

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*, 2d edition (Rothman and Greenland, 1998).

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	t+u = a ₁
	Non-exposed	v	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

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

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, $u_{\text{herd}(i)}$ refers to the herd containing the i^{th} individual. If there are 40 herds, 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 herd and i to the i^{th} individual in the j^{th} herd. 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 Causal 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
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 (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 j
<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	observation counter
I	incidence rate
I	Moran's I (spatial autocorrelation coefficient)
i	Incidence = rate at which new infections are occurring in the population - this is the population incidence rate (designated i to differentiate it from 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_{dir}	directly standardised rate
I_e	expected incidence rate
I_h	indicator for K-function
IIA	independence of irrelevant alternatives
I_{ind}	indirectly standardised rate
IPTW	inverse probability of treatment weighted
IQR	interquartile range
IR	incidence rate ratio
<i>IR_G</i>	incidence ratio based on group-level data
<i>Is</i>	standard population incidence rates
j	designated for strata
j	designator for categories
j	designator for covariate patterns in a dataset
j	designator for time intervals (actuarial life table) or time points (KM life table)
j	sampling interval in systematic random sample
J	total number of j

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
lnL	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
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
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)
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_{ti}	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
SMR	standardised morbidity/mortality ratio
so	sampling odds

<i>Sp</i>	specificity
<i>sr</i>	sampling risk
<i>SRR</i>	Standardized risk ratio
<i>SS</i>	subject specific
<i>STROBE</i>	STrengthening the Reporting of OBServational studies in Epidemiology
<i>t</i> or <i>T</i>	animal-time
<i>TCE</i>	True causal effect
<i>t_d</i>	doubling time
<i>T_i</i>	study outcome for study <i>i</i> in meta-analysis
<i>t_j</i>	time of event (KM life table)
<i>t_{j-1}, t_j</i>	time span in the interval (actuarial life table)
Δt	length of period
<i>TP</i>	true prevalence
<i>TR</i>	time ratio
<i>T_S</i>	standard population animal-time at risk
<i>T_{scan}</i>	Spatial scan statistic
<i>TVC</i>	time-varying covariate
<i>U</i>	measure of confounding bias (meta-analysis)
<i>u_i</i>	random effect of study <i>i</i>
<i>var</i>	variance
<i>V_i</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
<i>w_j</i>	subjects withdrawn during interval (censored observations) (actuarial life table) or censored observations at time <i>t_j</i> (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
<i>Z_α</i>	standard normal percentile for $\alpha/2$ Type I error (for sample size calculations)
<i>Z_β</i>	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 $Y \sim N(0,1)$)
≈	approximately equal to
α	level of significance (Type I error)
β	regression coefficient or vector ($1 \times n$) 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
$\gamma(\text{h})$	empirical semi-variogram
$\delta\delta$	spatial edge correction factor
Δ_i	Glass's I for study i (meta-analysis)
$\varepsilon\varepsilon$	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
$\pi\pi$	3.14159 (natural number)
ρ	correlation - intra-class correlation coefficient (r also used)
ρ_{ec}	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
$\tau\tau$	cutpoint for proportional odds
τ	distribution of survival times
τ^2	between study variance in meta-analysis
$\varphi\varphi$	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 location and animal-health problems

AID	autoimmune disease
Ap	<i>Actinobacillus pleuropneumonia</i>
AVC	Atlantic Veterinary College, at University of Prince Edward Island, Canada
BRD	bovine respiratory disease

BRSV	bovine respiratory syncytial virus
BSE	bovine spongiform encephalopathy
bst	bovine somatotropin (growth hormone)
BVD	bovine viral diarrhea
BVDV	bovine viral diarrhea virus
d	day(s)
EBL	enzootic bovine leukosis
ELISA	enzyme-linked immunosorbent assay
IBR	infectious bovine rhinotracheitis (Herpes 1)
IFAT	indirect fluorescent antibody test
ISA	infectious salmon anemia
Map	<i>Mycobacterium avium</i> subsp <i>paratuberculosis</i>
Mh	Mannheimia hemolytica
mo	month(s)
MUN	milk urea nitrogen
OD	optical density
Ont.	Ontario (large province in Canada)
OVC	Ontario Veterinary College at University of Guelph, Ontario
PCR	polymerase chain reaction
PEI	Prince Edward Island (smallest province in Canada)
PI	persistently infected (eg with BVDV)
ppb	parts per billion
ppm	parts per million
TAC	total aerobic bacterial count
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 herd size (-numcow-) broken into four categories (0-29, 30-59, 60-89, 90+) would result in the following four variables:

```
-numcow_0-
-numcow_30-
-numcow_60-
-numcow_90-
```

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

INDEX

A

accelerated failure time models 507
 acceleration parameter 507
 coefficients in aft models 508
 generalised gamma model 509
 log-logistic model 508
 log-normal model 509
 time ratio 508
 accuracy 93
 accuracy and precision 93
 actuarial life tables 473
 adaptive design 220
 adjacent-category model 430, 441
 adjusted odds ratio 285
 admission risk bias 252
 agreement 94
 bias 98
 factors affecting kappa 97
 kappa 96
 McNemar's χ^2 98
 multiple raters 98
 weighted kappa 98
 Akaike's Information Criteria (AIC) 384
 all possible/best subset regressions 385
 alternative hypothesis 145
 analytic control of confounding 288
 analytic sensitivity and specificity 93
 analytic study 34, 47
 apparent prevalence 102
 apparent prevalence 124
 assumptions in logistic regression 399
 atomistic fallacy 783
 attack rates 83
 attributable fraction (exposed) 140
 attributable risk 140
 autocorrelation 609

B

backward elimination 385
 backward stepwise 385
 basic reproductive number (R_0) 722
 Bayesian analysis 638
 burn-in period 644
 choice of prior distributions 640

conjugate priors 640
 Gibbs sampling 644
 homogeneous chains 642
 improper prior 642
 latent class models for diagnostic test
 evaluation 656
 Markov Chain Monte Carlo 643
 Markov chains 642
 measurement errors and imperfect tests 655
 Metropolis-Hastings sampling 644
 missing data 654
 statistical analysis based on MCMC
 estimation 646
 Bayesian information criteria (BIC) 384
 Bayesian paradigm 639
 Bayes' theorem 639
 posterior distribution 639
 prior distribution 639
 Berkson's fallacy 252
 beta-binomial model 593
 binary data 580
 blinding 175
 Bonferroni adjustment 229
 Bonferroni correction 613

C

case-case studies 203
 basis 203
 design issues 204
 case-cohort studies 206
 analysis 207
 basis 206
 design issues 206
 case-control study 182
 admission 189
 analytic control 194
 case series 183
 closed source population 183
 exclusion and inclusion criteria 193
 matching 194
 neighbourhood controls 190
 nested 182
 number of control groups 193
 number of controls per case 193
 open populations 183

- population-based controls 190
- primary study base 182
- principles of control selection 184
- random digit dialling 190
- sampling controls from a primary-base open population 188
- sampling controls from a secondary base 189
- secondary study base 182
- selecting controls in rate-based designs 187
- selecting controls in risk-based designs 185
- source population 182
- study base 182
- subject's exposure 190
- case-crossover studies 200
 - analysis 202
 - design issues 200
- case fatality rates 83
- case-only studies 208
 - analysis 209
- case-series studies 204
 - analysis 206
 - basis 204
 - design issues 205
- categorising continuous predictors 377
- causal diagram 284, 307
- causal diagram 21
- causal model 367
- causal relationships
 - causal diagram 307
 - distorter variable 313
 - explanatory antecedent variable—complete confounding 310
 - explanatory antecedent variable—
 - incomplete confounding 311
 - exposure-independent variable(s) 308
 - extraneous variable 308
 - graphical aids 307
 - intervening variable 312
 - moderator variable 315
 - simple antecedent variable 309
 - spurious relationships 307
 - summary of effects 315
 - suppressor variables 314
 - Venn diagrams 307
- causation
 - causal criteria 23
 - coherence or plausibility 26
 - consistency 26
 - dose-response relationship 26
 - statistical issues 25
 - strength of association 25
 - study design 25
 - time sequence 25
- causes 9
 - causal complement 13
 - component-cause model
 - component-cause 10
 - component-cause model 10
 - direct causes 16
 - indirect cause 16
 - necessary
 - necessary cause 10
 - proximal causes 16
 - sufficient
 - sufficient cause 10
- censoring 471
 - gaps 471
 - interval censoring 471
 - left censoring 471
 - right censoring 471
 - truncation 471
- census 34
- centring 339
- clinical trial 214
 - participants 214
 - phases of clinical research 214
 - subjects 214
- closed population 77
- cluster sampling 39
- clustered data 530
- clustering 536
 - adjustment by overdispersion factor 546
 - adjustment by the design effect 544
 - clustering and confounding 541
 - clustering for binary outcome 540
 - clustering for continuous data 537
 - clustering for discrete data 537
 - fixed effects and stratified models 543
 - intra-class correlation coefficient (ICC) 544
 - Mantel-Haenszel procedure 544
 - overdispersion 544
 - predictor variables 534
 - simulation studies on the impact of
 - clustering 540
 - stratified analysis 544

- variance adjustment factor 536
- variance inflation as a result of clustering 537
- Cochran's Q statistic 751
- coefficient of determination 330
- cohort study 168
 - analytic control 175
 - blinding 175
 - confounding 174
 - diagnostic criteria 175
 - exclusion or restricted sampling 174
 - exposure status 174
 - exposure threshold 171
 - exposure time 173
 - fixed cohorts 170
 - follow-up period 175
 - longitudinal study 168
 - matching 174
 - measuring the outcome 175
 - non-permanent exposures 173
 - permanent exposures 172
 - rate-based (incidence density) designs 170
 - rate-based cohort analyses 177
 - reporting of cohort studies 177
 - risk-based (cumulative incidence) designs 170
 - risk-based cohort analysis 176
 - single cohort 168
 - STROBE 177
- compartmental models 717
- complete case analysis 374
- concordance correlation coefficient 95
- conditional association 284
- conditional logistic regression 422
- conditionally independent tests 110
- confidence intervals 143
- confidence intervals 147
- confidence intervals 83
- confounding 272
 - analytic control 275
 - confounders 272
 - exclusion 275
 - exclusion (restricted sampling) 275
 - exposure of interest 272
 - extraneous factors 272
 - intermediate factor 274
 - intervening factor 274
 - matching 275
 - population confounder 274
 - standardised risks/rates 295
- contextual effects 564
- continuation-ratio model 431, 443
- continuous spatial fields 703
- controlled trials 214
 - allocation of study subjects 221
 - alternatives to randomisation 221
 - analysis
 - intent-to-treat 227
 - multiple comparisons 229
 - per-protocol 227
 - subgroup analyses 229
 - analysis 227
 - are 225
 - clinical trial designs for prophylaxis of communicable organisms 230
 - cluster randomisation 223
 - cross-over studies 223
 - eligibility criteria 217
 - ethical considerations 233
 - factorial designs 223
 - follow-up/compliance 226
 - masking (blinding) 225
 - measuring the outcome 227
 - multicentre trials 224
 - other sample size issues 220
 - random allocation 222
 - reporting of clinical trials 235
 - sample size 219
 - sample size for sequential and adaptive designs 220
 - sample size for the allocation of clusters of subjects 220
 - specifying the intervention 225
 - split-plot designs 224
 - stating the objectives 215
 - the study group 216
 - unit of concern 217
- convenience sample 37
- convergence criterion 399
- correlated data 530
- correlated test results 111
- correspondence analysis 372
- count 74
- count data 580
- counterfactual 290
- counterfactual concepts 17

- Cox proportional hazards model 486
 - baseline hazard 489
 - evaluation
 - assumption of independent censoring 497
 - assumption of proportional hazards 494
 - Cox-Snell residuals 499
 - delta-beta 502
 - deviance residuals 501
 - goodness-of-fit 500
 - graphical assessment 494
 - graphical assessment 494
 - Harrell's *c* concordance statistic 500
 - martingale residuals 500
 - overall fit of the model 498
 - R^2 500
 - scaled Schoenfeld residuals 495
 - scaled score residual 502
 - Schoenfeld residuals 495
 - score residuals 502
 - time-varying effects 495
 - fitting the Cox proportional hazards model 487
 - handling of ties 487
 - hazard ratios 486
 - model-building 490
 - stratified analysis 490
 - ties
 - Breslow method 489
 - Efron method 489
 - marginal calculation 487
 - partial calculation 489
 - time-varying covariates 491
 - time-varying effects 492
 - time-varying predictors 492
 - validating the model 492
- Cronbach's alpha 370
 - cross-classification 531
 - cross-classified 559
 - cross-classified and multiple membership models 651
 - cross-over studies 223
 - cross-sectional studies 158
 - assessing exposure 160
 - inferential limitations of cross-sectional studies 161
 - repeated cross-sectional versus cohort studies 162
 - sample-size aspects 160
 - source population 159
 - study group 159
 - target population 159
 - crude odds ratio 284
 - cumulative hazard function 483
 - cumulative incidence 76
 - cutpoint 104
- ## D
- deductive reasoning 6
 - descriptive studies 152, 156
 - descriptive study 34
 - detection bias 251
 - detection of confounding 283
 - causal diagrams 284
 - change in measure of association 284
 - change-in-estimate 286
 - directed acyclic graph 284
 - non-collapsibility of odds ratios 286
 - statistical identification of confounders 286
 - deviance 401
 - diagnostic tests 92
 - differential equation models 732
 - direct effects 286
 - directed acyclic graph 284
 - discrete repeated measures data
 - adding correlation structure to a GLMM 624
 - GLMMs without explicit correlation structure 627
 - discrete-time survival analysis 518
 - basis for analysis 520
 - complementary-log-log regression 521
 - logistic regression 521
 - disease frequency 74
 - count 74
 - odds 75
 - proportion 74
 - rate 75
 - distorter variable 313
 - dummy variables 333
 - dummy variables 334
 - Durbin-Watson test 361
- ## E
- ecologic studies 772

analytic 772
 confounding by group 777
 effect modification 777
 exploratory 772
 inferences 776
 modelling approaches 774
 ecologic variable 773
 aggregate 773
 derived variable 773
 environmental or contextual 774
 group or global 774
 effect modification 294
 empirical semi-variogram 704
 etiologic fraction 141
 evidence
 experimental evidence 20
 limits of experimental study evidence 21
 observational evidence 21
 exact logistic regression 421
 exact probabilities 146
 exclusion (restricted sampling) 275
 explanatory antecedent variable—complete
 confounding 310
 explanatory antecedent variable—incomplete
 confounding 311
 explanatory studies 152
 exposure variable 324
 exposure-independent variable(s) 308
 external validity 155
 external validity 244
 external validity 35
 extraneous variable 324

F

factor analysis 372
 failure function 481
 fixed effects 555
 focus groups 59
 force of infection 722
 forest plot 749
 forward selection 385
 forward stepwise 385
 fractional polynomials 379
 frailty models and clustering 510
 clustering in survival data 512
 individual frailty models 511
 shared frailty models 513
 Cox model 514

Cox model—Poisson regression 515
 interpretation of coefficients 516
 funnel plot 758

G

generalisability 244
 generalised estimating equations (GEE) 627
 estimating equations 629
 GEE for multilevel data structures 630
 statistical inference using GEE 629
 generalised gamma model 509
 generalised linear mixed models 580
 complementary log-log function 589
 confidence intervals and tests 588
 GLMMs for binary data 588
 GLMMs for categorical data 592
 GLMMs for count data 589
 maximum likelihood estimation 594
 over- and underdispersion in GLMMs 600
 population-averaged versus
 cluster-specific parameters 587
 prediction 599
 quasi-likelihood estimation 596
 residuals and diagnostics 600
 robustness against model
 misspecification 600
 statistical analysis of GLMMs 593
 generalised linear models 409
 canonical link 409
 distribution 409
 estimation methods for GLMs 410
 GLM model evaluation 410
 link function 409
 generalised negative binomial models 459
 geographic information system 664
 global-influence property 378
 gold standard 100
 group level studies 772
 group-level effects 773

H

hazard function 482
 constant hazard 484
 Cox regression model 485
 gamma, log-normal and log-logistic
 hazards 485
 proportional hazards model 485

Weibull hazard 484
 herd immunity 232
 herd immunity 723
 herd sensitivity 124
 herd specificity 125
 herd-level testing 123
 apparent prevalence 124
 variance of the estimate of the A_p 125
 heterogeneity 750
 hierarchical data 530
 hierarchical indicator variables 333, 337
 historical control trials 221
 history
 multiple causation
 history of multiple causation concepts 2
 scientific inference
 history of scientific inference 5
 hurdle models 462

I

ideal experiment 272
 imputation 375
 incidence 75
 calculation of risk and rate 82
 cumulative incidence 76
 incidence count 76
 incidence density 76
 incidence proportion 76
 incidence rate 76
 incidence risk 76
 incidence times 75
 risk 76
 incidence rate 76, 722
 animal-time unit 78
 approximate calculation 78
 exact calculation 78
 incidence rate difference 140
 incidence rate ratio 138, 449
 indicator variables 333
 indicator variables 334
 indirect standardisation 85
 individual frailty models 511
 induction period 171
 inductive reasoning 5
 infection vs disease 717
 infectious diseases 716
 infectious disease epidemiology
 contact rate 720
 contacts 719
 direct transmission 719
 effective contact rate 721
 horizontal transmission 719
 infectious period 718
 latent period 718
 probability of transmission 721
 probability of transmission 725
 susceptible 718
 vertical transmission 719
 influential observations 358
 information bias 255
 differential misclassification of exposure or
 outcome 262
 misclassification of both exposure and
 disease 262
 misclassification of extraneous
 variables 263
 non-differential misclassification of disease
 case-control studies 261
 non-differential misclassification of disease-
 cohort studies 260
 non-differential misclassification of
 exposure 257
 recall bias 262
 reducing misclassification errors 262
 validation studies to correct
 misclassification 263
 information criteria 384
 instrumental variables to control
 confounding 302
 intensity models 717
 interaction 291, 340, 381
 additive scale 291
 effect modification 293
 multiplicative scale 293
 intermediate variable 274
 internal validity 155
 internal validity 35
 interquartile ranges 390
 intervening variable 312
 intervening variable 274
 intervention 214

J

judgement sample 37

K

Kaplan-Meier estimate of survivor function 475
 Kaplan-Meier function and estimator 476
 kappa 96

L

L'Abbé plot 756
 leverage 357
 likelihood ratios 107

- category specific LR 109
- cutpoint specific LR 108

 likelihood ratio statistics 147, 400
 limits of agreement plot 95
 linear mixed model 555

- best linear unbiased predictors (BLUPs) 571
- Box-Cox transformation for linear mixed models 573
- empirical Bayes estimates 571
- fixed versus random effects 573
- full information maximum likelihood 567
- inference for fixed part of model 568
- inference for random part of model 569
- likelihood-based analysis 567
- prediction 570
- residuals and diagnostics 571
- restricted maximum likelihood 567
- robustness against model misspecification 575
- sample size 576
- statistical analysis of linear mixed models 565

 linear regression

- ANOVA table 326
- assumptions
 - homoscedasticity 349
 - independence 344
 - linearity 347
 - normal distribution 347
- Box-Cox transformations 351
- Breusch-Pagan test 349
- causal interpretation 341
- Cook-Weisberg test 349
- Cook's distance 357
- deletion residuals 348
- delta-beta 359
- DFITs 357
- errors in the X-variables 337
- estimates and intervals for prediction 329
- evaluating the least squares model 343
- F-statistic 328
- homoscedasticity 349
- influential observations 357
- interaction 340
- interpreting transformed models 353
- jackknife residuals 348
- leverage 348, 357
- linearity of predictor-outcome
 - association 352
- mean square 327
- mean square error 327
- measurement error models 338
- multivariable models 325
- normality of residuals 349
- outliers 356
- R^2 330
- raw residual 347
- regression coefficient 324
- residuals 347
- root MSE 327
- scaling variables 334
- significance of a regression coefficient 329
- significance of groups of predictor variables 331
- standard error of prediction 327
- transformations 350
- X-variables 333

 link function 409
 Ljung-Box Q-test 361
 local polynomial smoother 376
 local-influence property 376
 log-logistic model 508
 log-normal model 509
 logistic regression 396

- apparent overdispersion 415
- assumptions
 - independence 399
 - linearity 399
- assumptions in logistic regression 399
- categorical predictor 403
- confounding 405
- continuous predictor 403
- covariate patterns 411
- delta-betas 418
- dichotomous predictor 402
- evaluating logistic regression models 410

- goodness-of-fit tests 412
 - hat matrix and leverage 418
 - Hosmer-Lemeshow goodness-of-fit test 413
 - MCMC
 - autocorrelation (ACF) and partial autocorrelation functions (PACF) 650
 - Brooks-Draper 650
 - diagnostics 648
 - effective sample size 650
 - kernel density plot 648
 - Monte Carlo standard error 650
 - Raftery-Lewis 650
 - trace plot 648
 - variance partition coefficient 651
 - measurement error 264
 - measurement error 255
 - measures of association 136
 - attributable fraction (exposed) 140
 - attributable risk 140
 - confidence intervals 143
 - etiologic fraction 141
 - hypothesis testing 143
 - incidence rate difference 140
 - measures of effect 139
 - presentation of incidence rate data 136
 - presentation of incidence risk data 136
 - relationships among RR, IR and OR 139
 - risk difference 140
 - significance (hypothesis) testing 145
 - standard error 145
 - strength of an association 137
 - study design and measures of association 143
 - vaccine efficacy 141
 - measures of effect in the population 141
 - meta-analysis 744
 - fixed-effects model 747
 - heterogeneity
 - evaluation 752
 - graphical assessment 753
 - meta-regression 753
 - stratified analysis 752
 - subgroup analyses 752
 - underlying risk 756
 - heterogeneity 750
 - imputing 2x2 table cell frequencies 763
 - imputing missing variance estimates 763
 - influential studies 759
 - inverse variance weighting 748
 - propensity scores 281
 - maximum likelihood estimation 398
 - interaction 405
 - interpretation of coefficients 402
 - interpretation of the intercept 404
 - model-building 408
 - outliers 416
 - overdispersion 414
 - Pearson and deviance residuals 411
 - Pearson χ^2 statistic 413
 - predictive ability of model 416
 - receiver operating characteristic curves 416
 - sensitivity and specificity 416
 - presenting effects of factors on the probability scale 405
 - R^2 (pseudo- R^2) 415
 - real overdispersion 415
 - sample size 421
 - logistic regression with random effects 580
 - longitudinal study 168
 - longitudinal study 608
 - lowess smoother 376
- M**
- Mallow's C_p 384
 - Mantel-Haenszel (MH) procedure 289
 - Mantel-Haenszel estimator 289
 - Mantel-Haenszel odds ratio 290
 - marginal structural models 298
 - marginalised models 603
 - marginally independent 284
 - Markov Chain Monte Carlo (MCMC) 638
 - masking (blinding) 225
 - matching 276
 - analysis
 - frequency-matched data 280
 - Mantel-Haenszel matched 280
 - Mantel-Haenszel χ^2 test 280
 - McNemar's χ^2 280
 - pair-matched data 280
 - blocking 276
 - caliper-matching 279
 - frequency and pair matching 279
 - general guidelines for matching 278
 - matching on propensity scores 282
 - overmatching 278

- Mantel-Haenszel 748
 - mean difference 748
 - outcome scales 760
 - Peto 748
 - process 746
 - random-effects model 748
 - sparse data 763
 - standardised mean difference 748
 - summary estimate of effect 748
 - types of data 745
 - meta-analysis of diagnostic tests 766
 - meta-analysis of observational studies 764
 - misclassification bias 255
 - missing values 374
 - mixed models 554
 - mixed models for discrete repeated measures data 624
 - model-building
 - causal model 367
 - cautions in using any automated selection procedures 386
 - correlation analysis 369
 - creation of indices 370
 - cross-validation correlation 388
 - goals of the analysis 366
 - non-statistical considerations 383
 - number of predictors 368
 - P-values and automated selection procedures 388
 - parsimony vs fit 366
 - reliability 388
 - role of subject matter knowledge 366
 - screening predictors 369
 - screening variables based on unconditional associations 370
 - shrinkage on cross-validation 388
 - specifying the maximum model 367
 - split-sample analysis 388
 - statistical considerations—non-nested models 384
 - statistical criteria—nested models 383
 - validity 388
 - models of causation 10
 - moderator variable 315
 - mortality rate 81
 - mortality statistics 81
 - multinomial logistic model 429
 - independence of irrelevant alternatives (IIA) 435
 - interpretation of coefficients 432
 - models for outcomes with alternative specific data 436
 - obtaining predicted probabilities 434
 - regression diagnostics 436
 - testing significance of predictors 433
 - multiple comparisons 229
 - multiple membership 531
 - multiple outcome event data 517
 - Anderson-Gill model 518
 - Prentice-William-Peterson model 518
 - multiple tests 109
 - multistage sample 33
 - multistage sampling 40
 - multivariable modelling to control confounding 301
 - multivariate 325
- N**
- narrative reviews 740
 - negative binomial distribution 455
 - negative binomial regression 454
 - alternative variance functions 459
 - evaluating overdispersion 459
 - generalised negative binomial models 459
 - negative binomial regression diagnostics 459
 - negative binomial regression modelling 457
 - Poisson-gamma mixture distribution 456
 - zero-inflated models 462
 - Nelson-Aalen estimate of cumulative hazard 478
 - non-probability sampling 37
 - null hypothesis 145
 - null hypothesis 36
- O**
- observational studies 157
 - odds 75
 - odds ratio 138
 - odds ratios 397
 - odds ratios 432
 - offset 730
 - density dependent 730
 - density independent 730
 - open population 77

orthogonal polynomials 379

outbreak 727

 doubling time 727

 exponential phase 727

 generation interval 727

 regression modelling 729

outcome 324

outliers 356

overdispersion 452

 apparent overdispersion 452

 dealing with overdispersion 453

 evaluating overdispersion 453

 real overdispersion 452

P

P-value 146

parallel interpretation 110

parsimony 366

Pearson χ^2 146

Poisson distribution 447

Poisson regression model 448

 Anscombe residuals 451

 assessing overall fit 451

 deviance residuals 451

 evaluating Poisson regression models 451

 extra-Poisson variation 452

 influential points and outliers 454

 interpretation of coefficients 449

 overdispersion 452

 Pearson residuals 451

 residuals 451

 risk ratios 451

Poisson regression with random effects 584

polynomial models 378

pooled samples 126

pooled testing and HSe 127

population attributable fraction 142

population incidence rate 722

power 36, 47

 power calculation by simulation 52

precision 93

predictor 324

predictive values 102

 effect of prevalence on predictive
 values 103

 increasing the predictive value of a positive
 test 104

 predictive value negative 102

 predictive value positive 102

prevalence 80

 relationship between prevalence and
 incidence 80

 apparent prevalence 102

 post-test prevalence 102

 pre-test prevalence 101

prevention paradox 15

primary sampling unit 39

principal components analysis 372

probability density function 482

probability sample 37

profile plots 609

propensity scores 281

 analysis of propensity score matched
 data 282

 average treatment effect in treated
 individuals 282

 balancing of exposure groups 281

 computing propensity scores 281

 kernel matching 282

 matching on propensity scores 282

 multivariable modelling using propensity
 scores 301

 nearest-neighbour matching 282

 radius matching 282

 region of common support 282

 stratification using propensity scores 295

proportional hazards model 485

proportional morbidity/mortality rates 83

proportional odds model 430, 437

 Brant (Wald) test 440

 dealing with non-proportional odds 441

 evaluating the proportional-odds
 assumption 440

 generalised ordinal logistic regression
 model 441

 heterogeneous choice logistic model 441

 partial proportional odds model 441

 predicted probabilities 437

 regression diagnostics 441

 stereotype logistic model 441

pseudo-population 298

publication bias 757

purposive sample 37

Q

quadratic models 378

- questionnaire 58
 - data-coding and editing 68
 - methods of administration 59
 - pre-testing 66
 - qualitative 59
 - quantitative 59
 - questions
 - checklist question 62
 - closed question 61
 - designing 60
 - open question 61
 - ranking question 64
 - rating question 62
 - two-choice/multiple-choice question 62
 - visual analogue scale 64
 - wording the question 65
 - response rate 67
 - structure 65
 - types 59
 - validation 67
- R**
- R_0
 - average age of infection 729
 - critical proportion susceptible 727
 - endemic equilibrium 729
 - estimating R_0 726
 - R_0 for endemic diseases 727
- R_0 and vaccination 723
 - critical percentage 723
 - herd immunity 723
- R^2 330
- random coefficients 562
- random digit dialling 190
- random effects 555
- random effects logistic regression
 - cluster-specific 582
 - ICC 583
 - interpretation of fixed effects parameters 582
 - interpretation of variance parameter(s) 582
 - latent response variables 583
 - population-averaged 582
 - subject-specific 582
 - variance components 583
- random effects Poisson regression
 - interpretation of fixed effects parameters 584
 - interpretation of variance parameters 584
- random intercept model 555, 563
- random slopes 560
 - caveats of random slopes modelling 561
 - random slope models as hierarchical models 562
 - random slopes as non-additive herd effects 560
- randomised controlled trial 272
- randomised controlled trials 154
- rate 75
- recall bias 262
- receiver operating characteristic curves 105
- reference category 335
- regular indicator variables 333
- repeatability 94
- repeated measurements 532
- repeated measures data 554, 608
 - AIC 622
 - arma(1,1) 619
 - autoregressive 618
 - compound symmetry 618
 - correlation structure 618
 - covariance matrix 615
 - exchangeable 618
 - homogeneous 619
 - ICC 621
 - linear mixed models with correlation structure 616
 - longitudinal versus cross-sectional study designs 611
 - multivariate analysis 615
 - repeated measures ANOVA 613
 - residual autocorrelation function 622
 - Toeplitz 619
 - trend models 622
 - univariate methods 612
 - unstructured correlation structure 619
- reproducibility 94
- risk 76
 - calculation of risk and rate 82
 - closed population 77
 - effect of risk factor prevalence on disease risk 11
 - open population 77
 - proportion of disease explained by risk factors 15
 - stable population 77

risk and rate 79
 risk difference 140
 risk period 74
 risk ratio 137
 risk ratios 451
 robust standard errors 353
 robust variance estimation 547
 ROC - receiver operating characteristic
 curves 104, 105
 running line smoother 376
 running mean smoother 376

S

sample size
 adjustment in multivariable studies 50
 adjustment for clustering 48
 comparing proportions or means 48
 estimating proportions or means 47
 expected variation in the data 46
 formulae 47
 general approaches to sample-size
 estimation 51
 impact of information bias on sample
 size 266
 level of confidence 47
 power 47
 power calculation by simulation 52
 precision of the estimate 46
 precision-based sample-size
 computations 51
 sample-size determination 46
 sampling from a finite population 48
 variance inflation factor 50
 sampling
 census vs sample 34
 cluster sampling 39
 convenience sample 37
 hierarchy of populations 34
 judgement sample 37
 multistage sampling 40
 power 36
 probability sample 37
 proportional sampling 38
 purposive sample 37
 sampling frame 35
 sampling to detect disease 53
 sampling units 35
 simple random sample 38

 source population 35
 stratified random sample 38
 study sample 35
 systematic random sample 38
 target population 34
 scatterplots 375
 screening and diagnostic tests 92
 agreement
 concordance correlation coefficient 95
 correlation coefficient 95
 kappa 96
 limits of agreement plot 95
 agreement 94
 clustering of test results 122
 coefficient of variation 94
 conditional independence 111
 conditionally dependent 111
 evaluating effects of factors on Se and
 Sp 122
 evaluation
 composite reference standard 115
 discrepant resolution 115
 gold standard populations 113
 gold standard reference test 113
 latent class models 116
 pseudo-gold standard procedures 115
 reference test with known Se and Sp 115
 evaluation of diagnostic tests 113
 factors that affect Se and Sp 121
 latent class models
 dependence among test results 121
 estimation procedures 118
 goodness-of-fit 119
 prevalence of the two populations
 differs 119
 Se and Sp constant across
 populations 120
 sample size
 gold-standard based procedures 123
 latent class models 123
 sample size requirements 123
 secondary attack rate (SAR) 725
 secondary attack rates 83
 secondary sampling units 40
 SEIR model 732
 selection bias 244
 admission risk bias 252
 Berkson's fallacy 252

- bias variable 245
- comparison groups 249
- detection bias 251
- evaluating and correcting selection bias 254
- non-response 249
- reducing selection bias 254
- sampling fractions 246
- sampling odds 246
- selective entry or survival bias 250
- selective entry or survival bias 250
- sensitivity 99
- sensitivity-specificity plot 105
- sequential design 220
- sequential design 229
- serial correlation 362
- series interpretation 110
- significance (hypothesis) testing 145
- simple antecedent variable 309
- simple random sample 38
- simple regression model 324
- SIR model 732
- smoothed lines 376
- smoothed lines on a logit scale 377
- source population 244
- source population 35
- space-time interaction
 - bivariate space-time K-function 702
 - Jacquez K-nearest neighbour test 702
 - Knox test 701
 - Mantel test 702
- spatial clustering 531, 687
 - Cuzick and Edwards test 691
 - Diggle-Chetwynd test statistic 692
 - focussed statistics 689
 - Geary's c 695
 - global statistics 689
 - K-function 691
 - Lawson-Waller local score test 700
 - local indicators of spatial association (LISA) 698
 - local Moran test 699
 - local statistics 689
 - Moran's I 694
 - space-time interaction tests 700
 - spatial correlogram 695
 - spatial scan statistic 697
- spatial data 664
 - continuous features 667
 - discrete features 667
 - raster format 665
 - vector format 665
- spatial data analysis 667, 680
 - area or polygon data 670
 - boundaries 682
 - cartogram 672
 - choropleth maps 670
 - descriptive risk mapping of area data 686
 - descriptive risk mapping of point data 684
 - difference function 691
 - diffusion cartogram 672
 - dot maps 669
 - edge effects 681
 - empirical Bayesian analysis 686
 - exploratory spatial analysis 682
 - first-order neighbours 680, 682
 - kernel smoothing 685
 - level of aggregation 682
 - log relative risk 685
 - modifiable area unit problem 670
 - point data 669
 - second-order spatial effects 680
 - spatial autocorrelation 680
 - spatial connectivity matrix 682
 - spatial dependence 680
 - spatial effects 680
 - spatial heterogeneity 680
 - spatial weights matrix 682
 - stationary 681
 - visualisation 668, 673
 - visualising aggregated spatial data 669
 - visualising point patterns 669
 - zoning effect 682
- spatial variation - modelling 704
 - Bayesian spatial regression model 707
 - conditional autoregressive (car) 705
 - generalised additive mixed models (GAM) 707
 - geographically weighted regression (GWR) 707
 - kriging 704
 - simultaneous autoregressive (SAR) 705
 - trend surface regression 704
- spatio-temporal data 673
- specification bias 355
- specificity 99
- splines 380

split-plot design 532
 standard errors 83
 standardisation of risks and rates 85
 direct standardisation of rates 86
 indirect standardisation of rates 85, 87
 indirect standardisation of risks 86
 standardised morbidity/mortality ratios 85
 standardised coefficients 390
 standardised morbidity/mortality ratios 85
 standardised risk ratio 297
 Stein's paradox 752
 stepwise regression 385
 stratified analysis 289
 stratified random sample 38
 STROBE 163
 study design 34, 154
 analytic study 34
 descriptive study 34
 descriptive versus analytic studies 34
 experimental versus observational studies 153
 prospective 157
 reporting of observational studies 158
 retrospective 157, 169
 study group 245
 study period 74
 study sample 35
 suppressor variables 314
 survey 58
 survey data – analysis 42
 clustering 43
 design effect (deff) 45
 finite population correction 45
 sampling weights 42
 stratification 42
 variance linearisation 44
 survey methods 548
 survival analyses - non-parametric
 confidence intervals 479
 log-rank test 479
 tests of the overall survival curve 479
 Wilcoxon test 480
 ‘point-in-time’ comparisons 479
 survival data 468
 survival models - parametric
 exponential model 503
 Gompertz model 507
 Weibull model 505

survival models – semi-parametric – see Cox
 proportional hazards model
 survival time 469
 incidence rate 470
 mean time to recurrence 470
 median time to recurrence 470
 n-year survival risk 470
 overall probability of recurrence 470
 survivor function 481
 survivor, failure and hazard functions 480
 survivor, failure and hazard functions 483
 susceptible-infectious-recovered (SIR) 722
 systematic random sample 38
 systematic reviews 741

T

target population 244
 target population 34
 targeted (risk-based) sampling 41
 test statistic 146
 threshold effects 716
 time-series data 361
 tolerance 339
 total causal effects 286
 transition models 626
 trim and fill 758
 true causal effect 303
 true prevalence 101
 truncation 471
 two-graph ROC plot 105
 two-stage sampling designs 209
 types of error 36
 Type I (α) error 36
 Type II (β) error 36

U

understanding causal relationships 306
 unmeasured confounders 304

V

vaccine efficacy 141
 vaccine efficacy 231
 validity 35, 244
 external validity 35
 internal validity 35
 variance component models 554
 variance inflation factor 339

Venn diagrams 307

W

Wald statistics 147

Wald tests 402

Wald χ^2 test for homogeneity 290

wide data format 609

X Y Z

zero-inflated models 462

zero-truncated models 465