
Author Index

Note: with references having more than three authors, the names of all authors appear in this index, but only the first authors appear in the in-text citations.

- | | |
|---------------------------|---|
| Aaltonen, M., 310 | Best, N., 180 |
| Aaronson, D., 449 | Betancourt, M., 76, 128, 180 |
| Abadie, A., 416 | Bingel, U., 356 |
| Abayomi, K., 333 | Bishop, C., 471 |
| Achen, C., 355 | Blalock, H., 211 |
| Adar, E., 33 | Blattner, F., 286 |
| Agodini, R., 417 | Bloom, H., 417 |
| Agresti, A., 67, 259, 286 | Bogart, W., 449 |
| Ainsley, A., 356 | Bogatz, G., 379, 449 |
| Akaike, H., 180 | Boscardin, W., 47 |
| Albert, A., 259 | Boshuizen, H., 333 |
| Aldy, J., 381 | Bosker, R., 471 |
| Ali, I., 492 | Box, G., 18, 210, 310, 471 |
| Almond, D., 449 | Bradley, R., 286 |
| Amemiya, T., 286 | Bradlow, E., 417 |
| Andersen, M., 180 | Brainard, J., 47 |
| Anderson, D., 286 | Brilleman, S., 492 |
| Anderson, J., 259 | Brillinger, D., 471 |
| Angrist, J., 416, 449 | Bring, J., 210 |
| Anoke, S., 90 | Brookmeyer, R., 379 |
| Ansolabehere, S., 382 | Brooks-Gunn, J., 150, 416 |
| Arce, A., 47 | Browner, W., 67 |
| Aronow, P., 356 | Bryk, A., 471 |
| Ashenfelter, O., 449 | Buenconsejo, J., 417 |
| Asimov, D., 33 | Buerkner, P., 180, 471, 492 |
| Assmann, S., 310 | Buja, A., 33 |
| Athey, S., 416 | Burmaster, D., 47 |
| Atkinson, A., 210 | Bush, R., 76 |
| Auerbach, J., 34 | Button, K., 67, 310 |
| Bafumi, J., 47 | Campbell, D., 355, 449 |
| Baird, A., 67 | Campbell, J., 454 |
| Balke, A., 449 | Campbell, S., 355 |
| Ball, S., 379, 449 | Card, D., 449 |
| Bannerjee, S., 471 | Carlin, J., 111 |
| Barnard, J., 333, 449 | Carlin, B., 180, 471 |
| Becker, R., 492 | Carlin, J., 47, 67, 76, 128, 150, 180, 310, 471 |
| Beecher, H., 356 | Carnegie, N., 416, 417 |
| Belin, T., 333 | Carpenter, B., 471 |
| Bennett, C., 67 | Carroll, R., 210, 449, 471 |
| Berglund, P., 310 | Carvalho, C., 211, 416, 471 |
| Berk, R., 180 | Case, A., 31, 34 |
| Berry D., 471 | Casella, A., 286 |
| Bertin, J., 33 | Cervone, D., 416 |
| Besag, J., 356 | Chambers, J., 33, 492 |

- Chan, V., 379
 Chang, M., 286
 Chang, S., 17
 Chapman, R., 286
 Chay, K., 449
 Chay, K. Y., 449
 Chipman, H., 471
 Clark, M., 417
 Clayton, D., 333
 Cleveland, W., 33, 471
 Clinton, J., 288
 Clogg, C., 333
 Cochran, W., 47, 310, 416
 Colloca, L., 356
 Cook, D., 33
 Cook, R., 259
 Cook, S., 76
 Cook, T., 355
 Cortez, P., 211
 Cortina, J., 237
 Costa, D., 381
 Coull, B., 67
 Couper, M., 310
 Cox, D., 210, 379, 471
 Cox, G., 382
 Crainiceanu, C., 449, 471
 Cramer, C., 237
 Cressie, N., 471
 Criqui, M., 211
 Croissant, Y., 449
 Cromwell, B., 449
 Crump, R., 416

 Dallal, G., 90
 David, M., 333
 Dawid, A., 379, 450
 de Groot, A., 67
 De Leeuw, J., 471
 Deaton, A., 31, 34
 Dehejia, R., 416, 417
 Diakow, R., 417
 Diamond, A., 417
 Diffendal, G., 333
 Diggle, P., 417
 Ding, P., 417
 Dobson, A., 286
 Dodhia, R., 34
 Dorie, V., 416, 417
 Dorman, P., 381
 Drake, C., 417
 Dreze, J., 449
 Duncan, D., 286
 Dunn, G., 333
 Dunson, D., 47, 76, 111, 128, 180, 471
 Dyke, G., 356
 Dynarski, M., 417

 Eddy, W., 180
 Edlin, A., 47
 Efron, B., 76, 90
 Ehrenberg, A., 34
 El-Bassel, N., 286
 Endicott, J., 417
 Enos, L., 310
 Erikson, R., 101, 288, 382
 Evans, W., 449

 Fagan, J., 17, 286
 Fair, R., 101
 Fay, R., 333
 Felton, J., 151
 Fienberg, S., 286, 450
 Finkelstein, M., 333
 Firth, D., 259
 Flaten, M., 356
 Flint, J., 310
 Forbes, A., 150
 Fortna, V., 7, 17
 Fowler, F., 310
 Fox, J., 33, 101, 150, 180
 Francis, G., 67
 Frangakis, C., 379, 449
 Franke, R., 356
 Friedlander, D., 417
 Friedman, J., 286, 471
 Friedman, M., 286
 Frolich, M., 417
 Fuller, W., 471
 Fung, K., 450

 Gabry, J., 180, 492
 Gal, D., 67
 Galea, S., 17
 Galton, F., 85
 Garfinkel, I., 333
 Gebler, N., 379
 Gelfand, A., 471
 Gelman, A., 17, 34, 47, 67, 76, 111, 128, 180, 211,
 237, 259, 286, 310, 333, 356, 379, 449,
 471, 492
 George, E., 471
 Gerber, A., 454
 Gertler, P., 17
 Geweke, J., 76
 Gilbert, L., 286
 Glickman, M., 449
 Goegebeur, Y., 259
 Goel, S., 67, 333
 Goldin, C., 379
 Goodrich, B., 180, 333, 492
 Graddy, K., 449
 Gramacy, R., 492
 Grantham-McGregor, S., 17
 Greenland, S., 111, 211, 417

AUTHOR INDEX

519

- Greevy, R., 356
 Grolemond, G., 34, 492
 Groves, R., 310
 Guo, J., 492
- Haavelmo, T., 379, 449
 Hagstrom, P., 381
 Hahn, J., 449
 Hahn, R., 416
 Hahn, T., 416
 Halloran, E., 356
 Hamermesh, D., 151
 Han, W., 150
 Handcock, M., 471
 Harada, M., 417
 Harrell, F., 101, 180, 211
 Harrison, J., 471
 Hasbun, R., 417
 Hastie, T., 286, 471
 Hauer, E., 286
 Healy, K., 33
 Heckman, J., 17, 379, 416, 417
 Heeringa, S., 310
 Heinze, G., 259
 Heisey, D., 310
 Heitjan, D., 78, 333
 Hernan, M., 355, 356
 Hibbs, D., 93, 101
 Higdon, D., 356
 Hill, C., 417
 Hill, J., 150, 286, 333, 356, 416, 417, 449, 471
 Hirano, K., 417, 449
 Hoenig, J., 310
 Hoerl, A., 471
 Hoeting, J., 471
 Hoff, P., 471
 Hofmann, H., 33
 Hong, G., 356
 Hoogerheide, L., 449
 Hotz, V., 416
 Huang, Z., 382
 Hudgens, M., 356
 Hullman, J., 33
 Hunter, C., 47
 Hunter, J., 18, 310
 Hunter, W., 18, 310
 Hurley, C., 33
 Hwang, J., 180
- Ichimura, H., 416
 Imai, K., 286, 417
 Imbens, G., 379, 416, 449
 Ioannidis, J., 310
- Jackman, S., 288
 Jacob, B., 449
 Jakulin, A., 259
- Jenkins, G., 471
 Jenkyn, J., 356
 Joffe, M., 417
 Jonasson, J., 180
 Jones, S., 356
- Kahn, M., 381
 Kahneman, D., 89, 90
 Kalesan, B., 17
 Karagas, M., 449
 Kasten, L., 310
 Katz, J., 47, 382
 Kaul, J., 356
 Kay, M., 492
 Kazarow, K., 379
 Keele, L., 417
 Keiser, O., 17
 Keller, M., 417
 Kendziorski, C., 286
 Kennard, R., 471
 Kennedy, P., 355
 Kenward, M., 417
 King, G., 47, 76, 382
 Kiss, A., 286, 333
 Klebanov, P., 416
 Kleibergen, F., 449
 Kleiner, B., 33
 Knook, D., 333
 Krantz, D., 17, 67, 381
 Krantz-Kent, R., 67
 Kreft, I., 471
 Kropko, J., 333
 Krueger, A., 449
 Kutner, M., 111, 149
- LaLonde, R., 417
 Landwehr, J., 259
 Lange, K., 286
 Lavori, P., 417
 Lawrence, M., 33
 Lax, J., 333
 Leamer, E., 379
 Lechner, M., 417
 Lee, A., 85, 90
 Lee, D., 449, 492
 Lee, E., 33
 Lefgren, L., 449
 Lenth, R., 310
 Lepkowski, J., 310
 Lessafre, E., 259
 Levin, B., 333
 Levine, D., 449
 Levitt, S., 356, 382
 Levy, M., 333
 Li, F., 416
 Liang, K., 286
 Liaw, F., 416

520

Liebman, J., 286, 333
 Lin, C., 17, 381
 Linero, A., 471
 List, J., 356
 Little, R., 286, 333
 Little, T., 333
 Liu, C., 286, 333
 Lohr, S., 310
 Loken, E., 67
 Lord, F., 90
 Lovell, J., 286
 Lu, B., 356
 Lu, H., 237
 Lumley, T., 492

Ma, G., 333
 Mack, S., 333
 Maclure, M., 211
 Maddala, G., 286
 Madigan, D., 471
 Magnusson, M., 180
 Mahboobeh, S., 379
 Mahr, T., 492
 Mallows, C., 180
 Manski, C., 379, 449
 Mason, G., 310
 McCullagh, P., 237, 286
 McCulloch, R., 471, 492
 McDonald, G., 212
 McDonald, J., 33
 McElreath, R., 128, 150
 McEwan, P. J., 449
 McFadden, D., 286
 McGonagle, K., 379
 McKeague, I., 333
 McShane, B., 67
 Meehl, P., 67
 Meissner, K., 356
 Meng, X., 76, 180, 333
 Meyers, M., 333
 Michalopoulos, C., 417
 Middleton, J., 417
 Miller, M., 67
 Millo, G., 449
 Mitchell, J., 151
 Mitnik, O., 416
 Mobily, M., 17
 Mokrysz, C., 310
 Montgomery, D., 310
 Montgomery, J., 379
 Morgan, S., 379, 449
 Morgenstern, H., 211
 Morris, C., 356
 Moskowitz, A., 78
 Mosteller, F., 33, 76, 210, 286
 Motyl, M., 67

AUTHOR INDEX

Mueller, K., 492
 Mulligan, C., 47
 Munafa, M., 310
 Murphy, R., 449
 Murray, J., 416, 471
 Murrell, P., 33
 Musio, M., 450

Nachtsheim, C., 111, 149
 Nallamotheu, B., 47
 Nelder, J., 237, 286
 Nelson, L., 67
 Neter, J., 111, 149
 Newman, T., 67
 Newton, M., 286
 Neyman, J., 379
 Ng, J., 286
 Nolan, D., 18, 47, 181, 211, 471
 Normand, S., 449
 Nosek, B., 67, 310
 Nyhan, B., 379

O’Keefe, S., 417
 Oakes, D., 471
 Ojanen, J., 180
 Open Science Collaboration, 67
 Orben, A., 67
 Oudshoorn, C., 333

Pagano, M., 90
 Palfrey, T., 286
 Pallesen, J., 379
 Palmer, S., 417
 Pardoe, I., 259, 288
 Park, D., 237
 Parker, A., 151
 Pasarica, C., 34
 Pearl, J., 449
 Pearson, K., 85, 90
 Persico, N., 211
 Petrosino, A., 310
 Phillips, J., 333
 Pickles, A., 333
 Piironen, J., 180, 211, 471
 Pinto, R., 17
 Pischke, W., 355
 Pittau, M., 259
 Plurphanswat, N., 34
 Pocock, S., 310
 Pole, A., 471
 Polson, N., 211, 471
 Poole, C., 211
 Postlewaite, A., 211
 Pratola, R., 492
 Pregibon, D., 259
 Price, P., 17, 381
 Przybylski, A., 67

AUTHOR INDEX

521

- Quagliarello, V., 417
 R Core Team, 111, 492
 Raftery, A., 471
 Raghunathan, T., 333, 379
 Ramsey, F., 47, 149
 Rasmussen, C., 471
 Raudenbush, S., 356, 471
 Reiter, J., 417
 Resnick, P., 33
 Richmond, C., 286
 Ridder, G., 417
 Ripley, B., 76, 492
 Ritov, Y., 417
 Rivers, D., 101, 288, 333
 Robert, C., 67
 Robins, J., 333, 355, 379, 417, 449
 Robins, P., 417
 Robinson, E., 310
 Rodu, B., 34
 Romero, D., 288
 Rosenbaum, P., 356, 379, 417
 Rosenstone, S., 101
 Ross, C., 90, 492
 Rothschild, D., 67, 333
 Rotnitzky, A., 417
 Rouse, C., 379
 Roy, A., 379
 Rubin, D., 47, 76, 111, 128, 180, 333, 356, 379,
 416, 449, 471
 Ruppert, D., 210, 449, 471
 Samii, C., 356
 Samuhel, M., 333
 Sarsons, H., 449
 Schafer, D., 47, 149
 Schafer, J., 333
 Schemper, M., 259
 Schenker, N., 333
 Schlessman, J., 211
 Schmid, C., 34
 Schmidt-Nielsen, K., 39, 47
 Schultz, B., 333
 Schwing, R., 212
 Scott, J., 211, 471
 Scott, M., 416, 417
 Sekhon, J., 417
 Shadish, W., 355, 417
 Shalit, U., 416
 Shirani-Mehr, H., 67
 Shirley, K., 67
 Shoemaker, A., 259
 Shor, B., 237
 Silber, J., 356
 Silva, A., 211
 Silver, N., 47
 Silverman, D., 211
 Simmons, J., 67
 Simonsohn, U., 67
 Simonton, D., 288
 Simpson, D., 76, 128, 180, 471
 Singer, E., 310, 379
 Skardhamar, T., 310
 Small, D., 356
 Smith, J., 417
 Snedecor, G., 47
 Snijders, T., 471
 Snyder, J., 382
 Sobel, M., 356, 450
 Solenberger, P., 333
 Spanbauer, R., 492
 Sparapani, R., 492
 Spiegelhalter, D., 180, 333
 Spitzer, E., 286
 Stan Development Team, 111, 492
 Stanley, J., 355
 Stefanski, L., 471
 Steiner, P., 417
 Steinglass, P., 286
 Stern, H., 47, 76, 111, 128, 180, 333, 471
 Stevens, M., 379
 Stewart, C., 382
 Stigler, S., 90, 149, 164, 286
 Stinson, M., 151
 Stone, M., 180
 Strathdee, S., 379
 Stuart, E., 417
 Su, Y., 259, 333, 416, 492
 Swartz, T., 47
 Swayne, D., 33
 Tackett, J., 67
 Talts, S., 76
 Taylor, J., 286
 Terry, M., 286
 Thistlethwaite, D., 449
 Thomas, L., 416
 Thomas, N., 356, 416
 Thurstone, L., 286
 Tibshirani, R., 76, 211, 286, 471
 Tinbergen, J., 449
 Tobin, J., 286
 Todd, P., 416, 417, 449
 Torres, M., 379
 Tosteson, T., 449
 Tourangeau, R., 310
 Traskin, M., 356
 Trevisani, M., 237
 Triest, R., 333
 Troxel, A., 333
 Tsai, W., 333
 Tsui, K., 286
 Tuerlinckx, F., 67, 259

522

Tufte, E., 33
 Tukey, J., 33, 34, 76, 210, 286
 Tukey, P., 33
 Tversky, A., 89, 90

Unwin, A., 33, 34
 Urbanek, S., 34
 Urquiola, M., 449

Van Buuren, S., 333
 van der Klaauw, W., 449
 van der Linde, A., 180
 van Dijk, H., 449
 van Dyk, D., 286, 417
 van Geen, A., 237
 Van Hoewyk, J., 333, 379
 Van Mechelen, I., 259
 VanderWeele, T., 355, 356, 417
 Varadhan, R., 379
 Vasilatos, N., 101
 Vehovar, V., 333
 Vehtari, A., 47, 76, 111, 128, 180, 211, 471, 492
 Venables, W., 492
 Vermeerch, C., 17
 Versteeg, R., 237
 Vikram, H., 417
 Viscusi, W., 381
 Vlahov, D., 379
 Volinsky, C., 34, 471
 von Mises, R., 68

Wachsmuth, A., 90
 Wager, S., 416
 Wager, T., 356
 Wahba, G., 471
 Wahba, S., 417
 Wainer, H., 33, 417
 Waldfogel, J., 150, 416
 Walker, S., 17, 286
 Wallis, W., 286
 Wang, N., 333
 Wang, W., 333

AUTHOR INDEX

Wasserman, W., 111, 149
 Watson, A., 356
 Wattenberg, L., 28, 34
 Weakliem, D., 128, 310
 Wedderburn, R., 286
 Weisburd, D., 310
 West, B., 310
 West, M., 471
 West, V., 286, 333
 Whang, W., 78
 Wickham, H., 33, 34, 492
 Wideman, L., 333
 Wiens, B., 286
 Wilkinson, L., 34, 90
 Wilks, A., 492
 Williams, C., 471
 Winkler, S., 34
 Winship, C., 379, 449
 Witte, S. S., 286
 Wlezien, C., 101
 Wolford, G., 67
 Wolfram, C., 382
 Woolridge, J., 149, 286, 449
 Wu, E., 286

Xie, Y., 492

Yajima, M., 333
 Yao, Y., 180, 492
 Yates, F., 310

Zanolini, A., 17
 Zanutto, E., 417
 Zaslavsky, A., 333
 Zelizer, A., 449
 Zellner, A., 286
 Zhang, K., 356
 Zhao, L., 417
 Zheng, Y., 237
 Zhou, A., 449
 Zimmerman, P., 449
 Zorn, C., 259

Subject Index

- 1.6, converting from logit to probit, 227, 272, 273, 286
 10 quick tips, 493–496
 2 standard deviations, scaling by, 186–187
 2.8 standard errors for power analysis, 295
 3 challenges of statistical inference, 3
 4, divide by for interpreting logistic regression coefficients, 220
 55,000 residents need your help, 63–65
- abline(), 4, 156
 Academy Awards example, 288
 additivity assumption, 154
 age adjustment, 31–33
 age as regression input, 196, 197
 age-period-cohort example, 32
 Akaike information criterion (AIC), 175, 180
 always-taker, 422, 423, 439, 449
 Alzheimer's disease, 23
 analytic calculations and simulation, 144
 arm (R package), 475
 arsenic example, 232–238, 242–249, 279, 283, 379
 as.matrix(), 113
 assignment variable in regression discontinuity, 433, 434
 assumptions, 14, 153–156
 additivity, 154
 causal inference, 155, 350–355
 equal variance of errors, 154
 failures, 155
 independence of errors, 154
 linearity, 154
 normality of errors, 154
 of regression analysis, ranked in importance, 153–155
 reliability, 23–25
 representativeness, 25, 154
 validity, 23–25, 153
 autoregression, 166–168
 available-case analysis, 324
 average predictive comparison, 249–252, 261
 average treatment effect (ATE), 342–344, 348, 358–361, 372, 397–399, 418, 420
- balance, *see* causal inference, imbalance
 Bangladesh, *see* arsenic example
 baseline factor level, 138
- basketball, 44, 47
 bayes_R2(), 171, 491
 bayes_sim(), 229
 Bayesian additive regression trees (BART), 466, 471
 Bayesian inference, 15–16, 106–107, 109, 113–129, 228–230
 bayesplot (R package), 475, 492
 beauty and sex ratio, *see* sex ratio example
 beauty and teaching evaluations example, 151, 213
 Beck Depression Inventory, 24
 before-after studies, 27, 341, 357, 367–370
 beta-binomial regression, 272
 bias, 55, 346–347, 385
 binary data, 217
 binned averages, 241
 binned residuals, 253–255, 259
 graph, 253
 binomial distribution, 43, 69
 compared to Poisson, 268
 naive application of, 46–47
 binomial regression, 287
 births, 68–71
 black box, 376
 bootstrapping, 73–76
 limitations, 75
 with structured data, 75
 brm(), 272, 278, 284, 285, 430
 brms (R package), 264, 430, 492
 building generalized linear models, 232–237, 242–247, 283–285
 building linear regression models, 199–206
 bypass example, 452
- caliper matching, 409
 categorical variables
 as regression inputs, 196
 ordered or unordered, 273
 causal inference, 5, 10–11, 134, 155, 246, 496
 adjusting for post-treatment variable, don't do it, 369, 442, 445
 assumptions, 350–355, 380, 386
 average treatment effect (ATE), 342–344, 348, 358–361, 372, 397–399, 418, 420
 balance, *see* imbalance
 basics, 339–342
 causes of effects, 415, 446

- challenges of interpreting regression
 - coefficients, 84–85, 373–374
- complier average causal effect (CACE), 422
- conditional average treatment effect (CATE), 343
- conditional independence, 389
- confounding covariate, 317, 380, 384–386, 392, 393, 395–396, 438
- constructed observational study, 405, 417, 418
- continuous treatment, 342
- counterfactual, 394, 399, 414, 423, *see also* potential outcomes
- difference-in-differences, 442–445, 449
- direct and indirect effects, 377–379
- effect of the treatment on the controls and treated, 398
- effects of causes, 415, 446
- exclusion restriction, 422–423
- fixed effects, 440–442, 449
- fundamental problem, 340–342, 377, 379
- fuzzy regression discontinuity, 434
- hypothetical example, 339–340
- ignorability, 388–389, 413, 417, 418, 422–423, 431, 444
- imbalance, 344, 347, 358, 388, 391–397, 411–413
 - graph, 393, 395, 396, 402, 403
- individual treatment effect, 342–344
- infinite regress of causation, 415
- instrumental variables, 421–432, 438, 449–450, 454
 - adjusting for covariates, 428
 - assumptions, 422–423, 427–428, 431–432
 - derivation, 423–428
 - identifiability, 427–428, 430
 - standard error, 429
 - two-stage least squares, 428–430
 - Wald estimate, 426, 428
 - weak instruments, 431
- intent-to-treat (ITT) effect, 421, 423, 426, 428
- intermediate outcomes, 374–379
- local average treatment effect (LATE), 416, 426
- mapping a hypothetical intervention, 414
- mapping your study to an ideal controlled experiment, 415–416
- matching, 399–405
 - propensity score, 399–405, 407, 417, 418
- mediation, 374, 377–379
- model extrapolation, 385, 392, 394
- monotonicity, 422–423
- multilevel model, 342, 440
- multiple treatments, 342, 413, 415
- natural experiment, 431
- observational study, 7, 383–416
- overlap, 344, 388, 391–397, 411–413, 416, 438
 - graph, 392, 393, 396, 403
- paths, 376–379
- population average treatment effect (PATE), 343, 348
- post-treatment variable, do not adjust for, 374–379
- poststratification, 342, 370, 372, 397–399
- potential outcomes, 339–342, 356, 358, 363–364, 375, 377, 379, 383, 384, 388, 414, 418, 423
 - close substitutes, 341
 - interpreting regression coefficients, 134
- predictive comparison, 383, 384
- principal stratification, 378
- propensity score matching, 399–405, 407, 417, 418
- propensity score weighting, 417
- randomized encouragement, 421
- randomized experiment, 389
 - cluster-randomized, 349
 - completely randomized, 388, 389
 - matched pairs, 140, 349, 352, 366
 - randomized blocks, 140, 347–349, 351, 352, 360, 366
- regression discontinuity, 432–438, 445, 449, 452
- sample average treatment effect (SATE), 343, 348
- selection bias, 383
- selection model, 417
- selection on observables, 389, *see also* ignorability
- sensitivity analysis, 417
- simultaneous causality, 416
- sketch, 364
- stable unit treatment value assumption (SUTVA), 350, 353–354, 356
- structural equation model, 432
- subclassification, 397–398, 416
- support of the data, 385
- treatment interaction, 370–371, 398, 436–438
- treatment variable, 414
 - using regression, 363–382
- varying-intercept model, 440–442
- weighting, 397, 398, 409–411
- causal statements and causal questions, 448
- causes of effects, 415, 445–448
- `cbind()`, 143, 156
- censored data, 323, 458
 - imputation for, 332–333
 - picture, 332
- centering, 137, 185–187

SUBJECT INDEX

525

- interactions and, 243
- Central Limit Theorem, 41–42, 77
- child care example, 374–379, 394–397, 399, 416, 430
- children with HIV, 297–300
- children’s test scores, 151, 185–187
- Chile schools example, 433–440
- choice models, 287, 288
- classical inference, 14–15
- cleaning data, 485–488
- cluster sampling, 311
- cluster-randomized experiment, 349
- co-op election example, 63–65
- cockroaches, *see* roaches example
- coef(), 82, 98, 156
- collinearity, 146, 198
- colochos, xv
- colors, 156
- combining inferences from multiple imputations, 326
- common causal support, 416
- communication, 23
- comparing distributions, 45
- comparing two surveys, 296
- comparisons as regression models, 99–102
- comparisons, graphs as, 31
- compatibility interval, 111
- complete-case analysis, 324
- compliance, 380, 421
- complier, 422, 431, 439, 449
- complier average causal effect (CACE), 422, 450
- compound model
 - imputing a variable that can be positive or zero, 329–330
 - simulation of, 70, 72, 285–286, 329–330
- computing, 16–17, 109–111, 467–470, 475–492, 495
- conditional average treatment effect (CATE), 343
- conditional independence, 389
- confidence interval, 51–53, 58
 - continuous data, 53
 - coverage, 52, 77
 - discrete data, 54
 - for proportion when $y = 0$ or n , 52, 67
 - proportions, 52
- confounding covariate, 317, 380, 384–386, 392, 393, 395–396, 438
- Congress example, 140–144, 287, 335, 381
- constructed observational study, 405
- constructive choice model, 279–283
- contaminated data, 278
- contamination model, 283
- continuous and discrete predictors, 196
- continuous probability simulation, 77
- controlled experiment, 415
- correlated error, 459
- correlation, 43, 48, 169, 179, 187–189
- cost-benefit analysis, 78, 280, 311–312
- count data, 263, 264, 266, 272
- count data and binary data, 272
- counterfactual, *see* causal inference
- counterfactual and predictive interpretations of regression, 134
- coverage example, 51
- coverage of confidence intervals, 52, 77, 98, 295
- cows example, 381
- criminology, 291
- cross validation, 172–180, 230–232, *see also* leave-one-out cross validation
- crossover trials, 341
- Current Population Survey, 417
- curve(), 37, 476
- cutpoints, 274, 276
 - ordered logit or probit, 274
- data matrix, 145
- data subsetting, 495
- data.frame(), 82
- dead salmon, 61
- death penalty, 54, 271, 332–333
- decision analysis, 129
 - arsenic example, 279–283
- decline effect, 311
- deep learning, 466
- default prior distribution, 124
- defier, 422, 423, 439
- degrees of freedom, 53, 110, 168, 169, 175
- dependent variable, why we avoid the term, 145
- design analysis, 291–312
 - after the data have been collected, 304
 - for continuous outcomes, 297–300
 - for proportions, 293–297
 - for regression coefficients, 300
 - graph, 292
 - inherently speculative, 299
 - unequal sample sizes, 296
- design of sampling and experiments, 291–312
- deterministic or random imputation, 327, 329, 334
- deviance, 174, 230–232
- diagnostics, 156–180
 - external validation, 171
 - residuals, 161, 253
 - simulation-based, 69–78, 288
- diarrhea, zinc, and HIV, 297–300
- difference-in-differences, 442–445, 449
- different software example, 492
- discrete and continuous predictors, 196, 260
- discrete probability simulation, 76
- distance metrics, 406
- distance to the nearest safe well, 234
- distribution, 40–45, 47
 - binomial, 43

- comparing, 45
- exponential, 124
- logistic, 226
- lognormal, 43
- negative binomial, 263
- normal, 41–43, 145
- Poisson, 44, 263, 265
- t, 53, 278
- χ^2 , 53
- divide-by-4 rule, for interpreting logistic regression, 220
- dividing by two standard deviations, 186–187
- dummy variable, *see* indicator
- dynamic graphics, 33
- earnings
 - height and, 183, 284
 - logarithmic models, 189–195
 - mixed discrete/continuous data, 283–284
- earnings example, 74, 84–85, 91, 189–195, 284
- econometrics and biostatistics, 417
- education as categorical input variable, 245
- effect size, why more important than sample size, 293
- effects of causes, 415, 446
- effects, why this term can be misleading, 84–85
- efficiency, 352
- elasticity, 195, 213
- election fraud, 63–65
- elections, 4, 62, 93–99, 101, 120, 140–144, 257, 335, 454
- elections and economy example, 4, 93–99, 113, 115, 120, 469
- Electric Company example, 6, 312, 364–372, 379, 386–394
 - graph of data, 365
- elephants, 24, 39
- encouragement design, 358
- ensemble learning, 466
- equal variance of errors, 154
- error rate, 255
 - where it can mislead, 255
- estimands and estimates, 50
- examples, 4, 476
 - Academy Awards, 288
 - age-period-cohort, 32
 - arsenic, 232–238, 242–249, 279, 283, 379
 - beauty and teaching evaluations, 151, 213
 - bypass, 452
 - child care, 374–379, 394–397, 399, 416, 430
 - Chile schools, 433–440
 - co-op election, 63–65
 - Congress, 140–144, 287, 335, 381
 - coverage, 51
 - cows, 381
 - death penalty opinions, 54
 - different software, 492
 - earnings, 74, 84–85, 91, 189–195, 284
 - elections and economy, 4, 93–99, 113, 115, 120, 469
 - Electric Company, 6, 312, 364–372, 379, 386–394
 - fake K-fold CV, 179
 - fake midterm and final, 88–89, 306–310
 - flight school, 90
 - gay marriage, 198, 467, 472
 - girl births, 68
 - golf, 461–464, 471
 - health expenditure, 25
 - height and weight, 42, 117–119, 136–139, 484–488
 - helicopter, 17–18
 - Human Development Index, 21–22
 - imputation, 322, 335
 - incentives, 373–374
 - intro class, 161
 - kid IQ, 131–135, 151, 156, 161, 185
 - LaLonde, 287, 417
 - logistic priors, 229
 - logit graphs, 241
 - mesquite bushes, 200–206
 - metabolic rate, 39–40
 - mile run, 37–38, 482
 - names, 27–28, 34
 - NES, 149, 217, 237, 257, 287
 - Newcomb, 163–165
 - parabola, 488–489
 - peacekeeping, 7–8
 - Pearson and Lee, 85–88
 - Pew, 23
 - Poisson, 265, 266
 - pollution, 212
 - poststratification, 313–316, 320–322
 - predicting student grades, 206–210
 - probability of a tied election, 45–47, 144
 - probability simulation, 69
 - pyth, 181
 - R^2 , 170
 - residuals, 158
 - restaurant, 489–490
 - risky behavior, 286, 288
 - roaches, 268–270, 284–285
 - rodents, 239
 - sample size, 302–304
 - scalability, 470
 - Sesame Street, 357, 422, 428
 - sex ratio, 121–122, 125–128, 304–306
 - simple causal, 10–11
 - simplest, 82, 99
 - stents, 45
 - storable votes, 274–278, 286
 - unemployment, 166–168

SUBJECT INDEX

527

- exclusion restriction, 422, 423, 431, 439
 exercise time, 45
 expectation, 41
 experimental benchmark, 405
 experimental design, 291–312, 347–350
 explained variance, *see* R^2
 exploratory data analysis, 30
 exponential distribution, 124
 exponential growth and decline, 38
 exposure, analogue to treatment in observational studies, 386, 414
 exposure, in Poisson or negative binomial regression, 267–268
 expression(), 248
 external validity, 171, 354, 380
 extra-virgin olive oil, 23
 extrapolation, 5, 6, 354

 F test, why we do not use it, 147
 factor variable, 198
 fake K-fold CV example, 179
 fake midterm and final example, 88–89, 306–310
 fake-data simulation, 76, 82–83, 91, 97–99, 101, 102, 150, 161–163, 176, 181, 239, 242, 265, 302–304, 312, 320–322, 334, 358, 494, 496
 feeling thermometer, 227
 file-drawer effect, 61
 fit many models, 495
 fitting the wrong model, 181
 fixed effects, 440–442, 449
 why we avoid the term, 440
 flight school example, 90
 forcing variable, *see* assignment variable
 forking paths, 61, 67
 forward and reverse causal inference, 445–448
 functions in R, 72, 326–328
 fundamental problem of causal inference, 340–342, 377, 379
 fuzzy regression discontinuity, 434, 438
 connection to instrumental variables, 438

 gain score, 369, 380
 garden of forking paths, 61
 Gaussian process, 465
 gay marriage example, 198, 467, 472
 General Social Survey, 261, 288
 generalized least squares, 148
 generalized linear model, 263–288
 beta-binomial, 272
 binomial, 270–272, 287
 building, 283–285
 cutpoints, 274
 logistic, 217–261
 logistic-binomial, 263, 270–272
 multinomial logit and probit, 264, 273–278, 286, 287
 negative binomial, 263–270, 286
 exposure, 267–268
 ordered logistic, 264, 273–278, 286, 287
 Poisson, 263–270, 286
 compared to binomial, 268
 exposure, 267–268
 probit, 264, 272–273
 probit or logit, 272
 robit as robust alternative, 278–279
 robust, 264
 simulation, 247–286
 thresholds, 274
 Tobit, 284, 286, 287
 generative model, 50
 genetic algorithms, 466
 geometric mean and standard deviation, 43
 ggplot2 (R package), 475
 girl births example, 68
 glm(), xiv
 goals of regression, 9–13, 18
 golf example, 461–464, 471
 goodness-of-fit, *see* model checking
 grades, predicting, 161
 graphics, 21–23, 25–31, 33, 156–163, 482–484, 493
 alternative displays of the same data, 365
 as comparisons, 31
 finding unexpected patterns, 31
 jittering, 132
 R, 33
 residual plot, 160
 scatterplot with regression lines
 superimposed, 135, 156–163
 showing fitted models, 31
 why, 30
 graphing a line, 37–38
 guessing on a multiple-choice test, 283
 gun control, 8–9, 17

 Hamiltonian Monte Carlo, 468–469
 handedness, 196
 Hawthorne effect, 354, 356
 health expenditure example, 25
 height, 72, 77, 85, 90
 earnings and, 183, 284
 logarithmic models, 189–195
 mixture model for, 41
 parents and children, 188
 weight and, 169, 212
 height and weight example, 42, 117–119, 136–139, 484–488
 helicopter example, 17–18
 heteroscedasticity, 154, 283
 hierarchical model, 350, 440, 460
 hist(), 69, 328
 history of regression, 85–87, 90

- holding other predictors constant, difficulty of, 133
 homoscedasticity, 154, 283
 horror movie, 269
 horseshoe prior for sparse regression, 179,
 208–210, 460
 hot-deck imputation, 330
 Human Development Index example, 21–22, 34
 hypothesis testing, 57–68, 147
 moving beyond, 66–67
- I(), 330
 ideal-point model, 288
 identifiability, 146
 categorized predictors and, 146, 198
 instrumental variables, 427–428, 430
 likelihood and, 146
 linear regression, 146, 198
 logistic regression, 227–228, 256–259
 idiosyncratic transformation, 213
 ifelse(), 327, 479
 ignorability, 323, 333, 350–352, 386, 388–389,
 413, 417, 422
 imbalance, *see* causal inference
 imputation, 322–335
 impute(), 328
 incentives example, 373–374
 income and voting, 217–222, 237
 incremental cost-effectiveness ratio, 78
 incumbency, 381, 454
 congressional elections, 141
 independence of errors, 154
 independent variable, why we avoid the term, 145
 index variable, 197
 indicator variable, 99–101, 132, 136–139, 197
 default, reference, or baseline condition, 198
 individual treatment effect, 343
 Inf, 478
 Infant Health and Development Program, *see* child
 care example
 inference, *see* statistical inference
 influence, 107–108, 111
 instrumental variables, *see* causal inference
 intent-to-treat (ITT) effect, 426, 428
 interactions, 134–136, 151, 185–187, 193, 360
 centering the input variables, 243
 compared to main effects, 301–302
 graphing, 245
 logistic regression, 242–247
 sample size and, 300–304, 312
 smoking and cancer, 10, 135
 treatment effects, 370–371, 398
 when to look for, 135
 intercept, 86, 95, 133, 135
 intermediate outcome, *see* causal inference
 internal validity, 354
 interpreting regression coefficients, 95, 205, 243,
 244
 interquartile range (IQR), 201
 intro class example, 161
 inverse estimated probability of treatment
 weighting, 409–411, 420
 invlogit() and logit(), 217
 iterative regression imputation, 331
 iteratively weighted least squares, 228
- Jacobian, 202–203
 Jamaica childhood intervention, 15, 17
 jittering, 132, 234
- K-fold cross validation, 178–179
 kid IQ example, 131–135, 151, 156, 161, 185
 knitr (R package), 475
- LaLonde example, 287, 417
 lasso prior for sparse regression, 210, 460
 latent-data formulation for logistic regression,
 226–228, 274
 Latinos, hypothetical survey of, 310
 least squares, 103–109, 111, 112, 146
 weighted, 147–148
 leave-one-out cross validation, 112, 172–178,
 201–206, 232, 235–236, 244
 loo (R package), 475
 legislative redistricting, 27
 level of a factor variable, 198
 library(), 476
 likelihood
 function, 146
 logistic regression, 228
 surface, 105–106
 linear predictor, 116, 219, 223
 linear regression, 131–214
 assumptions, 153–161
 compared to principal component line,
 187–188
 correlation and, 187–189
 counterfactual interpretation, 134
 diagnostics, 153–161
 displaying several, 148–149
 general principles for building models, 199
 interactions, 134–136
 interpreting coefficients, 133–136
 interactions, 135–136
 missing-data imputation, 326–330
 multiple predictors, 131–152
 notation, 145–146
 picture of matrix, 145
 prediction, 171–172, 200–206
 predictive interpretation, 134
 sample size and power calculations, 300
 simulation, 142–144
 statistical inference, 144–147

SUBJECT INDEX

529

- transformation, 183–214
 - validation, 171–172
 - weighted, 147–148
- linear transformation, 42–43, 48, 54, 137, 183–185, 193–194, 227, 234, 243, 245, 276
- linearity assumption, 154
- link function, 263
- lm(), compared to stan_glm(), xiv, 16, 109–112
- local average treatment effect (LATE), 416, 426
- loess(), 465, 467
- log score, 174–179, 230–232
- log(), 189
- log-log transformation, 194, 212, 213
- log10(), 192
- logarithmic transformation, 38–40, 43, 68, 189–195, 212, 254
 - even when not necessary, 195
 - interpreting coefficients, 190, 192
 - picture, 190
 - why base e and not base 10, 191–192
- logistic curve, 218
- logistic distribution, close to normal with standard deviation 1.6, 227, 273, 286
- logistic priors example, 229
- logistic probability density function, 226
- logistic regression, 217–261
 - binned residual plot, 259
 - choice models in one and multiple dimensions, 280–282
 - compared to probit, 272, 281, 287
 - divide-by-4 rule for interpreting coefficients, 220
 - graphing, 218, 234, 236, 239, 241–242, 245, 259
 - graphing data and fitted curve, 219
 - identifiability, 227–228, 256–259
 - inference, 221
 - interactions, 242–247
 - interpreting coefficients, 220–222, 235, 238
 - latent-data formulation, 226–228
 - logit and logit⁻¹ functions, 219
 - missing-data imputation, 326–330
 - odds ratio, 220–221
 - pictures, 218
 - predictive comparison, 220, 249–252
 - propensity score, 401, 407
 - separation, 256–259
 - simulation, 247
 - standard error, 221
 - two predictors, 235–237
 - using Stan, 461
 - well-switching in Bangladesh, 232–237
- logistic-binomial regression, 263, 270–272
 - overdispersion, 271
- logit, *see* logistic regression
- logit graphs example, 241
- logit() and invlogit(), 217
- lognormal distribution, 43
- loo(), 172, 177–178, 475
- loo_predict(), 174
- looking at the data, 488
- loops, 479
- lurking variable, 385
- machine learning, 464–467
- mad sd, 73
- mad(), 73
- Mahalanobis distance, 407
- many predictors, 459–460
- MASS (R package), 179, 285
- matched pairs experiment, 140, 349, 352, 366
- matching, *see* causal inference
- matching(), 401, 402
- maternal IQ, 132
- matrix notation, 145–146
- matrix of simulations, 143, 248
- maximum likelihood, 105–107, 110, 111, 146, 229–230
 - logistic regression, 228
- mean, 41
- mean imputation, 325
- mean(), 186
- measurement, 23–25
- measurement error, 49, 458, 471
- median, 73
- median absolute deviation, 73, 113
- mediation, *see* causal inference
- mesquite bushes example, 200–206
- messy data in R, 484–488
- metabolic rate example, 39–40
- mice, external validity of studies on, 23
- midterm and final exams, 161–163
- mile run example, 37–38, 482
- millimeters, inches, and miles, 183
- missing at random (MAR), 323, 333
 - impossibility of verifying, 323
- missing completely at random (MCAR), 323
- missing data, 322–335, 355, 458–459, 478
 - in R, 322
- missing-data imputation, 322–335
 - available-case analysis, 324
 - complete-case analysis, 324
 - congressional elections, 142
 - deterministic or random, 327, 329
 - iterative, 331
 - many variables, 330–332
 - matching, 330
 - models, 322–324
 - multiple imputation, 326–335
 - simple methods, 324–325
 - topcoding, 327
- mixed discrete/continuous data, 283–284, 329

- model checking, 12–13, 33, 494, 496
 using simulation, 163–168
 model error, 49
 model extrapolation, 385, 392, 394
 model selection, 180
 monotonicity assumption for instrumental
 variables, 422, 423, 439
 Monte Carlo, 468
 mortality rates, 31–33
 mothers and children, 85–88, 151, 185–187
 multilevel model, 350, 440, 460
 multinomial logit and probit regression, 264,
 273–278, 286, 287
 storable votes, 274–278
 multiple comparisons, 66
 multiple imputation, 326–335
 combining inferences, 326
 multiplicative model, 189
 multivariate imputation, 330–332
 mvnrm(), 179
 NA, 151, 478
 names example, 27–28, 34
 naming inputs, 192
 NaN, 478
 National Election Study, 149
 National Longitudinal Survey of Youth, 131, 405
 National Supported Work program, 417
 natural experiment, 431
 natural log, 191–192
 negative binomial regression, 263–270, 286
 exposure, 267–268
 interpreting coefficients, 267
 zero-inflated, 284–285
 NES example, 149, 217, 237, 257, 287
 never-taker, 422, 423, 439, 449
 Newcomb example, 163–165
 noncompliance, 355
 nonidentifiability, *see* identifiability
 nonlinear modeling, 460–464, 471
 nonparametric regression, 464–467
 nonresponse weighting, 325
 normal distribution, 41–43
 estimated regression coefficients, 221
 notation, 145
 regression errors, 154
 normality of errors, 154
 not missing at random (NMAR), 323
 notation for regression models, 81, 144–147
 observational study, 7, 383
 observed measurements, generalizing from, 3
 odds ratio, 220–221
 offset in Poisson or negative binomial regression,
 267, 287
 omitted variable bias, 385
 omniscience, 357
 open data, 66
 optimal design, 312
 optimizing in R and Stan, 488–490
 optimizing setting for `stan_glm()`, 16, 109, 470
 options(), 30
 ordered and unordered categorical outcomes, 273,
 277
 ordered logistic regression, 264, 273–278, 286, 287
 outcome variable, 4
 outlier, 178
 outliers in binary data, 278
 overdispersion, 266, 271
 simulation, 285
 overestimates, published results tend to be, 62, 293
 overfitting, 104, 175
 overlap, *see* causal inference
 ovulation and voting, 62, 66
p-hacking, 61, 67
p-values, *see* hypothesis testing
 packages in R, 475
 paired comparisons, 140, 366
 Panel Study of Income Dynamics, 417
 parabola example, 488–489
 parameters, 50
 partial pooling, 119, 128
 partisan bias, 27
 party identification, 23, 315
 paste(), 478
 pch, 156
 peacekeeping example, 7–8, 17
 Pearson and Lee example, 85–88
 Pew example, 23
 phase diagram for decision analysis, 282
 plm (R package), 440, 449
 plogis(), 217
 plot(), 156, 483
 plots of replicated datasets, 164–166
 plotting a series of regressions, 149
 plotting the data and fitted model, 156–163
 pnorm(), 96
 point prediction, 115–116, 223
 Poisson distribution, 44
 Poisson example, 265, 266
 Poisson regression, 263–270, 286
 checking using simulation, 268–269
 compared to binomial, 268
 exposure, 267–268
 interpreting coefficients, 267
 zero-inflated, 284–285
 political ideology, 23, 149
 political party identification, 148–149
 polling, 6, 54, 120, 149, 222, 257, 314–320
 pollution example, 212
 population average treatment effect (PATE), 343,
 348

SUBJECT INDEX

531

- Portugal students example, 206–210
 post-treatment outcomes, 445
 posterior predictive check, 163–168, 180
 posterior simulations, 113–115
 posterior uncertainty, 117, 247
 posterior_epred(), 116, 223, 315
 posterior_linpred(), 116, 223
 posterior_predict(), 116, 224
 poststratification, 313–322, 370, 372, 397–399
 poststratification example, 313–316, 320–322
 potential outcome, *see* causal inference
 power analysis, 291–312
 2.8 standard errors, 295
 classical, 293–300
 for continuous outcomes, 297–300
 for proportions, 293–297
 for regression coefficients, 300
 general concerns, 294
 graph, 292, 296
 inherently speculative, 299, 300
 picture, 295
 unequal sample sizes, 296
 power-law growth and decline, 38–40
 pre-test, 363–364, 367–370, 380
 sketch, 364
 pre-treatment variables, 352
 predict(), 115–116, 223, 401
 predicting student grades example, 206–210
 prediction, 12, 95, 96, 102, 113–129, 142, 151,
 171–172, 199–206, 222
 interpreting regression coefficients, 134
 prediction error, 174
 predictive check, *see* posterior predictive check
 predictive comparison, 220, 249–252, 259
 formula, 249
 graph, 250
 interactions, 252
 predictive comparison in causal inference, 383, 384
 predictive comparison in regression, *see* average
 predictive comparison
 predictive distribution, 115–119, 224
 predictive simulation, 128, 143, 163–168, 247–286
 binomial distribution, 248
 generalized linear models, 247–286
 latent logistic distribution, 248
 linear regression, 142–144, 151
 model checking and, 163–168
 predictors, 4, 144
 presidential elections, 93–99, 217–222
 principal component line, 187, 188
 principal stratification, 379
 prior distribution, 109–110, 123–127, 129, 259
 default, 124
 for regression with many predictors, 206–210
 horseshoe, 208–210, 460
 informative, 125–128
 lasso, 210, 460
 sex ratio example, 125
 uniform, 123
 weakly informative, 124–125, 208
 prior information, 14, 15, 62, 119–121, 229
 sex ratio example, 121
 prior_summary(), 491
 probability, *see* distribution
 probability modeling, 45–47
 probability models, simulation of, 69–72
 probability of a tied election, 45–47, 144
 probability simulation example, 69
 probit regression, 264, 272–273, 286, 287
 compared to logit, 272, 281
 profile likelihood, 257, 258
 programming in R, 488–492
 programming in Stan, 461–464, 492
 propagating uncertainty, 56, 77, 113–119, 247–249
 propensity score, *see* causal inference
 psychology, 62
 publication bias, 59, 305
 pyth example, 181

 Q-Q plots, why we do not recommend, 155
 qlogis(), 217
 quantile(), 73
 quick tips, 493–496

 R, 16–17, 475–492
 abline(), 4, 156
 arm package, 475
 as.matrix(), 113
 bayes_R2(), 171, 491
 bayes_sim(), 229
 bayesplot package, 475, 492
 brm(), 272, 278, 284, 285, 430
 brms package, 430, 492
 cbind(), 143, 156
 coef(), 82, 98, 156
 colors, 156
 curve(), 37, 476
 data types, 477
 data.frame(), 82
 digits, 30
 expression(), 248
 functions, 72, 326–328, 477, 488
 getting started, 475
 ggplot2 package, 475
 glm(), xiv
 graphics, 33, 482–484
 hist(), 69, 328
 I(), 330
 ifelse(), 327, 479
 impute(), 328
 Inf, 478
 invlogit(), 217, 248
 knitr package, 475

- library(), 476
- lm(), xiv, 16, 109–111
- loess(), 465, 467
- log(), 189
- log10(), 192
- logit(), 217
- loo package, 172, 475
- loo(), 172, 177–178
- loo_predict(), 174
- loops, 479
- mad(), 73
- MASS package, 179, 285
- matching(), 401, 402
- mean(), 186
- messy data, 484–488
- mvrnorm(), 179
- NA, 151, 478
- NaN, 478
- optimizing in, 488–490
- optimizing setting for stan_glm(), 16, 109, 470
- options(), 30
- packages, 475
- paste(), 478
- pch, 156
- plm package, 440, 449
- plogis(), 217
- plot(), 156, 483
- pnorm(), 96
- posterior_epred(), 116, 223, 315
- posterior_linpred(), 116, 223
- posterior_predict(), 116, 224
- predict(), 115–116, 223, 401
- prior_summary(), 491
- programming, 488–492
- qlogis(), 217
- quantile(), 73
- rbinom(), 69, 72, 248
- rbounds package, 417
- rbrms package, 264
- read.csv(), 482
- read.fwf(), 485
- read.table(), 4, 181, 482
- reading data, 481
- replicate(), 72
- rnegbin(), 285
- rnorm(), 72, 97, 260
- rowMeans(), 372
- rowSums(), 143
- rpois(), 72
- rprojroot package, 476
- rstan package, 475
- rstanarm package, xiv, 82, 264, 475, 476, 490–492
- RStudio, 475
- runif(), 248, 477
- sample(), 68, 70, 326, 477
- sd(), 186
- se(), 98
- seq(), 477
- shinystan package, 492
- stan_gamm4(), 467
- stan_glm(), xiv, 16, 82, 109–111, 218
- stan_polr(), 276
- subscripting, 480
- subset option in stan_glm(), 330
- sum(), 143
- survey package, 404, 475
- table(), 488
- working directory, 481
- writing data, 482
- R^2 , 84, 168–171, 193, 491
 - Bayesian, 170
 - example, 170
 - graph, 169
- radon, 11, 17
- random imputation, 326, 327, 329
- randomization, 78, 346–347, 357, 381
 - distribution, 346
- randomized blocks experiment, 140, 347–349, 351, 352, 360, 366
- randomized experiment, 345–355
 - limitations, 350–355
- randomized-encouragement design, 421
- ratio of parameters, 78
- rats, 39, 239
- rbinom(), 69, 72, 248
- rbounds (R package), 417
- read.csv(), 482
- read.fwf(), 485
- read.table(), 4, 181, 482
- reading data in R, 481
- redistricting, 27
- regression, *see* linear regression, logistic
 - regression, generalized linear models
 - why learn, 4–5
- regression discontinuity, *see* causal inference
- regression to the mean, 85–91, 187–189
- regularization, 459–460
- reliability, 23–25, 34
- replicate(), 72
- replicated datasets, plotted, 164–166
- replication, 148–149, 493
- replication crisis in science, 62, 67, 293
- representativeness, 25, 154
- researcher degrees of freedom, 61, 67
- residual standard deviation, 84, 104, 168–171
 - models of, 283
- residual sum of squares, 103
- residuals, 145, 253
 - binned, 253
 - graph, 161–163, 172, 253

SUBJECT INDEX

533

- graph vs. predicted values, not vs. observed values, 162, 163
- standard deviation of, 169
- residuals example, 158
- restaurant example, 489–490
- retrospective observational study, 415
- risky behavior example, 286, 288
- rnegbin(), 285
- rnorm(), 72, 97, 260
- roaches example, 268–270, 284–285
- robit regression, 278–279, 287
 - generalization of logit and probit, 279
 - picture, 278
- robust regression, 264, 278–279, 286
- rodents example, 239
- rowMeans(), 372
- rowSums(), 143
- rpois(), 72
- rprojroot (R package), 476
- rstan (R package), 475
- rstanarm (R package), xiv, 82, 264, 475, 476, 490–492
- RStudio, 475
- runif(), 248, 477

- sample average treatment effect (SATE), 343, 348
- sample mean, 53
- sample size calculation, 102, 291–312
 - 2.8 standard errors, 295
 - classical, 293–300
 - for continuous outcomes, 297–300
 - for proportions, 293–297
 - for regression coefficients, 300
 - general concerns, 294
 - inherently speculative, 299, 300
 - interactions, 300
 - picture, 295, 296
 - unequal sample sizes, 296
- sample size example, 302–304
- sample to population, generalizing from, 3
- sample(), 68, 70, 326, 477
- sampling distribution, 50, 53, 102, 346–347, 357, 361
- sampling model, 49
- sampling with replacement, 74, 477
- sampling, design for, 291–312
- scalability example, 470
- scaling of predictors, 183, 211
- scatterplot, 21–22, 25–27
 - data and regression lines superimposed, 135
- sd(), 186
- se(), 98
- secret weapon, 148–149, 213, 493
 - pictures, 54, 149, 222
- selection bias, 15, 25, 59, 62, 91, 305, 308, 323, 344, 374, 383–385, 389, 417
 - simulation, 309
- selection on observables, 389, *see also* ignorability
- sensitivity, 391
- separation in logistic regression, 256–259
- seq(), 477
- Sesame Street example, 357, 421–426, 428, 449
- sex ratio example, 121–122, 125–128, 304–306
- shinystan (R package), 492
- significance, *see* statistical significance
- significant digits, 29, 30, 34
- simple causal example, 10–11
- simplest example, 82, 99
- simulation, 69–78, 129, 142–144, 151, 288
 - combined with analytic calculations, 144
 - comparing simulated to actual data, 163–168
 - compound model, 70, 72, 285–286, 329–330
 - coverage of confidence intervals, 98
 - displaying uncertainty in a fitted model, 247
 - fake-data, 76, 97–99, 101, 102, 150, 161–163, 176, 320–322, 358, 494, 496
 - generalized linear models, 247–286
 - how many draws are needed, 152
 - logistic regression, 247
 - matrix of simulated parameters and predictions, 143, 248
 - nonlinear predictions, 140–144
 - overdispersed models, 285
 - posterior predictive check, 163–168
 - predictive, 247–286
 - probability models, 69–72
 - regression inferences, 142–144
 - replicated datasets, plotted, 164–166
- simultaneous-equation model, 459
- smoking and cancer, 10, 135
- Social Indicators Survey, 322–332, 335
- speed of light, 164
- spline, 465
- sports, 44, 47, 461–464
- square root transformation, 328
- stable unit treatment value assumption (SUTVA), 350, 353–354, 356
- Stan, 283, 468–470, 475, 476
 - building models in, 460–464
 - editing code from brms, 492
 - getting started, 476
 - optimizing in, 488–490
- stan_gamm4(), 467
- stan_glm(), xiv, 16, 82, 109–111, 218
 - why we use it, xiv, 109
- stan_polr(), 276
- standard deviation, 41, 53
- standard error, 51, 52
 - for a proportion, 51
- standardizing predictors, 184–187, 211, 247
- statistical error, 57–60
- statistical inference, 50–60, 144–147

- 3 challenges of, 3
 - graph of uncertainty, 221
 - standard error, 221
- statistical significance, 57–62, 221, 244, 493
 - problems with, 60–62
 - sample size and power, 294
- statistical significance filter, 59, 62, 305
- stents example, 45
- storable votes example, 274–278, 286
 - data and fitted curves, 275
- structural equation model, 432, 449
- subclassification, 416
- subscripting in R, 480
- subset option in `stan_glm()`, 327, 330
- subsetting data, 495
- sum of squares in least squares estimation, 104
- `sum()`, 143
- superpopulation, 49
- surly people, 323
- survey (R package), 404, 475
- survival analysis, 458
- switches, 167
- symbols on graphs, 27
- t distribution, 53, 278, 286
- t regression, 278–279
- t test, why we do not use it, 147
- `table()`, 488
- tables, displaying, 30, 34
- teaching evaluations and beauty, 151, 213
- test scores in Chile, 433–440
- test summary
 - graphical, 164–166
 - numerical, 63, 165, 269
- threats to validity, 354
- thresholds for ordered logit or probit, 274
- tied election, probability of, 45–47, 144
- time series, checking a fitted model, 166–168
- Tobit regression, 284, 286, 287
- topcoding for missing-data imputation, 327
- traffic accidents, 264–267
- transformation, 183–214, 496
 - idiosyncratic, 196, 213
 - linear, 42–43, 48, 183–185
 - log-log, 212, 213
 - logarithmic, 43, 68, 189–195, 212, 254
 - square root, 195, 328
- treatment, *see* causal inference
- treatment to control group, generalizing from, 3
- tree model, 466
- true values in fake-data simulation, 97
- twins, 70
- two-sided test, 57
- two-stage least squares, *see* causal inference, instrumental variables
- two-stage model for mixed discrete/continuous data, 283, 329
- type 1 and 2 errors, 58–59
- type M and S errors, 59, 67, 292, 305
- unbiased estimation, 55, 346–347
- uncertainty interval, 51, 68, 73, 111
- uncontested elections as missing data, 142, 335
- underdispersion, 266
- unemployment example, 166–168
- unequal variances, *see* heteroscedasticity
- unexpected patterns in graphs, 31
- unexplained variance, *see* R^2
- United Nations, 7–8, 17
- unordered categorical regression, 278
- utility theory, 280
- validation, 171–172
- validity, 23–25, 34, 153
- value function, 279
- value of a statistical life, 381
- variance, 41
 - explained and unexplained, 84, 168–171
 - models of, 283
- varying-intercept model, 440–442
- vector-matrix notation, 36–37, 145–146
- Vietnam War draft lottery, 431, 449
- visual and numerical comparisons of replicated to actual data, 167
- vote share, 96–97
- voting and income, 217–222, 237
- Wald estimate for instrumental variables, 426, 428
- weak instruments, 425, 431, 449
- weakly informative prior, 208, 259
- weight
 - age and, 211
 - example of a lognormal distribution, 43
 - height and, 212
- weighted average, 35–36, 47, 54, 108–109, 314, 397, 398
 - formula for Bayesian inference, 119
- weighted least squares, 147–148
- weighted regression, three models for, 147–148
- well-switching in Bangladesh, *see* arsenic example
- why ask why, 445–448
- winner's curse in low-power studies, 292
- within-person controls, 442
- workflow, 3–496
- working directory in R, 481
- writing data in R, 482
- χ^2 , 53, 65
- Xbox survey, 6, 316–320
- z-score, 150, 184
- zero-inflated negative binomial or Poisson regression, 284–285
- zinc and HIV, 297–300