Minimize risk. Maximize potential.
ICEAA Technology Showcase Webinar Series
Modeling OLS-based CER Errors

Dr. Steve Van Drew
March 24, 2021
Webinar Topics

• Regression Fundamentals
  – OLS assumptions
  – useful Excel regression functions and tool

• Linear CER Demonstration
  – Excel regression analysis
  – prediction interval
  – @RISK implementation

• Nonlinear (intrinsically linear) Regression
  – transformations
  – lognormal error term

• Power CER Demonstration
  – Excel regression analysis
  – @RISK implementation
Ordinary Least Squares (OLS) Regression Fundamentals

\[ Y = a + b \, X + \varepsilon \]

where:
- \( a \) = Y intercept
- \( b \) = slope
- \( \varepsilon \) = error / residual

Assumptions about \( \varepsilon \):
- normally distributed
- mean = 0
- constant variance
- individual \( \varepsilon_i \) independent

Linear w/ Error Term

\[ Y = a + bX + \varepsilon \]
Useful Excel Regression Functions and Tool

• Functions
  – INTERCEPT(known_y's, known_x's)
  – SLOPE(known_y's, known_x's)
  – LINEST(known_y's, [known_x's], [const], [stats])
  – FORECAST.LINEAR(x, known_y's, known_x's)

• Tool
  – Data → Data Analysis → Regression
Cost = f(Weight)

\[ y = 0.0831x + 2.4765 \]

\[ R^2 = 0.3251 \]

### Linear CER Data Set & XY Scatter Plot

<table>
<thead>
<tr>
<th>System</th>
<th>Cost ($K)</th>
<th>Weight (lbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.2</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>17.3</td>
<td>161</td>
</tr>
<tr>
<td>3</td>
<td>11.8</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>9.6</td>
<td>108</td>
</tr>
<tr>
<td>5</td>
<td>8.8</td>
<td>82</td>
</tr>
<tr>
<td>6</td>
<td>7.6</td>
<td>135</td>
</tr>
<tr>
<td>7</td>
<td>6.8</td>
<td>59</td>
</tr>
<tr>
<td>8</td>
<td>3.2</td>
<td>68</td>
</tr>
<tr>
<td>9</td>
<td>1.7</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>1.6</td>
<td>24</td>
</tr>
</tbody>
</table>

Possible outlier left in for demonstration purposes only.
Estimated values for $a (2.4765)$ and $b (0.0831)$

Cost ($\textdollar K$) = 2.4765 + 0.0831 * Weight (lbs)

The estimate of $\sigma_\varepsilon$ is statistically significant at $\alpha = 0.10$.
Prediction Interval

\[
\hat{y} \pm t(\alpha / 2, n - 2)S_e \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}
\]
Linear CER @RISK Implementation

Cost ($K) = 2.4765 + (.0831 * Weight) + \epsilon
\epsilon \sim\text{Normal}\ (0,\ 5.8076)

CAUTION: Almost 6% of simulated costs came in less than zero! Could use output filter to discard simulated negative costs.
Intrinsically Linear Transformations

**Linear**

\[ Y = a + bX \]

**Power**

\[ Y = aX^b \]

**Exponential**

\[ Y = ae^{bx} \]

\[ \ln Y = \ln a + b \ln X \]

\[ \ln Y = \ln a + bX \]
ε’ in power model is a multiplicative error term that is lognormally distributed, but in the transformed model ε is additive and normally distributed with same characteristics assumed as linear regression error term.

Power Model Basics

Power Model

\[ Y = aX^b \varepsilon' \]

Transformed Model

\[ \ln Y = \ln a + b \ln X + \varepsilon \]
if $Y \sim N(\mu, \sigma)$ then $X = e^Y \sim \ln(\mu, \sigma)$  
$E(X) = e^{\mu + \sigma^2/2}$; $V(X) = [e^{\sigma^2}(2\mu + \sigma^2)](e^{\sigma^2}-1)$  

to model in @RISK use  
RiskLognorm2 $(\mu, \sigma)$ or RiskLognorm $(E(X), \{V(X)\}^{0.5})$  

Lognorm2(0,.6917) $\equiv$ Lognorm(1.2703,.9950)
Power Model, Cost = f (Power Out)

Power CER Data Set, XY & InX/InY Scatter Plots

<table>
<thead>
<tr>
<th>System</th>
<th>Cost ($K)</th>
<th>Power Out (watts)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22.2</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>17.3</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>11.8</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>9.6</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>8.8</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>7.6</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>6.8</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>3.2</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>1.7</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>1.6</td>
<td>0.5</td>
</tr>
</tbody>
</table>

possible outliers left in for demonstration purposes only

Power Model, Cost = f (Power Out)

![Graph showing Power Model, Cost vs Power Out](image)

![Graph showing In Cost vs In Power Out](image)
Transformed CER
\[ \text{ln Cost (\$K)} = 0.9513 + 0.3519 \times \text{ln(Power Out)} \]

Power CER
\[ \text{Cost (\$K)} = 2.5892 \times \text{Power Out}^{0.3519} \]

Estimated values for \( \ln a \) (.9513) and \( b \) (.3519)
\[ a = e^{\ln a} = e^{0.9513} = 2.5892 \]

Statistically significant at \( \alpha = .05 \)

Estimate of \( \sigma_e \) is Standard Error = Root Mean Square Error

Poor \( R^2 \)
Power CER @RISK Implementation

Cost ($K) = 2.5892 * (Power Out ^ 0.3519) * ε’
ε’ ~ Lognorm2(0, 0.6917)

In this case ignoring the multiplicative error term would result in cost estimates that are 27% low on average since ε’ has a mean of 1.2703.
Questions?

https://www.palisade.com/trials.asp

sales@palisade.com
Thank You!

Steven L. Van Drew, PhD, PE
svandrew@palisade.com
Back Up Slides
if $Y \sim N(\mu, \sigma)$ then $X = e^Y \sim \ln(\mu, \sigma)$

$E(X) = e^{\lambda(\mu+\sigma^2/2)}; V(X) = [e^{\lambda(2\mu+\sigma^2)}](e^{\sigma^2}-1)$

to model in @RISK use
RiskLognorm2 ($\mu, \sigma$) or RiskLognorm ($E(X), \{V(X)\}^{1/2}$)

$E(X) = e^{\mu + \frac{\sigma^2}{2}} = e^{0 + \frac{.6917^2}{2}} = 1.2703;$

$V(X) = \left( e^{2\mu + \sigma^2} \right) \left( e^{\sigma^2} - 1 \right) = \left( e^{2(0) + .6917^2} \right) \left( e^{.6917^2} - 1 \right) = .9900;$

$\sqrt{V(X)} = \sqrt{.9900} = .9950$

Lognorm2(0,.6917) ≡ Lognorm(1.2703,.9950)
Power Model, Cost = f (Power Out)

Cost ($K)
- Predicted Cost ($K)
- 90% Prediction Interval (lower)
- 90% Prediction Interval (upper)