

INDEX

Page references followed by *fig* indicate an illustrated figure; followed by *t* indicate a table. Symbols are sorted within the index.

A

A priori probability, 149
Absolute cell references: description of, 63; \$ symbol and, 63–64; F4 function key for, 64
Absolute differences, 195–196*fig*, 197*fig*
Add Trendline option, 445
Add-Ins dialog box, 59*fig*
Admission data: calculating age using
 =YEARFRAC() function, 96*fig*–97*fig*; DOA (date of admission), 97–98; DOB (date of birth), 93–94. *See also* LOS (length of stay)
Alpha (α): adjusting *F* probability to protect against, 329–330; cost of research/ intervention and small, 260–264; definition of, 250–251; mutual dependence of beta and, 253–254; specifying, 255. *See also* Type I errors
Alternative hypothesis: calculating beta (β) for, 258–260; calculating *t* value for, 259–260
Analysis ToolPak, 58, 59*fig*
Analysis ToolPak-VBA, 58
=AND() function, 245–246
ANOVA: Two-Factor with Replication add-in, 347, 350*fig*–352*fig*
ANOVA (analysis of variance): applications of, 70; comparing *t* tests to, 11–12; definition of, 26; emergency trauma (ED) color code, 7; establishing statistical difference using, 9; *F* test for differences in, 322, 324; factorial, 343–355; independence assessed by, 270; one-way, 318–332*t*; for repeated measures, 334–342*t*; *t* test relationship to, 319. *See also* Variance
ANOVA Single-Factor Data Analysis add-in, 324–326

Arguments: data range, 105; definition of, 40; Excel functions (f_x and number of, 40. *See also* =FREQUENCY() function
Arrays: determinant of, 402; Excel's ability to manipulate, 403; Excel's functions related to, 402–403*fig*, 404; solving regression equations in Excel for inverse, 404*fig*–405. *See also* Matrix/matrices
=AVERAGE() function: calculation of average, 39–40*fig*; central tendency measured by, 189; coefficient of determination using, 371*fig*; frequency distributions and standard deviations using, 199–200; Function arguments dialog box for, 38–40*fig*; multiple regression calculated using, 407, 409*t*; number of arguments used in, 40; variance calculated using, 194–195
Average. *See* Mean (or average)

B

b coefficients: calculating standard errors of, 407*fig*; coefficient b_j , 396*fig*–397; Coefficient b_o , 374; generating regression results using, 406–407
Backward stepwise elimination, 430, 431*fig*, 432–433
Bar chart, 112–113*fig*
Bartlett test: description of, 330; for homogeneity of variance, 330–331*fig*; interpreting the, 331–332
Bayes' theorem, 159–160*fig*
Bell-shaped curve, 181, 203*fig*. *See also* Normal frequency distribution
Bernoulli distribution, 83
Best-fitting straight line, 361–362*fig*

Beta (β): calculating for any alternative hypothesis, 258–260; cost of research/intervention and small, 260–264; definition of, 251–253; determining, 256; mutual dependence of alpha and, 253–254; one-tail test and, 256–258; reality of many, 254–255; specifying, 255. *See also* Type II errors

Between group variance, 319–321

=BINOMDIST() function, 169–170*fig*

Binomial distribution: applications to health care professionals, 170–171*fig*; calculating correctly documented Medicare claims using, 172; for emergencies in an eight-hour shift, 171*fig*; empirical probability and, 153; equations for, 166, 168–171; random numbers from the, 83–84; visualizing correctly documented Medicare claims using, 172–173*fig*

Binomial distribution equations:

=BINOMDIST() function, 169–170*fig*;
combinatorial formulas, 166, 168; =FACT()
function, 168–169*fig*

Binomial probability: applied to flips of a coin, 164–166; applying to emergency and nonemergency visits, 166, 167*fig*; definition of, 164; distribution, 83–84, 164–173*fig*

Bins: chart of age showing ranges of, 111*fig*;
defined for the frequency distribution,
105–106*fig*; definition of, 105

Breast cancer education, statistical study of
effectiveness of, 11

Buttons: Excel Options, 57*fig*–58; Layout, 125;
Undo, 36–37*fig*

C

Cancer education intervention sample, 79–80*fig*

Carteret Falls regional hospital, establishing
statistical difference using regression, 9–10

Categorical data, emergency trauma (ED) color
code, 7

Categorical variables: used in data sets,
67; definition of, 21; using pivot table
to generate frequencies of, 123–134;
transforming into numerical variables,
98*fig*–99

Causality caveats, 8

Cell references: absolute and relative, 63; \$
symbol and absolute, 63–64; F4 function key
for absolute, 64

Cells: contiguous and noncontiguous,
41; copying to a range of cells, 34–36;
functions that give results in more than one,
60*fig*–61; highlighting, 33–34*fig*; result of
Ctrl+Shift+Right arrow, 33–34*fig*; summing
two noncontiguous areas, 41*fig*; Undo
command for undoing changes in, 36–37*fig*.
See also Spreadsheets

Central tendency: definition of, 186; dispersion
measure of, 189–192; Excel functions for,
188–189; mean measurement of, 186*fig*,
187; median measurement of, 186*fig*, 187;
mode measurements of, 186*fig*, 187–188;
Random Number Generation dialog box
for calculating, 217–220*fig*; variance
measurement of, 190–195. *See also* Standard
deviation

Chart formatting pop-up menu, 51*fig*

Chart Wizard icon, 129

Charts: bar, 112–113*fig*; column, 110–111;
line, 112*fig*; Pareto, 132, 134, 135*fig*; pie,
113*fig*–114; pivot tables generated by
option of automatic, 129–130; XY(Scatter),
114*fig*–115

Chi-square analyses: dichotomous dependent
variable treatment using, 455–458; small
expected values in cells and, 286–289*fig*;
statistical versus practical significance,
281–282; two-by-two table to understand,
280–285*fig*; Yates's correction used for, 457

Chi-square tests: applications of, 70; Bartlett
test form of, 330–332; definition of, 25;
establishing statistical difference using,
10–11; for statistical independence, 273–274;
template for chi-square statistic, 275*fig*–276;
Type I and Type II errors, 277–279; value
calculated for Logit analysis, 473–474*fig*

=CHIDIST() function, 274–275, 474

=CHINV() function, 275

=CHITEST() function, 276–277*fig*

Cluster samples: definition of, 19–20; stratified
versus, 20

Coefficient of b: b_1 , 396*fig*–397; b_0 , 374;
calculating standard errors of, 407*fig*;
generating regression results using, 406–407.
See also Correlation coefficient

Coefficient of determination (R^2), 369–370*fig*

Coefficients: of b, 374, 396*fig*–397,
406–407*fig*; log, 470, 471–477*fig*

- Coin flips: applying binary logic to, 164–166; possible combinations of five, 148*fig*. *See also* Probability
- Column chart, 110
- Conditional probability: Bayes' theorem and, 159–160*fig*; data frequencies and, 160; definition of, 159; independent events and, 160–161; mathematical independence link to, 273; spreadsheet to test independence using tables of, 161–162*fig*
- Confidence intervals: definition of, 238; from point estimates to, 239*fig*; using representative sample to infer a mean for, 238–239
- Confidence limits: calculating for multiple samples, 243–246; calculating *t* values for, 240–241; determining 95 percent, 242; emergency trauma (ED) color code, 7; one-tail versus two-tail *t* values for, 241; sample size and, 265
- Contingency table, 155–157
- Continuous numerical variables: used in data sets, 67–68; definition of, 23
- Continuous probability distributions, 180
- Copying: Paste Special dialog box for, 77*fig*; random numbers, 76
- Correlation coefficient: calculating the, 438; example of multicollinearity and, 437; finding Logit models, 468. *See also* Coefficient of *b*; Relationships; Variables
- Cost per case: equation for predicted values using dummy variable for total, 416–418*fig*; graph of LOS (length of stay) data for, 418*fig*; statistical applications related to, 8–9
- =COUNTIF() function, 245
- Create PivotTable dialog box, 138*fig*
- Cross-tab (cross-tabulation), *See also* Pivot table
- Ctrl+Shift+Right arrow, 33–34*fig*
- Cultural predictors, 435*fig*
- Cumulative differences, 195
- Cumulative frequency distribution: creating, 115*fig*; creating percentage and, 116; modifying chart of, 116–117
- Cumulative normal distributions: approximating area under normal curve, 206–207; description of, 205–206*fig*; Excel and the, 207–209; for the weight of newborns, 208*fig*
- Cumulative percentage distributions: creating, 115, 116; modifying a, 116–117
- ## D
- Data: access and preparation of, 86–99; best-fitting straight line to, 361–362*fig*; missing, 100–101; nature of, 66–71; secondary, 87; *t* test comparison of two groups of, 299–308; *t* tests for related, 313–316. *See also* Sampling
- Data access/preparation: adjusting dates in Excel, 91–92; inspecting, formatting, and modifying data after import, 90–91*fig*; overview of, 86–87; performing data checks, 93; secondary data used for, 87; Text Import Wizard for importing data, 87–90; transforming categorical variables into numerical variables, 98*fig*–99; values versus formulas, deleting extra columns, \$ formatting, 92–93*fig*
- Data analysis add-in: description and use of, 80–81; drawing samples using, 216–217; multiple regression using, 390*fig*–391
- Data Analysis add-in dialog box, 375, 376*fig*
- Data Analysis dialog box, 59*fig*, 81, 107*fig*, 217
- Data Analysis package: description of, 56–57*fig*; Excel Options button, 57*fig*–58; Excel Options dialog box, 58*fig*; factorial analysis of variance using, 350–351*fig*; installing the, 57–58; linear regression using the, 377; multiple regression analysis using, 391. *See also* Excel functions (f_x); Microsoft Excel
- Data checks: calculating age at admission using =YEARFRAC() function, 96*fig*–97*fig*; calculating length of stay (LOS), 95–97*fig*; =IF() statement used for, 94–95; =OR() statement used for, 94–95; performing, 93
- Data display: copying a cell to range of cells, 34–36; drag and drop for moving data, 36; Excel Charts function for, 46–51*fig*; in a graph, 46–51*fig*; highlighting a cell containing, 33–34*fig*; moving around a data set, 33*fig*; sorting a string, 53*fig*–56*fig*; statistical applications related to, 6; undoing changes, 36–37*fig*

- Data range argument, 105
- Data Ribbon: Data Analysis menu, 107; Sort dialog box on the, 53*fig*–56*fig*, 77*fig*; *t* test add-ins, 308–3011
- Data sets: =MIN() and =MAX() to determining values of, 104–105*fig*; performing checks of, 93; scatterplots of four, 381–382*fig*; screenshot of small, 66*fig*–67. *See also* Variables
- Degrees of freedom: definition and example of, 229*fig*; *F* test for differences in ANOVA and, 324; factorial analysis of variance in two-factor, 346–347*fig*, 353*fig*; in repeated measures, 340*fig*
- Deleting: columns, 92–93; missing data remedy of case, 100
- Dependent variables: definition of, 20; distinction between independent and, 69–70; dummy variable interaction effect and, 418–419*fig*; linear regression between independent and, 361; multiple regression for relationship of independent and, 389*fig*–390; terms as caused variables, 69. *See also* Independent variables
- Descriptive statistics color code, 7
- Determinant: array, 402; finding matrix, 401–402; nonzero, 402
- Dialog boxes: Add-Ins, 59*fig*; Create PivotTable, 138*fig*; Data Analysis, 59*fig*, 81, 107*fig*, 217*fig*; Data Analysis add-in, 375, 376*fig*; Excel Options, 58*fig*; Format Trendline, 445–448; Function arguments, 38–39*fig*; Histogram, 108*fig*; Initial Regression, 390*fig*–391; Insert Chart, 47*fig*; Insert Function, 37–38*fig*; Layout, 125–128, 131*fig*; Normal distribution, 82–83; Paste Special, 77*fig*; pivot table, 124*fig*–125*fig*, 128, 136*fig*; Random Number Generation, 81*fig*, 82, 217–220*fig*; Regression, 375, 376*fig*, 377, 377–379, 462; Select Data Source, 48–50; Sort, 53*fig*–56*fig*, 77*fig*; *t* test for equal variance, 309*fig*; Value Field, 141*fig*
- Dichotomous dependent variables: chi-square analysis of, 455–458; *F* statistic approach to, 462–463; introduction to, 454–455; Logit models for, 465; nonlinear relationships and, 466; OLS (ordinary least squares) with, 458–459; R square statistic approach to, 463–465; traditional treatments of, 455–465; WLS (weighted least squares) for estimating linear probability model for, 460; WLS (weighted least squares) using Excel’s Regression package, 461–462*fig*, 463*fig*; WLS (weighted least squares) weight calculation example, 460–461*fig*
- Difference. *See* Statistical difference
- Discrete distribution: binomial, 83–84, 164–173*fig*; normal, 118*fig*, 120–121, 153, 180–182; random numbers from, 84–85*fig*. *See also* Frequency distribution
- Discrete numerical variables: calculating mean and standard deviation of, 222–223*fig*; definition of, 22–23, 221–222
- Dispersion measure: absolute differences issue of, 195–196*fig*, 197*fig*; being centered at the mean, 196–197; cumulative differences issue of, 195; definition of, 189; Random Number Generation dialog box for calculating, 217–220*fig*; range measurement of, 189, 190; standard deviation measurement of, 189, 190; variance measurement of, 189, 190–195
- Distribution. *See* Frequency distribution
- DOA (date of admission) column, 97–98
- DOB (date of birth) column: checking data in the, 93; =IF() statement for checking, 93–94; when the information is not needed, 97–98
- DOD (date of discharge), 97–98
- \$ formatting, 92–93
- \$ symbol, 63–64
- Drag and drop, moving data with, 36*fig*
- DRGs (diagnostic-related groupings), 9
- Dummy variables: definition of, 98; equation for predicted values of total costs using, 416–418*fig*; for fitting nonlinear relationships, 443–444; general multiple regression analysis formula using, 418; interaction effects of, 418–424*fig*; in multiple regression, 416–425*fig*; with no interaction, 416
- E**
- Emergency room data: binomial distribution in eight-hour shift, 171*fig*; binomial probability

- applied to, 166, 167*fig*; color coding, 6–8; contingency table on, 155–157; first twenty observations in an, 155*fig*; formulas used for calculating probabilities of, 169*fig*; Poisson distribution applications to, 176*fig*–179*fig*; probabilities of number of visits using formulas, 169*fig*. *See also* Nonemergency visits
- Emergency trauma (ED) color code: description and function of, 6–7; descriptive statistics, confidence limits, and categorical data on, 7; establishing significant difference and causality caveats for, 8; inferential statistics, analysis of variance, and regression on, 7
- Empirical probability: definition of, 149–150; frequency of occurrence and, 150–151; independence and, 152; probability distributions and, 152–153; sequential events and, 151–152*fig*
- Error variance, 320
- Events: conditional probability and independent, 160–161; definition of, 146; empirical probability and sequential, 151–152*fig*; independent, 146–147; Venn diagram for two mutually exclusive, 148–149*fig*; Venn diagram of not mutually exclusive, 157–158*fig*
- Excel: advantages of using as statistical operations tool, 4–5; basics of, 30–34; Text Import Wizard of, 87–90. *See also* Data Analysis package; Excel functions (f_x)
- Excel basics: arithmetical conventions used in, 31*fig*–44; Excel functions (f_x), 31, 37–41*fig*; Formula Ribbon, 31, 37; Formulas Ribbon and Excel Functions menu, 32; spreadsheet format of, 30, 31*fig*; workbook groups of spreadsheets in, 30; working and moving around in a spreadsheet, 32–37*fig*
- Excel Charts function: basic bar graph, 48*fig*; identifying and formatting chart data for, 47–48; Insert Chart dialog box, 47*fig*; Select Data Source dialog box, 48–50; uses of, 46–47*fig*
- Excel functions (f_x): =AND() function, 245–246; =AVERAGE() function, 38–40*fig*, 188, 194–195, 199–200, 371*fig*, 407, 409*t*; =BINOMDIST() function, 169–170*fig*; central tendency, 188–189; =CHIDIST() function, 274–275, 474; =CHINV() function, 275; =CHITEST() function, 276–277*fig*; =COUNTIF() function, 245; description of, 31; direct input of, 40–41; Excel Charts function, 46–51*fig*; =EXP() function, 178; =FACT() function, 168–169*fig*; =FDIST() function, 324, 371*fig*, 408, 410*t*; =FINV() function, 326–327; =FREQUENCY() function, 60*fig*, 104–121*fig*, 201–202, 476; =IF() function, 43*fig*–45*fig*; Insert Function dialog box, 37–38*fig*; matrix math functions, 61, 402–405; =MDETERM() function, 402–403; =MEDIAN() function, 188; =MINVERSE() function, 402–403, 405, 477; =MMULT() function, 61–62*fig*, 402–403, 407, 409*t*, 475–477; =MODE() function, 188; =NORMDIST() function, 203–204*fig*, 207–209, 224–227*fig*; number of arguments and, 40; =POISSON() function, 176*fig*–177; =RAND() function, 61, 74–76, 217; spelling and syntax of function name, 40–41; =SQRT() function, 192, 410*t*; =STDEV() function, 190–192; =STDEV(P) function, 200, 224; =SUMPRODUCT() function, 366–367*fig*, 394–395*fig*; =SUMSQ() function, 366–367*fig*, 369, 371*t*, 395*fig*, 407, 408, 409*t*; =TDIST() function, 230–231*fig*, 260, 293–294, 314, 408, 410*t*; =TINV function, 231–232, 240–241; =TRANSPOSE() function, 402–403, 404, 409*t*, 475–477; =VAR() function, 190–191. *See also* Data Analysis package; Icons; Microsoft Excel
- Excel functions menu, 30
- Excel Options button, 57*fig*–58
- Excel Options dialog box, 58*fig*
- Excel's chart pop-up menu, 49*fig*
- Excel's Sort routine: random sample using, 76, 78; result of, 78*fig*; Sort dialog box, 77*fig*
- =EXP() function, 178
- Explained variance, 319–320
- Exponential model, 449, 450*fig*
- F**
- F* tests: adjusting *F* probability to protect against alpha inflation, 329–330; ANOVA Single-Factor Data Analysis add-in for,

- F* tests: (*continued*) 324–326; definition of, 322; dichotomous dependent variable OLS confirmed by, 462–463*fig*; for differences in ANOVA, 322, 324; differences in one-way analysis of variance using, 328–330; factorial analysis of variance and use of, 347; =FDIST() function for, 324; independence hypothesis testing using both *t* and, 373–374; interpreting independence hypothesis, 372–373; rejecting/accepting null hypothesis using, 329; testing for equal variance using, 306, 307*fig*
- =FACT() function, 168–169*fig*
- Factorial analysis of variance: ANOVA:
Two-Factor with Replication add-in for, 347, 350*fig*, 350*fig*–352*fig*; Data Analysis package for, 350–351*fig*; degrees of freedom in two-factor, 346–347*fig*, 353*fig*; description of, 343; example applied to hospital admissions, 343–347
- Factorial design: calculating within cell variation, 353–355; dividing within cell variation in repeated measures, 352–353; repeated measures in a, 351–355
- =FDIST() function, 324, 371*fig*, 408, 410*t*
- Finite population correction (fpc), 215
- =FINV() function, 326–327
- Format Data Series menu, 444, 445
- Format Trendline dialog box, 445–448
- Formatting: data after importing into Excel, 90–91*fig*; \$, 92–93
- Formula Ribbon: description of, 32; Excel functions (f_x menu through, 31, 37–41*fig*)
- Forward stepwise inclusion, 430–432, 433–434
- Frequency distribution: binomial, 164–173*fig*; chart function to graph, 109–115; conditional probabilities and, 160; cumulative, 115*fig*–117*fig*; defining bins for, 105–106*fig*; definition of, 199; histogram tool to graph, 107*fig*–109*fig*; normal, 118*fig*, 120–121, 180–182, 202–209; using pivot table to generate categorical variables, 123–134; Poisson, 84, 175–179*fig*; proportion, 224–227*fig*; skewed, 118*fig*–120; standard deviations and, 199–202; *t* distribution, 228–232, 295; uniform, 81*fig*–82, 118*fig*, 121. *See also* Discrete distribution
- =FREQUENCY() function: description of, 60*fig*, 104; frequency distribution using, 105–106*fig*, 201–202; =MIN() and =MAX(), 104–105*fig*; putting values into more than one cell using, 476; structure and use of the, 107. *See also* Arguments
- Function arguments dialog box, 38–39*fig*
- Function name syntax, 40–41
- f_x icon. *See* Excel functions (f_x)
- ## G
- Graphs: basic bar, 48*fig*; Chart function for frequency distribution, 109–115; Excel Charts function for working with, 47–50; formatting finished, 50–51*fig*
- ## H
- HDI (Human Development Index), 199–202*fig*
- Health care personnel's hourly rate, 121
- Health predictors, 435*fig*
- Healthcare Financial Management*, 12
- Highlighting: a cell, 33–34*fig*; an entire column, 34–35*fig*
- Histogram dialog box, 108*fig*
- Histogram tool: final output of example using, 109*fig*; graphing frequency distribution using, 107–109
- Human Development Index (HDI), 199–202*fig*
- Hypothesis testing: acceptance or rejection of initial hypothesis, 248–249; calculating value of beta for any alternative hypothesis, 258–260; creating hypothesis statements for, 247–248; definition of statistical context of, 246–247; of independence, 368–374; mutual dependence of alpha and beta, 253–254; reality of many betas for, 254–258; sample size and, 261–263; small alpha or small beta and cost of, 260–263. *See also* Type I error; Type II error
- ## I
- ICD-9-CM code, 21
- Icons: Chart Wizard, 129; PivotTable, 138*fig*. *See also* Excel functions (f_x)
- =IF() function: for making decisions within spreadsheet, 43*fig*–44; nested, 44–45*fig*
- =IF() statement: checking date of birth with an, 93–94; checking M or F with an, 94–95
- Imputation, 101

- Independence: ANOVA assessment of, 270; determination of two variables, 238, 270–279; mathematical, 270*fig*–279; statistical, 272–279; testing hypothesis of, 368–374
- Independence hypothesis testing: calculating coefficient of determination for, 369–370; calculating standard error and *t* tests, 373; calculation of total variance for, 368; coefficient b_0 and, 374; dividing total variance for, 368–369; interpreting *F* test for, 372–373; testing null hypothesis, 372; understanding explained versus unexplained variance for, 370–372*fig*
- Independent events: conditional probability and, 159; conditional probability tables via spreadsheets to test, 161–162*fig*; definition of, 146–147; mathematical versus statistical, 162
- Independent variables: definition of, 20; distinction between dependent and, 69–70; linear regression between dependent and, 361; multiple regression for relationship of dependent and, 389*fig*–390; termed as causal variables, 69. *See also* Dependent variables
- Infant mortality graphs: for 149 countries, 119*fig*; for the United States, 120*fig*
- Inferential statistics color code, 7
- Information matrix: calculation of the, 475; Excel functions used for, 475–477; Logit model use of, 474; *t* tests and, 477*fig*
- Initial Regression dialog box, 390*fig*–391
- Insert Chart dialog box, 47*fig*
- Insert Function dialog box, 37–38*fig*
- Interaction effects: multiple regression using dummy variables and, 418–424*fig*; repeated measure in Factorial design and no, 352
- Interval scale, 22
- Interval variables, 68–70
- Inverse array, 404*fig*–405
- J**
- Joint probability: “or” value for shift and emergency status, 158*fig*–159; definition of, 154–155; mutual exclusivity and addition rule for, 157–158*fig*; mutual exclusivity and simple addition rule on, 157; using spreadsheet for building contingency table of, 155–157
- L**
- Layout button, 125
- Layout dialog box, 125–128, 131*fig*
- Likert scale, 22
- Line chart, 112
- Linear probability model, 454
- Linear regression: best-fitting straight line to data, 361–362*fig*; calculating regression coefficients, 365–367*t*; definition of, 360; between dependent and independent variables, 361; determining, 364–365; examples of, 360*fig*; Excel Data Analysis package used for, 377; meaning and calculation of, 360–367*fig*; meaning in practical terms, 362–364*fig*; multiple regression as extension of simple, 388–389; Regression dialog box used for, 375, 376*fig*, 377, 377–379; slope intercept form, 362. *See also* Regression analysis
- Log coefficients: Excel’s solver add-in for solving, 470, 471–474; information matrix based on, 474–477*fig*; significance of individual, 474–475*fig*; standard errors for, 477–478
- Log nonlinear relationships, 442*fig*–443*fig*
- Log-likelihoods: Excel for finding, 468–469*fig*; Excel’s solver add-in for solving coefficients of, 470, 471–474; significance of individual coefficients, 474–475*fig*; spreadsheet layout for maximizing, 469–470*fig*
- Logarithmic model, 448
- Logarithmic option, 445–446
- Logit models: equation of the, 467–468; finding coefficients of, 468; introduction to, 467
- Logit relationship analysis: applications of, 70; calculation of chi-square for, 473–474*fig*; comparing OLS, WLS and, 479–480*fig*; definition of, 26; dichotomous dependent variables estimated using, 465, 466–478; finding log-likelihoods using Excel, 468–478; setting up the, 466. *See also* Regression analysis
- LOS (length of stay): calculating linear relationship of, 364–365; calculating the, 95–96*fig*; comparison of models in graphical form for, 424–425*fig*; establishing statistical difference between hospitals, 9; factorial analysis of variance for, 343–350*fig*; graph of cost data by, 418*fig*; for one year of

- LOS (length of stay): (*continued*)
 discharges from 200-bed hospital, 211*fig*;
 using only interaction term as predictor of,
 423–424*fig*; statistical applications related to,
 8–9; XY(Scatter) chart of, 114*fig*–115. *See also* Admission data
- M**
- Main effects, 346, 347
- Marginal probability: definition of, 154;
 mathematical independence link to, 273;
 mutual exclusivity and, 157
- Mathematical independence: concept of,
 270*fig*–273; marginal and conditional
 probabilities linking to, 272; variables and
 statistical, 272–279
- Mathematical operations: Excel arithmetical
 conventions used for, 31*fig*; matrix math
 functions, 61; overview of, 31–32
- Matrix math functions, 61
- Matrix/matrices: addition and subtraction of,
 399; definition of, 398; Excel capabilities for,
 61, 402–405; finding determinant of,
 401–402; finding the determinant of a,
 401–402; introduction to, 398–399;
 multiplication of, 400; scalars
 and multiplication of, 400. *See also* Arrays
- Maximum likelihood estimators, 455
- =MDETERM() function, 402–403
- Mean (or average): calculating discrete
 numerical variable, 222–223*fig*; definition
 of, 186*fig*, 187; measures of dispersion
 and being centered at the, 196–197; using
 representative sample to infer, 238–239;
 sampling distribution of the, 210–220*fig*
- Median, definition of, 186*fig*, 187
- =MEDIAN() function, 188
- Medicare payments graph, 119*fig*
- Medicare reimbursement claims (FORM
 CMS-485): binomial probability of correctly
 documented, 172; description of, 5;
 statistical application for, 5–6; visualizing
 binomial distributions of correctly
 documented, 172–173*fig*
- Microsoft Excel. *See* Excel
- =MINVERSE() function, 402–403, 405, 477
- Missing data: case deletion remedy for,
 100; description of, 100; imputation remedy
 for, 101
- =MMULT() function, 61–62*fig*, 402–403, 407,
 409*t*, 475–477
- Mode, definition of, 186*fig*, 187–188
- =MODE() function, 188
- Modifying: data after importing into Excel,
 90–91*fig*; dates in Excel, 91–92
- Monte Carlo technique, 101
- Moving Average model, 449
- MS-DRG categories: automatic chart option
 for generating pivot table, 129–130; chart by
 size, 133*fig*; first ten records showing broad,
 130*fig*; Layout dialog box for, 131*fig*; Pareto
 chart of, 133*fig*; pivot table for, 131–132*fig*;
 pivot table for sex and, 136*fig*–137
- Multicollinearity: calculating correlation
 coefficient, 438; description of, 437;
 examining, 437*fig*; testing for standard error
 inflation, 438–439*fig*
- Multiple R, 437. *See also* R square
- Multiple regression analysis: using b
 coefficients to generate, 406–407*fig*;
 calculating all results of, 407–410*r*;
 calculating the x and y summations,
 394–395*fig*; calculus behind the formulas
 for, 393–394; comparison of models in
 graphical form, 424–425*fig*; correlation and
 multicollinearity extension of,
 437*fig*–439*fig*; cultural predictors used in,
 435*fig*; definition of, 388; dummy variables
 in, 416–425*fig*; Excel output with scientific
 notation, 405–406; Excel's Data Analysis
 add-in for, 390–391; as extension of linear
 regression, 388–389; health predictors used
 in, 435*fig*; interpreting the output of Excel's,
 391–392*fig*; matrices and, 398–410*r*; purpose
 of, 389–390; R Square obtained from,
 391–392*fig*; solution to problem of,
 392–393; solving simultaneous equations
 using Excel, 394; successive elimination for
 b_j, 396–397
- Multiple regression models: alternative look
 at importance of theory in, 434–435*fig*;
 arguments for theory of, 427; backward
 stepwise elimination, 430, 431*fig*,
 432–433; Excel's stepwise regression,
 432–434; forward stepwise inclusion,
 430–432, 433–434; regression variables
 versus no regression variables use in,
 428*fig*–430

Mutual exclusivity: addition rule and, 157–158*fig*; definition of, 148–149; simple addition rule and, 157; Venn diagram of two events without, 157–158*fig*

N

n-by-*n* tables, 284–285

n-by-two table, 282*fig*–283

Nested =IF() functions: description of, 44–45; screenshot of, 45*fig*

Noncontiguous cells: description of, 41; summing two noncontiguous areas, 41*fig*

Nonemergency visits, 166, 167*fig*. *See also* Emergency room data

Nonlinear relationships: dichotomous dependent variables and, 466; dummy variables for fitting, 443–444; estimating, 440*fig*; Excel's automatic curve fitter for, 444–445*fig*; Excel's Format Trendline dialog box for, 445–448; Log, 442*fig*–443*fig*; using second-degree curves to fit, 440–442

Nonzero determinant, 402

Normal distribution dialog box, 82–83

Normal frequency distribution: characteristics of, 182, 204–205; continuous, 180; definition of, 181*fig*, 202–203; description of, 118*fig*, 120–121; empirical probability and, 153; Excel and cumulative, 205–209; =NORMDIST() function for, 203–204*fig*; standard deviation and, 181; *t* distribution approximating the, 229–230. *See also* Bell-shaped curve

=NORMDIST() function: calculating population proportions using, 224–227*fig*; constructing normal distribution using, 203–204*fig*; cumulative normal distribution using, 207–209

Null hypothesis H_0 : beta (β), one-tail test and, 256–258; calculating beta (β) for any alternative, 258–260; controlling beta (β) and, 258; description of, 247–248; *F* tests and rejecting/accepting, 329; independence hypothesis testing of, 372; reality of many betas and, 254–255. *See also* *t* tests

#NUM!, 277, 407, 460, 472

Numerical variables: used in data sets, 67; definition of, 21; transforming categorical variables into, 98*fig*–99

O

OLS (ordinary least squares) dichotomous dependent variable and, 458–459; comparing Logit analysis, WLS and, 479–480*fig*; immunization data results using, 459*fig*

One-tail tests: association between hypothesis and, 294; beta (β) and, 256–258; confidence limits for one-tail versus, 241; region of rejection for, 297*fig*; *t* test, 296–298

One-way analysis of variance: calculated for four hospitals, 321*fig*–322; definition of, 318; differences in, 327–332*t*; equations for, 319–321; =FDIST() function for, 324; =FINV() function for, 326–327

=OR() statement, 94–95

Ordinal scales, 22

Ordinal variables, 68

Outcomes (mutually exclusive), 148–149, 157–158*fig*

P

Pareto chart: creating a, 132, 134; MS-DRG categories, 133*fig*

Paste Special dialog box, 77*fig*

Pentad Home Health Agency, 5–6

Percentage distributions: creating, 115, 116; modifying chart of, 116–117

Pie chart, 113*fig*–114

Pivot table: automatic chart option for generating, 129*fig*–130; Create PivotTable dialog box, 138*fig*; creating Pareto chart, 132, 134; description of, 124; using Excel 2007 to generate one variable, 137–140*fig*; using Excel 2007 to generate two variable, 140–142*fig*; frequency distribution represented by, 128–129; Layout dialog box for working with, 125–128; logical extension to two variables, 135–137; for MS-DRG categories, 131–132*fig*; for MS-DRG categories and sex, 136*fig*–137; Pivot table dialog box for working with, 124*fig*–125*fig*, 128, 136*fig*; using pre-2007 versions of Excel, 124–125; for sex, 140*fig*; for sex and DRG category, 143*fig*. *See also* Cross-tab (cross-tabulation)

Pivot table layout screen, 139*fig*

PivotTable icon, 138*fig*

- Poisson distribution: applied to emergency room visits, 176*fig*–179*fig*; definition of, 175–176; equations for calculating, 177–178*fig*; =POISSON() function for, 176*fig*–177; random numbers from, 84; spreadsheet used to calculate, 178–179*fig*
- =POISSON() function, 176*fig*–177
- Polynomial option, 448
- Populations: calculating proportions of, 224–227*fig*; comparison of variance, 214*fig*; confidence interval of, 238–242; definition of, 14–15; drawing a random sample from a, 73–78*fig*; finite population correction (fpc) of, 215; parameters as referring to, 73; sampled, 17; target, 16.
See also Samples
- Power option, 448–449
- Practical significance, 281–282
- Predictor variable, 373–374, 388
- Prenatal visit probabilities, 216*fig**gw*
- Probability: binomial, 83–84, 164–173*fig*; concepts and definitions related to, 146–153; conditional, 159–162*fig*, 272; definition of, 146; empirical, 149–153; joint, 154–159; marginal, 154, 272; normal distribution of, 118*fig*, 120–121, 153, 180–182; Poisson distribution of, 84, 175–179*fig*; of prenatal visits, 216*fig*; a priori, 149; statistical applications related to, 6; stochastic process and, 146; *t* distribution, 231*fig*. *See also* Coin flips; Statistics
- Probability terminology/concepts: empirical probability, 149, 150–153; events, 146, 149*fig*; independent events, 146–147; mutually exclusive outcomes, 148–149; possible combinations of five coin flips, 148*fig*; priori probability, 149–150; sample space, 147
- Probit relationship, 455, 466
- Proportion: calculating population, 224–226; subtracting normal distribution from, 226–227*fig*
- Pseudo R square, 478
- R**
- R square: dichotomous dependent variable using, 463–465; generating regression coefficients and, 446; Logit problem and pseudo, 478; multiple regression using, 391–392*fig*. *See also* Multiple R
- R² (coefficient of determination), 369–370*fig*
- =RAND() function: copying random numbers using, 76; description of, 61; drawing samples using, 217; random number generation via spreadsheet using, 74–76; random number regeneration using, 76; using the, 75*fig*–76
- =RANDBETWEEN() function, 80
- Random number generation add-in, 216
- Random Number Generation dialog box, 81*fig*, 82, 217–220*fig*
- Random number tables, 74
- Random numbers: Bernoulli distribution generating, 83; Binomial distribution generating, 83–84; Discrete distribution generating, 84–85*fig*; five sets of, 83*fig*; Normal distribution generating, 82–83; Poisson distribution generating, 84; Uniform distribution generating, 81*fig*, 82
- Random sample techniques: data analysis add-in as, 80–81; generated from Uniform distribution, 81*fig*, 82; home health agency records example of, 78–79; =RAND() function as, 74–76; =RANDBETWEEN() function as, 80; Random Number Generation dialog box as, 81*fig*, 82; random number tables, 74; selecting cancer education intervention example of, 79–80*fig*
- Random samples: definition of, 15–16; drawing from population, 73–78*fig*
- Ratio scale, 22
- Readmission rates, 8–9
- Regression analysis: applications of, 70; definition of, 26; emergency trauma (ED) color code, 7; establishing statistical difference using, 9–10; problems and issues of, 12–13; relationship between *t* test and, 382–384. *See also* Linear regression; Logit relationship analysis
- Regression coefficients: calculating, 365–366*fig*; definition of, 365; with dummy and interaction, 420–421*fig*; with dummy only, 420*fig*; using Excel to calculate, 366–367*fig*; formulas for, 393–394
- Regression dialog box, 375, 376*fig*, 377, 377–379, 462

- Regression package, 461–462
- Relationships: Logit, 26, 70, 465–480*fig*;
 multiple regression of independent/
 dependent variables, 389*fig*–390; nonlinear,
 439–450*fig*; Probit, 455, 466; regression
 analysis and *t* test, 382–384; visual
 inspection of scatterplots, 380–381. *See also*
 Correlation coefficient
- Repeated ANOVA: calculating effect versus
 residual variation in, 335–340; degrees of
 freedom in, 340*fig*; description of, 334–335;
 in factorial design, 351–355; formulas for,
 337*t*–338*fig*, 342*t*; observed differences
 in, 340–342*t*; results of, 337*fig*; variation
 within, 335
- S**
- Sample size: confidence level and, 265; how
 to determining, 266–267; hypothesis testing
 and, 261–263; standard deviation and,
 265–266*fig*; *t* test comparing groups of
 different, 302–303
- Sample space, 147
- Sampled population, 17
- Samples: calculating and interpreting means of,
 212*fig*–213; calculating confidence limits for
 multiple, 243–246; cluster, 19–20; definition
 of, 15; finite population correction (fpc) of,
 215; inferring a mean using representative,
 238–239; integrating standard error and,
 240; nonrandom, 15–20; random, 15–20,
 73–85*fig*; random number generation of,
 216–219*fig*; simple random, 18; statistics
 as referring to, 72–73; stratified, 18–19, 20;
 systematic, 17–18. *See also* Populations
- Sampling: distribution of the mean,
 210–220*fig*; statistical applications related
 to, 6; *t* test comparing two groups using
 stratified, 299. *See also* Data
- SAS, 4
- Scales: distinguishing variables as, 68; interval,
 22; ordinal, 22; ratio, 22
- Scatterplots: constant variance assumption
 and, 381–382*fig*; definition of, 380; visual
 inspection to identify relationships, 380–381
- Sea Coast Alliance: establishing statistical
 differences between hospitals of, 9–11;
 statistical applications used by, 8–9
- Secondary data, 87
- Select Data Source dialog boxes, 48–50*fig*
- Simple random samples, 18
- Skewed frequency distribution, 118*fig*–120
- Solver add-in, 470, 471–474
- Sort dialog box, 53*fig*–56*fig*, 77*fig*
- Sorting data, 53*fig*–56*fig*
- Spreadsheets: arithmetical conventions used in,
 31*fig*; description of, 30, 31*fig*; highlighting
 cells, 33–34*fig*; =IF() function to make
 decisions within, 43*fig*–45*fig*; maximizing
 log-likelihoods using, 469–470*fig*; moving
 around a data set, 33*fig*; moving data with
 drag and drop, 36; Poisson distribution
 calculating using, 178–179*fig*; =RAND()
 function for generating random number
 generation via, 74–76; testing independence
 using conditional tables via, 161–162*fig*;
 Undo command for undoing changes in,
 36–37*fig*; workbooks groups of, 30; working
 and moving around, 32–37. *See also* Cells
- SPSS, 4
- =SQRT() function, 192, 410*t*
- Standard deviation: calculating discrete
 numerical variable, 222–223*fig*; description
 of, 181; as dispersion measurement, 189,
 190; frequency distribution and, 199–202;
 independence hypothesis testing, 373;
 normal distribution and, 181; sample size
 and, 265–266; variance calculated by,
 190–192. *See also* Central tendency
- Standard error inflation test, 438*fig*–439*fig*
- Standard errors: calculating b coefficient,
 407*fig*; concept of, 213–214; definition of,
 213; with estimated variance, 214–215;
 integrating sample mean and, 240; with
 known variance, 211–214; log coefficients,
 477–478; multicollinearity and testing for
 inflation of, 438–439*fig*; when sample size is
 large relative to population size, 215
- STATA, 4
- The State of the World’s Children, 119,
 428*fig*–429*fig*, 437
- Statistical applications: using analysis of
 variance versus *t* tests, 11–12; the “big
 picture” of using, 13–14; calculating
 standard hourly rate for health care
 personnel, 12; descriptive statistics,

- Statistical applications: (*continued*)
 confidence limits, and categorical data, 7; emergency trauma color code, 6–7; establishing statistical differences between two groups, 9–11; inferential statistics, analysis of variance, and regression, 7; length of stay, readmission rates, and cost per case in hospital alliance, 8–9; Medicare reimbursement claims, documentation, 5–6; relating variables via regression analysis, 12–13; sampling, data display, and probability, 6; study of effectiveness of breast cancer education, 11; two caveats of using statistics and, 8
- Statistical difference: caveats of establishing, 8; using regression, 9–10; using statistical tests, 10–11; using *t* tests and analysis of variance, 9
- Statistical independence: chi-square and Type I and Type II errors, 277–279; chi-square critical value for, 275*fig*–276; chi-square test for, 273–274; =CHIDIST() function, 274–275; =CHINV() function for, 275; =CHITEST() function for, 276–277*fig*; description of, 272–273
- Statistical significance, 281–282, 308
- Statistical terminology: interval scale, 22; ordinal scale, 22; populations and samples, 14–15; random and nonrandom samples, 15–20; ratio scale, 22; variables, 20–21, 22–23
- Statistical tests: introduction to ANOVA (analysis of variance), 7, 9, 11–12, 26, 70; introduction to chi-square, 10–11, 25, 70; introduction to establishing statistical difference using, 10–11; introduction to Logit analysis, 26, 70; introduction to regression analysis, 7, 9–10, 12–13, 26, 70; introduction to *t* tests, 9, 11–12, 25–26, 70. *See also specific test*
- Statistics: caveats to establishing causality using, 8; probability basis of, 146; as referring to samples, 72–73. *See also Probability*
- Statistics for Managers Using Microsoft Excel* (Levine, Stephan, Krehbiel, and Berenson), 4, 380
- =STDEV() function, 190–192
- =STDEVP() function, 200, 224
- Stepwise regression: backward stepwise elimination model for, 430, 431*fig*, 432; definition of, 430; Excel's approach to, 432–434; forward stepwise inclusion model for, 430–432, 433–434
- Stochastic process, 146
- Stratified samples: cluster versus, 20; definition of, 18–19; *t* test comparing two groups using, 299
- Studies in Family Planning*, 283
- Sum of squared differences, 197*fig*
- =SUMPRODUCT() function, 366–367*fig*, 394–395*fig*
- =SUMSQ() function, 366–367*fig*, 369, 371*fig*, 395*fig*, 407, 408, 409*t*
- Syntax (function name), 40–41
- SYSTAT, 4
- Systematic samples, definition of, 17–18
- ## T
- t* distribution: approximating normal distribution, 229–230; definition of, 228; degrees of freedom concept and, 229*fig*; exact probabilities of, 231*fig*; finding percentage of observations along, 230–232; illustration of two, 230*fig*; shape of the, 295; =TDIST function for, 230–231*fig*; =TINV function for, 231–232
- t* test add-in: description and differences in results of, 308–311; for related data, 315–316
- t* test comparisons: assumption of equal variance in, 305–306; *F* test for testing equal variance, 306, 307*fig*; pooled variance for unequal sample sizes, 303–305; samples of unequal size, 302–303; stratified sampling approach to selecting groups, 299; for two means, 299–302; unequal variances and, 306–308
- t* test for equal variance dialog box, 309*fig*
- t* tests: ANOVA (analysis of variance) relationship to, 319; applications of, 70; comparing ANOVA to, 11–12; comparing two groups using, 298–311*fig*; definition of, 25–26, 292; dichotomous dependent variable OLS confirmed by, 462–463*fig*; differences in results of Excel, 309–311;

- establishing statistical difference using, 9; independence hypothesis testing, 373; independence hypothesis testing using both *F* and, 373–374; information matrix and, 477*fig*; interpreting results of, 294; origins and meaning of, 294; for related data, 313–316; relationship between regression analysis and, 382–384; results for unequal variance, 311*fig*; setting up hypotheses for, 293; =TDIST() function for finding, 293–294; Type I and Type II errors links to, 295–296. *See also* Null hypothesis *H*(0)
- t* values: calculating alternative hypothesis, 259–260; finding confidence level, 240–241; one-tail versus two-tail, 241
- Target populations, 16
- =TDIST() function, 230–231*fig*, 260, 293–294, 314, 408, 410*t*
- Template for chi-square statistic, 275*fig*–276
- Text Import Wizard: description of, 87–88*fig*; steps for importing data with the, 88–90*fig*
- =TINV function, 231–232, 240–241
- =TRANSPOSE() function, 402–403, 404, 407, 475–477
- =TRUNC() function, 97
- Two-by-*n* table, 283–284
- Two-by-two table, 280–285*fig*
- Two-tailed tests: association between hypothesis and, 294; confidence limits for one-tail versus, 241; region of rejection for, 297*fig*; *t* tests, 296–298
- Type I errors: chi-square and, 277–278; definition of, 250–251; mutual dependence of beta and, 253–254; small alpha/beta and avoiding, 263–264; *t* tests link to, 295–296. *See also* Alpha (α); Hypothesis testing
- Type II errors: chi-square and, 277, 278–279; definition of, 251–253; mutual dependence of alpha and, 253–254; reality of many, 254–255; small alpha/beta and avoiding, 263–264; *t* tests link to, 295–296. *See also* Beta (β); Hypothesis testing
- U**
- Undo button, 36–37*fig*
- UNICEF, 119, 428
- Uniform frequency distribution: creating a, 118*fig*, 121; random numbers generated from, 81*fig*, 82; selecting corresponding parameters for, 81–82
- United Nations Development Program (UNDP), 199
- U.S. National Center for Health Statistics, 120
- V**
- Value Field dialog box, 141*fig*
- =VAR() function, 190–191
- Variables: categorical, 21, 67, 98–99, 123–134; continuous numerical, 23, 67–68; definition of, 20; dependent, 20, 69–70, 389*fig*–390, 418–419*fig*; determination of independence of, 238; dichotomous dependent, 454–480*fig*; discrete numerical, 22–23, 221–223*fig*; distinguishing as a scale, 68; dummy, 98, 416–425*fig*, 443–444; using Excel 2007 to generate pivot table with one, 137–140*fig*; using Excel 2007 to generate pivot table with two, 140–142*fig*; independent, 20, 69–70, 389*fig*–390; interval, 68–69; logical extension of pivot table to two, 135–137; numerical, 21, 67, 98–99; ordinal, 68; predictor, 373–374, 388; result of data sorting on two, 56*fig*; statistics that apply to different types of, 70–71; *x*, 381; *Y*, 375. *See also* Correlation coefficient; Data sets
- Variance: Bartlett test for homogeneity of, 330–332; calculating standard deviation with, 190–192; calculation of total, 368; comparison of population, 214*fig*; dividing total, 368–369; explained, 319–320; explained versus unexplained, 370, 372; *F* test for testing equal, 306; reasons for dividing by *n*-1, 192–195; results of *t* test for unequal, 311*fig*; screenshot of calculation of, 191*fig*; standard error with estimated, 214–215; standard error with known, 211–214; *t* test comparing unequal sample sized and pooled, 303–305; *t* test comparison and assumption of equal, 305–306; *t* test comparison and unequal, 306–308; testing hypothesis of independence and, 368–369. *See also* ANOVA (analysis of variance)
- Vectors, 403

Venn diagrams: of two events that are not mutually exclusive, 157–158*fig*; two mutually exclusive events, 148–149*fig*

W

Within group variance, 320–321

WLS (weighted least squares): comparing Logit analysis, OLS and, 479–480*fig*; example of using, 460–461*fig*; Excel's Regression package for using, 461–462; for generating weights greater than 1/less than 0, 460; linear probability models estimated using, 460

X

x axis, 364, 446

x variable, 381

XY(Scatter) chart, 114*fig*–115

Y

y axis, 362

y variable, 375

Yates's correction, 287*fig*–288, 457

=YEARFRAC() function statement, 96*fig*–97*fig*