Surveying Asteroid Configus with Statistical Methods

A Parable Showing Why the World Will Always Look Flat to Brunswikians

Gary H. McClelland
University of Colorado

Brunswik Society Meetings
Los Angeles, CA
November 1995

Question:

People of Asteroid Configus want to know whether the surface of the asteroid on which they live is uniformly flat or whether it has hills and valleys?

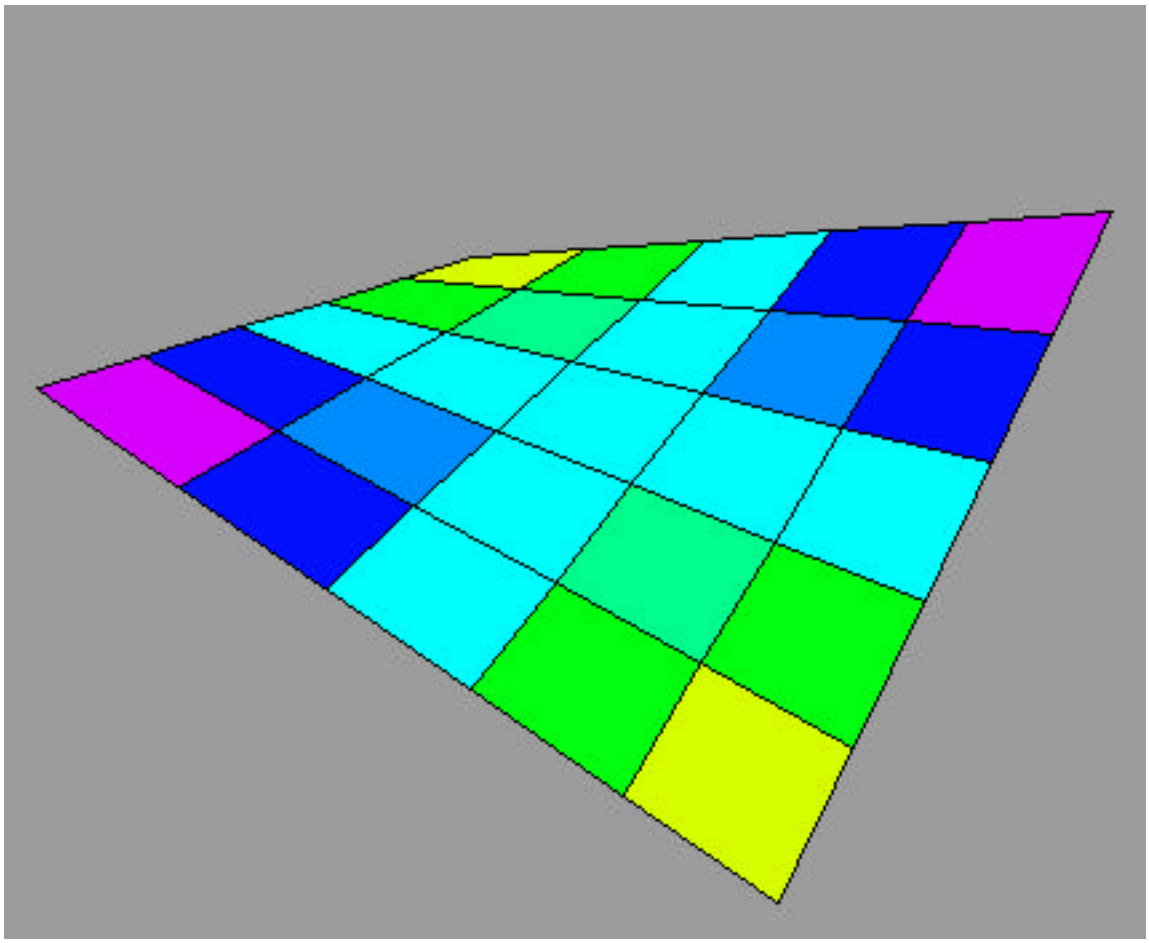
Problem:

This isolated society has not yet developed geometry and associated methods for land surveys

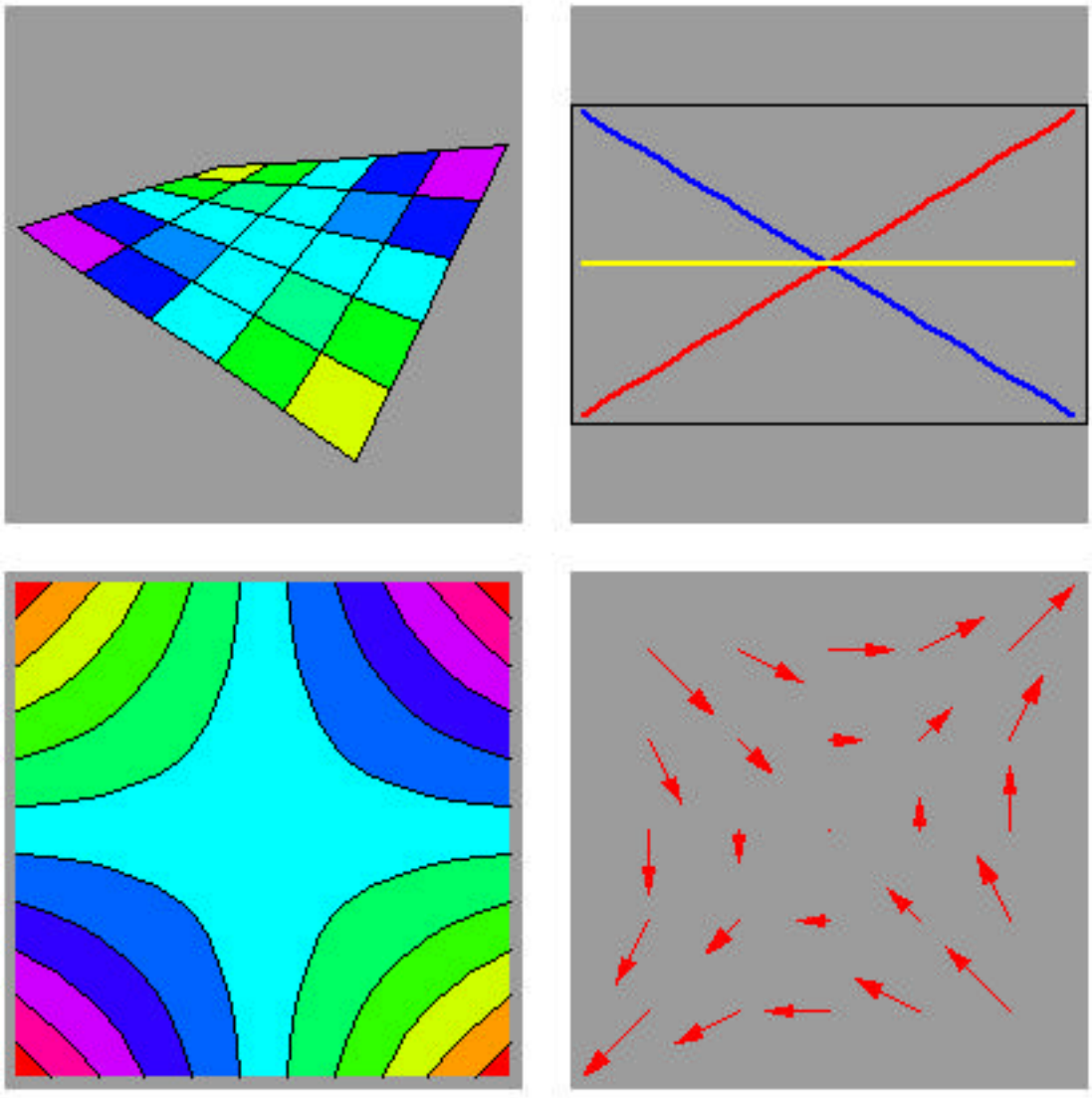
Solution:

Fortuitously, this society has developed sophisticated statistical methods, especially multiple regression

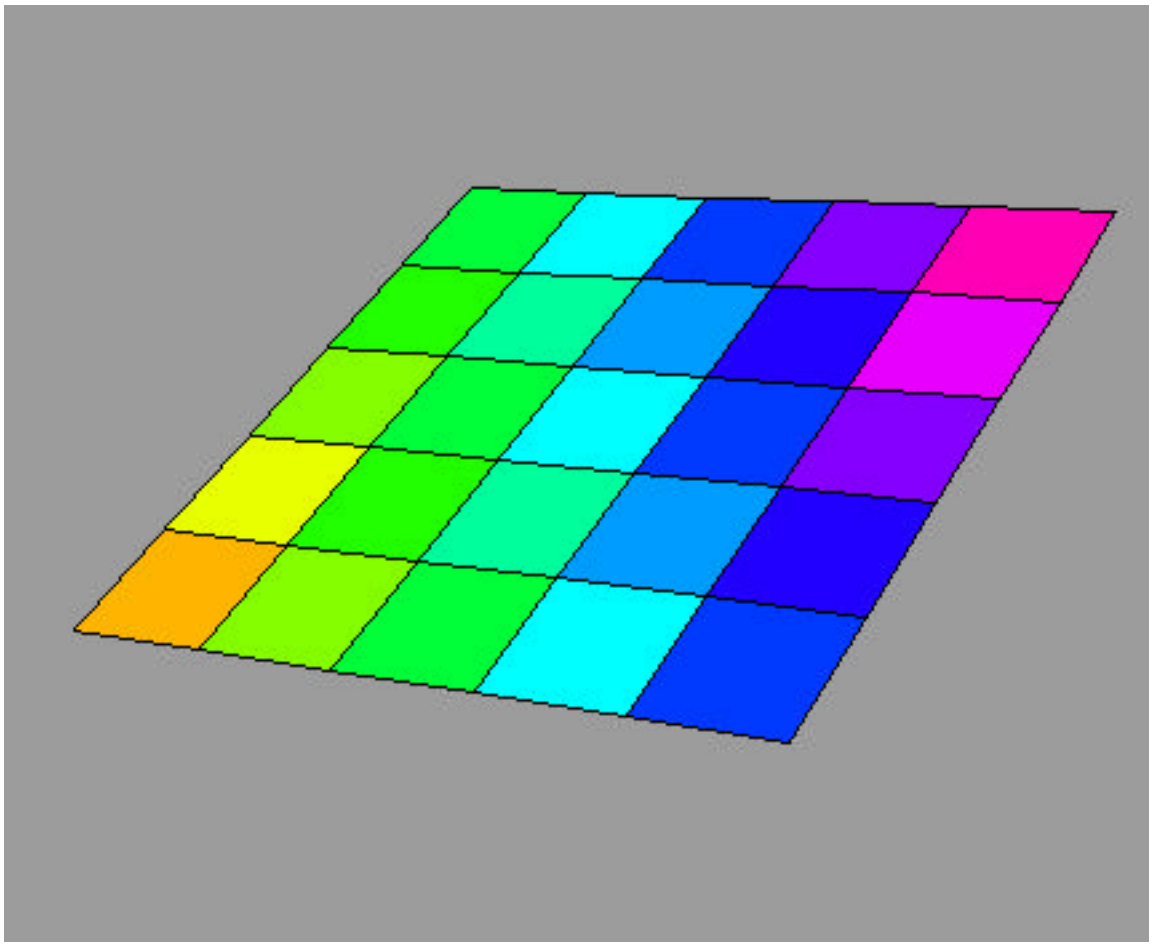
Surface of Asteroid Configus



Alternative Surface Views

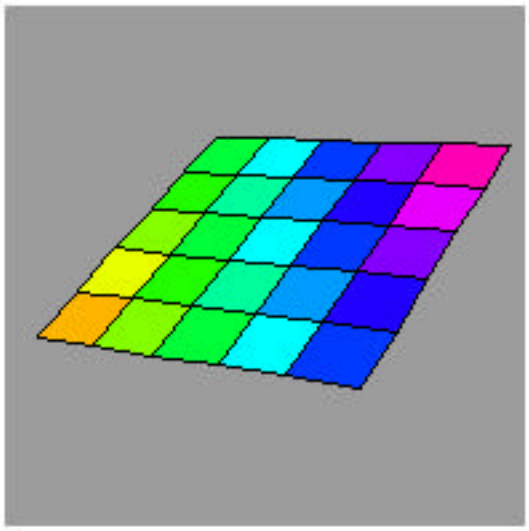


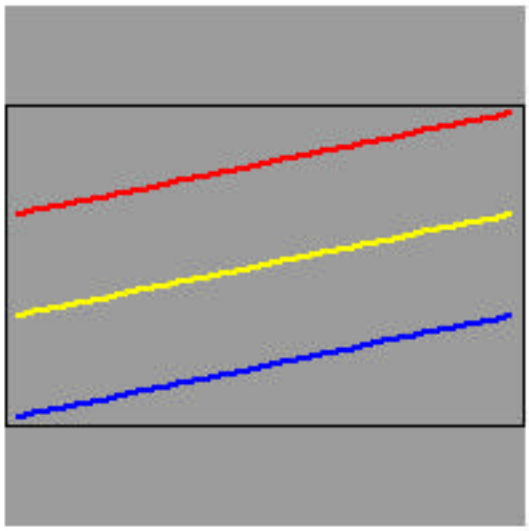
Alternative Uniform Slope Model for the Surface of Asteroid Configus

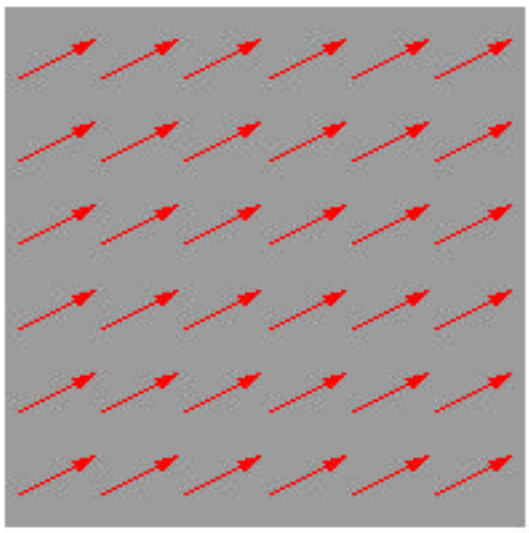


Gary.McClelland@Colorado.edu

Alternative Views of the Uniform Slope Model

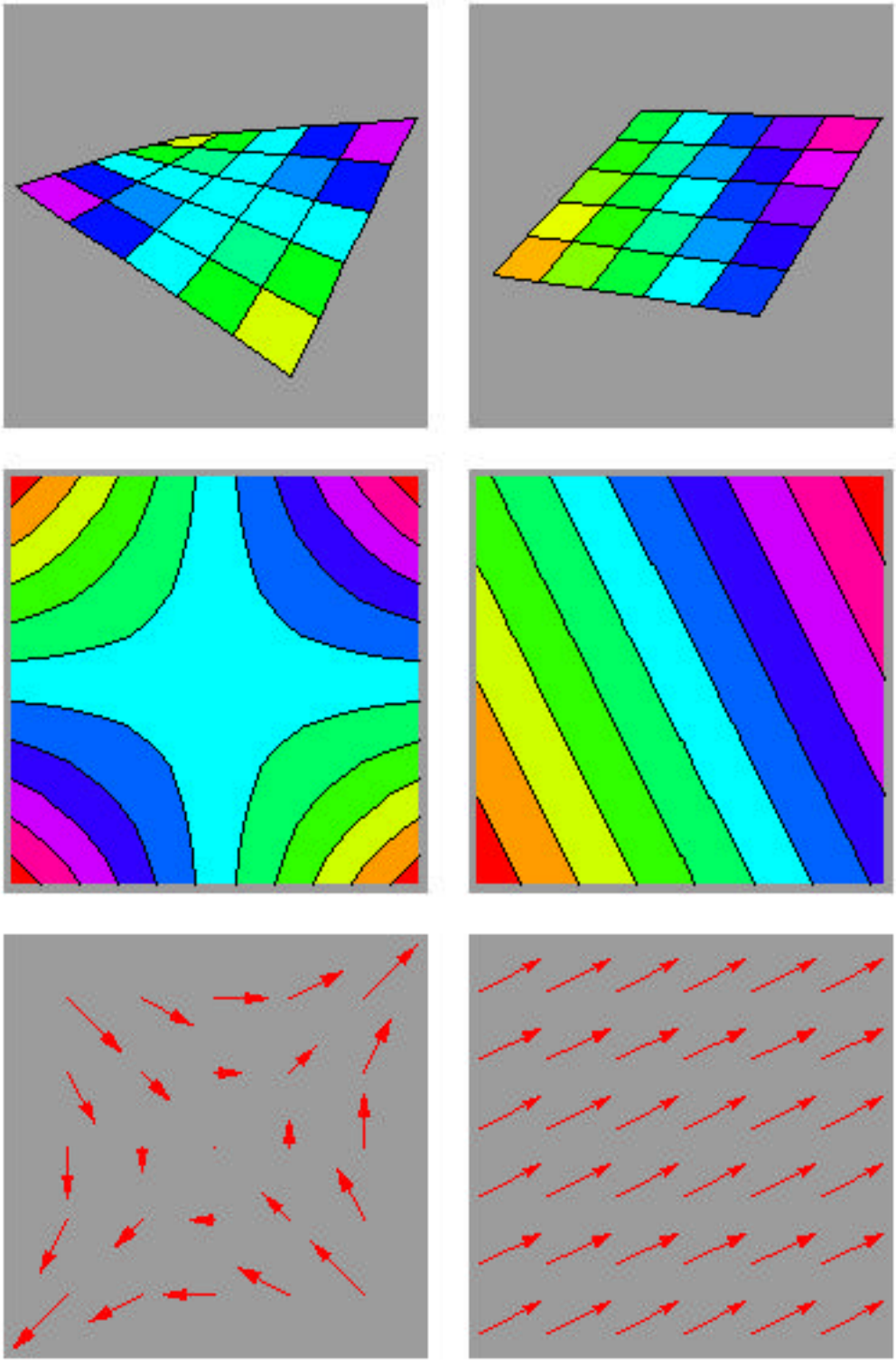






Gary.McClelland@Colorado.edu

Comparison of Hilly and Flat Models



Gary.McClelland@Colorado.edu

Statistical Model for Detecting Hills and Valleys

X = Longitude

Z = Latitude

Y = Elevation Measure

$$Y_i = \beta_0 + \beta_X X_i + \beta_Z Z_i + \beta_{XZ} X_i Z_i + \varepsilon_i$$

$$\beta_{XZ} \neq 0$$

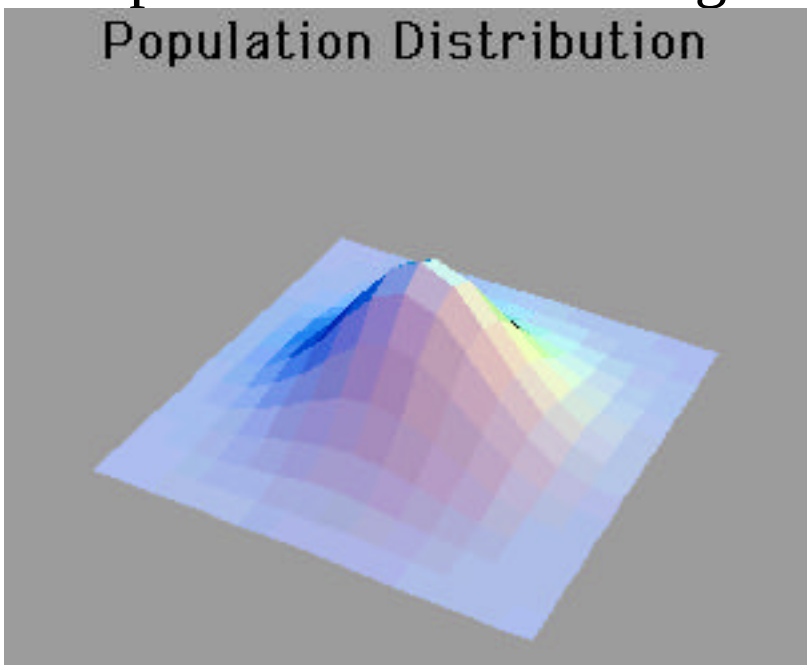
Implies World is Not Flat

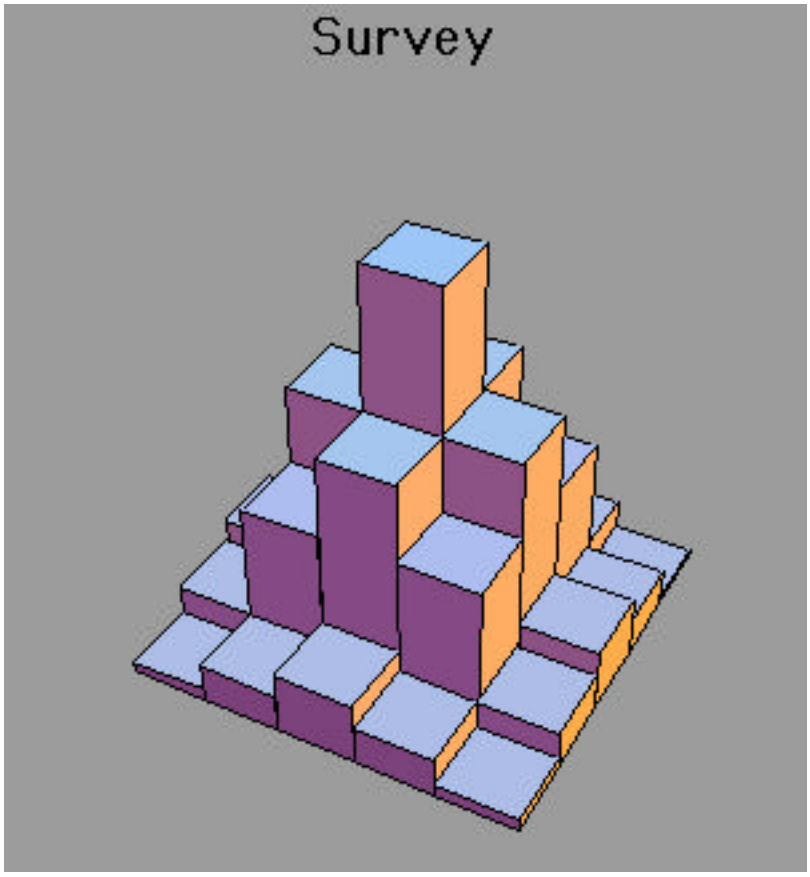
National Institute of Land Surfaces (NILS)

Commissions Brunswikian to do a Survey

- Send Altimeter Watch to Representative Sample of Configians
- Respondents Send Back Altimeter Watch Readings
- For 2 "Cues," 60 Observations Should Be Plenty
- Fit Regression Model

Representative Design





Gary.McClelland@Colorado.edu

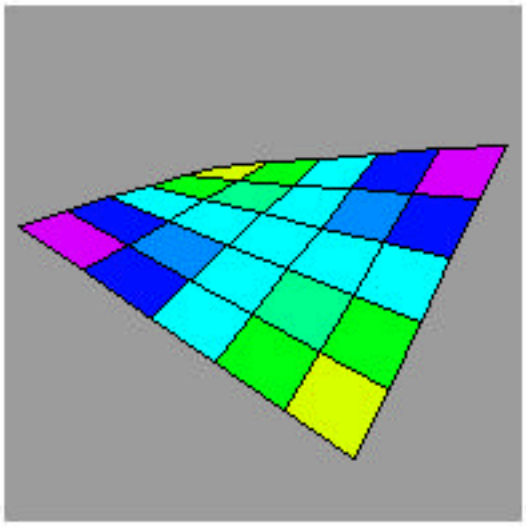
Representative Design Results

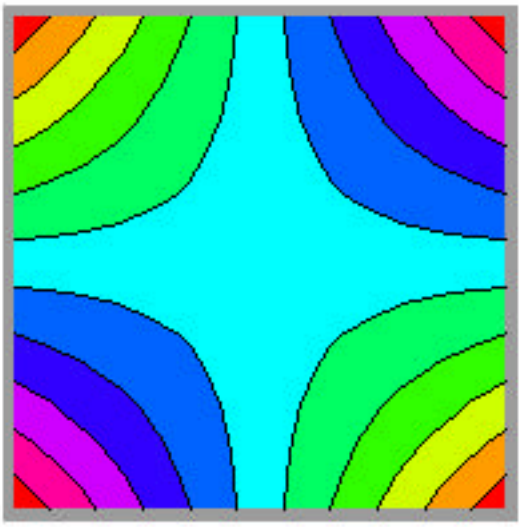
$$\hat{Y}_i = 0 + 0X_i + 0Z_i + 1X_iZ_i$$

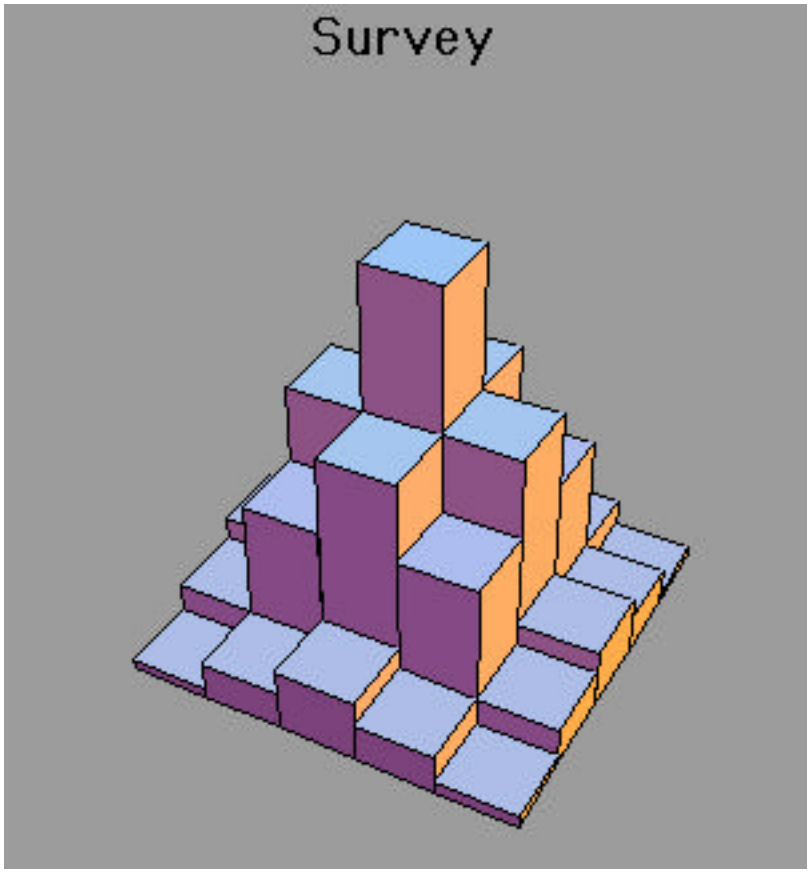
$$R^2 = .05 \qquad p = .09$$

- Therefore, no statistically significant evidence that Asteroid Configus is hilly
- Low R^2 implies that even if there are hills, they aren't important

Representative Design Puts Most Observations Where It Is Flat

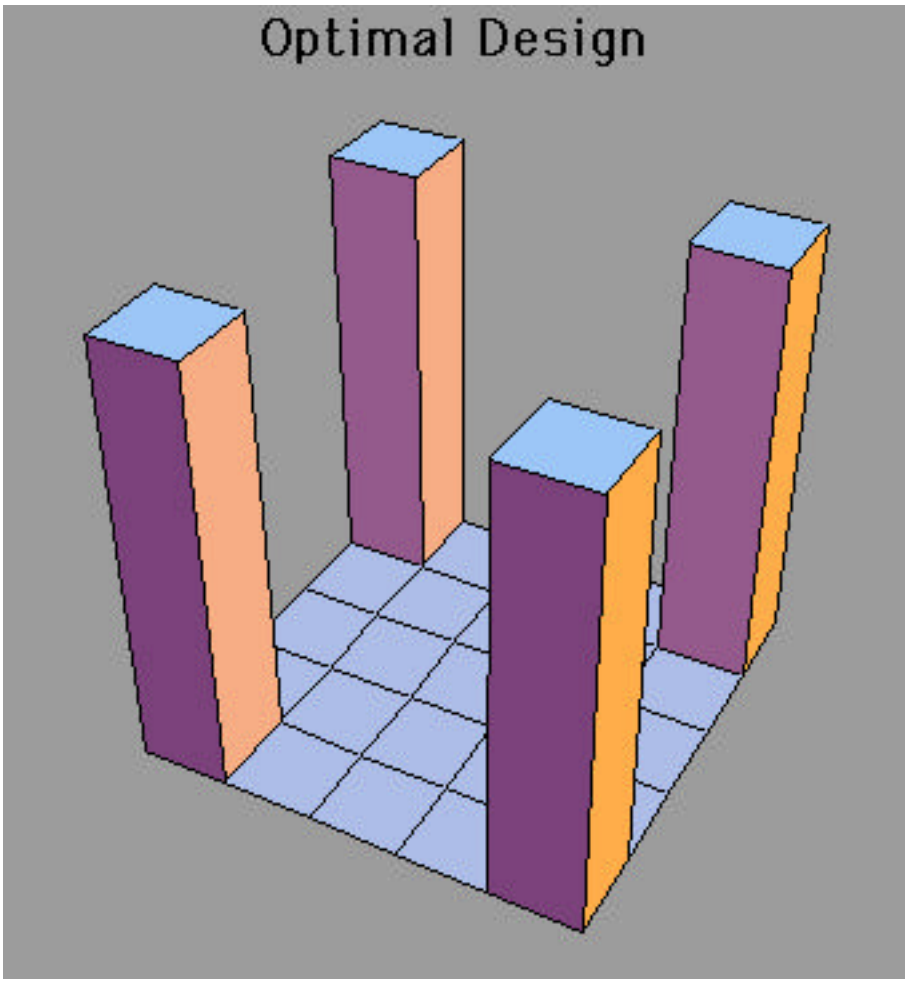




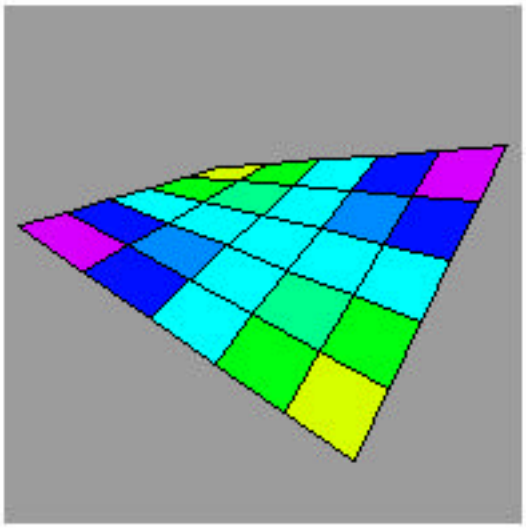


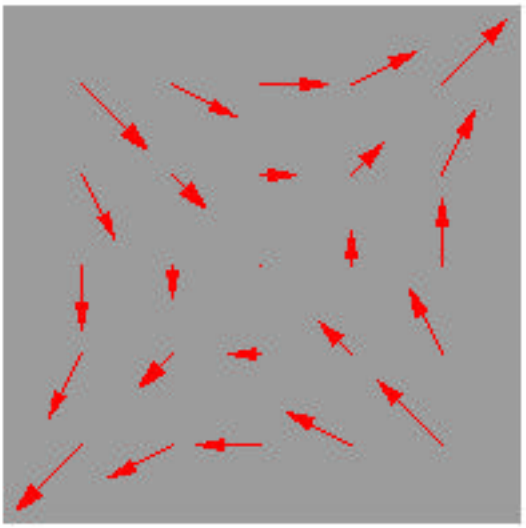
Configus Science Foundation (CSF)

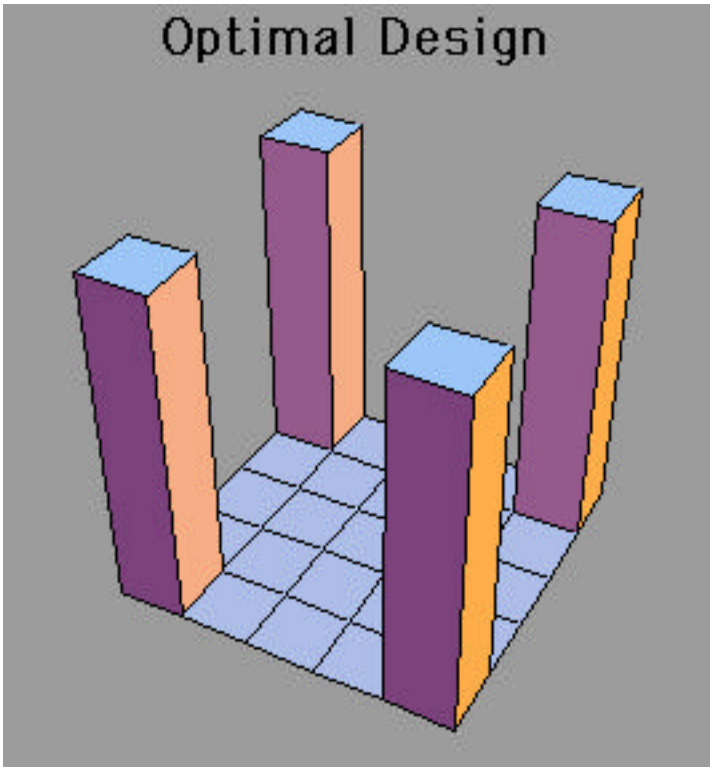
- Notes that NILS has 12 altimeter watches left over
- Awards small grant to researcher who uses optimal design



Optimal Design Puts Observations Where the Action Is







Optimal Design Results

$$\hat{Y}_i = 0 + 0X_i + 0Z_i + 1X_iZ_i$$

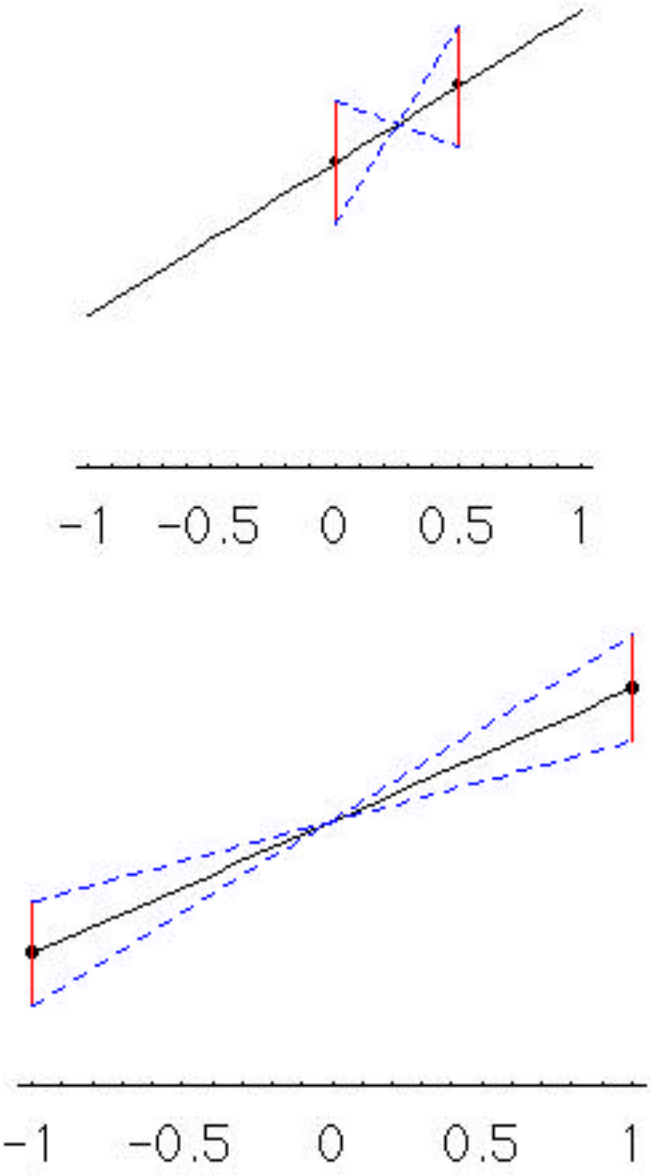
$$R^2 = .47 \qquad p = .03$$

- Therefore, there is statistically significant and strong evidence that Asteroid Configus is NOT flat
- Optimal Design easily detected an effect with only n = 12, that Representative Design could not detect with n = 60.

Scientific Conference to Reconcile the NILS and CSF Results

Proportion of Remaining Variance Explained (after Other Terms Included in the Model)

- It's the Variance!


-1
-0.5
0
0.5
1
-1
-0.5
0
0.5
1

Variance of the Estimate of β_X in Simple Regression:

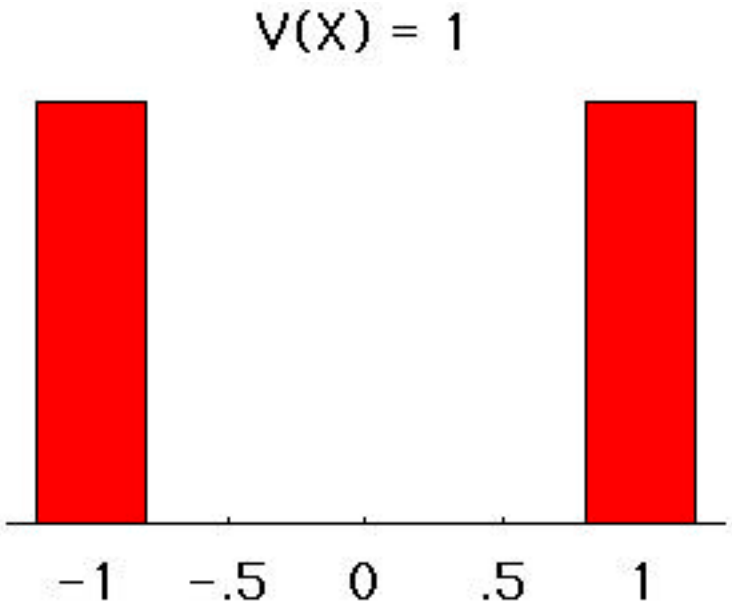
$$Y_i = \beta_0 + \beta_X X_i + \varepsilon_i$$

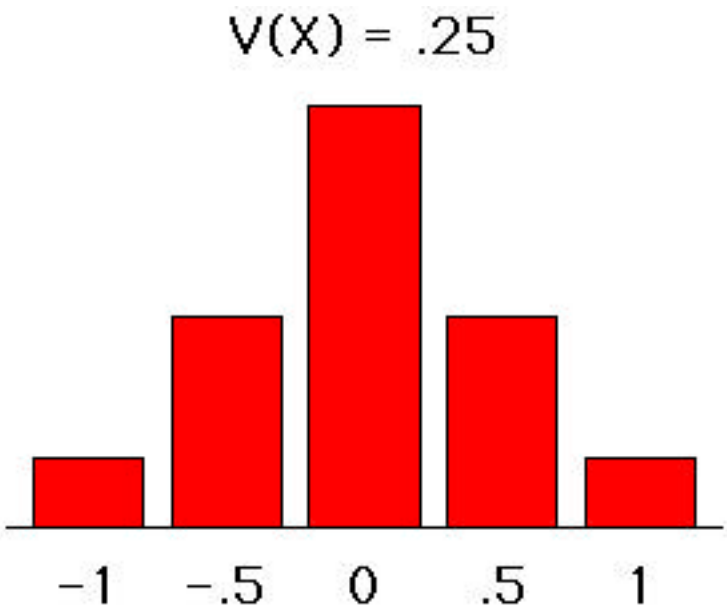
$$\frac{MSE}{V(X)\,n}$$

where $$MSE = \frac{\sum_i (Y_i - \hat{Y}_i)^2}{n - p}$$

Note tradeoff between V(X) and n

All else equal, increasing V(X) produces better estimate of β_X


V(X) = 1
−1
−.5
0
.5
1


V(X) = .25
−1
−.5
0
.5
1

Variance of the Estimate of β_X in Multiple Regression:

$$Y_i = \beta_0 + \beta_X X_i + \beta_Z Z_i + \varepsilon_i$$

$$\frac{MSE}{V(X.Z)n}$$

V(X.Z) is the variance of X after linearly adjusting for Z

$$V(X.Z) = V(X)(1 - r_{XZ}^2)$$

$$\frac{MSE}{V(X)(1 - r_{XZ}^2)n}$$

All else equal, <u>increasing</u> **V(X)** and <u>decreasing</u> r_{XZ}^2 produces better estimate of β_X

Variance of the Estimate of β_{XZ} in Moderator Regression:

$$Y_i = \beta_0 + \beta_X X_i + \beta_Z Z_i + \beta_{XZ} X_i Z_i + \varepsilon_i$$

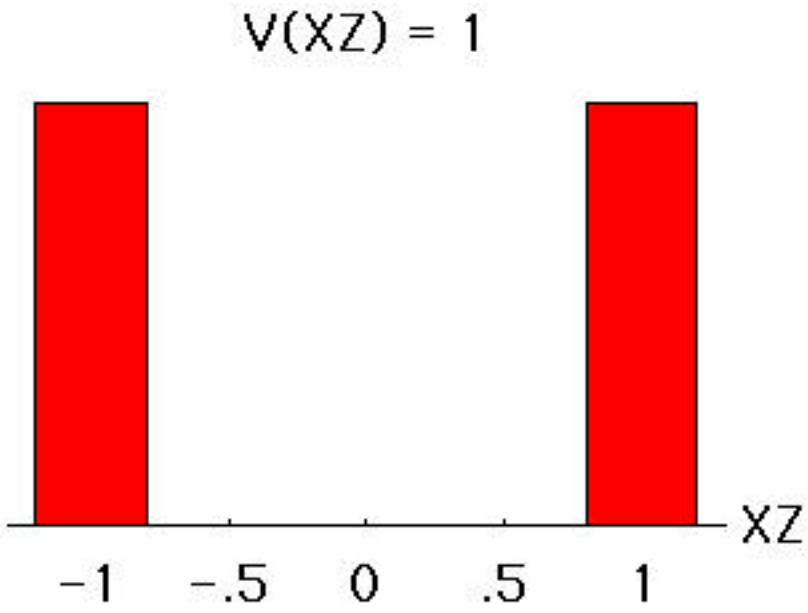
$$\frac{MSE}{V(XZ.X,Z)\,n}$$

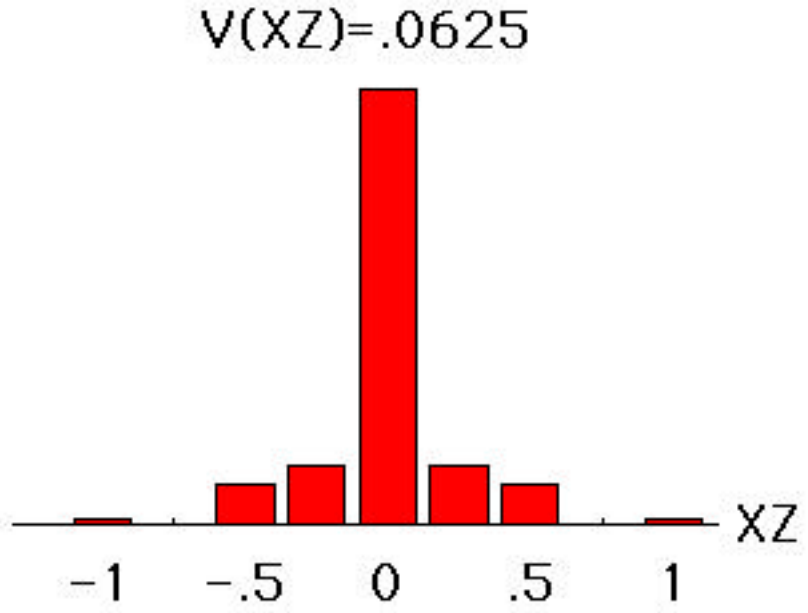
V(XZ.X,Z) = is the variance of XZ after linearly adjusting for X and Z

$$V(XZ.X,Z) = V(X)V(Z) + C(X^2,Z^2) - C^2(X,Z)$$

$$-\frac{C^2(X^2,Z)V(Z) + C^2(X,Z^2)V(X)}{V(X)V(Z) - C^2(X,Z)}$$

$$+\frac{2C(X,Z)C(X^2,Z)C(X,Z^2)}{V(X)V(Z) - C^2(X,Z)}$$


V(XZ) = 1
XZ
-1
-.5
0
.5
1


V(XZ)=.0625
XZ
-1
-.5
0
.5
1

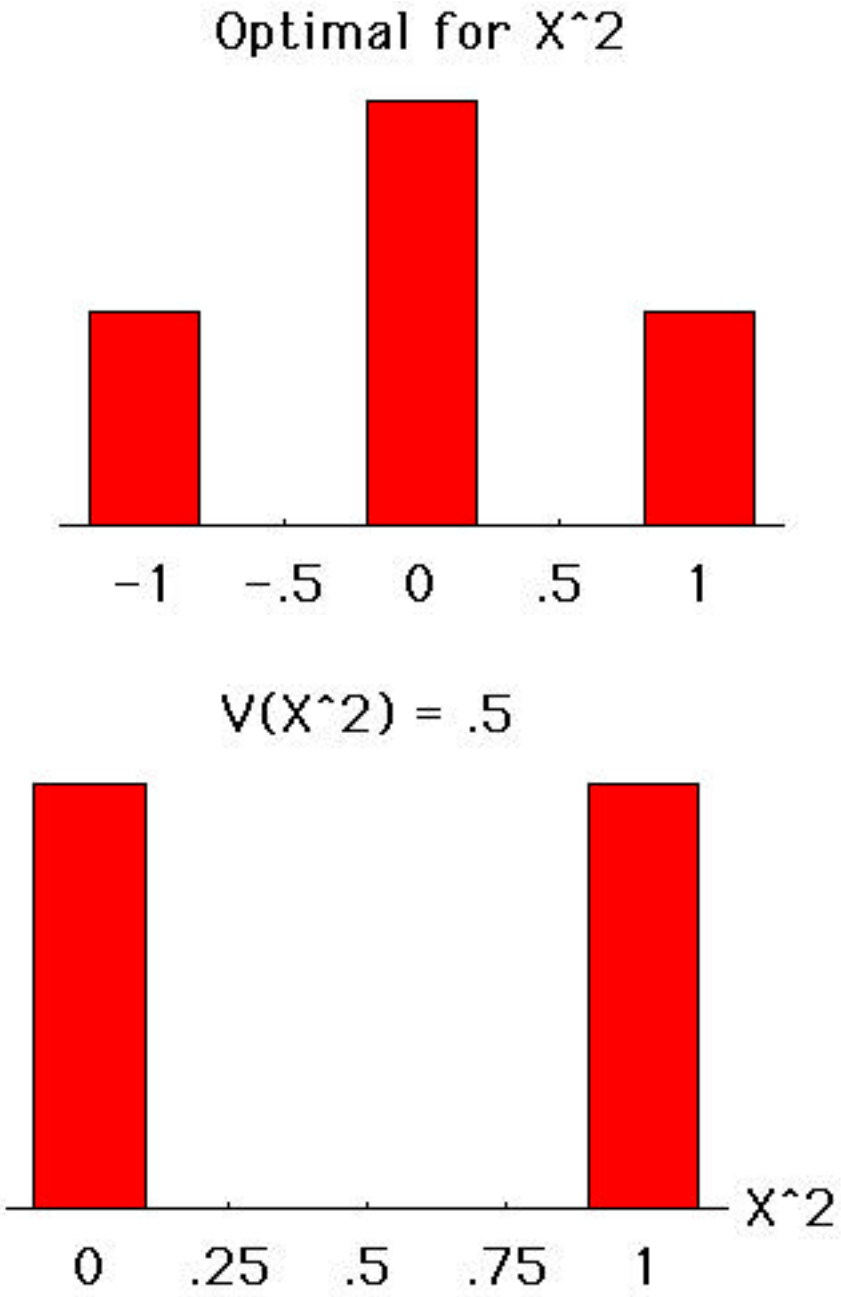
Variance of the Estimate of β_{X^2} in Quadratic Regression:

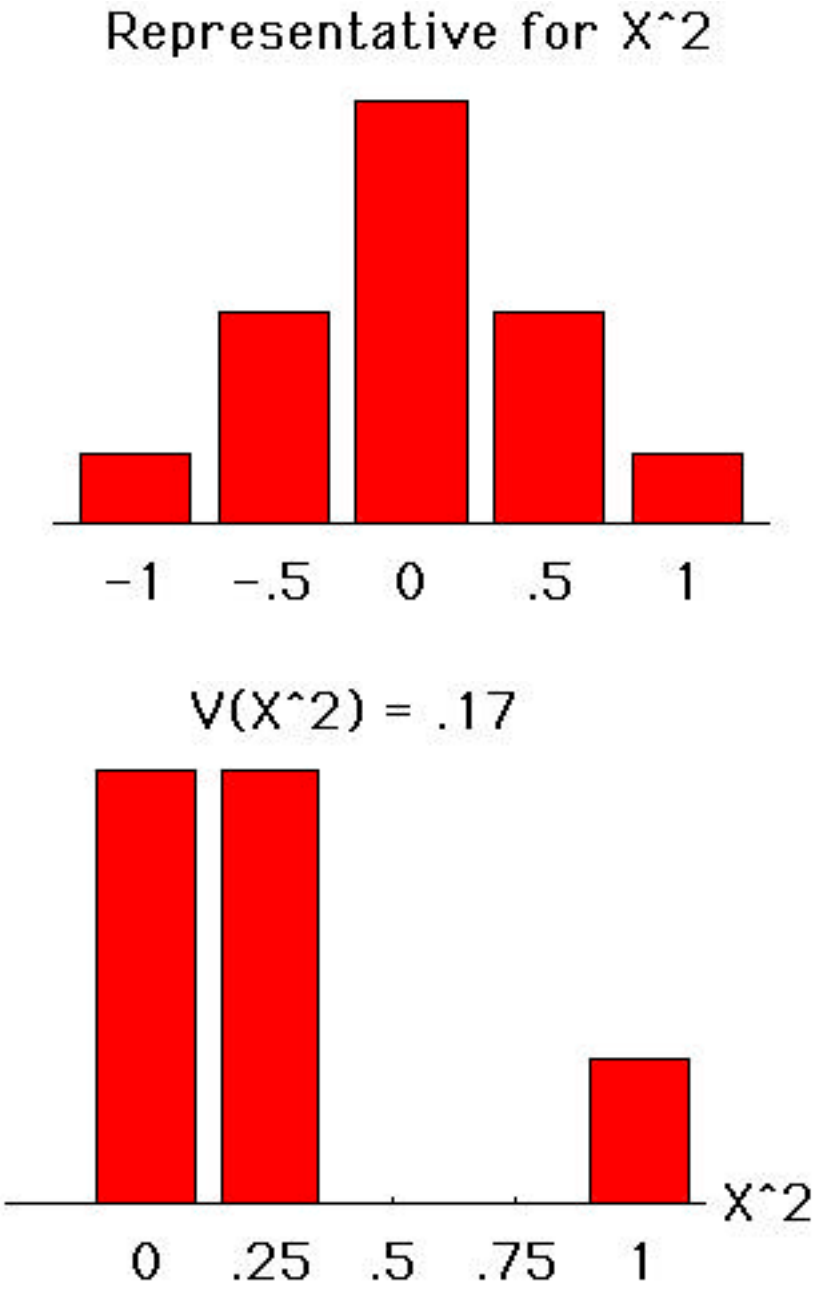
$$Y_i = \beta_0 + \beta_X X_i + \beta_{X^2} X_i^2 + \varepsilon_i$$

$$\frac{MSE}{V(X^2.X)\, n}$$

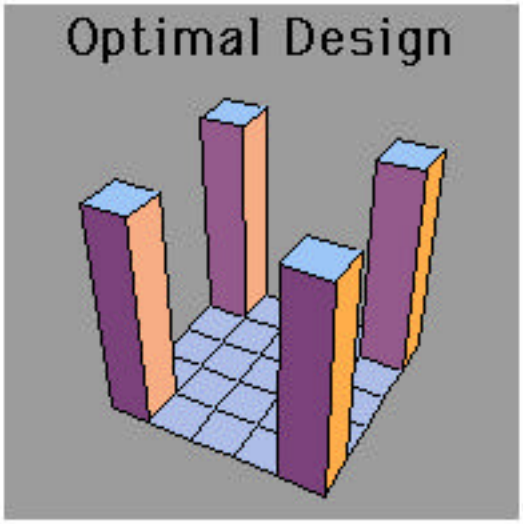
V(X2.X) = is the variance of X2 after linearly adjusting for X

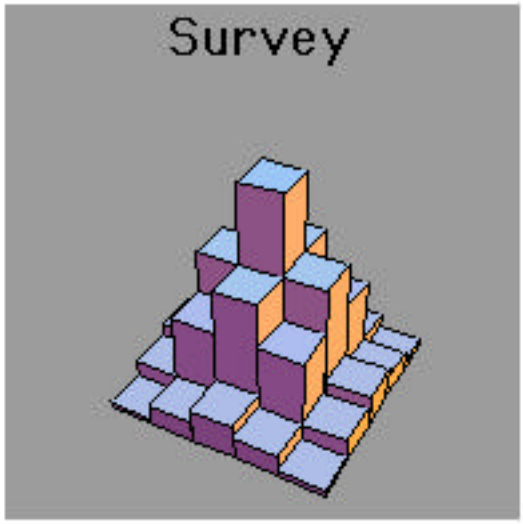
$$V(X^2.X) = V^2(X)[\text{Kurtosis}(X) + 2 - \text{Skew}^2(X)]$$


Optimal for X^2
-1
-.5
0
.5
1
V(X^2) = .5
X^2
0
.25
.5
.75
1


Representative for X^2
-1
-.5
0
.5
1
V(X^2) = .17
X^2
0
.25
.5
.75
1

Summary





- In this case, Optimal Design is 16 Times More Efficient for Detecting Configural Model than Representative Design

- To have the same efficiency as the Optimal Design with n = 12, the Representative Design would require n = 192

- To ensure adequate power for detecting configural effects, most representative designs would require 100s of observations, possibly as many as 1500.

Why Does It Matter?

- Both Designs give correct results!

  Both correctly estimated same model

  Optimal Design reveals that the world is hilly

  Representative Design shows that most people need not worry about the hills

How We Get Ourselves in Trouble

- Clinical Judges and Policy Makers tell us the world is hilly
- We use Representative Designs to model their judgments and tell them they are wrong—their world is really flat
- Extreme Cases are more important today, so important to know whether world is flat or hilly
- Reminder: not finding configural effects with Representative Designs does not mean the world is flat