ADSURGO
ascend with analytics

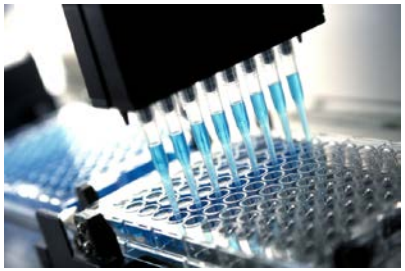
DISCOVERY SUMMIT

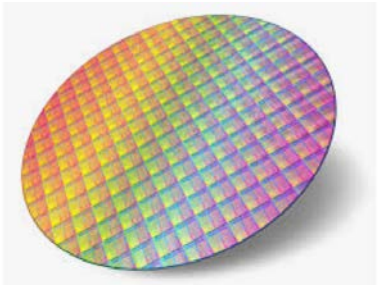
Establishing Equivalence in Practical Applications

October 17, 2019

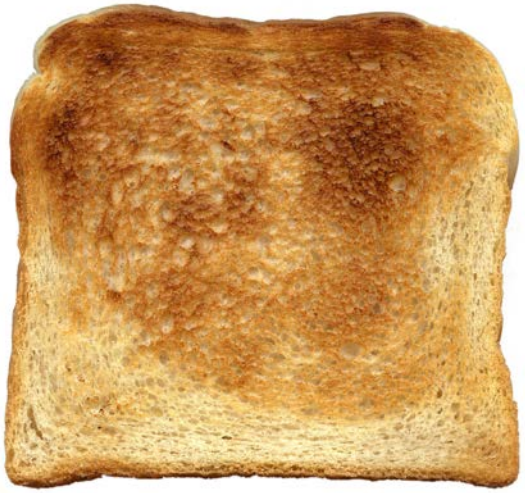
Jim Wisnowski, Adsurgo LLC
james.wisnowski@adsurgo.com

Andrew Karl, Adsurgo LLC
andrew.karl@adsurgo.com





Some Initial Thoughts



https://commons.wikimedia.org/wiki/File:Toast-2.jpg

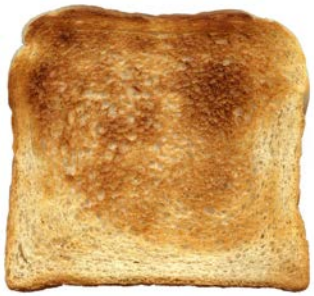
www.jif.com

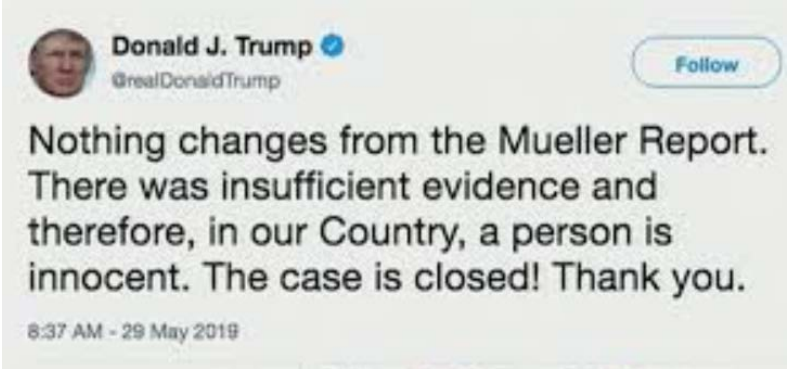
Equivalence Introduction

- Problem: Some batches are experiencing fill weight that differs from 28 ounces. A consultant has recently implemented design changes and SPC methods.
- Data: We sample n= 20 jars and they have a mean of $\bar{y}$ and standard deviation of *s*.
- Method: Two-sided one-sample t-test
- Conclusion: Consultant …"with a t-statistic of 0.77 and p-value of 0.45 we have proven our mean is equal to 28 ounces"

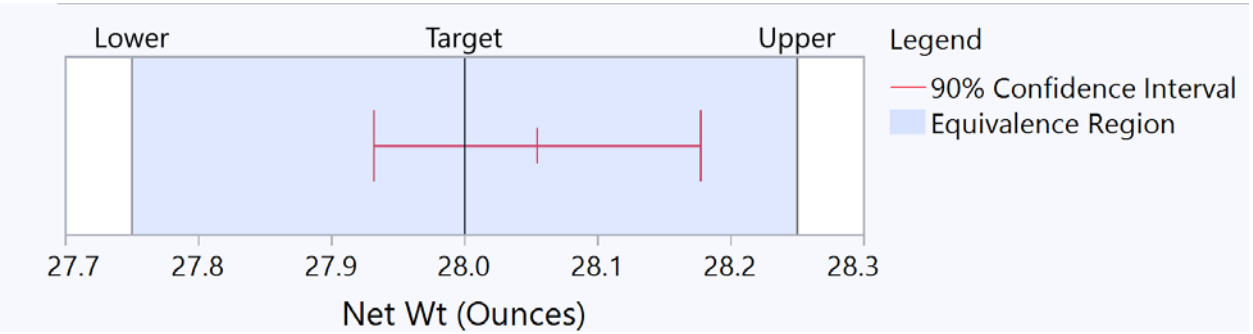
- Question: Have we really established the equivalency of the mean of 28 ounces?
- Practitioners may criticize us for the stats term ***fail to reject the null*** believing you either accept the null or accept the alternate hypothesis; but it is quite descriptive.

Establishing Equivalence



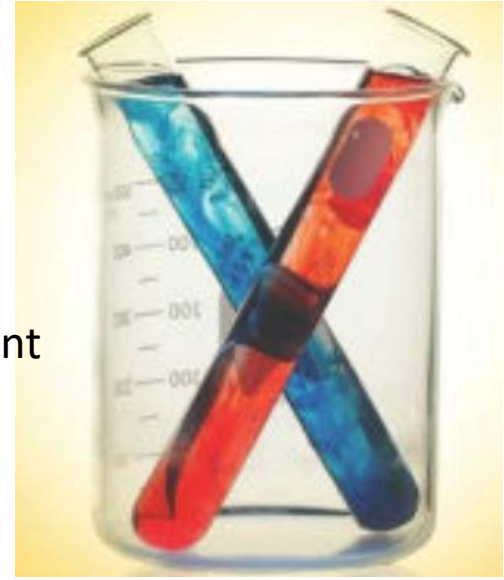


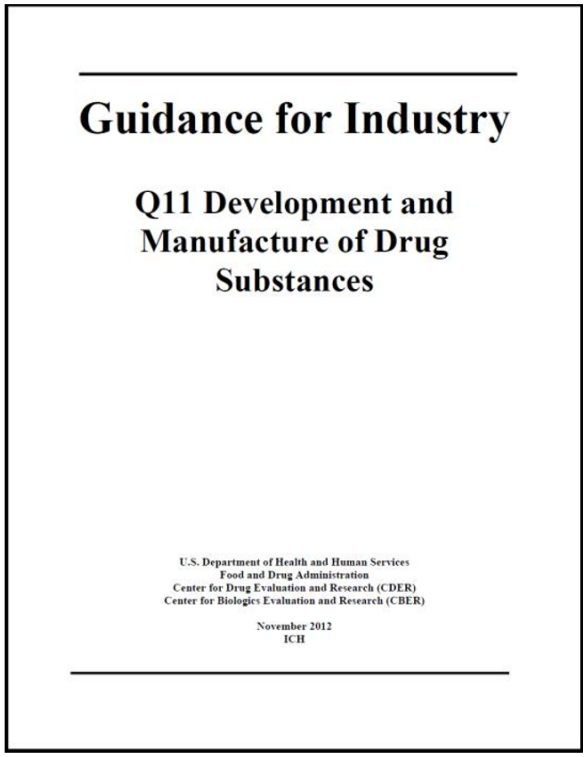
- Failing to reject is a good start! Target value should fall within Confidence Intervals too.
- Need to determine what difference Δ from 28oz is practically significant. Is it 0.000001 oz, 0.1oz, 1oz, 10oz?
- Conduct two one-sided tests (TOST) by adding and subtracting this delta value to the desired target (28 oz).
- JMP Demonstration to show Distribution platform
  - Test equivalence for a quarter ounce
  - Test equivalence for a tenth of an ounce



Establishing Equivalence: Pharma Example

- Problem: Impurity must be consistent between lab results and when scale to a pilot plant for drug substances
- Data: 30 observations from each
- Method: TOST for two samples (Fit Y by X, Fit Model)
- Conclusion: For a difference of 0.3; the two scales are equivalent





Guidance for Industry

Q11 Development and Manufacture of Drug Substances

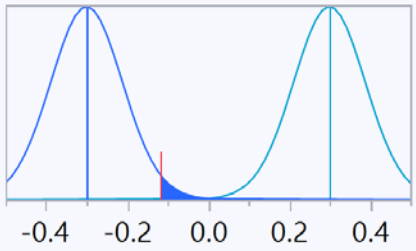
U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

November 2012
ICH

Practical Equivalence between Pilot and Lab

Specified Practical Difference Threshold	0.3
Actual Difference in Means	-0.11767
Std Error of Difference	0.089236

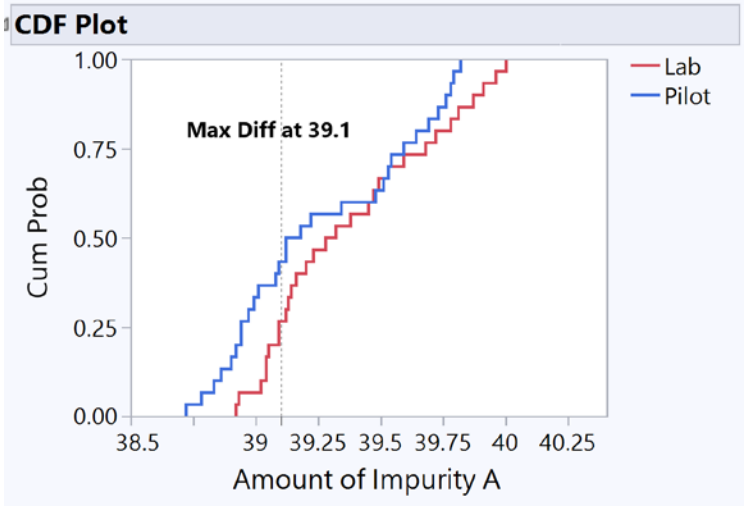
Null Hypothesis	DF	t Ratio	p-Value
Mean Difference ≥ 0.3	58	-4.68045	<.0001*
Mean Difference ≤ -0.3	58	2.043261	0.0228*
Max over both			0.0228*

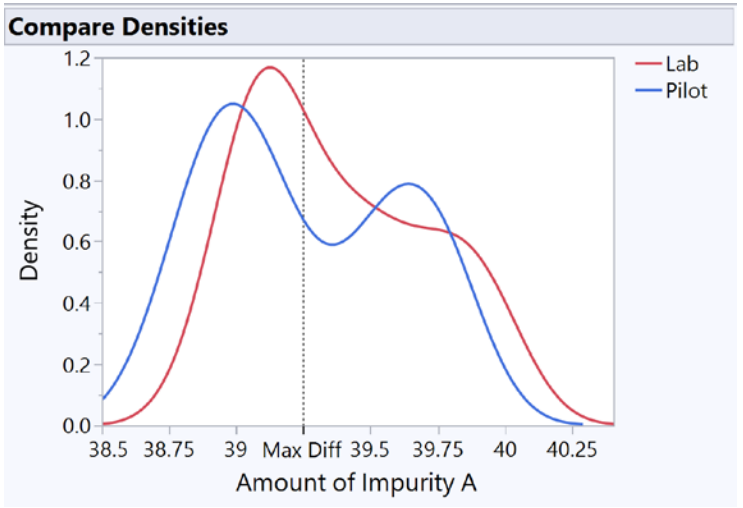


Reference: *Guidance for Industry Q11 Development and Manufacture of Drug Substances* Nov 2012.

Establishing Equivalence: Equal Distributions

- Are the probability density functions the same or close enough?
- Kolmogorov-Smirnoff test looks at max difference between the two CDF curves



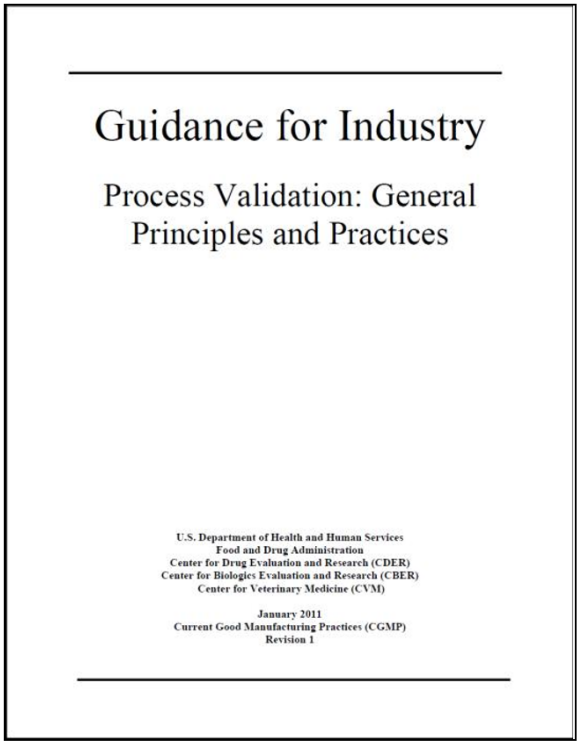


Kolmogorov-Smirnov Asymptotic Test

KS	KSa	D=max|F1-F2|	Prob > D	D+=max(F1-F2)	Prob > D+	D-=max(F2-F1)	Prob > D-
0.15	1.161895	0.3	0.1344	0.0333333	0.9672	0.3	0.0672

Establishing Equivalence: Pharma Homogeneity Example

- Problem: Homogeneity within a drug substance batch and consistency between batches are required process validation activities expected by FDA
- Data: For sampling a formulated drug product from the hold vessel or during final container filling (vials/syringes), sample from a divided filling period (Beginning, Middle, and End)
- Method: TOST for two samples (Fit Y by X, Fit Model)

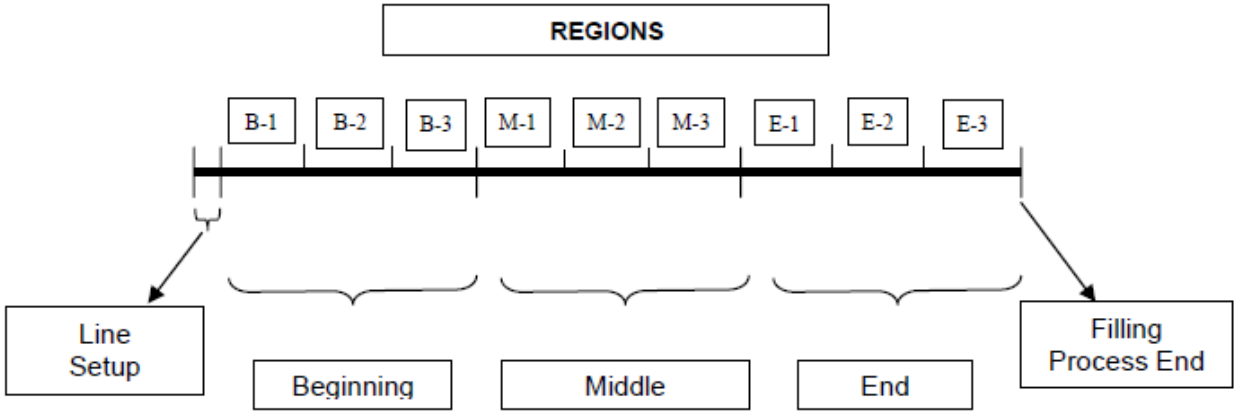


Guidance for Industry

Process Validation: General Principles and Practices

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Veterinary Medicine (CVM)

January 2011
Current Good Manufacturing Practices (CGMP)
Revision 1

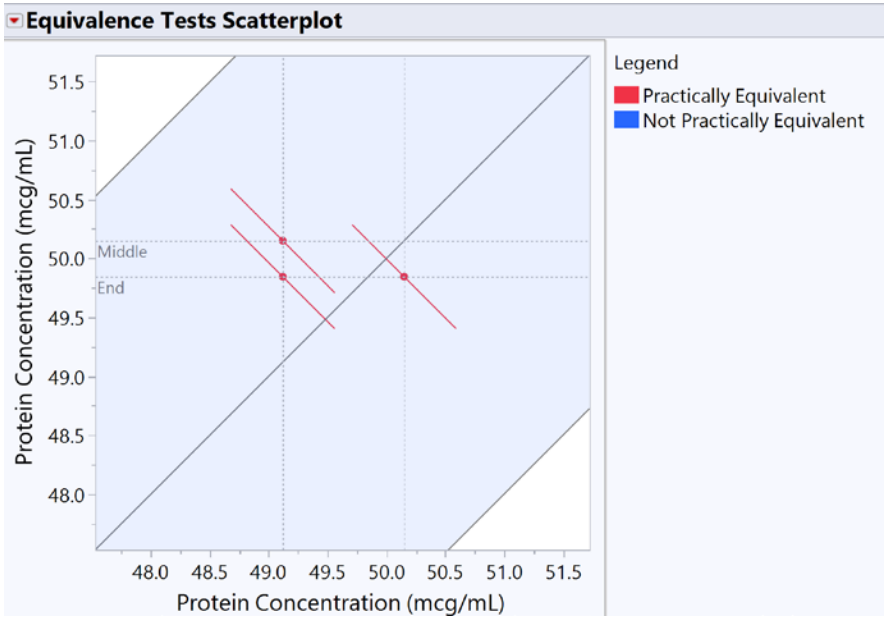


Reference: *Guidance for Industry Process Validation: General Principles and Practices*. Jan 2011.

Establishing Equivalence: Pharma Homogeneity Example

- Conclusion: the product is homogeneous between the three intervals.

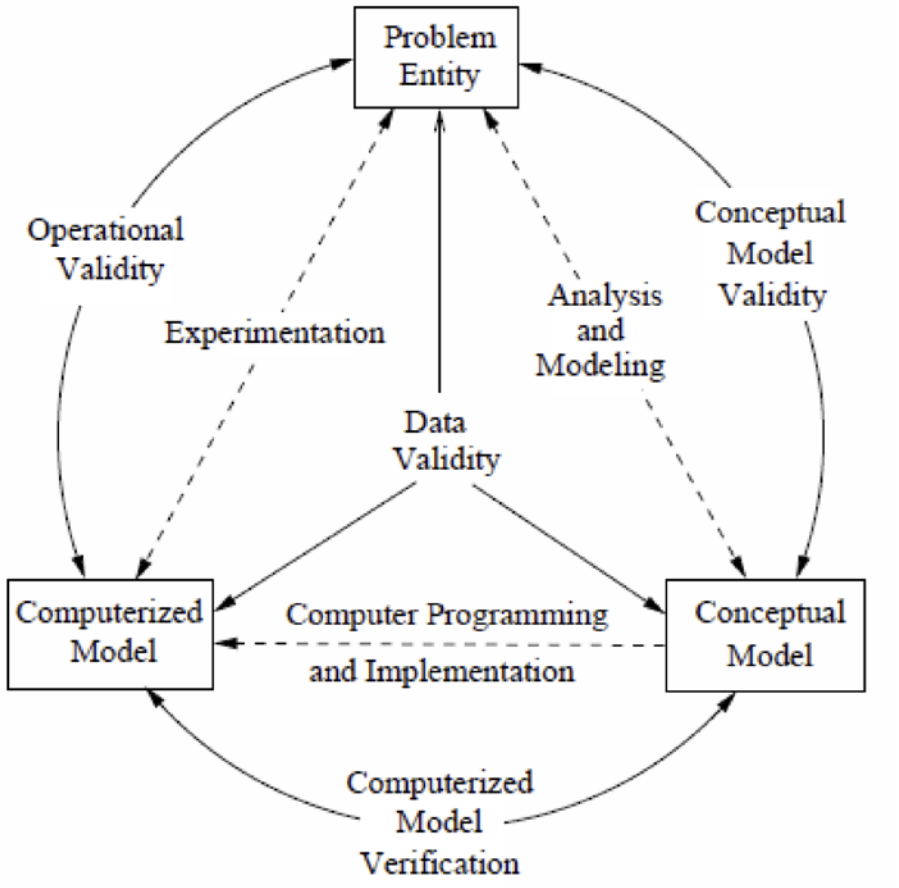
*note it is not required to adjust for experimentwise error (e.g. Tukey HSD) as each contrast must individually pass an average acceptance criterion (EAC)



Equivalence TOST Tests

Sampling Stage	-Sampling Stage	Difference	Lower Bound t Ratio	Upper Bound t Ratio	Lower Bound p-Value	Upper Bound p-Value	Max p-Value	Lower 90%	Upper 90%
Beginning	Middle	-1.03129	-7.91831	3.866977	<.0001*	0.0006*	0.0006*	-1.91411	-0.14846
Beginning	End	-0.72557	-7.31782	4.467465	<.0001*	0.0001*	0.0001*	-1.60840	0.15726
Middle	End	0.30571	-5.29215	6.493131	<.0001*	<.0001*	<.0001*	-0.57711	1.18854

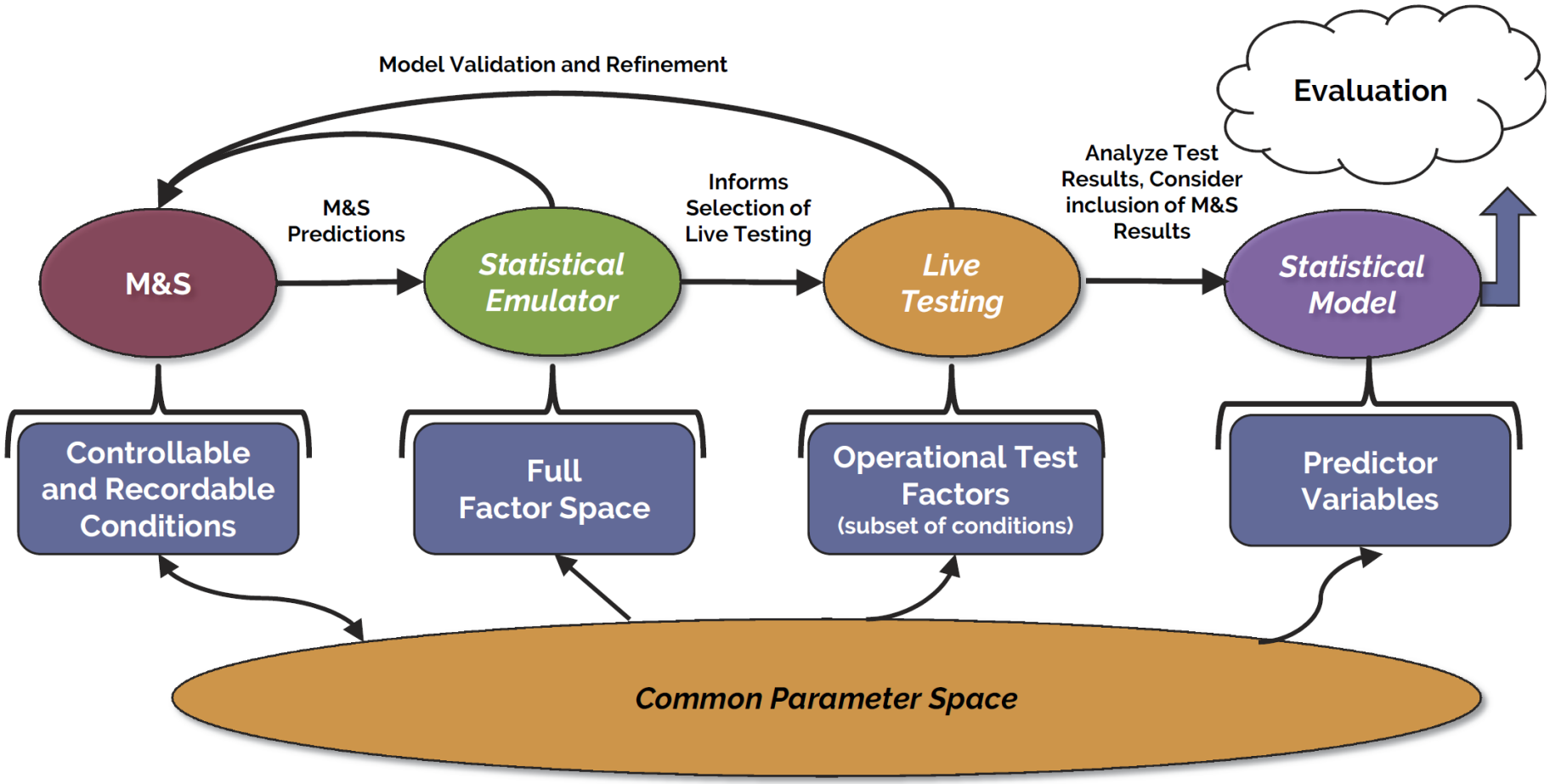
M&S Verification & Validation Process



Conceptual Framework for Modeling and Simulation

Source: DoD M&S VV&A Recommended Practice Guide

Generic M&S Framework



Modeling & Simulation Verification & Validation

- Establish the responses from the live-test data are “equal” to the simulation model

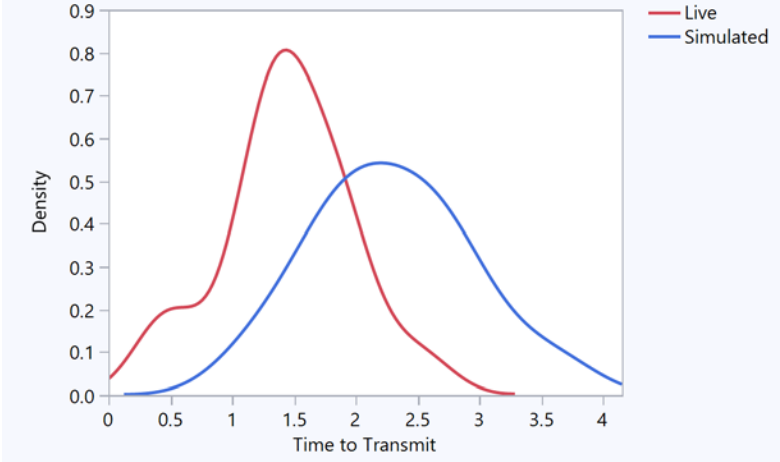
  H_o: The live data is equal to the simulation data

  H_A: The live data is not equal to the simulation data

- What is equal?
  - Means
  - Variances
  - Distribution

$$H_o: \mu_{sim} = \mu_{live}$$

$$H_A: \mu_{sim} \neq \mu_{live}$$



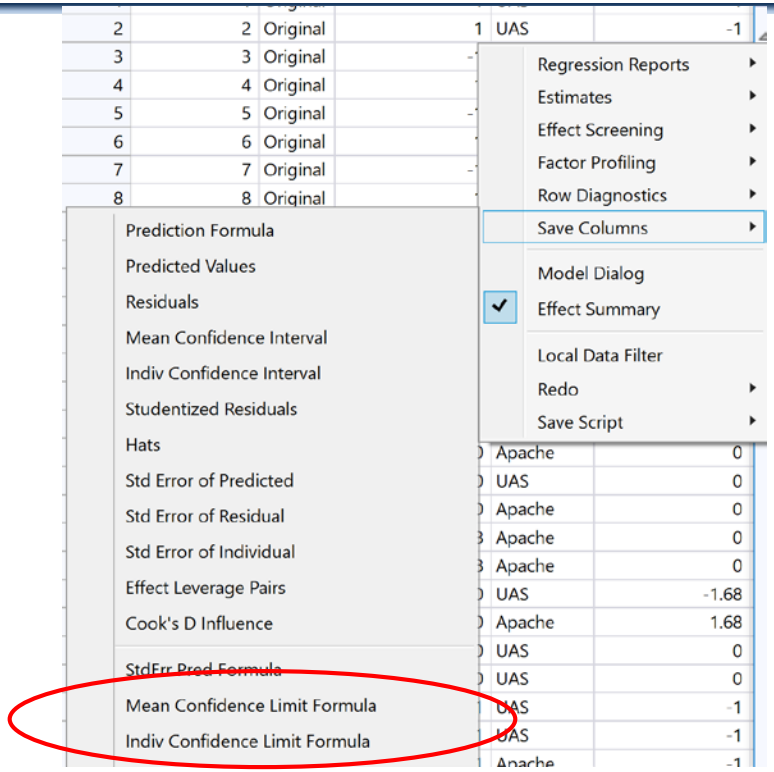
- How close is close enough?
- Want high power and confidence

For M&S V&V, use equivalence methods discussed so far!

M&S V&V Excursion 1: What If Only a Few Live Tests?

- Some programs have complex simulation models with only a few live test events available for validation
- Not possible to have enough runs to cover the operational envelope or make a credible statistical model
- What can be done with these observations for the validation effort?
  - Compare live tests with prediction intervals generated from statistical emulator
  - Generally going to be evaluated on a case by case basis if the point fell in or out of the interval
  - Create plot of actual versus predicted—looking for slope close to 1 with intercept at 0 (anything else is bias)
  - Could do some binomial analysis on the percentage that fall in the prediction interval or not
- Aggressive root cause analysis and investigation needed to determine what happened for those that fell outside the interval—should inform model update

A Single New Live Test: JAGM Example



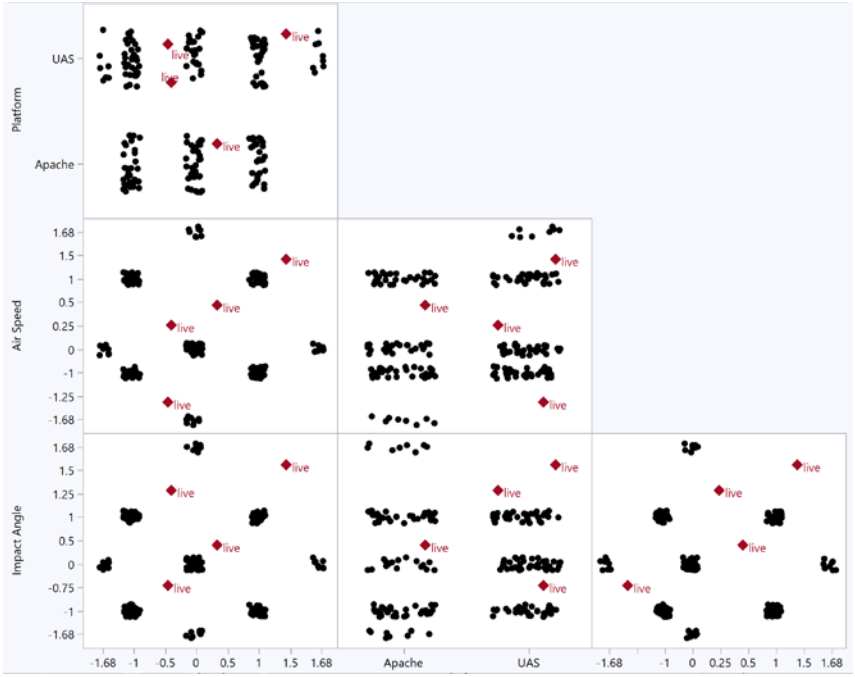
- Mean Confidence Limit (71.6, 73.8)
- Indiv Confidence Limit (65.0, 80.4)

- Consider live test at (.5, Apache, .5, .5)
- Use Save Columns to determine prediction intervals
- Lower 95% Conf Interval Formula

```
(61.1208540437949 + 10.3982831817751 * :Altitude
+Match( :Platform, "Apache", -0.480264606328478, "UAS", 0.480264606328478, . )
+4.36489120144274 * :Airspeed + 7.01550919935969 * :Impact Angle + :Altitude * (
:Airspeed * -9.23618322336606) + Match( :Platform,
    "Apache", :Airspeed * 0.920088354109923,
    "UAS", :Airspeed * -0.920088354109923,
    .
) + :Airspeed * (:Airspeed * 4.72985382124648) + :Altitude * (:Impact Angle *
8.36060184033829) + :Airspeed * (:Impact Angle * -0.866697020012297))
-1.97252818200132 * Sqrt(
    Vec Quadratic(
        [0.0127715158311945 0 -0.000521746026394556 0.0000809925085325671 0 0
        0.00208116725024335 -0.00927515485078563 0 0,
        0 0.00577505913660556 0 0 0 0 0 0 0,
        -0.000521746026394556 0 0.00512785428769679 -0.000796015227983683 0 0
        -0.0000850205063697125 0.000378911575693439 0 0,
        0.0000809925085325671 0 -0.000796015227983683 0.00589862743781533 0 0
        0.0000131980383926964 -0.0000588198040328372 0 0,
        0 0 0 0 0.00577505913660556 0 0 0 0,
        0 0 0 0 0 0.0078125 0 0 0,
        0.00208116725024335 0 -0.0000850205063697125 0.0000131980383926964 0 0
        0.00697803403437223 -0.00348307007249801 0 0,
        -0.00927515485078563 0 0.000378911575693439 -0.0000588198040328372 0 0
        -0.00348307007249801 0.011236097128989 0 0,
        0 0 0 0 0 0 0 0 0.0078125 0,
        0 0 0 0 0 0 0 0 0 0.0078125],
        [1] || :Altitude || Design Nom( :Platform, {"Apache", "UAS"} ) || :Airspeed
         || :Impact Angle || H Direct Product( :Altitude, :Airspeed ) ||
        H Direct Product( Design Nom( :Platform, {"Apache", "UAS"} ), :Airspeed )
         || H Direct Product( :Airspeed, :Airspeed ) ||
        H Direct Product( :Altitude, :Impact Angle ) ||
        H Direct Product( :Airspeed, :Impact Angle )
    ) * 14.8947832299308
)
```

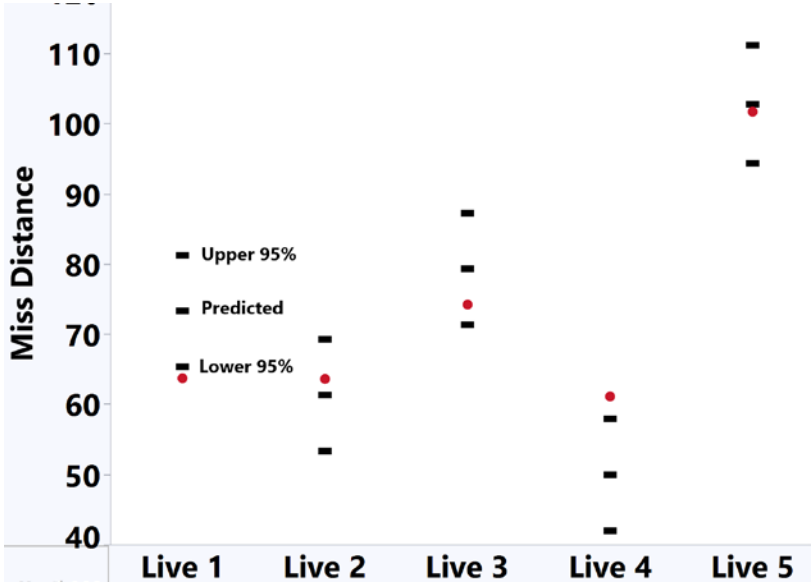
5 New Live Tests: JAGM Example

Design Space for Live



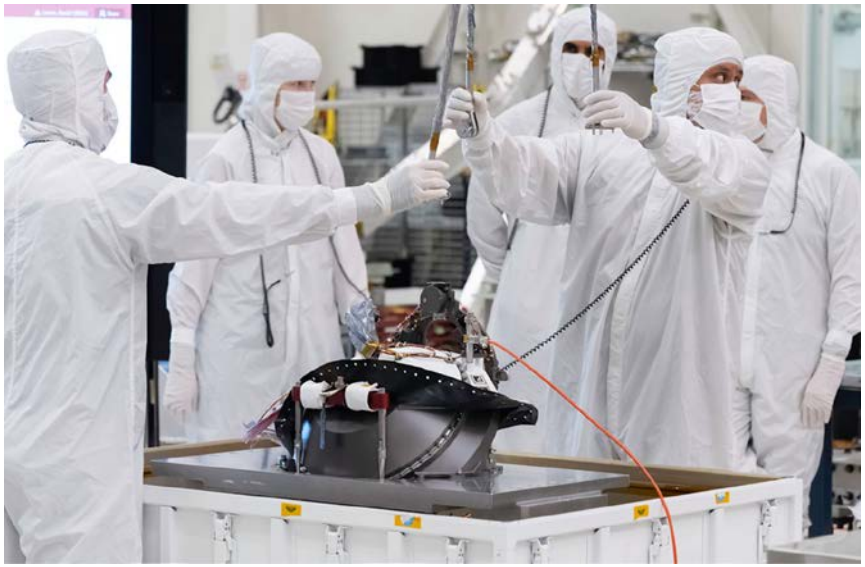
- 5 live tests conducted, not necessarily at the recommended locations
- 2 of 5 fell out of prediction intervals generated by the emulator
- No consistent pattern in misses

Predicted Miss Distance	Lower 95% PI	Upper 95% PI	Actual
73.3	65.3	81.2	63.7
61.3	53.3	69.3	63.6
79.3	71.3	87.2	74.2
49.9	41.9	57.9	61.1
102.7	94.3	111.2	101.7



M&S V&V Excursion 2: Equivalence of Parameters for Characterization

- Good to have response values approximately equal between live and simulated
- Often want to show factors and interactions are approximately equivalent between the two for characterization
- We could test to see if the simulated slope is equal to the live value
- Need to consider joint region

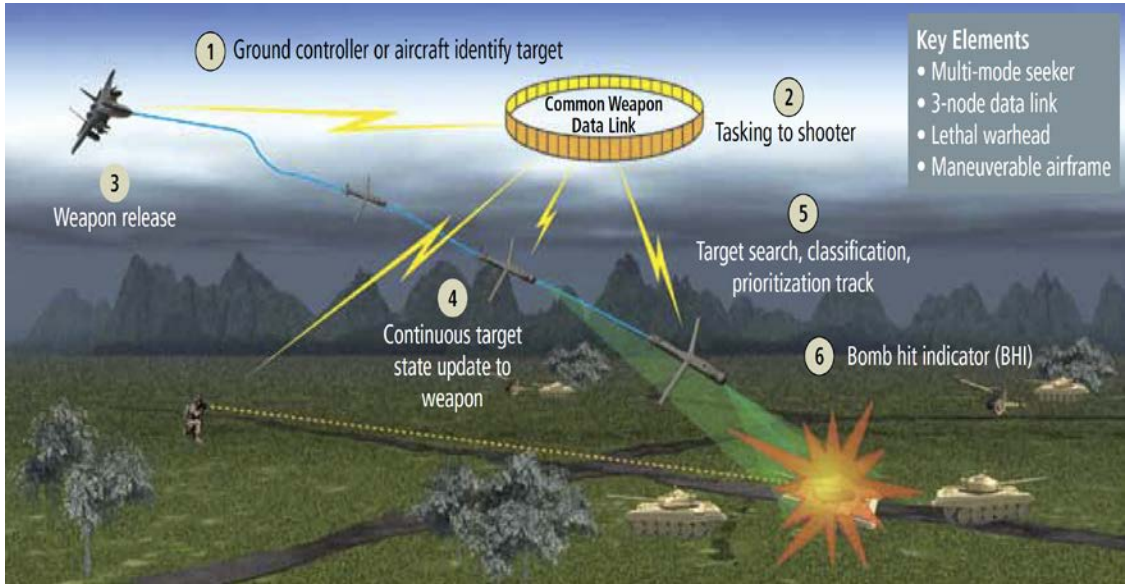


https://www.jpl.nasa.gov/spaceimages/details.php?id=PIA23320

Example: SDB II Weapon Effectiveness

- Problem: Small Diameter Bomb II is a multi-billion dollar system with very expensive test cases; M&S helps characterize performance, but it must be V,V&A'd
- Data: 5 factor response surface design for both live and simulated
- Method: Quick look profiler consistency, compare prediction interval accuracy, parameterize as test type, joint test all slopes

www.airforcemag.com

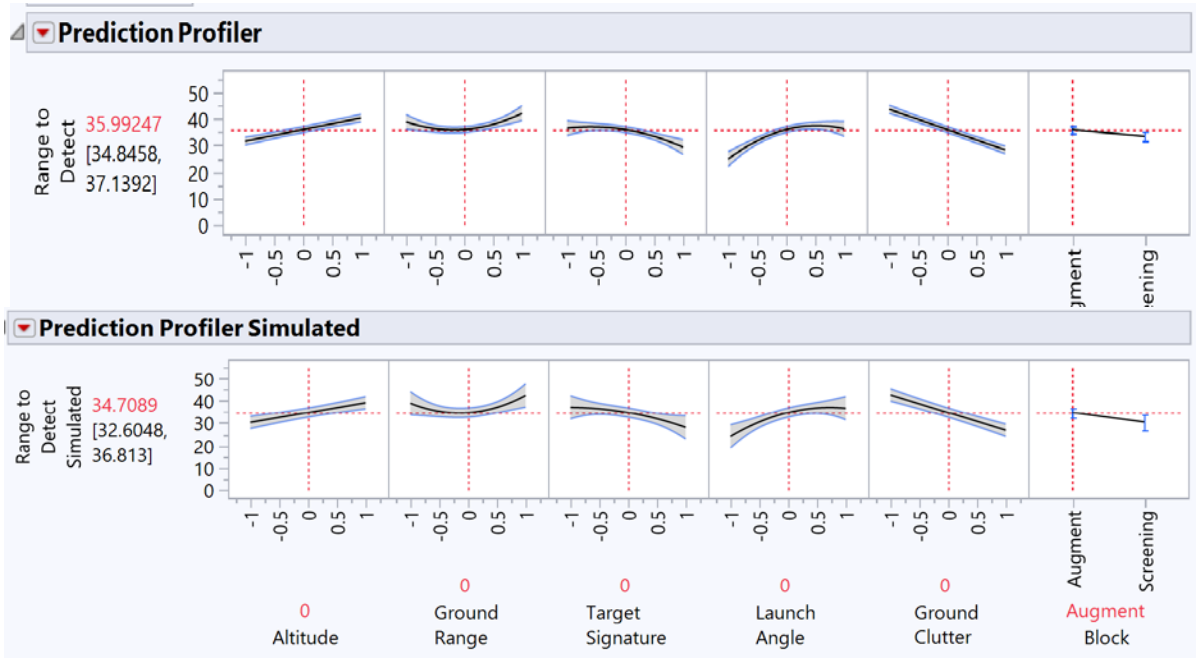


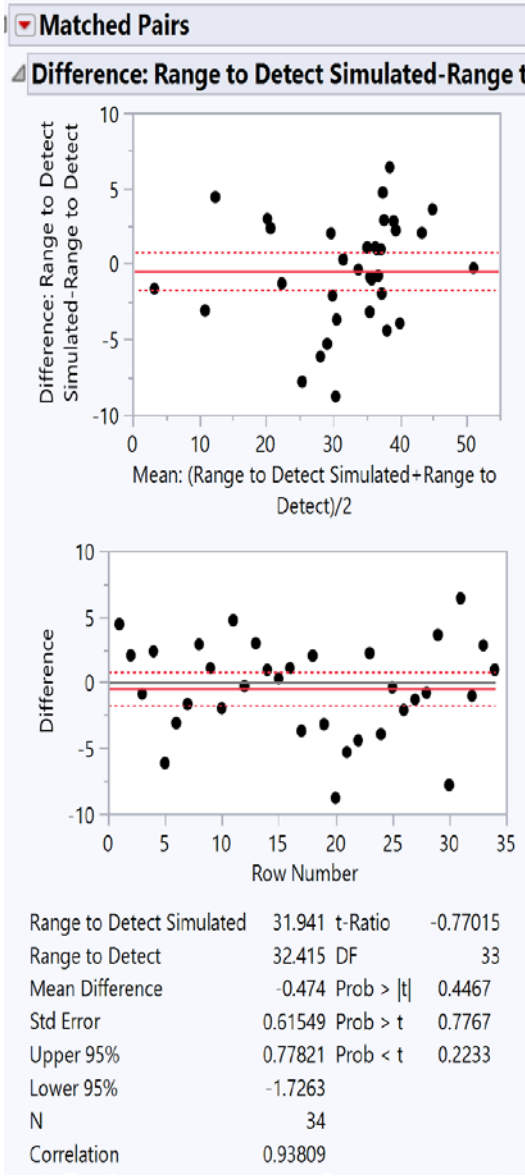
SDB II Quick Comparison

- Take a first cut at determining if the two models are similar

Summary of Fit (Live)	
RSquare	0.972247
RSquare Adj	0.95837
Root Mean Square Error	1.990974
Mean of Response	32.415
Observations (or Sum Wgts)	34

Summary of Fit Simulated	
RSquare	0.917108
RSquare Adj	0.875663
Root Mean Square Error	3.653195
Mean of Response	31.94098
Observations (or Sum Wgts)	34





Range to Detect Simulated	31.941	t-Ratio	-0.77015
Range to Detect	32.415	DF	33
Mean Difference	-0.474	Prob > |t|	0.4467
Std Error	0.61549	Prob > t	0.7767
Upper 95%	0.77821	Prob < t	0.2233
Lower 95%	-1.7263		
N	34		
Correlation	0.93809		

SDB II M&S V&V: Parameterizing Live vs Simulated

- Pool live and M&S data to build statistical model
  - Create binary indicator *TestType* for live or M&S
  - If statistically significant then not getting consistent results
  - Use indicator with interactions also to see if sensitive to some conditions
  - Method works best if you have a designed experiment for both live and simulated
  - Example:

$$Detection\ Range = \beta_0 + \beta_1 TestType + \beta_2 Threat + \beta_3 (TestType * Threat) + \epsilon$$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	34.630217	0.626894	55.24	<.0001*
Altitude	3.4949434	0.500566	6.98	<.0001*
Ground Range	1.3366537	0.500566	2.67	0.0099*
Target Signature	-4.739092	0.500566	-9.47	<.0001*
Launch Angle	5.6682212	0.500566	11.32	<.0001*
Ground Clutter	-7.687441	0.500566	-15.36	<.0001*
Ground Range*Ground Range	2.7978296	1.189794	2.35	0.0223*
Altitude*Launch Angle	1.4849838	0.530931	2.80	0.0071*
Ground Range*Launch Angle	1.7780303	0.530931	3.35	0.0015*
Launch Angle*Launch Angle	-6.192733	1.189794	-5.20	<.0001*
Launch Angle*Ground Clutter	5.0953628	0.530931	9.60	<.0001*
Model[Live]	-0.082358	0.364216	-0.23	0.8219
Block[Augment]	1.901491	0.511187	3.72	0.0005*

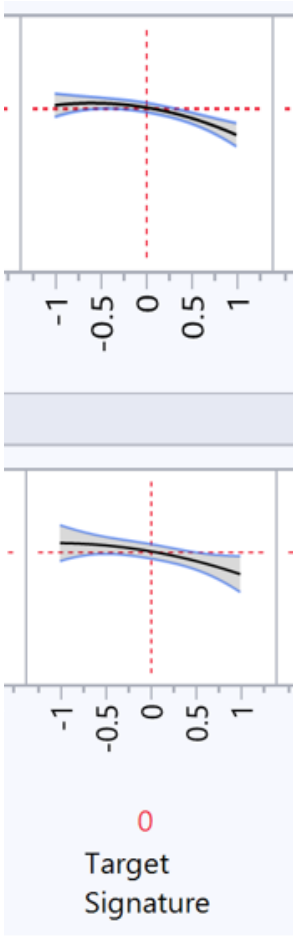
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	34.630217	0.601434	57.58	<.0001*
Altitude	3.4949434	0.480237	7.28	<.0001*
Ground Range	1.3366537	0.480237	2.78	0.0076*
Target Signature	-4.739092	0.480237	-9.87	<.0001*
Launch Angle	5.6682212	0.480237	11.80	<.0001*
Ground Clutter	-7.687441	0.480237	-16.01	<.0001*
Ground Range*Ground Range	2.7978296	1.141473	2.45	0.0178*
Altitude*Launch Angle	1.4849838	0.509368	2.92	0.0053*
Ground Range*Launch Angle	1.7780303	0.509368	3.49	0.0010*
Launch Angle*Launch Angle	-6.192733	1.141473	-5.43	<.0001*
Launch Angle*Ground Clutter	5.0953628	0.509368	10.00	<.0001*
Model[Live]	-0.082358	0.349424	-0.24	0.8146
Block[Augment]	1.901491	0.490426	3.88	0.0003*
Ground Clutter*Model[Live]	-0.017559	0.480237	-0.04	0.9710
Launch Angle*Model[Live]	0.0812233	0.480237	0.17	0.8664
Target Signature*Model[Live]	1.1674253	0.480237	2.43	0.0187*
Ground Range*Model[Live]	0.2861241	0.480237	0.60	0.5540
Altitude*Model[Live]	0.8933899	0.480237	1.86	0.0687

SDB II M&S V&V: Comparison of a Single Beta

- We can also formally test the differences in slopes between the live and simulated value

Parameter	Estimate Live	Estimate Simulated	Std Error Live
Intercept	34.73	32.60	0.61
Altitude	4.39	4.39	0.47
Ground Range	1.62	1.74	0.47
Target Signature	-3.57	-4.46	0.47
Launch Angle	5.75	6.27	0.47
Ground Clutter	-7.71	-7.85	0.47
Ground Range*Ground Range	4.58	5.91	1.20
Target Signature*Target Signature	-3.07	-2.13	1.20
Ground Range*Launch Angle	1.63	1.46	0.50
Launch Angle*Launch Angle	-5.47	-4.33	1.20
Launch Angle*Ground Clutter	4.63	4.67	0.50



Custom Test

Tgt Sig^2

Parameter	
Intercept	0
Altitude	0
Ground Range	0
Target Signature	0
Launch Angle	0
Ground Clutter	0
Ground Range*Ground Range	0
Target Signature*Target Signature	1
Ground Range*Launch Angle	0
Launch Angle*Launch Angle	0
Launch Angle*Ground Clutter	0
Block[Augment]	0
=	-2.13

Value	-0.935771277
Std Error	1.2017993793
t Ratio	-0.778641837
Prob>|t|	0.4444855545
SS	2.4032928735

Sum of Squares	2.4032928735
Numerator DF	1
F Ratio	0.6062831106
Prob > F	0.4444855545

SDB II M&S V&V: Comparison of a Single Beta

- Test for all parameters from live design are equal to the values given by simulated—Custom Test (*F*)
- Not enough evidence to suggest the joint regression surface differs between the two

Custom Test

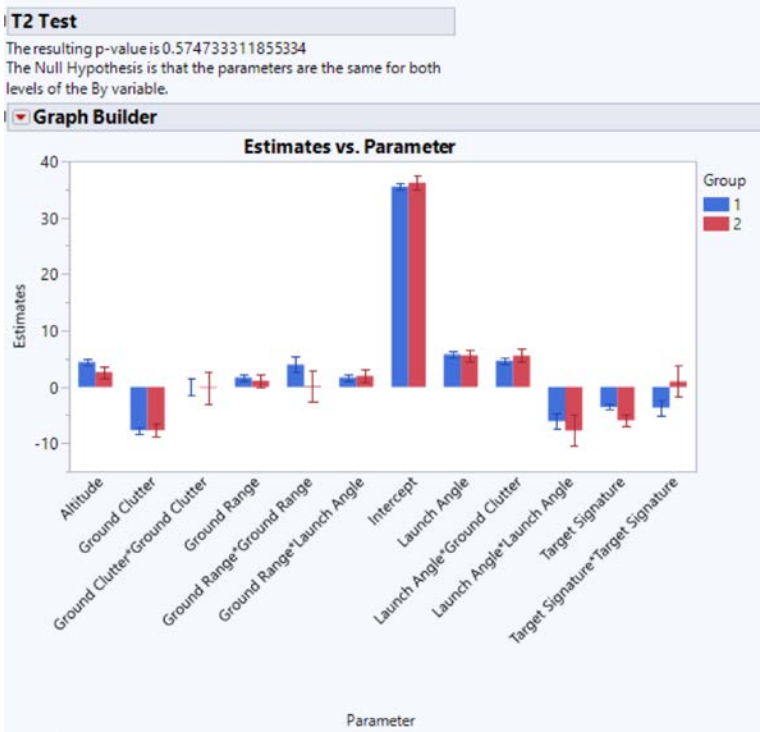
All

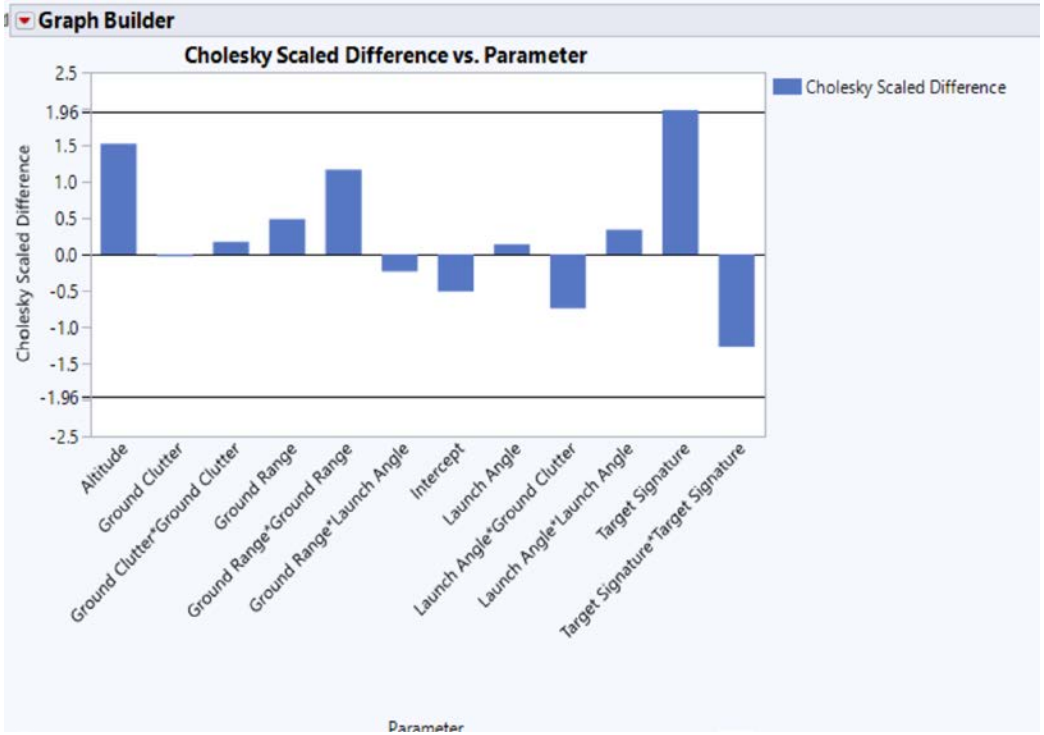
Parameter											
Intercept	1	0	0	0	0	0	0	0	0	0	0
Altitude	0	1	0	0	0	0	0	0	0	0	0
Ground Range	0	0	1	0	0	0	0	0	0	0	0
Target Signature	0	0	0	1	0	0	0	0	0	0	0
Launch Angle	0	0	0	0	1	0	0	0	0	0	0
Ground Clutter	0	0	0	0	0	1	0	0	0	0	0
Ground Range*Ground Range	0	0	0	0	0	0	1	0	0	0	0
Target Signature*Target Signature	0	0	0	0	0	0	0	1	0	0	0
Ground Range*Launch Angle	0	0	0	0	0	0	0	0	1	0	0
Launch Angle*Launch Angle	0	0	0	0	0	0	0	0	0	1	0
Launch Angle*Ground Clutter	0	0	0	0	0	0	0	0	0	0	1
Block[Augment]	0	0	0	0	0	0	0	0	0	0	0
=	32.6	4.39	1.74	-4.46	6.27	-7.85	5.91	-2.13	1.46	-4.33	4.67
Value	2.127912234	-0.001666667	-0.117222222	0.8883333333	-0.520555556	0.145	-1.325771277	-0.935771277	0.170625	-1.135771277	-0.039375
Std Error	0.6052699268	0.4692770999	0.4692770999	0.4692770999	0.4692770999	0.4692770999	1.2017993793	1.2017993793	0.4977435294	1.2017993793	0.4977435294
t Ratio	3.5156417656	-0.003551562	-0.249793187	1.8929824907	-1.109271166	0.308985885	-1.103155235	-0.778641837	0.3427970228	-0.945058964	-0.079107005
Prob>|t|	0.001949524	0.9971982781	0.8050649546	0.0715898633	0.2792921838	0.7602394455	0.2818810958	0.4444855545	0.7350049285	0.3548933995	0.9376626652
SS	48.99372486	0.00005	0.2473388889	14.20445	4.8776055556	0.37845	4.8239687859	2.4032928735	0.46580625	3.5403731655	0.02480625

Sum of Squares	70.806241284
Numerator DF	11
F Ratio	1.6238564213
Prob > F	0.1603510952

Hotelling T^2 to Test $\boldsymbol{H_0}$: $\boldsymbol{\beta_{Sim} = \beta_{Live}}$

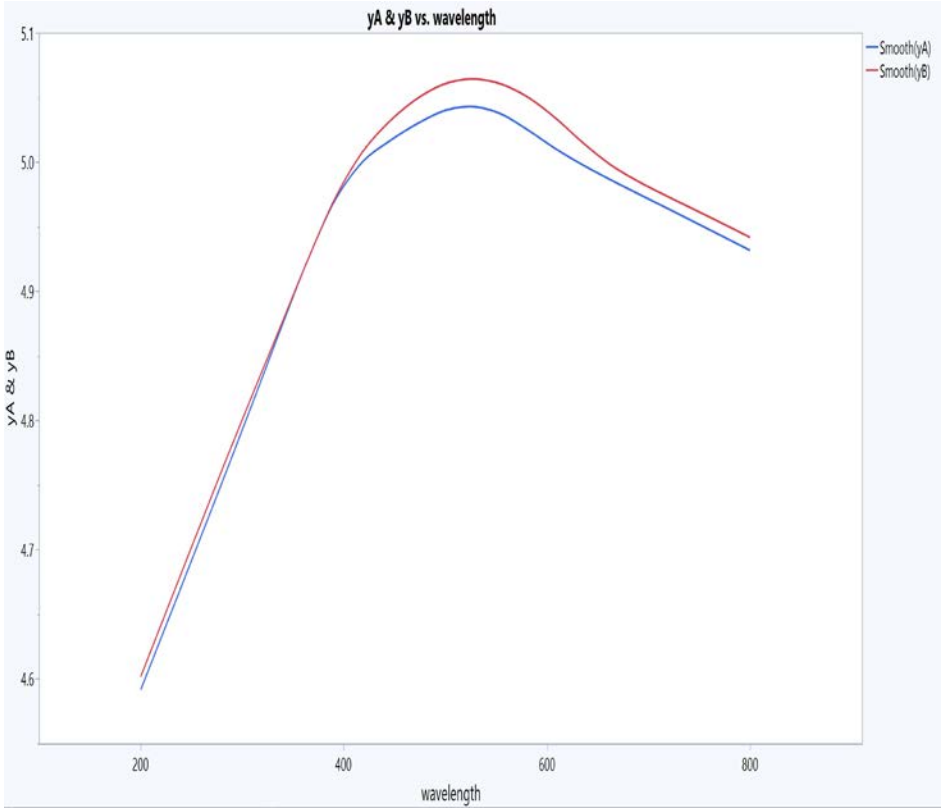
- JMP does not compute the combined covariance matrix
- JSL script uses the correct combined covariance structure to determine the T^2 test statistic and reports a *p*-value based on the Chi-Square

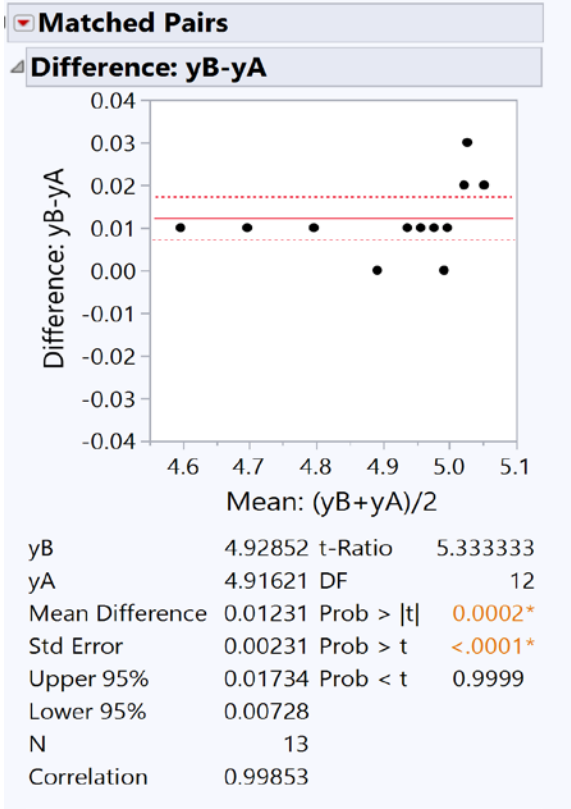




Equivalence in Curves

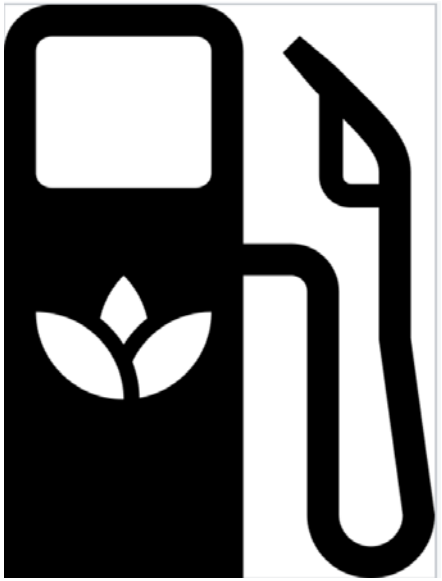
- Often we need to establish two or more responses over a continuum are equal (e.g. time series, instrumentation data,
- Possible to take differences at discrete points or min, max, average etc, but truly miss the functional form





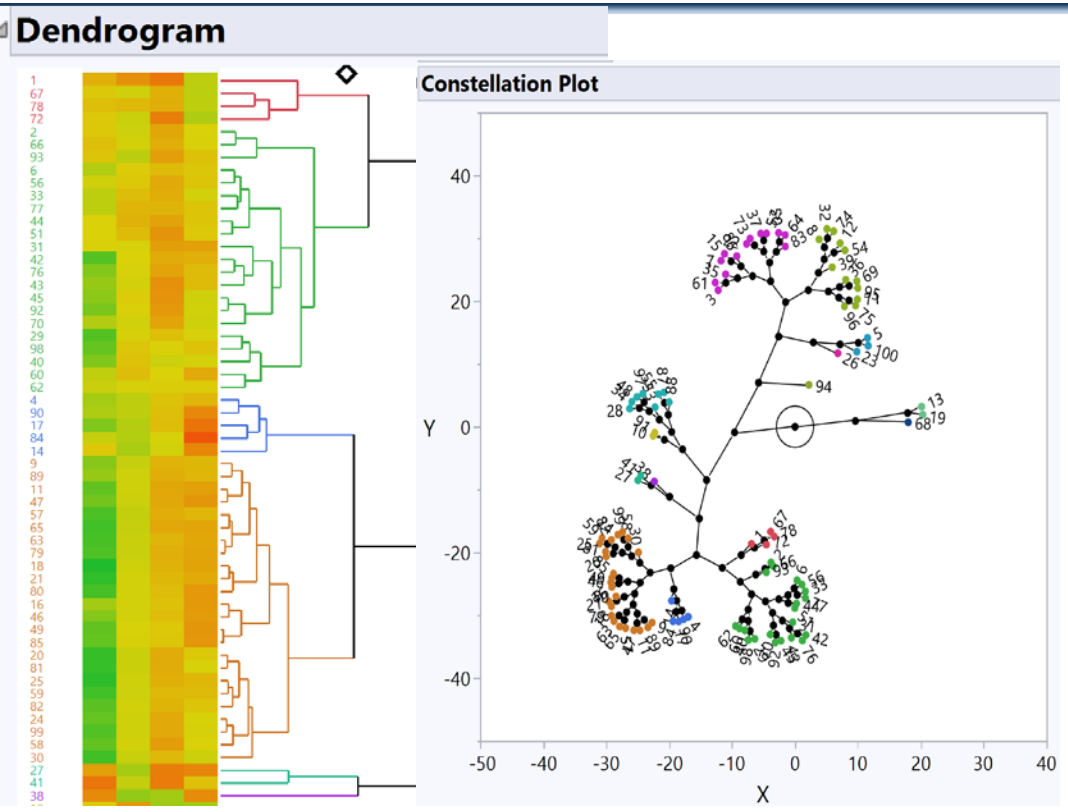
Equivalence in Curves

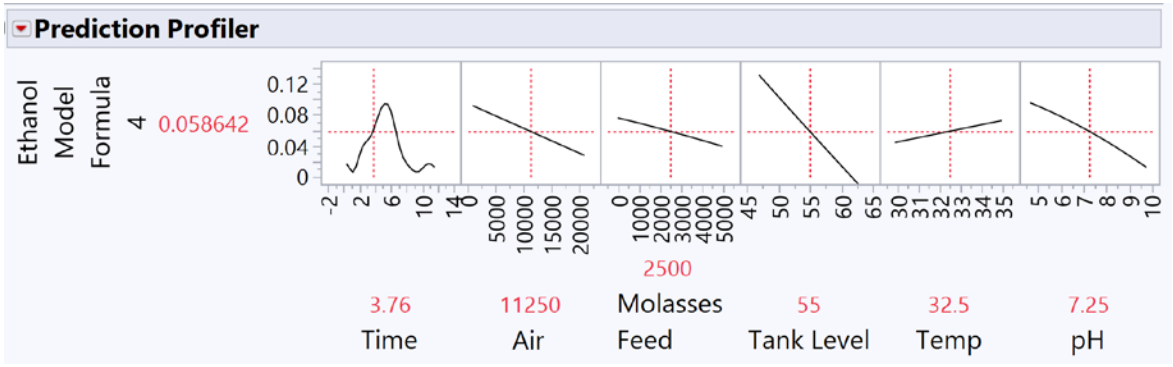
- Functional data as responses is prevalent across many industries
- Same need to establish parameters equal to specific values or response curves are equivalent to one another or a standard
- Use example data set Fermentation that looks ethanol production



https://en.wikipedia.org/wiki/Biofuel

Equivalence in Curves





- Cluster analysis of FPCs can group like curves with many graphics and metrics
- May have “Ideal” curve you want to establish for equivalence
- Profiler links factors to original functions
- Could have put in factor for Live or Simulated

Summary

- Analysis objectives are often in practice to demonstrate that a process is within certain levels of equivalence
- *Fail to reject* alone is a necessary condition, but not sufficient
- JMP has many platforms where the workflow is already integrated with proper test statistics and visuals to tell the story

Questions?

Hotelling T^2 to Test $\boldsymbol{H_0}$: $\boldsymbol{\beta_{Sim} = \beta_{Live}}$

- Given the simulation distribution (1) and live distribution (2)

$$\hat{\boldsymbol{\beta}}_{\mathbf{1}} \dot{\sim} N_p(\boldsymbol{\beta}_{\mathbf{1}}, \boldsymbol{\Sigma}_{\mathbf{1}}) \qquad \hat{\boldsymbol{\beta}}_{\mathbf{2}} \dot{\sim} N_p(\boldsymbol{\beta}_{\mathbf{2}}, \boldsymbol{\Sigma}_{\mathbf{2}})$$

$$\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \hat{\boldsymbol{\beta}}_{\mathbf{2}} \dot{\sim} N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{1}} + \boldsymbol{\Sigma}_{\mathbf{2}}) \text{ if } \hat{\boldsymbol{\beta}}_{\mathbf{1}} = \hat{\boldsymbol{\beta}}_{\mathbf{2}}$$

- Still assuming $\hat{\boldsymbol{\beta}}_{\mathbf{1}} = \hat{\boldsymbol{\beta}}_{\mathbf{2}}$, the quantity

$$\left(\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \hat{\boldsymbol{\beta}}_{\mathbf{2}}\right)^{\mathsf{T}} \left(\boldsymbol{\Sigma}_{\mathbf{1}} + \boldsymbol{\Sigma}_{\mathbf{2}}\right)^{-1} \left(\hat{\boldsymbol{\beta}}_{\mathbf{1}} - \hat{\boldsymbol{\beta}}_{\mathbf{2}}\right)$$

follows the Chi-Square Distribution with *p* degrees of freedom

Note: estimates are asymptotically normal around true estimates; procedure does not account for variability in covariance matrices which may lead to slightly increased Type I error rate=>consider using α=.01 to .025 for small samples to approximate a .05 error