INDEX

Access, to healthcare
  interval scales of, 80
  logistic regression analysis of, 311
Accountability, for quality of care, 4
Acyclical
  definition, 492
  directed acyclical graphs (DAGs), 494
Addition rule, 59, 60
Adjusted R-squared, 267–269
Age, as variable
  countable discrete levels of, 174
  distribution, 90–92
  out-of-range, 42
  as predictor of multiple outcomes, 487–488
  sequence among variables, 515
  stratified regression analysis, 430–436
Agency for Healthcare Research and Quality (AHRQ)
  chart review recommendations, 43
  claims data, 2
  Clinical Classification Software, 265, 312
  Consumer Assessment of Healthcare Providers and Systems (CAHPS), 158–159
  Healthcare Costs and Utilization Project, 365
  time to pain medication measure, 224–225
Aggregate functions, 22–23
  maximum, 25
  minimum, 25
AHRQ. *See* Agency for Healthcare Research and Quality
Alemi, F., 392
American National Standards Institute, 14
Analysis of variance (ANOVA), 164
Ancestors of variables, 496
Anderson-Darling test of normality, 289
Angular root transformation, 290
Area under receiver operating curve (AROC), 114–116, 117–118
  of sequential network learning algorithms, 516
AROC. *See* Area under receiver operating curve
Association
  between cause and effect, 240
  likelihood ratios and, 110
  of variables, 7
Association networks, 459, 468–469
  construction with Poisson regression, 481–484
  strata-conditioning relationship and, 463
Attributable risk, 241–242, 243–244
Attributes, 78. *See also* Variables
Aurora Health Care System, 2
Autocorrelation plots, 285–286, 287
Average, 82–83
  dispersion around, 85–87
  weighted, 84–85
Avramovic, Ivan, 392

Back-door path, 507–510
  blocked, 493, 509–510, 517, 520, 522
  covariates, 507, 508–509
  definition, 491, 493

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Backward stepwise selection of independent variables, 281
"Bad data" file, 24
Bar charts
of binomial distribution, 178, 179
as histograms, 92, 93
Barthel index, 80
Bayes, Thomas, 63
Bayesian probability models, 102
Bayes's data mining model, 105–106, 111–112
Bayes's formula, 68
conditional independence applied to, 64–66
odds form, 105–106, 110, 111–112
posterior odds of mortality, 112
prior opinion revision and, 62–63
of probability, 62–66
Bayes's theorem, 63
BayseiaLab, 511
Benchmarking, of physician performance, 2–3, 409–425
data balancing in, 409–411
distribution switching, 410, 411–413
limitations, 420–421
multilevel modeling of, 346
multiple comorbidities, 409, 413–416
overlap in, 416–418
patient matching in, 413–416, 421
SQL codes, 422–423
steps in, 409–410
stratified covariate balancing in, 409, 411
synthetic controls, 410, 416, 418–420, 421
terminology, 411
Bernoulli, Jacob, 175
Bernoulli density functions, 205–207
geometric distribution and, 206–209
Bernoulli distribution, 5
Bernoulli trials, 175, 206–207, 208
Big data
applications, 1
definition, 1
Binary variables, 79, 80, 309
common odds ratio, 386–389
as indicators, 79
logistic regression of, 309
main and interaction effects, 274–275
rate of occurrence, 174
Binomial distribution, 5, 90, 175–179
bar charts of, 178, 179
binomial parameters, 175–176
conditions required for, 175
definition, 175
example, 176–179
in Excel, 176, 181
normal approximation, 179–181, 182
normal distribution, 176, 179–181
skewed, 179
Binominal parameters, 175–176
Birth, clinical encounters before, 40
Birthdates, wrong, 40
Blocking the back-door path, 493, 509–510, 517, 520, 522
Breush test, 288
Budget
control of, 310
Tukey's charts of, 234–235
Bullying, 310
Bundled prices, 489
Buspirone, weighted propensity scoring study, 338–342

CAHPS. *See* Consumer Assessment of Healthcare Providers and Systems
Case, definition, 385
Case-control design, 49, 51, 52. *See also* Matched case control studies
backward look, 51
forward look, 48–51
with multiplicative stratified regression, 438
retrospective, 48

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

stratified, chi-square test of, 386–387
Case mix index, 83–85
Casual relationships, display of, 468
Categorical variables, 69, 79
Causal analysis, 7, 9. *See also* Causal control charts; Causal networks
regression use in, 296–297
Causal chains, 494, 495
Causal control charts, 239–254
assumptions of causality, 240–241
comparison with traditional control charts, 244–245
control limits calculations, 246, 247–250
lower control limit, 246, 249–250
upper control limit, 246, 249–250
weighted controls, 246, 249–250, 251
counterfactual, 245–251
data balancing in, 244, 245–246
Causal impact, 506–507
example, 516–519
Causal interpretations, 7
Causal networks, 487–526
back-door paths, 507–510
blocking of, 493, 509–510, 517
covariates, 507, 508–509
definition, 491, 493
common effect in, 495
direct cause in, 494–495
as directed acyclical graphs (DAGs), 494
directed arcs in, 493–494, 495, 514–516
"do operation" manipulation, 507
key concepts, 491–497
Markov blankets, 492, 493, 510, 513–514, 516–517, 510
network parameters specification, 510
network structure identification, 510–514
constraint-based algorithms, 512–513
multivariate methods, 513–514
search-and-score methods, 511
parameters identification, 516–517
parent–child node relationships, 501–502
probability calculations, 501–506
eliminating steps, 506
joining steps, 506
relationship with regression, 497–501
spurious correlation removal in, 496–497
terminology, 496
Causation
principles, 7–8, 489–491
relationship with correlation, 7, 476, 489
Cause and effect, 506
Centers for Disease Prevention and Control (CDC), Data to Care (D2C) procedures, 4
Centers for Medicaid & Medicare Services (CMS)
bundled payments, 499–501
Death Master List, 39
healthcare database, 2
hierarchical condition category (HCC), 416–419
Hospital Compare website, 224–227
patient satisfaction scores, 159
Central limit theorem, 141–144
Central tendency, 137
Characteristics, 78. *See also* Variables
Charlson comorbidity index, 102–103, 119
Chart reviews, 43
Children, of Markov blankets, 492, 493, 496, 510

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Chi-square test, 71–73
for goodness of fit, 322
for homogeneity of treatment impact, 386–389
for Hosmer-Lemshow tests, 324
of independence, 467, 478
for three variables, 471–475
Clinical encounters
chronology of events in, 47–51
erroneous data removal, 38–45
absence of encounters, 40–41
encounters after death, 38–39
encounters before birth, 40
patient-reviewed, 212–219, 221
Clinical practices, Big data-based evaluation of, 4
Clues, 78. *See also* Variables
CMS. *See* Centers for Medicaid & Medicare Services
Coefficient of determination, 277–280
maximization, 280
in predictive medicine, 281
Cohen's *d,* 149
Cohort study, 48–51, 52
Collider, definition, 492
Collider test, 515
Collinearity effects, 291–292
Common cause, 496
Common effect, 495
Community hospitals, survival rate analysis, 349–354
Community living centers, 520, 522
Comorbidities, 46, 47
causal network analysis, 518–519
as lung cancer prognostic factor, 439–447
multiple
in clinician benchmarking, 409, 413–416
risk assessment. *See* Multimorbidity index
stratification of, 488
Comparative effectiveness studies, 4. *See also* Matched case control studies
contradictory conclusions, 364
Compensation, performance-based, 2–3
Complaints, from patients, 212–219, 221
Complications, 46–47, 498–499
Conditioning
definition, 385, 493
through reduction in strata, 463–464
Confidence intervals, 5, 150–151
95 percent, 295
estimation, 6
for odds ratio, 186–189, 374–376
of Tukey's charts, 223–224
Confidence value approach, 150
Confounding, 1
in regression analysis, 498
removal
automated, 403–406
in causal analysis, 498
with propensity scoring, 327, 337–343, 384
with randomization, 384
Congestive heart failure, length of stay, 413–416
Consistency tests, 43–44
Constant probability *p,* 206
Constant variance, 161
Consumer Assessment of Healthcare Providers and Systems (CAHPS), 158–159
Contingency tables, 57, 66–71, 323
definition, 66–67
in Excel, 68–71
marginal, 473, 474
multidimensional, 471–473
partial, 473
Continuous distribution, 136, 138. *See also* Normal distribution
Continuous interval scale, 160
Continuous outcomes, stratified covariate balancing
difference models, 389–390
weighted data, 390–392, 393–395

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Continuous probability density function, 136
Continuous variables, 80, 309
  average of, 141
  random, 136
  relationships among, 73–74
Contradictory data, 42–44
Control, definition, 385
Control charts, 98-100. *See also* Causal control charts; p-charts; Tukey's control charts; XmR control charts; X-bar control charts
  control limits calculations, 154–156
    lower control limits, 98, 99, 153, 154
    upper control limits, 98, 99
  elements of, 153–156
  lower control limit (UCL), 154
  mix-adjustment of, 210
  with normal distribution
    X-bar charts, 152, 153, 158–171
  probability, 5
  risk-adjustment of, 194–199, 201, 210, 244–245
  stratification in, 245
  time-between, 209–212, 236
    of exercise resolutions, 219–220
    lower control limits, 212
    of patient reviews, 212–219, 221
    of sentinel events, 203–222
    upper control limits, 212, 216–217, 218, 219
  time changes on, 153
  upper control limit (UCL), 153, 154
  X-axis, 153
  Y-axis, 153
Control charts, 98-100. *See also* Causal control charts; P-charts; Tukey's control charts; XmR control charts; X-bar control charts
  applications, 152–153
Controls, synthetic, 410, 416, 418–420, 421
Co-parents, of Markov blankets, 492, 496, 510
Copay, stratified regression analysis, 430–436, 439–440
Corner cases, 355
Corner stratum, 437–440, 442
Correlation
  causation relationship, 7, 476, 489
  conditional, 475
  spurious, 475–477, 496–497
Correlation coefficients, 74–75
Cost
  as continuous variable, 79, 80
  as interval variable, 79
Cost data analysis
  of joining insurance plans, 84–85
  log transformation, 97
  with single cost predictor, 266–271
  statistical evaluation, 264–266
  transformation, 87–88
Cost-effectiveness analysis
  matched case controls of, 363
  value-based reimbursement and, 4
Cost overruns, frequency, 57
Counterfactual assumption, 240, 490, 498
Counterfactual causal control charts, 245–251
Counterfactual groups, 244
Covariates, 52. *See also* Stratified covariate balancing
  in causal analysis, 240
  combinations, 384, 385–386
  in control charts, 245
  definition, 491, 492
Cox, Louis Anthony Jr., 203–204
Cox's hazards regression, 258
Creating dummy variables, 69
Critical value approach, 146
Critical value of the test statistic, 145

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Cross-validation, 292–293
of multimorbidity index, 109–110
c-statistic, 323
Cumulative distribution function, 241, 242–243
Cumulative probability distribution, 139–140

Data
cleaning of. *See* Erroneous data
collection, 7
on the internet, 2
discarding of, 45–47
errors in entry, 24
filtering, 23–24
merging
matrix format, 12
in SQL, 5–6, 12
preparation
decision-making in, 12–13
errors in, 13
importance, 11–13
with SQL, 11–12
Data analysts, clinical roles of, 4
Data balancing, 244, 245-246. *See also* Propensity scoring
in benchmarking, 409–411
definition, 410
Data-driven companies, 2–3
Data to Care (D2C) procedures, 4
Dates
of birth, wrong, 40
conversion from text, 35–36
of death, wrong, 39–40
out-of-range, 42
Death, probability of, 80. *See also* Prognosis
Deceased patient data, 13
coding, 25
deletion, 38–39
erroneous, 38–40
Decision-making
clinical, 4
in data preparation, 12–13
Decision trees, 415
Definition of the variables method, 515
Density plots, 290, 293
Dependent samples, 151
Dependent samples test, 151–152
Dependent variables, 68
as count of events. *See* Logistic regression
in straight line equation, 259
Descendants of treatment/variables, 492, 496
Deviance (D), 323
expected, 197–199, 200
Diabetes mellitus
as complication of treatment, 47
hemoglobin A1C assessment, 163–171
multimorbidity index, 117, 118
Diagnosis
association with outcome, 110
missing, 13, 41–42
post-outcome, 46
post-treatment, 46–47
Diagnosis-related groups (DRGs)
case mix index, 83–84
in clinician benchmarking, 416, 418, 419
Diagnostic codes, 365. *See also* International Classification of Diseases (ICD)
interaction of, 428
Dichotomous events, 189
Directed acyclical graphs (DAGs), 494
Directional separation of network arcs, 514–516
Disabilities progression, casual analysis, 511, 512
Discrete variables, 69, 73, 79
histogram of, 90, 91
in propensity scoring, 331–332
rate of occurrence. *See* Rates, comparison of
Dispersion, around averages, 85–87

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Distribution, 6, 71, 90-92. *See also* Frequency distribution; Normal distribution
continuous, 136, 138
definition, 90
exponential, 142, 143
parabolic, 142, 143
switch, 410, 411–413
uniform, 142, 143
DNR (do not resuscitate) orders, causal analysis of, 503, 505–506
"Do operation," 507
Double-counted data, 13
DRGs. *See* Diagnosis-related groups
Dummy variables, 69, 271
Durbin-Watson statistic, 285–286, 287

Eating disabilities–mortality relationship
causal analysis, 507–509, 520, 522
stratified covariate balancing, 392–395
EHRs. *See* Electronic health records
Electronic health records (EHRs)
comprehensive, 7
erroneous data removal, 38–45
absence of encounters, 40–41
encounters after death, 38–39
encounters before birth, 40
free text, 30–33
as quality of care measures, 43
tables
encounter, 19–21
foreign keys, 15
joining of, 25–29
multiple, 5–6, 12
patient, 17–21
patient fields, 15
primary key, 15
provider, 19, 20–21
relationships among, 15, 20–21
Electronic health records (EHRs) data
merging, 5–6
observational nature, 5
random sampling, 81
Electronic health records (EHRs) systems, matched case controls of, 363
Elixhauser index, 102–103, 117–118, 119
Elopement, of patients, 206, 207–208
ElRafey, Amr, 392
Emergency departments
patient boarding time analysis, 247–250, 251, 489
time to pain medication performance measure, 224–227
Encounter tables, 19–21
End nodes, 502
Erroneous data
about clinical encounters, 38–45
absence of encounters, 40–41
encounters after death, 38–39
encounters before birth, 40
missing data, 41–42
Errors
in data entry, 24
Goodman and Kruskal reduction of, 515
medication, 204–205, 206, 232–233, 243–244
medication errors, 43
normally distributed, 288–289
in R software, 302
sum of squares of errors (SSE), 279
type I, 145, 146
type II, 145, 146–147
Error terms
heteroscedasticity of, 286–287
homoscedasticity of, 286
normal distribution of, 288–289
Event trees, 413–416, 415–416, 418
Excel. *See* Microsoft Excel
Exercise–weight loss relationship, Tukey's chart of, 228–230
Expected deviance, 197–199, 200
Exponential distribution, log transformation of, 97

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

False negative, 146
False positive, 146
FDA. *See* Food and Drug Administration
Fee-for-service organizations, logistic regression analysis of, 311
Finite population correction factor, 141
Fmax test, 152
Food and Drug Administration (FDA), 2
Forecasting, with multilevel modeling, 346
Forward stepwise selection of variables, 281
Fourth spread, 223–224, 226, 229, 230–231, 232–233
Fractures, time to pain medication study, 224–227
Free text, 30–33
Frequency distribution, 57, 87, 90–92, 174–175
  cumulative, 90, 91
  with histograms, 92–97
  relative, 90, 91
F-statistic, 279
F-test, 145
F value test, 280

Gamma distribution, 236–237
Gender, as variable
  countable discrete levels of, 174
  stratified regression analysis, 430–436
George Mason University Dataverse, 119
Goldfeld-Quant test, 288
Goodman and Kruskal errors reduction, 515
Goodness of fit
  adjusted R-squared, 267–269
  chi-square test, 322
  G-squared, 481, 482
  in logistic regression, 322–324
  in Poisson regression, 481
  R^2 statistic, 277–278
Grouping, of data, 22–23, 24–25
Grow-shrink algorithm, 512–513
G-squared statistic, 481, 482

Hartley's Fmax test, 152
Hazard rate, 241–244
HCC. *See* Hierarchical condition category
Healthcare, overuse of, 311
Healthcare team members, contributions to patient satisfaction
  propensity scoring, 329
  stratified covariate balancing, 387–389
Health insurance companies
  data analysis use, 3
  risk assessment methods, 3
  start-up, 3
  stock market price analysis, 250, 252
Health insurance cost, stratified regression analysis, 430–436, 441–450
Health insurance status, as categorical variable, 79
Heteroscedasticity, 286–288
  weighed regression and, 294
Hierarchical condition category (HCC), 416–419
Hierarchical regression. *See* Multilevel regression
Histograms, 92–93
  in Excel, 94–97
HIV. *See* Human immunodeficiency virus (HIV) infection
Home health care organizations, 4
Homogeneity of variance, 152
Homoscedasticity, 286
Hosmer-Lemshow tests, 324
Hospice care, effect on hospital readmissions, 48–52, 466–467
Hospital acquisitions, logistic regression analysis of, 311
Hospital Compare website, 224–227

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Hospital discharge, number of diagnostic codes in, 428
Hospital occupancy, 1
Humana, 250, 252
Human immunodeficiency virus (HIV) infection, 4
  multimorbidity index of, 117
Human resources
  logistic regression use, 310–311
  matched case controls use, 364
Hypothesis testing, 1, 5, 9. *See also* Null hypothesis
  chi-square test, 71–73
  comparison of two-sample means, 151–152
  conclusion of, 145
  context-specific, 295–296, 321–322
  definition, 144
  in Excel, 183, 186, 187
  level of significance, 145
  of logistic regression coefficients, 321–322
  one-sample *z* test of population means, 147–148
  of population proportion, 181–186, 187
  regression coefficients in, 295–296, 321–322
  steps in, 72–73
  test statistic, 145
    critical value, 145
    observed value, 145
  type I errors, 145, 146
  type II errors, 145, 146–147

ICC. *See* Interclass correlation
Identity theft, 39
Improvements, risk-adjusted P-charts of, 194–199, 201
Incongruous data, 13
Inconsistency, of data, 13, 42–44
Inconsistent data form, 44–45
Independence
  chi-square test of, 73, 467, 478
  for three variables, 471–475
  complete, 470, 474, 475
  concept, 460–461
  conditional, 64–66, 461, 464, 465
    causal network arcs and, 514, 515
  definition, 460
  joint, 470, 473–475
  joint preferential, 437, 442
  marginal probabilities and, 464–467
  of multiple morbidity index, 105–106, 107
  mutual information and, 477–478, 479
  of observation, 52
  order of parsimony, 470
  for three variables, 469–471
    chi-square test, 471–475
  visual network display of, 467–469
Independence assumption, 152
Independence samples tests, 151–152
Independent events, 160–161
Independent variables
  backward stepwise selection, 281
  in chi-square test, 72
  collinearity, 291–292
  with interactions, 428–429
  monotone relationships, 437
  in ordinary regression, 271
    binary variables, 271–273
    categorical variables, 271, 273
    continuous variables, 271
    interaction variables, 271
    preparation, 271–274
  squaring, 290
  unconfounded impact estimation, 433–436
Inferential statistics, 141–142
Influenza outbreaks, prediction of, 281
Interaction effect, 274–277
Interaction terms, 276–277, 428
  higher order, 428
  pair-wise, 428
  in propensity scoring, 331

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Interaction variables, 271
Intercept coefficient, 295
Interclass correlation (ICC) coefficient, 348
International Classification of Diseases (ICD)
codes, 24
use in case identification, 365
International Classification of Diseases (ICD-9), 103–104, 106, 108, 109
International Classification of Diseases (ICD-10), 106, 108–109
Internet
data collection on, 2
patient reviews posted to, 82–83
of things, 2
Interrelated variables, 493–494
Interval scales, of healthcare access, 80
Interval variables, 79, 87
Intuition, 98, 152–153
Inverse probability of treatment weighing (IPTW), 337–342
IPTW. *See* Inverse probability of treatment weighing
Item set, 403–404

Jarque-Bera test of normality, 289
Joint events, probability of, 464, 467
Joint preferential independence condition, 437, 442

Kaiser Permanente, 2
Kaplan-Meier estimator, 377–378
k binary covariates, 385
k constant, 439, 442, 444
k-fold cross-validation, 293
Kolmogorov-Smirnov test of normality, 289

Landmark time, 47–48
LASSO regression. *See* Least absolute shrinkage and selection operator (LASSO) regression
Law of large numbers, 142–144
Lean, definition, 3
Least absolute shrinkage and selection operator (LASSO) regression, 294–295, 399, 513–514, 518
Leave-one-out cross-validation, 293
Length of stay
case mix index, 83–84
casual network analysis, 494–496
clinician benchmarking, 413–416, 419–420
as continuous variable, 80
patient comorbidities effect, 413–416
Levene's test, 152
Likelihood ratios, 63
as association measurement, 110
conditional independence-based calculation, 64–66
contingency table-based, 69–71
definition, 68
of multimorbidity index, 102, 106–109
combination of diseases adjustment, 107
confounding, 111–112
detection applications, 110
diseases with no or complete mortality, 107–108
predictive applications, 110–111
rare diseases, 108
repeated diseases adjustments, 106–107
of zero value, 111
Linear equations, 259–260
Linear regression, 105
assumptions, 282–289
correctness of model form, 282–285
data transformations, 290–291, 292, 293
heteroscedasticity, 286–288
homoscedasticity, 286
independence of error terms, 285–286

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

normally distributed errors, 288–289
ordinary
logit transformation in, 315–317
R software for, 305–307
Linear transformation, of variables, 87–90
Link functions, 314
Logarithm transformation, 290–291, 292, 293
Logistic regression, 309–325
applications, 310–311
calculation, 313–315
case study, 312–315
coefficients, 320–321
hypothesis testing of, 321–322
estimation of parameters, 320–321
goodness of fit, 322–324
natural logarithm function, 313–314
propensity scoring with, 313–319, 331, 338, 339
R code, 318–319
SQL code, 315–316
Logit function of π, 314
Logit transformation, in ordinary regression, 315–317
Log transformation
of cost data, 290–292
of skewed distribution, 97
Longitudinal data, multilevel modeling of, 358
Lung cancer prognosis, stratified regression analysis, 439–458
SQL codes, 447–450
strata, 452–458

Main effect, 274–277
Marital status, as categorical variable, 79
Marketing
logistic regression use in, 311
matched case control use in, 362
Markov blankets, 399, 401–403
children, 492, 493, 496, 510
co-parents, 492, 496, 510
definition, 493
parents, 492, 493, 496, 499–501, 510, 513–514
Matched case control studies, 361–382
applications, 362–364
case definition, 364
controls definition, 364
enrollment period, 366
exposure to treatment measurement, 365–366
identification of cases, 364–365
matching controls to cases, 368–371
observation period, 366–368
outcomes analysis, 373–377
outcomes measurement, 371–373
overlap in, 378
person-time, 366, 377
time-to-event analysis, 377–378
verification of matching, 373
Matching, adequate, 343
Matrix format, of data merging, 12
Maximum–minimum hill-climbing algorithm, 511
Maximum values, 22
Mean, 82
arithmetic, 82
sampling distribution, 141–144
weighted, 82–83
comparison of, 135–172
in propensity scoring, 331–332
dispersion of data around. *See* Standard deviation
of normal distribution, 137
population (μ), 82
sample *(X),* 82
Median
fourth spread, 223–224, 226–227
of normal distribution, 137
Tukey's control limits, 224, 226–227

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Medical centers, cancer survival rates, 355–358
Medical errors
cumulative distribution function, 204
gradation, 206
hazard rate, 243–244
probability density function, 204, 205
time to, 204–205
Tukey's charts of, 232–233
Medical foster homes (MFHs)
case identification, 365
cost-of-care analysis, 264–271
logarithm transformation, 290–291, 292, 293
with single cost predictor, 266–271
definition, 264–265
matched case control study, 365–366
normal probability plots, 289
propensity scoring, 333–337
logistic regression analysis with, 313–319
Medicare, 310
health maintenance organizations (HMOs), 311
Medication errors, 43
conditional probabilities, 463–464, 465
EHR-related decrease, 3
Medications
advertising of, 311
comparative effectiveness studies, 4
data analysis, 2
as diagnostic indicators, 42
EHR test fields for, 30
generic, 346
pharmacovigilance of, 488
postlaunch effectiveness of, 2
propensity scoring, 328
Metformin, 42
MFHs. *See* Medical foster homes
Michigan Medicine, 328
Microsoft Excel
Analysis ToolPak Histogram tool, 96–97
binomial probability distribution, 176, 177–178, 181
central limit theorem simulation, 142–143
contingency tables, 68–71
correlation calculations, 74–75, 475
histograms, 94–97
hypothesis testing, 183, 186, 187
law of large numbers simulation, 142–143
linear transformation of variables, 89–90
logarithm transformation with, 290, 292
median function, 226
pivot table feature, 69–71
regression analysis, 266–271, 300
scatter plots, 74, 260–262
Microsoft SQL Server Management Studio, 21
Minimum values, 22
Missing data, 13, 41–42
Mode, of normal distribution, 137
Moffitt Cancer Center, Personalized Medicine Institute, 2
Monotone relationships, 437
Morbidity scales, 80. *See also* Multimorbidity (MM) index
Mortality, as binary event, 80
Mortality risk. *See also* Multimorbidity (MM) index
definition, 102
six-month, countable discrete levels of, 174
Multicollinearity, 294–295
Multilevel regression, 345–360
applications, 345–346
assumptions, 348
coefficient models, 347
corner cases, 355
examples, 349–354

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

intercept regression, 347–348, 352–358
SQL code, 354–358
interclass correlation (ICC) coefficient, 348
of longitudinal data, 358
macrolevel (practice level), 346–347, 348, 353–354
measurement issues, 358–359
microlevel (patient level), 346–347, 348, 349–353
stratification in, 354–358
Multimorbidity (MM) index
accuracy, 105, 112–118, 119
adjustment for ICD-10, 108–109
alternatives, 102–105, 117–118, 119
Bayes data mining model, 105–106, 111–112
comparison to other diagnosis-based indexes, 117–118, 119
comparison to physiological markers, 117, 118
cross-validation, 109–110
estimation of parameters, 106
example of use, 119, 120
likelihood ratios, 106–108, 111
combination of diseases adjustment, 107
confounding, 111–112
detection applications, 110
diseases with no or complete mortality, 107–108
multimorbidity score, 119
posterior odds, 119, 120
predictive applications, 110–111
rare diseases, 108
repeated diseases adjustments, 106–107
of zero value, 111
recommendations for improvement, 120–121
sample size, 109
sensitivity, 112–114
specificity, 112–114
SQL coding, 125–131
theoretical basis, 105–106
time interval calculations, 112
Multiple regression, 266–271
causal networks as, 499–501
definition, 258
equations, 261
interaction effect, 274–277
main effect, 274–277
terminology, 491, 492–493
Multiplication rule, 59–60
Multivariate regression, 6, 9
independent variables in, 110
multilinearity, 476
terminology, 491, 492–493
Mutual information, 477–478, 479
conditional, 478
Mutually exclusive events, 189
Myocardial infarction, length of stay, 413–416

National Institute of Mental Health, 338–339
Natural logarithm function, 313–314
Near-miss events, logistic regression of, 310–311
Network modeling, 487-488. *See also* Association networks; Causal networks
as extension of regression, 488
stratification in, 385
Never events, 3
Noise, random, 109
Nonlinear equations, 260
transformation, 260–261
Normal distribution, 90, 136–144
bell-shaped, 136, 137
in control charts, 152
definition, 136
examples, 136–139
properties, 136–137
standard, 138–139
examples, 139–140
Normality assumption, 152

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Normal probability plots, 289
Null hypothesis, 73, 144–145
  alternative hypotheses, 144–145
  confidence intervals, 150–151
  critical value approach, 150
  definition, 144
  failure to reject, 146, 147–148
  p-value approach, 150
  rejection, 146, 279
    in matched case control studies, 373
    statistical significance of, 149
Null model, 481
Null values, 21
Numbers
  conversion from text, 44–45
  conversion to text, 31
Nurse retention/turnover, 310, 345–346
Nursing Home Compare website, 311
Nursing homes, propensity scoring, 328–329, 333–337

Obama, Barack, 250
Observational data, propensity scoring, 327–344
Observation per person, 52
Observations
  independence of, 52
  over-time periods. *See* Tukey's charts
  single, 223
  weight of, 82
Observed variables, in weighed regression, 294
Odds
  Bayes's formula, 62–66
  of mortality, 112
  posterior, 63, 65–66, 112
  as ratios, 62–63
  relationship to probability, 62
Odds ratio
  common, 386–389
  confidence intervals, 188–189, 374–376
  definition, 186
  of mortality, 104–105
  of observed outcomes, 373–377
  in stratified covariate balancing, 386–389
One-sample tests, 152
One-sample *t* test, 148, 149
One-sample *z* test, 147–148
One-sided tests, 139
Operating room fires
  geometric distribution, 206
  hazard rate, 242–243
Order, of records, 36–38
Ordinal variables, 79, 80
Outcome
  association with diagnosis, 110
  causal relationship with treatment
    back-door path, 507–510
    numerical example, 503–506
    probability prediction, 501–503
    regression and, 497–501
  control charts of, 98–99, 153
  definition, 492
  diagnosis after, 46
  impact of treatment on, 48
  in matched case control studies
    analysis, 373–377
    measurement, 371–373
  multiple, 487-488. *See also* Causal networks
  patient-reported, 43–44
  predicted and actual, 323
  therapeutic ranges, 371–373
  trend analysis of, 98–99
  variables for measurement of, 80
  in Veterans Health Administration system, 3
Outliers, detection of, 144
Out-of-range data, 42
Overfitting, of data, 281
Overlap, case-control
  in benchmarking, 416–418
  in matched case control studies, 378
  in propensity scoring, 343

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

in stratified covariate balancing, 398–403
calculation, 398
definition, 398
Markov blanket of treatment, 399, 401–403
partial matches, 398–403
synthetic controls, 400

Pain levels, patient-reported, 43–44
Pain medications, 224–227
Paired samples test, 151–152
Paired *t*-test, 333, 373
Parabola, equation for, 260
Parametric distribution, 90
Parents, of Markov blankets, 492, 493, 496, 499–501
Path, definition, 492
Patient-centered medical homes (PCMHs), 328
Patient online review analysis, 82–83, 212–219, 221
Patient satisfaction
network models of, 488–489
with nursing home care, 329
as ordinal variable, 80
patient online review analysis, 82–83, 212–219, 221
propensity scoring, 328, 329
stratified covariate balancing of, 387–389
X-bar control chart measurement, 158–162
Patient tables, creation of, 17–21
Paxil, 4
Pay-for-performance schemes, propensity scoring of, 328–329
PCMHs. *See* Patient-centered medical homes
P (probability) control charts, 5, 189–194
control limits calculations, 190–194, 210
errors in display of, 194
lower control limits, 192–193, 194
observations per period, 236
observed rates, 193–194
risk-adjusted, 194–199, 201
expected deviance, 195, 197–199, 200
expected rate, 198, 199, 200
lower control limit, 195
t-statistic, 198, 199
upper control limit, 195
upper control limits, 192–193, 194
x-y plots, 190–191, 194
Pearl, Judea, 507
Pearl's collider test, 515
Pearson correlation, 74
Pharmacovigilance, 488
Physician performance. *See* Benchmarking, of physician performance
Physiological markers, as prognostic indicators, 117, 118
Plots, 153
Poisson distribution, 90
Poisson regression, 258, 478–484
for association network construction, 481–484
response variables in, 480
Population mean, one-sample *z* test of, 147–148
Population of interest, 80–81
Pravastatin, 4
Prediction, causal, 501–506
Predictive medicine
matched case controls of, 363
variables of, 281
Predictive models
evaluation, 46
training-data set, 46
validation-data set, 46
Predictors
discarding of, 45
in multimorbidity models, 110–112
obvious, 45
rare, 45
relationship to outcomes, 46
single cost, 266–271

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Presidential election (2016), 250, 252
Price, Richard, 63
Pricing violations, 209
Probability, 55–62
addition rule, 59, 60
calculus of, 58–61
causal network-based calculation, 501–506
conditional, 61–62, 67–68
Bayes's formula for, 62–66
independence and, 64–66, 461, 464, 465, 514, 515
joint probability-based calculation, 464–467
marginal probability-based calculation, 466–467
of death, 80
decimal expression of, 56, 62
definition, 56, 58, 461
empirical, 56, 57
expected values in, 84–85
frequency distribution, 58
graphical representation, 59–61
of joint events, 464, 467
marginal, 64–66, 466–467
multiplication rule, 59–60
odds-based calculations, 66
posterior, 105
predicted, comparison with events, 43–44
random variations, 71–73
relationship to odds, 62
subjective or personal, 56, 57
theoretical, 56–57
Probability control charts. *See* P (probability) control charts
Probability density functions, 241, 242–243
Bernoulli, 205–207
binomial, 205
geometric, 205, 206–208
Poisson, 205
Probability functions, 57–58
Probability networks, strata-conditioning and, 463
Process control, 9
Process improvement, matched case control use in, 362–363
Productivity, of data-driven organizations, 1, 3
Prognosis
applications, 102
definition, 80, 102
in multiple morbidity. *See* Multimorbidity index
Propensity scoring, 7, 384, 410–411
applications, 328–329
comparison with stratified covariate balancing, 392–397
data balancing in, 330–332
double regression in, 338
extreme weights in, 343
interaction terms, 331
inverse probability of treatment weighing (IPTW), 337–342
logistic regression with, 313–319, 338–339
of medical foster homes, 333–337
with logistic regression analysis, 313–319
overlap in, 343
quintile matching, 332–333 343
as a simulation, 329–331
steps, 330–331
verification of propensity scores, 342–343
Propensity to participate in treatment, 329
Provider networks, 489
Provider tables, 19, 20–21
Pseudo-R^2, 323
p-value, 149, 295
p-value approach, 149, 150

Q–Q plots, 283–285, 288, 289
Quality control, 136
matched case control use in, 362–363
Quality improvement, 153
Quality of care

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

accountability for, 4
of data-driven organizations, 2–3
measures, 3
Quality of life, multilevel modeling of, 346
Quartiles, fourth spread, 223–224, 226
Quintiles, in propensity score matching, 332–337, 343

R (software)
anova function, 279–280
correlation calculations, 475
cost data log transformation, 290–291
definition, 300
downloading of, 300
errors in, 302
heteroscedasticity, 287
linear Poisson distribution models, 481–485
logit transformation, 318–319
propensity score matching, 333–336
regression analysis tools, 300–307
Shapiro-Wilk test of normality, 289
stratified covariate balancing, 406
weighted propensity scoring, 339–343
weighted regression performance, 294
R^2 coefficient, 277–278, 281
Race, as variable
as categorical variable, 79
countable discrete levels, 174
Randomization, for removal of confounding, 384
Random noise, 109
Random sampling, 81
Random seed values, 110
Rank order functions, 36–38
Rare events. *See also* Sentinel events analysis
geometric distribution-based probability, 208
Rates, comparison of, 173–201
Bernoulli distribution, 175–179
binomial probability distribution, 175–179
comparison of two rates, 183–186
confidence interval for odds ratio, 186–189
discrete variables summarization, 174–175
inference for a single rate, 183–186
normal approximation, 179–181
p- (probability) control charts, 189–201
control limits calculations, 190–194
errors in display of, 194
lower control limits, 192–193, 194
observed rates, 193–194
risk-adjusted, 194–199, 201
upper control limits, 192–193, 194
x-y plots, 190–191, 194
statistical significance, 181–183
Ratios, odds as, 62–63
Readmission
effect of hospice care on, 48–52, 466–467
rates, 4
Reasoning, causal, 515
Receiver operating curve (ROC), 114, 323
Regression, 255–307. *See also* Multiple regression; Multivariate regression; Stratified regression
applications, 256–258
cause-or-effect interpretation, 296–297
collinearity effects, 291–292
confounding in, 498
Cox's hazards, 258
cross-validation of, 292–293
definition, 256, 261
error terms

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Regression (*continued*)
heteroscedasticity of, 286–287
homoscedasticity of, 286
normal distribution of, 288–289
Excel use, 266–271
forward, 280–281
hierarchical, 280–281
logistic, 258
model building, 280–281
multicollinearity of, 294–295
ordinary/standard, 258
log transformation in, 315–317
parameters
effect of interaction terms on, 276
estimation, 295
tests of, 255
for prognostic predictive models, 102
relationship between causal networks, 497–501
residuals, 262–264
autocorrelation, 285–286, 287
diagnostic plots, 282–283
squared, 264
tests of parameters of, 256
types, 258
weighted, 294, 337–343
Regression coefficients, 262
collinearity and, 291–292
in cost data evaluation, 268–270
definition, 262
in hypothesis testing, 295–296
interaction terms and, 276–277
multiple variables and, 428–429
stratified covariate balancing, 427–428
in stratified regression, 427–458
impact of correction factors, 434–435
impact of independent variables, 433–434
stratified regression equation, 436
unconfounded impact, 427–458
units of measurement, 296
Regression equations, 256–258, 259–264
multilinear form, 429
network representation, 497, 498
Repeated measures test, 151–152
Resampling, 390–391
Reserved words, 22
Residuals, 262–264
autocorrelation, 285–286, 287
diagnostic plots, 282–283
squared, 264
Restricted maximization algorithm, 511
Rise, definition, 259
Risk adjustment, of control charts, 194–199, 201
Risk assessment, 8–9
with binary data, 203–204
of health insurance companies, 3
of mortality. *See also* Multimorbidity (MM) index
differential point systems, 103–104
selective methods, 104
ROC. *See* Receiver operating curve
Rosenbaum, Paul R., 329–330, 384
Rubin, Donald B., 245, 329–330, 384
Run, definition, 259

Samples and sampling
adaptive, 81
complete, 81
convenience, 81
not representative, 82–83
random, 81
representative, 80–81
Sample size, 466
distribution of mean and, 141–144
in sentinel event analysis, 203
Scatter plots, 73–74
creation with Excel, 260–262

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

creation with R, 304–305
Seed values, 110
Sentinel events analysis, 203–222
Bernoulli distribution function, 205–209
cumulative distribution function, 204
days to event, 208–209
expected value, 204–205
geometric distribution function, 206–209
probability density function, 204
with time-between control charts, 209–212
exercise resolution example, 219–220
patient reviews example, 212–219, 221
Severity of illness, 80, 102
Shapiro-Wilk test of normality, 289
Shewart, Walter A., 156
Shewart charts, 156
Significance, statistical, 149
Simdata, 294
Skewed distribution
binomial probability distribution, 179
log transformation, 97
Tukey's chart sensitivity to, 236–237
Slope
calculation of, 259–260
definition, 259
Slope coefficient, 295
Southeast Alabama Medical Center, 225–227
SQL (Structured query language)
definition, 14
versions of, 14
web-based searching of, 14
SQL (Structured query language) codes, commands, and functions, 6, 8–9, 11–53
BY, 13
FROM, 16–17
INTO, 39–40
BETWEEN, 42
for benchmarking, 422–424
CAST, 44–45
CONCAT, 30–31
for conditional probability calculations, 462–463
CONVERT, 35–36, 44–45
CREATE TABLE, 17–21
data control language, 14
data definition language, 14
data functions, 33–36
data manipulation functions, 32–33
data manipulation portion, 14
data merging function, 5–6, 12
DATEADD, 33–34
DATEDIFF, 33, 34–35
DATEPART, 33, 34
FROM dbo.data, 16
referencing temporary tables, 16–17
for deletion of erroneous data, 38–45
GETDATE, 33
GROUP, 13
GROUP BT ID, 39–40
GROUP BY, 22–23, 24, 37, 38
HAVING, 24–25, 39
IIF manipulation functions, 33
INSERT VALUE, 17, 21–23
for intercept regression modeling, 353–358
JOIN, 13
for joining of tables, 25–29
full join, 25, 28–29
inner join, 25–27
join statements, 26–27
left or right join, 25, 27–28
no join (cross join), 25, 29
for logit transformation, 315–316
manipulation functions, 13
Microsoft SQL Server Management Studio, 21

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

SQL (*continued*)
for multimorbidity index, 106, 125–131
ICD-9-based, 106, 108, 109
ICD-10-based, 106, 108–109
likelihood ratio calculations, 106, 107–108, 111
sensitivity and specificity measures, 112–114
for mutual information calculation, 479
NULL VALUES, 21
ORDER BY, 22, 38
for prognostic predictive models, 102
random seed values, 110
RANK, 36–38
RANK_DENSE, 36–38
rank order functions, 36–38
risk measurement, 195
SELECT, 13, 15–23
field name deletion with, 16
purpose, 15
reserve words, 15
TOP 20* FROM #temp, 16–17
SELECT ID, 39–40
standardized functions, 14
for stratified covariate balancing, 395–397, 403–406
for stratified regression, 439
confounded impact of variables, 433–434
correction factor estimation, 434–435, 446
k constant, 439, 450
STUFF, 32–33
text functions, 30–33
time to pain medication, 225–227
USE Database 1, 16
WHEN, 13
WHERE, 22, 23–25, 37, 39–40, 47
Square root transformation, 290
Squiggly symbol, 387
SSE. *See* Sum of squares of errors
SST. *See* Sum of squares total
Standard deviation
calculation, 86–87
definition, 85–87
weighted, 87
Standardized normal distributions, 5
Statistical analysis, steps in, 78
Statistical process control, 3, 5
Statistical significance, 145, 149
Stock market prices, causal control chart analysis, 250, 252
Straight line equation, 259–260
Strata/stratum, definition, 355, 386
Strategic planning, matched case control use in, 362
Strategy, logistic regression analysis of, 311
Stratification, 81
in control chart construction, 245
definition, 386, 492, 493
history, 385
in multilevel modeling, 354–358
in network modeling, 385
relationship to conditioning, 463–464
as subgroup analysis, 386
Stratified covariate balancing, 383–407, 463
in benchmarking, 409, 411
case-control overlap, 398–403
calculation, 398
definition, 398
Markov blanket of treatment, 399, 401–403
partial matches, 398–403
synthetic controls, 400
of causal networks, 517
comparison with propensity scoring, 392–397
of continuous outcomes
difference models, 389–390
weighted data, 390–392, 393–395
definition, 384

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

examples, 392–395, 392–397
of patient satisfaction, 387–389
in propensity scoring, 332
SQL code, 403–406
SQL codes, 395–397
Stratified regression, 427–458
multilinear form, 429–430
calculation of parameters, 430–436
comparison with multiplicative form, 436–437
correction factors, 429–430, 434–435
definition, 429
example, 430–436
impact of independent variables, 434–436
multiplicative form, 446–447
case and control strata, 438–440
comparison with multilinear form, 436–437
corner stratum, 437–440, 442
correction factors, 446
estimation of parameters, 437–439, 447–450
examples (health insurance cost), 430–436, 441–450
examples (lung cancer prognosis), 439–458
joint preferential independence condition, 437, 442
k constant, 439, 442, 444
SQL (lung cancer prognosis), 447–450, 452–458
Structured query language. *See* SQL (structured query language)
Student's *t*-distribution, 148, 236
with four degrees of freedom, 236
for mortality risk, 195, 199
tables, 195
for treatment effects, 390
Study design, 6, 48–51
case-control design, 48, 49, 51, 52
cohort design, 48–51, 52
observation per person, 52
unit of analysis, 52
Sturges formula, 92–93
Subsets, selection of, 23–25
Substrings, 30
Suicide risk model, 490
Sum of squares, R anova function, 279–280
Sum of squares of errors (SSE), 279
Sum of squares total (SST), 278–279
Surgery
robotic, propensity scoring of, 328
wrong-side, 203, 208, 250
Survival function, 242–243
Switching distribution, 410, 411–413
Synthetic controls, 410, 416, 418–420, 421
Synthetic minority oversampling technique, 416, 418

Tables, of EHRs
encounter, 19–21
foreign keys, 15
joining of, 25–29
full join, 25, 28–29
inner join, 25–27
join statements, 26–27
left or right join, 25, 27–28
no join (cross join), 25, 29
multiple, 5–6, 12
patient, 17–21
patient fields, 15
primary key, 15
provider, 19, 20–21
relationships among, 15, 20–21
Taboo algorithm, 511
Tercero-Gomez, Victor, 236–237
Tertiary care centers, survival rate analysis, 349–354
Text, conversion to dates, 35–36
Text fields, combing of, 30–31
Text processing, 2
Therapeutic ranges, of outcomes, 371–373

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Tilde symbol, 387
Time periods, for observations, 223–227
Time-stamped data, 47–51
Time to pain medication (OP_21) measure, 224–227
Training-data set, 46
Treatment, definition, 491, 492
Treatment impact on outcome
  back-door path, 507–510
    blocked, 493, 509–510, 517
    covariates, 507, 508–509
    definition, 491, 493
  stratified covariate balancing
    binary outcomes, 386–389
    continuous outcomes, 389–395
    difference models, 389–390
    weighted data, 390–392, 393–395
Treatment participation. *See* Propensity scoring
Trend line regression, 260–262
Trump, Donald, 250, 252
t-statistic, 148, 295, 373, 390
t-tests, 152
t-tests
  one-sample, 148, 149
  paired, 333, 373
Tukey, John, 223–224
Tukey's control charts, 210, 223–228
  comparison with other charts, 236–237
  confidence interval limits, 223–224
  control limits calculations
    with fourth spread, 224, 226
    lower control limits, 224, 231, 233, 234, 235, 236
    with post-intervention period data, 229, 230
    with pre-intervention period data, 229, 230, 231
    tightness of, 229
    upper control limits, 224, 226, 229, 230, 231, 233, 234, 236
    without post-intervention period data, 232–233, 234
    without pre-intervention period data, 232–233, 234, 235
  examples
    budget variations, 233–235
    exercise time/weight control, 227–232
    medical errors, 232–233
    time to pain medication, 224–227
  fourth spread calculations, 223–224, 226, 229, 230–231, 232–233
  gamma distribution sensitivity, 236–237
  observed to expected values comparison, 233–235
  reference point, 234
Two-sided tests, 139
Type I errors, 145, 146
Type II errors, 145, 146–147

Uniform distribution, 90
US Department of Veterans Affairs
  cancer comorbidities analysis, 518–519
  eating disabilities–mortality analyses
    causal analysis, 507–509, 520, 522
    stratified covariate balancing, 392–395
  patient outcomes, 3
  performance measurement in, 3
  progression of disabilities analysis, 511–512
  suicide risk model, 490

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

Veterans Affairs Informatics and Computing Infrastructure (VINCI), 2
US Food and Drug Administration (FDA), 2
Unit of analysis, 52
Units of measurement, 296
Univariate data analysis, 78
Univariate methods of inference, 6
Universe of possibilities, 461–463, 461–464, 466, 467
University of California, medical centers' database, 2

Validation-data set, 46, 293
Value-based reimbursement, 4, 186, 188
propensity scoring of, 329
Values, expected, 83–85
Variable character data type, 35
Variables. *See also* Binary variables; Dependent variables; Discrete variables; Independent variables
association of, 7
constant, 78
contingency table–based relationships, 66–71
correlation between, 74–75
counterfactual effects of, 7
definition, 78
dummy, 69, 271
examples, 78
expected values, 83–85
fluctuations of, 135–136
forward stepwise selection, 281
interval, 79, 87
levels of, 78–80
probability of observation, 90–92
linear transformation of, 87–90
mechanism of, 7
multicollinearity, 294–295
from multiple tables, 12
new, calculation of value, 30–33
nominal, 80
optimal class interval size, 92–93, 94
ordinal, 79, 80
ratio, 79–80, 87
restriction of number of, 294–295
sequence of, 7, 514–516
standard deviation of, 85–87
values over time, 98–99
X, 81
Variance, 86, 87
of sum, 88
Variation
random, 244
with special or assignable causes, 244
Veterans Affairs Informatics and Computing Infrastructure (VINCI), 2, 441
Veterans Health Administration. *See* US Department of Veterans Affairs

Wald test, 320–321
Weighted covariates, 390–392, 393–395
Weight loss, Tukey's charts of, 228–232
White test, 288
Wilcoxon signed-rank test, 373

X^2 test, 145
X-bar control charts, 152, 153, 158–171
assumptions of, 160–161
comparison with Tukey's charts, 236
distribution of findings, 171
example, 158–162
lower control limit, 159–162, 168, 170
risk-adjusted, 162–171
upper control limit, 159–162, 168, 170–171

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com

XmR control charts, 152, 153, 156–158, 210
comparison with Tukey's charts, 236–237
lower control limit, 158
outliers, 236
Shewart charts, 156
upper control limit, 157–158

z statistic, 148
z tests, 147–148, 152
one-sample, 147–148

Copying and distribution of this PDF is prohibited without written permission.
For permission, please contact Copyright Clearance Center at www.copyright.com