

# Index

Page numbers followed by “f” indicates figures and “t” indicates tables.

## A

Absolute values, 733–734  
 of sample correlation coefficients, 126–127  
 Addition rule of probabilities, 154–155  
 definition of, 154  
 example of, 155–156, 163  
 summarized, 199  
 Additive property of normal random variables, 279–281  
 problems for, 281–284  
 Algorithms, for random sample generation, 736  
 $\alpha$  test of significance levels  
 definition of, 391  
 types of tests (summary tables), 391, 406t, 417t, 450t, 459t, 468t, 488t  
 Alternative hypothesis  
 definition of, 389  
 establishing, 391  
 types of tests (summary tables), 406t, 417t, 429t, 450t, 459t, 468t, 488t  
 Analogous hypothesis, in two-factor ANOVA tests, 521  
 Analysis of variance (ANOVA)  
 column factors in, 515  
 double-summation notation, 517  
 Fisher, invented by, 504, 529  
 grand mean, 516  
 introduction to, 504–505  
 key terms for, 530  
 one-factor, 504, 505–510  
 problems for  
 one-factor tests, 510–514  
 two-factor, 518–520, 527–529

review problems, 533–536  
 row factors in, 515  
 summary of, 530–533  
 two-factor, 514–518  
 two-factor, testing hypotheses, 520–527  
*Analytical Theory of Probability* (Simon), 310  
 Approximation rule  
 empirical rule and, 268  
 normal curves and, 269f  
 for normal random variables, 268  
 Arbuthnot, John, 430  
*Ars Conjectandi* (Bernoulli), 244  
 Averages, weighted, 76

## B

Bacon, Francis, 457, 672  
 Bar graphs, 19–21  
 example of, 20f  
 first use of, by Guerry, 59  
 relative frequency, 24f  
 summary of, 60, 61f  
 and symmetry, 21f  
 Bayes' theorem, 185–187  
 conditional probability and, 185–186  
 definition of, 186  
 false-positive results and, 187  
 problems for, 187–189  
 Bell, E.T., 292  
 Bell-shaped curves. *See also* Normal curves  
 density curves, 266–267  
 empirical frequencies and, 305  
 Bernoulli, Jacob, 9, 244  
 Bernoulli, Jacques, independent trials and, 244  
 Bernoulli, Jean, 244

Bernoulli, Nicholas, 244  
 Bimodal data sets, 112  
 histograms of, 112f  
 Binomial distributions, 749–754  
 of hypergeometric random variables, 249  
 normal approximation to, 317–319  
 Binomial parameters, hypothesis test of, 423–425  
 Binomial probabilities, 241f  
 hypothesis tests, 488t  
 Binomial random variables, 238–243  
 definition of, 238  
 expected value and variance of, 243–244, 316  
 independent trials for, 239  
 with parameters  $n$  and  $p$ , 238, 239  
 Poisson random variables and, 250–251  
 probabilities, with various parameters, 241f  
 probability formula, 239  
 probability mass functions of, 318f  
 problems for, 244–247  
 standard deviation of, 316  
 Biological data sets, normal distribution of, 569–570  
 Biometrics, 127–128  
 Birth rates  
 histograms for, 37f  
 history of mortality tables, 7–9  
 Blood cholesterol levels, 33t  
 frequency tables of, 33t  
 Box plot, definition of, 104  
 Braudel, F., 59  
 Burt, Cyril, 606

## C

- Categorical data, 590–592
- Causation, correlation and, 128
- Center of gravity, 78, 79f, 221
- Central limit theorem, 115, 304–311
  - approximately normal data sets and, 306
  - definition of, 304
  - error measurement in, 305–306
  - historical perspective on, 307, 310
  - normal curves and, 306
  - problems for, 311–313
  - sample mean distribution and, 306–310
  - sample sizes for, 310–311
    - for various populations, 307
- Chance variation, 700
- Chi-squared density function, 324f
- Chi-squared distributions, degrees of freedom of, 324, 360
- Chi-squared goodness-of-fit tests
  - concepts, 609–615
  - definition of, 610
  - independence in contingency tables with fixed marginal totals, 631–634
  - independence of two
    - characteristics, 620–626
    - contingency tables, 622
    - degrees of freedom, 623
    - summary table, 626t
  - introduction to, 606–609
  - key terms for, 637
  - p* value of, 612
  - Peterson's analysis, 608, 610, 616
  - problems for, 615–620
    - contingency tables with fixed marginal totals, 634–637
    - independence of two
      - characteristics, 626–631
    - review problems, 640–645
    - summary of, 638–640
    - summary table, 615t
    - testing null hypothesis in, 610
- Chi-squared percentiles, 611f, 741–742t
- Chi-squared random variables, 323–325, 360
- Class boundaries, 32
- Class intervals, 32
- Coefficient of determination, 578–580
  - definition of, 579
  - problems for, 580–581
- Column factors, in ANOVA, 515
- Combinations, 193
- Comparison rankings, 685–688
- Complements, 149
- Conditional probability, 167–176
  - Bayes' theorem and, 185–186
  - definition of, 167–168
  - problems for, 177–184
- Confidence
  - centered intervals, 353f
  - definition of, 347
  - interval estimators, 347, 349, 362, 371
  - level percentiles of, 350t
  - 90, 95, and 99 percent of, 350t, 351f
- Confidence bounds
  - for interval estimation of population proportions, 375–377
  - upper and lower, 355–357, 364–366
- Confidence intervals, lengths of, 373–375
- Constants and properties of
  - expected value, 221
  - sample variance, 102
- Contingency table
  - definition of, 622
  - independence testing, 620–626
    - with fixed marginal totals, 631–634
  - problems for independence testing, 626–631, 634–637
- Continuity correction, 319
- Continuous distributions, 262
- Continuous random variables, 262–264
  - definition of, 262
  - probability density function in, 262–263
  - problems for, 264–266
- Control charts
  - cumulative sum, 722–724
  - EWMA, 717–721
  - for fraction defective, 715–716
  - problems for
    - cumulative sum, 725
    - EWMA, 721–722
    - for fraction defective, 717
- Control/control group, 4
  - Arthur Young (historical perspective), 457
  - hypothesis testing, 455, 486
  - normal distributions and, 704
  - samples as, 457
- Correlations. *See also* Sample correlation coefficient
  - associations measured by, 128
  - negative, 121
  - positive, 121
- Counting arguments, 194
- Counting principles, 189–195
  - basic, 189
  - generalized basic, 190
  - notation of, 193
  - problems for, 195–198
- Course of Experimental Agriculture* (Young), 457
- Critical region, definition of, 390
- Cumulative relative frequency table, 43
- Cumulative sum control charts, 722–724
  - problems for, 725

## D

## Data

- approximately symmetric, 21
- collection, 3–4
- detected by histograms, 35f
- manipulation of, and scientific fraud, 606
- modern approach to, 2–3
- paired, 51–54
- symmetric, 20–21
- Data mining, 408
- Data sets, 729–732
  - approximately normal, 110f
  - bimodal, 112
  - biological, 569–570
  - central limit theorem and, 306
  - central tendencies of, 84
  - comparison rankings, 685–688
  - constructing histograms from, 34
  - finding sample variance for, 100–102
  - frequency tables and graphs, 18–32
  - grouped data and histograms, 32–43
  - histograms of, 109f, 110f
  - introduction to, 18
  - key terms for, 59–60

- normal, 73, 109f, 138
  - paired data, 51–58
  - problems for
    - frequency tables and graphs, 25–32
    - grouped data and histograms, 39–43
    - paired data, 54–58
    - stem-and-leaf plots, 47–51
  - review problems, 63–70
  - skewed, 110f, 111f
  - stem-and-leaf plots, 44–51
  - summarizing. *See* Statistics
  - De Mere, Chevallier, on probability, 157
  - De Moivre, Abraham, normal distribution and, 262, 263–264
  - Degrees of freedom
    - of chi-squared distributions, 324, 360
    - definition of, 323
    - error random variables, 554–555
    - one-factor ANOVA, 506–508
    - of random variables, 360
    - remarks on  $F$  random variable, 507–508
    - values of  $F$  and, 507t
  - Densities
    - of sample means, 300f
    - symmetric about zero, 658f
  - Density curves. *See also* Probability density function
    - of uniform random variables, 265f
  - Density percentile, in interval estimations, 361f
  - Dependent variables, 539
  - Descartes, René, 58
  - Descriptive statistics, definition of, 4
  - Deviations
    - definition of, 78–79
    - historical perspective on, 78
    - $i$ th deviation, 135
  - Dice, fair, 167
  - Discrete random variables
    - binomial, 238–243
    - concepts, 211–214
    - definition of, 210, 212
    - expected value, 218–225
    - hypergeometric, 248–249
    - key terms for, 254
    - Poisson, 250–253
    - probability distribution and, 215
    - problems for
      - binomial random variables, 244–247
      - expected value, 225–231
      - hypergeometric random variables, 249–250
      - Poisson random variables, 253–254
      - random variables, 215–217
      - variance, 236–238
    - review problems, 256–259
    - summary of, 254–256
    - variance of, 231–236
  - Disjoint events, 149, 150
  - Distributions
    - central limit theorem and sample mean, 306–310
    - chi-squared, 324, 360
    - continuous, 262–263
    - key terms for, 325–326
    - preview of, 298
    - review problems, 327–330
    - of sample variance of normal populations, 323–325
    - summary of, 326–327
  - Doll, R., 72
  - Don-Hill study, 72
  - Double-summation notation, 517
  - Dummy variables for categorical data, 590–592
- E**
- Empirical frequencies, bell-shaped curves and, 305
  - Empirical rule, 109–114
    - approximation rule and, 268
    - definition of, 111
    - historical perspective on, 115
    - normal data sets and, 114
    - problems for, 114–120
  - Equality testing
    - of equality of means
      - known variances, 446–450
      - small-sample, with equal unknown variances, 463–468
      - unknown variances and large sample size, 453–459
    - of multiple probability distributions, 683–688
    - of population proportions, 481–489
  - Equally likely outcomes, 161–164
    - random selection, 162
  - Error random variable, 553–555
    - multiple linear regression model, 587–588
    - problems for, 556
  - Error sum of squares, in two-factor ANOVA tests, 522
  - Estimated regression line
    - definition of, 545
    - scatter diagram of, 547f
  - Estimates, definition of, 332
  - Estimation
    - interval estimators of mean, 347–371
    - interval estimators of population proportion, 371–380
    - introduction to, 332–333
    - key terms for, 380–381
    - point estimator for population mean, 333–334
    - point estimator of population proportion, 336–338
    - of population variance, 342–344
    - of probability of sensitive events, 341–342
    - problems for
      - interval estimators, known variance, 357–359
      - interval estimators of population proportion, 377–380
      - interval estimators, unknown variance, 366–371
      - point estimator, population mean, 334–336
      - point estimator, population proportion, 336–338
      - population variance, 344–347
      - sensitive event probability, 342
    - review problems, 383–386
    - summary, 381–383
  - Estimators
    - case studies for, 373–374
    - control charts and unknown mean and variance, 707–709
    - definition of, 332
    - least-square, 544–545
    - in one-factor ANOVA, 505–506
    - point, of population mean, 333–334
    - population variance, 463–464
    - in two-factor ANOVA, 521, 523
    - unbiased, definition of, 333
  - Events, 146–150
    - complements, 149

Events (*continued*)

- definition of, 148
- disjoint or mutually exclusive, 149, 150
- independent, 167–176
- intersection/union, 148–150
- null, 148–149
- EWMA control charts. *See* Exponentially weighted moving-average control charts
- Expectation, definition of, 218
- Expected value, 218–225
  - of binomial random variables, 243–244, 316
  - center of gravity analogy, 221
  - of chi-squared random variables, 324
  - definition of, 218
  - frequency interpretation of probabilities, 218, 224
  - of hypergeometric random variables, 248
  - of Poisson random variables, 252
  - population means and, 301, 302
  - problems for, 225–231
  - properties of, 221–225
  - of sample means, 300
  - of sums, 222, 224
  - of sums/products using constants, 221
  - of zero, 231
- Experiment, 146–150
  - definition of, 146
  - equally likely outcomes in, 161–164
  - problems for, 150–153, 164–167
- Exponential bell-shaped curve, 291
- Exponential distribution, 310
- Exponential random variables, density of average of, 310f
- Exponentially weighted moving-average (EWMA), 719
- Exponentially weighted moving-average control charts, 717–721
  - definition of, 718
  - example of, 721f
  - problems for, 721–722
  - standard, definition of, 719

**F**

- F* distribution
  - definition of, 506–507
  - degrees of freedom, 507–508
  - example of, 507f
  - percentiles, 743–748
- Factorial notation (!), 191
- Fair dice, 167
- False-positive results, Bayes' theorem and, 187
- Fermat, Pierre, 157
- Finite populations
  - problems for, 319–323
  - proportions in, 314
  - random variables in, 315
  - sampling proportions for, 313–319
- First quartile, definition of, 93
- First-generation hybrids
  - crossing, 607f
  - described, 606
- Fisher, Ronald A.
  - analysis of Mendel's data, 608
  - ANOVA and, 504, 529
  - role, in history of statistics, 10
  - on sample correlation coefficients, 128
  - significance levels and, 393
  - t* test and, 418
- Fixed margins, 631
- Fraction defective
  - control charts for, 715–716
  - problems for control charts for, 717
- Fraud and data manipulation, 606
- Freedman test, 688
- Frequency histograms, 34
  - example of, 34f
  - summary of, 60
- Frequency interpretation, of expected value, 218, 224
- Frequency polygons, 19–21, 36–37
  - example of, 20f
  - relative, 21, 22f, 37, 38f
  - summary of, 60, 61f
- Frequency, relative, 153, 168. *See also* Probabilities
- Frequency tables, 18–25
  - of blood cholesterol levels, 33t
  - constructing, 22
  - definition of, 18
  - problems for, 25–32
  - sample means and, 75

- of sick leave, 19t
- summary of, 60
- of symmetric data, 21t
- Future responses, prediction intervals for, 573–575

**G**

- Galileo, 157
- Galton, Francis
  - on frequency of error, 307
  - on heredity, 127
  - regression and, 115, 538, 564, 570–571
  - role, in history of statistics, 9–10
- Gauss, Karl Friedrich and normal curves, 291–292
- Genes, 606
- Geometrically weighted moving-average, 719
- Goodness-of-fit tests, 606–609. *See also* Chi-squared goodness-of-fit tests
- Gosset, W.S.
  - role, in history of statistics, 10
  - on sample correlation coefficients, 128
  - t* statistic distributions and, 418
- Grand mean, in ANOVA, 516
- Graphical plotting, of Edmund Halley, 58–59
- Graphs, 18–25
  - bar, 19–21
  - frequency polygons, 19–21
  - line, 19–21
  - problems for, 25–32
  - relative frequency, 21–24
- Graunt, John
  - life table, 159
  - mortality table of, 9t
  - role, in history of statistics, 7–9
- Grouped data, 32–39
  - problems for, 39–43
- Guass, Karl Friedrich, 9, 589–590
- Gaussian distribution, 291
- Guerry, A.M., bar graphs used by, 59

**H**

- Halley, Edmund, 8–9
  - graphical plotting and, 58–59
- Hawthorne effect, 456
- historical perspective, 458
- Heredity, Galton, Francis, on, 127
- Hill, A.B., 72

- Histograms, 32–39
  - of approximately normal data sets, 110f
  - of bimodal data sets, 112f
  - of birth rates, 37f
  - characteristics of data detected by, 35f
  - constructing, 34
  - definition of, 33
  - importance of, 34
  - of normal data sets, 109f
  - Pearson and, 59
  - problems for, 39–43
  - of skewed data sets, 110f, 111f
  - summary of, 60, 62f
- Huyghens, Christian, 85
- Huyghens, Ludwig, 85, 157
- Hybrid genes, 240
- Hypergeometric random variables, 248–249
  - binomial distributions of, 249
  - definition of, 248
  - expected value and variance of, 248
  - problems for, 249–250
- Hypothesis
  - alternative, 389
  - establishing, 391
  - null. *See* Null hypothesis
  - proving, 391, 405
  - statistical, 389
- Hypothesis tests
  - analysis of variance. *See* Analysis of variance (ANOVA)
  - of binomial parameters, 423–425
  - errors, 391
  - first published (historical perspective), 430
  - goodness of fit. *See* Chi-squared goodness-of-fit tests
  - introduction to, 388
  - key terms for, 433
  - key terms for, concerning two populations, 493
  - linear regression and  $\beta$  equals zero, 557–560
  - mean of normal populations, 394–400
  - misinterpreting rejections, 489
  - for non-normal distributions, 406
  - nonparametric. *See* Nonparametric hypotheses tests
  - observational studies and, 486
  - one-sided, 403–406
    - of median, 655
  - one-sided, defined, 405
  - $p$  values, 398
    - of population proportion, 422
  - pictorial depiction of, 397f, 404f, 411f, 415f
  - point estimators for, 392
  - population proportion and, 421–429
  - population proportion equality tests, 481–489
  - problems for, 392–394
    - normal population with known variance, 400–403
    - one-sided tests, 406–409
    - paired-sample  $t$  tests, 476–481
    - population proportions, 429–433, 490–493
    - small-sample, with unknown population variances, 468–471
    - $t$  tests, 417–421
    - two normal populations with known variances, 450–453
    - two-factor ANOVA, 527–529
    - unknown variance and large sample sizes, 459–463
  - proposed case studies for, 437–442
  - randomizing sets in, 486
  - review problems, 437–442
  - review problems for, concerning two populations, 498–502
  - significance levels and, 388–392
  - summary of, 433–437
  - summary of, concerning two populations, 493–498
  - $t$  tests, 409–417
  - Three Mile Island case, 407–408
  - two binomial probabilities, 488t
  - two population tests
    - introduction, 444–445
    - paired-sample  $t$  test, 471–476
  - two population tests of equality of means
    - known variances, 446–450
    - small-sample, with equal unknown variances, 463–468
    - unknown variances and large sample size, 453–459
  - two-factor ANOVA, 520–527
  - types of (summary tables), 406t, 417t, 429t, 450t, 459t, 468t, 488t
  - unknown variance, 409–417
  - $Z$  test, 396, 397f
- I**
  - In control process, 700
  - Independent events, 167–176
    - any number of, and probability, 176
    - definition of, 173–174
    - problems for, 177–184
    - testing in contingency tables, 631–634
    - testing in two characteristics of populations, 620–626
  - Independent random variables, 234
  - Independent trials
    - Bernoulli, Jacques, and, 244
    - for binomial random variables, 239
  - Independent variables, 539
  - Inferential statistics, overview of, 4–5
  - Input variables
    - definition of, 539
    - in simple linear regression, 540–542
  - Interquartile range, 103–104
  - Intersection of events, 148–150
  - Interval estimators of mean, 347–371
    - confidence and, 347, 349, 362
    - confidence bounds in, 355–357
    - definition of, 347
    - introduction to, 333
    - 90, 95, 99 percent confidence, 350t, 351f
    - of normal populations with known variance, 347–357
    - of normal populations with unknown variance, 359–366
    - for population means, 362
    - problems for, 357–359, 366–371
    - sample size for, 353
    - $t$  random variable, 359–361
  - Interval estimators of population proportion, 371–380
    - case studies for, 373–374
    - confidence and, 371

Interval estimators of population proportion (*continued*)  
 length of confidence interval, 373–375  
 problems for, 377–380  
 IQ, stem-and-leaf plots for, 52f

## K

Kruskal-Wallis test, 684

## L

Laplace, Pierre Simon, 9  
 Law of frequency of error, 307  
 Least squares method, 589–590, 596  
 Least-square estimators, 544–545  
   of regression parameters, 588  
 Left-end inclusion convention, 32  
 Legendre, Adrien Marie, 589–590  
 Levels of significance. *See* Significance levels  
 Life table, 159  
 Line graphs, 19–21  
   example of, 19t  
   summary of, 60, 61f  
 Linear regression  
   biological data sets, 569–570  
   dummy variables for categorical data, 590–592  
   estimated regression line, 545, 547f  
   estimating regression parameters, 544–548  
   Galton and, 538  
   introduction, 539–540  
   key terms for, 595  
   linear equation, 539  
   multiple, 586–592  
   prediction intervals for future responses, 573–575  
   problems for  
   analysis of residuals, 586  
    $\beta$  equals zero, 560–564  
   coefficient of determination, 580–581  
   error random variable, 553–555  
   estimating parameters, 548–553  
   multiple regression model, 592–594  
   prediction intervals, 575–578  
   regression to mean, 570–573

sample correlation coefficient, 583  
 simple model, 542–544  
 regression to the mean, 539, 564–570  
 residuals, analysis of, 584–586  
 review problems, 599–604  
 simple model for, 540–542  
 summary of, 595–599  
 testing  $\beta$  equals zero hypothesis, 560–564  
 Linear relationships  
   equation for, 539  
   sample correlation coefficient, 123, 126  
 Lower confidence bounds  
   for interval estimation of means, 355–357, 364–366  
   for interval estimation of population proportions, 375–377  
 Lower control limit (LCL), 700, 702

## M

Mann-Whitney test, 669  
 Margin of error, 383  
 Marlowe, Christopher, 672  
 Mathematical preliminaries, summation of, 733  
 Mean. *See also* Population mean; Sample mean  
   definition of, 218  
   detecting shifts in, 700–705  
   regression to, 564–570  
    $\bar{X}$  control charts for detecting shifts in, 700–705  
   unknown mean and variance, 707–709  
*Men of Mathematics* (Bell), 292  
 Mendel, Gregor, 606–608  
   data manipulation, 608, 613–614  
 Mendenhall, Thomas (historical perspective), 671–672  
 Method of least squares, 589–590, 596  
 Method of maximum likelihood, 393  
 Method of movements, 393  
 Modal values, 97  
 Mortality tables, history of, 7–9  
 Mosteller, Frederic, 672  
 Moving average. *See* Exponentially weighted moving-average control charts

Multiple linear regression model, 586–592  
   definition of, 587  
   problems for, 592–594  
 Multiplication rule of probability, 172  
 Mutually exclusive events, 149, 150

## N

*Natural Inheritance* (Galton), 307  
 Negative correlations, 121  
   in sample correlation coefficients, 122–123  
 Newton, Isaac, 8–9  
 Neyman, Jerzy, 10, 393  
 Nightingale, Florence (historical perspective), 627  
 Noise, 349  
   random, 396  
 Nonparametric hypotheses tests  
   comparison rankings, 685–688  
   definition of, 648  
   equality of multiple probability distributions, 683–688  
   Freedman test, 688  
   key terms for, 693  
   Kruskal-Wallis test, 684  
   normal distribution tests  
   compared to, 672–673  
   permutation tests, 689–692  
   problems for  
   equality of multiple probability distributions, 688  
   permutation tests, 692–693  
   rank-sum test, 673–675  
   runs test for randomness, 681–683  
   sign test, 655–657  
   signed-rank test, 664–667  
   rank-sum test, 667–673  
   review problems, 696–697  
   runs test for randomness, 676–681  
   sign test, 648–655  
   null hypothesis, 650f  
    $p$  value, 650  
   signed-rank test, 657–663  
   summary of, 693–696  
 Normal approximation, binomial distribution and, 317–319  
 Normal curves  
   approximate areas under, 269f  
   approximation rule and, 268, 269f



- central limit theorem and, 306
  - Gauss (historical perspective on), 291–292
  - standard, 268f
  - Normal data sets, 109–114
    - definition of, 109
    - empirical rule and, 114
    - historical perspective on, 115
    - problems for, 114–120
    - summary of, 138
  - Normal distribution
    - biological data sets, 569–570
    - control and, 704
    - historical perspective on, 263–264
    - introduction to, 262
    - normal random variables, 266–269
  - Normal populations
    - interval estimators of, with known variance, 347–357
    - interval estimators of, with unknown variance, 359–366
    - one-sided tests of, 403–406
    - problems for distribution in, 325
    - problems for mean tests of, 400–403
    - problems for testing two, with known variances, 450–453
    - sample variance distribution in, 323–325
    - tests. *See* Hypothesis tests
  - Normal probabilities, finding, 277–279
  - Normal random variable, 266–269
    - approximation rule for, 268
    - key terms for, 290
    - percentiles of, 284–288
    - probabilities associated with, 271–276
    - problems for, 269–271
      - additive property, 281–284
      - continuous random variables, 264–266
      - percentiles of, 289–290
      - probabilities associated with, 276–277
    - review problems, 293–296
    - standard, 267
    - standard deviation of, 277, 279
    - standardizing, 277
    - summary of, 290–293
  - Null event, 148–149
  - Null hypothesis
    - appropriate, 416
    - classical testing procedure for, 391
    - definition of, 389
    - discrediting, 391
    - not rejecting, 399–400
    - rejection of, 390
    - significance levels necessary for rejection of, 404
    - statistical test of, 390
    - testing, in chi-square
      - goodness-of-fit tests, 610
    - testing, with unknown variance, 414–415
    - types of tests (summary tables), 406t, 417t, 429t, 450t, 459t, 468t, 488t. *See also* Hypothesis tests
  - Numerical science, 9
- O**
- Observational studies, for hypothesis tests, 486
  - One-factor analysis of variance, 505–510
    - definition of, 504
    - problems for, 510–514
    - summary of, 530–531
    - summary table, 510t
  - One-sided tests, 403–406
    - definition of, 405
    - problems for, 406–409
    - sign tests, 653–655
    - two population tests, 468t, 474, 484, 487
  - Outcomes
    - definition of, 146
    - equally likely, 161–164
    - problems for equally likely, 164–167
    - sample space and events of experiment, 146–150
  - Outliers, 40, 54, 69
- P**
- p* value
    - of chi-squared test, 612
    - hypothesis testing, 398
    - in population proportion hypothesis tests, 422, 483
    - in sign tests, 650
    - of signed-rank tests, 660
    - summary of, 434
    - two-sided tests of, 425–429
  - types of tests (summary tables), 406t, 417t, 429t, 450t, 459t, 468t, 488t
  - Paired data, 51–54
    - notation, 51
    - paired sample *t* test, 471–476
    - problems for, 54–58, 476–481
  - Parallel circuit, 183
  - Pascal, 157
  - Pearson, Egon, 393
  - Pearson, Karl
    - chi-squared goodness-of-fit test and, 608, 610, 616
    - on De Moivre, 264
    - histograms used by, 59
    - on Nightingale, 627
    - product-moment correlation coefficient of, 128
    - regression to the mean and, 565
    - role, in history of statistics, 10, 292, 393
  - Percent confidence interval estimator, 350
  - Percentiles
    - chi-squared, 611f, 741–742t
    - confidence levels, 350t
    - definition of, 285
    - F* distributions, 743–748
    - finding, by conversion to standard normal, 287
    - of normal random variables, 284–288
    - problems for, of normal random variables, 289–290
    - sample, 90–93, 136–137
    - t* distribution, 740–741t
  - Permutation tests, 689–692
    - problems for, 692–693
  - Permutations, 191–192
  - Philosophical Transactions of the Royal Society* (Arbuthnot), 430
  - Piazzi, Giuseppe, 589
  - Pie charts, 24–25
    - example of, 25f
  - Placebo effect, 3, 445, 456
  - Placebos, 4
  - Plague and history of statistics, 7–9
  - Playfair, William, pie charts used by, 59
  - Poincaré, Henri, 311
  - Point estimator
    - hypothesis testing, 392
    - introduction to, 333
    - of population mean, 333–334

- Point estimator (*continued*)
  - for population proportions, 336–338
  - problems for, of population means, 334–336
  - problems for, of population proportions, 338–340
  - standard errors and, 334
- Poisson random variables, 250–253
  - binomial random variables and, 250–251
  - definition of, 250
  - expected value and variance of, 252
  - graphs of, 251f
  - problems for, 253–254
- Poisson, Simeon, 140
- Pooled estimator
  - testing equality of means, 464
  - testing population proportions, 483
- Population distributions
  - introduction to, 298–299
  - probability distributions of
    - sample mean and, 302f
  - sign test of, 652–653
- Population mean
  - confidence bounds for, 355–357
  - definition of, 300
  - hypothesis tests for normal, 394–400
  - interval estimators for, 362
  - obtaining, 301, 302
  - one-sided tests concerning two, 468t
  - point estimator of, 333–334
  - problems for point estimation of, 334–336
  - sample means and, 333
  - t* test for, 409–417
- Population proportions
  - case studies for, 373–374
  - hypothesis tests concerning, 421–429
  - interval estimators of, 371–377
  - one-sided hypothesis and, 487
  - point estimators for, 336–338
  - pooled estimators in, 483
  - problems for hypothesis tests of, 429–433, 490–493
  - problems for interval estimators of, 377–380
  - problems for point estimation of, 338–340
  - testing equality of, 481–489
- Population size, sample size and, 316
- Population standard deviation, 302
  - estimating, 334, 343–344
- Population variance
  - definition of, 300
  - estimating, 342–344
  - hypothesis tests for small sample
    - with equal unknown, 463–468
  - interval estimators of normal
    - populations with known, 347–357
  - interval estimators of normal
    - populations with unknown, 359–366
  - obtaining, 301, 302–303
  - pooled estimators, 464
  - problems for estimation of, 344–347
  - standard normal distributions of
    - equal, 464
- Populations
  - central limit theorem for various, 307
  - definition of, 5
  - densities of sample means of, 300f
  - finite, 313–319
  - normal, 323–325
  - numerical values associated with, 299
  - sampling from correctly, 320
- Positive correlations, 121
  - in sample correlation coefficients, 122–123
- Prediction intervals
  - definition of, 540, 573
  - for future responses, 573–575
  - problems for, 575–578
- Probabilities
  - addition rule for, 154–155
  - Bayes' theorem, 185–187
  - binomial, 241f
  - of binomial random variables, 238–243
  - conditional, 167–176
  - counting principles, 189–195
  - definition of, 5, 146
  - of equally likely outcomes, 161–164
  - finding normal, 277–279
  - historical perspective on, 157
  - key terms for, 198–199
  - as long-run relative frequencies, 153, 168
  - multiplication rule, 172
  - for negative *x*, 273
  - problems for
    - Bayes' theorem, 187–189
    - conditional probability and independence, 177–184
    - counting principles, 195–198
    - equally likely outcomes, 164–167
    - properties, 156–161
    - sample space, 150–153
  - problems for, of sensitive events, 342
  - properties of, 153–156
  - review problems, 201–208
  - sample proportions and, 317–319
  - sample space and events of
    - experiment, 146–150
  - of sensitive events, 341–342
  - standard normal, 272t, 739–740t
  - of standard normal random
    - variables, 271–276
  - summary of, 200–201
- Probability density function
  - definition of, 262–263
  - density curves, 263f, 264f, 360f
  - of sample means from standard
    - normal populations, 300f
  - of *t* random variables, 360f
- Probability distribution
  - discrete random variables and, 213
  - introduction to, 298–299
  - population distributions and, 302f
  - of sample means, 301
- Probability mass functions
  - of binomial random variables, 318f
  - graph of, 302f
- Probability models, overview of, 4–5
- Problem of the points, 157
- Product-moment correlation
  - coefficient, Pearson's, 128
- Programs, 755
  - Program 3-1, 137, 458
  - Program 3-2, 133
  - Program 5-1, 242, 243, 319, 422, 423, 424, 426, 437, 651, 655, 694
  - Program 6-1, 275, 276, 277, 291, 308, 319



Program 6-2, 286, 292  
 Program 8-1, 361, 475  
 Program 8-2, 361, 412, 415, 416, 558  
 Program 8-3, 362, 363f, 364f, 366  
 Program 9-1, 413, 414, 415, 417, 473, 474, 476  
 Program 10-1, 466–467  
 Program 11-1, 509, 525, 531  
 Program 11-2, 525, 527, 533  
 Program 12-1, 546, 555, 566, 568, 575, 598  
 Program 12-2, 588, 590, 599  
 Program 14-1, 660–661, 662, 663, 694  
 Program 14-2, 671, 672, 695  
 Program 14-3, 677, 679, 680, 681, 696  
 Program A-1, 314, 736, 737  
 Proportions, in finite populations, 314  
 Pure dominance, 240  
 Pure recessiveness, 240

## Q

Quality control methods, 504  
   cumulative sum control charts, 722–724  
   exponentially weighted moving-average control charts, 717–721  
   introduction to, 700  
   key terms for, 725  
   problems for  
     cumulative sum control charts, 725  
     exponentially weighted moving-average control charts, 721–722  
      $\bar{S}$  control charts, 713–715  
      $\bar{X}$  control chart for detecting mean shifts, 705–706  
   review problems, 726  
    $S$  control charts, 710–712  
   Shewhart control chart, 704  
   summary of, 725–726  
    $\bar{X}$  control charts, 700–705  
     unknown mean and variance, 707–709  
 Quartiles  
   definition of, 93

interquartile range, 103–104  
 summary of, 136–137  
 Quetelet, Adolphe, 59, 627  
   on normal data sets, 115

## R

Random error, 539  
 Random noise, 396  
 Random numbers, 736  
 Random samples  
   algorithms for generating, 736  
   choosing, 314–315, 735–737  
   definition of, 6, 314  
 Random selection, 5  
   equally likely outcomes, 162  
 Random variables, 211–214  
   assigning values to, 211  
   binomial, 238–243  
   chi-squared, 323–325, 360  
   continuous, 262–264  
   definition of, 211  
   degrees of freedom, 323, 360  
   density curves of, 264f, 265f  
   discrete, definition of, 210, 212  
   error, 553–555  
   expected value, 218–225  
    $F$ , with degrees of freedom, 507f  
   in finite populations, 315  
   hypergeometric, 248–249  
   independent, 234  
   normal, 266–269  
   Poisson, 250–253  
   problems for, 215–217  
   review problems, 256–259  
   standard normal, 267  
   summary of, 254–256  
    $t$  random variable, 360  
   uniform, 264  
   variance of, 231–236  
   of zero expected value, 231  
 Random walk model, 440  
   geometric, 369  
   linear, 346, 369  
 Randomization  
   in hypothesis testing, 486  
   technique for sensitive events, 341–342  
 Randomness, runs test for, 676–681  
 Range, 137  
 Rankings, comparison, 685–688  
 Rank-sum test  
   mean and variance of TS, 668  
   normal distribution tests  
     compared to, 672–673  
   problems for, 673–675  
   summary of, 695  
   for two populations, 667–673  
   two-sample, 669  
 Referents, 486  
 Regression fallacy  
   definition of, 539  
   regression to the mean and, 569  
 Regression line  
   estimated, 545, 547f  
   standardized residuals, 584, 585f, 586f  
 Regression parameters  
   errors and, 544f  
   estimating, 544–548  
   least-square estimators of, 588  
   multiple linear regression model, 587–588  
   problems for, 548–553  
 Regression to the mean, 539, 564–570  
   definition of, 539, 564  
   Galton's regression to mediocrity, 538  
   historical perspective on, 115  
   regression fallacy, 569  
   scatter diagram, 564f, 565f  
 Rejection region, 390  
 Rejections, misinterpreting, 489  
 Relative frequency graphs, 21–24  
   sick leave data, 22t  
 Relative frequency histograms, 34  
   summary of, 60  
 Relative frequency polygons, 21, 37  
   examples of, 22f, 38f  
 Residuals, 554  
   analysis of, 584–586  
   standardized, 584, 585f, 586f  
 Response variables  
   definition of, 539  
   in simple linear regression, 540–542  
 Row factors, in ANOVA, 515  
 Row sum of squares  
   definition of, 524  
   two-factor ANOVA tests, 523–524  
 Runs test for randomness, 676–681  
   definition of, 677  
   problems for, 681–683  
   summary of, 695–696

## S

- S control charts, 710–712
  - examples of, 712f
  - problems for, 713–715
- Salaries, stem-and-leaf plots for, 52f
- Sample correlation coefficient, 120–128, 582
  - absolute values of, 126–127
  - computational formula for, 124
  - definition of, 582
  - definition of (formula), 123
  - described, 73, 120
  - example scatter plots, 127f
  - historical perspectives on, 127–128
  - linear relationship of data, 123, 126
  - negative correlations in, 122–123
  - Pearson's product-moment, 128
  - positive correlations in, 122–123
  - problems for, 128–134, 583
  - properties of, 123
  - summary of, 138
  - variable values in, 122
- Sample mean, 73–78, 299–303
  - central limit theorem and
    - distribution of, 306–310
  - central tendencies described by, 84
  - compared to sample median, 86
  - definition of (formula), 73
  - densities of, 300f
  - deviations and, 78–79
  - expected values of, 300
  - frequency tables and, 75
  - historical perspective on, 78, 85
  - population means and, 333
  - probability distribution of, 301
  - problems for, 79–83, 86–90, 303–304
  - standard deviation of, 302
  - summary of, 136
  - variance of, 300
- Sample median, 83–86
  - central tendencies described by, 84
  - compared to sample mean, 86
  - definition of, 83
  - historical perspective on, 85
  - as percentile (50th), 90
  - problems for, 86–90
  - summary of, 136
- Sample mode, 97–98
  - definition of, 97
- problems for, 98–99
- summary of, 136
- Sample percentiles, 90–93
  - calculating, 90–91
  - of data set size  $n$ , 91
  - definition of, 90
  - problems for, 93–97
  - quartiles of, 93
  - summary of, 136–137
- Sample proportions
  - probabilities and, 317–319
  - problems for, 319–323
- Sample size
  - hypothesis tests for large, 453–459
    - appropriate sizes for, 455
  - hypothesis tests for small, 463–468
  - for interval estimation, 353
  - population size and, 316
  - problems for hypothesis tests for
    - large, 459–463
    - problems for hypothesis tests for small, 468–471
- Sample space, 146–150
  - definition of, 146
  - problems for, 150–153
- Sample standard deviation, 99–105
  - definition of (formula), 102–103
  - problems for, 105–109
  - summary of, 137
- Sample variance, 99–105
  - computing by hand, 101, 137
  - definition of (formula), 100
  - distribution of, in normal
    - populations, 323–325
  - problems for, 105–109
  - problems for distribution of, in
    - normal populations, 325
  - unchanged, when adding
    - constants to data, 102
- Samples
  - as controls, 457
  - from correct populations, 320
  - definition of, 5, 299
  - key terms for, 325–326
  - preview of, 298
  - random, 314–315
  - representative, 6
  - review problems, 327–330
  - sizes of, for central limit theorem, 310–311
  - summary of, 326–327
- Sampling proportions, for finite
  - populations, 313–319
- Scatter diagram
  - definition of, 51
  - estimated regression line, 547f
  - IQ vs. income, 53f
  - regression to the mean, 564f, 565f
  - simple linear regression, 540, 541f
  - standardized residuals and, 585f, 586f
  - summary of, 63
- Second quartile, definition of, 93
- Sensitive events
  - estimating probability of, 341–342
  - problems for estimation of
    - probability of, 342
- Series circuit, 183
- Set notation, 734
- Shewhart control chart, 704
- Shewhart, Walter, 704
- Sign test, 648–655
  - equality of population
    - distributions, paired
      - samples, 652–653
  - of null hypothesis, 650f
  - one-sided, 653–655
  - $p$  value in, 650
  - problems for, 655–657
  - summary of, 693–694
- Signed-rank test, 657–663
  - $p$  value, 660
  - problems for, 664–667
  - summary of, 694–695
  - summary table, 663
  - zero-differences and ties in, 662–663
- Significance levels
  - $\alpha$  test, 391
  - correct, 397–398
  - definition of, 391
  - historical perspective on, 393
  - hypothesis test concepts, 388–392
  - hypothesis tests (summary tables), 391, 406t, 417t, 450t, 459t, 468t, 488t
  - null hypothesis and, 404
  - $p$  value, 398
  - in paired-sample  $t$  tests, 473
  - problems for, 392–394
  - of two-sided  $t$  test, 411f
- Simon, Pierre, central limit theorem
  - and, 310
- Simple linear regression model
  - assessing, 584–586

definition of simple linear regression, 540  
 input variable in, 540–542  
 problems for, 542–544  
 response variable in, 540–542  
 scatter diagram and, 540, 541f  
 Simple random samples, definition of, 6  
 Skewed data, 110f, 111f  
 Standard deviation  
   of binomial random variables, 316  
   definition of, 235  
   of normal random variables, 277, 279  
   point estimators for population mean, 334  
   point estimators for population proportion, 337  
   population, estimate, 343  
   of random variables, 235–236  
   sample, 99–105  
   of sample mean, 302  
 Standard error  
   for estimating sensitive events, 342  
   in point estimation of population mean, 334  
   in point estimation of population proportion, 337  
 Standard exponential weighted moving-average control chart, 719  
 Standard normal  
   conversion to, 277–279  
   finding percentiles and, 287  
   in interval estimations, 360f  
 Standardized residuals, 584  
   problems for, 586  
   regression line and, 584–586  
   scatter diagrams of, 585f, 586f  
 Standardizing, normal random variables, 277  
 Statistical hypothesis, definition of, 389  
*Statistical Methods for Research Workers* (Fisher), 418  
 Statistical test, of null hypothesis, 390  
 Statistics  
   changing definition of, 11  
   definition of, 3, 72  
   history of, 7–10  
   key terms for, 10, 134–136

Statistics, sampling  
   Central limit theorem, 115, 304–311  
   distribution of sample variance, 323–325  
   normal data sets and empirical rule, 109–120  
   problems for  
     central limit theorem, 311–313  
     distribution of sample variance, 325  
     empirical rule, 114–120  
     finite populations, 319–323  
     sample correlation coefficient, 128–134  
     sample means, 79–83, 86–90, 303–304  
     sample median, 86–90  
     sample mode, 98–99  
     sample variance and standard deviation, 105–109  
   sample mean, 73–83, 299–303  
   sample median, 83–97  
   sample mode, 97–99  
   sample standard deviation, 99–109  
   sample variance, 99–109  
   sampling proportions, 313–319  
 Stem-and-leaf plots, 44–47  
   definition of, 44  
   examples of, 45–47  
   introduced by John Tukey, 59  
   for IQ scores, 52f  
   problems for, 47–51  
   for salaries, 52f  
   summary of, 62  
   uses of, 46  
 Stratified random sampling, definition of, 6–7  
 Sums, expected value of, 222, 224  
 Symmetry, 20–21  
   approximately symmetric, 21  
   bar graphs and, 21f  
   histograms, 34–35

## T

*t* distribution, 360f  
   percentiles, 740–741t  
*t* random variable, 359–361  
*t* test  
   historical perspective on, 418  
   for non-normal distributions, 416

for normal population means, 409–417  
   paired-sample, 471–476  
   problems for, 417–421  
   types of, 417t  
 Test statistic, definition of, 389  
 Third quartile, definition of, 93  
 Trials, 238  
   for hypergeometric random variables, 248  
 Tukey, John, stem-and-leaf plots used by, 59  
 Two-factor analysis of variance, 514–529  
   definition of, 504  
   error sum of squares, 522  
   estimators, 521  
   problems for, 518–520  
   problems for testing hypotheses and, 527–529  
   row sum of squares, 523–524  
   summary of, 531–533  
   summary table, 525t  
   testing hypotheses and, 520–527  
   unbiased estimators in, 523  
 Two-sample rank-sum test, 669  
 Two-sided tests  
   of *p* value, 425–429  
   significance level of *t* tests, 411f  
   *t* tests, 410–413  
   two population tests, 450t, 468t  
 Type I error, 391  
 Type II error, 391

## U

Unbiased estimator  
   definition of, 333  
   in two-factor ANOVA, 523  
 Uniform random variables, 264  
 Union of events, 148–150  
 Upper confidence bounds  
   for interval estimation of means, 355–357, 364–366  
   for interval estimation of population proportions, 375–377  
 Upper control limit (UCL), 700, 702

## V

Variance  
   of binomial random variables, 243–244  
   computational formula for, 231

Variance (*continued*)  
 definition of, 231  
 estimating population, 342–344  
 of hypergeometric random variables, 248  
 hypothesis tests for large sample sizes with unknown, 453–459  
 hypothesis tests for unknown, 409–417  
 of independent random variable sums, 234  
 measuring, in response values, 578  
 of Poisson random variables, 252  
 population mean tests with known, 394–400  
 problems for, 236–238  
 properties of, 233–236  
 of random variables, 231–236

reduction of, 344  
 sample, 99–105  
 of sample means, 300  
 standard deviation and, 235–236  
 two population tests of equality of means, known variances, 446–450  
 $\bar{X}$  control charts and unknown, 707–709  
 Variation, 73  
   chance, 700  
 Venn diagrams, 148, 149f, 150f

## W

Wallace, David, 672  
 Weighted averages, 76  
 Weiss, A., 76  
 Wilcoxon sum-of-ranks test, 669  
 Wright, Sewell, 393

## X

$\bar{X}$  control chart  
   for detecting shifts in mean, 700–705  
     unknown mean and variance, 707–709  
   examples of, 702f, 712f  
   problems for  
     detecting shifts in mean, 705–706  
     fraction defective, 717  
     S control chart, 713–715  
     subgroup size in, 702f

## Y

Young, Arthur, controls used by, 457

## Z

Z test, 396, 397f