

GLOSSARY AND TERMINOLOGY

There is considerable variation in terminology and methods of presenting data among epidemiology texts and other information sources. In general, the terminology and data layouts used in this book will conform to those used in *Modern Epidemiology*, 3rd edition (Rothman and Greenland, 2008).

GT. 1 DATA LAYOUT

The outcome variable is listed in the rows of the table, the predictor variable is listed in the columns.

Risk calculations (2X2 table)

	Exposure		
	Exposed	Non-exposed	
Diseased	a_1	a_0	m_1
Non-diseased	b_1	b_0	m_0
	n_1	n_0	n

where:

a_1	=	the number of subjects that have both the disease and the risk factor.
a_0	=	the number of subjects that have the disease but not the risk factor.
b_1	=	the number of subjects that have the risk factor but do not have the disease.
b_0	=	the number of subjects that have neither the disease nor the risk factor.
m_1	=	the number of diseased subjects.
m_0	=	the number of non-diseased subjects.
n_1	=	the number of exposed subjects.
n_0	=	the number of non-exposed subjects.
n	=	the number of study subjects.

In general, no distinction is made between values derived from a sample and population values as it is usually easy to determine what is being referred to from the context. In select situations where the distinction is necessary, upper-case letters (*eg* A_1) will be used for population values and lower case (*eg* a_1) for sample values.

Rate calculations (2X2 table)

Here, subject-time replaces the number of non-diseased.

	Exposure		
	Exposed	Non-exposed	
Number of cases	a_1	a_0	m_1
Animal-time at risk	t_1	t_0	t

where:

a_1	=	the number of cases of disease in the exposed group.
a_0	=	the number of cases of disease in the non-exposed group.
t_1	=	the animal-time accumulated in the exposed group.
t_0	=	the animal-time accumulated in the non-exposed group.
t	=	the total animal-time accumulated by the study subjects.

Diagnostic tests (2X2 tables)
Gold standard layout

	Test result		
	Positive	Negative	
Disease positive	a	b	m ₁
Disease negative	c	d	m ₀
	n ₁	n ₀	n

Note The marginals are the same as for risk calculations; the inner cell values are denoted as a, b, c, d.

Test comparison layout

	Test 2 positive	Test 2 negative	Total
Test 1 positive	n ₁₁	n ₁₂	n _{1.}
Test 1 negative	n ₂₁	n ₂₂	n _{2.}
Total	n _{.1}	n _{.2}	n

Correlated data
Matched-pair case-control data layout

		Control pair		Case totals
		Exposed	Non-exposed	
Case pair	Exposed	t	u	t+u = a ₁
	Non-exposed	v	w	v+w = a ₀
	Control totals	t+v = b ₁	u+w = b ₀	

Note If pair-matching is used in a cohort study, the same format is used but the case (rows)-control(columns) status is replaced by exposed (rows) non-exposed (columns) and the exposure status is replaced by disease status.

Significant digits

Throughout the text, data are often presented with more significant digits than normally would be warranted. This is done for clarity and to avoid rounding errors.

GT. 2 MULTIVARIABLE MODELS

In general, multivariable models will be presented as follows, with explicit subscripting (*eg* for observation number) used only if absolutely necessary for clarity:

$$\text{outcome} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

where the outcome may be a variety of parameters (*eg* for logistic regression outcome = $\ln(p/1-p)$) and k is the number of parameters in the model (excluding the intercept).

In some situations, βX or μ will be used to represent the entire right-hand side of the model (*ie* the linear predictor) to simplify presentation:

$$\beta X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

The terms predictor, exposure, risk factor and independent variable will all be used to designate factors that ‘cause’ the outcome of interest, although in general we prefer to use one of the first

two terms. These will be designated X .

The terms outcome and dependent variable will both be used for the response, but the former term is used most commonly. These will be designated Y .

GT. 3 MULTILEVEL MODELS

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + u_{fam(i)} + \varepsilon_i$$

Note For the sake of simplicity, a single index notation will be used for all multilevel data. The subscript i denotes the individual (lowest level) observation. In the example above, $u_{fam(i)}$ refers to the family containing the i^{th} individual. If there are 40 families in the study, u could have one of 40 values. An alternative notation, used in some texts, has multiple indices such as $u_j + i_j$ where j refers to the family and i to the i^{th} individual in the j^{th} family. We will use this notation for repeated measures data where Y_{ij} = measurement for subject i at time j .

GT. 4 GLOSSARY

Terms related to formulae and methods

a	number of cases
ACF	autocorrelation function
AF_e	attributable fraction in the exposed group
AF_p	attributable fraction in the population
AFT	accelerated failure time
AIC	Akaike's Information Criteria
ALR	alternating logistic regression
ANOVA	analysis of variance
AP	apparent prevalence
AR	autoregressive
ARMA	autoregressive moving average
AUC	area under ROC curve
BIC	Bayesian Information Criteria (Schwartz Bayesian Criteria)
BLUP	best linear unbiased predictor
BUGS	Bayesian analysis using Gibbs sampling
c	constant (eg baseline hazard)
c	cost of sampling
c	Geary's c for correlation between values at pairs of spatial points
c	rate of contacts an animal makes with other animals in one time period
CAR	conditional autoregressive
CCC	concordance correlation coefficient
χ^2	chi-square (χ^2)
CI	confidence interval
corr (Y)	correlation matrix of Y
cov (Y)	covariance matrix of Y
covar	covariance

covar (+)/covar(-)	covariance in test positive (+)/negative (-) sample results
cp	cutpoint
C_p	Mallow's statistic
cp	rate of 'effective' contacts
CRS	composite reference standard
CV	coefficient of variation
D	deviance statistic ($-2 \cdot \ln L$)
D	minimum number diseased
D	duration
D	disease
d	duration = duration of the infectious period
d	standardised difference in propensity scores
D-	subjects not having a specified disease/condition
D'	classified (not necessarily correct) disease state
D(h)	difference function (between K-functions for cases and controls)
D+	subjects having a specified disease/condition
DAC	directed acyclic graph (aka cusal diagram)
DB	delta-beta
deff	design effect
df	degrees of freedom
DFITS	difference in fit statistic
d_i	Cohen's d for study i (meta-analysis)
DIC	deviance information criterion
$d_{ii'}$	distance between point i and i'
d_j	outcome events (failures) during the interval (actuarial life table) or number of events at time t_j (K-M life table)
DOR	diagnostic odds ratio
e	2.71828 (natural number)
E	expected value (eg $E(Y)$ = expected value of Y)
E	exposure factor
E-	subjects not exposed
E+	exposed subjects
ES	effect size
ESS	effective sample size
EV	extraneous variable
exp	expected cell number
exp	exponential function (ie $\exp(x) = e^x$)
f	proportion of population vaccinated
$F(t)$	failure function
$f(t)$	probability density function
$f(\theta)$	prior distribution for θ (Bayesian analysis)
$f(\theta Y)$	posterior distribution for θ (Bayesian analysis)
FNF	false negative fraction

FP	fractional polynomial
<i>FPC</i>	finite population correction
<i>FPF</i>	false positive fraction
G^2	likelihood ratio statistic
GEE	generalised estimating equations
g_i	Hedge's adjusted g for study <i>i</i> (meta-analysis)
GLM	generalised linear model
GLMM	generalised linear mixed model
GWR	geographically weighted regression
$h(t)$	hazard function
$H(t)$	cumulative hazard function
$h_0(t)$	baseline hazard function
h_i	leverage
H_j	distribution of host factor and/or subject time in stratum <i>j</i>
<i>HR</i>	hazard ratio
<i>Hs</i>	standard population distribution of host factor
<i>HSe</i>	herd sensitivity
<i>HSp</i>	herd specificity
<i>i</i>	observation counter
<i>I</i>	incidence rate
<i>I</i>	Moran's I (spatial autocorrelation coefficient)
<i>i</i>	incidence = rate at which new infections are occurring in the population - this is the population incidence rate (designated <i>i</i> to differentiate it from <i>I</i>)
IC	information criteria
<i>ICC</i>	intra-class correlation coefficient
<i>ID</i>	incidence rate difference
<i>ID_G</i>	incidence rate difference based on group means
<i>I_{dir}</i>	directly standardised rate
<i>I_e</i>	expected incidence rate
<i>I_h</i>	indicator for K-function
IIA	independence of irrelevant alternatives
<i>I_{ind}</i>	indirectly standardised rate
IPTW	inverse probability of treatment weighted
IQR	interquartile range
<i>IR</i>	incidence rate ratio
<i>IR_G</i>	incidence ratio based on group-level data
<i>Is</i>	standard population incidence rates
<i>j</i>	designated for strata
<i>j</i>	designator for categories
<i>j</i>	designator for covariate patterns in a dataset
<i>j</i>	designator for time intervals (actuarial life table) or time points (KM life table)
<i>j</i>	sampling interval in systematic random sample
<i>J</i>	total number of <i>j</i>

k	cutpoint for herd-level testing (number of positives required for positive herd classification)
k	number of predictors in a model (not including intercept)
k	number of spatial clusters or groups
k	number of studies in a meta-analysis
$K(h)$	K-function for spatial density of events per distance h
$K(ht)$	bivariate space-time K-function
KM	Kaplan-Meier (life table or survival model)
L	allowable error (one-half the length of a confidence interval)
L	likelihood function (eg $L(Y \theta)$)
L	lag-time in case-crossover studies
L_0	null or baseline likelihood function
LCM	latent class model
L_{full}	likelihood function from full model
LISA	local indicator of spatial association
l_j	subjects at risk of failure at the start of the time interval (actuarial life table)
$\ln$	natural log
$\ln L$	$\ln$ (likelihood function)
$\log$	natural log (also $\ln$)
LR	likelihood ratio
LR_{cat}	likelihood ratio for defined category of result
LR_{cp}	likelihood ratio based on defined cutpoint(s)
L_{red}	likelihood function from reduced (smaller) model
LRT	likelihood ratio test
m	number of matched controls per case
m	number of observations in a covariate pattern
m	number of samples in a pooled sample
m	number of subjects per cluster (group)
MANOVA	multivariate analysis of variance
MAR	missing at random
MAUP	modifiable areal unit problem
MCA	multiple correspondence analysis
MCAR	missing completely at random
MCMC	Markov chain Monte Carlo
MCSE	Monte Carlo standard error
MD_i	mean difference in study i in a meta analysis
ML	maximum likelihood
MM	method of moments
MNAR	missing not at random (also NMAR)
MOR	median odds ratio
MOR_c	cluster median odds ratio
MQL	marginal quasi-likelihood
MSE	mean square error

n	number
n	sample size
N	population size
n'	adjusted sample size
NB-1 etc	negative binomial models – see Chapter 18 for details
o	odds
obs	observed cell number
OD	optical density
OR	odds ratio
OR(ABC)	odds ratio for factor ABC
OR(ABC D)	odds ratio for factor ABC conditional on D
OR_a	odds ratio - adjusted
OR_c	odds ratio - crude
OR_j	stratum-specific odds ratio
OR_{MH}	Mantel-Haenszel adjusted odds ratio
OR_{sf}	odds ratio of sampling fractions
p	probability as in $p(D+ E+)$ or $p(Y=1)$
p	proportion as in $\ln(p/1-p)$
p	shape parameter for Weibull distribution
p	probability of transmission of the infection if one animal is infectious and one is susceptible
p'	classified (not necessarily correct) proportions with exposure or disease
p_j	probability of surviving interval j (actuarial life table) or survival at time t_j (KM life table)
P	P-value
P	prevalence
PA	population average
PACF	partial autocorrelation function
par	population at risk
par	parameter
PAR	population attributable risk
PD	prevalence difference
PE	prediction error
$pl(\lambda)$	profile likelihood function
$PlSe$	pooled-sample sensitivity
$PlSp$	pooled-sample specificity
$PPV-$	positive predictive value of a negative test
PQL	penalised quasi-likelihood
PR	prevalence ratio
PS	propensity score
PSU	primary sampling unit
PV	predictive value
$PV-$	negative predictive value

$PV+$	positive predictive value
q	$1-p$
q_j	risk of event during interval j (actuarial life table) or at time t_j (KM life table)
Q	Cochrane's Q statistic
QIC	quasi-likelihood under the independence model information criterion
r	correlation coefficient (ρ also used)
R	incidence risk
R	spatial region
R_0	R_0 = basic reproductive number = # of new cases that arise from an infectious individual in a completely susceptible population.
r^2	squared correlation (R^2 also used)
R^2	coefficient of determination (r^2 also used)
RCT	randomised controlled trial
RD	risk difference (also know as attributable risk)
RDD	random digit dialling
REML	restricted maximum likelihood
res_p	Pearson residual
r_i	raw residual
r_j	average number of subjects at risk during a time interval (actuarial life table) or at time t_j (KM life table)
ROC	receiver operating characteristics
RR	risk ratio (alternatively known as relative risk)
RR_a, RR_u	adjusted and unadjusted RR (meta-analysis)
Rs	standard population incidence risk
r_{si}	standardised residual
R_t	R_t = effective reproduction number = # of new cases arising from each infectious individual at time t .
r_{ii}	studentised residual
$s = S/N$	proportion of the population that is susceptible Note: in a completely susceptible population $S_0=N$ so $s_0=1$
S, I, R, N	the numbers of susceptible, infectious, removed and total number of animals in the population, respectively
$S(t)$	survivor function
SAR	Simultaneous autoregressive
SD	standard deviation
SE	standard error
Se	sensitivity
Se_{corr}/Sp_{corr}	corrected Se/Sp based on cross-sectional validation
Se_{new}/Sp_{new}	Se/Sp of current test adjusted for Se/Sp of referent test
Se_p/Sp_p	Se/Sp in parallel interpretation of test results
Se_s/Sp_s	Se/Sp in series interpretation of test results
sf	sampling fraction
sf_{T+}/sf_{T-}	sampling fractions for cross-sectional validation
S_i	value of latent variable for individual i

s_i	spatial points
<i>SMR</i>	standardised morbidity/mortality ratio
<i>so</i>	sampling odds
<i>Sp</i>	specificity
<i>sr</i>	sampling risk
<i>SRR</i>	standardized risk ratio
<i>SS</i>	subject specific
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
<i>t</i> or <i>T</i>	animal-time
TCE	true causal effect
t_d	doubling time
T_i	study outcome for study <i>i</i> in meta-analysis
t_j	time of event (KM life table)
t_{j-1}, t_j	time span in the interval (actuarial life table)
Δt	length of period
<i>TP</i>	true prevalence
<i>TR</i>	time ratio
T_s	standard population animal-time at risk
T_{scan}	spatial scan statistic
TVC	time-varying covariate
<i>U</i>	measure of confounding bias (meta-analysis)
u_i	random effect of study <i>i</i>
var	variance
V_i	within study variance for study <i>i</i> in meta-analysis
<i>VIF</i>	variance inflation factor
<i>W</i>	sampling weights based on probability of exposure
w_j	subjects withdrawn during interval (censored observations) (actuarial life table) or censored observations at time t_j (KM life table)
<i>X</i>	predictor variable or design matrix of predictors
<i>Y</i>	outcome variable or vector of outcome values
<i>Z</i>	design matrix for random effects
<i>Z</i>	extraneous variable, factor or confounder
<i>Z</i>	standard normal deviate
Z_α	standard normal percentile for $\alpha/2$ Type I error (for sample size calculations)
Z_β	standard normal percentile for one-tailed β Type II error (for sample size calculations)

Note Acronyms are not italicised in Arial font (tables and figures).

Symbols

*	multiplication symbol
/	division
#	number
~	approximate symbol or distributed as (eg $Y \sim N(0,1)$)

$\approx$	approximately equal to
α	level of significance (Type I error)
β	regression coefficient or vector (1*n) of coefficients
β	Type II error (power=1- β)
β	frailty factor
β_{aft}	coefficient from accelerated failure time model
β_{ph}	coefficient from proportional hazards model
γ	prior (spatial) disease rate
$\gamma(h)$	empirical semi-variogram
δ	spatial edge correction factor
Δ_i	Glass's Δ for study i (meta-analysis)
ε	error (or vector (1*n) of error values
θ	posterior value for local (spatial) disease rate
θ	a specified or assumed value
θ_0	null specified value
λ	kernel density
λ	hazard
λ	rate at which susceptible animals becomes infectious
λ	power transformation
μ	random group effect
μ	mean
π	3.14159 (natural number)
ρ	correlation - intra-class correlation coefficient (r also used)
ρ_{ce}	confounder-exposure correlation
σ	standard deviation
σ^2	variance
σ^2_h	herd variance
σ^2_i	random slope variance for β_1
σ^2_r	regional variance
τ	spatial bandwidth
τ	cutpoint for proportional odds
τ	distribution of survival times
τ^2	between study variance in meta-analysis
φ	dispersion parameter in GLM(M)
φ	variance of prior disease rate
χ^2	chi-square statistic
χ^2_{homo}	χ^2 test for homogeneity
χ^2_{Wald}	Wald chi statistic

Terms related to time, location and specific health problems

AID	autoimmune disease
BC	British Columbia (Canada's most westerly province)
bp	blood pressure
bwt	birth weight
CRD	childhood respiratory disease
d	day(s)
EIA	enzyme immunoassay
ELISA	enzyme-linked immunosorbent assay
HIV	human immunodeficiency virus
HPV	human papilloma virus
IFAT	indirect fluorescent antibody test
mi	myocardial infarction
MMR	measels, mumps and rubella
mo	month(s)
nv	norovirus
Ont.	Ontario (large province in Canada)
PCR	polymerase chain reaction
PEI	Prince Edward Island (smallest province in Canada)
ppb	parts per billion
ppm	parts per million
RSV	respiratory syncytial virus
SARS	severe acute respiratory syndrome
STREP	<i>Streptococcus pneumoniae</i>
VI	virus isolation
yr	year(s)

GT. 5 PROBABILITY NOTATION

$E(Y)$ = expected value of Y

$p(D+)$ = probability of having the disease of interest

$p(T+|D+)$ = probability of being test positive given the animal had the disease of interest

$p(D+|E+)$ = probability of having the disease of interest in an exposed group

$p(D+|T+)$ = probability of having the disease of interest given the animal was test positive

c_k^n = the number of combinations of k items from n items

GT. 6 NAMING VARIABLES

Variable names in the text will be set between pairs of dashes (*eg* -varname-) but the dashes will not be included in tables and figures or if the variable is used in an equation.

Modifications of variables will generally (but not always – you wouldn’t expect us to be totally consistent, would you?) be named by adding a suffix to the original variable name. For example:

varname_ct	centred version of the variable
varname_sq	squared version of the variable
varname_c#	a categorical version of -varname- with n = # categories
varname_ln	log transformed version of the variable

Indicator variables will usually be named by appending the category value (or left-hand end of the category range if it is a continuous variable). For example, a variable representing birth weight (-bwt-) broken into four categories (0-2499, 2500-2999, 3000-3499, 3500+) would result in the following four variables:

- meduc_0-
- meduc_2500-
- meduc_3000-
- meduc_3500-

Note Unless otherwise specified, values falling exactly on the dividing point will fall in the upper category.

INDEX

A

accelerated failure time model 541
 coefficients in aft models 541
 generalised gamma model 544
 log-logistic model 542
 log-normal model 542
 time ratio 542
 acceleration parameter 541
 accuracy 97
 actuarial life tables 507
 adaptive design studies 253
 additive interaction 326
 adjacent-category model 464, 475
 adjusted odds ratio 318
 admission risk bias 284
 aggregate variables 816
 agreement 98
 air/water borne transmission 758
 Akaike's Information Criteria (AIC) 419
 all possible/best subset regressions 420
 ALR 669
 alternating logistic regression 669
 alternative hypothesis 149
 analogy 29
 analysis of spatial data 718
 analytic control 216
 analytic control of confounding 322
 analytic sensitivity 97
 analytic specificity 97
 analytic study 36, 50, 157, 815
 Anderson-Gill model 552
 ANOVA table 362
 Anscombe residuals 485
 apparent prevalence 107
 ar(1) 657
 area or polygon data 708
 arma(1,1) 658
 artifactual heterogeneity 791
 assumptions in logistic regression 433
 atomistic fallacy 819, 828
 attack rates 87
 attributable fraction (exposed) 144
 attributable risk 144
 autocorrelation 646
 autocorrelation function (ACF) 688

automated selection procedures 422
 autoregressive 657

B

backward elimination 422
 backward stepwise 422
 Bacon, Francis 7
 basic reproductive number (R_0) 762
 Bayes, Thomas 8
 Bayes' theorem 677
 Bayesian analysis 676
 burn-in period 682
 choice of prior distributions 678
 conjugate priors 678
 Gibbs sampling 682
 homogeneous chains 680
 improper prior 680
 latent class models for diagnostic test evaluation 694
 Markov chain Monte Carlo 681
 Markov chains 680
 measurement errors and imperfect tests 693
 Metropolis-Hastings sampling 682
 missing data 692
 statistical analysis based on MCMC estimation 685
 Bayesian Information Criteria (BIC) 419
 Bayesian paradigm 677
 Bayesian spatial regression model 745
 Begg's test 799
 Berkson's fallacy 284
 best linear unbiased predictors (BLUPs) 605
 beta-binomial model 628
 bias variables 278
 binary data 616
 bivariate space-time K-function 740
 Bland-Altman plot 100
 blinding 192, 256
 BLUPs 605
 Bonferroni adjustment 261, 651
 boundaries 720
 Box-Cox transformation 386, 606
 Brant (Wald) test 474
 Breslow method 523
 Brooks-Draper 688

burn-in 682

C

caliper-matching 315

cartogram 710

cartography 706

case fatality rates 87

case-case studies 228

basis 228

design issues 228

case-case-control studies 229

case-cohort studies 233

analysis 234

basis 233

design issues 233

case-control study 202

admission 212

analytic control 216

case series 205

closed source population 204

exclusion and inclusion criteria 216

matching 216

neighbourhood controls 214

nested 202

number of control groups 215

open population 204

primary study base 202

principles of control selection 207

random-digit-dialling 214

sampling controls and data layout in rate-based designs 209

sampling controls from a secondary base 212

sampling from a primary-base open population 211

secondary study base 202

selecting controls and data layout in risk-based designs 207

source population 202

study base 202

subject's exposure 214

case-crossover studies 224

analysis 227

design issues 225

case-only studies 235

analysis 236

case-series studies 231

analysis 232

basis 231

design issues 231

casual contact 758

categorising continuous predictors 412

causal complement 15

causal diagram 23, 317, 343, 834

causal model 403

causal relationships

causal diagram 343

distorter variable 349

explanatory antecedent variable—complete

confounding 346

explanatory antecedent variable—

incomplete confounding 347

exposure-independent variable(s) 344

graphical aids 343

intervening variable 348

moderator variable 351

simple antecedent variable 345

spurious relationships 343

summary of effects 351

suppressor variables 350

Venn diagrams 343

causation

causal criteria 25

coherence or plausibility 28

consistency 28

dose-response relationship 27

statistical issues 26

strength of association 27

study design 26

time sequence 27

causes 10

causal complement 15

component-cause model 11

direct causes 17

indirect cause 17

necessary cause 11

proximal causes 17

sufficient cause 11

censoring 504

gaps 506

interval censoring 505

interval truncation 506

left censoring 505

left truncation 506

right censoring 505

truncation 505

- census 36
- centring 375
- chain binomial models 767
- checklist question 65
- cholera 3
- choropleth maps 708
- clinical heterogeneity 791
- clinical trial 244
 - phases of clinical research 245
 - subjects or participants 244
- close contact 758
- closed cohorts 184
- closed population 81, 204
- closed question 65
- cluster randomisation 255
- cluster sampling 41
- cluster-median odds-ratio 620
- cluster-specific 618, 623
- clustered data 564
- clustering - effects 570
 - adjustment by overdispersion factor 580
 - adjustment by the design effect 578
 - clustering and confounding 575
 - clustering for binary outcome 574
 - clustering for continuous data 571
 - clustering for discrete data 571
 - fixed-effects and stratified models 577
 - intra-class correlation coefficient 571, 578
 - Mantel-Haenszel procedure 578
 - overdispersion 578
 - simulation studies on the impact of clustering 574
 - stratified analysis 578
 - variance adjustment factor 570
 - variance inflation as a result of clustering 571
- clustering of predictor variables 568
- Cochran's Q statistic 792
- coding 70, 72
- coefficient of determination 366
- coefficient of variation 98
- Cohen's d 802
- Cohen's kappa 101
- coherence 28
- cohort 180
- cohort study 180
 - analytic control 191
 - blinding 192
 - confounding 190
 - diagnostic criteria 192
 - exclusion (also called restricted sampling) 190
 - exposure status 189
 - exposure threshold 187
 - exposure time 189
 - fixed cohorts 184
 - follow-up period 191
 - longitudinal study 180
 - matching 190
 - measuring the outcome 191
 - non-permanent exposures 188
 - permanent exposures 187
 - rate-based (incidence density) designs 185
 - rate-based cohort analyses 193
 - reporting of cohort studies 194
 - risk-based (cumulative incidence) designs 184
 - risk-based cohort analysis 192
 - single cohort 180
 - STROBE 194
- collinearity 374
- compartmental models 755
- complementary log-log function 625
- complete case analysis 408
- compliance 258
- component-cause model 11
- composite reference standard 119
- compound symmetry 656
- concordance correlation coefficient 98
- conditional association 317
- conditional autoregressive (CAR) 742
- conditional independence 116
- conditional logistic regression 456
- conditional risk sets model 552
- confidence intervals 87, 147, 151
- confounding 308, 439, 575, 818
 - analytic control 311
 - confounders 308
 - exclusion 311
 - exclusion (restricted sampling) 311
 - exposure of interest 308
 - extraneous factors 308
 - intermediate factor 310
 - intervening factor 310
 - matching 311
 - population confounder 310

- standardised risks/rates 329
- conjugate priors 678
- consensus 8
- consistency 28
- CONSORT 244, 265
- constrained cumulative logit model 471
- contact rate 758, 763
- contextual effects 598
- continuation-ratio model 465, 476
- continuous features 705
- continuous spatial fields 741
- controlled field trials 158
- controlled trials
 - allocation of study subjects 254
 - alternatives to randomisation 254
 - analysis
 - multiple comparisons and assessments 261
 - intent-to-treat 259
 - per-protocol 259
 - subgroup analyses 261
 - background, objectives and summary trial design 246
 - clinical trial designs for prophylaxis of communicable organisms 262
 - cluster randomisation 255
 - cross-over studies 255
 - eligibility criteria 249
 - factorial designs 255
 - follow-up/compliance 258
 - masking (blinding) 256
 - measuring the outcome 251
 - multicentre trials 256
 - other sample size issues 254
 - participants: the study group 247
 - random allocation 255
 - reporting of clinical trials 265
 - sample size 252
 - sample size for sequential and adaptive designs 253
 - sample size for the allocation of clusters of subjects 252
 - specifying the intervention 250
 - split-plot designs 256
 - statistical methods and analysis 259
 - unit of concern 248
- convenience sample 39
- convergence criterion 433
- Cook's distance 392
- Cornfield's approximation 152
- correlated data 564
- correlated test results 116
- correlation structure 656
- correspondence analysis 408
- count 78
- count data 616
- counterfactual 18, 323
- covariance pattern model 659
- covariate pattern 441
- Cox proportional hazards model 519
 - baseline hazard 523
 - evaluation
 - assumption of independent censoring 530
 - assumption of proportional hazards 527
 - Cox-Snell residuals 532
 - delta-beta 536
 - deviance residuals 535
 - goodness-of-fit 533
 - graphical assessment 528
 - Harrell's C concordance statistic 534
 - martingale residuals 534
 - overall fit of the model 532
 - r^2 534
 - scaled Schoenfeld residuals 530
 - scaled score residual 536
 - Schoenfeld residuals 530
 - score residuals 536
 - time-varying effects 528
 - fitting the Cox proportional hazards model 522
 - handling of ties 522
 - hazard ratios 519
 - model-building 524
 - stratified analysis 524
 - ties
 - Breslow method 523
 - Efron method 523
 - marginal calculation 523
 - partial calculation 523
 - time-varying covariates 524
 - time-varying effects 526
 - time-varying predictors 524, 525
 - validating the model 527
- Cox-Snell residuals 532

critical percentage 764
 critical proportion susceptible 769
 Cronbach's alpha 406
 cross-classification. 565, 593
 cross-classified and multiple membership models 691
 cross-over studies 255
 cross-sectional studies 164
 assessing exposure 165
 inferential limitations of cross-sectional studies 169
 repeated cross-sectional versus cohort studies 170
 sample-size aspects 166
 source population 164
 study group 165
 target population 164
 cross-validation correlation 424
 crude odds ratio 318
 cumulative hazard 512, 516
 cumulative incidence 80
 cutpoint 109
 Cuzick and Edwards test 729

D

data coding 835
 data editing 838
 data entry 835
 data processing—multilevel data 840
 data processing—outcome variable(s) 839
 data processing—predictor variables 840
 data verification 839
 data-collection sheets 834
 deductive reasoning 7
 deff 47
 delta-beta 395, 453, 536
 delta-deviance 453
 delta- χ^2 453
 derived variable 816
 descriptive studies 36, 156, 161
 detection bias 286
 detection of confounding 316
 causal diagrams 317
 change in measure of association 318
 change-in-estimate 319
 directed acyclic graph 317
 non-collapsibility of odds ratios 319
 statistical identification of confounders 319

deterministic models 761
 deviance 435
 deviance residuals 447, 485, 535
 DFITS 392
 diagnostic criteria 192
 diagnostic odds ratio (DOR) 806
 diagnostic test 96
 sample size 127
 difference function 729
 differential equations 762
 differential misclassification bias 295
 diffusion cartogram 710
 Diggle-Chetwynd test statistic 730
 direct cause 17
 direct effects 321
 direct standardisation 90
 direct transmission 758
 directed acyclic graph 317
 discrepant resolution 119
 discrete features 705
 discrete repeated measures data
 adding correlation structure to a GLMM 663
 GLMMs without explicit correlation structure 665
 transition models 664
 discrete-time survival analysis 552
 complementary-log-log regression 555
 logistic regression 555
 disease frequency
 count 78
 odds 79
 proportion 78
 rate 79
 dispersion 580
 distorter variable 349
 dose-response relationship 27
 dot maps 707
 doubling time 768
 dummy variables 369
 duration 84
 Durbin-Watson test 396

E

ecologic bias
 confounding by group 821
 effect modification (interaction) by group 824

- within-group bias 820
- ecologic fallacy 819
- ecologic studies 814
 - analytic 815
 - confounding by group 820
 - effect modification 820
 - exploratory 814
 - inferences 819
 - modelling approaches 817
- ecological perspective 826
- ecologic variable 816
 - aggregate 816
 - derived variable 816
 - environmental or contextual 817
 - group or global 817
- edge effects 719
- effect modification 327
- effective contact rate 758
- effective reproduction number (R_t) 762
- effective sample size 688
- Efron method 523
- Egger's test 799
- eligibility criteria 249
- empirical Bayes estimate 605
- empirical Bayesian analysis 724
- empirical semi-variogram 741
- environment 758
- environmental or contextual variables 817
- epigenesis 5
- etiologic fraction 145
- evidence
 - experimental evidence 22, 29
 - limits of experimental study evidence 23
 - observational evidence 22
- exact confidence intervals 151
- exact logistic regression 456
- exact probabilities 150
- exchangeable 656
- exclusion (restricted sampling) 190, 311
- exclusion criteria 216, 782
- expected variation in the data 49
- experimental studies 157
- explanatory antecedent variable 346
- explanatory antecedent variable—complete confounding 346
- explanatory antecedent variable—incomplete confounding 347
- explanatory studies 157

- exploratory spatial analysis 720
- exploratory studies 814
- exponential model 517, 537
- exposure 141, 186
- exposure homogeneity 816
- exposure status 189
- exposure threshold 187
- exposure time 189
- exposure variable 360
- exposure-independent variable 344
- external validity 37, 276
- extra-Poisson variation 486
- extraneous variable 360

F

- factor analysis 408
- factorial designs 255
- failure function 515
- failure functions 514
- field trial 244
- finite population correction 47
- first-order neighbours 720
- first-order spatial effects 718
- Fisher's exact P-value 151
- fixed cohorts 184
- fixed effects 589, 618
- fixed-effects model 577, 786
- focus groups 62
- force of infection 762
- forest plot 789
- forward selection 420
- forward stepwise 422
- fractional polynomials 414
- frailty models
 - clustering in survival data 547
 - individual frailty models 545
 - shared frailty models 547
 - Cox model—Poisson regression 549
 - Cox models 548
 - interpretation of coefficients 550
- Framingham Heart Study 4
- frequency-matching 315
- full information maximum likelihood 601
- funnel plot 798

G

Galbraith plot 794
 Geary's contiguity ratio (or Geary's c) 733
 GEE 667
 generalisability 276
 generalised additive mixed models (GAM) 745
 generalised estimating equations 665

- estimating equations 667
- GEE for multilevel data structures 668
- statistical inference using GEE 667

 generalised gamma model 544
 generalised linear mixed model 616, 623

- complementary log-log function 625
- confidence intervals and tests 634
- GLMMs for binary data 625
- GLMMs for categorical data 627
- GLMMs for count data 625
- maximum likelihood estimation 631
- over- and underdispersion in GLMMs 638
- population-averaged versus cluster-specific parameters 623
- prediction 635
- quasi-likelihood estimation 632
- residuals and diagnostics 635
- robustness against model misspecification 638
- statistical analysis of GLMMs 630

 generalised linear models 444

- canonical link 444
- distribution 444
- estimation methods for GLMs 445
- GLM model evaluation 445
- link function 444

 generalised negative binomial models 495
 generation interval 769
 geographic information system 702
 geographically weighted regression (GWR) 745
 Gibbs sampling 682
 Glass's Δ 803
 GLMM 616
 global statistics 727
 global-influence property 413
 gold standard 105

- gold standard populations 118
- gold standard reference test 118
- pseudo-gold standard 119

Gompertz model 540
 goodness-of-fit test 533
 Greenwood model 768
 group-level studies 815
 group-level testing

- apparent prevalence 128
- group sensitivity 128
- group specificity 129

 group or global variables 817
 group-level effects 816
 grouped binary (binomial) data 580

H

Harrell's C concordance statistic 534
 Hausman and McFadden test 469
 Hawthorne effect 285
 hazard 508
 hazard function 515

- constant hazard 517
- Cox regression model 519
- gamma, log-normal and log-logistic hazards 518
- proportional hazards model 519
- Weibull hazard 517

 Hedges' adjusted g 803
 herd immunity 263, 764
 heterogeneity 791
 heterogeneous mixing 760
 hierarchical data 564
 hierarchical indicator variables 369, 372
 Hippocrates 3
 historical control trials 254
 history

- multiple causation 2
- scientific inference 6

 homogeneous correlation 658
 homogeneous mixing 759
 homoscedasticity 382
 horizontal transmission 758
 Hume, David 7
 hurdle models 496
 hybrid study designs 223
 hyperpriors 724
 hypothesis testing 149

I

ICC 98, 100, 571, 578, 591, 620, 659

- ideal experiment 308
 - immunisation 763
 - imputation 410
 - incidence 79
 - calculation of risk and rate 86
 - cumulative incidence 80
 - incidence count 80
 - incidence density 80
 - incidence proportion 80
 - incidence rate 80
 - incidence risk 80
 - incident times 79
 - risk 80
 - incidence density 80
 - incidence density sampling 211
 - incidence rate 80
 - approximate calculation 82
 - exact calculation 82
 - person-time unit 82
 - incidence rate difference 144
 - incidence rate ratio 142, 483
 - incidence rates 762
 - incidence risk 80
 - incident times 79
 - inclusion criteria 216, 782
 - incubation period 757
 - independence 433
 - independence of irrelevant alternatives 469
 - independent censoring 530
 - indicator variables 369
 - indices 405
 - indirect cause 17
 - indirect standardisation 89
 - indirect transmission 758
 - individual frailty 545
 - induction period 187
 - inductive reasoning 6
 - infectious disease 754
 - infectious disease transmission
 - air/water borne transmission 758
 - casual contact 758
 - close contact 758
 - direct transmission 758
 - environment 758
 - horizontal transmission 758
 - indirect transmissio 758
 - sexual transmission 758
 - vector borne 758
 - vertical transmission 758
 - infectious period 756
 - influential observations 392
 - influential studies 801
 - information bias 288
 - differential misclassification of exposure or outcome 295
 - misclassification of both exposure and disease 295
 - misclassification of extraneous variables 296
 - non-differential misclassification of disease in case-control studies 294
 - non-differential misclassification of disease in cohort studies 294
 - non-differential misclassification of exposure 290
 - recall bias 295
 - reducing misclassification errors 295
 - validation studies to correct misclassification 297
 - information criteria 419
 - instrumental variables to control confounding 334
 - intensity models 755
 - intent-to-treat 259
 - interaction 376, 418, 439, 818
 - interaction 324
 - additive scale 326
 - effect modification 327
 - multiplicative scale 326
 - intercept 360
 - interference 262
 - intermediate variable 310
 - internal validity 37, 276
 - interquartile ranges 426
 - intervening variable 310, 348
 - intervention 244
 - interview 63
 - intra-class correlation coefficient 98, 100, 571, 578, 591, 620, 659
- J**
- Jacquez k-nearest neighbour test 740
 - Jenner, Edward 7
 - judgement sample 39

K

K-function 729
 Kaplan-Meier function and estimator 511
 kappa 101
 prevalence and bias adjusted kappa 102
 weighted kappa 103
 kernel density plot 688
 kernel smoothing 723
 Knox test 739
 Koch, Robert 4
 kriging 742
 Kuhn, Thomas 8
 kurtosis 385

L

L'Abbé plot 797
 latent class models 121
 Bayesian estimation 122
 clustering of test results 126
 expectation-maximisation 122
 goodness-of-fit 124
 group-level testing 128
 maximum likelihood 122
 latent period 756
 latent response variables 620
 Lawson-Waller local score test 738
 leave-one-out analysis 424
 leverage 392, 453
 likelihood 432
 likelihood ratio 112
 likelihood ratios
 category-specific LR 114
 cutpoint-specific LR 113
 likelihood ratio test (*LRT*) 434, 151
 limits of agreement plot 100
 linear regression
 ANOVA table 362
 assumptions
 homoscedasticity 382
 independence 379
 linearity 383
 normal distribution 383
 Box-Cox transformations 386
 Breusch-Pagan test 385
 causal interpretation 377
 Cook-Weisberg test 385
 Cook's distance 392
 deletion residuals 384
 delta-beta 395
 DFITS 392
 errors in the x-variables 373
 estimates and intervals for prediction 366
 evaluating the least squares model 379
 F-statistic 363
 homoscedasticity 385
 influential observations 392
 interaction 376
 interpreting transformed models 389
 jackknife residuals 384
 leverage 384, 392
 linearity of predictor-outcome association 387
 mean square 363
 mean square error 363
 measurement error models 373
 multivariable models 361
 normality of residuals 385
 outliers 391
 R^2 366
 raw residual 383
 regression coefficient 360
 residuals 383
 root mse 363
 scaling variables 369
 significance of a regression coefficient 364
 significance of groups of predictor variables 367
 standard error of prediction 363
 transformations 386
 X-variables 368
 linearity 383, 433
 linear mixed model 588
 best linear unbiased predictors (BLUPs) 605
 Box-Cox transformation for linear mixed models 606
 empirical Bayes estimates 605
 fixed versus random effects 609
 inference for fixed part of model 602
 inference for random part of model 603
 likelihood-based analysis 601
 prediction 605
 residuals and diagnostics 605
 robustness against model misspecification 609

- sample size 610
 - statistical analysis of linear mixed models 601
 - link function 444
 - literature review 782
 - Ljung-Box Q-test 396
 - local indicators of spatial association or LISAs 736
 - local Moran test 737
 - local polynomial smoother 412
 - local statistics 727
 - local-influence property 411
 - log likelihood 432
 - log odds 431
 - log relative risk 723
 - log-cumulative hazard plot 528
 - log-linear models 462
 - log-logistic model 542
 - log-normal model 542
 - log-rank test 513
 - logistic model 431
 - logistic regression
 - apparent overdispersion 449
 - assumptions
 - independence 433
 - linearity 433
 - assumptions in logistic regression 433
 - categorical predictor 438
 - confounding 439
 - continuous predictor 437
 - covariate patterns 446
 - delta-betas 453
 - dichotomous predictor 436
 - evaluating logistic regression models 445
 - goodness-of-fit tests 447
 - hat matrix and leverage 453
 - Hosmer-Lemeshow goodness-of-fit test 447
 - interaction 439
 - interpretation of coefficients 436
 - interpretation of the intercept 438
 - model-building 441
 - outliers 452
 - overdispersion 449
 - Pearson and deviance residuals 446
 - Pearson χ^2 statistic 447
 - predictive ability of model 450
 - receiver operating characteristic curves 451
 - sensitivity and specificity 450
 - presenting effects of factors on the probability scale 439
 - R^2 450
 - real overdispersion 450
 - sample size 455
 - logistic regression with random effects 617
 - logit 431
 - long data format 647
 - longitudinal data 260
 - longitudinal study 180, 646
 - loss to follow-up bias 285
 - lowess smoother 411
- ## M
- macroparasitic 755
 - main effects 376
 - Mallows' Cp 420
 - MANOVA 654
 - Mantel test 739
 - Mantel-Haenszel (MH) procedure 322, 578
 - Mantel-Haenszel odds ratio 323
 - Mantel-Haenszel χ^2 150
 - MAR 410
 - marginal calculation 523
 - marginal structural models 331
 - marginalized models 639
 - marginally independent 317
 - Markov chain Monte Carlo 676
 - martingale residuals 534
 - masking (blinding) 256
 - matching 190, 216, 311
 - analysis
 - frequency-matched data 315
 - Mantel-Haenszel matched OR 316
 - Mantel-Haenszel χ^2 test 316
 - McNemar's χ^2 316
 - pair-matched data 316
 - blocking 312
 - caliper-matching 315
 - frequency- and pair-matching 315
 - general guidelines for matching 313
 - matching on propensity scores 337
 - overmatching 314
 - propensity scores 335
 - maximum likelihood 601

- maximum likelihood estimation 432, 631
- maximum model 403
- MCAR 408
- MCMC 676
- McNemar's χ^2 103
- mean difference 788
- mean plot 647
- measurement error 288, 297
- measures of association 141
 - attributable fraction (exposed) 144
 - attributable risk 144
 - confidence intervals 147
 - etiologic fraction 145
 - hypothesis testing 147
 - incidence rate difference 144
 - measures of effect 144
 - presentation of incidence rate data 140
 - presentation of incidence risk data 140
 - relationships among RR, IR, and OR 143
 - risk difference 144
 - significance (hypothesis) testing 149
 - standard error 149
 - strength of an association 141
 - study design and measures of association 147
 - vaccine efficacy 145
- measures of effect 144
- measures of effect in the population 146
- meta-analysis 785
 - fixed-effects model 786, 787
 - forest plot 790
 - heterogeneity
 - evaluation 792
 - graphical assessment 793
 - meta-regression 794
 - stratified analysis 793
 - subgroup analysis 792
 - underlying risk 797
 - heterogeneity 791
 - imputing 2x2 table cell frequencies 804
 - imputing missing variance estimates 803
 - influential studies 801
 - inverse variance weighting 787
 - Mantel-Haenszel 787
 - mean difference 788
 - outcome scales 801
 - Peto 788
 - process 786
 - random-effects model 786
 - random-effects model 788
 - sparse data 804
 - standardised mean difference 788
 - summary estimate of effect 787
 - types of data 785
- meta-analysis of diagnostic tests 806
- meta-analysis of observational studies 804
- meta-regression 794
- Metropolis-Hastings sampling 682
- microparasitic 755
- Mill, John Stuart 7
- misclassification bias 288
- missing at random (MAR) 410
- missing completely at random (MCAR) 408
- missing data bias 286
- missing not at random (MNAR, NMAR) 410
- missing values 408
- mixed models 588
- mixed models for discrete repeated measures data 662
- MNAR 410
- model-building
 - causal model 403
 - cautions in using any automated selection procedures 422
 - correlation analysis 405
 - creation of indices 405
 - cross-validation correlation 424
 - goals of the analysis 402
 - non-statistical considerations 418
 - number of predictors 404
 - P-values and automated selection procedures 423
 - parsimony vs fit 402
 - reliability 424
 - role of subject matter knowledge 402
 - screening predictors 405
 - screening variables based on unconditional associations 406
 - shrinkage on cross-validation 424
 - specifying the maximum model 403
 - split-sample analysis 424
 - statistical considerations—non-nested models 419
 - statistical criteria—nested models 419
 - validity 423

- model misspecification 638
- models of causation 11
- moderator variable 351
- modifiable area unit problem 708
- Monte Carlo simulation 727
- Monte Carlo standard error 688
- Moran scatterplot 738
- Moran's I 731
- morbidity 78
- mortality 78
- mortality rate 85
- mortality statistics 85
- multicentre trials 256
- multinomial logistic model 463, 466
 - independence of irrelevant alternatives (IIA) 469
 - interpretation of coefficients 467
 - models for outcomes with alternative specific data 470
 - obtaining predicted probabilities 469
 - regression diagnostics 470
 - testing significance of predictors 468
- multiple comparisons 261
- multiple comparisons and assessments 261
- multiple membership 565
- multiple outcome event data 551
 - Anderson-Gill model 552
 - Prentice-William-Peterson model 552
- multiple tests 115
- multiple-choice question 65
- multiplicative interaction 326
- multistage sampling 42
- multivariable modelling to control confounding 328
- multivariate 361
- multivariate analysis of variance 654

N

- narrative reviews 780
- necessary cause 11
- negative binomial distribution 488
- negative binomial regression 488
 - alternative variance functions 491
 - evaluating overdispersion 493
 - generalised negative binomial models 495
 - negative binomial regression diagnostics 493
 - negative binomial regression modelling

- 491
- Poisson-gamma mixture distribution 490
- zero-inflated models 496
- neighbourhood controls 214
- Nelson-Aalen estimate of cumulative hazard 512
- nested case-control study 202
- nested models 434
- networks 772
- non-collapsibility of odds ratios 319
- non-differential misclassification bias 290
- non-informative prior 678
- non-parametric analysis 506
 - confidence intervals 512
 - log-rank test 513
 - tests of the overall survival curve 513
 - Wilcoxon test 513
- 'point-in-time' comparisons 512
- non-parametric kernel density 722
- non-permanent exposures 188
- non-probability sampling 39
- non-response bias 282
- normal probability plot 385
- normality 383
- null hypothesis 38, 149

O

- observational evidence 22
- observational studies 157, 162
- odds 79
- odds ratio 142, 431, 466
- offset 483
- one-tailed or 2-tailed 149
- open population 81, 204
- open question 65
- ordinal data 462
- orthogonal polynomials 414
- outcome scales 801
- outcome variable 360
- outliers 391
- overdispersion 486, 578, 638
 - apparent overdispersion 486
 - dealing with overdispersion 487
 - evaluating overdispersion 487
 - real overdispersion 487

P

- P-value 150
- PACF 688
- pair-matching 315
- parallel interpretation 115
- parametric analysis 507
- parametric survival models 537
 - exponential model 537
 - Gompertz model 540
- Weibull model 538
- parametric models 536
- parsimony 402
- partial autocorrelation functions (PACF) 688
- partial calculation 523
- partial ecologic studies 814
- Pearson correlation 98
- Pearson residuals 446, 485
- Pearson χ^2 150
- per-protocol basis 259
- permanent exposures 187
- Peto-Peto-Prentice test 514
- phases of clinical research 245
- piecewise-constant baseline hazard 537
- platykurtic 385
- point data 707
- point patterns 707
- Poisson distribution 481
- Poisson regression 482
 - Anscombe residuals 485
 - assessing overall fit 485
 - deviance residuals 485
 - evaluating Poisson regression models 485
 - extra-Poisson variation 486
 - influential points and outliers 488
 - interpretation of coefficients 483
 - overdispersion 486
 - Pearson residuals 485
 - residuals 485
 - risk ratios 485
- Poisson regression with random effects 621
- polynomial models 413
- pooled samples 130
- Popper, Karl 7
- population attributable fraction 146
- population attributable risk 146
- population-averaged estimate 618, 623
- post-test prevalence 107
- posterior distribution 677
- power 38, 49
 - power calculation by simulation 55
- pre-test prevalence 107
- precision 97
- predicted probabilities 469, 472
- predictive values 107
 - effect of prevalence 108
 - increasing the predictive value 108
 - predictive value negative 108
 - predictive value positive 107
- predictor variable 360
- Prentice-William-Peterson model 552
- prevalence 84
- prevalence and incidence 84
- prevention paradox 17
- primary sampling unit 41
- primary study base 202
- principal components analysis 407
- prior distribution 677
- probability density function 515
- probability of transmission 758
- probability sample 39
- probit function 625
- product-limit estimate 510
- profile plots 647
- profile-likelihood intervals 605
- propensity scores 191, 335
 - analysis of propensity-score-matched data 337
 - average treatment effect in treated individuals 337
 - balancing of exposure groups 336
 - computing propensity scores 336
 - kernel-matching 337
 - matching on propensity scores 337
 - multivariable modelling using propensity scores 338
 - nearest-neighbour-matching 337
 - radius matching 337
 - region of common support 336
 - stratification using propensity scores 337
- proportional hazards model 519
- proportional morbidity/mortality rates 87
- proportional-odds assumption 474
- proportional-odds model 464, 471
- proportional odds model
 - Brant (Wald) test of proportional-odds

- assumption 474
- dealing with non-proportional odds 475
- evaluating the proportional-odds assumption 474
- generalised ordinal logistic regression model 475
- heterogeneous choice logistic model 475
- partial proportional-odds model 475
- predicted probabilities 472
- regression diagnostics 475
- stereotype logistic model 475
- prospective studies 162
- prospective study design 181
- proximity matrix 720
- pseudo-gold standard 119
- pseudo-population 331
- pseudo- R^2 450
- publication bias 798
- purposive sample 39

Q

- QIC 668
- quadratic models 414
- quadrature 631
- qualitative 63
- quasi-likelihood estimation 632
- questionnaire 62
 - data-coding and editing 72
 - methods of administration 63
 - pre-testing 70
 - qualitative 63
 - quantitative 63
 - questions
 - checklist question 66
 - closed question 65
 - designing 64
 - open question 65
 - ranking question 68
 - rating question 66
 - two-choice/multiple-choice question 66
 - visual analogue scale 67
 - wording the question 69
 - response rate 71
 - structure 69
 - types 63
 - validation 71

R

- R_0 762
- R_0 - estimating 768
- R_0 - limitations 766
- r^2 366, 534
- Raftery-Lewis 688
- random allocation 255
- random coefficients 597
- random effect 589, 617, 621
- random intercept model 589, 598
- random-digit-dialling 214
- random-effects model 786
- random effects logistic regression
 - cluster-specific 618
 - ICC 620
 - interpretation of fixed effects parameters 618
 - interpretation of variance parameter(s) 619
 - latent response variables 620
 - population-averaged 618
 - subject-specific 618
 - variance components 619
- random effects Poisson regression
 - interpretation of fixed effects parameters 622
 - interpretation of variance parameters 622
- random slopes 594
 - caveats of random slopes modelling 595
 - random slope models as hierarchical models 598
 - random slopes as non-additive group effects 594
- randomised controlled trial 158, 244, 308
- ranking question 65
- raster format 703
- rate 79
- rate-based (incidence density) cohort studies 185
- rate-based approach 209
- rate-based cohort analyses 193
- rate-based designs 209
- rating scale question 65
- recall bias 295
- receiver operating characteristic curve 110, 451
- recurrence data 551
- Reed-Frost model 768

- reference category 370
 - refutationism 7
 - regression calibration estimate 298
 - regular indicator variables 369
 - reinfection threshold 772
 - reliability 98, 424
 - repeatability 98
 - repeated measurement data 566, 588, 646
 - AIC 660
 - arma(1,1) 658
 - autoregressive 657
 - compound symmetry 656
 - correlation structure 656
 - covariance matrix 653
 - exchangeable 656
 - homogeneous 658
 - ICC 659
 - linear mixed models with correlation structure 654
 - longitudinal versus cross-sectional study designs 647
 - multivariate analysis 653
 - repeated measures ANOVA 652
 - residual autocorrelation function 660
 - Toeplitz 658
 - trend models 661
 - univariate methods 649
 - unstructured correlation structure 658
 - repeated measures ANOVA 652
 - reproducibility 98
 - residual autocorrelation function 660
 - residuals 383, 446, 485, 360, 635
 - response feature 651
 - restricted maximum likelihood 601
 - retrospective studies 162, 181
 - reverse-causation 170
 - risk 80
 - calculation of risk and rate 86
 - closed population 81
 - effect of risk factor prevalence on disease risk 12
 - open population 81
 - proportion of disease explained by risk factors 16
 - risk and rate 83
 - risk difference, 144
 - risk period 78
 - risk ratio 141, 485
 - risk set 211
 - risk-based (cumulative incidence) cohort studies 184
 - risk-based cohort analysis 192
 - risk-based designs 207
 - robust standard errors 388
 - robust variance estimation 581
 - ROC 110, 451
 - root MSE, 363
 - running line smoother 411
 - Russell, Bertrand 7
- ## S
- sample 36
 - sample size
 - estimating proportions or means 50
 - expected variation in the data 49
 - general approaches to sample-size estimation 53
 - level of confidence 49
 - power 49
 - power calculation by simulation 55
 - precision of the estimate 49
 - precision-based sample-size computations 53
 - sample-size determination 48
 - sampling from a finite population 51
 - variance inflation factor 53
 - sample size - impact of information bias 299
 - sample size - RCTs 252
 - sample size - survival analyses 557
 - sampling
 - census vs sample 36
 - cluster sampling 41
 - convenience sample 39
 - hierarchy of populations 36
 - judgement sample 39
 - multistage sampling 42
 - probability sample 39
 - proportional sampling 41
 - purposive sample 39
 - sampling frame 37
 - sampling to detect disease 55
 - sampling units 37
 - simple random sample 40
 - source population 37
 - stratified random sample 40
 - study sample 37

- systematic random sample 40
- target population 37
- sampling fractions 278
- sampling odds 278
- saturated model 435
- scaled Schoenfeld residuals 530
- scaled score residual 536
- scatterplots 411, 794
- Schoenfeld residuals 530
- scientific inference 6
- score residuals 536
- screening tests 96
- screening vs diagnostic tests 96
- second-order (local or small-scale) spatial effects 718
- secondary attack rate (SAR) 87, 766
- secondary sampling units 42
- secondary study base 202
- SEIR 771
- selection bias 277
 - sampling fractions 278
 - admission risk bias 284
 - Berkson's fallacy 284
 - bias variable 278
 - comparison groups 281
 - detection bias 286
 - evaluating and correcting selection bias 287
 - non-response 282
 - reducing selection bias 287
 - sampling odds in selection bias 278
 - selective entry and survival bias 283
- selective entry bias 283
- semi-parametric analysis 506
- sensitivity 104, 450
- sensitivity analyses 774
- sensitivity-specificity plot 110
- sequential design 253, 261
- sequential testing 116
- serial correlation 397
- series interpretation 115
- sexual transmission 758
- shared frailty 547
- shrinkage on cross-validation 424
- significance (hypothesis) testing 149
- simple antecedent variable 345
- simple random sample 40
- simple regression model 360
- simultaneous autoregressive (SAR) 742
- single cohort study 180
- SIR 761, 771
- skewness 385
- Small-Hsiao test 469
- smoothed lines 411
- smoothed lines on a logit scale 412
- Snow, John 3
- source population 37, 202, 276
- space-time clusters 738
- space-time interaction
 - bivariate space-time K-function 740
 - Jacquez k-nearest neighbour test 740
 - Knox test 739
 - Mantel test 739
- spatial autocorrelation 718
- spatial cluster analysis 725
 - Cuzick and Edwards test 729
 - Diggle-Chetwynd test statistic 730
 - focussed statistics 727
 - Geary's c 733
 - Global statistics 727
 - K-function 729
 - Lawson-Waller local score test 738
 - local indicators of spatial association (LISA) 736
 - local Moran test 737
 - local statistics 727
 - Moran's I 731
 - space-time interaction tests 738
- spatial correlogram 733
- spatial scan statistic 735
- spatial clustering 566
- spatial connectivity matrix 720
- spatial correlogram 733
- spatial data 702
 - continuous features 705
 - discrete features 705
 - raster format 703
 - vector format 703
- spatial data analysis 705
 - area or polygon data 708
 - boundaries 720
 - cartogram 710
 - choropleth maps 708
 - descriptive risk mapping of area data 724
 - descriptive risk mapping of point data 722
 - difference function 729
 - diffusion cartogram 710

- dot maps 707
- dynamic visualisation of spatial data 711
- edge effects 719
- empirical Bayesian analysis 724
- exploratory spatial analysis 720
- first-order neighbours 720
- first-order spatial effects 718
- kernel smoothing 723
- level of aggregation 720
- log relative risk 723
- modifiable area unit problem 708
- point data 707
- second-order (local or small-scale) spatial effects 718
- second-order spatial effects 718
- spatial autocorrelation 718
- spatial connectivity matrix 720
- spatial dependence 718
- spatial effects 718
- spatial heterogeneity 718
- spatial weights matrix 720
- stationary 719
- visualisation 706
- visualisation of spatially continuous data 711
- visualising aggregated spatial data 707
- visualising point patterns 707
- zoning effect 720
- spatial dependence 718
- spatial heterogeneity 718
- spatial modelling 742
 - Bayesian spatial regression model 745
 - conditional autoregressive (CAR) 742
 - generalised additive mixed models (GAM) 745
 - geographically weighted regression (GWR) 745
 - kriging 742
 - simultaneous autoregressive (SAR) 742
 - trend surface regression 742
- spatial scan statistic 735
- spatial weights 720
- spatio-temporal data 711
- specification bias 390
- specificity 104, 450
- specificity of association 28
- splines 416
- split-plot design 256, 566
- split-sample analysis 424
- standard error of prediction 363
- standard errors 87
- standardisation of risks and rates 89
 - direct standardisation of rates 90
 - indirect standardisation of rates 89
 - indirect standardisation of risks 90
- standardised coefficients 425
- standardised mean difference 788
- standardised morbidity/mortality ratios 89
- standardised residuals 384
- standardised risk ratio 330
- stationary correlations 657
- statistical heterogeneity 791
- Stein's paradox 793
- stepwise regression 422
- stochastic models 761
- stratified analysis 322, 578, 793
- stratified random sample 40
- strength of association 27
- STROBE 172, 194, 218
- studentised residuals 384
- study base 202
- study design 159
 - analytic study 36
 - descriptive study 36
 - descriptive versus analytic studies 36
 - experimental versus observational studies 157
 - prospective 162
 - reporting of observational studies 170
 - retrospective 162
- study group 247, 277
- study period 78
- study quality 783
- study sample 37
- subgroup analyses 261, 792
- subject-specific 618
- subplots 567
- sufficient cause 11
- summary statistic 651
- suppressor variables 350
- survey 62
- survey data
 - clustering 47
 - design effect (deff) 47
 - finite population correction 47
 - sampling weights 45

stratification 44
 variance linearisation 47
 survey methods 582
 survival bias 283
 survival data 502
 survival time - quantifying 503
 incidence rate 504
 mean time to recurrence 503
 median time to recurrence 504
 n-year survival risk 504
 overall probability of recurrence 504
 survivor function 508, 515
 susceptible individual 756
 Susceptible-Infectious-Recovered (SIR) model 761
 systematic random sample 40
 systematic reviews 781

T

target population 37, 276
 targeted (risk-based) sampling 43
 targeted interventions 763
 Tarone-Ware test 514
 test statistic 150
 time ratio 542
 time sequence 27
 time-series data 396
 time-to-event data 502
 time-varying effects 524
 time-varying predictors 524
 Toeplitz 658
 tolerance 375
 total causal effects 321
 trace plot 686
 transition models 664
 trend models 661
 trend surface regression 742
 trim-and-fill 800
 true prevalence 106
 truncation 505
 two-choice/multiple-choice question 66
 two-stage sampling designs 237
 two-graph ROC plot 110
 types of data 785
 types of error 38
 Type I (α) error 38
 Type II (β) error 38

U

unconditional associations 406
 underdispersion 580, 638
 underlying risk 797
 understanding causal relationships 342
 unit of concern 248
 unmeasured confounders 340
 unstructured correlation 658

V

vaccine efficacy 145, 263
 vaccine trials 262
 vaccines 764
 validation 774
 validation studies 297
 validity 276
 external validity 37
 internal validity 37
 variable names 836
 variance component models 588
 variance components 619
 variance inflation factor 53, 375
 variance partition coefficient 689
 vector borne 758
 vector format 703
 venn diagrams 343
 vertical transmission 758
 visual analogue scale 67
 visualisation 705

W

waiting time 481
 Wald statistics 151, 436
 Wald χ^2 test for homogeneity 323
 Weibull model 538
 whole-plots 566
 wide data format 647
 Wilcoxon test 513
 withdrawals 81
 Woolf's approximation 152
 working correlation matrix 667

XYZ

zero-inflated models 496
 zero-truncated models 499
 zoning effect 720