

Index

A

a posteriori comparison 113, 114, 115, 130–34, 137
a priori comparison 113, 114, 137
 absolute measurement error 76
 agreement index 89, 91, 92, 93–97, 98–102
 alpha coefficient 64, 76, 83–84, 87
 alpha level 106, 108
 alternative hypothesis 106
 analyses, running 8–9
 analysis of variance *see* ANOVA
 ANALYZE command 8
 ANOVA 113, 115, 128–34
 assessment 1, 2
 attenuation 29
 authenticity 1

B

bar chart 12, 13, 18, 19, 25–28, 41–43
 between-groups design 106
 between-groups variance 113
 bimodal distribution 12, 13
 biserial correlation 29
 bivariate distribution 1, 29, 46–47, 51, 52, 53
 box-and-whisker plot 12, 20, 21
 Brown-Forsythe F statistic 129, 130

C

categorical variable 1
 CI *see* confidence interval
 classical item analysis 54, 55, 56, 57
 classical item statistics 62–65, 66–75
 classical test theory (CTT) 76, 78
 classification error 89
 coefficient alpha (α) 64, 76, 83–84, 87
 coefficient kappa (k^{\wedge}) 89, 91, 98
 coefficient of determination 29, 30
 command syntax 7

compensatory composite score 150, 152, 154
 composite score 150, 152, 158–61
 COMPUTE command 162, 164
 confidence interval (CI) 76, 77, 89, 91, 93, 113, 114, 115, 122–24, 126, 182
 confidence level 113, 115, 116, 117, 135, 137
 confirmatory factor analysis 139, 143
 construct definition 1, 2, 4, 139, 140
 construct validity 1, 2
 continuous variable 1
 correlated sample design 106
 correlation 29, 30, 32–37, 51–53, 139
 correlation coefficient 29, 30, 33, 41, 47–49
 see also Pearson product-moment correlation coefficient; Spearman rank-order correlation coefficient
 CORRELATIONS procedure 48, 50–51, 51–53
 counterbalanced design 76, 80
 covariance 29
 criterion domain 150
 criterion level 89, 90
 criterion-referenced approach 1, 2, 89–105
 Cronbach's alpha *see* coefficient alpha
 curvilinear relationship 29, 31
 cut score 89, 90, 91, 92

D

Data Editor file 9
 Data Editor window 8, 9
 data entry 8, 11
 Data View window 8
 decision study (D-study) 76
 degrees of freedom 12, 13, 130
 dependability indices 76, 80, 102–4
 dependent sample design 106
 dependent variable 113, 114
 DESCRIPTIVE command 63, 64

descriptive statistical analysis 1, 3
 DESCRIPTIVES procedure 22, 23, 163–65
 Detrended Normal Q-Q plots 45–46
 deviation score method 38
 dichotomous scale 1, 2, 64
 directional research hypothesis 106, 107
 discrete variable 1
 distracters 54, 55
 distribution-free statistical tests 113
 distribution of sample means 106, 107
 domain-referenced score 150, 151
 domain score 89, 91

E

ecological correlation 29
 effect size 113
 effective weight 150, 151, 160
 equivalence reliability estimate 76, 77
 estimated proportion of agreement 89, 90, 98
 evaluation 1, 2
 EXAMINE command 20n1, 43, 44
 experimental design 139, 141, 143, 145, 148, 149
 experimental hypothesis 106
 exploratory factor analysis 139
 EXPLORE procedure
 confidence intervals 122–24
 descriptive statistics 20–21, 23
 relationships 43–46, 50, 51–53
 skewness and kurtosis 128
 extrapolation 139, 140
 EXTREME command 44–45
 extreme values 44–45, 46

F

facet of measurement 54, 55, 56, 78
 facilitating cases 29
 factor analysis 29, 30
 false negative 89, 90
 false positive 89, 90, 91
 filenames 9
 fixed facet 76, 77
 flat distribution 12
 formative evaluation 1
 FREQUENCIES procedure
 descriptive statistics 18–20, 22–28, 68–69, 71–72
 percentile ranks 162–63, 164, 165
 relationships 51–53
 skewness and kurtosis 41–43, 49–50
 frequency distribution 12
 frequency polygon 12, 13
 fully crossed design 76

G

generalizability coefficient (r^2) 76
 generalizability study (G-study) 76
 generalizability theory (G-theory) 76, 78, 80, 102
 generalization 139, 140

GENOVA program 102, 103
 grading 1
 Guttman split-half reliability estimate 76, 81, 82, 85

H

heteroscedasticity 29, 47, 49
 high-stakes decisions 1, 4
 histogram 12, 13, 18, 19, 25–28, 41–43
 homogeneity of variance 129–30
 homoscedasticity 113, 114

I

ID= ID command 44
 impact 1
 independent sample design 106, 113
 independent score 113
 independent variable 113, 114
 index of dependability (F) 76, 78, 89, 90, 92
 individual specific standard error of measurement
 ($SE_{meas(X_i)}$) 89, 93
 inferential statistical analysis 1, 2
 interactiveness 1
 internal consistency reliability 66, 76
 internal scale 12
 interpretative argument 139, 140, 141, 142
 interrater reliability estimate 76
 interval estimate 113
 interval scales 1, 2, 152
 intrarater reliability estimate 76, 77
 IRT standard error of estimation 76
 item characteristic function 54
 item difficulty 54, 55
 item difficulty index (p_i) 54
 item discrimination 54, 64
 item discrimination index (D_{m-nm}) 54, 55
 item information function 54
 item parameters 54
 item response theory 54, 55, 56, 57
 item statistics 54
 item-total correlation 64

K

kappa coefficient (k^{\wedge}) 89, 91, 98
 kappa squared ($k^2(X, Tx)$) 89, 93
 Kendall's tau b correlation coefficient 48
 key 54, 55
 Kuder-Richardson formula 20 (KR-20) 76, 77, 103–4
 Kuder-Richardson formula 21 (KR-21) 76
 kurtosis *see* skewness and kurtosis

L

leptokurtic distribution 12
 level of significance 106, 108
 levels of measurement 1, 3
 linear relationship 29, 30, 31
 linear T-score 150, 151, 152, 155, 164, 165
 linear transformation 12, 13

170 INDEX

LISTWISE command 44, 47
 logit scale 54
 low-stakes decisions 1, 4

M

many-facet Rasch measurement 54
 mastery level 89, 90
 mean 12, 13, 14, 15, 17
 MEAN command 63
 measure type, assigning 10
 measurement 1, 2
 measurement error 29, 76, 78
 measurement scale 1, 3
 median 12, 13, 14, 15, 17, 18
 menu bar 8
 mesokurtic distribution 13, 19, 24–25, 106 12
 mGENOVA program 102
 MISSING PAIRWISE command 44
 missing scores 47
 mode 12, 13, 14, 15, 17, 18
 MODEL command 83, 85, 86
 multiple-groups design 106, 108, 115, 116, 117
 multiple linear regression 29
 multitrait-multimethod correlation matrix 139
 multivariate distribution 1, 3

N

'N' symbol 13
 negatively skewed distribution 12, 47
 nested design 76
 nominal scale 1, 2, 5, 12
 nominal weight 150, 151, 158, 159, 160
 non-compensatory composite score 150, 151, 154
 non-directional research hypothesis 106, 107, 108
 non-equivalent group design 139, 149
 non-parametric statistical tests 113
 NONPAR CORR command 48
 norm group 150
 norm-referenced approach 1, 2, 76–88
 normal distribution 13, 19, 24–25, 106 12
 normal probability plots 44, 45–46
 Normal Q-Q plots 45–46
 normalized standard score 150
 NR item discrimination index (D_i) 54, 55, 58–60
 null hypothesis 106, 107, 108, 114, 115, 116, 117, 122, 134, 136, 137

O

object of measurement 76
 observed score 139
 odd-even split-half reliability estimate 76
 one-tailed probability 106, 107, 108, 115, 116, 117, 135, 136, 137
 One-Way ANOVA procedure 128–34
 ONEWAY procedure *see* One-Way ANOVA procedure
 operational hypothesis 106, 107, 108–12
 ordered domain 150

ordinal scale 1, 5, 12
 outliers 12, 13, 21, 29, 46
 output, examining 9
 Output file 9
 Output window 9

P

parallel forms reliability estimate 76, 77
 parallel measures 76, 77
 parametric statistical tests 113, 114
 partial-credit score items 55
 path analysis 29
 peaked distribution 12
 Pearson product-moment correlation coefficient (r) 29, 30, 31, 38–39, 47–49
 percentage correct score 150, 151, 162
 percentile rank 150, 151, 152, 155, 162–63, 164, 165
 phi coefficient (F) 92, 93, 103
 phi lamda (F_i) 89, 90, 91, 93, 103
 planned (*a priori*) comparisons 113, 114, 137
 platykurtic distribution 12
 PLOT command 44
point-and-click method 8–9
 point-biserial correlation 29
 point estimate 113, 114
 population 1, 108
 population parameter 1
 positively skewed distribution 12
 post hoc (*a posteriori*) comparisons 113, 114, 115, 130–34, 137
 practicality 1
 pre-post-test difference index (DIS_{ppd}) 54, 61–62
 PRINT command 9, 48
 procedure, use of term 7
 program, use of term 7

Q

quantitative approaches 140, 143, 148–49

R

random assignment 141
 random facet 76
 random variance 76
 range 12, 13, 14, 15, 17, 18
 range restriction 29, 30, 32
 rankings 150
 ratio scale 1, 12
 raw score 12, 13, 150, 151
 raw score method 39
 regression equation 29
 regression line 29, 31
 relative measurement error 76, 77
 reliability 1, 2
 reliability coefficients 103
 RELIABILITY procedure
 classical item statistics 62–65, 66–75
 coefficient alpha method 83–84, 87

odd-even split method 86, 87
 split-half method 84–85, 87
 research design 2, 145–48
 right-wrong score items 55
 robustness 113, 114, 129–30

S

sample 1
 sample statistic 1, 3
 sampling error 1, 2, 3
 saving files 9
 SCALE command 63, 64, 84, 85
 scale mean 64
 scatterplot 29, 30, 33, 34–37, 46–47
 SCATTERPLOT procedure 46–47, 50, 51–53
 score *see* test score
 SEM *see* standard error of measurement
 semi-interquartile range 12
 single-group design 106, 107, 114, 115, 116, 117
 skewness and kurtosis 12, 13, 18–19, 24–28, 41–43, 49–50, 124, 128
 Spearman-Brown prophecy formula 76, 77
 Spearman-Brown split-half reliability estimate 76, 81, 82, 85
 Spearman rank-order correlation coefficient (ρ) 29, 30, 31, 40–41, 47–49
 split-half reliability estimate 76, 77, 81, 87
 SPSS
 opening and using 8–9
 terms 7
 squared-loss agreement index 89
 squared multiple correlation 64
 stability reliability estimate 77
 standard deviation 12, 13, 14, 15, 16, 17, 18
 standard error 19
 standard error of measurement for absolute decisions
 (SEM_{abs}) 89, 93
 standard error of measurement (SEM) 76, 78, 80
 standard error of the mean 106, 107, 113, 114
 standard score 150, 152
 statistical analyses, running 8–9
 statistical research hypothesis 106, 107, 108–12, 115, 116, 117, 122
 STATISTICS command 44–45, 63, 64, 85
 stem-and-leaf plots 20–21
 structural equation modeling 29
 SUMMARY command 63, 64
 summative evaluation 1
 Syntax file 9
syntax method 8, 9
 Syntax window 8–9
 systematic error variance 76

T

t-test 113, 114, 115, 118–22, 136
 T-TEST command

dependent samples 124–28
 independent samples 124–26
 target language use domain 139
 target score 139, 140
 test content 139, 140, 141, 148
 test information function 54
 test method variance 76
 test-retest reliability estimate 77
 test score distribution 12, 13, 14, 17, 18–22, 24–28
 test score interpretation 139, 142, 149
 test score use 139
 test-taking process 139, 149
 test usefulness 1
 tetrachoric correlation 29
 theoretical hypothesis 106, 107, 108–11
 threshold loss agreement index 89, 91
 TOTAL command 63, 64
 ‘true score’ variance 76, 77
 two-tailed probability 106, 107, 115, 116, 117, 135, 136, 137
 type I error (a) 106, 107
 type II error (b) 106, 108

U

uncorrelated sample design 106
 unidimensionality 54, 55
 univariate distribution 1, 3, 41–46
 universe of admissible measures 76
 universe of generalization 76
 universe score 76, 139
 unordered domain 150, 151, 152
 unsystematic error variance 76
 urGENOVA program 102

V

validation 139
 validation claims 139, 140, 146, 147
 validation counterclaims 139, 140, 146, 147
 validation study 139, 142
 validity 139, 140, 141
 variable 1, 5, 8, 10, 113, 114
 Variable View window 8
 VARIABLES command 44, 48, 63, 84, 86
 variance 12, 18, 19, 64
 VARIANCE command 63
 variance component (s^2) 76
 verbal protocol analysis 139, 140

W

Welch t statistic 129, 130
 within-group design 106
 within-groups variance 113

Z

z-score 150, 151, 155, 163–65
 z-tests 118–22