Palisade

Minimize risk. Maximize potential.

ICEAA Technology Showcase Webinar Series
Modeling OLS-based CER Errors

Dr. Steve Van Drew

March 24, 2021

Palisade

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

Palisade

Ordinary Least Squares (OLS) Regression Fundamentals

$Y = a + b X + \varepsilon$

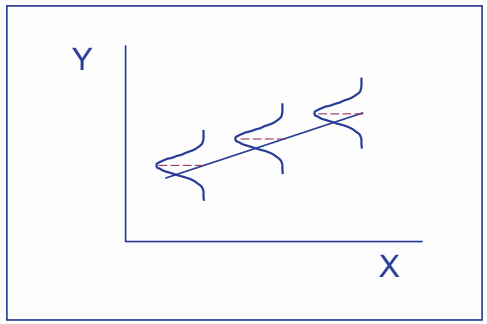
where:

- a = Y intercept
- b = slope
- ε = error / residual

Assumptions about ε:

- normally distributed
- mean = 0
- constant variance
- individual ε_i independent

Linear w/ Error Term

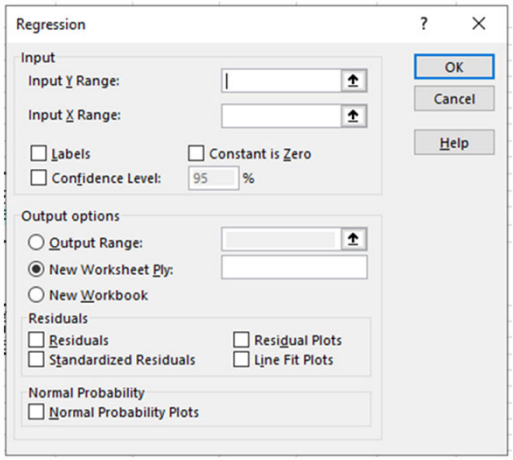


$$Y = a + bX + \varepsilon$$

Palisade

Useful Excel Regression Functions and Tool

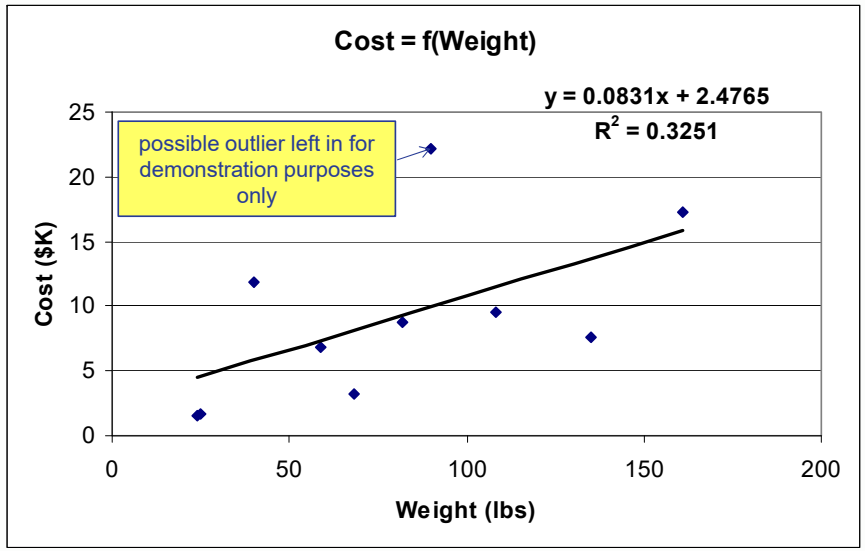
- Functions
  - INTERCEPT(known_y's, known_x's)
  - SLOPE(known_y's, known_x's)
  - LINEST(known_y's, [known_x's], [const], [stats])
  - FORECAST.LINEAR(x, known_y's, known_x's)
- Tool
  - Data → Data Analysis → Regression



Palisade

Linear CER Data Set & XY Scatter Plot

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



Palisade

Linear CER Excel Regression Output

	I	J	K	L	M	N	O	P	Q
3	SUMMARY OUTPUT								
4									
5	*Regression Statistics*								
6	Multiple R	0.570207654							
7	R Square	0.325136769							
8	Adjusted R Square	0.240778865							
9	Standard Error	5.807608335							
10	Observations	10							
11									
12	ANOVA								
13		*df*	*SS*	*MS*	*F*	*Significance F*			
14	Regression	1	129.9974834	129.9974834	3.854253766	0.08523197			
15	Residual	8	269.8265166	33.72831457					
16	Total	9	399.824						
17									
18		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
19	Intercept	2.476502105	3.823373453	0.647726971	0.535303869	-6.34021288	11.29321709	-6.34021288	11.29321709
20	Weight (lbs)	0.083124973	0.042341025	1.963225348	0.08523197	-0.014513605	0.180763552	-0.014513605	0.180763552

Poor R^2

Estimate of σ_ε is Standard Error = Root Mean Square Error

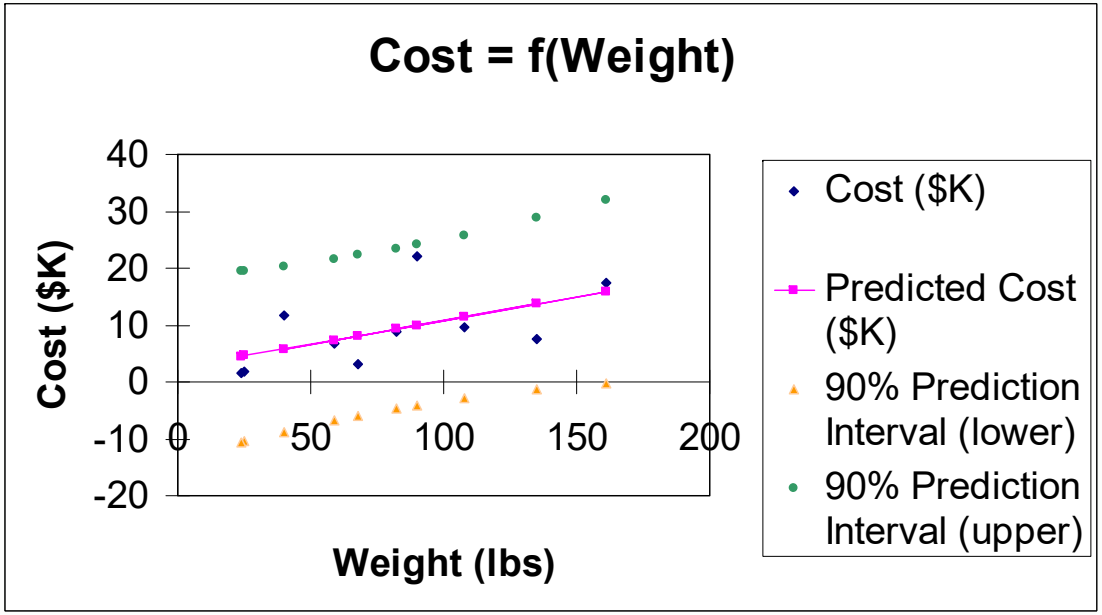
statistically significant at $\alpha = .10$

Estimated values for a (2.4765) and b (.0831)

Cost ($K) = 2.4765 + .0831 * Weight (lbs)

Palisade

Prediction Interval



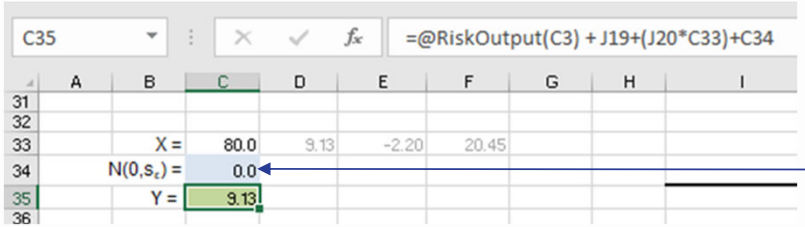
$$\hat{y} \pm t_{(\alpha/2,\, n-2)} S_{\varepsilon} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}$$

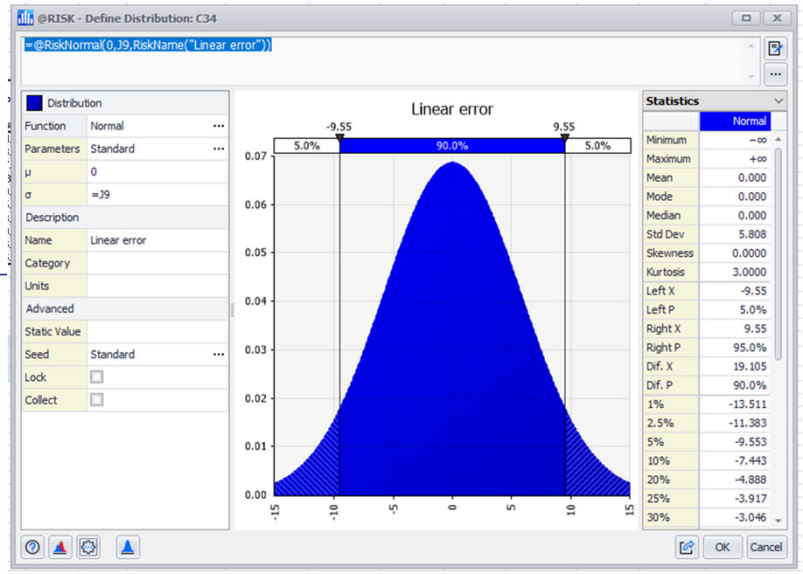
Palisade

Linear CER @RISK Implementation

Cost ($K) = 2.4765 + (.0831 * Weight) + ε

ε ~ Normal (0, 5.8076)

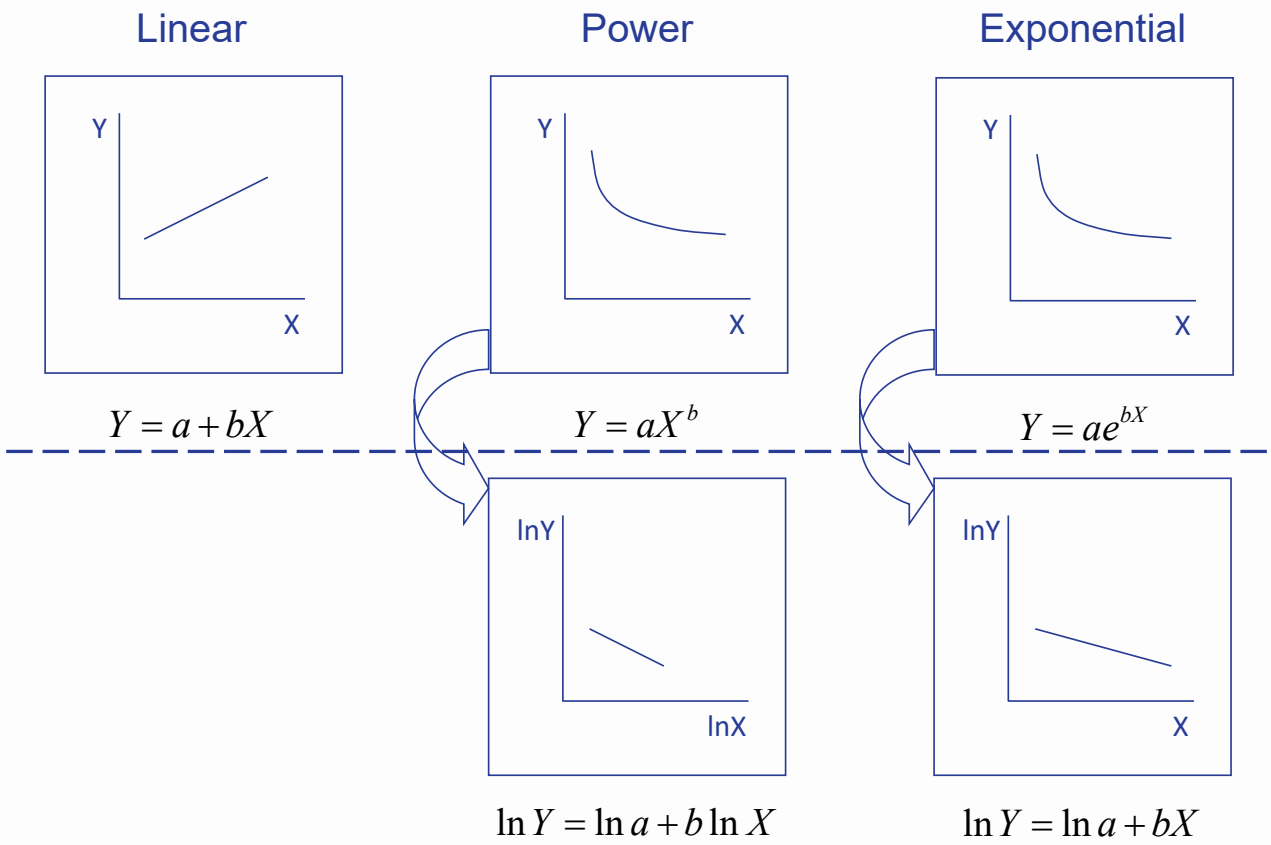




CAUTION: Almost 6% of simulated costs came in less than zero!
Could use output filter to discard simulated negative costs.

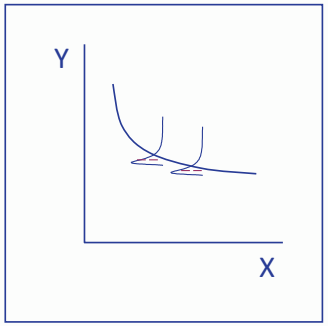
Palisade

Intrinsically Linear Transformations



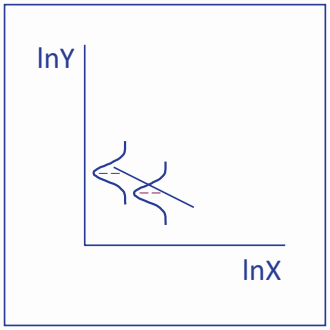
Power Model Basics

Power Model



$$Y = aX^b \varepsilon'$$

Transformed Model



$$\ln Y = \ln a + b \ln X + \varepsilon$$

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

Palisade

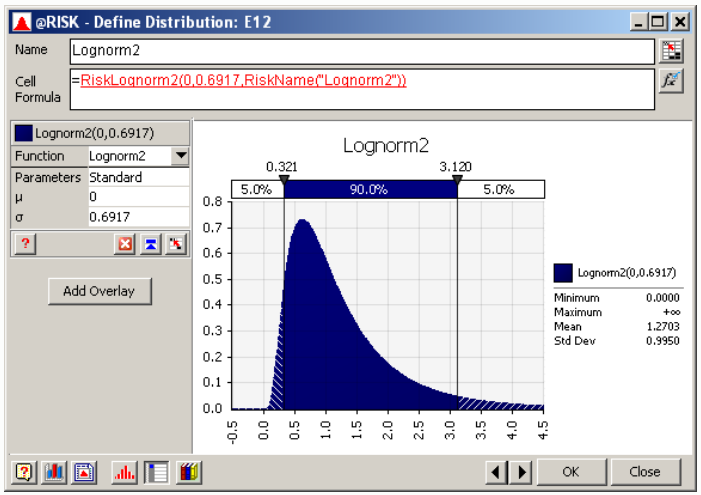
Lognormal Error Term

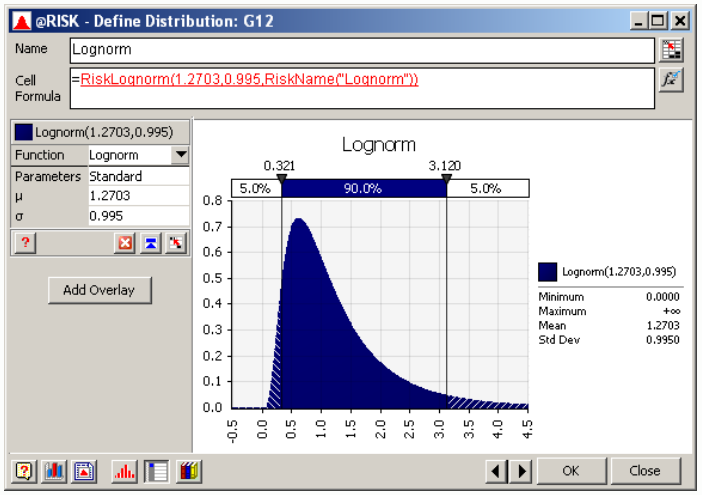
if $Y \sim N(\mu,\sigma)$ then $X = e^Y \sim \mathit{ln}(\mu,\sigma)$

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

to model in @RISK use

RiskLognorm2 (μ,σ) or RiskLognorm $(E(X), \{V(X)\}^{.5})$



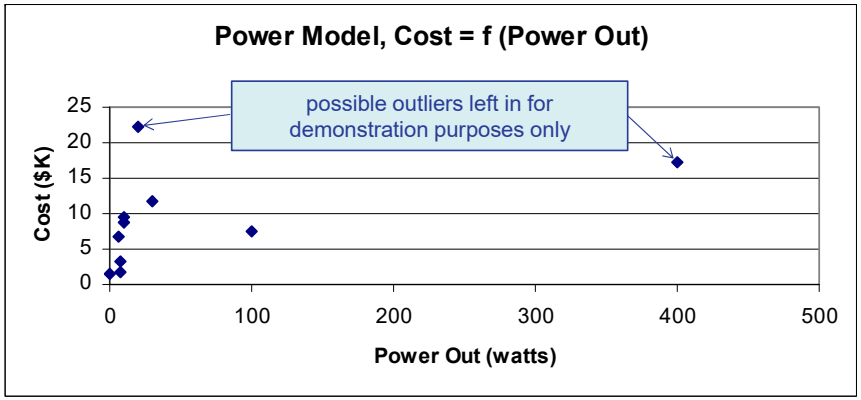


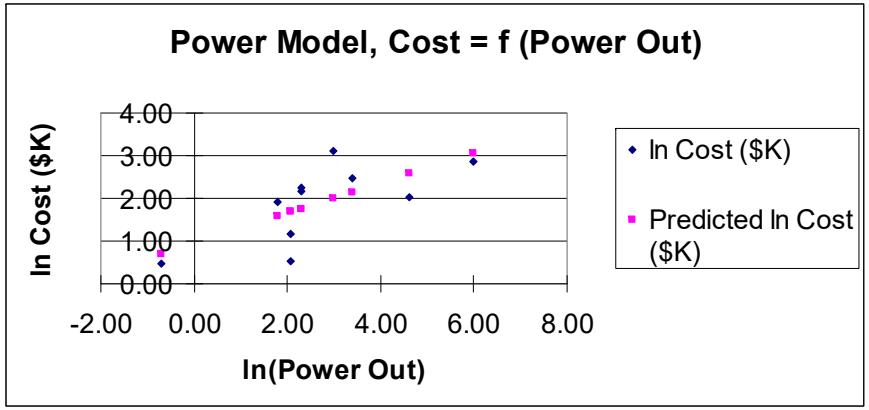
Lognorm2(0,.6917) ≡ Lognorm(1.2703,.9950)

Palisade

Power CER Data Set, XY & *lnX lnY* Scatter Plots

System	Cost ($K)	Power Out (watts)
1	22.2	20
2	17.3	400
3	11.8	30
4	9.6	10
5	8.8	10
6	7.6	100
7	6.8	6
8	3.2	8
9	1.7	8
10	1.6	0.5





Palisade

Power CER Excel Regression Output

	I	J	K	L	M	N	O	P	Q
3	SUMMARY OUTPUT								
4									
5	*Regression Statistics*								
6	Multiple R	0.692400707							
7	R Square	0.479418739							
8	Adjusted R Square	0.414346082							
9	Standard Error	0.691710586							
10	Observations	10							
11									
12	ANOVA								
13		*df*	*SS*	*MS*	*F*	*Significance F*			
14	Regression	1	3.525050199	3.525050199	7.367437523	0.026481169			
15	Residual	8	3.827708278	0.478463535					
16	Total	9	7.352758477						
17									
18		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
19	Intercept	0.95133719	0.411195499	2.31358853	0.049411509	0.00311867	1.89955571	0.00311867	1.89955571
20	ln(Power Out)	0.351906981	0.129649143	2.714302401	0.026481169	0.052935521	0.650878441	0.052935521	0.650878441

Poor R^2

Estimate of σ_ε is Standard Error = Root Mean Square Error

statistically significant at $\alpha = .05$

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

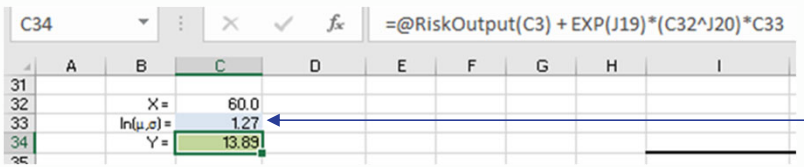
Transformed CER
ln Cost ($K) = .9513 + .3519 * *ln*(Power Out)
Power CER
Cost ($K) = 2.5892 * Power Out $^{.3519}$

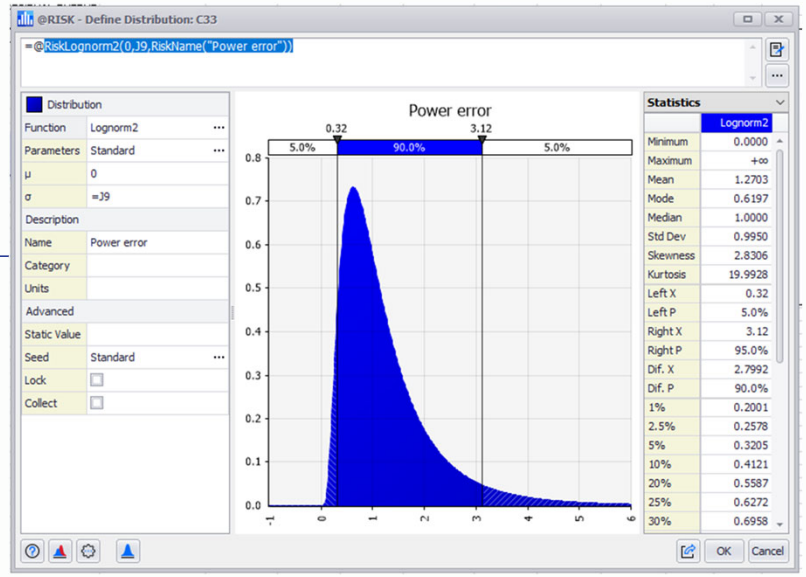
Palisade

Power CER @RISK Implementation

Cost ($K) = 2.5892 * (Power Out $^{.3519}$)* ε'

ε' ~ Lognorm2(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

Palisade

Questions?

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

sales@palisade.com

Palisade

Thank You!

Steven L. Van Drew, PhD, PE
svandrew@palisade.com

Palisade

Back Up Slides

Palisade

Lognormal Error Term Calculations

if Y ~ N(μ,σ) then X = e^Y ~ *ln*(μ,σ)
E(X) = e^(μ+σ^2/2); V(X) = [e^(2μ+σ^2)](e$^{\sigma^2}$-1)

to model in @RISK use
RiskLognorm2 (μ,σ) or RiskLognorm (E(X), {V(X)}$^{.5}$)

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

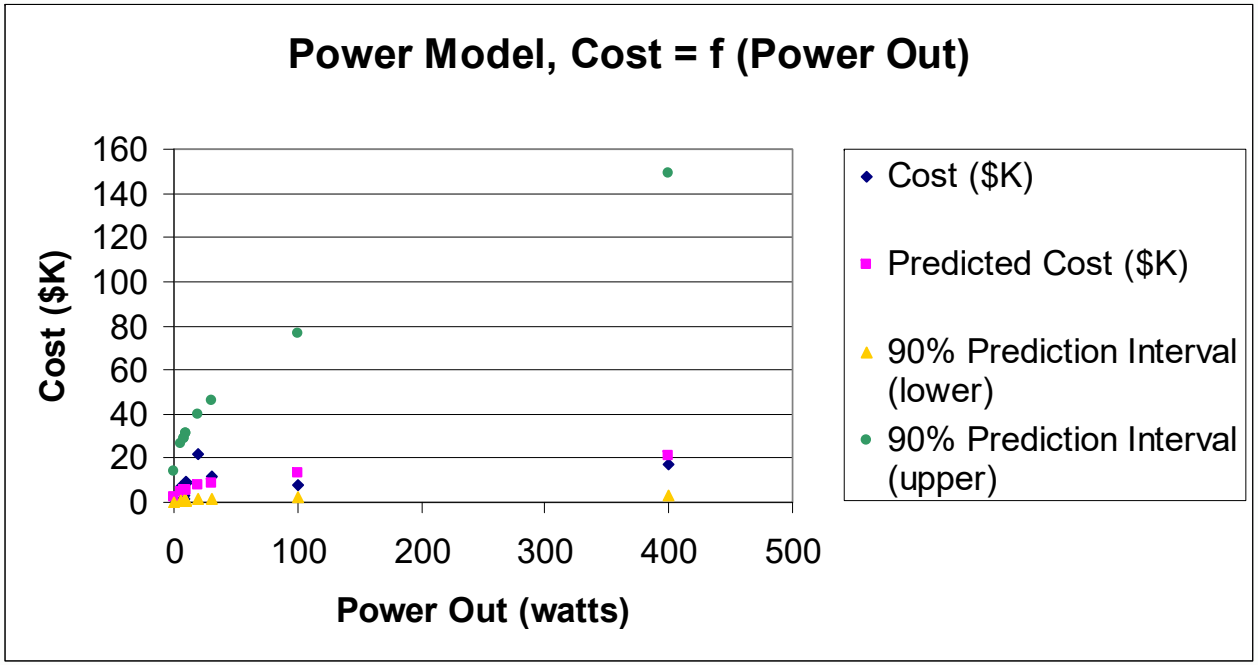
$$V(X) = \left(e^{2\mu+\sigma^2}\right)\left(e^{\sigma^2}-1\right) = \left(e^{2(0)+.6917^2}\right)\left(e^{.6917^2}-1\right) = .9900;$$

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

Lognorm2(0,.6917) ≡ Lognorm(1.2703,.9950)

Palisade

Power Model 90% Prediction Interval



Palisade