

# Index

Numbers: **bold** = table, *italics* = figure

- 2 × 2 tables, 116–119, **120** *passim*  
 3 × 3 tables, 134–135, **135**, **138**
- adjustment for covariance method,  
 49–55, 130  
 advantage, 51  
 comments, 54  
 example, 50–54  
 extension of mixed model analysis  
 (MMA), 54
- age-period-cohort models, 2
- Akaike's information criterion  
 (AIC), 100, **151**, 227, 230,  
 233–234, 238, 240
- Altman, D.G., 8
- Amsterdam Growth and Health  
 Longitudinal Study, 4, 64, 85, 143
- analysis of covariance, 167–170, 172,  
 181–186 *passim*
- analysis of variance (ANOVA), 4,  
 10, 12–13, 19, 29, 220 *passim*  
 cross-sectional, 25  
 naive, 18–19
- antiretroviral therapy (ART), 108
- Applied Longitudinal Data*  
*Analysis*  
 book aim, xi  
 book emphasis, xi  
 descriptive information, 5  
 examples, 4  
 general approach, 4  
 main interests, 2  
 most important feature, xi  
 new material in third edition, 6  
 prior knowledge assumption, 4  
 study design, 1–4
- area under curve (AUC), 178, 178  
 calculation (Equation 10.9), 177  
 use as summary statistic, 177
- autoregression coefficient, 88, 167,  
 181
- autoregressive models, 88–90, 130,  
 181–183  
 data structure needed, **88**  
 example, 88  
 first-order (Equation 5.5), **88**  
 auxiliary variables  
 definition, 210
- back pain, 128, 187, 189, 190, **219**
- Barthel index, 153–156, 157
- baseline, 164–165, 169  
 adjustment for, 190, 197  
 definition, 164
- baseline hazard function, 111, 113,  
 115
- baseline measurement, 164–165,  
 169–170, 172, 178, 186
- baseline value, 165, 167–168, 170,  
 172, 176–178, 181–184, 186,  
 198, 216
- Bayesian information criterion  
 (BIC), 100, **151**, 155, 161, 227,  
 229, 233–234, 238, 240
- between-subjects and within-  
 subjects relationship, 6  
 disentanglement (models), 76–91
- between-subjects effects, 17, 22, 24,  
 179, 181
- between-subjects factor, 19–21, 25
- between-subjects relationship, 38,  
 41, 67, 141
- between-within models. *See* hybrid  
 models
- binomial distribution, 122, 152,  
 159–160
- birth cohort, 2–3
- blood pressure, 143, 172, 175–178,  
 179–180, 182, 184–185, 187,  
 198–199, 201, **218**
- body height  
 development with age, 3  
 body mass index (BMI), 201  
 body weight, 67, 67, 69, 71, 96
- Bozdogan's criterion (CAIC), 232,  
 234, 238
- broad data structures, 5
- cardiopulmonary fitness, 96, 143
- case-control studies, 1
- categorical outcome variables, 5,  
 134–141  
 comparing groups, 135  
 example, 135–136  
 more than two measurements,  
 135  
 outputs, 136, 139–142
- regression based methods (with  
 example), 136–141  
 two measurements, 134–135
- categorical variable, 20, 32–33, **59**,  
 60–62, 65–66, 72–73, 132, 133,  
 161, 183, 190, 221  
 imputation methods, 207
- causality  
 criteria, 1, 92
- causality in observational  
 longitudinal studies, 92–115
- G-methods, 107–110
- joint models, 110–115
- longitudinal mediation models,  
 94–105
- outputs, 94–95, 99–110, 112–114
- time-lag models, 92–94
- cause and effect, 1
- caution required, 19, 105, 133, 190,  
 212, 217, 242
- CD 4 count, 108
- censoring, 162  
 outcome measure with floor  
 effect, 153
- CHD. *See* coronary heart disease
- Chi-square, 29–30, 36–37, 46, 51,  
 56, 78, 118, 126, 134, 222
- Chi-square test, 35, 116, 188, 202
- cholesterol, 5, 5, 7–9, 15, 18, 26, 28,  
 49, 51–55, 57–59, 64, 70–72,  
 74, 76–77, 79, 81, 83–84,  
 94–95, 220  
 and body weight, 67  
 categorical variable, 135, 135,  
 137–138, 138, 140–141, 150
- Friedman test, 30
- linear development over time, 57
- longitudinal development, 17
- longitudinal regression methods,  
 34
- CI. *See* confidence interval
- Claassen, E., 235
- clustering, 143–144, 147–148, 196,  
 226  
 definition, 143
- clustering score, **143**
- Cochran's Q, 118  
 disadvantage, 117

## Index

- coefficient of interest, 31, 67, 121, 169–170, 191
- cohort effects, 1, 3, 3
- common trend assumption, 198
- comparison of changes, 169–170
- comparison of changes (adjusted for baseline)  
 basically same as ‘analysis of covariance’ ( $qv$ ), 168
- compound symmetry assumption, 11
- conditional frailty model, 194
- confidence interval, 7–8, 36, 46, 100, 114, 116–119, 121, 185, 189, 190 *passim*
- confounders, 68, 105, 183, 196–198
- CONSORT statement, 216
- contemporaneous relationship, 96–97, 101
- continuous outcome variables, 4, 7–30, 61, 122, 124, 136, 216
- comments, 25
- comparing groups, 19–25
- components, 14
- different contrasts, 26–29
- floors and ceilings, 152
- GEE analysis (software), 220–223
- imputation methods, 207–210, 214
- mixed model analysis (software), 235
- modelling of time, 56–75
- more than two measurements, 9–19
- non-parametric equivalent of GLM for repeated measures, 29–30
- outputs, 9–10, 12–16, 18–19, 21–22, 24, 27, 29
- paired *t*-test (non-parametric equivalent), 8–9
- post-hoc procedures, 25
- RCT with more than one follow-up measurement, 172–186
- RCT with one follow-up measurement, 165–172
- regression-based methods, 31–55
- statistical significance (sample size needed), 217
- time, 13–15
- transformation factors, 14
- two measurements, 7–8
- univariate versus multivariate approach, 19
- continuous variables (imputation of missing data), 205–207
- comments, 206
- cross-sectional methods, 205
- longitudinal methods, 206
- multiple imputation, 206–207
- control group, 164–166, 168–169, 186, 216
- coronary heart disease, 143–144, 143, 147–148
- correlated observations, 32, 44, 49, 54, 84, 89, 109, 168, 200
- correlated residuals, 44, 54, 89
- correlation coefficients, 43–44, 50
- correlation matrix, 49, 54–55, 87, 222
- correlation structures, 48, 50, 123, 124
- autoregressive, 43, 45–46
- exchangeable, 42, 44–46, 48, 55, 57, 121, 190, 192, 205, 220
- five-dependent, 43, 44–45
- four-dependent, 122
- independent, 42, 44, 46, 87, 90, 122, 124
- m*-dependent, 44
- Poisson GEE analysis, 145, 146
- stationary dependent, 42, 46
- two-dependent, 43
- unstructured, 43, 45–46
- within-subjects, 43–44, 54
- count outcome variables, 141–151
- comments, 150
- example, 143–147
- GEE analysis, 137–146
- GEE versus mixed model analysis, 147
- mixed model analysis, 146–147
- negative binomial regression analysis, 148–150
- outputs, 144–145, 147–148, 150–151
- counterfactuals, 197
- counting method, 192, 195
- covariance. *See also* adjustment for covariance method  
 negative versus positive, 39
- covariance equation, 49
- covariance matrix, 19, 22
- purpose, 49
- covariance parameters, 51, 229, 232
- covariance structures
- exchangeable, 50, 51, 53
- five-dependent, 53
- unstructured, 50, 52–54
- covariate of interest, 56, 80–81, 89, 103, 108
- time-dependent, 82
- time-independent, 82
- covariates, 1, 37, 92, 94
- can be continuous, dichotomous or categorical, 142
- definition, 1
- time-dependent, 41, 80–81
- Cox regression analysis, 110–112, 114
- use, 111
- Cox regression for recurrent events, 194–195, 195
- problem, 191
- use, 191
- cross-over design, 195
- cross-sectional linear regression analysis, 31–32, 49, 165
- naive, 36
- cross-sectional logistic regression analysis, 121
- cross-sectional methods, 1, 4, 152, 164, 178
- cross-sectional problems, 7, 31
- cross-sectional regression analysis, 32–33, 38, 196
- cubic relationship, 15
- cubic S-shaped function, 58
- Data Augmentation (DA), 207
- data structures, 5, 6
- datasets
- six-repeated-measurement, 7
- without missing data, 49
- datasets with missing data, 49, 202, 203
- analysis performed on, 204–205
- GEE versus mixed model analysis, 214, 214
- degrees of freedom, 7–8, 10–16, 19, 29–30, 36–37, 51, 56, 58, 78
- dependency of repeated observations within subjects (longitudinal problem), 7
- descriptive statistics, 196
- deviation score, 76–78, 80–82, 130, 132, 141 *passim* definition, 76, 79
- diabetes, 160
- dichotomous outcome variables, 1, 5, 116–133, 135–136, 147, 204, 216
- comparing groups, 117, 119
- GEE analysis (software), 224–226
- imputation methods, 207
- imputation methods (missing data), 210–212
- longitudinal intervention studies, 187–194
- longitudinal regression methods, 119–133
- mixed model analysis (software), 235–242
- more than two measurements, 117
- outputs, 118–119, 122–123, 125, 127, 131–132
- regression-based methods, 189–193
- simple analysis, 188
- software (further reading), 235
- statistical significance (sample size needed), 218

- total change over time, 117  
 two measurements, 116–117
- dichotomous outcome variables  
 (example), 117–119  
 comparing groups, 119  
 development over time, 117–118  
 two measurements, 117
- dichotomous outcomes, 106, 108, 196
- difference contrast, 26–27
- difference in difference method, 197–200  
 limitation, 199
- difference in proportions  
 problem, 116
- direct effect, 94, 96, 98–100, 102
- Directed Acyclic Graph (DAG), 103
- Disability Index of Health  
 Assessment Questionnaire (HAQ-DI), 153, 158, 159–160
- disease activity, 155, 159–160
- drop-outs, 201, 202, 207, 218–219
- dummy variable coding, 59, 61
- dummy variables, 32–33, 60–62, 65–66, 161, 183, 186, 190–191
- econometrics, 79, 152, 213
- educational research, 32
- effect (outcome variable), 1
- effect estimates, 184–185, 199, 206, 219
- effect modifiers, 196
- effect size, 216–217
- error mean squares, 15
- error sum of squares, 12–13, 19–20, 25
- error variance, 36, 126, 233
- eta squared, 18
- excess of zeros, 150, 152, 161
- factor of interest, 10
- fixed effects, 41, 79
- flexible parameter method, 113–114
- floor or ceiling effects  
 longitudinal two-part models (with example), 159–163  
 outcome variables, 152–163  
 outputs, 154, 156, 159–160, 162–163  
 tobit mixed model analysis (with example), 153–159
- forced expiratory volume in one second (FEV1), 85, 88
- forced vital capacity (FVC), 85, 88
- Friedman test, 29–30  
 equation, 29
- Gauss–Hermite quadrature method, 235–236, 238, 240, 242
- Gaussian family, 45, 47, 49, 57, 87, 90, 121
- GEE. *See* generalised estimating equations
- general group effect, 20
- generalised estimating equations, 42–46  
 autoregressive linear analysis, 90  
 comparison with mixed model analysis, 48–49  
 comparison with mixed model analysis (count outcome variables), 147  
 comparison with mixed model analysis (longitudinal regression methods), 129
- continuous outcome variables (software), 220–223
- correlation structures, 42–44
- count outcome variables, 137–146
- datasets with missing data, 205, 205
- dichotomous outcome variables (software), 224–226
- example, 44–46
- further reading, 133
- linear analysis, 45, 47, 49, 55–57, 87, 121–122
- logistic analysis, 122–123, 129–130, 131–132, 189–190, 192
- longitudinal intervention studies, 168
- longitudinal regression methods, 121–124
- generalized linear model (GLM), 10  
 assumptions, 10  
 graphical representation, 180–181  
 limitations, 179  
 mixed model analysis ‘extension of’, 33  
 non-parametric equivalent of for repeated measures, 29–30  
 RCT with more than one follow-up measurement, 178
- repeated measures, 10, 12–13, 15, 18–21, 25–26, 33, 60, 178, 205, 214, 220 *passim*
- repeated measures (adjusted for baseline), 178
- repeated measures (basically ‘testing method’), 25
- repeated measures (complex study designs), 25
- repeated measures (problems), 25
- repeated measures (simple longitudinal dataset), 13
- univariate approach (numerical example), 12–13
- genmod procedure, 220, 225
- glimmix procedure, 236
- GLLAMM, 159, 162  
 key to abbreviation, 138
- G-methods, 107–110, 197  
 types, 108
- Greenhouse–Geisser method, 15–16, 18, 21, 23–24, 179–180
- group by time interaction effect, 20
- growth curve analysis, 56–60
- growth curve parameters, 71–72
- growth trajectories  
 subjects with different (classification), 70–75
- Helmert contrast, 26, 28
- high density lipoprotein cholesterol, 143
- HIV patients, 108
- hot-deck method, 205, 207
- Hotelling’s  $T^2$ , 10
- Hotelling’s Trace, 16, 22
- Huber–White sandwich estimator, 43, 46
- Hurvich and Tsai’s criterion (AICC), 227–228, 230, 233–234, 238
- Huynh–Feldt method, 15–16, 18, 23–24, 179–180
- hybrid models (to disentangle between- and within-subjects relationship), 6, 76–82  
 comments, 82  
 data structure needed, 77  
 dichotomous covariate, 85  
 direct estimation of, 78–80  
 example, 76–78  
 with time-dependent covariates, 80–81
- hypercholesterolemia, 111–114, 117–119, 122–123, 125–127, 132, 213, 220
- hypoglycaemic events, 161, 162–163
- id number, 33
- imputation methods, 205–213  
 additional analyses, 210  
 categorical variables, 207  
 comments, 212  
 continuous variables, 205–209, 214  
 cross-sectional, 215  
 cross-sectional versus longitudinal, 205  
 dichotomous variables, 207, 210–212  
 multiple imputation in combination with mixed model analysis, 203–209  
 problem, 206
- independent sample  $t$ -test, 18, 20, 176–178, 196, 202–204, 204
- indirect effect, 94, 96, 98–99, 101–102

## Index

- individual mean value, 76–77,  
79–80, 82, 130, 141
- intercepts, 33–34, 48, 126  
definition, 14
- intervention effect, 165, 167–170,  
172, 181, 183, 186, 191,  
196–198
- intervention group, 165–166, 168,  
195, 216, 218  
short-term versus long-term  
effects, 176–177
- intervention studies, 1, 61, 70, 92,  
117, 177, 185, 216
- intervention variable  
time-dependent versus time-  
independent, 195
- intraclass correlation coefficient  
(ICC), 36
- inverse probability weighting, 109,  
197
- Laplace method, 239, 242
- last observation carried forward  
(LOCF) method, 206
- last value carried forward. *See* LVCF
- latent class growth analysis (LCGA),  
72
- latent class growth mixture  
modelling (LCGMM), 73
- latent growth curve parameters,  
74
- latent normal distribution, 159
- latent variable modelling  
framework, 74
- latent variables  
definition, 103
- learning effects, 4
- lifestyle parameters, 94
- likelihood ratio test, 36–37, 49–51,  
54, 56, 58, 125–126, 138, 147,  
159  
nature and purpose, 35
- linear development, 14–15, 20, 24,  
26, 56–58, 61, 63, 71, 111–114,  
149, 154
- linear interpolation method, 206
- linear regression analysis, 4, 136,  
167, 199
- linear relationship, 13–14, 31
- linear transformation factors,  
14
- linear transformed dataset, 14
- link function, 45, 121, 144
- log likelihood, 36–37, 51, 90, 125,  
140, 161  
definition, 35  
tobit versus linear mixed model  
analysis, 155
- logistic mixed model analysis, 127,  
190  
effect of new treatment on  
recovery, 191
- logistic regression analysis, 4, 35,  
109, 119, 121, 126, 131, 191,  
202  
naive multinomial, 139
- loneliness, 148, 151, 151
- long data structure, 5, 31, 39, 198
- Longitudinal Aging Study  
Amsterdam (LASA), 148
- longitudinal and cross-sectional  
relationships  
pooled analysis, 38
- longitudinal data analysis, 136  
application of statistical methods,  
xi  
correct interpretation, 201, 204  
mixed models, 32–34  
software, 220–242
- longitudinal intervention studies, 6,  
164–200, 216  
beyond RCT, 197–198  
comments, 196, 199  
dichotomous outcome variables,  
187–194, 217  
difference in difference method,  
198–200  
outputs, 169–171, 173–174,  
179–180, 182–185, 187–188,  
190–193, 195, 198–199  
regression based methods, 168  
stepped wedge designs, 196
- longitudinal linear regression  
analysis, 119, 142
- longitudinal logistic regression  
analysis, 142
- longitudinal log-linear regression  
analysis, 142
- longitudinal mediation models,  
94–105  
comments, 103  
data structure needed, 98, 106  
example, 96  
summary, 107
- longitudinal multinomial logistic  
regression analysis, 142
- longitudinal problem, 31
- longitudinal regression analysis, 110  
one of most difficult parts, 76
- longitudinal regression imputation  
method, 206
- longitudinal regression methods,  
31–32  
adjustment for covariance  
method, 130  
comments, 133  
dichotomous outcome variables,  
119–133  
example, 34–37  
fixed part of model, 36  
GEE, 121–124  
GEE versus mixed model analysis,  
127–129  
mixed model analysis, 124–127
- models to disentangle between-  
and within-subjects  
relationship, 130–132  
random part of model, 36
- longitudinal relationship  
continuous outcome and  
covariates (equation), 31
- longitudinal studies  
advantage, 1  
causality, 92–115  
definition, 1  
missing data, 201–215  
popularity, 1  
problems, 4
- longitudinal two-part models,  
159–163  
biggest problem, 163  
comments, 162  
lower-bound method, 15–16, 18,  
23–24, 179–180
- lung function, 85, 88
- LVCF, 206–207, 208, 208, 211, 213
- Mann-Whitney *U* tests, 196
- MAR, 201, 203, 204–205, 205,  
207–209, 208, 210, 212,  
213–214, 213
- marginal structural models, 197
- Markov Chain Monte Carlo  
(MCMC), 207
- Markov models, 194  
same as ‘autoregressive models’  
(*qv*), 88
- Mauchly’s Test of Sphericity, 15–16,  
22
- maximum likelihood, 51, 209,  
226–227, 229  
definition, 35
- maximum likelihood estimation  
method, 41–42
- MCAR, 201, 203–204, 203, 205, 207,  
208, 209, 210, 213–214, 213
- McNemar test, 117–118, 134  
change over time in dichotomous  
outcome variable, 116  
disadvantage, 116  
example, 118  
limitation, 118
- mean or median of series method,  
205, 207
- mean squares, 12, 15, 27
- mean value, 80
- mean variance adaptive Gauss–  
Hermite quadrature, 125
- mediation analysis, 94, 96, 98–100  
cross-sectional studies, 96  
important statistical tool, 94  
linear mixed model, 101–110  
types, 96
- medical ethics committees,  
216
- medical science, 71, 82, 152

- groups with same growth trajectory, 75
- medical studies, 110, 152–153, 163, 165, 209, 213
- control studies versus cohort studies, 1
- designs (schematic illustration), 2
- observational versus intervention, 1
- melegit procedure, 235–236
- memory measurement (older subjects), 4
- missing at random. *See* MAR
- missing completely at random. *See* MCAR
- missing data, 149, 156, 197
- alternative methods, 213
- determinants, 203–204
- examples, 202
- GEE versus mixed model analysis, 214
- generating datasets with, 202, 203
- imputation methods, 205–213
- informative versus non-informative, 201–202
- informative versus non-informative (definitions), 201
- intermittent, 201, 202
- logistic regression analyses, 204
- longitudinal studies, 201–215
- subjects with versus subjects without, 204
- missing data mechanisms, 201, 208
- missing not at random. *See* MNAR
- mixed model analysis, 32–42, 56, 90, 110, 173–174, 182, 184–185, 187, 200
- combination with multiple imputation (missing data), 209
- comparison with GEE, 48–49
- comparison with GEE analysis (longitudinal regression methods), 127–129
- continuous outcome variables (software), 226–235
- count outcome variables, 142, 146–147
- datasets with missing data, 205, 205
- development over time of continuous outcome, 34
- dichotomous outcome variables (software), 235–242
- extension, 54
- extension of GLM for repeated measures, 33
- further reading, 133
- general idea, 33, 36
- idea behind, 33
- imputation of missing data, 207
- linear, 34–35, 41, 50, 54, 59, 71–72, 77–79, 83, 86, 89, 95, 99, 103, 112, 155, 208
- linear time-lag, 94
- logistic, 129, 213
- longitudinal data, 32–34
- longitudinal intervention studies, 168
- longitudinal regression methods, 124–127
- more efficient than multiple imputation, 209
- multinomial logistic, 137–141
- prior multiple imputation ‘not necessary’, 210
- regression coefficient (interpretation), 38–41
- software (further reading), 235
- standard (drawback), 107
- steps, 33
- ML. *See* maximum likelihood
- MNAR, 201, 203, 204–205, 208, 208, 210, 212–214, 213
- model fit statistics, 90, 100
- model of changes, 83–85, 87, 88, 90–91, 130–131
- data structure needed, 86
- models to disentangle between- and within-subjects relationship, 76–91, 130–132
- estimation of within-subjects part, 82–91
- hybrid, 76–82
- outputs, 77–81, 84–87, 89–90
- models to estimate within-subjects part of longitudinal relationship, 82–91
- autoregressive (with example), 88–90
- comments, 90–91
- model of changes (with examples), 87–91
- monotone missing data definition, 210
- Mplus software, 74
- multilevel analysis same as ‘mixed model analysis’ (*qv*), 32, 138
- multiple imputation, 207, 208, 208, 211, 213, 215
- combination with mixed model analysis, 209
- datasets with missing data, 206–207
- illustration of, 206
- ‘potential advantage’ over mixed model analysis, 210
- within-imputation versus between-imputations variance, 207
- multiplication factor, 218–219
- multivariate analysis of covariance (MANCOVA), 178
- multivariate analysis of variance (MANOVA), 10
- multivariate tests, 15, 21
- mvaghermite integration method, 125, 127, 147, 151, 154–155, 159–160, 191
- naive linear regression analysis, 44
- naive logistic regression analysis, 125–126
- negative binomial GEE analysis, 148
- negative binomial mixed model analysis, 151
- negative binomial regression analysis, 143, 151
- count outcome variables, 148–150
- no confounding assumptions, 105
- noise, 206, 209
- non-collapsibility problem, 130–131, 133
- normal distribution, 33, 126, 139, 150, 152, 155, 163, 210, 222
- latent, 153
- used in STATA, 229
- null hypothesis, 11, 16, 22
- number of observations, 35, 46, 90, 100–101, 125
- number of subjects, 35, 46, 67
- object-oriented software, 222
- observational cohort studies types, 1
- observational longitudinal studies, 1–4, 70, 164
- causality, 6, 92–115
- observational studies, 197–198
- odds ratio, 121, 126, 138, 189, 190–191, 193–194
- pools two interpretations, 121
- one-within design, 19, 21, 25, 29
- definition, 10
- one-within, one-between design, 19–21, 25
- OR. *See* odds ratio
- outcome variable of interest, 34, 128, 155, 187, 202
- standard deviation, 217
- outcome variables, 1, 92, 94, 108, 121, 128, 197, 204
- characteristics, 4
- floor or ceiling effects, 6, 152–163
- overall error variance divided into two parts, 36
- overall regression coefficient definition, 77
- overdispersion, 143, 148–149
- pair of regression analyses, 95
- paired *t*-test, 7–10, 18–19, 31
- assumptions, 7

## Index

- paired *t*-test (cont.)  
 cholesterol (difference  $t=1$  and  $t=6$ ), 9  
 limitation, 7  
 non-parametric equivalent, 8–9  
 normality assumption, 7  
 not possible (situation with more than two repeated measurements), 10  
 parametric survival analysis, 113  
 path analyses, 104  
 pattern mixture models, 213  
 period effects, 2–3  
 physical activity, 3, 143–148, 168, 168, 170–171, 173–174, 176  
 longitudinal development with age, 3  
 physical fitness, 94, 100–102, 104, 106–114  
 Pillai's Trace, 16, 22  
 placebo, 164, 172, 175–176, 177–178, 180  
 Poisson distribution, 143–144, 148, 152, 159, 161  
 characteristics, 141, 143  
 skewness to right characteristic, 150  
 Poisson family, 144–145  
 Poisson GEE analysis, 144–145, 146, 148  
 Poisson mixed model analysis, 146–147, 149, 151, 161  
 Poisson regression analysis, 142–143, 147–148, 151–152  
 polynomial contrast, 26  
 polynomial function  
 second order, 56  
 third degree, 58  
 population average method, 128, 129  
 post-hoc procedures, 25–26  
 potential outcome analysis, 104, 197–198  
 propensity score adjustment, 197  
 Proper, K.I., 168  
 proportion of change, 116–119, 134–136  
 standard error, 118  
 prospective cohort studies  
 design, 1  
 pseudo-likelihood method, 242  
*p*-values, 7–9, 13, 15, 25–26, 29, 36–37, 46, 58, 61, 100, 121, 138, 149, 155, 161, 178, 189, 191, 206, 222, 229  
 quadratic development, 20, 23–24, 26, 56, 59–62, 64  
 quadratic function, 18, 58  
 quadratic relationship, 13, 15  
 quadratic term, 56, 58  
 quadrature points, 236, 242  
 quasi-causal relationships, 92, 106  
 quasi-likelihood, 42  
 R software, 6, 220  
 GEE analysis with continuous outcome variable, 221–222, 224  
 GEE analysis with dichotomous outcome variable, 224, 227  
 linear mixed model analysis with only random intercept, 235  
 linear mixed model analysis with random intercept and random slope, 235  
 logistic mixed model analysis with random intercept, 242  
 mixed model analysis with continuous outcome variable, 229–231  
 mixed model analysis with dichotomous outcome variable, 239–240  
 random allocation at baseline, 164–165  
 random coefficient analysis, 33, 36  
 random effects, 80  
 variances and covariances, 139  
 random intercept, 35–37, 39, 41, 48, 54, 58–60, 62–64, 66, 71, 78–79, 112, 125–127, 139–140, 147, 154, 156, 159, 162, 214, 228, 232–233, 235, 237–238, 240–241  
 software syntax, 226  
 random intercept variance, 36, 42, 51, 80, 84, 89, 126, 128, 155–156, 233, 240  
 random slope, 36–37, 39, 49, 56, 58–60, 71, 78, 112, 125, 127, 140, 160, 214, 232, 235  
 definition, 33  
 for deviation score, 80  
 software syntax, 226  
 random slope variance, 42, 126, 233  
 random variances, 79  
 randomized control trial with more than one follow-up measurement  
 continuous outcome variables, 172–186  
 GLM for repeated measures, 178  
 GLM for repeated measures (adjusted for baseline), 178  
 regression-based methods, 178–186  
 simple analysis, 175–177  
 summary statistics, 177–178  
 randomized control trial with one follow-up measurement  
 continuous outcome variables, 165–172  
 randomized control trials, 1–2, 128, 164  
 stepped wedge designs, 196  
 rank numbers  
 mean rank, 9  
 RCT. *See* randomized control trials  
 real life, 9, 45, 54, 105, 110, 161, 165, 172, 197, 203, 213  
 reference category, 60, 138, 185  
 regression based methods, 59, 204  
 advantage, 178  
 comparing groups, 61  
 regression coefficient, 31–33, 36, 46, 50–51, 56, 58, 60, 67, 69–70, 76, 80–84, 88, 92–93, 121, 163, 191, 208  
 calculation, 36  
 interpretation, 67  
 interpretation (mixed model analysis), 38–41  
 one for each covariate preferable, 153  
 pooled longitudinal, 86  
 random variation, 48  
 range (after multiple imputation), 211  
 standardized, 88  
 varying between subjects, 48  
 regression model  
 fixed versus random parts, 79  
 regression to mean, 6, 165–168, 166, 172, 177, 182, 184, 186, 196–197  
 regression-based methods, 196  
 adjustment for covariance, 49–55  
 categorical outcome variables (with example), 136–141  
 comments, 41–42  
 continuous outcome variables, 31–55  
 dichotomous outcome variables, 189–193  
 GEE, 42–46  
 longitudinal, 31–32  
 mixed model analysis (mixed model analysis), 32–42  
 mixed model analysis versus GEE, 48–49  
 outputs, 35, 37, 41, 44–47, 49, 51–53  
 RCT with more than one follow-up measurement, 178–186  
 relative change, 166, 167  
 relative risk, 189, 192  
 definition, 188  
 repeated contrast, 26, 28–29  
 repeated measures method, 186–187  
 non-parametric equivalent of GLM, 29–30  
 research grants, 216  
 research question, 69, 75, 90, 152, 154, 177, 185

- as basis for analysis, 4  
 residual change  
   general idea, 167  
 residual pseudo-likelihood method, 236, 240  
 residual variance, 36, 79–80, 155–156, 227, 232  
 restricted maximum likelihood, 35, 41–42, 50, 229  
   default estimation method (software packages), 226  
 Roy's Largest Root, 16, 22
- sample size, 7, 9, 11, 19, 164  
 sample size calculations, 216–219  
   example, 218  
   information needed, 218–219  
 SAS, 6, 220  
   GEE analysis with continuous outcome variable, 220–221, 224  
   GEE analysis with dichotomous outcome variable, 224, 227  
   linear mixed model analysis with only random intercept, 235  
   linear mixed model analysis with random intercept and random slope, 235  
   logistic mixed model analysis with random intercept, 242  
   mixed procedure, 227–229  
   mixed model analysis with continuous outcome variable, 227–229  
   mixed model analysis with dichotomous outcome variable, 236–237  
 scale parameter  
   purpose, 45  
 SD. *See* standard deviation  
 selection models  
   general idea, 213  
 semi-parametric survival analysis, 111  
 serum cholesterol, 4, 143, 168–169, 168, 171, 173–174, 176  
 sex, 5, 5, 21–22, 24, 61–63, 82, 120, 138, 150–151, 154–156, 203, 204  
 signed rank sum test, 8–9  
 significance level, 8, 21, 56, 63, 118, 216–218  
 simple contrast, 26–27, 29  
 single imputation methods, 206  
 smoking, 85, 87, 88, 152, 163  
 social sciences, 32, 71  
 software, 5, 8–9, 35  
 software for longitudinal data analysis, 220–242  
   GEE with continuous outcome variable, 220–223  
   GEE with dichotomous outcome variables, 224–226  
   mixed model analysis with continuous outcome variables, 226–235  
   mixed model analysis with dichotomous outcome variables, 235–242  
   outputs, 221–223, 225–228, 230–234, 237–241  
   software packages, 6, 8–9, 11, 26, 34, 42, 121, 126, 137, 178  
   sphericity assumption, 11, 15–16, 18–19, 21, 23, 179–180  
   definition, 19  
   violation, 19  
 sphericity coefficient (epsilon), 11, 15  
 SPIRIT study, 160  
 SPSS, 6, 178, 220  
   GEE analysis with continuous outcome variable, 222–223, 224  
   GEE analysis with dichotomous outcome variable, 223–224, 227  
   linear mixed model analysis with only random intercept, 235  
   linear mixed model analysis with random intercept and random slope, 235  
   logistic mixed model analysis with random intercept, 242  
   mixed model analysis with continuous outcome variable, 231–235  
   mixed model analysis with dichotomous outcome variable, 240–242  
   squared differences, 11  
   standard deviation, 8, 46, 49, 115  
   artificial decrease, 206  
   residual (or error), 229  
 standard errors, 36, 38, 56, 67, 80, 90, 124, 161, 199  
   difference in proportions (calculation), 116  
   distribution (after multiple imputation), 212  
   estimation (model-based procedure), 220  
   GEE versus mixed model analysis (mixed model analysis), 49  
   model-based method, 224  
   robust, 46, 49, 85, 220, 222, 224  
 standard longitudinal regression, 82, 85  
 standard model, 88, 90–93, 93, 94, 96  
 STATA, 5, 35, 79, 138, 159, 220, 235  
   default estimation method (ML), 226  
   GEE analysis with continuous outcome variable, 220, 224  
   GEE analysis with dichotomous outcome variable, 224, 227  
   linear mixed model analysis with only random intercept, 235  
   linear mixed model analysis with random intercept and random slope, 235  
   logistic mixed model analysis with random intercept, 242  
   mixed model analysis with continuous outcome variable, 226  
   mixed model analysis with dichotomous outcome variable, 235–236  
   use of *z*-distribution, 229  
 stepped wedge designs, 195–196  
   definition, 195  
   schematic illustration, 196  
   usually longitudinal, 196  
 structural equation model, 74, 104  
   general idea, 71  
   typical illustration, 74  
 Stuart–Maxwell test, 134–135  
 study design, 216  
 subject specific method, 128, 129  
 subject variance, 128  
 sum of skinfolds, 5, 35–37, 40–41, 45–47, 48, 49, 51–55, 64, 70, 72–73, 76–77, 79–81, 83–84, 87, 90, 94–95, 121, 131, 139, 220 *passim* longitudinal regression methods, 34  
 sum of squares, 11–15, 18, 21, 27, 33  
 summary statistics  
   examples (intervention studies), 177  
   general idea, 177  
 survival analysis, 108, 110–111, 114–115, 196  
 syntax  
   R, 222, 224, 229, 231, 239  
   SAS, 220, 224, 229, 236  
   SPSS, 223–224, 232, 235, 240, 242  
   STATA, 220, 226, 235–236  
*t*-distribution  
   R, 229  
   SAS, 229  
 temporality, 1, 92  
 tertiles, 81  
 test effects, 4  
 test statistic, 7–8, 10, 67  
 time, 2, 9–10, 26, 149, 151, 154–156, 169–170, 178  
   relationship (shape) between outcome variable and, 13–15  
 time at risk, 194–195, 195  
   definition, 192  
 time effect, 13, 15, 18, 20–21, 25, 178  
 time modelling, 56–75  
   adjustment for time, 64–69

**Index**

- time modelling (cont.)  
 adjustment for time (longitudinal data analysis), 70  
 comparing groups, 60–63  
 growth curve analysis, 56–60  
 interaction with time, 69–70  
 outputs, 57–60, 62–66, 68–69, 71–74  
 subjects with different growth trajectories (classification), 70–75  
 time versus age, 68  
 time of measurement effect, 3  
 time to event method, 192  
 time variable, 56, 58, 61, 64, 67–68, 82, 183, 196, 199  
 time-lag models, 93, 96, 99, 150  
 causality in observational longitudinal studies, 92–94  
 data structure needed, 92  
 occasion when useful, 93  
 time-lag relationship, 95–98, 101  
 time-varying confounders, 107, 109  
 tobit mixed model analysis, 152–159  
 general idea, 153
- Toeplitz 5 covariance structure, 53  
 total change over time, 116–118, 134–135  
 total time method, 192  
 treatment effect, 178–179, 182, 184, 192  
 two-part regression models, 152
- univariate approach, 11, 21  
 longitudinal dataset (four measurements, six subjects), 12  
 numerical example, 12–13  
 validity, 11  
 upper or lower censoring  
 same as ‘floor or ceiling effects’ ( $qv$ ), 152
- variance parameters, 50–51, 229  
 vector of covariate over time, 38  
 vector of outcome over time, 38
- Wald test, 125, 146, 148, 150, 222  
 Wang, N., 209  
 Weibull baseline hazard function, 113
- Weibull distribution, 113  
 Wilcoxon signed rank sum test, 8, 10  
 Wilks’ Lambda, 16, 22  
 within-subjects and between-subjects relationship  
 disentanglement (models), 76–91  
 within-subjects correlation, 43–44, 54–55  
 within-subjects effects, 16, 18, 23–24, 179–180  
 within-subjects error sums of squares, 20  
 within-subjects factor, 19, 21, 25  
 within-subjects interpretation, 126  
 within-subjects relationship, 26, 38, 41, 67, 126, 141  
 within-time sum of squares, 13
- xtreg* procedure, 79–81
- zero-inflated Poisson regression, 152  
 z-statistic, 36, 222  
 calculation, 36, 46  
 z-value, 9, 61, 67, 121, 138, 149, 161