LAB MATERIALS (SUPPORT) ON SELECTING APPROPRIATE TESTS, CHARTS, INTERPRETING EFFECT SIZES, SKEWNESS, AND SPSS PROCEDURES FOR DATA ANALYSIS

Table 3.3. ***Selection of Appropriate Descriptive Statistics and Plots***

Visual representation	***Nominal***	***Dichotomous***	***Ordinal***	***Normal***
Frequency Distribution	Yes	Yes	Yes	OK
Bar Chart	Yes	Yes	Yes	OK
Pie Chart	Yes	Yes	Yes	OK
Histogram	No	No	OK	Yes
Frequency Polygon	No	No	OK	Yes
Box and Whiskers Plot	No	No	Yes	Yes
Central Tendency				
Mean	No	OK	Of ranks, OK	Yes
Median	No	OK = Mode	Yes	OK
Mode	Yes	Yes	OK	OK
Variability				
Range	No	Always 1	Yes	OK
Standard Deviation	No	No	Of ranks, OK	Yes
Interquartile Range	No	No	Yes	OK
How Many Categories	Yes	Always 2	OK	Not if truly continuous
Shape				
Skewness	No	No	Yes	Yes

Figure. 5.1. A Decision Tree to Help you Select the Appropriate Inferential Statistic from Tables 5.1-5.4 (Morgan et al., 2020, p.90-92)

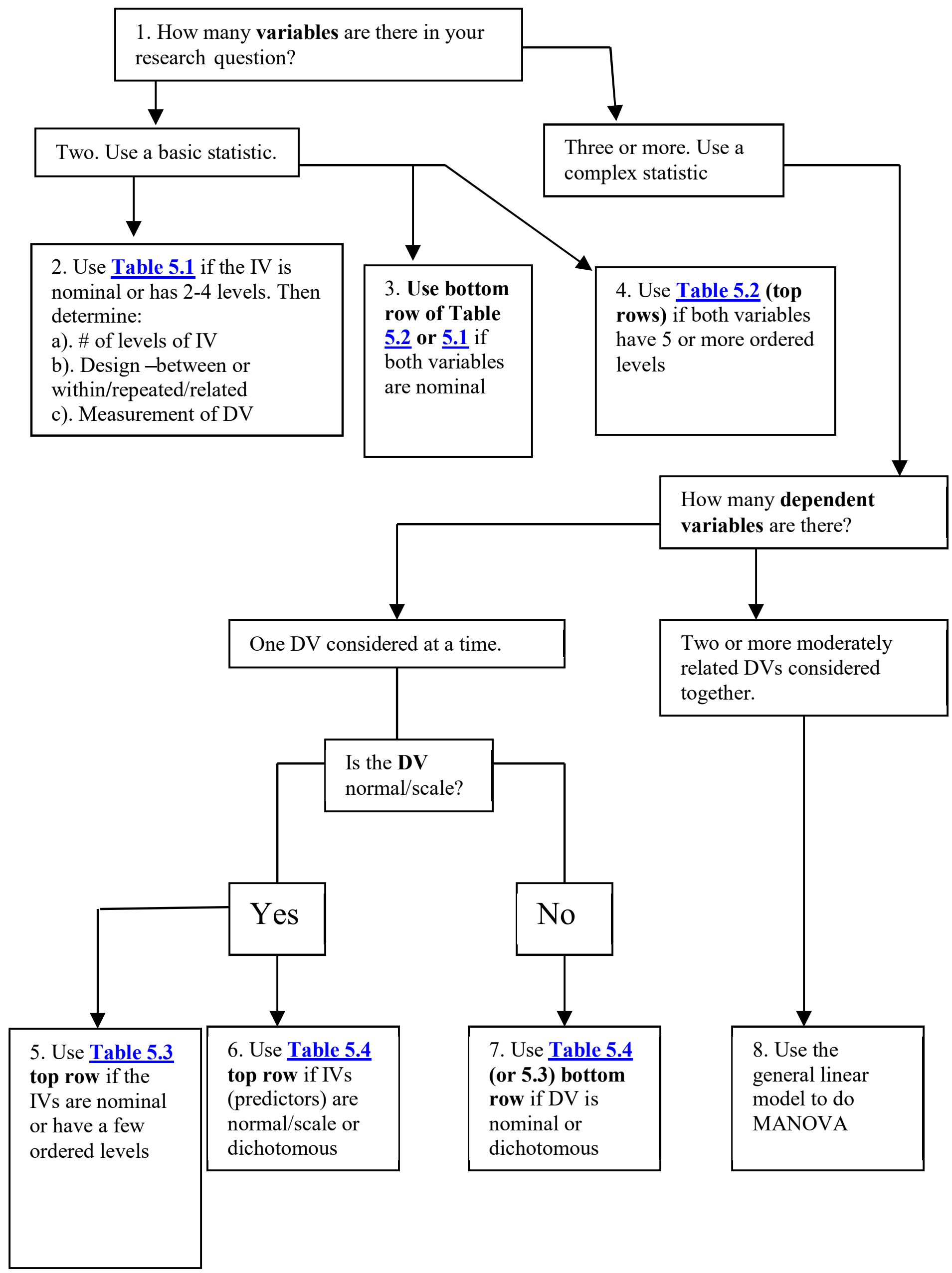


Table 5.1. Selection of an Appropriate Inferential Statistic for Bivariate (Two Variable), Difference Questions or Hypotheses[a] (Morgan et al., 2020, p.90).

Level of Measurement of the **Dependent Variable**	Appropriate Descriptive Statistics to report	One Factor or Independent Variable with **2 levels** or **Categories**/ Groups/ Samples		One Independent Variable **3 or more Levels** or Groups	
		Independent Samples or Groups **(Between)**	Repeated Measures or Related Samples **(Within)**	Independent Samples or Groups **(Between)**	Repeated Measures or Related Samples **(Within)**
Dependent Variable Approximates **Normal/Scale** Data and Assumptions are not Markedly Violated	MEANS, STANDARD DEVIATION	INDEPENDENT SAMPLES ***t*** Test Ch. 9	PAIRED SAMPLES ***t*** Test Ch. 9	ONE-WAY ANOVA Ch. 9	GLM REPEATED MEASURES ANOVA Ch. 9
Dependent Variables Clearly **Ordinal** or Parametric Assumptions Markedly Violated	MEDIAN (or MEAN RANKS), INTERQUARTILE RANGE	MANN-WHITNEY U Ch. 9	WILCOXON Ch. 9	KRUSKAL-WALLIS Ch. 9	FRIEDMAN
Dependent Variable **Nominal or Dichotomous**	MODE, MINIMUM & MAXIMUM	CHI-SQUARE Ch. 7	MCNEMAR	CHI-SQUARE Ch. 7	COCHRAN Q TEST

[a] It is acceptable to use statistics that are in the box(es) below the appropriate statistic, but there is usually some loss of power. It is not acceptable to use statistics in boxes above the appropriate statistic or ones in another column.

Table 5.2. Selection of an Appropriate Inferential Statistic for Bivariate (Two Variable), Associational Questions or Hypotheses (Morgan et al., 2020, p.91)

Level (Scale) of Measurement of **Both Variables**	RELATE	Two Variables or Scores for the Same or Related Subjects
Variables are both **Normal/Scale** and Assumptions not Markedly Violated	SCORES	PEARSON (r) or BIVARIATE REGRESSION (Ch. 8)
Both Variables at Least **Ordinal** Data or Assumptions Markedly Violated	RANKS	KENDALL'S TAU-B or SPEARMAN (Rho) Ch. 8
One Variable is **Normal/ Scale** and One is **Nominal**		ETA Ch. 7
Both Variables are **Nominal** or **Dichotomous**	COUNTS OF PEOPLE	PHI or CRAMER'S V Ch. 7

Table 5.3. Selection of the Appropriate Complex (Two or More Independent Variables) Statistic to Answer Difference Questions or Hypotheses (Morgan et al., 2020, p.92)

Dependent Variable(s) ⬇	**Two or More Independent Variables**		
	All Between Groups	**All Within Subjects**	**Mixed (Between and Within)**
One **Normal/ Scale** Dependent Variable	GLM, Factorial ANOVA or ANCOVA Ch. 9	GLM (ANOVA) With Repeated Measures on All Factors	GLM (ANOVA) with Repeated Measures on Some Factors
Ordinal Dependent Variable	Nonparametric Tests	Nonparametric Tests	Mixed Models (Generalized Linear)
Dichotomous Dependent Variable	LOG-LINEAR	Mixed Models	Mixed Models

Table 5.4. Selection of the Appropriate Complex Associational Statistic for Predicting a Single Dependent/ Outcome Variable from Several Independent Variables (Morgan et al., 2020, p.92)

One Dependent or Outcome Variable	**Several Independent or Predictor Variables**		
	Normal or Scale	**Some Normal Some Dichotomous (Two category)**	**All Dichotomous**
Normal/ Scale (Continuous)	MULTIPLE REGRESSION Ch. 8	MULTIPLE REGRESSION Ch. 8	MULTIPLE REGRESSION Ch. 8
Dichotomous	DISCRIMINANT ANALYSIS	LOGISTIC REGRESSION	LOGISTIC REGRESSION

Multivariate analysis of variance, **MANOVA**, is used if you have two or more normal (scale) dependent variables treated simultaneously.

Table 5.5. ***Interpretation of the Strength of a Relationship*** **(*Effect Sizes*) – (Morgan et al., 2020, p.98).**

General Interpretation of the Strength of a Relationship	***The d Family***[a]	***The r Family***[b]					
	d	r^2	***r, rho, Φ***	R^2	***R***	η^2	**η (eta)**[d]
Much larger than typical	≥|1.00|[c, e]	.49	≥|.70|	.49+	|.70|+	.21	|.45|+
Large or larger than typical	|.80|	.25	|.50|	.26	|.51|	.14	|.37|
Medium or typical	|.50|	.09	|.30|	.13	|.36|	.06	|.24|
Small or smaller than typical	|.20|	.01	|.10|	.02	|.14|	.01	|.10|

[a] *d* values can vary from 0.0 to + or – infinity, but *d* greater than 1.0 is relatively uncommon.

[b] *r* family values can vary from 0.0 to + or -1.0, but except for reliability (i.e., same concept measured twice), *r* is rarely above .70. In fact, some of these statistics (e.g., phi) have a restricted range in certain cases; that is, the maximum phi (Φ) is less than 1.0.

[c] We interpret the numbers in this table as a range of values. For example, a *d* greater than .90 (or less than -.90) would be described as "much larger than typical," a *d* between, say, .70 and .90 would be called "larger than typical," and *d* between, say, .60 and .70 would be "typical to larger than typical." We interpret the other three columns similarly.

[d] Partial etas are multivariate tests equivalent to *R*. Use *R* column.

[e] Note: | | indicates absolute value of the coefficient. The absolute magnitude of the coefficient, rather than its sign, is the information that is relevant to effect size. *R* and η usually are calculated by taking the square root of a squared value, so that the sign usually is positive.

Effect Size formula for Cohen's d using different statistical methods:

$$\text{Cohen's } d = \frac{M_1 - M_2}{SD_{pooled}}$$

$$\text{Glass's } \Delta = \frac{M_1 - M_2}{SD_{control}}$$

$$\text{Hedges' } g = \frac{M_1 - M_2}{SD^*_{pooled}}$$

$$\eta^2 = \frac{SS_{Effect}}{SS_{Total}}$$

$$\eta_P^2 = \frac{SS_{Effect}}{SS_{Effect} + SS_{Error}}$$

$$\omega^2 = \frac{SS_{Effect} - (df_{Effect})(MS_{Error})}{SS_{Total} + MS_{Error}}$$

Eta Squared, Partial Eta Squared, and Omega Squared Formulas

Cohen's $f2$ method of effect size: Cohen's $f2$ method measures the effect size when we use methods like ANOVA, multiple regression, etc. The Cohen's $f2$ measure effect size for multiple regressions is defined as the following:

$$f^2 = \frac{R^2}{1 - R^2}$$

Where R^2 is the squared multiple correlation.

$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{N_1+N_2-2}\right)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$ Social Science Statistics $t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{(N_1-1)s_1^2 + (N_2-1)s_2^2}{N_1+N_2-2}\right)\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$

$$r = \frac{N \sum xy - \sum (x)(y)}{\sqrt{N \sum x^2 - \sum (x^2)][N \sum y^2 - \sum (y^2)]}}$$

Cohen's d follows a classification system based on their effect sizes (Cohen, 1992) i.e.
Cohen's d = .10 = weak effect
Cohen's d = .30 = moderate effect
Cohen's d = .50 = strong effect

Cohen's d formula

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s_{pooled}}$$

where

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

https://www.psychometrica.de/effect_size.html

Effect Size for Chi-square Test

We review three different measures of effect size: Phi φ, Cramer's V and the Odds Ratio.

Phi φ

For the goodness of fit in 2 × 2 contingency tables, phi, which is equivalent to the correlation coefficient r (see Correlation), is a measure of effect size. Phi is defined by

$$\varphi = \sqrt{\frac{\chi^2}{n}}$$

where n = the number of observations. A value of .1 is considered a small effect, .3 a medium effect and .5 a large effect.

This is the effect size measure (labelled as w) that is used in power calculations even for contingency tables that are not 2 × 2 (see Power of Chi-square Tests).

Cramer's V

Cramer's V is an extension of the above approach and is calculated as

$$V = \sqrt{\frac{\chi^2}{n \cdot df^*}}$$

where $df^* = \min(r - 1, c - 1)$ and r = number of rows and c = number of columns in the contingency table. Per Cohen, you use the guidelines for phi divided by the square root of df^*.

SPSS CODEBOOK

This so-called codebook provides you with SPSS steps or procedures to assist you in getting your output from data analysis.

Frequency Distribution: Note that frequency table are only prepared for nominal & ordinal variables.

Analyze → Descriptive Statistics → Frequencies → move variable(s) you want to make frequency distributions for into the variable(s) box → Ensure Display frequency tables' box is checked → Click OK

Bar Chart: Note that bar charts are only prepared for SPSS nominal & ordinal variables.

Analyze → Descriptive Statistics → Frequencies → move variables(s) you want to make bar charts for into the variable(s) box → Click Charts → Select Bar Charts → Click continue → Ensure Display frequency tables box is NOT checked (unless you want to make a frequency distribution as well) → Click OK

Pie Chart: Note that pie charts are only prepared for SPSS nominal & ordinal variables.

Analyze → Descriptive Statistics → Frequencies → move variables(s) you want to make pie charts for into the variable(s) box → Click Charts → Select Pie Charts → Click continue → Ensure Display frequency tables box is NOT checked (unless you want to make a frequency distribution as well) → Click OK

(Note that you may double click on the output to make amendment to any of the charts)

Histograms with Normal Curve Superimposed: Note that histograms are only prepared for SPSS scale (normal/approximately normal/continuous) variables.

Analyze → Descriptive Statistics → Frequencies → move variable(s) you want to make a histogram for into the variable(s) box → Click Charts → Select Histograms → Ensure Show normal curve box is selected → Click continue → Ensure Display frequency tables box is NOT checked (unless you want to make a frequency distribution as well) → Click OK

Measures of Central Tendency and Measures of Variability

Analyze → Descriptive Statistics → Frequencies → move variable(s) you want to generate these statistics for into the variable(s) box → Click Statistics → Select all the statistics you desire from the various options shown → Click continue → Ensure Display frequency tables box is NOT checked (unless you want to make frequency distributions as well) → Click OK

<u>Phi and Cramer's V</u>: Used for dichotomous/nominal variables only

Analyze → Descriptive Stats → Crosstabs → move 1 of the variables to the Row(s) box and 1 of the variables to the Column(s) box (doesn't matter which of the variables you put in each of these boxes) → Statistics → Phi and Cramer's V → Continue → OK

<u>ETA</u>

Analyze → Descriptive Stats → Crosstabs → move 1 of the variables to the Row(s) box and 1 of the variables to the Column(s) box (doesn't matter which of the variables you put in each of these boxes) → Statistics → ETA → Continue → OK

<u>Kendall's Tau</u>

Analyze → Correlate → Bivariate Correlations → move the 2 variables of interest to the variables box → Click Kendall's tau-b → Click 2-tailed → OK

<u>Spearman's Rho</u>

Analyze → Correlate → Bivariate Correlations → move the 2 variables of interest to the variables box → Spearman → Click 2-tailed → OK

<u>Pearson r</u>

Analyze → Correlate → Bivariate Correlations → move the 2 variables of interest to the variables box → Click Pearson → Click 2-tailed → OK

<u>Bivariate Regression (also called Simple Linear Regression):</u>

Analyze → Regression → Linear → Move DV to Dependent variable box and IV to the Independent variable(s) box → Statistics → ensure Estimates and Model fit are selected (they should be already by default) → Continue → OK

<u>Multiple Regression</u>

Analyze → Regression → Linear → Move DV to Dependent variable box and IVs to the Independent variable(s) box → Statistics → Click on Estimates, Model fit, and Descriptives → Continue → OK

<u>Binary Logistic Regression</u>

Analyze → Regression → Binary Logistic → Move the DV to DV box and move the IVs to Covariates box → Leave the method as Enter but click Options → Check CI for exp(B), be sure that 95 is in the box (this will provide 95% CI for the odds ratio of each predictor's contribution to the equation) → Click OK.

Chi-Square Test

Analyze → Descriptive Statistics → Crosstabs → move 1 of the variables to the Row(s) box and 1 of the variables to the Column(s) box (doesn't matter which of the variables you put in each of these boxes) → Statistics → Chi-Square → Continue → Cells → Select observed and expected → Continue → OK

McNemar test

Analyze → Nonparametric Tests → Related Samples → Select "Customized analysis" under Objective → Click Fields, select "Use custom field assignments" and Move the 2 variables to Test Fields → Click Settings, select Customized tests → Select McNemar's test (2 samples) → click Run

Independent-samples t-test

Analyze → Compare means → independent samples t-test → Put the IV in the Grouping Variable box and the DV in the Test Variable(s) → Define groups (you may need to click on the Grouping Variable box first) → Type in the codes for the groups being compared (i.e., codes for the values of the IV) → Click OK

Mann-Whitney U Test

Analyze → Nonparametric Tests → Independent Samples → Select "Customized analysis" under Objective → Click Fields, select "Use custom field assignments" and Move the dependent variable to Test Fields and Independent variable to Groups → Click Settings, select Customized tests → Select Mann-Whitney U test (2 samples) → click Run

Alternative (old) route:

Analyze → Non-parametric → Legacy Dialogue → 2 Independent Samples → Move IV to grouping variable box and DV in test variable box → Define groups (you may need to click on the Grouping Variable box first) → Type in the codes for the groups being compared (i.e., codes for the values of the IV) → Continue → Select Mann Whitney → OK

Paired-samples t-test

Analyze → Compare means → paired samples t-test → Put the 2 variables side-by-side in the Paired Variables box → OK

Wilcoxon Matched-pair Signed-rank-test

Analyze → Nonparametric Tests → Related Samples → Select "Customized analysis" under Objective → Click Fields, select "Use custom field assignments" and Move the 2 variables to Test Fields → Click Settings, select Customized tests → Select Wilcoxon matched-pair signed-rank (2 samples) → click Run

One-Way ANOVA Test

Analyze → Compare means → One-way ANOVA → Move IV to Factor's box and DV to variables box → Post-hoc → Select Games Howell and Tukey → Options → Select Descriptives, Homogeneity of Variance Test, and Means Plot → Continue → OK

Kruskal-Wallis H Test

Analyze → Non-parametric tests → Independent Samples → select Customize analysis → Fields → select Use custom field assignments, moved DV to test fields and IV to Groups → Settings → Customize tests → select "Kruskal-Wallis 1-way ANOVA (k samples)" → Run

Alternative (old) route:

Analyze → Non-parametric → Legacy Dialogue → K Independent Samples → Move IV to grouping variable box and DV to test variable box → Define range (you may need to click on the Grouping Variable box first) → Type in highest code and the lowest code for the values of the IV → Continue → Kruskal-Wallis H → Options → Select Descriptives → Continue → OK

Friedman's Two-Way ANOVA by Rank Test

Analyze → Nonparametric Tests → Related Samples → Select "Customized analysis" under Objective → Click Fields, select "Use custom field assignments" and Move all the variables to Test Fields → Click Settings, select Customized tests → Select Friedman's Two-Way Analysis of Variance by Ranks → click Run

Two-Way (Factorial) ANOVA Test

Analyze → General Linear Model → Univariate → Move IVs to the Fixed Factors box and DV so the Dependent Variable box → Plots → Move one IV to the Horizontal Axis Box and the other IV to the Separate Lines Box (doesn't matter which variable you put in each of these boxes, but your graph might not look the same as others if you don't do it the same as others) → Add → Continue → Options → Descriptive Statistics, Estimates of Effect size, and Homogeneity Tests → Continue → Post-Hoc → Move variables to Post-Hoc tests box → Tukey (and Games-Howell if it allows this) → Continue → OK

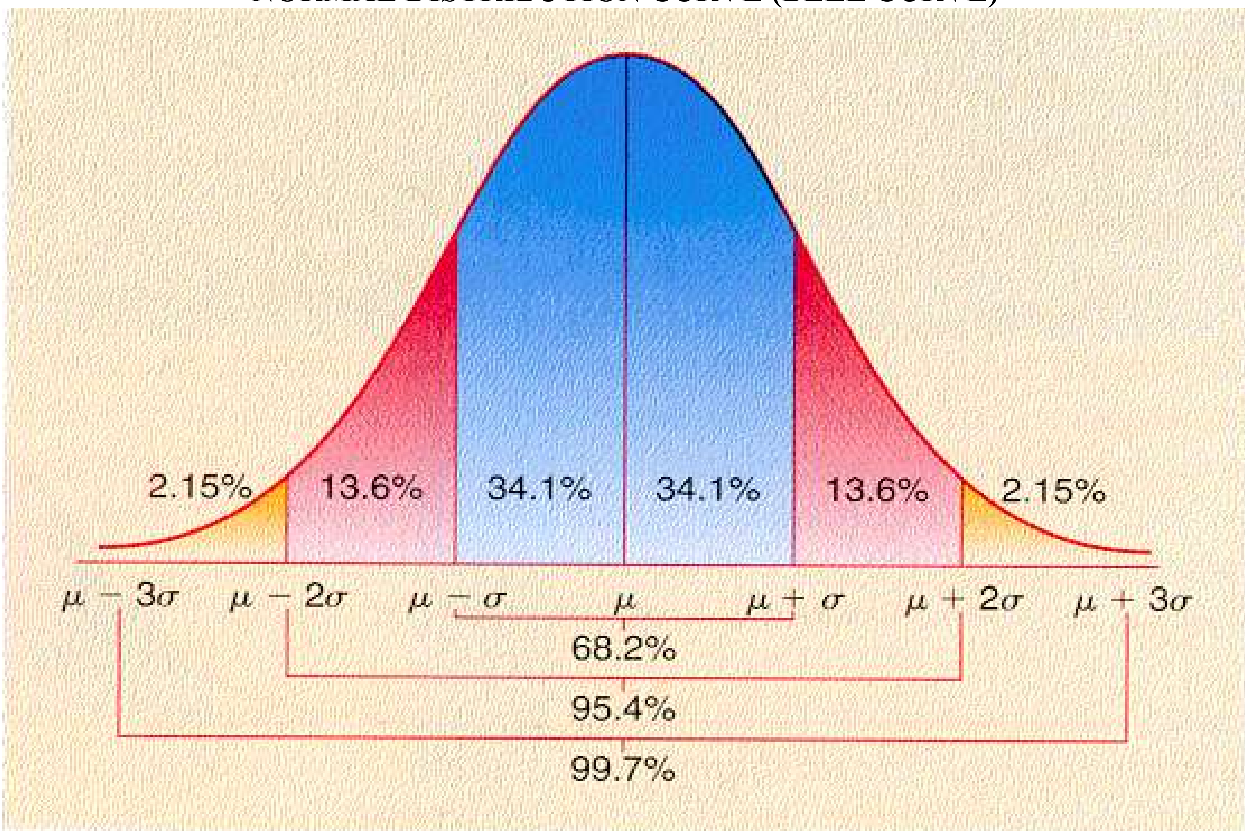
GLM Repeated Measures ANOVA Test

Analyze → General Linear Model → Repeated Measures → You may change the name of IV in the "Within-Subjects Factor Name" box from 'Factor' to 'Time' → Define the number of repetition in the Number of Levels box → click "Add" → Click Define → move your variables to the respective boxes (i.e., move all the repeated variable to "Within-Subjects Variables (factor 1 or Time) → (if there is between subjects variables and covariates in your dataset, please move them to their appropriate boxes) → Click Options → click the boxes for "Descriptive statistics, Estimates of Effect size, Homogeneity tests … → Continue → OK

Cochran Q Test

Analyze → Nonparametric Tests → Related Samples → Select “Customized analysis” under Objective → Click Fields, select “Use custom field assignments” and Move the 3 variables to Test Fields → Click Settings, select Customized tests → Select Cochran’s Q (k samples) → click Run

NORMAL DISTRIBUTION CURVE (BELL CURVE)



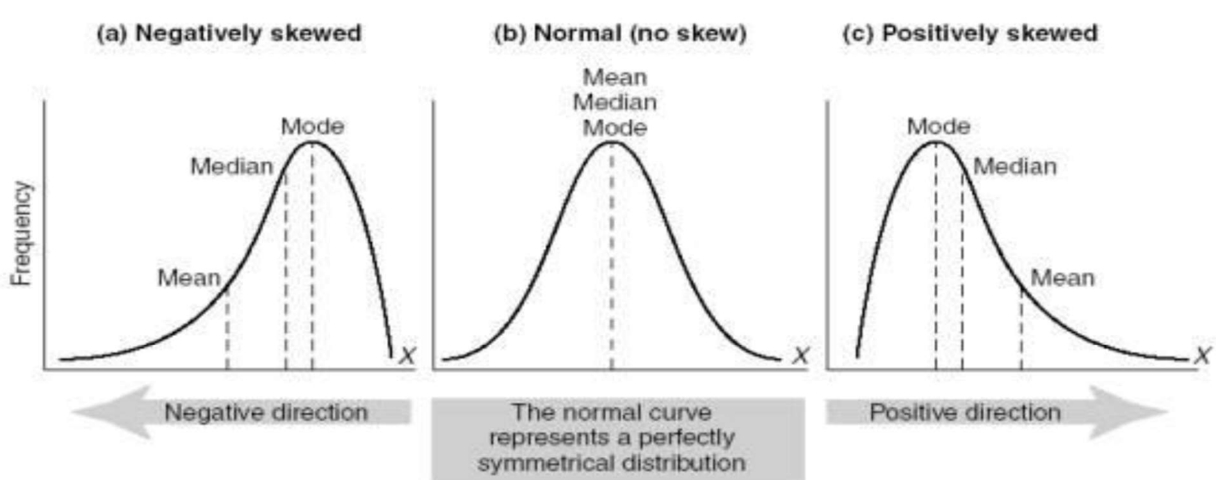


FIGURE 15.6 Examples of normal and skewed distributions