

Index

- Absolute values, 72
 mean, 72
- Addition process, 68
- Addition rule, 101
- Adjusted rate, *See* Standardized rate
- Age-adjusted death rate, 17
- AIDS, 12–14
 cases, 12–14
 transmissions, 12
- Alternative hypothesis, 173, 180
- Amount of uncertainty, 136
- Analogies, 177
 common expectations, 178
 medical screening tests, 177–178
 trials by jury, 177
- Analysis of variance (ANOVA), 225–231,
 240, 241
 approach, 252–254
 example, 227–229
 one-way, 225, 227, 229, 230, 240
 procedure, 230
- Analysis of variance table, 227, 228, 231, 253,
 258, 260–262
- ANOVA method, 225, 226, 229
 one-way, 229
- ANOVA model, 225, 226
 one-way, 225, 226
- Arithmetic average, *See* Arithmetic mean
- Arithmetic mean, 68, 69
- Association, 74–76, 78, 107, 248
 negative, 75, 76, 248
 positive, 75, 76, 78, 248
- Autosum icon, 30
- Average distance, 73
- Average gap, *See* Average
 distance
- Average infant mortality, 125
- Average variation, 227
- Bar chart(s), 8, 10, 28, 29, 79
 example, 8
 horizontal axis, 8
 use, 8
- Bayes' theorem, 106
- Before-and-after intervention, 145
- Bell-shaped distribution, 111
- Bi-modal distribution, 59
- Bi-modal pattern, 59
- Binary data, 1, 148, 186, 188, 191
 one-sample problem, 186
 sample, 186
- Binary variables, 196
- Binomial distribution, 121–124
 mean, 123
 variance, 123
- Breast cancer, 99, 186–188

- Cancer screening test, 101
- Case-control data, 20
- Case-control study(ies), 3, 19, 21, 22, 24, 146, 153, 194, 196, 197, 199
 - pair-matched, 146
- Categorical data set, *See* Discrete data set
- Cell frequencies, 30
- Cell probabilities, 107
- Centers for Disease Control (CDC), 12, 155
- Central limit theorem, 111, 112, 126, 140–142, 148, 149, 152, 172, 175, 176, 182, 187, 216
- Certainty of uncertainty, 99
- Cervical cancer, 6, 7
- Chance concordance, 108
- Change rate, 12
 - definition, 12
 - example, 12
- Chart wizard, 28, 79, 263
 - icon, 28, 79
- Chi-square distribution(s), 127, 202, 279
 - percentiles, 279
- Chi-square statistic, 193, 202
- Chi-square test, xx, xxii, 196
- Coefficient of determination, 251
- Cohort studies, 14
 - epidemiological designs, 14
 - termination, 14
- College algebra, xxi
- Comparative studies, 2
 - prospective, 2
 - retrospective, 2
- Computational aids, 26, 127, 201–202, 229–231, 262–264
 - bar/pie charts, 28
 - computer screen, 27
 - cut and paste, 27
 - forming 2×2 tables, 30
 - formula bar, 27
 - normal curves, 127
 - select and drag, 28
 - t curves, 128
- Computation note, 158
 - confidence intervals, 158
- Concordance, 107, 108
 - category-specific proportions, 107
 - proportion, 107
- Conditional probabilities, 103–105
 - negative predictivity, 105
 - positive predictivity, 105
 - sensitivity, 105
 - specificity, 105
- Confidence estimation, 135, 139
 - introduction, 139
- Confidence interval(s), 136, 140, 143, 151, 153, 156, 158, 182
 - computations, 158
 - relationship, 182
- Confounder, 3, 4, 15, 21, 22, 24, 25
 - potential, 3, 15
- Confounding variable, 146, 196
- Connecticut tumor registry, 190
- Continuous data, 125, 215
 - example, 216–217
 - one-sample problem, 215
 - probability models, 125
- Continuous data set, 52, 186, 215
- Continuous distributions, 126
 - Chi-square distribution, 126
 - F distribution, 126
 - t distribution, 126
- Correlation analysis, 74
- Correlation coefficient, *See* Pearson's correlation coefficient
- Covariates, 245, 254
 - continuous, 254
 - dichotomous, 254
 - polytomous, 254
- Criminal court, 10, 13, 15, 171, 172, 177
- Crude death rate, 13, 16, 52–53
 - age-specific, 13, 16
 - cause-specific, 13
 - definition, 13
- Cumulative distribution, 64
 - children weights, 64
- Cumulative frequency graph, 63, 65, 66
 - example, 66–68
 - percentile scores, 65
 - true boundaries, 63
 - upper class boundary, 63
- Cumulative frequency percentiles, 63, 65
- Cumulative frequency polygons, 64
- Cumulative relative frequency, 63
- Cytological test, 6
- Data analysis, 262
- Data file, 30, 93, 202
- Data organization, 52
 - graphical methods, 52
 - tabular methods, 52

- Data presentation, 52
 - graphical methods, 52
 - tabular methods, 52
- Data range, 80
- Data set(s), 3, 9, 30, 53, 54, 56, 59, 65, 68–71, 263
- Data summarization, 52, 68
 - numerical methods, 68
- Data transformation, 28
- Daycare children, 58, 60
 - distribution of weights, 58, 60
- Decision-making process, 173, 174
- Decision-making rule, 175–177, 179
- Degree of confidence, 99, 140, 143, 147
- Degree of freedom, 73, 126, 127, 144, 147, 222, 227, 250, 251, 253, 258, 259, 278, 280
- Degree of variation, 71, 226
- Density curve, 111, 112, 125, 127
- Department of Health and Human Services, xvi
- Dependent variable, 243, 244, 248, 249, 253, 263
- Descriptive measure, 98
- Deterministic relationship, 74
- Diagnostic procedures, 5, 6
 - sensitivity, 6
 - specificity, 6
- Dichotomous data, *See* Binary data
- Difference of means, 145, 147
- Difference of proportions, 152
- Discrete data, 7, 52, 79, 121, 186, 215
 - graphical methods, 79
 - probability models, 121
- Discrete data set, 52, 186, 215
- Discrete distribution(s), 121, 124
 - binomial distribution, 121
 - Poisson distribution, 121
- Discrete variable, 121
- Descriptive statistics, 80, 81, 98, 127, 158, 224, 229, 231
 - mean, 127, 158
 - standard deviation, 127, 158
- Disease prevalence, 5, 14, 105
 - definition, 5
 - proportions, 5
- Dispersion, 68, 71
 - measures, 71
- Displaying proportions, 7–12
- Dots clusters, 75
- Dummy factor, 30
- Dummy variables, 254
- Effect modification(s), 4, 255, 256
- Effect modifier, 22
- End-stage renal disease (ESRD), 14, 18
- Epidemiological studies, 5, 19, 220
- Error mean square, 253, 258
- Error sum of squares (SSE), 252, 253, 257, 258
- Excel file, 26
 - workbook, 26
- Excel's PivotTable, 202
- Excel worksheet, 26
 - cell reference, 26
 - columns, 26
 - rows, 26
- Expected number of deaths, 16
- Explanatory variable, *See* Independent variables
- Exponentiation, 124, 126, 152
- Exposure history(ies), 2, 4, 14
- F*-distributions, 127, 280
 - percentiles, 280
- Follow-up death rate(s), 14, 18
- Follow-up time, 14, 70
 - definition, 14
- Frequency distribution, 53, 55, 57
 - example, 53
 - menarchal age, 57
 - number of intervals, 54
 - widths of the intervals, 54
- Frequency polygon, 57, 59, 71
- Frequency table, *See* Frequency distribution
- F*-statistic, 231, 259
- F*-test, 227–230, 253, 259–261
- Gaussian distribution, 125
- Geometric mean, 69, 70, 152
 - calculation, 69
- Glaucomatous blindness, 5
 - example, 5
- Gold standard method, 107
- Graduate Record Exam, xix
- Graphical aids, 262–264
 - ANOVA table, 262
 - regression line, 263
- Hand calculation process, 73
- Health norms, 66
- Healthy limit, 120
- Height curves, 67

- Histogram, 57–59, 62, 111
 - boundary, 59
 - horizontal axis, 59
 - horizontal scale, 57
 - rectangular bars, 59
- Histogram polygon, 57
- Horizontal axis, 52
- Horizontal scale, 57
- Hospital-based study, 195
- How to Lie with Statistics, *the book*, xix
- Hypertension model, 119
- Hypotension model, 119
- Hypothesis test(s), 173–174, 179, 183
 - two-sided, 183
- Hypothesis testing process, 171, 175, 177
 - analogies, 177–179
 - errors, 174–175
 - introduction, 171
- Hypothetical data set, 183

- Incidence rate, 14
 - definition, 14
- Independence testing, 250–252
 - example, 251–252
- Independent variable(s), 243, 244, 248–250, 253, 256, 260, 263
- Infant mortality rate (IMR), 124, 125
- Interaction, 4, 254, 256
- Interaction term, 256
- Interval density, 59
 - definition, 59
 - unit, 59
- Interventions evaluation, 145

- Joint probabilities, 102, 104

- Kappa statistic(s), 108, 109
 - category-specific, 109
 - result, 109
- K sample means, 227

- Least squares method, 246, 257
- Line chart, 29
- Line graphs, 10–12, 64
 - example, 10–12
 - horizontal axis, 10
- Log scale, 61, 69, 70, 152–155, 223, 227
 - odds ratio, 153
 - variance, 153
- Lung cancer, 3, 4, 21, 197

- Malaria rates, 11
- Malignant melanoma, 150
 - prevalence, 150
- Mantel–Haenszel method, 21–26, 196–201
 - example, 22–26, 197–201
- Mantel–Haenszel procedure pools data, 22, 197
- Mantel–Haenszel test, 196
- Marginal probability, 102, 103
- McNemar Chi-square statistic, 202
- McNemar’s Chi-square test, 189, 191
- Mean, 69
 - arithmetic mean, 69
 - geometric mean, 69
- Mean concept, 67
- Mean deviation, 71
- Mean square, 226, 227
- Measures of dispersion, 71–73
 - definition, 71
- Measures of location, 70–71
 - mean, 71
 - median, 71
 - mode, 71
- Measures of morbidity, 12, 13
- Measures of mortality, 12, 13
- Measuring agreement, 106–110
 - reliability, 106
 - validity, 106
- Median value, 66, 71
 - advantage, 71
 - definition, 71
 - disadvantages, 71
- Mercaptopurine (6-MP), 6–, 70
- Microsoft’s Excel, 26, 79, 229, 262
 - worksheet/spreadsheet, 26
- Morbidity, 1, 12–15
- Mortality data, 15
- Multiple determination coefficient, 258
- Multiple linear regression model, 255, 256, 258
 - testing hypotheses, 258
- Multiple regression analysis, 254–262
 - example, 259–262
 - overall regression test, 259
 - single variables tests, 259
- Multiplication rule, 103, 106, 108
- Multivariate method, 254
- Myocardial infarction (MI), 24, 151, 152, 199
 - rates, 152

- National rate, 175, 183
- Natural logarithm, 124, 153

- Normal curve, 110–120, 127
 - mean, 127
 - normdist, 127
 - NORMINV, 128
 - shape, 110
 - standard deviation, 127
- Normal distribution, 110, 111, 125, 126, 149
 - mean, 149
 - role, 126
 - standard deviation, 149
 - use, 111
- Null hypothesis, 173, 175, 177–183, 187, 189, 191, 196, 215–221, 259
- Odds ratio (OR), 19–22, 24, 25, 30, 104, 152, 153, 199, 201
 - definition, 104
 - example, 20
- One-sample *t*-test, 218, 229
- One-sided test, 179, 181, 187, 188, 191
- One-way scatter plot, 52, 53
 - advantage, 53
 - horizontal axis, 52
- Oral contraceptives (OCs), 24, 146, 15, 173, 199, 201, 217, 219
 - users, 173, 201
- Pair-matched binary data, 188, 189, 217, 218
 - analysis, 188, 217
 - example, 190–191, 218–221
- Pair-matched case-control study, 188, 217
 - epidemiological design, 188, 217
- Parameters estimation, 135, 246–250, 257
 - example, 246–250
- Pearson's coefficient of correlation, 74, 76, 77, 81, 152, 155, 156, 247, 248, 251
 - ANOVA table, 262
 - regression line, 263
- Percentage saturation, 60, 143
 - frequency polygon, 60
- Percentile scores, 65
- Pie chart, 8–10, 28, 79
 - example, 9–10
 - horizontal axis, 8
 - results, 10
 - use, 8
- PivotTable report, 30
- PivotTable wizard, 30
- Placebo group, 70
- Point estimate, 140, 150, 183
- Poisson density, 124
- Poisson distribution, 121, 124, 126
 - mean, 124
 - variance, 124
- Polynomial regression, 256
- Pooled proportion, 191, 194, 195
- Population(s), 98, 106, 112, 119, 136, 139, 142, 174
 - binary characteristic, 98
 - mean, 112
 - parameter(s), 112, 136, 139, 174
- Population coefficient of correlation, 155, 157
 - confidence interval, 157
 - estimation, 155
 - example, 157–158
- Population mean(s), 135–138, 141–143, 145, 147, 172–174, 179, 186, 215, 221, 250
 - comparison, 215, 221–225
 - confidence intervals, 142
 - estimation, 140–148
 - example, 222–225
- Population odds ratio, 152, 154, 155
 - confidence interval, 154
 - estimation, 152
- Population proportion, 135, 136, 137, 148, 149, 186, 215
 - comparison, 186
 - estimation, 148–152
- Population variance, 126, 144, 216
- Positively skewed distributions, 61
- Predictive value, 105, 106
 - applicational stag, 105
 - developmental stage, 105
 - negative, 105, 106
 - positive, 105, 106
- Predictor variables, 243
- Presence-and-absence classification, 1, 19
- Primary risk factor, 21, 22, 194, 197
- Probability, 98–110, 141, 142
 - rules, 101, 104
 - statistical relationship, 103
- Probability density function, 124, 125, 126
- Probability models, 98, 118, 120–127
 - distributions, 121
 - variables, 121
- Proportion(s), 1–12, 98, 100
 - calculation, 9
 - category-specific, 7
 - comparative studies, 2
 - definition, 98

- Proportion(s) (*Continued*)
 - displaying, 7–12
 - example, 2
 - screening tests, 5
- Proportion of exposure, 3, 4
- Public health intervention, 192
- p*-value, 180–182, 199, 201, 202
 - criterion, 181
 - definition, 182
- Quadratic model, 256
- Quadratic term, 256
- Random sampling, 100, 101, 112
- Random selection, 98–100
 - definition, 98
- Random variable, 136, 142, 158
- Rate(s), 12, 13, 29
 - change, 12
 - crude, 13
 - specific, 13
 - standardized, 13
- Rates standardization, 15, 29
- Ratio(s), 18–20, 22
 - definition, 18
 - odds and odds ratio, 19–21
 - relative risk, 18
- Regional Disease Program, 14
- Registry data, 5
- Regression analysis, 77, 243, 253
- Regression approach, 252
- Regression coefficient(s), 245, 255, 259
- Regression line, 248, 250, 252, 263
- Regression model, 244, 245, 251, 253–256, 258
 - independent variables, 254
- Regression parameters, 245, 255
 - meaning, 245, 255
- Regression sum of squares (SSR), 252, 253, 257, 258
- Rejection region, 179, 180, 187, 188, 192, 216, 222
- Relative frequency, 56, 58, 59, 99–101, 110
 - definition, 56
 - example, 56
- Relative risk (RR), 18–21
- Reproducibility, 107, 140, 180
- Response variable, *See* Dependent variable
- Retrospective study(ies), 2, 3, 20
 - advantages, 2
 - example, 3
 - limitations, 2
- Risk factor(s), 2, 3, 18–20, 145, 152, 190, 243, 245
- Risk ratio, *See* Relative risk
- Sample mean, 138, 139, 141, 142, 148, 172, 186
 - standard error, 148
 - unbiasedness, 139
- Sample proportion, 148, 149, 175, 182, 187, 188
 - standard error, 148, 149
- Sample variance, 216
- Sampling distribution(s), 137–142, 149, 173, 175, 176, 187, 216
 - properties, 138
 - sample mean, 137
 - variance, 140, 141, 173
- Scatter diagram, 75–78, 244–245, 248
 - birth weight data, 75
- Scholastic Aptitude Test, xix
- Screening test(s), 5–7, 99, 104, 106, 178
 - evaluation, 105
- Sensitivity, 6, 7, 103, 105–107
 - definition, 6
- Serum-testosterone levels, 228, 229
- Shipbuilding, 3, 4, 24, 199
 - industry, 4, 24, 199
- Simple linear regression model, 244, 254
- Simple regression analysis, 244
- Small samples, 144–145
 - example, 144–145
 - use, 144
- Specific rates, 13
- Specificity, 6, 7, 105–107
 - definition, 6
- Standard deviation (SD), 72, 118, 140, 142, 143, 148, 217
 - calculation, 72
 - formula, 72
- Standard error, xxiii, 140–144, 147, 183, 187, 251
 - example, 143
- Standardized rates, 13
- Standard normal curve, 111–114, 277
 - area under curve, 112, 114, 277
- Standard normal distribution, 144, 189, 192, 216, 222
- Standard normal score, 123, 124
- Standard normal variate, 112
- Standard population, 16, 17
 - choice, 17
- Standard toolbar, 28

- Statistic(s), xvi, xvii, xviii, xix, 2
 definition, xvii, 2
- Statistical analyses, 61, 77
- Statistical analyses packages, 26
 BMDP, 26
 SAS, 26
 SPSS, 26
- Statistical evidence, 174
- Statistical formulas, xviii, xxi, xxiii
 dealing with, xxii
- Statistical inference, 139
- Statistical procedures, xv
 estimation, xv
 testing, xv
- Statistical relationship, 74, 76, 103, 104,
 173, 244
 strength, 76
- Statistical significance, 172
- Statistical test(s), 135, 171, 172, 178, 216
- Statistics learning, xix–xx, xxii
 emotional reasons, xix
 intellectual reasons, xix
 time required, xxii
- Stepping stone, 20
- Sub-population(s), 178, 215
- Sum of squared deviations, 226, 246,
 252, 257
- Surgeon General’s Report, 150
- Survival time, 70
- Symmetric distribution, 61
 positively skewed, 61
- Systolic blood pressure (SBP), 77, 78, 119–121,
 135, 146, 156, 219, 248, 249, 260
 example, 78
 limits, 120
 means, 120, 121
 standard deviation, 120, 121
- Target population, 99, 100, 102, 106, 112,
 118, 135
 definition, 99
 numerical characteristic, 135
- t*-coefficient, 145, 147, 158, 251, 252
 value, 252
- t* curves, 128
 example, 129
 TDIST, 128
 TINV, 128
- t* distribution(s), 126, 278
 percentiles, 278
- Test statistic, 174, 179, 180, 181, 218, 227, 250,
 253
F statistic, 227, 253
- Total sum of squares (SST), 226, 227, 253, 258
- Toxic shock syndrome (TSS), 155
- t*-test, xxii, xxiii, 218, 250, 251, 259
- Two population proportions, 191
 comparison, 191
 example, 192–196
- Two-sample *t*-test, 222–224
- Two-sided test, 179, 180, 181, 187, 191
- Type I error, 172, 174, 180, 225
- Type II error, 172, 174–177, 179
 probability, 176
- Unbiased estimator, 138
- Uni-modal pattern, 59
- Univariate probability, *See* Marginal probability
- University of Minnesota Adolescent Health
 Program, 2
- U.S. nonwhite families, 61, 62
 income, 62
- U.S. population, 16–18
- Variable, 21, 30, 52, 53, 57, 101, 102, 108, 112,
 121, 125, 136, 142, 158, 196, 243–245,
 248–251, 254, 256
 binary, 196
 confounding, 146, 196
 dependent, 243, 244, 248, 249, 253, 263
 independent, 243, 244, 248–250, 253, 256,
 260, 263
 random, 136, 142, 158
 statistics, 136–137
- Variance, 72, 139, 140, 225, 228
 calculation, 72
 one-way analysis, 225–229
- Variance approach analysis, 252–254, 257–258
 example, 253–254
- Variance formula, 73
- Variation coefficient, 73
- Visual aids, 26, 79
 bar/pie charts, 28
 computer screen, 27
 cut and paste, 27
 descriptive statistics, 80
 forming 2×2 tables, 30
 formula bar, 27
 histograms, 79
 select and drag, 28

- Vital statistics, xvii, 13
- Volume oxygen (VO_2), 144, 145
- Weight curves, 67
- White-and-Nonwhite classification, 1
- Zero frequency, 59
- z -score, 118, 187–189, 191, 192, 196, 201, 202
 - Chi-squared, 192
- z standard deviations, 277
- z -test, 189
- z -value, 115–118
 - probability, 116, 117