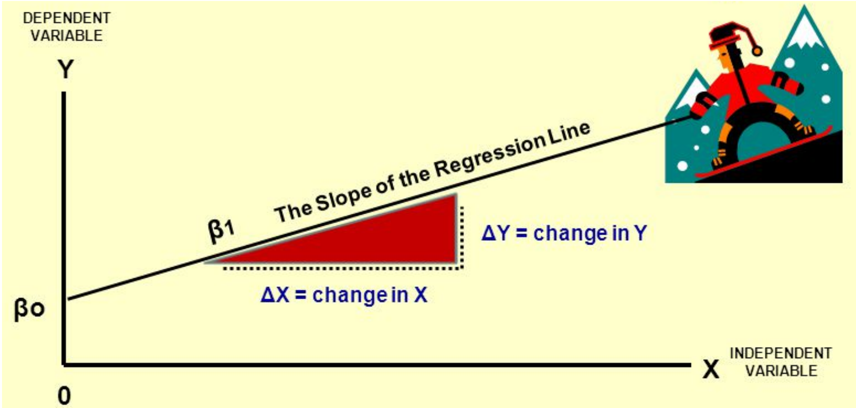
$$y_{ij} = \beta_0 + \beta_1\left(\frac{x_i - \bar{x}}{\Delta}\right) + \varepsilon_{ij}$$

$$X^* = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & +1 \end{pmatrix}$$

Orthogonal vectors

Estimation & Hypothesis Testing of Linear Model

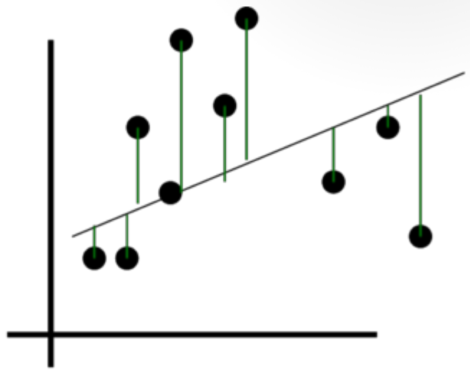
- Least square estimates:
  - $\hat{\beta} = (X^T X)^{-1} X^T y$
  - $\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y = Hy$
  - $e = y - \hat{y} = y - Hy$
  - $\hat{\sigma}^2 = \frac{e^T e}{n-p}$



- Hypothesis testing:
  - t-test for one coefficient

    $H_0: \beta_j = 0 \quad vs \quad H_1: \beta_j \neq 0, \quad j = 0 \text{ or } 1$
  - F-test for the whole model

    $H_0: \beta_0 = \beta_1 = 0 \quad vs \quad H_1: \beta_0^2 + \beta_1^2 \neq 0$



Regression Model vs. Fixed Effects Model

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}$$

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

Advantages	Limitations
▶ Fewer parameters ▶ More *df* to estimate σ^2 ▶ Needs no replicates ▶ Can be used for prediction	▶ Sensitive to linear assumption

# of free parameters in regression model is 3:

$\beta_0, \beta_1, \sigma^2$

# of free parameters in fixed effect model is $(a+1)$:

$\mu,\ \tau_1,\ \tau_2,\ \ldots,\ \tau_{a\text{-}1},\ \sigma^2$

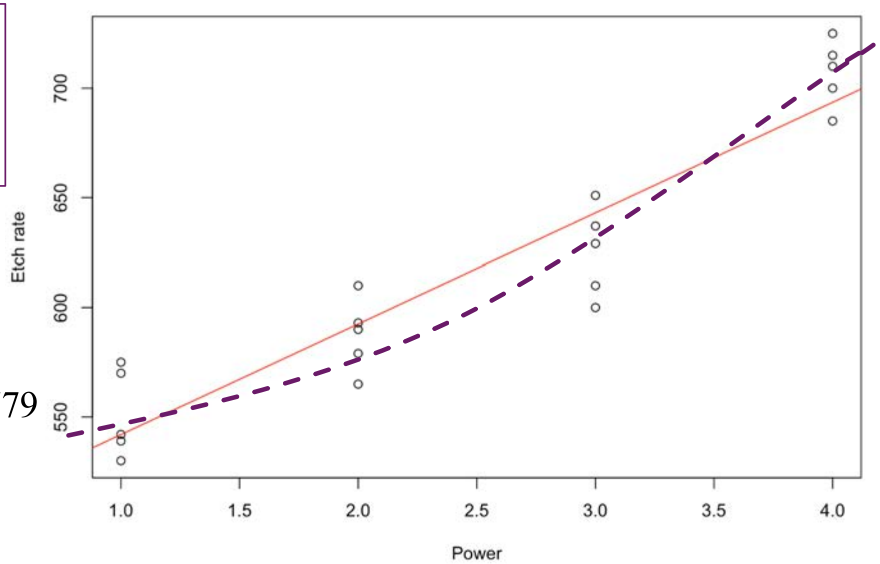
Fit the Etching Data via Linear Regression

Power (W)	Observations 1	2	3	4	5	Totals	Averages
160	575	542	530	539	570	2756	551.2
180	565	593	590	579	610	2937	587.4
200	600	651	610	637	629	3127	625.4
220	725	700	715	685	710	3535	707.0

Etch Rate Data (in Å/min) from the Plasma Etching Experiment)

```
> mod <- lm(Rate ~ as.numeric(Power),
                          data=Etch)
> summary(mod)
```

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  491.400     11.799   41.65  < 2e-16 ***
     x        50.540      4.308   11.73 7.26e-10 ***
Residual standard error: 21.54 on 18 degrees of freedom
Multiple R-squared:  0.8843,   Adjusted R-squared:  0.8779
F-statistic: 137.6 on 1 and 18 DF,  p-value: 7.263e-10
```

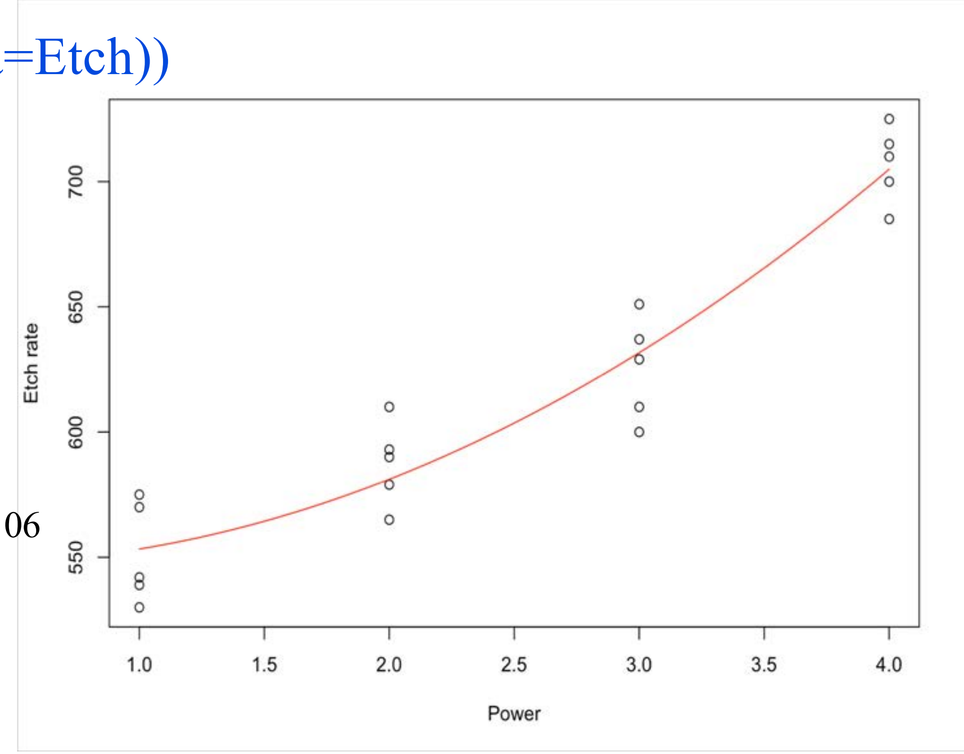


Adding a Quadratic Term

- > x <- as.numeric(Etch$Power); x2 <- x^2
- > summary(lm(Rate ~ x + x2, data=Etch))
- Or simply
- > summary(lm(Rate ~ x + I(x^2), data=Etch))

```
Coefficients:
             Estimate   Std. Error t value   Pr(>|t|)
(Intercept)  548.150    22.949    23.886    1.61e-14 ***
x             -6.210    20.936    -0.297    0.7703
x2            11.350     4.122     2.754    0.0136 *

Residual standard error: 18.43 on 17 degrees of freedom
Multiple R-squared:   0.92,        Adjusted R-squared:  0.9106
F-statistic: 97.76 on 2 and 17 DF,  p-value: 4.74e-10
```



Numerical Proof of Improvement

- We use the anova() function to further quantify the extent to which the quadratic fit is superior to the linear fit
- > x <- as.numeric(Etch$Power);
- > modf <- lm(Rate ~ x + I(x^2), data=Etch)
- > mod <- lm(Rate ~ x , data=Etch)
- > anova(mod, modf)

```
Analysis of Variance Table

Model 1: Rate ~ as.numeric(Power)
Model 2: Rate ~ x + I(x^2)
  Res.Df  RSS Df  Sum of Sq    F   Pr(>F)
1     18    8352
2     17    5776  1      2576      7.58  0.014 *
```

Note: Model Comparisons by R

- Regression Model

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}$$

- summary(lm(Rate ~ as.numeric(Power), data=Etch))

```
Coefficients:
                   Estimate  Std. Error  t value Pr(>|t|)
(Intercept)         491.400     11.799   41.65   < 2e-16 ***
as.numeric(Power)    50.540      4.308   11.73  7.26e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.54 on 18 degrees of freedom
Multiple R-squared:  0.8843,       Adjusted R-squared:  0.8779
F-statistic: 137.6 on 1 and 18 DF,  p-value: 7.263e-10
```

- Fixed Effect Model

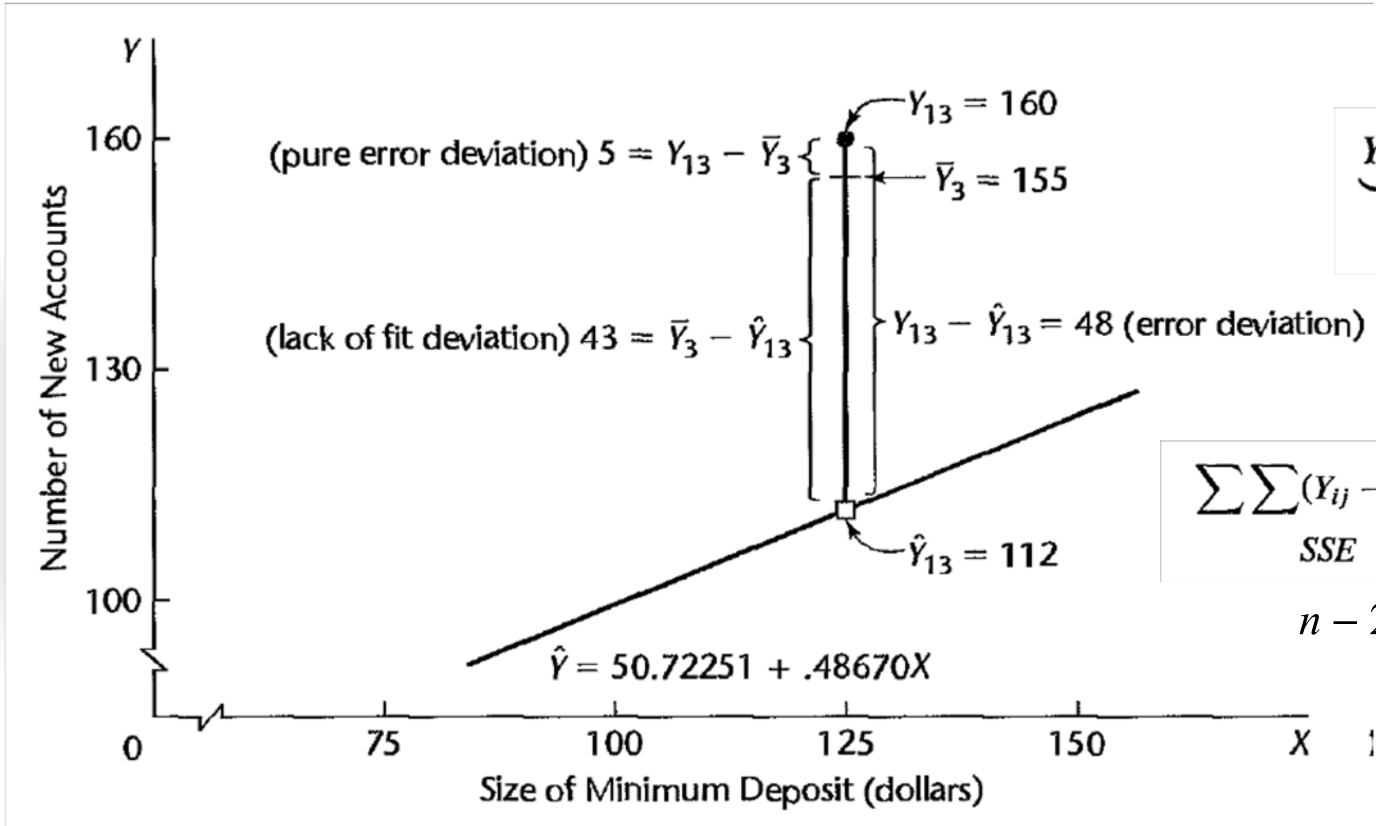
$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

- summary(lm(Rate ~ Power, data=Etch))

```
Coefficients:
            Estimate Std. Error  t value   Pr(>|t|)
(Intercept)  551.200      8.169  67.471  < 2e-16 ***
Power180      36.200     11.553   3.133  0.00642 **
Power200      74.200     11.553   6.422  8.44e-06 ***
Power220     155.800     11.553  13.485  3.73e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.27 on 16 degrees of freedom
Multiple R-squared:  0.9261,       Adjusted R-squared:  0.9122
F-statistic:  66.8 on 3 and 16 DF,  p-value: 2.883e-09
```

Illustration of Decomposition of Error Deviation



$$\underbrace{Y_{ij} - \hat{Y}_{ij}}_{\text{Error deviation}} = \underbrace{Y_{ij} - \bar{Y}_j}_{\text{Pure error deviation}} + \underbrace{\bar{Y}_j - \hat{Y}_{ij}}_{\text{Lack of fit deviation}}$$

$$\underbrace{\sum\sum(Y_{ij} - \hat{Y}_{ij})^2}_{SSE} = \underbrace{\sum\sum(Y_{ij} - \bar{Y}_j)^2}_{SSPE} + \underbrace{\sum\sum(\bar{Y}_j - \hat{Y}_{ij})^2}_{SSLF}$$

$$n - 2 = n - c + c - 2$$

ANOVA Table

- For testing lack of fit of simple linear regression function

Source of Variation	SS	df	MS
Regression	$SSR = \sum\sum(\hat{Y}_{ij} - \bar{Y})^2$	1	$MSR = \frac{SSR}{1}$
Error	$SSE = \sum\sum(Y_{ij} - \hat{Y}_{ij})^2$	$n-2$	$MSE = \frac{SSE}{n-2}$
Lack of fit	$SSLF = \sum\sum(\bar{Y}_j - \hat{Y}_{ij})^2$	$a-2$	$MSLF = \frac{SSLF}{a-2}$
Pure error	$SSPE = \sum\sum(Y_{ij} - \bar{Y}_j)^2$	$n-a$	$MSPE = \frac{SSPE}{n-a}$
Total	$SSTO = \sum\sum(Y_{ij} - \bar{Y})^2$	$n-1$	

Model with Non-linear Terms

- Naive model

$$y_{ij} = \underbrace{\beta_0}_{\text{Ground mean}} + \overbrace{\beta_1}^{\text{Linear effect}} x_i + \overbrace{\beta_2}^{\text{Quadratic effect}} x_i^2 + \underbrace{\varepsilon_{ij}}_{\text{Experimental error} \sim N(0,\sigma^2)}$$

- A better model with transformed data

$$y_{ij} = \beta_0 + \beta_1 x_i \left(\frac{x_i - \bar{x}}{\Delta}\right) + \beta_2 \left(\frac{x_i - \bar{x}}{\Delta}\right)^2 + \varepsilon_{ij}$$

Data Transformation in Quadratic Model

x	$x-\bar{x}$	$\frac{x-\bar{x}}{\Delta}$	$z=\left(\frac{x-\bar{x}}{\Delta}\right)^2$	$z-\bar{z}$
180	-20	-1	1	$+1/3$
200	0	0	0	$-2/3$
220	$+20$	$+1$	1	$+1/3$

$$y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_{ij}$$

$$\Rightarrow$$

$$y_{ij} = \beta_0 + \beta_1 \left(\frac{x_i - \bar{x}}{\Delta}\right) + \beta_2 \left(\frac{x_i - \bar{x}}{\Delta}\right)^2 + \varepsilon_{ij}$$

$$X = \begin{pmatrix} 1 & 180 & 180^2 \\ 1 & 200 & 200^2 \\ 1 & 220 & 220^2 \end{pmatrix}$$

$$X^* = \begin{pmatrix} 1 & -1 & +1 \\ 1 & 0 & -2 \\ 1 & +1 & +1 \end{pmatrix}$$

Orthogonal vectors

D-Optimal Design

Definition

- Regression model $y = X\beta + \varepsilon$
- Least square estimate $\hat{\beta} = (X^T X)^{-1} X^T y$

$$Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

- Let $D = |X^T X|$, we have $\left|Cov(\hat{\beta})\right| = \sigma^2 / D$
- Design X that maximizes D, and thus minimizes $\left|Cov(\hat{\beta})\right|$ is called the D-optimal design

D-Optimal Design for Simple Linear Regression

- Regression model：

$$y = \beta_0 + \beta_1 x + \varepsilon$$

- Assume $x \in [-1, 1]$
- D-optimal design is: half data points take -1, half take 1
- This can be easily proved based on the facts below:

$$X^T X = \begin{pmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \qquad D = |\mathbf{X}^T\mathbf{X}| = N\sum x_i^2 - \left(\sum x_i\right)^2$$

D-Optimal Design for Quadratic Regression

- Regression model:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$$

- Assume $x \in [-1, 1]$
- D-optimal design is: 1/3 take -1, 1/3 take 0, 1/3 take 1
- This can be proved similarly by checking:

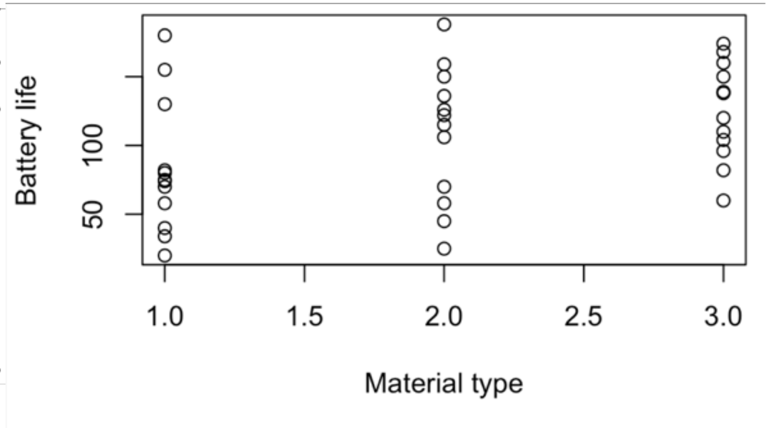
$$X^T X = \begin{pmatrix} N & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix}$$

One Quality Factor & One Quantitative Factor

- Elements:
  - Quality factor A with a levels
  - Quantitative factor B with b levels
  - Response Y
- Goals:
  - Evaluate the impact of factor A & B to response Y
  - Find the best combination of A & B
- Analysis strategies:
  - Treat levels of A & B as nominal variables (i.e., fixed effects)
  - Fit a regression model of B & Y for each level of A separately

The Battery Life Experiment: An Example

Material Type	Temperature (°F) 15		70		125	
1	130	155	34	40	20	70
	74	180	80	75	82	58
2	150	188	136	122	25	70
	159	126	106	115	58	45
3	138	110	174	120	96	104
	168	160	150	139	82	60

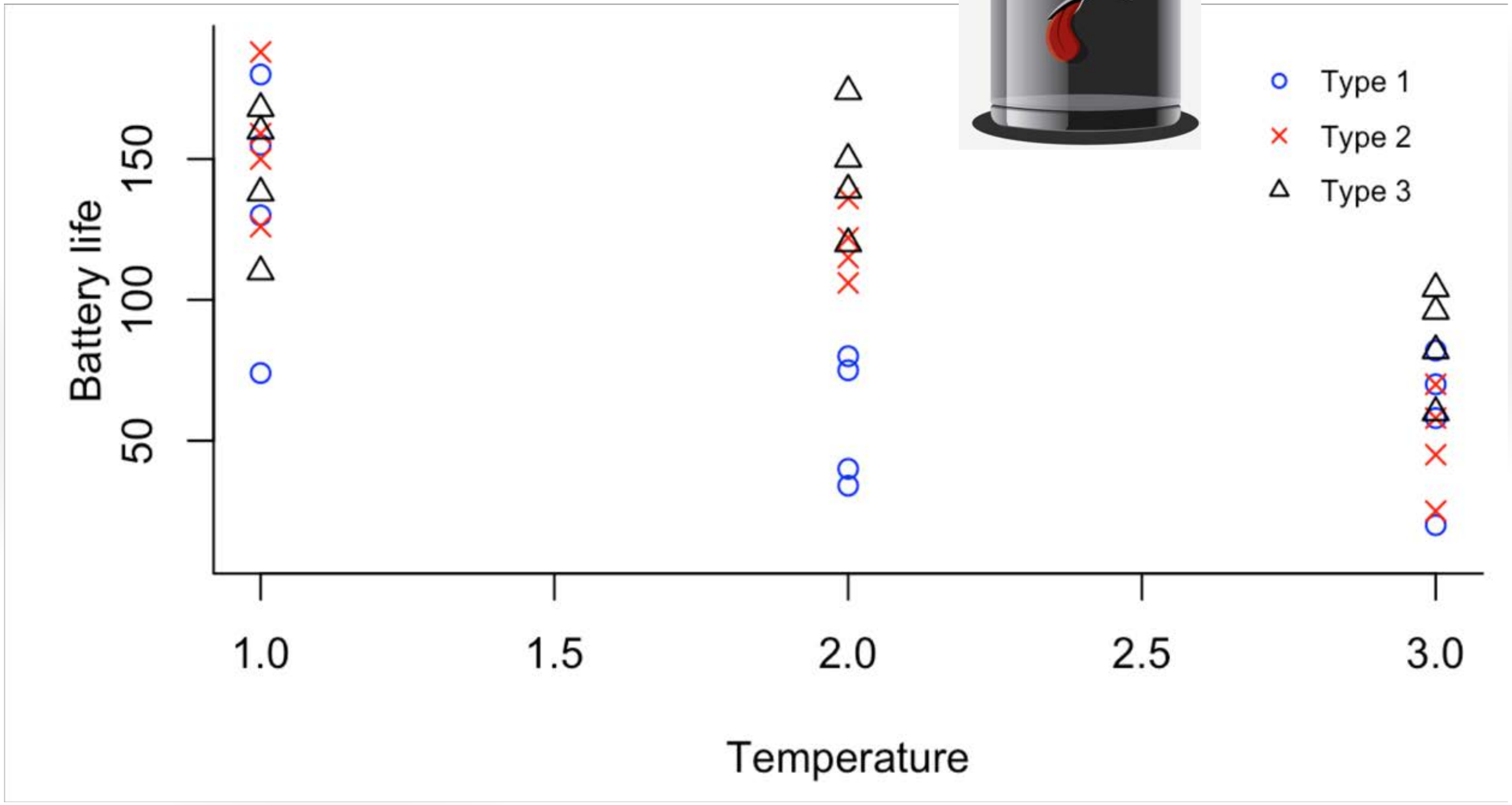


Factor A= Material type (a = 3); Factor B= Temperature (b = 3); n = 4

Q1.What **effects** do material type & temperature have on life?
Q2. Is there a choice of material that would give long life ***regardless of temperature***(a **robust** product)?

Linear Regression done by R

- See the data grouped by each level of *A*



Linear Regression done by R

- > mod <- lm(Life ~ Type, data= Battery)
- > summary(mod)

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   83.17      13.00   6.396 3.03e-07 ***
Type2         25.17      18.39   1.368   0.1804
Type3         41.92      18.39   2.279   0.0292 *

Residual standard error: 45.05 on 33 degrees of freedom
Multiple R-squared:  0.1376,        Adjusted R-squared:
0.08533
F-statistic: 2.633 on 2 and 33 DF,  p-value: 0.08695
```

- > with(Battery, tapply(Life, Type, mean))

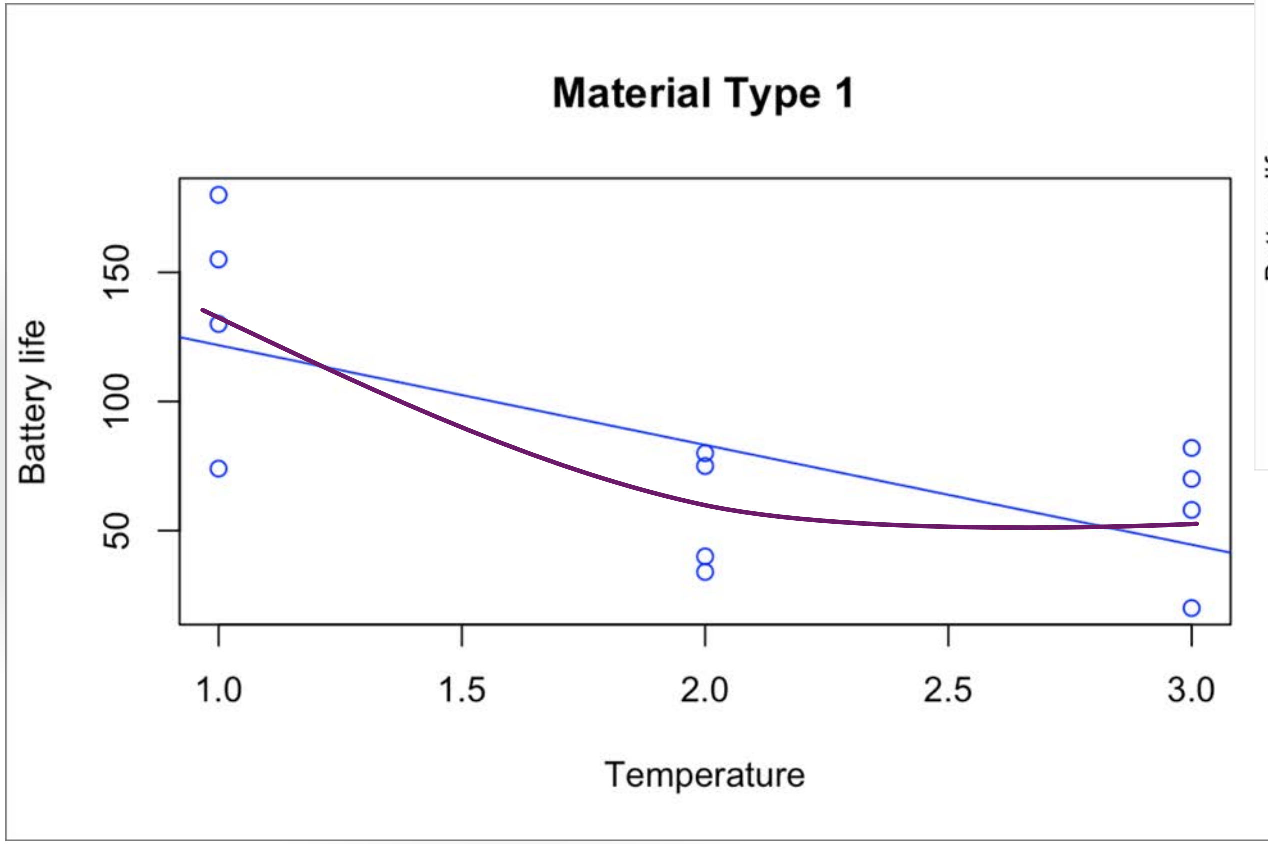
$\hat{\mu}_1 = 83.17$

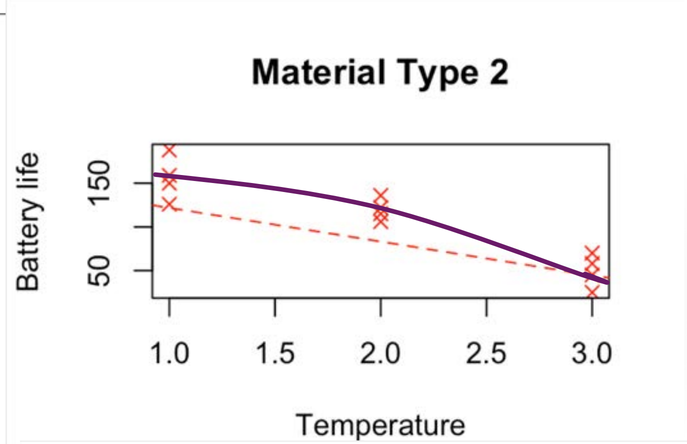
$\hat{\mu}_2 = 83.17 + 25.17$

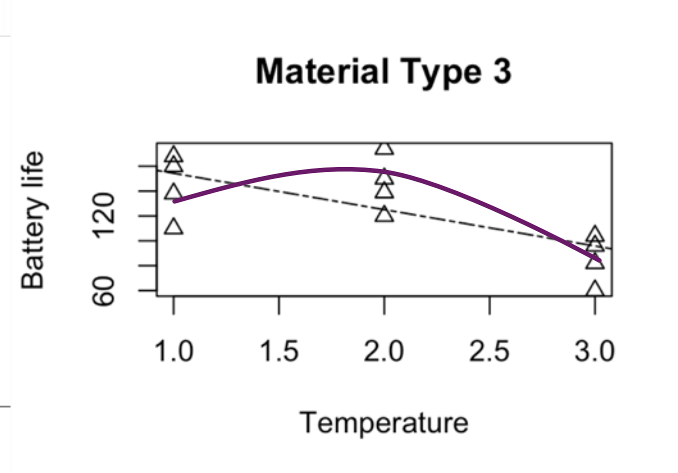
$\hat{\mu}_3 = 83.17 + 41.92$

Linear Regression done by R: Separate Lines

- > mod1 <- lm(Life ~ as.numeric(Temp), data=Battery[Type==1,])

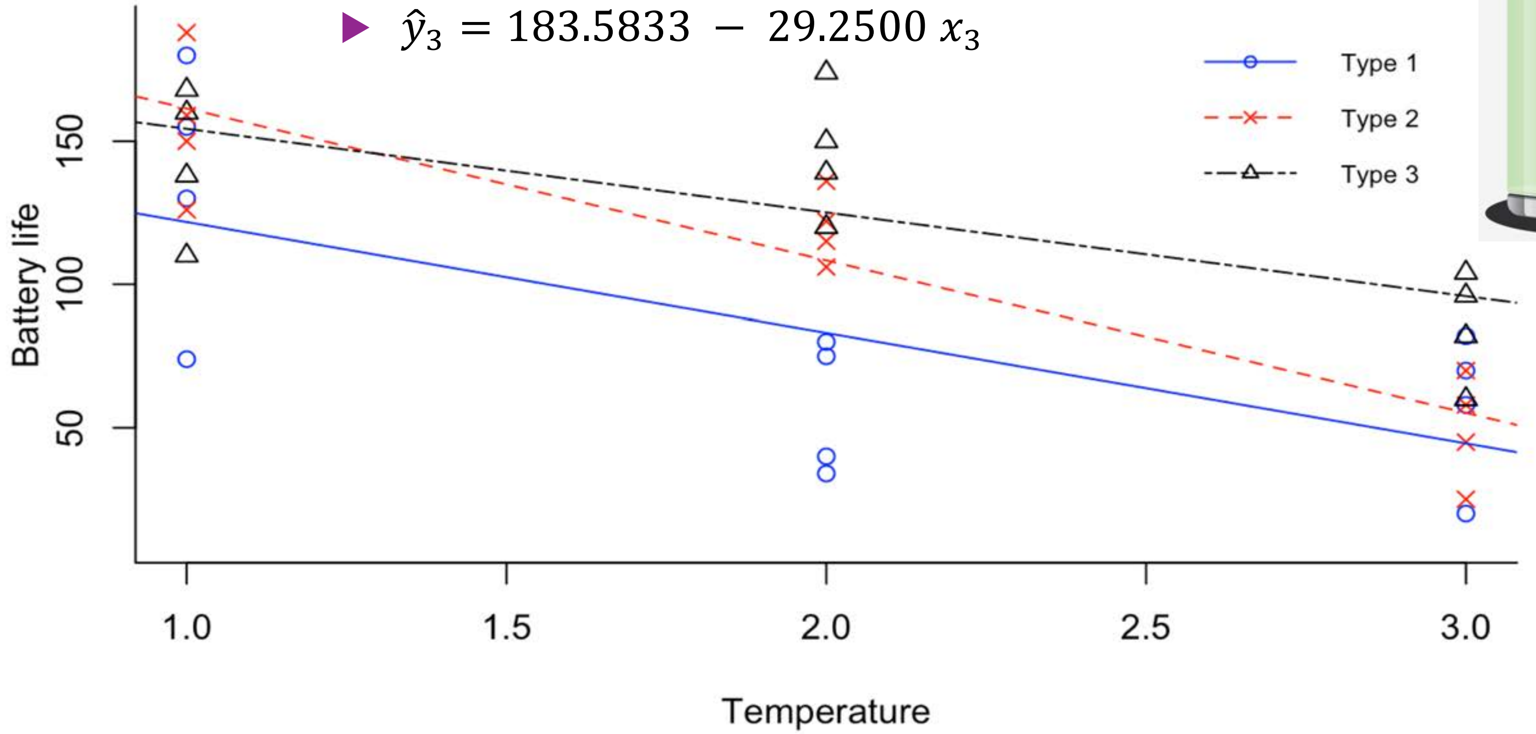






Linear Regression by R: All 3 in 1 Plot

- $\hat{y}_1 = 160.4167 - 38.6250\, x_1$
- $\hat{y}_2 = 214.5833 - 53.1250\, x_2$
- $\hat{y}_3 = 183.5833 - 29.2500\, x_3$



Quadratic Regression by R

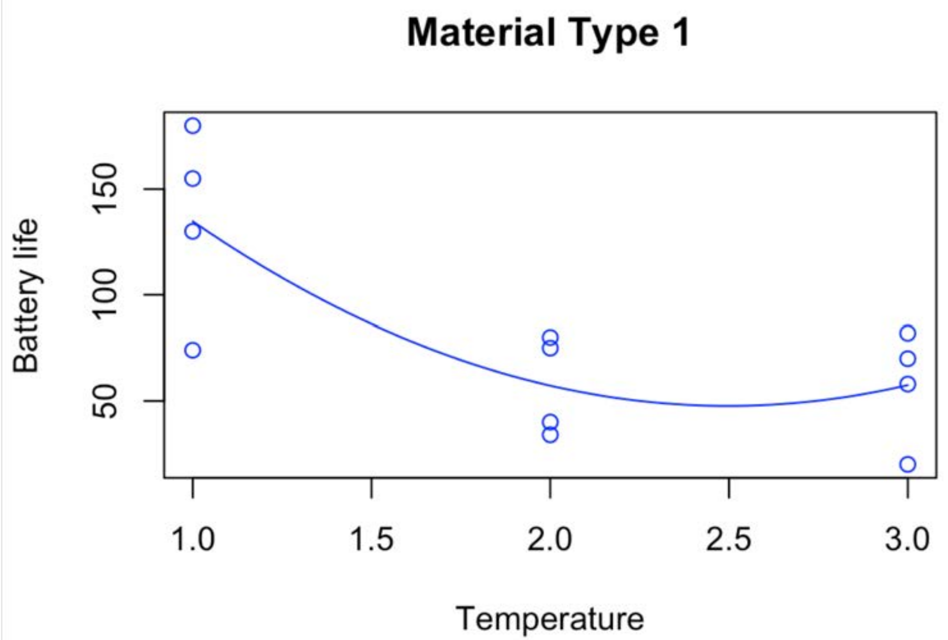
- $\hat{y}_1 = 57.2 - 109.2x_1 + 311.0x_1^2$

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	160.4	28.6	5.61	0.00023	***
x	-38.6	13.2	-2.92	0.01540	*

Residual standard error: 37.5 on 10 degrees of freedom
Multiple R-squared: 0.46, Adjusted R-squared: 0.406
F-statistic: 8.5 on 1 and 10 DF, p-value: 0.0154

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	57.2	16.7	3.43	0.0075	**
x	-109.2	33.3	-3.28	0.0096	**
I(x^2)	311.0	163.3	1.90	0.0893	.

Residual standard error: 33.3 on 9 degrees of freedom
Multiple R-squared: 0.615, Adjusted R-squared: 0.529
F-statistic: 7.18 on 2 and 9 DF, p-value: 0.0137



Full Regression Model

$$y_{ijk} = \beta_{i0} + \beta_{i1}x_{ijk} + \beta_{i2}x_{ijk}^2 + \varepsilon_{ijk}$$

- Separate regressions on each of the *a* levels of factor *A*

$$\begin{aligned} y_{1jk} &= \boxed{\beta_{10}} + \boxed{\beta_{11}}x_{1jk} + \boxed{\beta_{12}}x_{1jk}^2 + \boxed{\varepsilon_{1jk}} \\ y_{2jk} &= \beta_{20} + \beta_{21}x_{2jk} + \beta_{22}x_{2jk}^2 + \varepsilon_{2jk} \\ &\cdots \\ y_{ajk} &= \beta_{a0} + \beta_{31}x_{ajk} + \beta_{a2}x_{ajk}^2 + \varepsilon_{ajk} \end{aligned}$$

$$\hat{y}_1 = 83.17$$
$$\hat{y}_2 = 83.17 + 25.17$$
$$\hat{y}_3 = 83.17 + 41.92$$

? Main effects of A | Linear & quadratic effects of B | Experimental error ~ $N(0,\sigma^2)$

- Parameters: $(\vec{\beta}_1, \vec{\beta}_2, \cdots, \vec{\beta}_a, \sigma^2)$ $\quad \vec{\beta}_i = (\beta_{i0}, \beta_{i1}, \beta_{i2})$
- #Parameters: $3 \times a + 1 < a \times b$

$$\hat{y}_1 = 160.4167 - 38.6250\, x_1$$
$$\hat{y}_2 = 214.5833 - 53.1250\, x_2$$
$$\hat{y}_3 = 183.5833 - 29.2500\, x_3$$

- $\hat{y}_1 = 57.2 - 109.2x_1 + 311.0x_1^2$

Note: Interpreting the Intercept

- > data1 <- Battery[Type==1,]
- > data1$x <- with(data1,as.numeric(Temp)-mean(as.numeric(Temp)))
- > mod11 <- lm(Life~ x, data=data1)
- > summary(mod11)

$\hat{y}_1 = 83.17$

- Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	83.17	10.81	7.690	1.66e-05 ***
x	-38.62	13.25	-2.916	0.0154 *

Interpreting the Intercept

- $\hat{y}_1 = 57.2 - 109.2x_1 + 311.0x_1^2$

- > summary(lm(Life ~ poly(x, 2), data=Battery[Type==1,]))
- > summary(lm(Life ~ x + I(x^2), data=Battery[Type==1,]))

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	83.17	9.62	8.64	1.2e-05	***
poly(x, 2)1	-109.25	33.34	-3.28	0.0096	**
poly(x, 2)2	63.48	33.34	1.90	0.0893	.

Residual standard error: 33.3 on 9 degrees of freedom
Multiple R-squared: 0.615, Adjusted R-squared: 0.529
F-statistic: 7.18 on 2 and 9 DF, p-value: 0.0137

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	57.2	16.7	3.43	0.0075	**
x	-109.2	33.3	-3.28	0.0096	**
I(x^2)	311.0	163.3	1.90	0.0893	.

Residual standard error: 33.3 on 9 degrees of freedom
Multiple R-squared: 0.615, Adjusted R-squared: 0.529
F-statistic: 7.18 on 2 and 9 DF, p-value: 0.0137

Test Various Effects

$$y_{ijk} = \beta_{i0} + \beta_{i1}x_{ijk} + \beta_{i2}x_{ijk}^2 + \varepsilon_{ijk}$$

- Test main effects of factor A:

  H_0: $\beta_{10} = \beta_{20} = \cdots = \beta_{a0} = 0$ *vs* H_1: any of them $\neq 0$

- Test linear effects of factor B:

  H_0: $\beta_{11} = \beta_{21} = \cdots = \beta_{a1} = 0$ *vs* H_1: any of them $\neq 0$

- Test quadratic effects of factor B:

  H_0: $\beta_{12} = \beta_{22} = \cdots = \beta_{a2} = 0$ *vs* H_1: any of them $\neq 0$

- Test interaction of factor A & B:

  H_0: $\beta_{11} = \beta_{21} = \cdots = \beta_{a1} = \beta_1$ *vs* H_1: any pair are different

  H_0: $\beta_{12} = \beta_{22} = \cdots = \beta_{a2} = \beta_2$ *vs* H_1: any pair are different

$$\begin{aligned} y_{1jk} &= \beta_{10} + \beta_{11}x_{1jk} + \beta_{12}x_{1jk}^2 + \varepsilon_{1jk} \\ y_{2jk} &= \beta_{20} + \beta_{21}x_{2jk} + \beta_{22}x_{2jk}^2 + \varepsilon_{2jk} \\ &\cdots \\ y_{ajk} &= \beta_{a0} + \beta_{31}x_{ajk} + \beta_{a2}x_{ajk}^2 + \varepsilon_{ajk} \end{aligned}$$

<u>The hierarchy principle</u> indicates that if a model contains a high-order term, it should also contain all of the lower order terms that compose it

R Code

- full <- lm(Life ~ Type * poly(as.numeric(Temp), 2), Battery)
- r1 <- lm(Life ~ poly(as.numeric(Temp), 2), Battery)
- r2 <- lm(Life ~ Type * I(as.numeric(Temp)^ 2), Battery)
- r3 <- lm(Life ~ Type * as.numeric(Temp), Battery)
- r4 <- lm(Life ~ Type + poly(as.numeric(Temp), 2), Battery)
- anova(r1, full)
- anova(r2, full)
- anova(r3, full)
- anova(r4, full)