Subject Index

1 in 4 youths abused, 16, 55, 91-94 abcnews.com, 54 accuracy, 12, 211-213 addiction to the World Wide Web, 54addresses, street, 48-52 Africa, number of countries in, 67, 268age adjustment, 13, 32 ages, see guessing ages aggregation, 44 AIDS, 160, 164 air bags, 211 American Statistician, see magazines amoebas (example of logarithms), 33 analysis of variance, 259, 261 anchoring, 67 animals, 35-37 arcsine rule, 245-246, 275 artifacts in maps, 13-14, 274 assignments, see homework average absolute error, 12 average student, 24 babies, see births barroom brawls, 156 baseball, 45, 109-111, 151, 270 length of World Series, 109–111 hypothesis test, 132–133 basketball, 135, 213, 253, 269, 271 Bayesian statistics, 215-221, 274 before/after study, see observational study Benford's law of first digits, 52–54, 132, 268 data, 53 hypothesis test, 132 worksheet, 52 Berkeley, California, 150 best of the best, 4 beta distribution, 215

betting, 213, 274 bias, 13, 25, 50, 54, 56–57, 74, 121– 122, 130, 268 selection, 154–156 distinguishing from other sorts of bias, 154-155 puzzles, 155-156 binomial distribution, 122-123, 130-132, 239 biology, 35-37, 43-44, 105-106 birthday problem, 238-239, 241 births boys or girls, 105–106, 122–123, 169 twins, 106 bivariate normal distribution, 30 blackboard, 19 blindness, see experiments body measurements, 43-44 Boston Globe, see newspapers Boston Herald, see newspapers boys, see births branching processes, 250 Brier score, 211-212, 274 Buffon's needle, 243, 253 Buffon's noodle, 186 buzz groups, 187 calibration of age guesses, 12 of point spreads, 213 of probability estimates, 211-212, 274of subjective intervals, 274 of uncertainty intervals, 216-219 cancer rates, 13-14, 265, 274 Bayes-estimated map of highest, 216Bayesian inference for, 215 map of highest, 14 map of lowest, 15 simulation, 219-220

candy weighing, 121-122 data, 122 photo, 170 card games, why we do not use as probability examples, 104 case studies, 255-264, 275 helicopter design, 259–264 quality control, 258-259 categorical variables, 20–21 causation, 44 Census adjustment, 235 Chance, see magazines Chance News, 188 chi-square test, see hypothesis testing Chicago Sun-Times, see newspapers China, sex ratio of births, 123 church attendance, as categorical or continuous, 21 class participation, 172, 265, 273 clinician's illusion, 155 cluster sampling, see sampling Clusterville, 226-228 worksheet, 227 coalitions, 111, 270 coffee tasting, 70–72 cognitive illusions, 268 coin flipping and spinning, 115–120 arcsine law, 245-246 data, 118 for random numbers, 224–225 physical model, 115–120 probability of 6 straight heads in 100 flips, 108 real and fake sequences, 106–108, 269simulations, 108 collecting data, see data collection College Teaching, see magazines community service, see survey conditional probability, 112-115, 270 lie detection, 114–115, 163 Monty Hall, 113, 270 three cards, 112-113

three prisoners, 113–114, 270 confidence intervals, 124-129, 225, 271 basketball shooting, 135 coverage, 127-129 graph, 128 for regression coefficients, 138 golf putting, 127, 144 graph, 144 guessing ages, 124–125 inflatable globe, 125–126 memory quizzes, 134 graph, 134 noncoverage, 129 graph, 129 poll differentials, 126 subjective, 205, 216-219 United Nations experiment, 125 Congressional elections, 111 Consumer Price Index, 31 contingency table, 27-28 continuous variables, 20–21 correlation, 43-45, 47 and aggregation, 44 and causation, 44-45 coupling (technique in probability theory), 251 course packet, 189 course plan, 189-199 coverage, see confidence intervals crime, 150, 151, 162, 268, 270, 273 cubes (example of logarithms), 33 Daily Californian, see newspapers data cleaning, 42 data collecting by students body measurements, 43–44 golf putting, 146 data collection, 11 basketball shooting, 135 biased coins and dice, 116-118 candy weighing, 121–122 class survey, 17-18, 223-224

> coffee tasting, 70–72 family size, 56–57 group projects, 178–185

294 REFERENCES

guessed exam scores, 25–27 guessing ages, 11 handedness. 17, 21 helicopter design, 259–264 inflatable globe, 125–126 memory quizzes, 133-134 soda consumption, 22 spaghetti, 185–186 telephone numbers, 48-52United Nations experiment, 65-68 data display, 24-28, 266, 272 data dredging, see multiple comparisons deception as ethical issue, 161 confidence intervals demonstration, 129 United Nations experiment, 65-68 decision analysis, 204–215, 274 decision tree, see probability tree depression, 96 descriptive statistics, 19-37 design of experiments, see experimental design details, why they are important, 169 dice for random numbers, 48-52, 104-105loaded, 115-120 dice games, why we do not use as probability examples, 104 discrete data, 28 displaying graphs on the blackboard, 19 distribution, 20-22 Benford's law, 53 beta, 215 binomial, 122-123, 130-132, 239 bivariate normal, 30 continuous with a spike at zero, 21, 23gamma, 220 geometric, 239

heights, 29 hypergeometric, 70, 132 length of World Series, 110, 133 negative binomial, 239 normal, 28–30, 105, 145, 205, 207 Poisson, 105, 220, 243–244, 253 divorce, 157–159 Doubs, France, 29, 30

earnings, see height and income economic indexes, 267 economics, 31-32, 39-41, 99-103, 138-141, 207-211, 213, 267, 274 Economist, see magazines education and public opinion, 27–28 eggs, number produced in 1965, 218 elections, 111, 126, 188, 270, 271 elephant, metabolic rate of, 35 estimating a big number, 14–15 ethics, 159-164 adjusting for exam scores, 162-163and randomized experiments, 160-161background information, 162 misleading information, 163 privacy and confidentiality, 163-164 racial discrimination, 161–162, 273randomized experiments, 159-160ethnicity, 161–162 Euler's formula, 242 examactual vs. guessed scores, 25–27 as teaching tool, 175 correcting for guessing, 162–163, 273correlation, 30, 46 number of pages written, 44–45 psychometric study, 142–143 randomizing, 68-69, 142 scatterplot, 29, 30

scoring probabilistic answers, 211-212Excel, why we don't let students use, 179exit poll, 55-56 expectation, see probability experimental design, 259-264 graphs, 263 experiments, 65-72 biased coin, 115-120blindness, 161–162 embedded in surveys, 68 ethical issues, 159-161 helicopter design, 259–264 IV fluids for trauma victims, 86-91nonrandomized, 86-91 prayer, 136-137 protocol, 159-160 surgeons and music, 17 taste test, 69–72 United Nations, 65–68 using exams, 68–69, 142–143 extrapolation, 20 fake data, 3 false positives, 114–115 family size, 56–57 data from students, 57 favorite activities, 4

feedback, 188, 273 fires, 158 first digits, see Benford's law first week of class, 11–18, 189 football, 213 forecasting, 212–213 friends, popularity of, 155

Gallup poll, 28, 126, 266 gambler's ruin, 123–124, 240 gambling, 213, 239, 274 lotteries, 108–109 roulette, 123–124 systems, demonstration that they don't work, 123–124

gamma distribution, 220 GDP, see Gross Domestic Product generating functions, 250, 253 geometric distribution, 239 geometry, 145 getting started, 167–171 girls, see births globe, inflatable, 125-126, 130, 271 golf putting data and fitted model, 144, 146 diagram, 145 example of confidence intervals, 127modeling, 143-146 government statistics, 232, 274–275 grade point average (GPA), 31 graphics, 24-28, 266, 272 Gross Domestic Product, 31 group activities, 11, 32, 48-52, 57-65, 106-107, 114, 127-129, 135, 148 - 154, 167, 173 - 176,205-207, 219-220, 223-224, 228, 237-238, 256-257 photos, 169, 170, 173, 174 guessing ages, 11-13, 124-125, 265 confidence intervals, 124–125 data, 13 hypothesis tests, 130 photo, 174 worksheet, 12 guessing exam scores, 25–27 Gulf War, 149 hand span and height, 43–44 handedness, 21, 265, 266 as categorical or continuous, 21 association with age of death, 156, 272 data from students, 22 guessed distribution, 22 worksheet, 17 heart bypass surgery, blindness in, 161 - 162height and hand span, 43-44

296 REFERENCES

and income, 39-41, 138-141 data, 41 Stata log file, 42, 139, 141 conscripts, apparently bimodal, 30men and women, not bimodal, 29.266 men's exaggeration of, 130 parents and children, 46 helicopter design, 259–264, 276 data, 261 instructions, 260 hierarchical models, 220-221, 259 histogram. 20–22 candy weighing, 122 family size, 57 first digits, 53 handedness, 22 heights, 30 soda consumption, 23 student guess, 22 United Nations experiment, 68 Holocaust, 55 homework, 1, 78-79, 157, 176, 243-244hot hand in basketball, 253 how to do it, 237-238 how to use this book, v, 1-5, 167-175Hungarian maps, 251–252 hypergeometric distribution, 70, 132 hypothesis testing, 129–133 chi-square test baseball World Series, 132–133 Benford's law, 132 golf putting, 145–146 guessing ages, 130 heights of men, 130 inflatable globe, 130 random exams experiment, 69 taste testing, 132 telephone book sampling, 131– 132United Nations experiment, 130-131

illegal aliens, 99–103 income, see height and income India, sex ratio of births, 123 ineligible sampling units, 54 infant deaths, 16 inflatable globe, see globe inflation, 31, 213, 267, 274 instructions coffee tasting project, 71 helicopter design, 260 independent projects, 179, 180 probability projects arcsine law, 245-246 branching processes, 250 Markov chains, 247–251 martingales, 249-250 random permutations, 251–252 sampling and inference, 232-233 sampling project, 59-62 statistical literacy assignment, 80.81 Internet survey, 54 interpolation, 20, 34 IQ, 135–136, 219, 271 IV fluids for trauma victims, 86–91 jar of quarters, 204-207 Journal of Educational and Behavioral Statistics, see magazines Journal of Statistics Education, see magazines Jurimetrics, see magazines keys and boxes, 241, 251–252 Korea, sex ratio of births, 123 Las Vegas, 213 law of large numbers, 206, 245 least squares, see linear regression left-handedness, see handedness lie detection, see conditional probability life, see value linear combinations, 31–32

linear regression, 25, 38–47, see also multiple regression income and height, 39-41, 138-141log metabolic rate, 35 multiple, see multiple regression R-squared, 42, 139-141, 147 simple examples, 38–39 worksheet. 39 linear transformation, 31–32 mean and standard deviation, 105, 135-136 log-linear plot, 34 log-log plot, 35 logarithmic regression, 41 logarithmic transformation, 32-37, 53 income, 141 metabolic rate, 35-37 simple examples, 33 world population, 33-34 logistic regression, 142 London Observer, see newspapers lotteries, 108-109 lottery, 269 low R-squared, why it can be good, 140lurking variables, 41, 44-45, 72-74 lying with statistics, see statistical communication magazines American Statistician, 187 Chance, 187, 188 College Teaching, 187 Economist, 150 Journal of Educational and Behavioral Statistics, 187 Journal of Statistics Education, 187, 188 Jurimetrics, 235 Mathematics Teacher, 187 New England Journal of Medicine, 86-87, 269 New Scientist, 118

Pediatrics, 91-92, 95-96, 269

Salon, 136-137 Science News, 95, 269 Sports Illustrated, 55 Teaching Statistics, 187 man, metabolic rate of, 35, 37 maps, see also cancer rates artifacts in, 13-14, 150, 151, 216, 274Hungarian, 251-252 Markov chains, 247-251, 253 marriage, 157-159 martingale betting system, 123 martingales, 249-250 math majors, see survey Mathematics Teacher, see magazines McCarthy, Joseph, 148 mean, 22-24 after linear transformation, 105 weighted, 23, 234 measurement error, see bias and variation median, 22-24weighted, 234 memory guizzes, 133–134, 271 data, 46, 134 example of regression to the mean, 45 - 46hierarchical modeling, 220-221 metabolic rate, 35-37, 267 graph, 35 mile run, 20, 265 mistakes, see warnings model-based analysis, 99-103 money, see value monster in the crib, 94–99 Monty Hall, see conditional probability multiattribute value functions, 214 multilevel models, 220-221, 259 multiple comparisons, 135-137, 150-152, 160 multiple regression, 138–147 exam scores, 142 helicopter design, 263

income on height and sex, 138-141interactions. 140–141 interpretation, 138-140 Pythagoras, 146–147 residuals, 147 multitasking, 167 names, frequency of, 134-135, 271 National Rifle Association, 55 negative binomial distribution, 239 New England Journal of Medicine, see magazines New Scientist, see magazines New York Restaurant Association, 84 New York Times, see newspapers newspaper articles, 15-17, 54-55, 73-74, 77-103, 148-154 finding sources, 83–84 newspapers Boston Globe, 154, 272 Boston Herald, 152, 272 Chicago Sun-Times, 158, 272 Daily Californian, 151, 272 London Observer, 112 New York Times, 3, 86, 159, 229, 269, 274 San Francisco Chronicle, 3, 16, 65, 73, 236, 265 San Francisco Examiner, 3, 16, 91, 95, 100, 149, 153, 265, 269, 272 San Francisco Weekly, 153, 272 nonlinear regression golf putting, 143-146 nonresponse, 54, 56 normal distribution, 28-30, 105, 145, 205, 207 bivariate, 30 number theory, 242 numeracy, 14-15, 265 numerical variables, 20-22 Oakland, California, 15

observational data, 182-183 observational study, 44-45, 72-74 before/after comparisons, 74, 76 chlorinated tap water, 73–74 infant mortality, 17 monster in the crib, 94-99 SAT coaching, 74 smoking and cancer, 73, 75 optimization, 207, 262 order-of-magnitude estimation, 14-15outliers, 42 overconfidence, 25, 27, 216-219, 274 of engineering judgments, 217 of subjective probability intervals, 218 Palm Beach, 188 palm lines and age, 155 Patriot missile, 148, 149 Pediatrics, see magazines permutations, random, 239, 241, 251-252phone book, see telephone book physics, 119–120 pickle jar, 119 planning, 167-168 point spreads, 213 Poisson distribution, 105, 220, 243-244, 253 political science, 111, 126, 188, 270 poll differentials, 126 Polya's urn, 252 population, see world population power, see statistical power practice, 168 prayer, 136-137, 271 pre-test, 74, 162 prediction, 26 pregnant women, 73-74 preparation, 167-168, 189-190 Presley, Elvis, 106 probability, 104-120, 237-253 compound events, 105-109, 239-240

conditional, see conditional probability constructing examples, 104 expectation, 206-207, 240-241 modeling golf putting, 143–146 length of World Series, 109-111 Poisson distribution, 243–244 space shuttle failure, 111–112 voting and coalitions, 111 problems challenging, 241-242 introductory. 238–241 structured projects, 244-249 unstructured projects, 249-252 projects branching processes, 250 Markov chains, 250–251 martingales, 249-250 random permutations, 251–252 research papers, 252–253 rare events, 108-109, 111-112, 270overestimation by surveys, 55 subjective, 205, 216-219 worksheet, 218 tree assessing utility of money, 208 lie detector problem, 115 three cards problem, 113 projects, 57-65, 70-72, 175-185, 254-264, see also probability projects coffee tasting, 70-72ideas for topics, 181-183 literacy. 80-83 sampling, 230-235 propagation of uncertainty, 15 props, 172, 238-239 protocol, 179, 180, 273 psychiatry, 155 psychometrics, see exam public opinion, 126, 164, 271 about Vietnam War, 27-28 as categorical or continuous, 21

exit poll, 55-56 punchline, 170–171 puzzles estimating risk, 212 exit poll, 56 fires, 158 guessed exam scores, 26-27 heights of conscripts, 29 mile run. 20 order-of-magnitude estimation, 14 - 15Pythagoras, 147 real and fake coin flips, 107 selection bias, 155–156 subjective probability intervals, 218telephone book sampling, 51 three cards, 112–113 Vietnam War opinions, 28 Pythagoras goes linear, 146–147 quality control, 258-259, 275 quarters, jar of, 204-207 questionnaire math majors survey, 66 video game survey, 64 quotations from students, see students' experiences R, see S-Plus R-squared, see linear regression race, 161-162 raking, see sampling Rand Corporation, 65 random digits, 104-105

random digits, 104–105 random numbers, 104–105, 224–225 randomization, see experiments randomness, the look of, 106–108 rare events, see probability rat, metabolic rate of, 35 reflection principle, 245 regression, see linear regression regression to the mean, 26, 30, 45–47 Reno, Janet, 46

residual, 34, 40

residual plot, 147 response surface, 261-264 graph, 262 right-handedness, see handedness risk, 210-211, 274 aversion, 209-211 worksheet, 212 role playing, 106, 114, 137, 219-220 Roper poll, 55 roulette, 123–124 how to do the demonstration badly, 169rounding, 29, 185, 266, 267 S-Plus, 147, 265 Salon, see magazines sample size, see statistical power sampled population, 54 sampling, 17, 48-57, 91-94, 121-122, 222–235, see also survey bias, 50 class projects, 57-65 cluster, 226-228 data-collection grid, 184 distribution, 122, 226 exit poll, 55–56 frame, 54, 55, 230 network, 155 protocols, 48, 49, 184 raking adjustment, 233 simple examples of bias, 56 template, 54 triptych template, 226 U.S. government surveys, 231 unequal probabilities, 50, 55 wacky surveys, 54–55 weighting, 233-234 San Francisco Chronicle, see newspapers San Francisco Examiner, see newspapers San Francisco Hotel Association, 83 San Francisco Weekly, see newspapers SAT, see Scholastic Assessment Test

scaling, 35-37 scatterplot, 24-28 and regression line, 39, 40 body measurements, 44 coin flips, 108 exam scores, 25-27, 29, 30, 143 income, 41 memory quizzes, 46 metabolic rate, 35 use of different symbols, 24–27, 44 schedule, 189-199 Scholastic Assessment Test, 31, 74, 269school buses, 14-15 school reforms, Texas, 3-4, 73, 265 Science News, see magazines selection bias, see bias self-experimentation, 273 sex, 13, 29, 37, 41, 138-141 sex ratio, see births shrinkage, 220-221 siblings, see family size Simpson's paradox, 44 Smiths, 15, 216 smoking, 73 soccer, 112 soda consumption, 22 data from national survey, 23 data from students, 23 mean, 22soda tasting, 70 soft drinks, see soda sources, v, 186-188, 265-276 space shuttle, 111–112 spaghetti, 185-186, 243 sports baseball, 109-111, 132-133, 270 basketball, 135, 269, 271 golf, 127, 143-146, 271 mile run, 20, 265 soccer. 112 Sports Illustrated, see magazines squares (example of logarithms), 33 standard deviation, 31

after linear transformation, 105 Stata, 265 log files, 42, 139, 141 StatCity, 232-234, 275 Statistical Abstract of the United States, 24, 266, 273 statistical communication, 148-164, 272connection to rest of course, 156-157homework assignments, 157 ignoring the baseline, 150 meaningless numbers, 148 misinformation, 148–150 misleading comparisons, 152–154 outright lies, 148 selection bias, 154-156 statistical literacy, 77-103, 228 list of packets, 79 packets 1 in 4 youths abused, 91-94illegal aliens, 99–103 IV fluids for trauma victims, 86-91 monster in the crib, 94–99 statistical power basketball shooting, 135 births of boys and girls, 123 statistical significance, see hypothesis testing statistics, why it is important, 3–4 stereology, 275 Stirling's formula, 245 street addresses, see addresses students' experiences, 59-61, 72, 90-91, 94, 98-99, 102-103, 183-185, 188, 259-261, 263-264, 273students, early deaths of, 155 subjective probability, see probability Surgeon General's report, 73 surgeons and music, 16 surprise, 171-172 survey, see also sampling

1 in 4 youths abused, 55, 91-94addiction to Web, 54-55child abuse, 17 classroom, 17-18, 223, 224 community service, 65 do you love statistics, 56 exit poll, 55–56 Holocaust, 55 math majors, 65 monster in the crib, 94–99 telephone, 57, 164 U.S. government, 231 U.S. kids reading well, 148 video games, 63 wacky, 54-55 working women, 229 table, two-way, 27-28 target population, 54 taste tests, 69-72, 132 Teaching Statistics, see magazines telephone book sampling, 48–52 complexities, 50-51 data, 51 example, 50 hypothesis test, 131–132 photo, 173 worksheet, 49 telephone numbers seven digits in, 271 unlisted, 50 test scores in Boston schools, 152-154Texas school reforms, 3–4, 73, 265 three cards, see conditional probability three prisoners, see conditional probability time series, 19-20, 33-34 mile run, 20 world population, 34 Toxics Release Inventory, 4, 265 transformations, see linear and logarithmic transition probabilities, 247

302 REFERENCES

transparency, 19 trigonometry golf putting, 144–145 sampling from the globe, 126 triptych template, 226 twins, see births two-way table, 27-28 Vietnam War, 28 United Nations experiment, 65–68, 73, 125, 268 confidence interval, 125 data, 68 hypothesis test, 130–131 worksheet, 67 utility, see value value functions life, 210-211 money, 207-210 multiattribute, 214 variables, categorical and continuous, 20 - 21variation, 11, 13, 22, 23, 122, 261 video games, see survey Vietnam War, 27-28, 150 data, 266 worksheet, 28 voting, 111, 126, 188, 270, 271 wacky surveys, 54-55 warnings, 20, 136-137, 146-148, 154, 159, 168-171, 181, 228 watts, 35 weighing candies, see candy weighted average, 23, 27-28, 215, 234 weighted median, 234 weights candies, 121-122 students, 127 What's in the news?, 15–17 woman, metabolic rate of, 37 working women survey, 229 worksheet Benford's law, 52

cluster sampling, 227 estimating risk, 212 guessing ages, 12 handedness, 17 linear regression, 39 sampling, 54, 226 statistical literacy, 230 subjective probability intervals, 218telephone book sampling, 49 two-way table, 28 United Nations experiment, 67 world population, 33-34, 41-43 data, 34 graphs, 34 log regression, 41-43world records, 20 World Series, see baseball written reports, 61-63, 82, 176-177, 185, 257