

Applying Statistics Recommended by Regulatory Documents

Steven Walfish

President, Statistical Outsourcing Services

steven@statisticaloutsourcingservices.com

301-325-3129



About the Speaker

- Mr. Steven Walfish is the President of Statistical Outsourcing Services, a consulting company that provides statistical analysis and training to variety of industries. Prior to starting Statistical Outsourcing Services, Mr. Walfish was the Senior Manager Biostatistics, Non-clinical at Human Genome Sciences in Rockville MD. Prior to joining HGS, Mr. Walfish was a Senior Associate at PricewaterhouseCoopers specializing in the pharmaceutical industry.
- Mr. Walfish brings over 18 years of industrial expertise in the development and application of statistical methods for solving complex business issues including data collection, analysis and reporting. Mr. Walfish has held positions with Johnson & Johnson and Chiron Diagnostics where he worked with large data sets for monitoring process data.
- Mr. Walfish holds a Bachelors of Arts in Statistics from the University of Buffalo, Masters of Science in Statistics from Rutgers University and an Executive MBA from Boston University.



Learning Objectives

- Understand regulations that govern method validation.
- Using the correct statistical methods.
- How to Plan experiments



Agenda

- Regulatory Documents that Govern Method Validation
- Components of Method Validation
- Other Statistical Methods
- Integrating Statistical Methods with Regulatory Requirements
- Interactive Exercise: Case Study



Regulatory Documents

- USP
 - General Chapter <1225>, Validation of compendial methods, *United States Pharmacopeia XXIII*, National Formulary, XVIII, Rockville, MD, The United States Pharmacopeial Convention, Inc, 1995, 1710–1612
- ICH
 - Q2A
 - Q2B
- FDA Guidance on Bioanalytical Method Validation
- FDA Draft Guidance for Industry
 - **Analytical Procedures and Methods Validation; Chemistry, Manufacturing, and Controls Documentation**



Compendial Methods

- All USP methods are considered validated according to the guidelines in place at the time they were first published.
- Many USP methods are several years old, and what may have passed for validation in the past has changed significantly.
- A minimal re-validation of an USP method to prove suitability for your product should be conducted.



ICH Guidelines

- **Specificity**
 - **Linearity**
 - **Range**
 - **Accuracy**
 - **Precision**
 - Repeatability
 - Intermediate precision
 - **Detection limit**
 - **Quantification limit**
- 

Specificity

- Specificity is the degree of bias (or lack thereof) caused by expected sample components and common interferences, determined by measuring the analyte with and without anticipated interferences.



t-test

- The t-test is a statistical test for comparing two sample means.
- A common application of this process is to test if an interfering substance has a significant effect on the response variable (Specificity of the method).
- Compare the “spiked” samples to the neat samples.

t-test

- The t-test has the following general formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- where n_1 and n_2 are the sample sizes, \bar{X}_1 and \bar{X}_2 are the sample means, and s_1^2 and s_2^2 are the sample variances. The degrees of freedom are n_1+n_2-2

Dunnett's Test

- A procedure for comparing each experimental mean with the control mean.
- Dunnett test is more powerful than tests designed to compare each mean with each other mean.



Specificity Example

Sample	Response
0.5	1
0.5	1.02
0.5	1.04
0.5	1.04
0.5	1.04
0.25	0.95
0.25	0.97
0.25	0.93
0.25	0.93
0.25	0.97
0.125	0.84
0.125	0.86
0.125	0.84
0.125	0.84
0.125	0.85
0.0625	0.81
0.0625	0.83
0.0625	0.84
0.0625	0.83
0.0625	0.83

- Sample is serially diluted from a neat sample of 0.5 mg/ml
- 0.25 is the first sample that is statistically different than 0.0625. ($p < 0.001$)
- 0.125 sample p-value using Dunnett test is 0.185.

Linearity

- Linearity is the variance in the response with concentration, measured as the RSD of response factors or the correlation coefficient (R^2) from a linear regression fit.

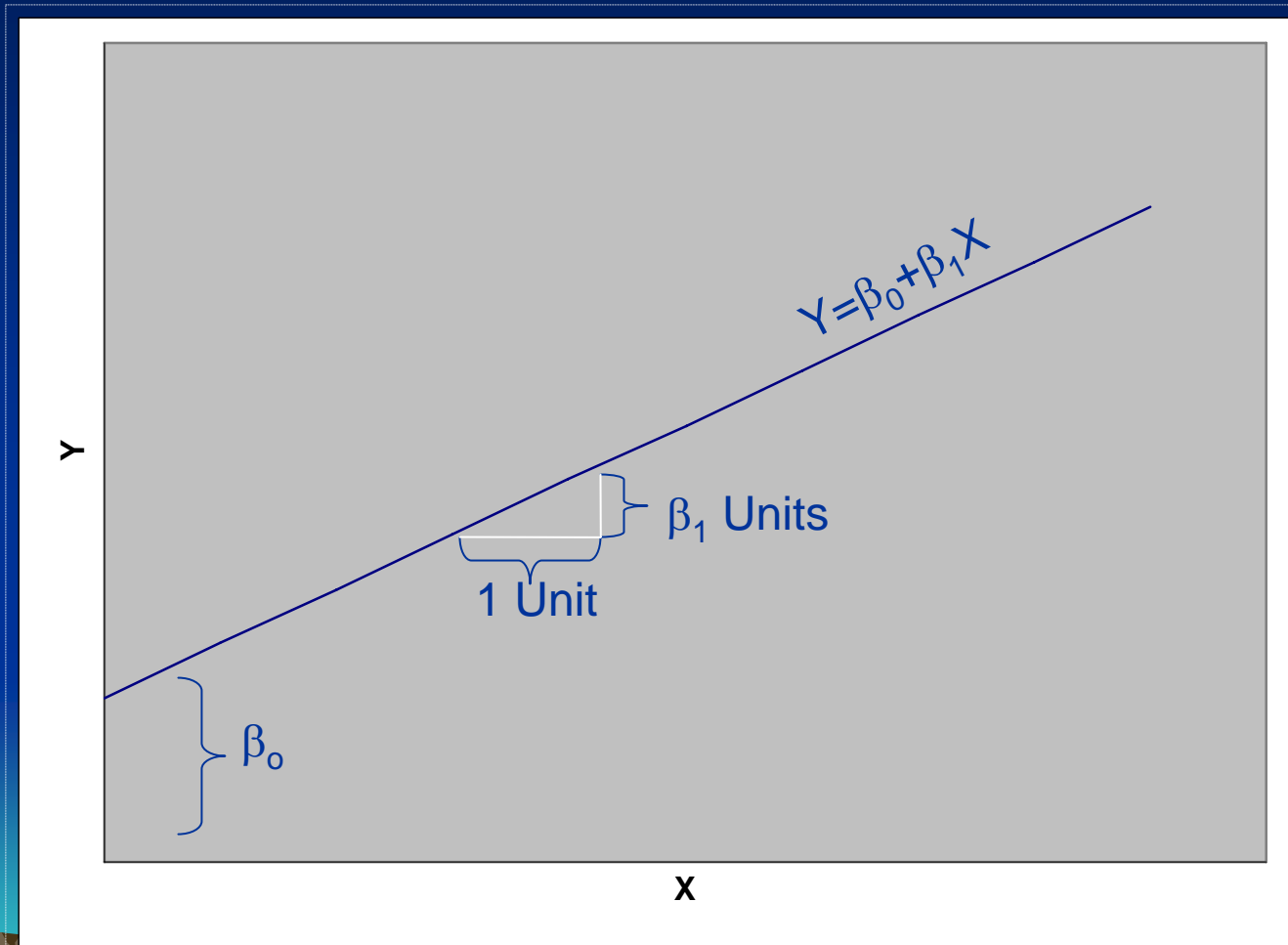


Regression

- The ordinary least squares regression model has a single regressor x that has a straight line relationship with respect to y .
- The intercept (β_0) is the average y -value when x is equal to zero.
- The slope (β_1) is the change in y for a unit change in x .
- All models have some random error component (ε_i).
- The simple linear model is: $y = \beta_0 + \beta_1 x + \varepsilon_i$



Simple Regression Model



Hazards of Regression

- Extrapolating beyond the range of the x values.
- Influential observations or outliers give misleading models.
- Regression does not mean cause and effect.
- The regression of $y=x$ is not the same as the regression of $x=y$.



Prediction Intervals

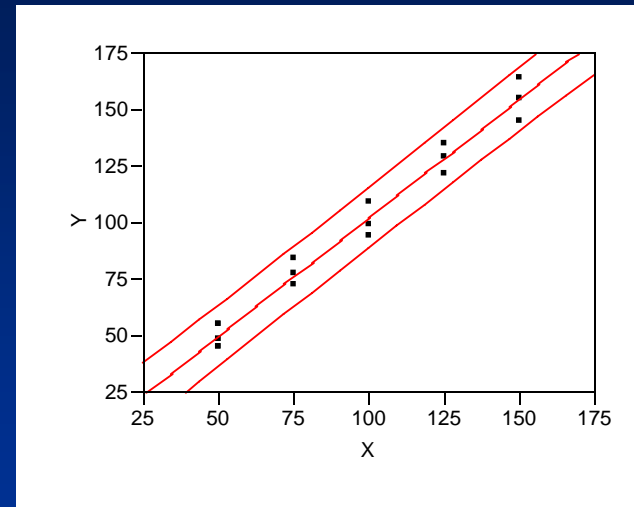
- A potential use for regression analysis is to predict future values of the response variable.
- The predicted values are computed by plugging the value of each x variable into the regression equation.
- An interval around the regression line can be calculated that gives a $(1-\alpha)\%$ prediction interval for a new observation.



Linearity Example

$$Y = -2.733333 + 1.048 X$$

X	Y
50	45
50	48
50	55
75	72
75	84
75	77
100	94
100	99
100	109
125	135
125	129
125	121
150	145
150	155
150	164



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	27457.600	27457.6	744.1084
Error	16	590.400	36.9	Prob > F
C. Total	17	28048.000		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-2.733333	3.801644	-0.72	0.4825
X	1.048	0.038419	27.28	<.0001

Range

- Range is the concentration upper and lower levels which meets the linearity, precision, and accuracy performance characteristics.



Accuracy

- **Accuracy** is closeness to the true value, measured by % recovery of sample spikes or % error in the analysis of a reference sample.



Tolerance Intervals

- Statistical tolerance limits are limits within which we expect a stated proportion of the population to lie.
- What interval will contain p percent of the population measurements?
- Tolerance intervals take the form of:

$$\bar{X} \pm ks$$

- k is a constant such that the interval will cover p percent of the population with at least p percent of the population at a certain confidence.



Accuracy Example

Target	Actual	Mean	Std Dev	% Recovery	Lower CI	Upper CI
50	45					
50	48					
50	55	49.33	5.13	98.67%	36.59	62.08
75	72				73%	113%
75	84					
75	77	77.67	6.03	103.56%	62.69	92.64
100	94				84%	120%
100	99					
100	109	100.67	7.64	100.67%	81.69	119.64
125	135				82%	110%
125	129					
125	121	128.33	7.02	102.67%	110.89	145.78
150	145				89%	120%
150	155					
150	164	154.67	9.50	103.11%	131.06	178.28
					87%	109%

Precision

- Precision is the degree of agreement between replicate analyses of an homogenous sample, usually measured as the relative percent difference (RPD) between duplicates or the relative standard deviation (RSD) of a set of replicates.
- The ICH Guideline defines three precision measurements:
 - repeatability or short term precision
 - intermediate precision which is essentially the same as intra-lab robustness
 - reproducibility which is essentially the same as inter-lab ruggedness.



Variance Components

- Variance components, or “decomposition” of variance is a method to partition the different sources of variation.
- Repeatability expresses the precision under the same operating conditions over a short interval of time. Repeatability is also termed intra-assay precision .
- Intermediate precision expresses within-laboratories variations: different days, different analysts, different equipment, etc.
- Q2A and Q2B talk about reproducibility being the lab to lab variability. Some might include it in the intermediate precision.

Precision Example

Analyst	Run	Rep	Purity
1	1	1	99
1	1	2	96
1	1	3	95
1	2	1	95
1	2	2	89
1	2	3	89
1	3	1	92
1	3	2	91
1	3	3	90
2	1	1	88
2	1	2	92
2	1	3	88
2	2	1	86
2	2	2	85
2	2	3	83
2	3	1	84
2	3	2	86
2	3	3	86

- Intermediate precision is analyst + run.
- Repeatability is rep.
- Mean Purity = 89.7%
- Variance Analyst = 17.92
- Variance Run (within Analyst) = 7.07
- Variance Rep = 4.39

Source	Variance	CV	% Variance
Analyst	17.92	4.7%	61.0%
Run	7.07	3.0%	24.1%
Error	4.39	2.3%	14.9%

Detection Limit

- Detection limit is the lowest concentration which can be detected with confidence (usually at the 99% confidence level), determined as three times the standard deviation (SD) of the background signal or the low level sample spikes.



Quantification Limit

- Quantification Limit is the concentration level above which the concentration can be determined with acceptable precision (usually $RSD < 10$ to 25%) and accuracy (usually 80 - 120% recovery), usually determined as ten times the SD of the background or low level sample spikes.



Robustness and Ruggedness

- Robustness is usually done in development to assess the performance of the test method after varying method parameters.
- Ruggedness is a measure of a test method's capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage
- Design of Experiments can help to assess the robustness and ruggedness of the method.



Non-Normal Data

- How do we validate methods with non-normal data?
 - Use Non-parametric statistics.
 - Cell based methods use logistic regression.
 - Some nonlinear regression models are applicable.



Nonparametric Statistics

- Nonparametric statistical methods assume no underlying distribution.
- Median and range replace mean and standard deviation.
- Nonparametric methods are useful when:
 - The measurements are only categorical; i.e., they are nominal or ordinal scaled
 - The assumptions underlying the use of parametric methods cannot be met.



Advantage of Nonparametric Tests

- They may be used on all types of data- categorical data, which are nominally scaled or are in rank form, called ordinally scaled, as well as interval or ratio-scaled data.
- For small sample sizes they are easy to apply.
- They make fewer and less stringent assumptions than their parametric counterparts.



Disadvantages of Nonparametric Tests

- If the assumptions of the parametric methods can be met, it is generally more efficient to use them.
- For large sample sizes, data manipulations tend to become more laborious, unless computer software is available.
- Often special tables of critical values are needed for the test statistic, and these values cannot always be generated by computer software.

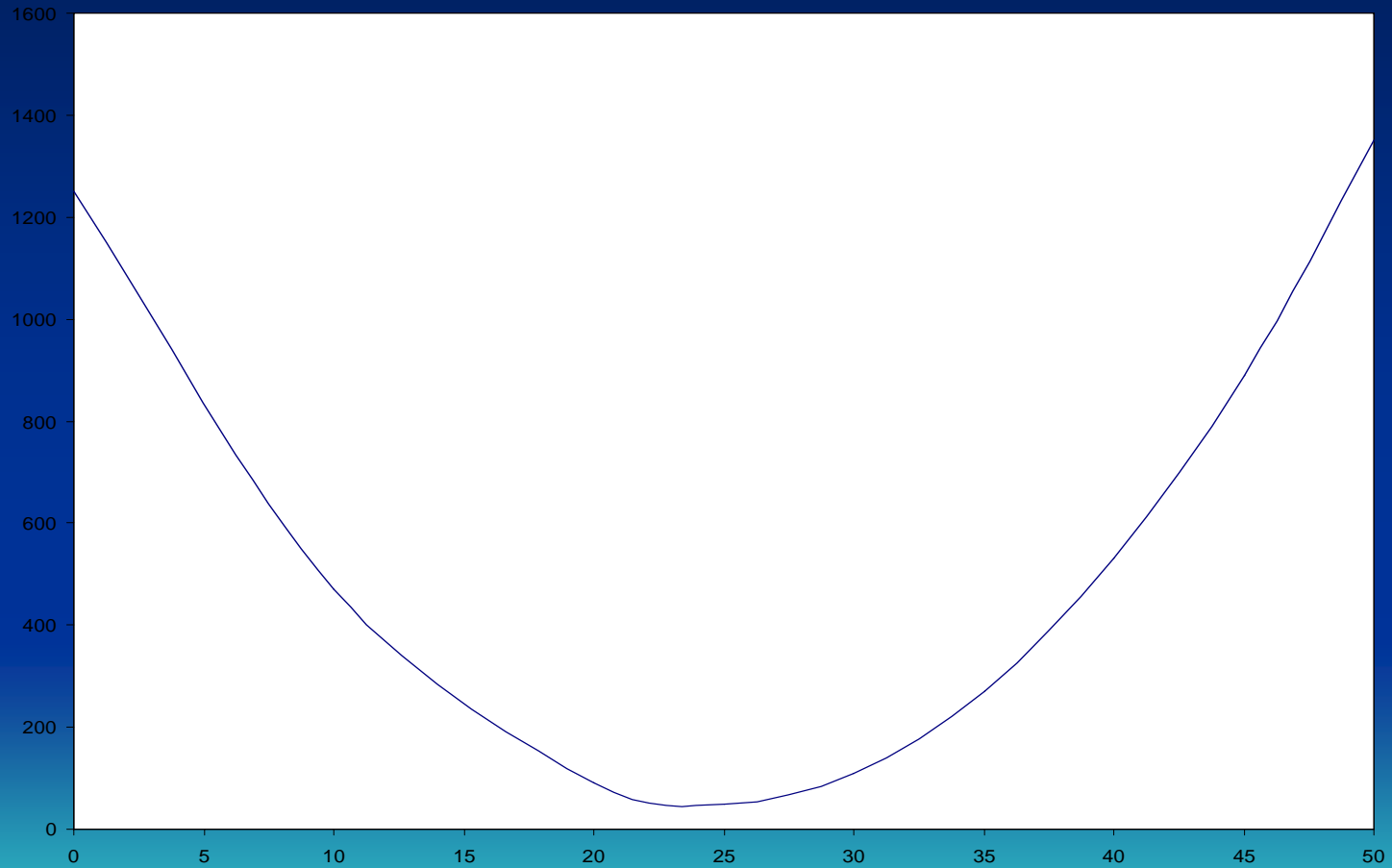
Nonlinear Regression

- Nonlinear models often occur as a result of a physical phenomenon such as exponential growth or decay.
- The nonlinear model **CANNOT** be linearized through transformation.



Example of a Nonlinear Model

$$Y = \beta_1(X - \beta_2)^2 + \beta_3X + \varepsilon$$



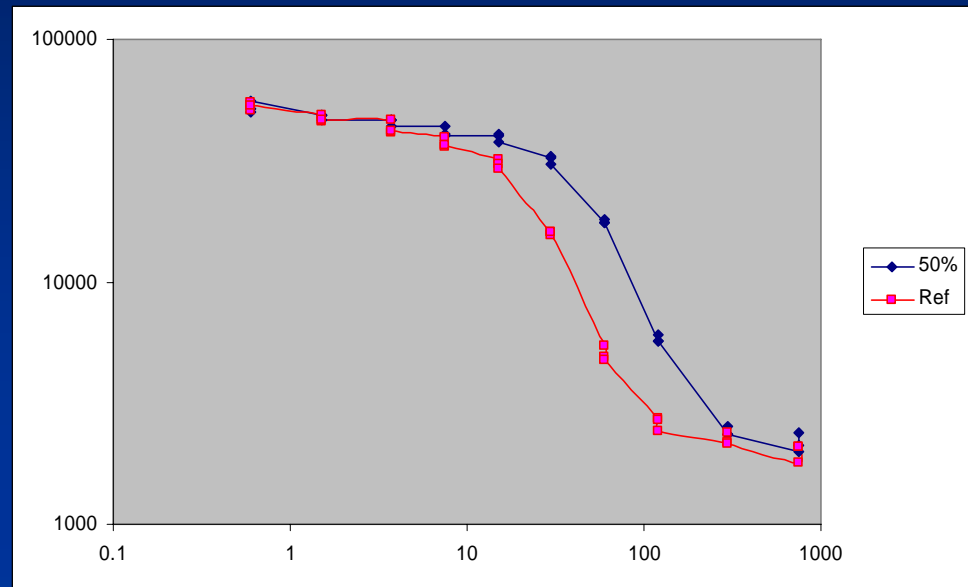
Estimation Techniques for Nonlinear Regression

- Nonlinear models cannot be solved explicitly, therefore iterative methods must be used.
- Convergence of the model might occur if the model is mis-specified or insufficient data.
- Similar to linear regression, model diagnostics can be applied to nonlinear regression.



Logistic Regression

- A bioassay is used to determine the strength or biological activity of a substance, such as a drug or hormone, by comparing its effects with those of a standard preparation on a culture of living cells or a test organism.
- Most bioassays follow a sigmoid dose response curve when plotted on a log-log scale.
- The relative potency is the relationship between the reference curve and test curve.



Parallelism

- Parallelism or dilutional similarity means that the two dilutional curves are proportional throughout the range of the assay.
- In the linear regression case, it would be similar slopes with different intercepts.
- Parallelism is essential to compute relative potency.

4-Parameter Logistic Model

- The model for the 4-parameter fit is:

$$Y = A + \frac{B - A}{\left[1 + \left(\frac{\text{conc}}{EC_{50}} \right)^{\text{Slope}} \right]}$$

- A = Lower Asymptote
- B = Upper Asymptote
- EC_{50} = 50% binding concentration
- Slope = shape parameter of the model
- Conc = X variable



Issues With the Approach

- The model fit is iterative.
- Model does not always converge.
- Reference and test curves might not be parallel.



Combined Protocol

- Test specificity using a minimum of 6 replicates per interfering substance level. Usually use a single analyst for this study.
- Test LOD and LOQ
- 2 analysts, 2 days, 3 samples at 5 different concentrations. These 60 observations can be used for linearity, accuracy, precision and range.



Data Monitoring

- How should method validation and method qualification data be monitored and handled?
- Would like to assess if any outliers exists before the study is complete.
- System suitability testing accomplishes this without the chance of biasing the data.

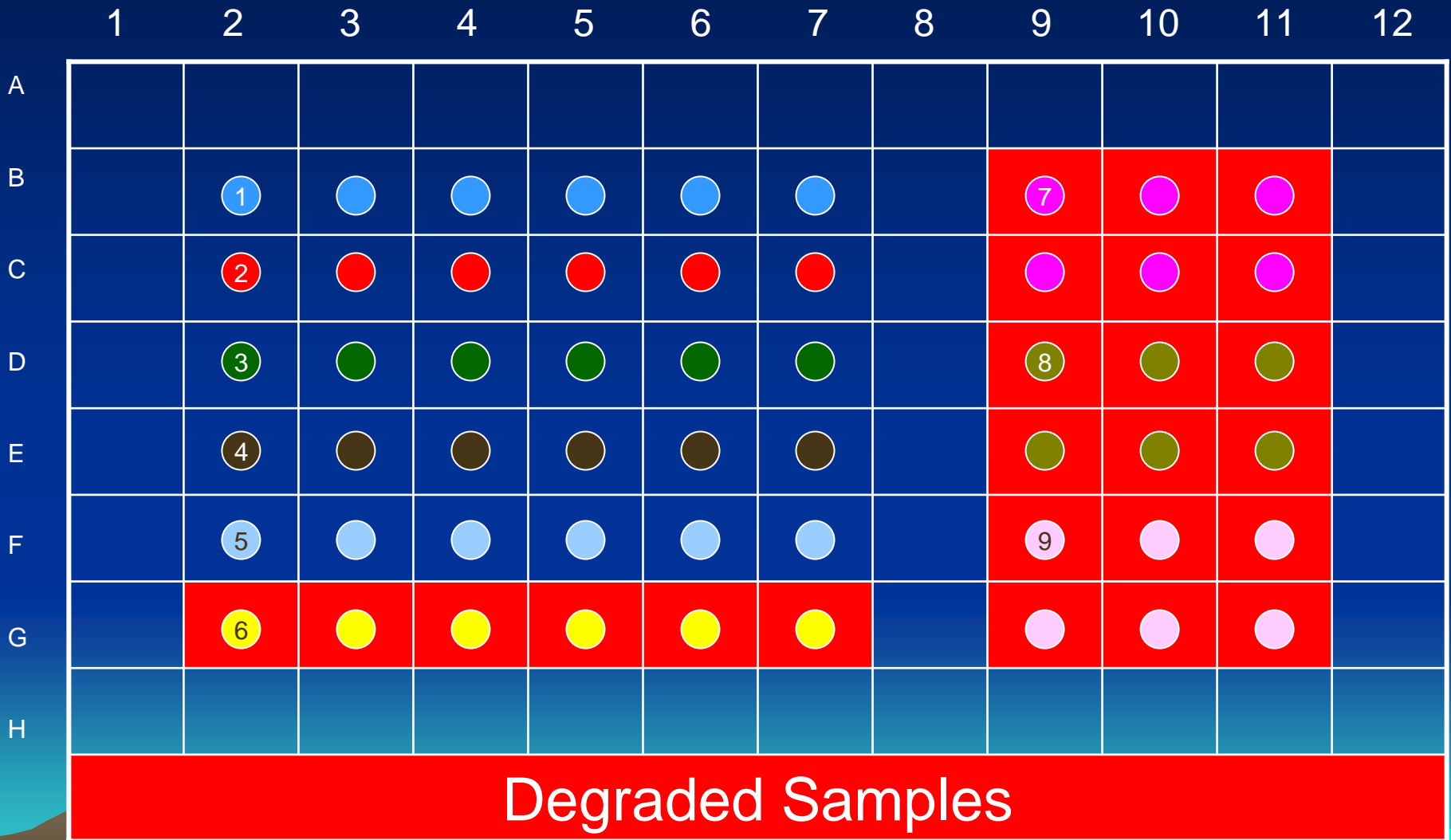


Case Study

- 9 samples each with 6 repeats.
- Assess accuracy and precision for each sample.
- Assess linearity



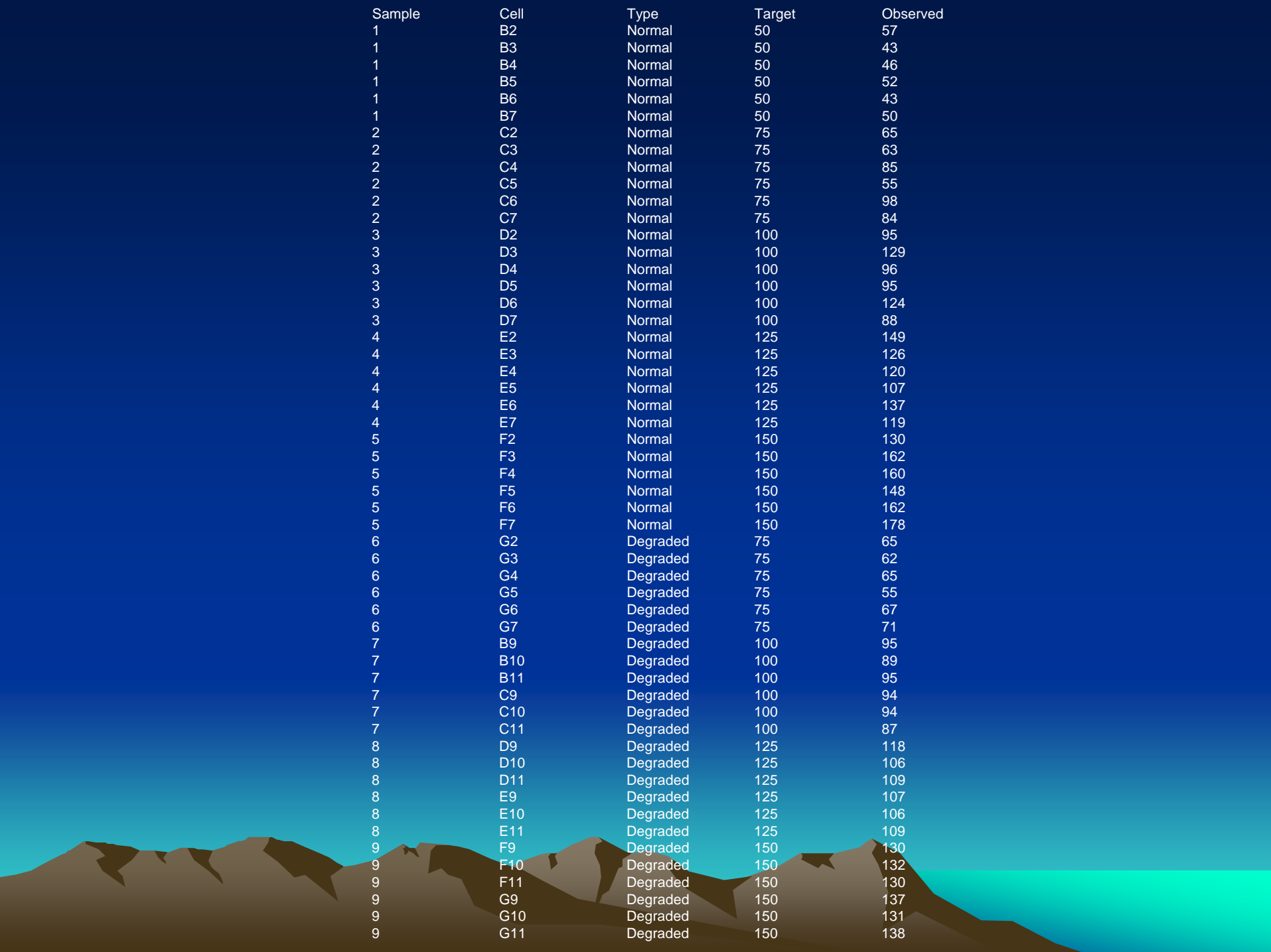
Case Study



Case Study

- Any issues with the experiment?
- What can be improved in the plate layout?
- What additional analysis should be conducted?





Accuracy and Precision

Sample Type	Spiked Concentration	Mean	Std Dev	% Recovery
Degraded	75	64.2	5.4	85.6
Degraded	100	92.3	3.4	92.3
Degraded	125	109.2	4.5	87.3
Degraded	150	133.0	3.6	88.7
Normal	50	48.5	5.5	97.0
Normal	75	75.0	16.5	100.0
Normal	100	104.5	17.4	104.5
Normal	125	126.3	14.8	101.1
Normal	150	156.7	16.2	104.4

Linearity

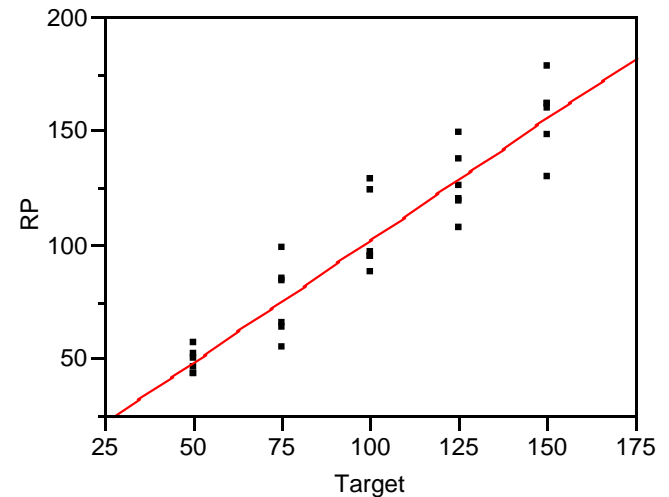
- $RP = -4.866667 + 1.0706667 \text{ Target}$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	-4.866667	7.674773	-0.63	0.5312
Target	1.0706667	0.072358	14.80	<.0001

Summary of Fit

RSquare	0.886613
RSquare Adj	0.882564
Root Mean Square Error	14.01215
Mean of Response	102.2
Observations (or Sum Wgts)	30



Row Effect

- Samples 7, 8 and 9 are tested in separate rows. Can do testing to see if there is a row effect.
- Only 3 observations in each row, so the power is not very good.
- All t-tests show no significant row effects.



Final Thoughts

- Statistical design should be integrated into any method validation.
- Do not have to do one at a time experimentation.
- Precision analysis is the most important component to any study.

