

Validation of Lineup Protocols for Visual Inference

PhD Proposal
by
Mahbubul Majumder

Program of Study Committee:
Dianne Cook, Major Professor
Heike Hofmann, Major Professor
Frederick O. Lorenz
Ulrike Genschel
Leslie Miller

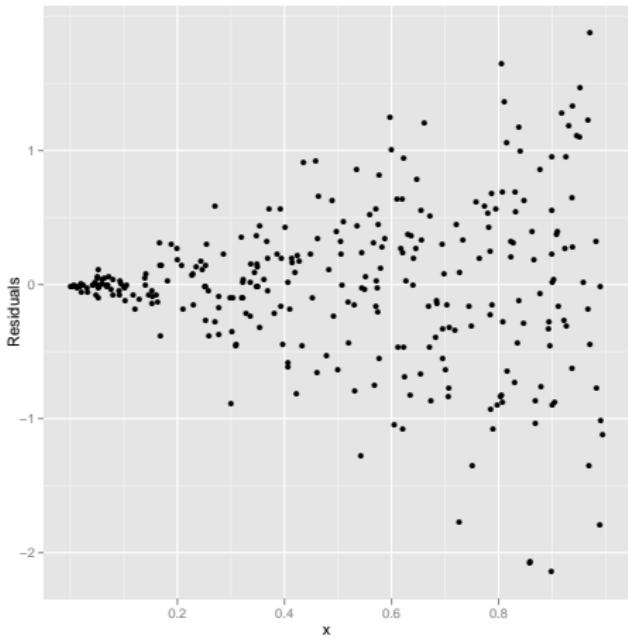
May 5, 2011

Statistical graphics

- A linear model is fitted to the data.
- $R^2 = 0.99$.
- Is it a good model or a bad model?
- Hold on. Show some plots. We want to see.

We want to see

- This residual plot shows the problem with the model.
- Statistical graphics have been used for
 - exploratory data analysis.
 - model checking and diagnostics.
- Can we use statistical graphics for inference?



Visual inference

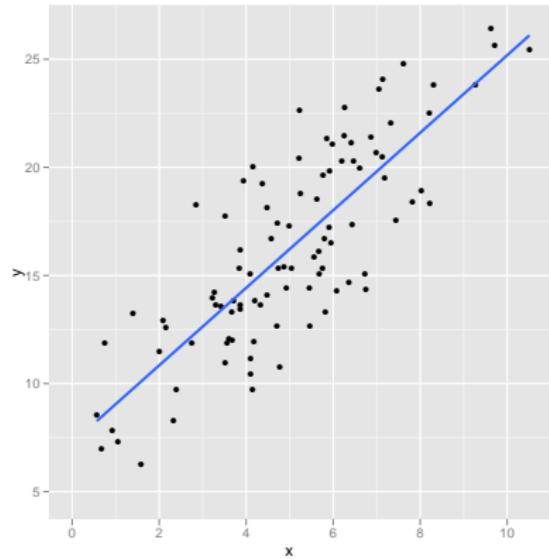
- Buja et al (2009)
 - Introduced method to test the significance of findings.
 - Demonstrated formal testing of overall model fitting.
- Protocols of Visual Inference.
 - Rorschach.
 - Lineup.
- Validation of lineup protocols.
 - the issues related to the performance.
 - has been the focus of this research.

Organization of the chapters

- Chapter 1
 - Introduced method to test the significance of findings.
 - Demonstrated formal testing of overall model fitting.
- Chapter 2
 - focuses on testing parameters related to regression models.
 - Demonstrated formal testing of overall model fitting.
- Chapter 3
 - Introduced method to test the significance of findings.
 - Demonstrated formal testing of overall model fitting.
- Chapter 4
 - Introduced method to test the significance of findings.
 - Demonstrated formal testing of overall model fitting.
- Chapter 5
 - Introduced method to test the significance of findings.
 - Demonstrated formal testing of overall model fitting.
- Chapter 6
 - Introduced method to test the significance of findings.
 - Demonstrated formal testing of overall model fitting.

Test Statistic

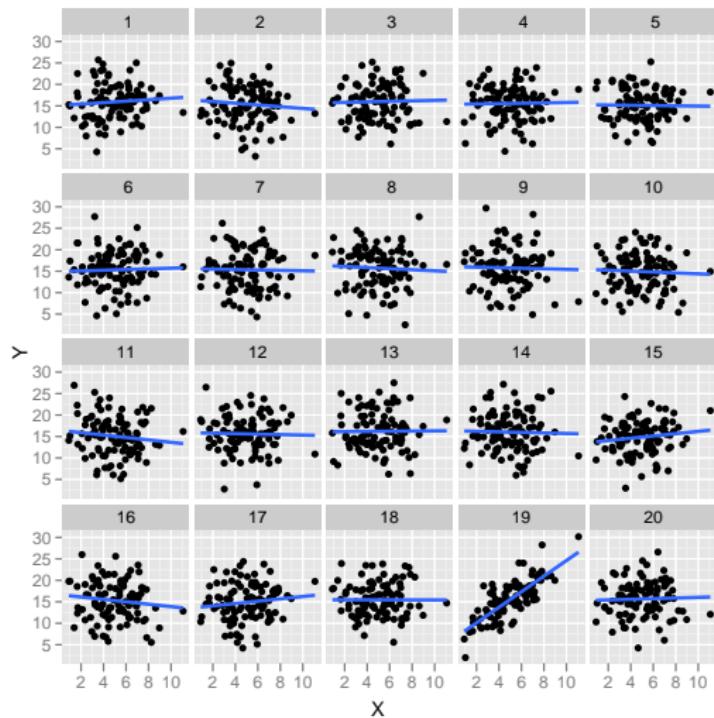
- Function $T(Y)$ that maps data Y to a plot.
- Associated with a specific null hypothesis.
- A good test statistic should display an extreme feature of the data if it exists.
- As an example, this test statistic investigates the existence of a non-zero slope.
- Testing $H_0 : \text{Slope}=0$ vs $H_1 : \text{Slope} \neq 0$



Compare test statistic with null distribution

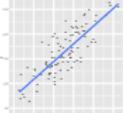
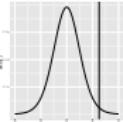
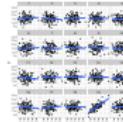
Lineup plot

- A layout of axb plots.
- One of the plots is of observed data
- All other plots are simulated from null model.
- Reject null hypothesis if observed plot is identified.
- To identify the observed plot, more than one person can be involved.



Comparison: Visual vs Mathematical Inference

Model: $Y = \beta_0 + \beta X + \epsilon; \epsilon \stackrel{iid}{\sim} N(0, \sigma^2)$

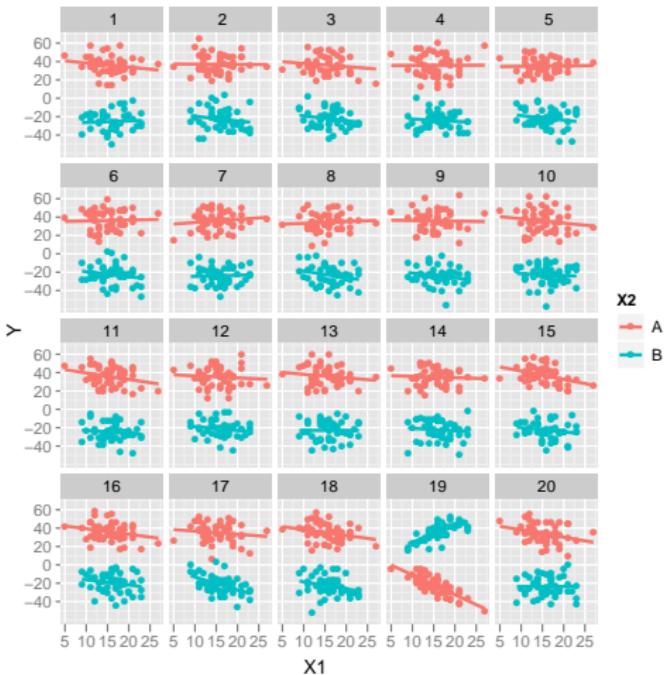
	Mathematical Inference	Visual Inference
Hypothesis	$H_0: \beta = 0$ vs $H_1: \beta \neq 0$	$H_0: \beta = 0$ vs $H_1: \beta \neq 0$
Test statistic	$T(y) = \frac{\hat{\beta}}{se(\hat{\beta})}$	$T(y) =$ 
Null Distribution	$f_{T(y)}(t);$ 	$f_{T(y)}(t);$ 
Reject H_0 if	observed T is extreme	observed plot is identifiable

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1} X_{i2} + \dots + \epsilon_i ; \epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

Null Hypothesis	Type	Test Statistic
$H_0 : \beta_k = 0$	Residual Plot	
$H_0 : X$ Linear	Residual Plot	
$H_0 : \beta_k = 0$ for categorical X_k	Boxplot	
$H_0 : \beta_k = 0$ (interaction with categorical X_k)	Scatter plot	
$H_0 : \text{Model Fits}$	Histogram	

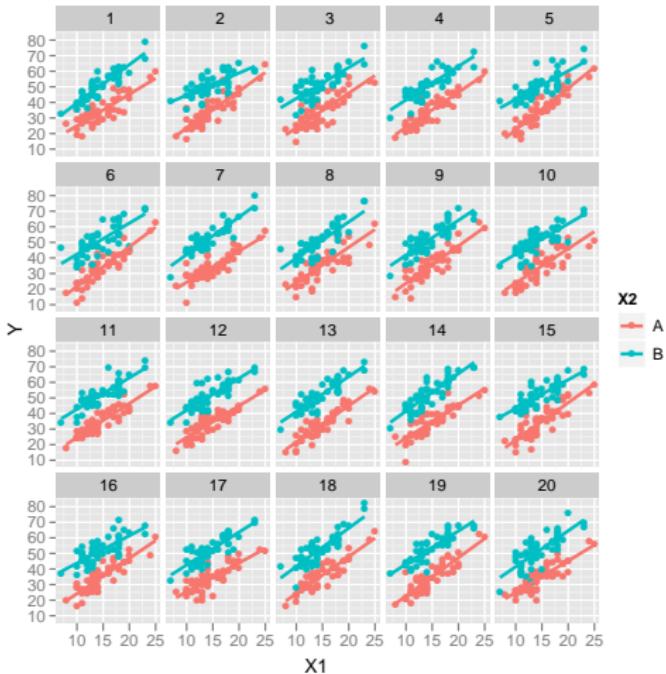
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon ; \epsilon \sim N(0, \sigma^2)$$

- $H_0 : \beta_3 = 0$
- Can you identify the observed plot?



$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon ; \epsilon \sim N(0, \sigma^2)$$

- $H_0 : \beta_3 = 0$
- Can you identify the observed plot?



P value and Type-I error

For a lineup of m plots

① p-value for an Individual evaluation

- When reject report p-value $\leq \frac{1}{m}$.
- When cannot reject report p-value $\geq 1 - \frac{1}{m}$.

② p-value for N independent evaluations

- Under Null hypothesis, $Pr(\text{Reject}) = \frac{1}{m}$ for each evaluation.
- number of success $U \sim \text{Binom}(N, \frac{1}{m})$.
- p-value = $Pr(U \geq u) = \sum_{k \geq u}^N \binom{N}{k} \left(\frac{1}{m}\right)^k \left(1 - \frac{1}{m}\right)^{(N-k)}$ where u be the observed number of success.
- Exact probability for discrete variable makes it conservative.
- When $N = 1$ this p-value matches with individual judgment p-value

③ Type-I error probability = $\frac{1}{m}$.

Power for testing $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_0^c$

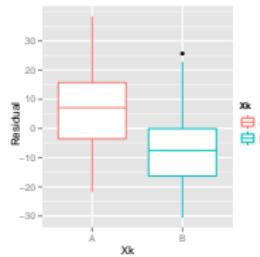
- For a lineup of m plots, power function of θ be defined as

$$\beta(\theta) = \begin{cases} \text{Type-I error} = \frac{1}{m} & \text{if } \theta \in \Theta_0, \\ Pr(\text{Reject } H_0) & \text{if } \theta \in \Theta_0^c. \end{cases}$$

- Estimated power = $\frac{u}{N}$
 u = number of successful evaluations
 N = number of independent evaluations.
- A generalized mixed linear model can be used to estimate power.

Simulation based experiment

- Model: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$; $\epsilon \stackrel{iid}{\sim} N(0, \sigma^2)$; X_2 categorical
- Hypothesis $H_0 : \beta_2 = 0$ vs $H_1 : \beta_2 \neq 0$
- Test statistic is the boxplot of residuals of fitted null model grouped by X_2 . For lineup plot we simulate data from $N(0, \hat{\sigma}^2)$



Survey Setting

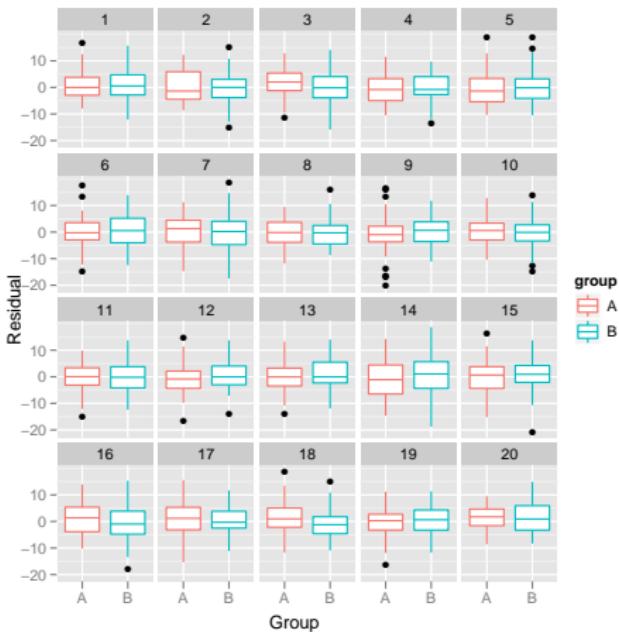
- Values of parameters considered for survey experiment.

Sample size (n)	σ	values for β_2					
100	5	0	1	3	5	8	
	12	1	3	8	10	16	
300	5	0	1	2	3	5	
	12	1	3	5	7	10	

- For each of the above combinations 3 independent lineup plots were generated.
- Recruited 324 participants through Amazon Mechanical Turk web site.

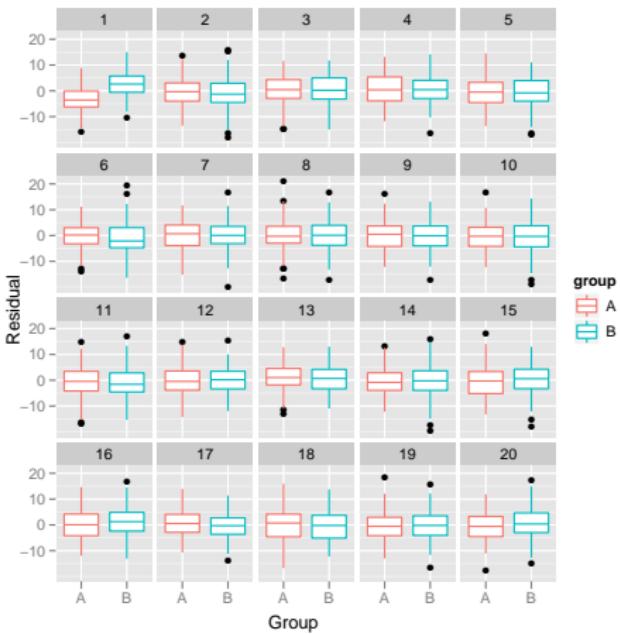
Survey results

- Sample size 100, $\beta = 1$ and $\sigma = 5$.
- For observed plot, p-value is 0.75
- Most of the responses are 18. Has p-value 0.028 which is minimum.
- Attempted 18 times with 5.5% success.



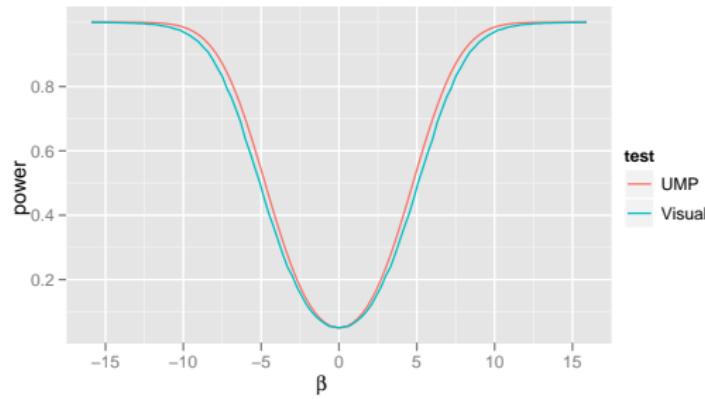
Survey results

- Sample size 300, $\beta = 5$ and $\sigma = 5$.
- For observed plot, p-value < 0.0001
- Attempted 23 times with 100% success.

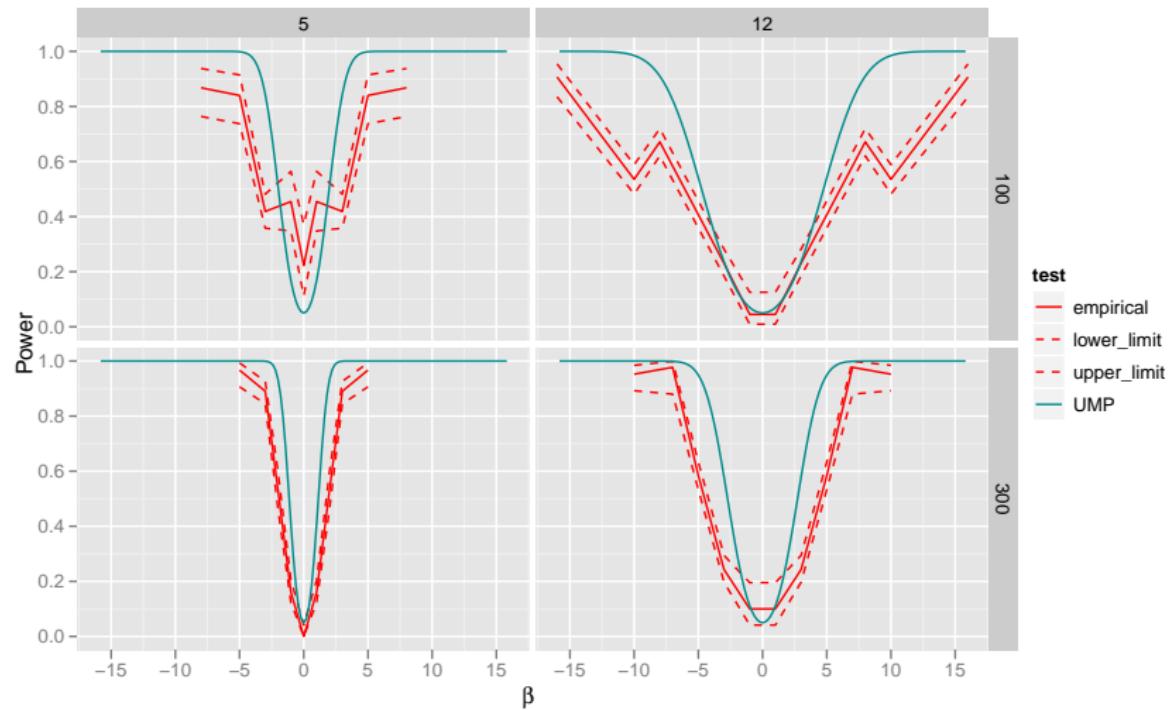


Expected power

- Under H_1 distribution of p-value p_m is right skewed
- Under H_0 $p_m \sim \text{Uniform}(0,1)$
- $p_0 = \min(p_m) \sim \text{beta}(1, m - 1)$
- Expected power = $\Pr(p_{obs} < p_0)$

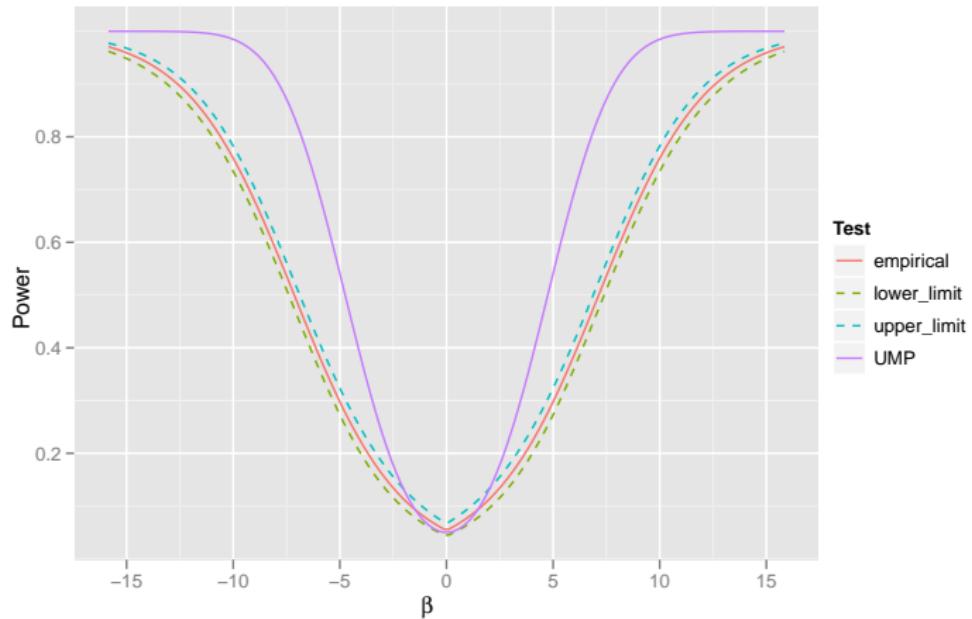


Observed power



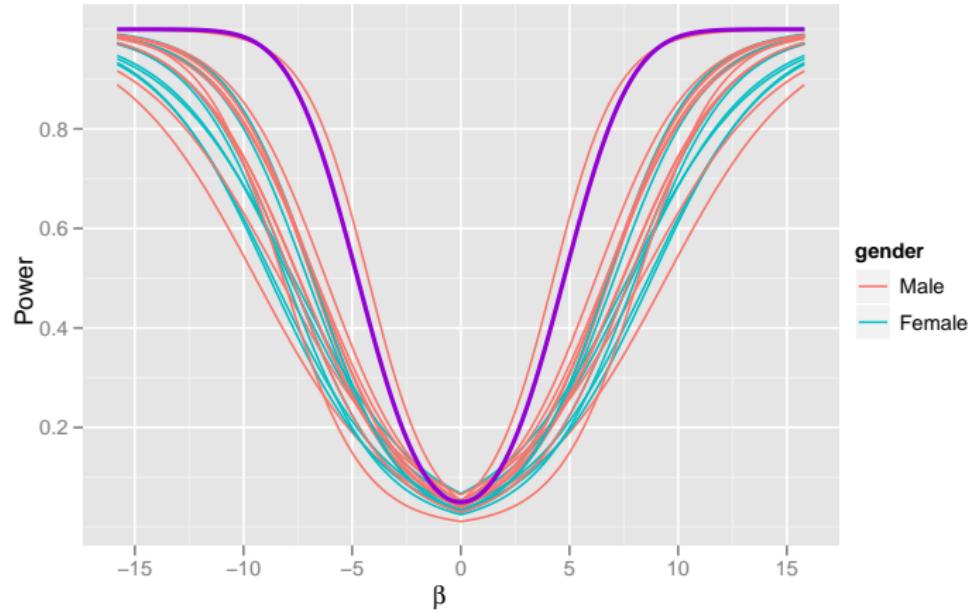
Power estimated from logistic model

Sample size = 100, standard deviation = 12



Power estimated from generalized mixed model

Sample size = 100, standard deviation = 12



Future work

- Visual inference is not the competitor to traditional inference.
- May use where traditional tests can't be used.
- What if not normal?
 - Extend this study for generalized linear model.
- Apply the procedure with real data.
- Conduct survey
 - Examine the other test statistics.
 - Assess the sensitivity of power to modeling conditions.
 - Discover the most effective specification of a plot.

Thanks

Question?