Best Practices in Academic/Clinical Department Administration and Scholarship of Discovery

Joseph Balogun
PT, PhD, FACSM, FNSP, FAS, FRSPH
Distinguish University Professor
College of Health Sciences

CHICAGO STATE UNIVERSITY

Best Practices in Academic/Clinical Department Administration and Scholarship of Discovery

Venue: University of Medical Sciences, Ondo City

Date: July 9 and 10, 2018

CHICAGO STATE UNIVERSITY

Shout out to My Friends in Chi-Town

11:00- 12 noon: Third Presentation

Scholarship of Discovery -

Design and Statistical Analysis

12 noon – 1:00 pm: Discussion and Demonstration

Presentation Overview

Covers the following 3 major topics:

1. Evaluating the Basic Assumptions of Parametric Test
2. Testing the basic assumptions of parametric tests - outliers, normality, homogeneity of variance
3. Statistical vs. practical significance – Effect size

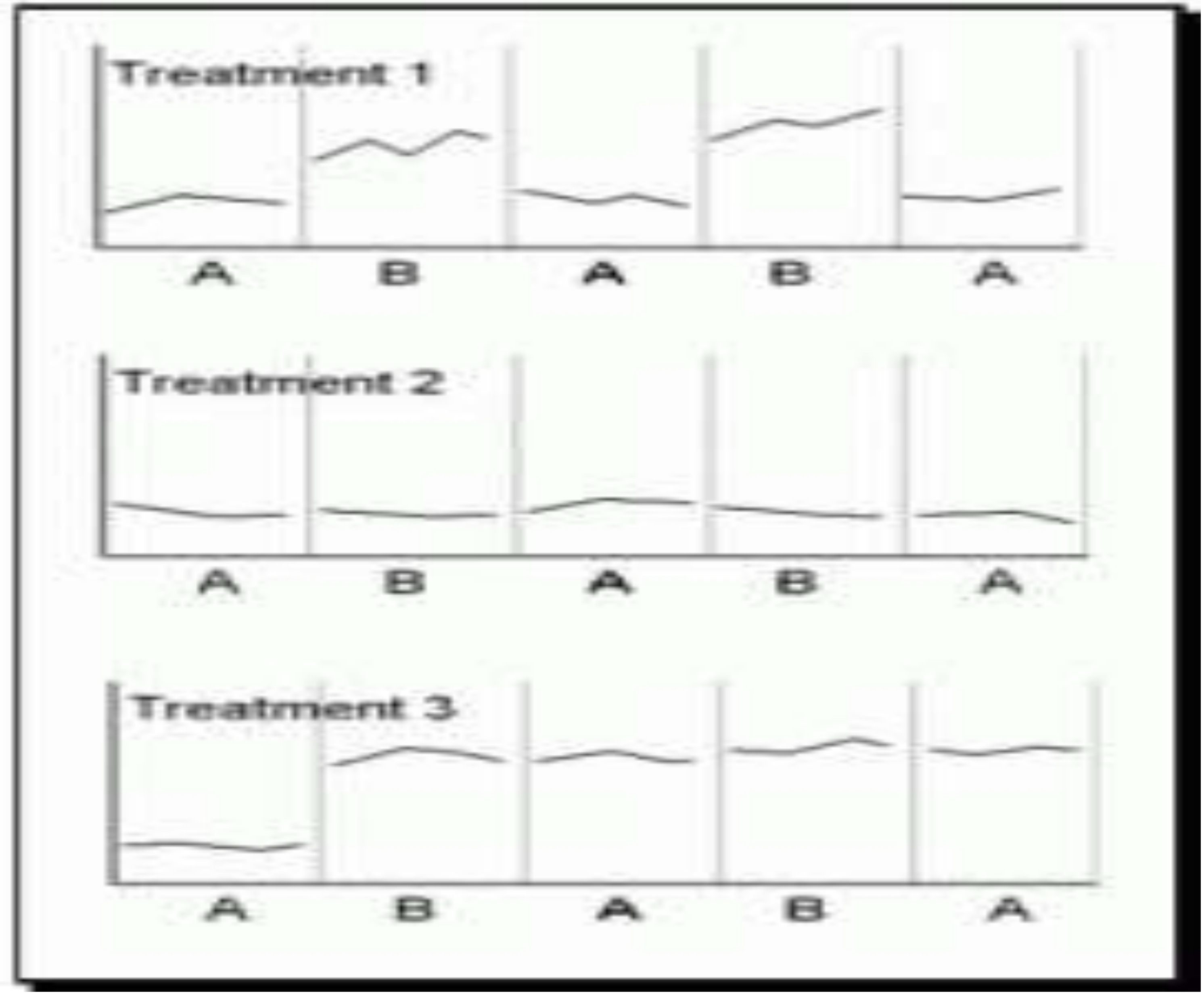
Common Pitfalls in Different Experimental Design Methods*

http://www.slideshare.net/abineshraja/errors-in-research

http://www.qualtrics.com/blog/5-common-errors-in-the-research-process/

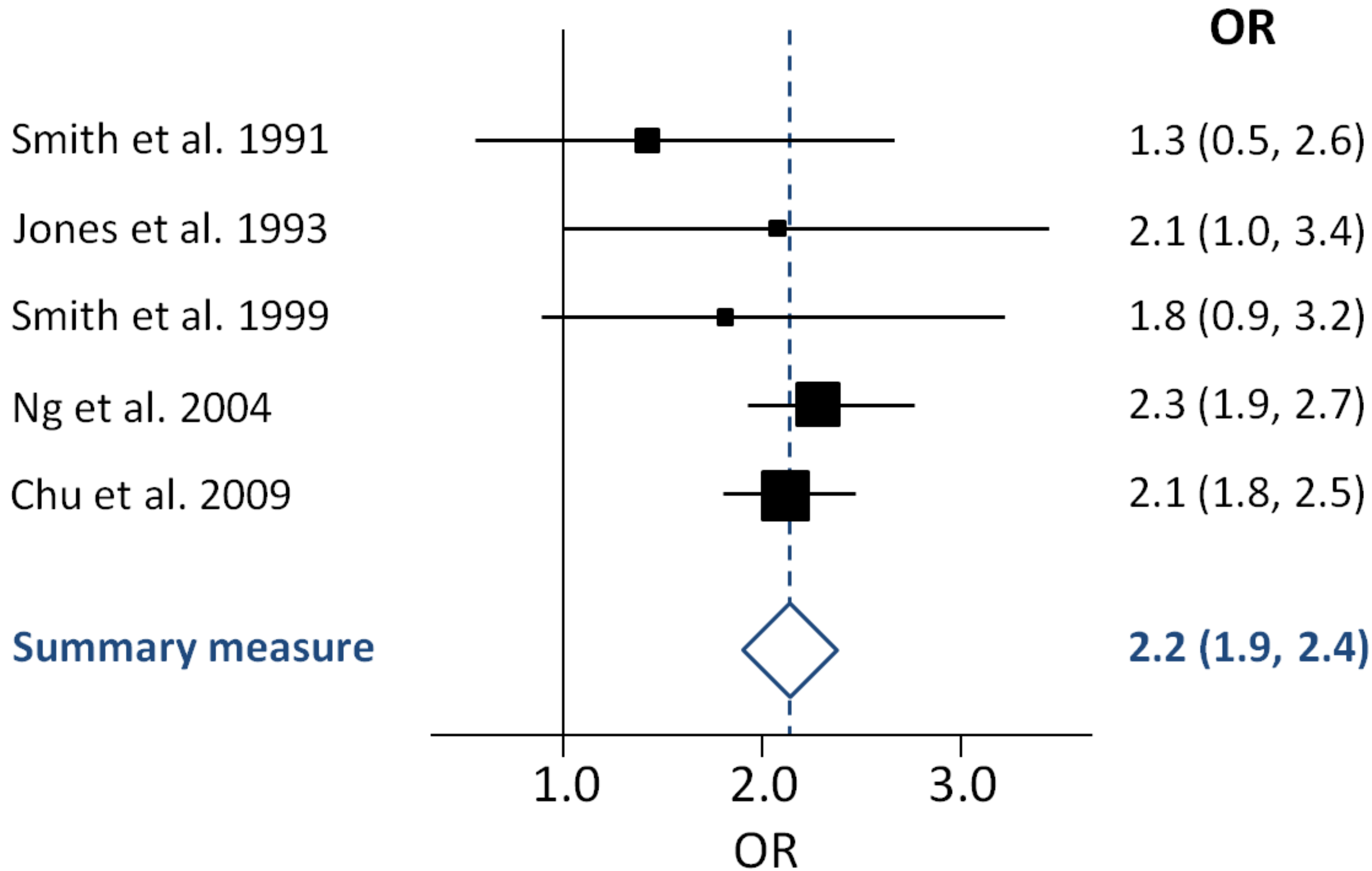
Study Design and Credibility Gap

Fig 3: Single-Subject Research Design (SSRD): Multiple Phases and Conditions--- E.G., Application of Three Different Treatments on Three Single Subjects



Forest Plots (Blobbogram)Used in Systematic Review and Meta-Analysis Studies

https://www.meta-analysis.com/pages/why_use.php?gclid=Cj0KCQjw6pLZBRCxARIsALaaY9Zf84jBreR-KtEgILJ__Mb7lDv5y0R6FXmXfFRXcRN7vSlRAPMmcSsaAphoEALw_wcB



http://adc.bmj.com/content/90/8/845

Evaluating the Basic Assumptions for Parametric Test

Basic Assumptions for Parametric Tests

	Assumptions	Test
1	The observations in the two groups must be independent	
2	Optimum sample size	
3	The scale of measurement must be interval or ratio.	
4	The DV should not contain any outliers.	Box Plot
5	The DV should be normally distributed.	K-S or SW
6	The distribution of the two groups should be homogenous.	Levene Test
	http://www.statisticssolutions.com/manova-analysis-one-sample-t-test	

Guide for Selecting the Appropriate Inferential Statistics

	Assumptions/Criteria	Parametric Test	Non-Parametric Test
1	The observations are	Independent on one another; groups	Dependent on one another; pre/post test or matched
2	Scale of measure'nt of DV	Interval/ratio	Nominal/Ordinal
3	Sample size	Large	Small
4	DV are normally distributed – Draw histogram, Q-Q plot. Test with Fisher's measures of Skewness/ Kurtosis; K-S/S-W	Ho accepted; data is Mesokurtic	Ho rejected; data is skewed or Kurtotic
5	Groups are of equal variance; Levene's Test	Ho accepted homogeneity	Ho rejected Heterogeneity
6	Outliers -Draw Box plot	No outlier	Outlier present

When one or more of the conditions are not met, proceed to use nonparametric stats

Inferential Statistics Equivalence

	Comparison	Parametric Test	Non parametric Test
1	One Sample	One-sample t-test	One-sample Wilcoxon/ signed rank test
2	Two groups	Student t-test	Mann-Whitney U Test
3	Repeated tests (Pre/post test)	Paired t-test	Sign Test; Wilcoxon Signed Ranks Test (T)
4	Three groups or more	One Way ANOVA	Kruskal-Willis Analysis of Variance by ranks (H or x2)
5	Three or more repeated tests	One way repeated measures ANOVA	Friedman two way Analysis of Variance by Ranks (x2)

Correlation Coefficient Statistics Equivalence

	Parametric Statistics	Non-parametric Statistics
1	Pearson Product-Moment Correlation Coefficient ® - For continuous (interval/ratio) variables	Spearman rank (Rho) correlation coefficient -- For discrete (ordinal/rank) variable
		Phi Coefficient For dichotomous variable
2	Intra-class correlation coefficient (ICC) – For two or more repeated ratings on interval/ratio scales	Point Biserial Correlation For dichotomous and Continuous variables

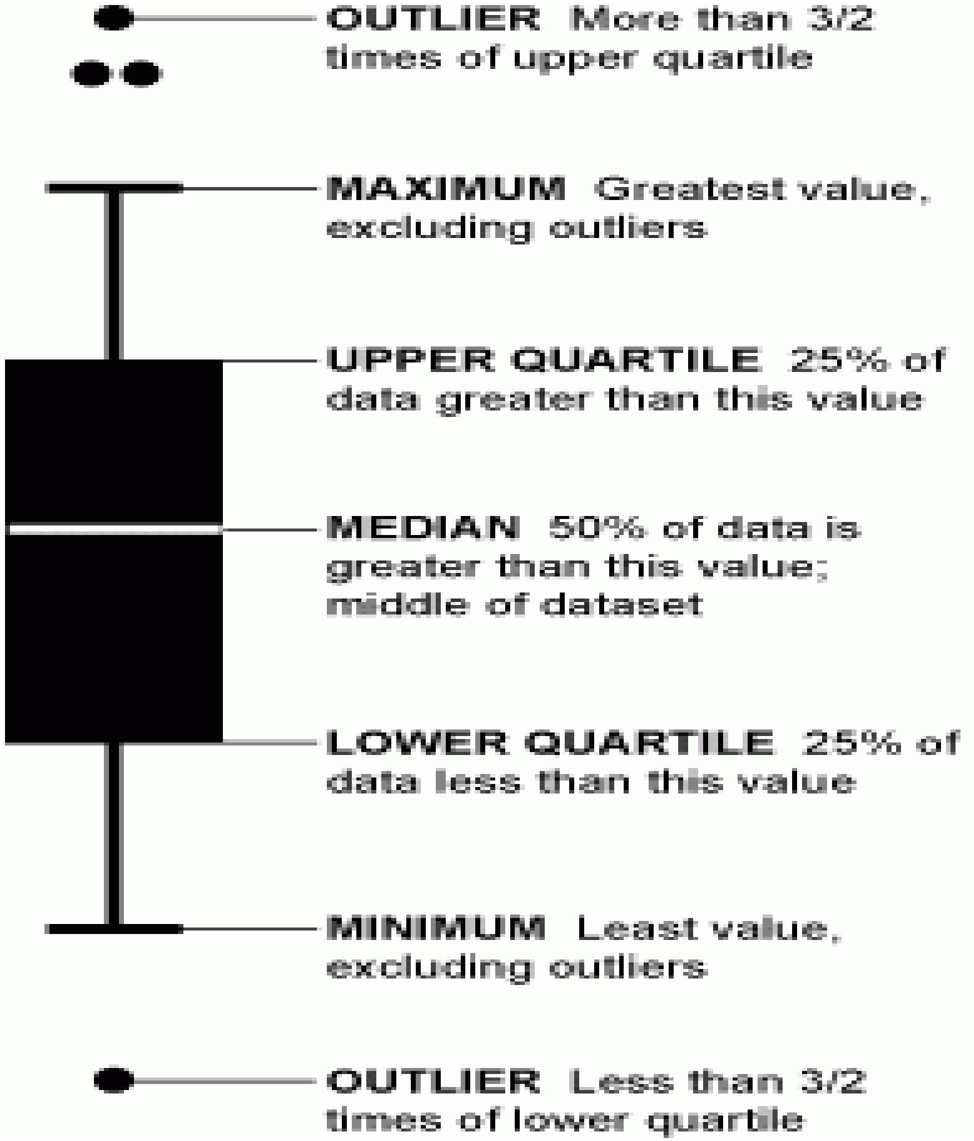
Recognition of Outliers in the Dependent Variable:

Reading a Box-and-Whisker Plot

Information Obtainable from the Box Plot: **Outliers**

- An outlier is a data value which is too extreme to belong in the distribution of interest
- Box-plots (obtained on SPSS through explore) are useful for visualizing the variability in a sample, as well as locating any outliers.
- The minimum, maximum, median, upper and lower quartiles data for the distribution can be obtained from the box plot
- The box plot diagram on the next page shows a sample with 3 outliers on top and one outlier below.

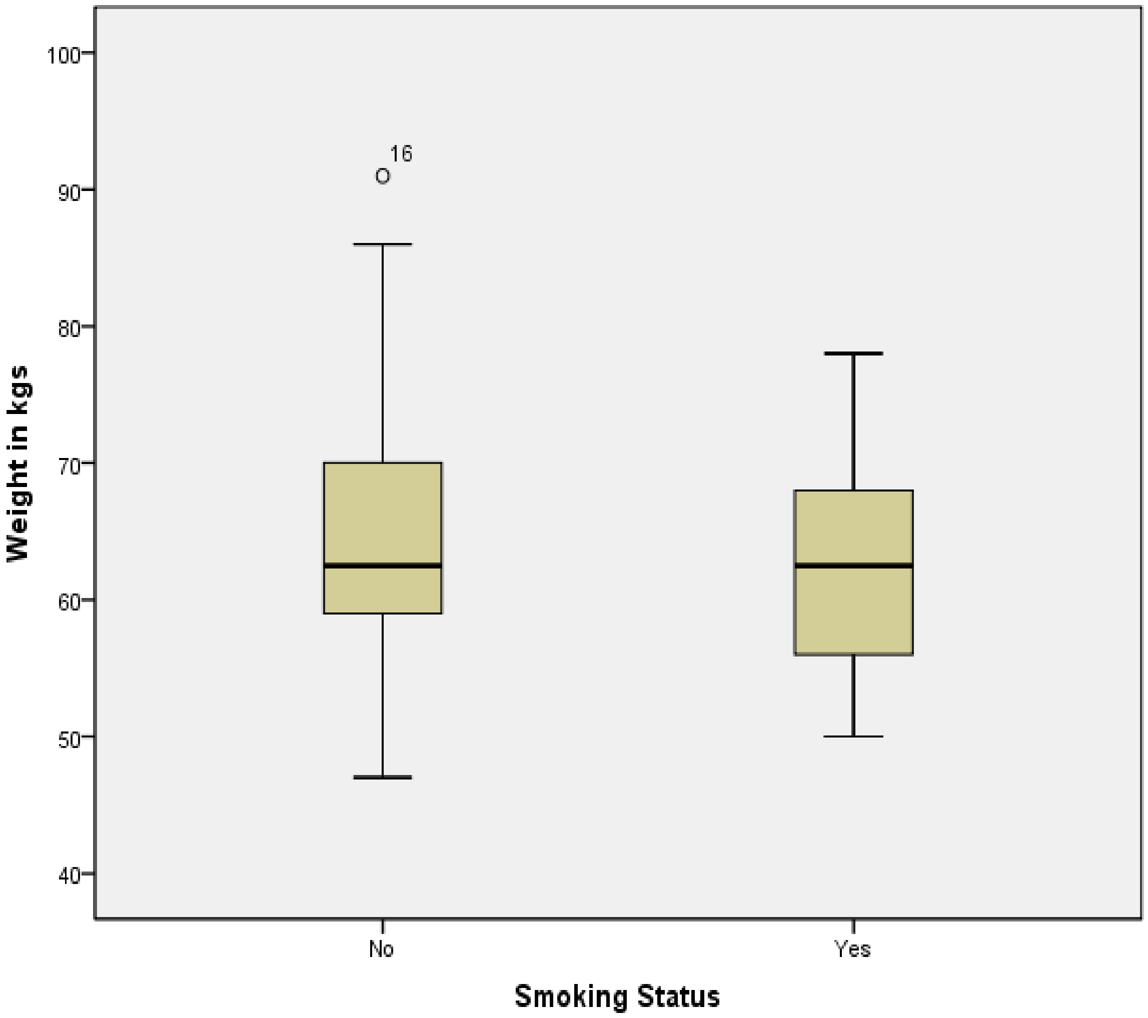
Reading a Box-and-Whisker Plot



Homogeneity of Variance Within Groups

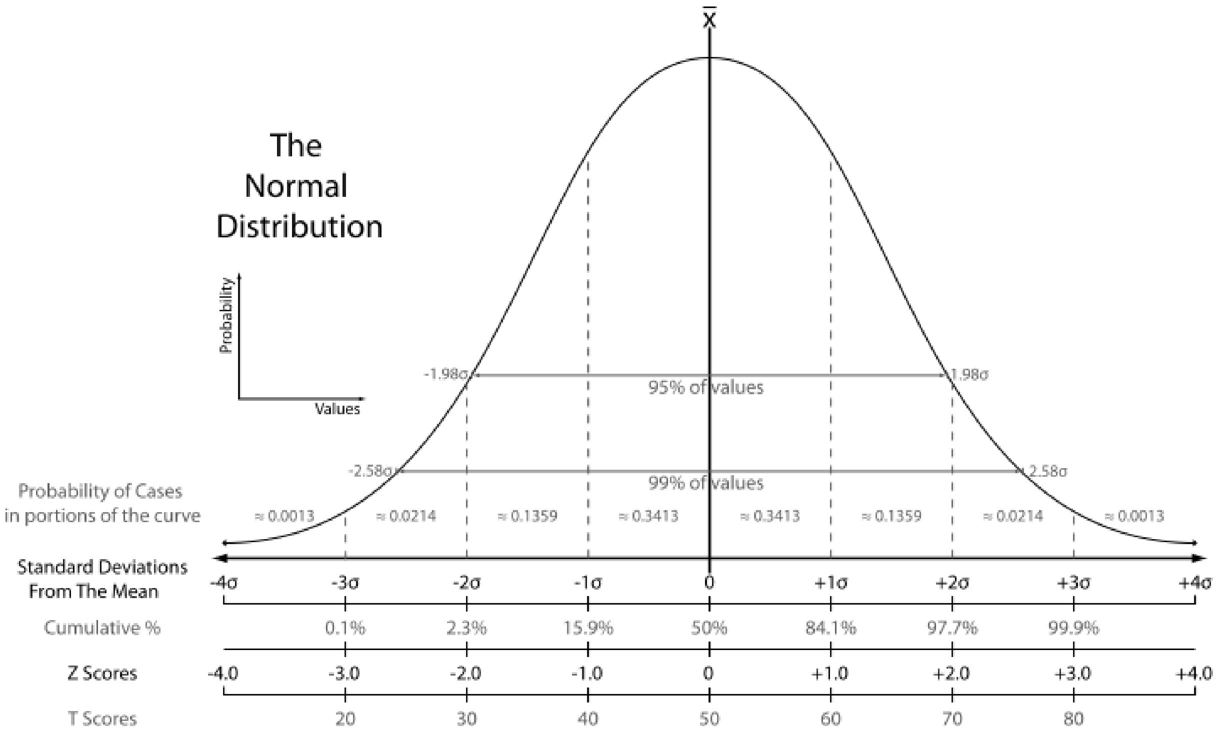
- The Box-plots (obtained on SPSS through explore) are also useful for visualizing the homogeneity of variance among groups
- The box plot on the next page shows a sample with one outlier (#16) and the right shows a sample with no outlier.
- The Box-plot also provides a graphical representation of the homogeneity with the groups. In the data displayed next page, the heights of the charts (minimum, maximum, median, lower and upper quartile) are similar, suggesting that the variances are homogeneous, but this is speculative. You need to use the Levene's test to confirm the speculation

Testing for Normality of the Weight Data with the Box Plot via Explore Analysis on SPSS



Testing for Normality of the Dependent Variable

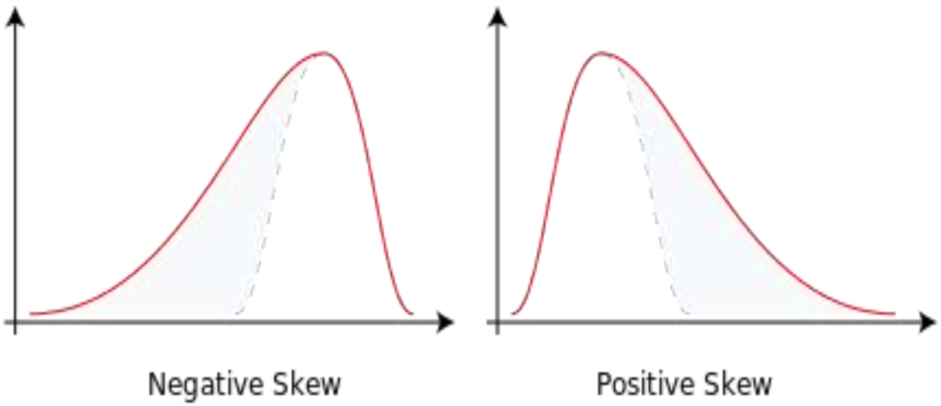
The Normal Distribution



Skewness

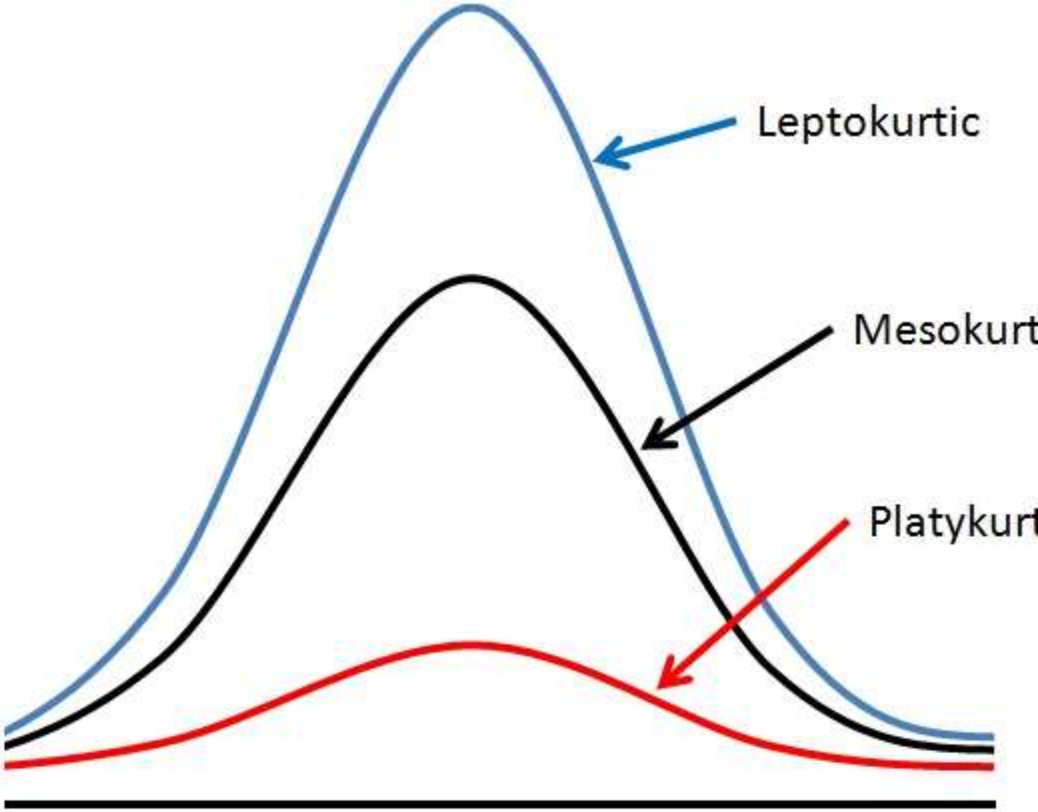
Clustering of scores at one end of a distribution, resulting in an asymmetrical distribution

- *Positive Skewness:* Scores are clustered at the *right end* of the distribution, trailing off into the high end
- *Negative Skewness:* Scores are clustered at the *left* end of the distribution, trailing off into the low end



Kurtosis

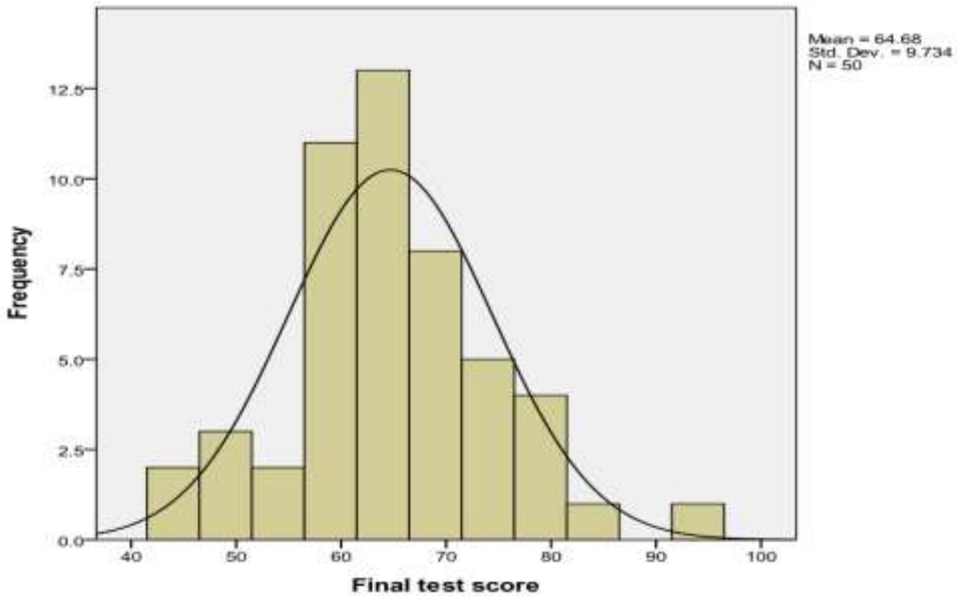
- Non-normal, but still symmetrical distributions
  - *Platykurtic:* Few scores in the middle, but more surrounding it (results in a "flatter" distribution)
  - *Leptokurtic:* Most scores in the middle, but with relatively few surrounding it (results in a "peaked" distribution)



Plot Histogram Curve

First, always plot the histogram with the normal distribution curve to inspect for skewness and kurtosis

Then proceed to test for normality to confirm or refute your speculation.



Testing for Normality of the Distribution of the Dependent Variables with Kolmogorov-Smirnov (K-S) and Shapiro-Wilk Tests

- The most widely used test of Normality by SPSS is the Kolmogorov-Smirnov (K-S) test
- The K-S test is more powerful for large samples
- Shapiro-Wilk Test is more appropriate for small sample sizes (< 50 samples) but can also handle sample sizes as large as 2000.
- The null hypothesis tested is that the distribution is normal; not skewed or kurtotic.
- The exact significant p level given in the SPSS print out will confirm if p is < or p >

When to use Shapiro Wilks and Kolmogorov-Smirnov (KS) Tests

- http://geography.unt.edu/~wolverton/Normality%20Tests%20in%20SPSS.pdf
- For n = 3 to 2000 use Shapiro Wilks test
- For n > 2000 use Kolmogorov-Smirnov test
- A significant KS or Shapiro Wilks tests mean the sample distribution is not shaped like a normal curve
- Null hypothesis (Ho) = normality If you accept, then assume normality
- If you reject Ho, then do not assume normality
- “Statistic” is the test statistic;“Sig” is the significance for the test (aka the p-value)
- If p < then 0.05, reject the H0 because the test is

Interpretation of the KS and Shapiro Wilks Tests

Tests of Normality

Course		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Time	Beginner	.177	10	.200*	.964	10	.827
	Intermediate	.166	10	.200*	.969	10	.882
	Advanced	.151	10	.200*	.965	10	.837

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

The above table presents the results from the KS and the Shapiro-Wilk Tests.

Interpretation of the KS and Shapiro Wilks Tests

- We can see from the above table that for the "Beginner", "Intermediate" and "Advanced" Course Group the dependent variable, "Time", was normally distributed.
- The significant (**Sig.)** value for K-S Test was 0.200 and the Sig. value for Shapiro-Wilk Test was 0.827
- The **Sig.** value of both the KS and the Shapiro-Wilk Test are greater than 0.05, the data is normal. If it is below 0.05, the data significantly deviate from a normal distribution.
- Because the N is less than 50, we will read from the Shapiro-Wilk Test

Manual Testing for Normality of a Distribution from SPSS Print out

- An easier Manual method of determining if a Skewness or Kurtosis of a distribution is significantly different from normal (Mesokurtic) distribution is to run the EXPLORE procedure with SPSS for the variable to obtain the Skewness Coefficient (SKC) or Kurtosis Coefficient (KtC) values and their respective standard error (SE) & proceed:

  Test for Skewness: Fisher's Measures of Skewness

  - Compute SKC/SE
  - Compare the sample statistic value obtained with the population critical parameter of 1.96 at .05 alpha level; 2.58 at .01 alpha level and 3.29 at .001 alpha level

  Test for Kurtosis: Fisher's Measures of Kurtosis

  - Compute KtC/SE
  - Compare the sample statistic value obtained with the population critical parameter of 1.96 at .05 alpha level; 2.58 at .01 alpha level and 3.29 at .001 alpha level

Examples of Manual Testing for Normality from Explore

- A general rule of thumb is to determine whether a distribution is significantly skewed is to divide the Skewness Coefficient (SKC) with Standard Error (SE) for Skewness
- If this results in a number that is between -1.96 and + 1.96, then the distribution is not significantly skewed.

Example #1

- Consider a distribution with SKC of 0.245 and SE of 0.441
- Compute Fisher's Measures of Skewness (SKC/SE) and compare to 1.96; 2.58; 3.29; (.05; .01; .001)
- Fisher's Measures of Skewness (SKC/SE) = 0.245/0.441 = 0.556

  0.556 falls within 1.96
- This means the null hypothesis was accepted; P>.05
- That is, the distribution is not significantly skewed from

Examples of Manual Testing for Normality from Explore

Example #2

- Consider a distribution with Skewness Coefficient (SKC) = 2.498 and SE for Skewness = 0.427
- 2.498/0.427 = 5.85
- Compute Fisher's Measures of Skewness (SKC/SE) and compare the value obtained 1.96; 2.58; 3.29; (.05; .01; .001)
- The computed index, 5.85, fell outside of 3.29 critical region
- This means the null hypothesis is rejected at 0.01 alpha level ($p<0.01$)
- The variable is significantly skewed; the variable

Evaluating the Homogeneity (Equality) of Variance in an Independent T-Test Design with Levene’s Test

Interpretation of the SPSS Printout of Levene's Test for Equality of Variances

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DVWORI	Equal variances assumed	1.493	.257	2.887	8	.020	2.0000	.6928	.4024	3.5976
	Equal variances not assumed			2.887	6.817	.024	2.0000	.6928	.3528	3.6472

Interpretation of the Levene's Test for Equality of Variances

- To find out which row to read from, look at the large column labeled Levene's Test for Equality of Variances.
- This is a test that determines if the two conditions have about the same or different amounts of variability between scores.
- Below it you will see two smaller columns labeled F and Sig.
- Look in the Sig. column. It will have one value. You will use this value to determine which row to read from. In this example, the value in the Sig. column is 0.26 (when rounded).

If the Sig. Value is greater than .05

Independent Samples Test

		Levene's Test for uality of Variance		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DWORI	Equal varian assumed	1.493	.257	2.887	8	.020	2.0000	.6928	.4024	3.5976
	Equal varian not assume			2.887	6.817	.024	2.0000	.6928	.3528	3.6472

Read from the top row. A value greater than .05 means that the variability in your two conditions is about the same. This means that the variability in the two conditions is not significantly different. This is a good thing. In this example, the Sig. value is greater than .05. So, we read from the first row.

If the Sig. Value is less than or equal to .05…

- Read from the bottom row. A value less than .05 means that the variability in your two conditions is not the same. This means that the variability in the two conditions is significantly different. This is a bad thing, but SPSS takes this into account by giving you slightly different results in the second row. If the Sig. value in this example was greater less than .05, we would have read from the second row.
- **So we've got a row**
- Now that we have a row to read from, it is time to look at the results for our T-test. These results will tell us if the Means for the two groups were statistically different or if they were relatively the

Interpretation of the Rx Effect: Sig (2-Tailed) value

- This value will tell you if the two condition Means are statistically different. Make sure to read from the appropriate row. In this example, the Sig (2-Tailed) value is 0.02. Recall that we have determined that it is best to read from the top row.
- Reading from the top row: The Sig. value of 0.020 is less than alpha level of 0.050. This indicates that the two mean scores are significantly different from one another ($t = 2.887$, $df = 8$ $p < .05$)
- Reading from the bottom row: The Sig. value of 0.024 is less than alpha level of 0.050. This indicates that the two mean scores are

Evaluating the Homogeneity (Equality) of Variance in an ANOVA Design with Levene’s Test

https://www.youtube.com/watch?v=XrG1HZo77U4

How to interpret the SPSS Printout from an independent samples T-test?

- The assumption of homogeneity of variance is that the variance within each of the populations is equal.
- This is an assumption of ANOVA.
- ANOVA works well even when this assumption is violated except in the case where there are unequal numbers of subjects in the various groups.
- Levene's test is an inferential statistic used to assess the equality of variances for a variable calculated for two or more groups.
- Some common statistical procedures assume that variances of the populations from which different samples are drawn are equal.

Intervention Studies:

Statistical Significance versus Practical Significance

https://www.google.com/search?q=practical+or+clinical+significance&oq=practical+or+clinical+significance+&aqs=chrome..69i59j69i57.32601j0j9&client=ms-android-hms-tmobile-us&sourceid=chrome-mobile&ie=UTF-8

Statistical Vs. Practical Significance

- Statistical significance refers to the unlikelihood that mean differences observed in the sample have occurred due to sampling error.
- Given a large enough sample, despite seemingly insignificant population differences, one might still find statistical significance.
- Practical significance looks at whether the difference is large enough to be of value in a practical sense.

- https://www.google.com/search?client=ms-android-hms-tmobile-us&q=What+is+the+difference+between+practical+and+statistical+significance%3F&sa=X&ved=0ahUKEwj_vtfs5pXXAhVs7IMKHQ1jDagQzmcIPg&biw=360&bih=517&dpr=2#xxri=2

Statistical Vs. Practical Significance

- While most published statistical work include information on significance, such measures can cause problems for practical interpretation.
- For example, a significance test does not tell the size of a difference between two measures (practical significance), nor can it easily be compared across studies.
- To account for this, it is generally recommended by most journals that all published statistical reports also include effect size

What is Effect Size?

- Effect size is the magnitude, or size, of an effect.
- Statistical significance (*e.g.*, $p < .05$) tells us there was a difference between two groups or more based on some treatment or sorting variable.
- For example, using a *t*-test, we could evaluate whether drug A is more effective in treatment of hypertension.
- For six weeks, patients with hypertension were randomly assigned into two groups, drugs A and B. Systolic and diastolic BP were taken at rest following 6 weeks regime of treatment with the two medication. The average BP were obtained. SBP for patients in Group A =132mmHg and Group B = 138 mmHg

What is Effect Size?

- If we were testing for **significance difference,** we would calculate standard deviation and evaluate the results using a *t*-test. The results give us a value for *p*, telling us (if $p <.05$, for example) drug A is superior to drug B in lowering BP.
- What this fails to tell us is the magnitude of the difference.
- In other words, *how much more effective* was drug A from drug B?
- To answer this question, we standardize the difference and compare it to 0.

Calculation of Effect Size

How to Calculate Cohen's Effect size

SPSS currently do not run test for Cohen's effect size.

- Cohen's effect size can be computed using the following methods
- Direct Manual Calculation- Laborious
- Indirect Manual Calculation using appropriate data from the T (Mean, SD), or ANOVA (SSB, SSW, SST) tests printout from SPSS
- Derivation from Online Calculators (Recommended)

Indirect Manual Calculation of Effect Size Using Appropriate data (Mean, SD) From the SPSS T -Test Printout

Calculating Effect Size (Cohen's *d*) for Two Groups

- Given mean (*m*) and standard deviation (*sd*), you can calculate effect size (*d*). The formula is:

Cohen's *d* = <u>***m1*** (group or Rx 1) **–** ***m2*** (group or Rx2)</u>
Pooled standard deviation (***SD*)**

Pooled SD = $\sqrt{[(sd_1^2 + sd_2^2) / 2]}$

The basic format for group comparison is to provide: population (*N*), mean (*M*) and standard deviation (*SD*) for both samples, the statistical value (*t* or *F*), degrees freedom (*df*), significance (*p*), and confidence interval (CI.95).

Calculation of the Effect Size (Cohen's *d, r*) & SD

- Effect size is a standard measure that can be calculated from any number of statistical outputs.
- One type of effect size, the standardized mean effect, expresses the mean difference between two groups in standard deviation (SD) units.
- Typically, you'll see this reported as Cohen's *d*, or simply referred to as "*d*."
- Though the values calculated for effect size are generally low, they share the same range as standard deviation (-3.0 to 3.0), so can be quite large.
- Interpretation depends on the research question.

Example

- The average BP were obtained. SBP for patients in Group A =132mmHg and Group B = 138 mmHg. Assuming there was a statistically significant difference between the two drugs, (*M* = 132, **sd_1** = 16.11) and team 2 (*M* = 138, **sd_2** = 14.09), $t(98) = 3.09$, $p \leq .05$, *CI.95* -15.37, -3.35.
- Pooled ***SD*** = $\sqrt{[(sd_1^2 + sd_2^2) / 2]}$ **= 15.13**
- **Cohen's *d* =** = 6/**15.13** = 0.40; Therefore, we accept the null hypothesis that there is no difference in BP between the two medication.
- Further, Cohen's effect size value *d* = 0.40; suggested a small to moderate practical significance

Interpretation Criteria for Effect size for T-Test

Cohen (1988) considers the following intervals for his d-values:

- 0.1 - 0.3 Small practical effect
- 0.3 - 0.5: Intermediate practical effect
- 0.5 > Large/Strong practical effect

Indirect Manual Calculation of Effect Size Using Appropriate data (SSB, SSW) From the SPSS ANOVA Printout

Calculation of the Effect size for a between groups ANOVA (Due to Treatment/Intervention)

Just remember to consider the design of the study – is it between groups or within subjects?

The formula for calculating η^2

$$\eta^2 = \frac{\text{Treatment Sum of Squares}}{\text{Total Sum of Squares}}$$

ANOVA

RECALL

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	31.444	2	15.722	7.447	.006
Within Groups	31.667	15	2.111		
Total	63.111	17			

Interpretation of the Effect size for a between groups ANOVA (Due to Treatment/Intervention)

- The treatment sum of squares is the first row: Between Groups (31.444)
- The total sum of squares is the final row: Total (63.111)

$$\eta^2 = \frac{31.444}{63.111}$$

$$\eta^2 = 0.498$$

This would be deemed by Cohen's guidelines as a very large effect size; 49.8% of the variance was caused by the IV (treatment).

Calculating the Effect size for a within subjects ANOVA

Just remember to consider the design of the study – is it between groups or within subjects?

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
SPEED1	Sphericity Assumed	31.444	2	15.722	7.183	.012
	Greenhouse-Geisser	31.444	1.732	18.158	7.183	.017
	Huynh-Feldt	31.444	2.000	15.722	7.183	.012
	Lower-bound	31.444	1.000	31.444	7.183	.044
Error(SPEED1)	Sphericity Assumed	21.889	10	2.189		
	Greenhouse-Geisser	21.889	8.658	2.528		
	Huynh-Feldt	21.889	10.000	2.189		
	Lower-bound	21.889	5.000	4.378		

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	600.889	1	600.889	307.273	.000	.984	307.273	1.000
Error	9.778	5	1.956					

a. Computed using alpha = .05

Calculating the Effect size for a within subjects ANOVA

The formula is slightly more complicated because you have to calculate the total Sum of Squares:

- Total Sum of Squares = Treatment Sum of Squares + Error Sum of Squares + Error (between subjects) Sum of Squares.

$$\eta^2 = \frac{\text{Treatment Sum of Squares}}{\text{Total Sum of Squares for within subjects ANOVA}}$$

Calculating the Effect size for a within subjects ANOVA

Total Sum of Squares: 31.444 (top table, SPEED 1) + 21.889 (top table, Error(SPEED1)) + 9.778 (Bottom table, Error) = 63.111

- The value is the same as the last example with between groups – so it works!
- Just enter the total in the formula as before:
- $\eta^2 = \dfrac{31.444}{63.111} = 0.498$
- Again, 49.8% of the variance in the DV is due to the IV.

Interpretation Criteria for Effect size for ANOVA (η^2)

- Cohen's (1988) guidelines for interpreting effect size:
- Small: 0.01
- Medium: 0.059
- Large: 0.138
- So if you end up with $\eta^2 = 0.45$, you can assume the effect size is very large. It also means that 45% of the change in the DV can be accounted for by the IV.

How to Calculate Effect size for a between groups ANOVA (η^2)

- Calculating effect size for between groups designs is much easier than for within groups.

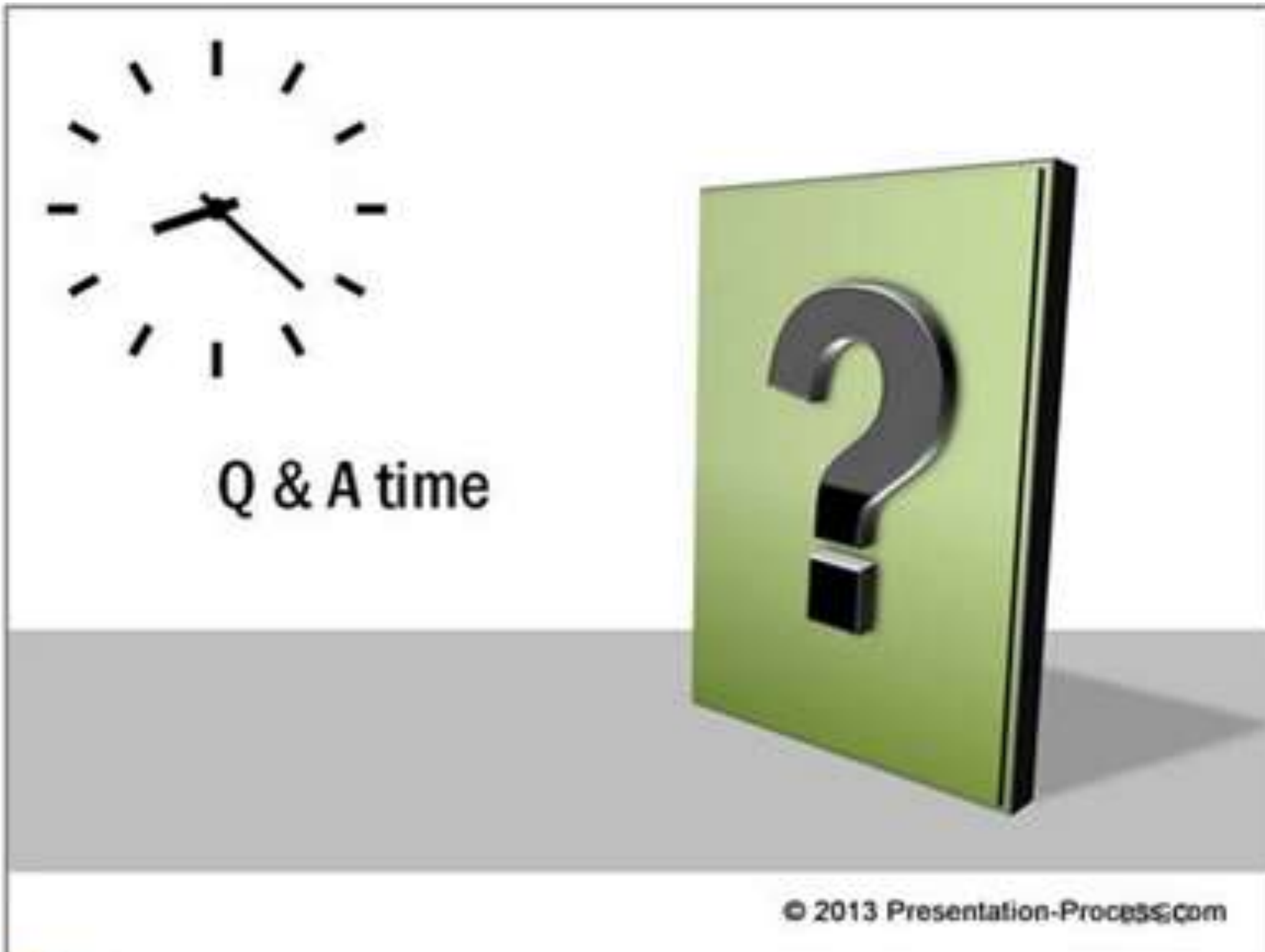
12 noon – 1:00 pm: Discussion and Demonstration

Feel free to contact me by email at jbalogun@csu.edu

Workshop Learning Objectives

At the end of the training, the learner will be able to:

- Create an academic culture of using evidence to make administrative decisions within their department.
- Design and implement a comprehensive assessment program for an academic department.
- Articulate evidence-based teaching strategies and recipe for high quality education.
- Construct measurable course objectives, and student learning outcomes for an academic program.
- Discuss different types of research approaches, experimental designs and quantitative data analysis, testing for the assumptions of parametric and non-parametric statistics.
- Discern areas of weakness in published manuscripts.
- Identify inappropriate use of statistics.
- Determine the clinical significance of an intervention study.



Feel free to contact me by email at jbalogun@csu.edu