

Index

- α , 106, 107, 160
 β , *see* power
 ϵ^2 , 60
 η^2 , 60
 ω , 199
 ω^2 , 60
 A' , *see* \hat{P}
 aligned rank transforms, 148
 alpha, 117
 alternative hypothesis, 27
 ANCOVA, 139, 158
 ANOVA, 139, 160
 assumptions of, 145, 146
 avoiding, 150
 effect size for, 140, 147
 exploration with, 158
 Greenhouse–Geisser correction, 147
 Huyhn–Feldt correction, 147
 interaction effects, 143
 main effect, 142
 mixed design, 147, 151
 as omnibus test, 140
 one-way, 140, 141
 repeated measures, 141, 146
 robust alternatives, 147
 robustness, 145
 split-plot design, *see* mixed design
 two-way, 72, 142, 160
 articulation, 73, 144
 axioms, 40
 of logic, 12
 of plausibility, 43
 of probability, 41

 Bayes factor, 45
 Bayes's rule, 39, 44
 Bayesian, 162
 null effects, 63

 Bayesian statistics, 35, 38
 over-testing, 162
 unconstrained, 44
 bifactor analysis, 199
 big data, 21, 35
 fishing in, 164
 binomial distribution, 88
 bloated specific, 190, 210
 Bonferroni correction, 161
 Type II errors, 161
 bootstrapping, 136
 Brunner–Munzel test, 59, 122, 215

 calculator
 new designs for, 20
 CCPIG questionnaire, 187
 central limit theorem, 89, 127
 chi-squared test, 92
 classical test theory, 192, 205, 223
 Cliff's test, 59, 122, 215
 closurization, 156
 closurize, *see also* openumaggen
 cluster, 171
 Cohen's d , 33, 57
 complexity in analysis, 77
 confidence interval, *see also* parameter
 estimation, 33, 61
 of effect size, 61
 matching to statistical test, 61, 62
 confirmatory factor analysis, 192
 confound
 in correlation, 173
 construct validity, 178
 contrast, 151
 correlation, 59, 159, 168
 effect of clustering, 171
 effect of outliers, 170
 effect size for, 59
 is not causation, 167
 is not correlation, 168

correlation (*cont*)
 Pearson's r , 168
 perfect, 168
 Spearman's ρ , 173
 spurious, 167
 zero, 168

Cronbach's α , 177, 204
 problems with, 199
 recommended values of, 206

data type, 64
 dependent variable, 27, 80
 directional hypothesis, 80, 151
 discussion, 75, 156
 distribution, 28
 distribution of data, 28
 Likert item, 28
 normal, 28

dominance, *see also* effect size, 54, 58, 93
 non-parametric test, 121
 tests for, 122
 vs location, 221, 227

effect size, 32, 47, 53, 74
 vs p -value, 33
 big, 62
 dominance, 54, 58, 73
 for power calculation, 107
 informed estimates, 108
 location, 54, 55, 73, 140
 matching to statistical test, 58,
 59, 66
 meaning, 62
 small, 34, 62
 variation, 54, 59, 73, 140
 zero, 62, 92

ensmartness, 205

errors in interaction, 19, 76, 88, 163

estimating parameters, 32

evidence, 27, 31, 35, 36, 48

experiment
 as first-class object, 19
 need for statistics, 18, 40, 72
 one is not enough, 22, 36, 51, 85, 106,
 157, 165
 as severe test, 16, 18

experiments in HCI, 18

exploratory factor analysis, 192

exploratory study, 157

exploring data, 76, 78, 85, 102, 155
 rules, 165

F -test, *see* ANOVA, 231

factor, 190

factor analysis, 174, 187
 g , 198

bifactor, 199
 interpreting loadings, 196
 methods, 192
 number of factors, 194
 rotation, 195

factor loading, 191
 sign of, 196, 197

factorial design, 72

falsification, 13, 15

Faraday's motor, 19

Feyerabend, Paul, 14

file-drawer effect, *see also* p -hacking, 26

fishing, 155, 162

Frequentist, 61

Frequentist statistics, 38, 106
 critique of, 47

Friedman test, 116, 120

General Aggression Model, 15

general factor, g , 198

Hacking, Ian, 19

heteroscedasticity, 129

homoscedasticity, 127, 145

honesty, 26, 35, 36, 50, 77, 85, 162, 164, 165

hypothesis
 alternative, 27
 null, 27, 28, 38
 statistical, 49
 substantive, 34, 49

immersion in games, 27

independence of observations, 127,
 145, 146

independent variable, 27, 80

induction, 12, 31

interquartile range, 97

isolation of phenomena, 19, 35

item response theory, 192, 222

Jaynes
 Edwin, 42

Kolmogorov–Smirnov test, 92, 130

Kruskal–Wallis test, 116, 160

Kuhn, Thomas, 14

Levene test, 130

Likert item, 87, 176
 labelling, 181
 midpoint, 178
 response format, 176
 statements, 189
 tests for, 215

Likert scale, 176, 188

Likert, Rensis, 176

- location, *see also* effect size, 54, 55, 93, 96, 118,
 126, 131, 140, 147
 vs dominance, 221, 227
 effect of outliers, 95
 mean as measure of, 55
 measures of, 55, 131
 non-parametric test, 121
 robust measures of, 131
 log-normal distribution, 223
- M-estimator, 134
 machine learning, 21
 magnitude-based inference, 35, 64
 MANCOVA, 75, 139
 Mann–Whitney test, 58, 109, 115, 215
 MANOVA, 139
 Mauchly's test for sphericity, 146
 Mayo, Deborah, 15
 mean, 87, 125
 geometric, 130
 M-estimator, 134
 trimmed, 133, 150
 Winsorized, 134
 median, 131
 lack of robustness, 131
 median absolute deviation, 97
 mediation, 174
 menu design, 72
 microwave racing, 74
 mixed normal data, 222
 moderation, 174
 moo, 157
- new experimentalism, 18
 NHST, 25, 26, 53, 162
 argument form, 30, 31
 Neyman–Pearson form, 31
 problems of, 25
 non-parametric test, 93, 114
 alternative to ANOVA, 148
 lower power, 119
 robustness to outliers, 119
 non-parametric tests
 (not) distribution free, 120
 normality, 60, 86, 114, 127, 145, 222
 central limit theorem, 89
 definition of outliers, 96
 equation of, 87
 multivariate, 146
 tails of tests, 81
 tests for, 91, 92, 130
 transformations for, 130
 null hypothesis, 27, 28
 null hypothesis significance testing, *see*
 NHST
- omnibus test, 140
 one-tailed test, *see also* directional
 hypothesis, 81
 optional stopping, *see also* *p*-hacking,
 47
 outlier, 75, 88, 95
 causes of, 98
 definition from normality, 96
 omitting from analysis, 102
 problem for correlation, 170
 robust definitions, 97
 over-testing, 140, 155, 159
 avoiding with ANOVA, 140, 160
 by ANOVA, 144, 160
- \hat{P} , 58, 123
 confidence intervals of, 61
p-hacking, 26, 31, 35, 47, 84, 106, 164
p-value, 25, 38, 53
 problems of, 47
 rituals for, 26
 vs effect size, 33
 parallel analysis, 194
 parameter estimation, 48, 54, 60
 parametric tests, *see also* normality,
 87
 Pearson's *r*, 168
 perceived aesthetics, 59
 personality and HCI, 158
 placebo effect in games, 48
 planned comparison, 161
 planning principle, 73, 79
 articulation, 73
 honesty, 77
 simplicity, 75
 Plato's cave, 190
 plausibility, 42
 Popper, Karl, 13
 power, 66, 104, 119, 215
 β , 106
 calculation, 108
 determinants of, 107
 sample size, 107
 and Type II errors, 106
 priming, 15
 principal component analysis, 192
 probability, 31, 40, 41, 43
 conditional, 41
 as plausibility, 42
 posterior, 39, 46
 prior, 39, 45
 probability distribution, *see also* distribution
 of data, 28
 binomial, 88
 chi-squared, 101

- probability distribution (*cont*)
 discrete, 87
 log-normal, 90, 101
 mixed normal, 91
 mixed-normal, 128
 non-parametric, 114
 normal, 86
 power law, 88
 probability theory, 28
 proportion of variance, *see* variation
 pseudo-independent variable, 158
 median split, 158
 tercile split, 158
 psychometrics in HCI, 187
- quartile
 lower, 97
 upper, 97
- questionnaire, 186
 tests for, 215
 validity, 187
- questionnaire development
 factor analysis, 190
 item generation, 188
 sample sizes, 190
- R, *see* statistics software, 61
- rank-based tests, *see* non-parametric tests
- ranking data, 115
- Rasch model, 222
- regression
 linear, 168
 multivariate, 174
 regression line, 169
- reification of concepts, 201
- reliability, 177, 204, 205
 ω , 199
 consistency, 177
 internal, 177
 split-half correlation, 177
 stability, 177
 test-retest, 177
- replication, 33
- reporting analysis, 73
- robustness
 alternatives to means, 133
 ANOVA, 145
 Kruskal–Wallis test, 148
 measures of location, 131
 non-parametric tests, 117
 t -test, 129
 robustness of tests, 86, 110
- sample size, 108
 big enough, 90
- sampling distribution, 89
 mean of, 89
- scatterplot, 168, 173
- science, 11
 in HCI, 1, 12
 paradigm shift, 14
 paradigms of, 14, 25
- scree plot, 194
- sea-pig, *see* CCPIG questionnaire
- semantic differential, 177
- severe test, 12, 15, 31, 48–50, 73, 75, 78, 151
 as experiment, 16
 vs exploratory study, 85, 157
 vs fishing, 163
 further exploration, 157
 need for ANOVA, 140
 other than an experiment, 17
 using directional hypothesis, 80
 with NHST, 33
 in relation to theory, 17
- Shapiro–Wilk test, 92, 130
- Sign test, 121
- significance, 83
 practical, 53, 56, 62
 statistical, 25
- simple structure, 191, 195
- simplicity, 75, 144, 174
- situated knowledge, 186
- skew, 88
- spatial presence, 115
- Spearman's ρ , 173
- sphericity, 146
- SPSS, 61
- standard deviation, 87, 141
- standard error, 127
- statistical argument
 gold standard, 34
 NHST, 30
- statistics
 need for experiments, 18, 40
 planning for, 72, 109
- statistics software, 5, 193
 R, 6
 WRS2 package for R, 6
 Structural Equation Modelling, 174
- t -test, 21, 29, 57, 74, 81, 91, 111, 125, 215, 228
 assumptions of, 127
 effect of outliers, 95
 effect size, d , 57
 equation of, 126
 robust alternatives, 135
 robustness, 129
 Welch's adjustment, *see also* Welch test
 Yuen–Welch test, 136

Index

235

- test assumptions, 74, 220
- test for equality of variance, 130
- transforming data, 130
 - ranks, 116
- two one-sided tests (TOST), 64
- two-tailed test, 81
- Type I error, 62, 91, 105, 140
 - actual error rate, 110, 118
 - non-parametric test, 117
 - quality of tests, 215
- Type II error, 91, 105
 - non-parametric test, 119
 - quality of tests, 215
- user experience, 186
- validity
 - of questionnaires, 187, 200
 - of a scale, 205
 - statistical, 75
- variance, 141
 - for reliability, 206
- variation, *see also* effect size, 54, 59, 140
 - and normality, 60
- Welch test, 131, 147, 215
- Wilcoxon test, 116, 120
- Yuen–Welch test, 136, 215