INDEX

A

Abadie, A., 44, 387, 393
Abortion acceptance; correlation ratios of, 101*fig*; by education and religious denomination, 131*fig*; factor loadings before and after rotation, 253*t*–254*t*; multiple-item scale on religiosity and, 250–257; percent by religion and education, 28*t*; percentage by religiosity and education, 27*t*
Academic Technology Services (UCLA), 186
Added-variable plots, 232–234, 234*fig*
Additive effects; definition of, 27; example of, 27–28*fig*
Additive scaling, 247
Adjusted R^2, 111–112
African Americans; distribution of religiosity and educational attainment, 13*t*–14; militancy by educational attainment, 14*t*–15; militancy by religiosity among urban, 10*t*–11; militancy distribution by religiosity among urban, 8*t*–9; militancy percent by religiosity and educational attainment, 15*t*–16, 17–18*t*; *See also* Racial differences
Age at first marriage; binomial logistic regression of, 318–327; discrete-time hazard-rate model of, 318–327; expected probabilities of, 320*fig*–322*fig*, 326*fig*; GSS (1994) data on, 319; odds ratios for model predicting likelihood of, 323*t*–324*t*
Age differences; curvilinear relationships of income and, 140–142; evolution beliefs by religion, sex, and, 37*t*–38*t*; expected probability of marrying by sex, mother's education, and, 326*fig*; joint probability distribution of education and, 36*t*; percentage of evolution beliefs by, 34*t*; proportion of evolution beliefs standardized for education and, 39*t*; relationship between 2003 income and, 141*fig*; religiosity association with, 17; religious denomination and, 35*t*
Aggregation; description of, 98–99; effect on correlations, 99*fig*
Agresti, A., 294
Alaniz, M. L., 355
Allison, P. D., 185, 363, 376, 380
Almond, D., 44, 385
Altman, D. G., 183
Aly, H. Y., 342
AMOS (software), 396
An American Dilemma (Myrdal), 397
American Sociological Association, 202
American Sociological Review, 2
American Statistical Association, 202
Ancillary parameters, 348
Anderson, A. B., 181
Andrews, F. M., 164
Angrist, A., 389
Annals of Eugenics (journal), 93
Annals of Human Genetics (journal), 93
Ansolabehere, S., 387
Anticommunist sentiment; expected percentage from agreement with, 276*t*; goodness-of-fit statistics for log-linear models on, 275*t*; log-linear analysis of, 273–277; predicting frequency distribution of, 272–273*t*
Approximate Bayesian bootstrap, 185
Arithmetic transformations, 75–76
Armed threats prevalence; effect parameters for models of, 310*t*–314; effect parameters for probit analysis, 331*t*–334; goodness-of-fit statistics for various models on, 308*t*–310; logistic regression example of, 304–314; percentage ever threatened by gun by selected variables, 306*t*–307*t*
Ashenfelter, O., 369, 380
Augmented component-plus-residual plots, 235*fig*
"Available from author" archive, 404

B

Baltagi, B. H., 384
Barkan, S. D., 400
Barker, D.J.P., 385
Barnes, S. H., 97
Baseline models, 268
Basilevsky, A., 181
Baum, C., 390
Baumer, E. P., 361
Bayesian Information Criterion *(BIC)*; comparing models using, 133–135, 151, 270–271; definition of, 133, 270; log-linear analysis of crossings models and, 286
Bayesian multiple imputation; comparing casewise deletion and, 191–193*t*; description of, 182, 185–186
Beardon, W. O., 246
Becker, M. P., 294
Becker, S. O., 44, 393
Berelson, B., 44
Berk, R. A., 182, 393
Bernstein, I. H., 246
Bias; omitted variable, 106, 364–380; sample-selection, 390
Bielby, W. T., 185, 395
Binary outcomes; FE (fixed effects) models for, 375–376; RE (random effects) model for, 377–380

COPYRIGHTED MATERIAL

Binomial logistic regression; algebra equivalencies to use for, 330; examples of, 304–329; introduction to, 302–303; probit analysis alternative to, 302, 330–334*fig*, 361; relation to log-linear analysis, 303–304; *See also* Dummy (dichotomous) variables; Logistic regression
Binomial logistic regression examples; age at first marriage (hazard-rate model), 318–327; appointment to *nomenklatura* position in Russia (case-control model), 327–329; predicting prevalence of armed threats, 304–314; schooling progression ratios in Japan, 314–318
Biometrika (journal), 93
Bishop, Y.M.M., 266, 294
Black, D. A., 393
Blau, P. M., 112, 393, 395
Blinder, A., 175
Blinder-Oaxaca decomposition, 175
Bollen, K. A., 367, 371, 375, 397
Bonferroni adjustment, 231
Bootstrapping standard errors, 238–240
Boshuizen, H. C., 186, 190, 191
Boskin, M. J., 361
Brand, J. E., 44, 367, 371, 375, 393, 397
Breen, R., 182, 342
Breslow, N. E., 327
Brick, J. M., 181
Bryk, A. S., 229, 387
Budig, M. J., 380
Bureau for Applied Social Research, 22
Burgess, E., 242
Burke, P. J., 263, 273, 278, 294
Buttenheim, A. M., 380

C

Caliendo, M., 44, 393
California Center for Population Research (CCPR) [UCLA], 404
California city populations, 201*t*–202
Cameron, A. C., 361
Cameron, L., 361
Campbell, C., 380
Campbell, D. T., 43, 44, 98
Cappella, J. N., 361
Card-image records, 72
Carmines, E. G., 243, 246
Case-control sampling, 327
Casewise deletion; comparing multiple imputation and, 191–193*t*; description of, 183
Categorical dependent variables; definition of, 335; multinomial logit analysis of, 336–342
Censored (or truncated) dependent variables; definition of, 335; tobit regression for, 353–361
Censored prediction, 358
Chamberlain, G., 380
Chattopadhyay, A., 361
Checking for errors, 404–405
Chen, M. D., 355, 382, 392
Chen, S., 400
Cheng, M.-T., 185
Chick, G., 278
Chinese Cultural Revolution, 158, 159, 160*t*, 163*fig*
Chinese educational attainment; quality of education, 158–164; by size of place of residence at age fourteen, 289*t*–292
Chinese family income; comparison of OLS and FE estimates for determinants of, 374*t*; estimating/interpreting FE and RE models for, 372–375; SES characteristics by size of place of residence, 373*t*
Chinese Health and Nutrition Survey (1994), 397
Chinese literacy rates; graphic representation of results on, 118–120*fig*; multiple regression equations on determinants of, 113–118; survey estimation procedures on, 215–219; urban versus rural residence and, 42*t*–43
Chinese occupational status; frequency distribution by father's occupation, 280*t*; goodness-of-fit statistics for models of intergenerational, 284*t*; RC model used to predict mobility of, 291–293
Clark, T. G., 183
Clogg, C. C., 294
Clustering; design effects and, 207; strategy to offset, 209–212
Codebook; data organization and corresponding, 68*e*; definition of, 67
Coefficient of determination (R^2), 111–112
Coefficients; constraining to zero or to equality, 132; of determinants of estimated political conservatism model, 260*t*–261; expressed as deviations from grand mean, 164–166; included in model of educational attainment by race, 177*t*; metric regression, 106–107; Pearson correlation, 91–92; standardized, 107–110; testing equality of, 147–149; testing significance of individual, 107; Zeller's seemingly unrelated regression test of, 257; *See also* Regression coefficients
Cohen, G., 393
Cohen, J., 166, 184
Cohen, P., 166, 184
Collinearity reduction, 145
Computer software. *See* Statistical package programs
Conditional logistic regression, 361
Conditional mean imputation, 184
Constant Flux (Erikson and Goldthorpe), 304
Contingency transformations, 76
Contrast coding, 166, 170–171
Control variables; combining matching with statistical, 44; definition of, 12; examples of using, 12–15; experiments vs. statistical, 43–45; propensity score matching and, 44; weighted net percentage difference of, 16

Cook's Distance measure (Cook's D), 231–232, 237
Correlation analysis; Bayesian alternative for comparing models, 133–135; coefficience of determination (r^2), 91, 111–112, 308–309; for determinants of literacy in China, 113–120; dummy variables and, 120–123; independent validation of, 135–136; introduction to multiple, 104–105; metric regression coefficients for, 106–107; multicollinearity, 108; process and purpose of, 91–93; standard error of estimate (root MSE), 112–113; standardized coefficients, 107–110; strategy for comparisons across groups, 124–132; testing significance of individual coefficients, 107; variance of dichotomous variables and, 110; *See also* Variable relationships
Correlation ratios; acceptance of abortion scale, 101*t*; description of, 99–102; testing linearity using, 102
Correlations; factors affecting size of, 94–99*fig*; included in model of educational attainment by race, 176*t*; pairwise-present, 183; regression coefficients relationship to, 94
Covariance structure modeling, 395
Covariate adjustment, 29; *See also* Direct standardization
Cronbach's alpha, 244, 246*t*
Cross-tabulations; additive and interaction effects, 26–28*fig*; direct standardization, 28–43; logic of elaboration and, 22–25*fig*; statistical controls versus experiments and, 43–45; statistical package programs used to produce, 69–70; suppressor variables, 25–26*fig*; *See also* Data analysis
Crossings models, 285–286
Curran, P. J., 397
Cut points, 348
Czech Republic foreign-language competence; effect parameters for model of determinants on, 339*t*–340*t*; multinomial logit analysis of, 337–341; *See also Social Stratification in Eastern Europe after 1989* (Treiman and Szelényi)

D

Data; log-linear analysis with individual-level, 277; pooled from multiple surveys, 399–400; transformation of, 72–79; understanding the properties of your, 400–401; weighting, 212–214; *See also* Missing data
Data analysis; -do- files used in, 30, 31; -log- files used in, 30, 31; statistical package programs used for, 65–86, 294, 396; tips on using Stata commands in, 80–86; *See also specific technique*
Data files; illustration of organized, 67*e*; machine readable/computer readable, 67–68; organization of, 67–70
Data sets; card decks and card-image computer file, 70–72; codebook on organized, 68*e*; combining multiple years of GSS, 222–223*t*
Data transformation; analyzing surveys with missing data, 78–79; arithmetic transformations, 75–76; contingency transformations, 76; description of, 72; handling missing data during, 77–78; recoding for, 73–75
Daula, T., 355
Davis, J. A., 6, 294
de Graaf, P., 40, 258, 287
Deck ID, 72
Deff statistics (design effect), 207–209, 210*t*–211*t*, 217–219
Dehejia, R. H., 393
Deng, Z., 159, 387
Dependent variables; categorical, 335, 336–342; censored (or truncated), 335, 353–361; definition of, 11; effect-proportional scaling for metric, 257–258; multinomial logit analysis of, 336–342; ordinal, 5–6, 335, 342–353; ordinal logistic regression for, 342–353; other models for analysis of limited, 361; regression analysis of, 89*fig*–91; standardization of, 94; tobit regression for, 353–361; *See also* Variables
Descriptive statistics, 114
Design effect (deff); description of, 207–209, 217–219; for selected statistics, 210*t*–211*t*
Dichotomous variables; definition and examples of, 121; variance of, 110; *See also* Dummy (dichotomous) variables
Dickinson, P., 179
DiPrete, T. A., 388
Direct standardization; description of, 28–29; in earlier survey research, 31; *See also* Covariate adjustment
Direct standardization examples; belief that humans evolved from animals, 31–39*t*; level of literacy by urban vs. rural residence in China, 42*t*–43; occupational status by race in South Africa, 39–42; religiosity by militancy among urban Blacks, 29–31
Discrete-time hazard-rate models; binomial logistic regression of, 318–327; description of, 318; -do-files. *See* Stata -do- files
Documentation, 403–404
Domanski, H., 278
Downey, D. B., 104
Dummy (dichotomous) variables; computing discrete change for, 333; contrast coding of, 166, 170–171; definition and examples of, 121; effect coding of, 166, 167–170; mean imputation with, 183; other ways of representing, 166–172; parameterization of, 271; used within regression framework, 120–123; sequential effects of, 166, 171–172; *See also* Binomial logistic regression; Dichotomous variables
Duncan, B., 110
Duncan, G. J., 342, 361
Duncan, O. D., 112, 179, 294, 393, 394–395, 395, 396
Duncan, O. D. (father), 395

E
Earls, F., 388
Educational attainment; abortion acceptance by religious denomination and, 131*fig*; among women, effect on income of race and, 220–222; coefficient of determination (R^2) on, 111–112; coefficients of model on U.S. adult, 223*t*; curvilinear relationship between income and, 144*fig*; effect of cultural capital on Russian, 187–193*t*; examining standardized coefficients on, 109–110; expected income by, 145*fig*; expected probability of marrying by sex, age, and mother's, 326*fig*; factors affecting racial differences in, 174–179; joint probability distribution of age and, 36*t*; linear splines used to calculate U.S. trends in, 152–158*fig*; mean years of schooling by year of birth, 155*fig*; percent accepting abortion by religion and, 28*t*; percent by religiosity adjusting for, 30*t*; percent militant by, 14*t*–15; percent militant by religiosity and, 15*t*–16, 17–18*t*; percentage distribution of religiosity by, 13*t*–14; percentage of evolution beliefs by, 34*t*; percentage of legal abortion support by religiosity and, 27*t*; proportion of evolution beliefs standardized for age, 39*t*; relation between father's years of schooling and, 90*fig*; religious denomination and, 35*fig*; by size of place of residence at age fourteen and Chinese, 289*t*–292; as socio-economic status (SES) indicator, 304; three-year moving average years of schooling by years of birth, 155*fig*; voting by race, volunteer association, and, 278*t*; years of school completed by year of birth, 154*fig*; *See also* Years of schooling
Educational transition; binomial logistic regression on models for Japanese, 314–318; effect parameters for models of Japanese, 316*t*, 317; goodness of fit statistics for models of Japanese, 315*t*–317
Effect coding, 166, 167–170
Effect parameters; binomial logistic regression, 310*t*–314; Czech Republic foreign-language competence model, 339*t*–340*t*; derivation of, 295–297; Japanese educational transition, 316*t*, 317; log-linear analysis, 271–272; for OLS model political party identification, 353, 354*t*; for ordered logit model of political party identification, 345*t*–346*t*; probit analysis of gun threat, 331*t*–334
Effect-proportional scaling, 257–258
Eight Nation Survey of Political Attitudes (Barnes and Kaase), 97;
Eliason, S. R., 297, 300;
Ellison, C. G., 392
Eltinge, J. L., 207
Endogeneity, 388–389
Endogenous switching regression, 391–392
England, P., 380
Entwisle, B., 388
EQS (software), 396
Erikson, R., 283, 294, 384, 403
Errors-in-variables regression, 258–261
η^2 formula, 101
"Ethnographic present," 398
Ettner, S., 390
Evans, M.D.R., 187
Event history models, 318
Evolution beliefs; percentage by age, 34*t*; percentage by level of education, 34*t*; percentage by religion, age, and sex, 37*t*–38*t*; percentage by religious denomination, 33*t*; percentage distribution of, 31–32*t*; proportion standardized for education and age, 39*t*
Experiments; natural, 44, 386–387; statistical controls vs., 43–45
Exploratory factor analysis (factor-based scaling), 247–250

F
Factor scores, 250
Factor-based scaling (exploratory factor analysis), 247–250
Fair, R. C., 355
FE (fixed effects) models; allowing the slopes of the X's to vary, 366; analyzing more than two time points, 367–369; for binary outcomes, 375–380; for continuous variables, 365–370; description of, 364–365; determinants of income in China example of, 372–375; fixing effects across individuals rather than over time, 369; fundamental equation of, 365–366; interactions between time-constant and time-varying variables, 367; limitations and cautions regarding, 369–370; testing effects of time-invariant variables over time, 366–367
Featherman, D. L., 112, 394
"Fetal origins" hypothesis, 44, 385–386
Fienberg, S. E., 266, 294
Fischer, C. S., 403
Fisher, R. A., 125–126
Fitted marginals notation, 265
Fitzgerald, J. M., 400
Flu pandemic (1918), 385–386
Fodor, E., 40, 112, 179
Forristal, J. D., 388
Fox, J., 108, 166, 231
Frankenberg, E., 369, 380, 388
Freese, J., 314, 334, 342, 361
Frequency of sex; alternative estimates of model of, 358*t*; GSS (2000) data on, 356–357*t*; three estimates for U.S. married women, 359*fig*; tobit regression for, 356–361
Fu, V. K., 294

G
Gamoran, A., 392
Ganzeboom, H.B.G., 40, 188, 258, 287, 292

Gardeazabal, J., 387
Gaziano, C., 204
Gelman, A., 133
Gender equality attitudes, 150–152; *See also* Sex
Generalized ordered logit model, 349, 350*t*–352*t*
Gerber, T. P., 391
Geronimus, A. T., 380
Gilbert, G. N., 294
GLIM (software), 294
Glymour**, 389
Goldberger, A. S., 147, 353
Goldthorpe, J. H., 283, 294, 384, 403
Goodman, L. A., 265, 291, 294
Goodness of fit; criterion for, 91–92; for educational transition in Japan models, 315*t*–317; effect of Cultural Revolution on education, 160*t*; for log-linear models of anticommunist sentiment, 275*t*; model selection based on, 266–271; for various armed threats prevalence models, 308*t*–310
Gould, W. W., 156, 300, 361
Granger**, 389
GRE test-items, 245
Greenberg, D. F., 361
Greene, W. H., 355, 390
Greenwood, S. F., 399
Gruenewald, P. J., 355
Grusky, D. B., 294, 388
GSS (*General Social Survey*); card-image records of the, 72; combining data sets for multiple years, 222–223*t*; description and obtaining data sets from, 31; on educational attainment by Blacks (1990–2004), 174; factor analysis opportunities using the, 250; frequency of sex data in 2000, 356–357*t*; household data from, 204, 205; independent validation using, 135–136; making cross-temporal comparisons using the, 304; occupation/education effect on vocabulary data in 2004, 166–167; questions about acceptability of abortion in 2006, 100; SAMPCODE variable included by, 218–219, 220; sample incomes used in 2006, 96; stratified sample strategy used by, 212; subset of respondents excluded in, 150; survey estimation using the, 215; treatment of dummy variables using data from, 120–121; weighting procedure used by, 212; *See also* National Opinion Research Center (NORC)
Guilkey, D. K., 400
Gupta, S., 392
Guseva, A., 392

H

Hagan, J., 391
Halaby, C. N., 44, 380, 393
Haller, A. O., 395
Hamerle, A., 380
Harding, D., 44
Hardy, M. A., 166, 391
Hauser, R. M., 112, 133, 202, 283, 294, 395, 403
Hausman, J., 342, 371
Haynes, S. E., 342
Heckman correction procedure, 391
Heckman, J. J., 391, 403
Heckman selection model, 391
Henderson, G., 397
Hendrickx, J., 274, 295
Herrnstein, R. J., 403
High leverage points, 91
High School and Beyond study, 387–388
Hildebrand, D. K., 6
Hodge, R. W., 396
Hoffman, S. D., 342, 361
Hofmeyr, J. F., 392
Holding constant variables, 45
Holland, P. W., 266, 294, 389
Hollerith, H., 70
Hosmer, D. W., 304, 361
Hotdeck imputation, 185
Hotz, V. J., 380
Household random samples, 203–205
Hout, M., 288, 294, 403
Hsiao, C., 384
Huang, M.-H., 403
Hum, D.P.J., 181
Human Relations Area Files (HRAF), 399

I

IBM card data set, 70–72
IBM (International Business Machine Company), 70
IBM punch card, 71*fig*, 72
Ichino, A., 44, 393
IIA (independence of irrelevant alternatives), 341–342
Income differences; curvilinear relationship between education and, 144*fig*; curvilinear relationships of age and, 140–142; effect on education and race among women on, 220–222; relationship between 2003 age and, 141*fig*; semilog transformations of, 142–146; years of school and expected, 145*fig*
Independent (or predictor) variables; definition of, 11; suppressor effect of two or more, 21
Independent validation, 135–136
Influence statistics, 231–232
Institute for Social Research, 208
Instrumental variable (IV) regression, 389–390
Interaction effects; definition of, 26; examples of, 26–27*t*
International Socioeconomic Index of Occupations (ISEI), 40, 258, 288
Interval variables; definition of, 6; ordinal variables treated as, 6

Intervening variables; description of, 22, 24; examples of, 24*fig*–25*fig*

J

Jacobs, D., 342, 355
Jahoda, M., 22
Japanese educational transition; binomial logistic regression on models of, 314–318; effect parameters for models of, 316*t*, 317; goodness of fit statistics for models of, 315*t*–317
Jasso, G., 75
Johnes, G., 361
Jones, F. L., 173, 179
Jonsson, J. O., 342
Jöreskog, K. G., 396

K

Kaase, M., 97
Kahn, J. R., 75
Kalton, G., 181
Kaufman, R. L., 179, 272
Keating, N. L., 393
Keeley, M., 355
Kelley, J., 173, 179
Kelly, J. M., 400
Kelly, N. J., 400
Kenney, D. A., 98
Kim, J.-O., 109
King, G., 297
Kish, L., 204, 208, 209
Kish table, 204
Kitagawa, E., 175
Knoke, D., 263, 273, 277, 278, 294
Knook, D. L., 186, 190, 191
Korenman, S., 380
Kraus, V., 179
Krueger, A., 369, 380

L

L^2 (likelihood ratio χ^2); definition of, 267; goodness of fit testing using, 266–271; likelihood function of, 297–298; maximum likelihood estimation, 297–300, 302; *See also* Maximum likelihood estimation
Lab notebook record, 405
Lassen, D. D., 387
Lazarsfeld, P. F., 22, 26
Least squares criterion, 91
Least-squares regression lines, 92*fig*
Lee, H.-K., 146, 179
Lee, J. Z., 380
Lemeshow, S., 304, 361
Leverage points; description of, 94–96; effect of single deviant high, 95*fig*
Levin, S. G., 355
Lewin, A., 39
Li, Q., 42
Liang, Z., 361, 382
Lichter, D., 392
Likelihood function. *See* L^2 (likelihood ratio χ^2)
Linear prediction, 358
Linear splines; alternative specifications of functions, 156; demonstration that inclusion of linear term does not affect predicted values, 153*t*; description of, 152; U.S. trends in educational attainment example of, 152–158*fig*
Linear-by-linear association models, 287–288
LISREL (Linear Structural Relations), 396
Listwise deletion, 183
Literacy rates. *See* Chinese literacy rates
Little, R.J.A., 181, 184, 185
Log Multiplicative (or RC) model, 291–293
-log- files. *See* Stata -log- files
Log-linear analysis; of anticommunist sentiment distribution, 273–277; binomial logistic regression relationship to, 303–304; effect parameters testing, 271–272, 295–297; with individual-level data, 277; interaction vs. association terminology in, 264; introduction to, 264–265; parsimonious models and, 277–294; with polytomous variables, 277; software available for, 294–295; using Stata for, 272–273
Logic of elaboration; description of, 22; intervening variables, 24*fig*–25*fig*; spurious association, 22–23*fig*
Logistic regression, 302; *See also* Binomial logistic regression
Logit analysis, 304
Long, J. S., 297, 304, 314, 334, 342, 348, 355, 361, 392, 397
Lovell, P. A., 355
"Low R-squares," 112
Lu, B., 393
Lu, Y., 39, 104, 187, 377, 380, 387
Lucas, R. E., 392
Luijkx, R., 292
Lundquist, J. H., 393

M

McFadden, D., 341, 342
Machine-readable data set, 4
McIvery, J., 246
McKeever, M., 40, 112, 179
McLaughlin, D., 392
Mail surveys, 198–199
Male disability, 385–386*fig*
Manning, W. D., 392
Manski, C. F., 391, 392
MAR (missing at random) data, 182
Maralani, V., 104, 187
Mare, R. D., 157, 278, 294, 314, 342, 355, 391, 392
Marriage age. *See* Age at first marriage
Martin, B., 243

Marx, G., 2, 3, 6
Marx, K., 93
Mason, W. M., 204, 369, 380, 388
Maximum likelihood estimation, 297–300, 302
MCAR (missing completely at random) data, 182, 184
Mean substitution, 184–185
Means; decomposing the difference between two, 172–179; decomposition of difference in years of school by race, 178*t*; expressing coefficients as deviations from grand, 164–166; included in model of educational attainment by race, 176*t*; ISEI by level of education, 258, 259*t*
Meff statistics, 208, 217
Metric regression coefficients, 106–107
Militancy; control variables for, 12; decision regarding search for additional variables on, 16–17; direction to percentage the table on, 11–12; distribution by religiosity among urban Negroes, 8*t*–9; Marx's construction of, 6–7; percent by religiosity among urban Negroes, 10*t*–11; percent militant by educational attainment, 14*t*–15; percent militant by religiosity and educational attainment, 15*t*–16, 17–18*t*; percentage distribution of religiosity by educational attainment, 13*t*–14; *See also* "Religion: opiate or inspiration of civil rights militancy among Negroes?" (Marx)
Miller, G., 402
Mincer, J., 142
Missing data; analyzing surveys with, 78–79; data transformation and, 77–78; defining, 75, 77; MAR (missing at random) type of, 182; MCAR (missing completely at random) type of, 182, 184; MNAR (missing not at random) type of, 182; NI (nonignorable) type of, 182; outliers and, 75; problem of, 182; *See also* Data
Missing data strategies; casewise deletion, 183; full Bayesian multiple imputation, 182, 185–186; hotdeck imputation, 185; mean substitution, 184–185; weighted casewise deletion, 183–184
Missing values, 79
Mixed models, 361
MNAR (missing not at random) data, 182
Mobility effects, 147
Model selection; anticommunist sentiment example of, 272–273; effect parameters factor in, 271–277, 295–297; goodness of fit basis for, 266–271; theory-based, 271–277
Models; acceptance of abortion, 127*t*–131*fig*; armed threats prevalence, 304–314; baseline, 268; Bayesian Information Criterion (BIC) for comparing, 133–135, 151, 270–271, 286; coefficients of educational attainment of adults, 223*t*; conditional logistic regression, 361; description and components of, 7; determinants of Chinese characters identified on ten-item test, 116*t*; determinants of estimated political conservatism, 260*t*–261; determinants of income of U.S. women, 221*t*–222; determinants of strength of occupation-education connection, 236*t*– 237; discrete-time hazard-rate, 318–327; estimating cultural capital effect on education attainment in Russia, 187–193*t*; expected percentage from anticommunist sentiment, 276*t*; generalized ordered logit model, 349, 350*t*–351*t*; goodness-of-fit statistics for log-linear anticommunist sentiments, 275*t*; Heckman selection, 391; L^2 (likelihood ratio χ^2) of, 266–271, 297–300; linear splines, 152–158*fig*; mixed, 361; nested logistic regression, 361; parsimonious, 277–294; poisson regression, 361; predicting abortion acceptance, 256*t*; predicting knowledge of Chinese characters by year of birth, 161*t*; probit analysis, 302, 330–334, 361; saturated, 130, 265–266; SEM (structural equation modeling), 393–397; of technology, political integration, stratification level relationships, 269*t*–271; tobit regression, 355–361; tolerance of atheists coefficients, 165*t*–166; vocabulary knowledge, 169*t*–170*t*
Monette, G., 108
Morgan, S. L., 389, 393
Moving average, 325
MPlus (software), 396
Mueller, C. W., 109
Müller, W., 235
Mullin, C. H., 380
Multicollinearity, 108
Multilevel analysis importance, 387–388
Multinomial logit analysis; description of, 336–337; foreign-language competence in Czech Republic example of, 337–341; IIA (independence of irrelevant alternatives) in, 341–342
Multiple classification analysis, 164–166
Multiple imputation; Bayesian method of, 182, 185–186; comparing casewise deletion and, 191–193*t*
Multiple-item scale construction; additive scaling, 247; factor-based scaling, 247–250; from incomplete information, 249
Multiple-item scales; construction methods for creating, 246–258; description and function of, 242; effect-proportional scaling, 257–258; errors-in-variables regression and, 258–261; Marx's questionnaire constructing, 6–7; realiabilty through, 243–245; religiosity and abortion attitudes example of, 250–257; validity of measures through, 242–243; values of Cronbach's alpha for, 244, 246*t*; Zeller's seemingly unrelated regression procedure for, 257
Multiple-punched columns, 71
Multiple-regression equations; Bayesian Information Criterion *(BIC)* for comparing models, 133–135, 151; coefficient of determination (R^2), 111–112; for comparisons across groups, 124–131*fig*; description

Multiple-regression equations (*continued*) of, 105; for determinants of literacy in China, 113–120*fig*; dummy variables and, 114, 120–123; independent validation of, 135–136; metric regression coefficients, 106–107; multicollinearity, 108; reexpressing variables as deviations from the means, 131–132; standard error of estimate (root MSE), 112–113; standardized coefficients, 107–110; testing additional hypotheses by constraining coefficients to zero or to equality, 132; testing significance of individual coefficients, 107; three-dimensional representation of, 105*fig*
Multiple-regression techniques; decomposing the difference between two means, 172–179; expressing coefficients as deviations from grand mean (multiple classification analysis), 164–166; linear splines, 152–164; nonlinear transformations, 140–147; other ways of representing dummy variables, 166–172; testing assumption of linearity, 149–152; testing equality of coefficients, 147–149
Multistage probability samples; description of, 196, 197–202; GSS (1994) use of, 220
Murdock, G. P., 268, 398–399
Murray, C., 403
Myrdal, G., 398

N

National Institutes of Health (NIH), 401
National Opinion Research Center (NORC), 401; *See also* GSS (*General Social Survey*)
National Science Foundation (NSF), 401
Natural experiments, 44, 386–387
Natural log, 144
Nee, V., 385
Nested logistic regression, 361
Net regression coefficients, 106
Netemeyer, R. G., 246
Newton, P. J., 393
NI (nonignorable) missing data, 182
Nobles, J., 380
Nomenklatura memberships (Russia); binomial logistic regression analysis of, 327–329; coefficients for model of determinants of, 328*t*
Nominal variables, 4
Nonlinear transformations; curvilinear relationships: age and income, 140–142; description of, 140; mobility effects, 147; semilog transformations: income, 142–146
Nonresponse sources, 207
Nord, R., 355
Nordholt, E. S., 181
Notes on Social Measurement (Duncan), 395
Nuisance effects, 214
Nunnally, J., 246

O

Oaxaca decomposition, 175
Oaxaca, R., 175
O'Brien, R. M., 355
Occupational status; frequency distribution by father's, 280*t*; goodness-of-fit statistics for models of intergenerational, 284*t*; by race in South Africa, 39–42
Odds ratios; antilogs of the *b*s interpreted as contributions to, 311; constructing, 348; for predicting likelihood of age at first marriage, 323*t*–324*t*; τs defined as functions of, 264
OLS (ordinary least-squares) regression; comparing FE estimates for Chinese family income determinants with, 374*t*; description of, 90; to determine best-fitting line, 91; errors-in-variables regression extension of, 258–261; limitations of, 370; linear relationship between two variables using, 92*fig*; misleading results of, 319; ordinal logistic regression alternative to, 353; regression diagnostics of underlying assumptions of, 225–240; relation between years of schooling and father's years of school, 90*fig*
Omitted variable bias; description of, 106, 364–365; fixed effects (FE) modeling for binary outcomes, 375–376, 377–380; fixed effects (FE) modeling for coping with (continuous outcomes), 364–370, 372–375; handling continuous variables of income in China, 372–375; random effects (RE) modeling for coping with (binary outcomes), 377–380; random effects (RE) modeling for coping with (continuous outcomes), 365, 370–375
Open-ended questions, 3
Ordinal dependent variables; definition of, 5–6, 335; ordinal logistic regression for, 342–353; treated as interval variables, 6
Ordinal logistic regression; comparisons to other estimating procedures, 348–353; constructing odds ratios, 348; converting logits to *Y**-standardized form, 347; description of, 342–343; getting predicted percentages, 348; OLS model as alternative to, 353; political party identification example of, 343–353
Ordinal logit equation, 342
Oster, E., 387
Outliers; description and effect of, 94–96; missing values and, 75; regression diagnostics and, 231

P

Pairwise-present correlations, 183
Park, H., 294
Parker, P., 391
Parsimonious models; crossings, 285–286; extensions of, 293–294; linear-by-linear association, 287–288; log-linear analysis of, 277–281; quasi-independence, 283–284*t*; quasi-symmetry, 285; row-effects (and

column-effects), 288–293; topological or levels, 281–283; uniform association, 286–287
Partial-regression leverage plots, 232
Partial-regression plots, 232
"Passively imputed" variables, 191
Path analysis, 394–395
Paul, C., 181, 183, 185
Pearson correlation coefficient, 91–92
Pearson, K., 93
Pearsons's *r*; computational formula for, 93; origins of, 93; *See also* R^2 (coefficient of determination)
Petersen, T., 334
Peterson, R., 391
Pierce, G. L., 278
Pisati, M., 294, 295
Poisson regression, 361
Polachek, S., 142
Political conservatism estimates, 260*t*–261
Political party identification; effect parameters for OLS model of, 353, 354*t*; effect parameters for ordered logit model of, 345*t*–346*t*; generalized ordered logit model of, 349, 350*t*–352*t*; ordinal logistic regression of, 343–353
Population Association of America, 202
Population register samples, 203
Population Working Paper archive, 404
Population/population subgroup issues, 383–386
Portes, A., 394
Powers, D. A., 263, 285, 294, 297, 300, 304, 334, 342, 361, 392
Powers, E. T., 400
PPS (probability proportional to size) sampling, 200
predictor variables. *See* independent (or predictor) variables
Principal factor analysis with iterations, 248
Probability samples; case-control sampling, 327; design effects of, 207–209; downweighting sample size in Stata, 219; household, 203–205; importance of, 397–400; multistage, 196, 197–202; population register, 203; quota, 206; random walk, 206; simple random samples, 196–197; stratified, 196, 206–207; stratifying to offset effect of clustering, 209–223*t*; superior sampling procedure for, 205–206; weighting, 212–214
Probit analysis; effect parameters for gun threat, 331*t*–334; introduction to, 330–334*fig*; as logistic regression alternative, 302, 361; marginal effect of, 331–333; probabilities associated with values of coefficients of, 334*fig*; Stata commands used to create model for, 334
Professional practices; conducting sensitivity analysis, 402–403; documenting your work, 403–404; doing a last check for errors, 404–405; exploring alternatives to your a priori hypotheses, 402; understanding the properties of your data, 400–401; *See also* Research design issues
Progression ratios, 314
Propensity score matching, description of, 44, 392–393
Proportional odds assumption, 348–349
Prouteau, L., 392
Provost, C., 268, 398, 399
Pseudo R^2, 308–309
Pseudo-log likelihood, 308
PSUs (primary sampling units), 200

Q

Quality of Employment Survey, 403
Quasi-independence models, 283–284*t*
Quasi-symmetry models, 285
Quester, A., 355
Questionnaires; Marx's construction of militancy concept using, 6–7; nonresponse to, 207; open-ended questions used in, 3; pooling data from multiple, 399–400; probability sampling for, 35, 196–224; sources of nonresponse in, 207; *See also* Survey estimation procedures; Surveys
Quota samples, 206

R

r; computational formula for, 93; formula for calculating, 91; origins of, 93
R^2 (coefficient of determination); formula for computing from correlations, 111
Racial differences; armed threats prevalence and, 304–314; factors affecting education attainment by, 174–179; South African occupational groups of males by, 41*t*; South African occupational status by, 39–42; tips on studying, 175; voting by education, volunteer association, and, 278*t*; women's income by education and, 220–222; *See also* African Americans; South African Blacks
Radelet, M. L., 278
Raftery, A. E., 133, 317
Ragin, C., 355
Random walk samples, 206
Rasler, K., 361
Ratio variables, 6
Raudenbush, S. W., 229, 387, 388
Raw (or metric) regression coefficients, 106–107
Ray, S. C., 182
RC (or Log Multiplicative) model, 291–293
RE (random effects) models; for binary outcomes, 377–380; for continuous variables, 370–372; determinants of income in China example of, 372–375
Recoding variables, 73–75
Reducing excessive collinearity, 145

Regression coefficients; Bayesian alternative for comparing models and, 133–135; constraining coefficients to zero or to equality, 132; correlation relationship to, 94; dichotomous (dummy) variables used in, 110, 120–123; factors affecting the size of, 94–99*fig*; independent validation of, 135–136; metric, 106–107; multicollinearity and, 108; multiple-regression equations, 105*fig*; standardized, 107–110; strategy for comparisons across groups and, 124–132; testing significance of difference between, 129; testing significance of individual, 107; *See also* Coefficients
Regression diagnostics; bootstrapping and standard errors, 238–240; definition of, 227; four scatter plot illustrating problems requiring, 225–228*fig*; robust regression estimates, 237*fig*–238; societal differences in status attainment example of, 228–237
"Regression toward the mean," 97–98
Reliabiity; definition of, 243; multiple-item scales to enhance, 243–245; ways to assess, 244
"Religion: opiate or inspiration of civil rights militancy among Negroes?" (Marx), 2–4; *See also* Militancy
Religiosity; age association with, 17; correlating militancy relationship to, 10*t*–11, 13*t*–14, 15*t*–16, 17–18*t*; by militancy among urban Blacks, 29–31; multi-item scale on abortion acceptance and, 250–257; percent accepting abortion by education and, 28*t*; percentage of legal abortion support by education and, 27*t*
Religious denomination; abortion acceptance by education and, 131*fig*; age differences and, 35*t*; educational attainment and, 35*fig*; evolution beliefs by, 33*t*; evolution beliefs by age, sex, and, 37*t*–38*t*
Relles, D. A., 182
Research design issues; comparisons, 382–388; endogeneity, 388–389; endogenous switching regression, 391–392; Heckman selection model, 391; instrumental variables regression, 389–390; multilevel analysis, 387–388; natural experiments, 386–387; population subgroups, populations, and historical periods, 382–386; propensity score matching, 392–393; sample-selection bias, 390; structural equation modeling (SEM), 393–397
Residual-versus-fitted plots, 233–234*fig*
Response category, 4
Ribar, D. C., 392, 400
Roberts, J. M., Jr., 278
Robust regression, 237*fig*–238
Rona-Tas, A., 392
Roncek, D. W., 355
Ronning, G., 380
Roos, P. A., 179, 383, 403
Root MSE (standard error of estimate), 112–113
Rosen, H., 355
Rosen, S., 392
Rosenbaum, P. R., 44, 392
Ross, H. L., 44
Rossi, A., 27
Rossi, P., 401
Rotated factor matrix, 248
Row-and-column-effects models; log-linear analysis of, 288–290; Row-and-Column-Effects Model I version of, 290–291; Row-and-Column-Effects Model II (RC or Log-Multiplicative) version of, 291–293
Royston, P., 185, 186
Rubin, D. B., 44, 133, 182, 185, 389, 392
Rubin's rules, 185
Russian educational attainment, 187–193*t*
Russian *nomenklatura* membership; binomial logistic regression analysis of, 327–329; coefficients for model of determinants of, 328*t*

S

Sakamoto, A., 392
Saltzman, G. M., 355
SAMPCODE variable, 218–219, 220
Sample-selection bias, 390
Samples. *See* Probability samples
Sampling frame, 202
Sampson, R. J., 388
SAS; data file organization in, 67–70; description of, 66–67
SAT test-items, 245
Saturated models; description of, 130, 265–266; interaction parameters for, 282*t*
Sayrs, L. W., 384
Scale construction. *See* Multiple-item scale construction
Schafer, J. L., 185
Schenker, N., 185
Schervish, P. G., 272
Scholz, J. K., 380
Schwartz, C. R., 278
Scribney, W., 300
SEM (structural equation modeling), 393–397
Semilog transformations, 142–146
Sensitivity analysis, 402–403
Sequential effects, 166, 171–172
Sewell, W. H., 395
Sex; evolution beliefs by age, religion, and, 37*t*–38*t*; expected probability of marrying by age, mother's education, and, 326*fig*; frequency distribution of graduate course program by, 265*t*; frequency of sex by marital status and, 360*fig*–361; South African occupational groups by race and, 41*t*; *See also* Gender equality attitudes
Sex (frequency of); alternative estimates of model of, 358*t*; expected by gender and marital status,

360*fig*–361; GSS (2000) data on, 356–357*t*; three estimates for U.S. married women, 359*fig*; tobit regression for, 356–361
Sharma, S., 246
Shavit, Y., 234
Shields, M. P., 342
Simple random samples, 196–197
Skaggs, S., 342
Sloan, J. H., 44
Smith, D. A., 355
Smith, H. L., 44, 392
Smith, J. A., 393
Smith, P. L., 156
Smith, T. W., 205
Smits, J., 294
Smock, P. J., 392
Smoothing distributions, 315
Smple selection bias, 182
Snyder, J. M., Jr., 387
Sobel, M. E., 294
Social Sciences Research Council, 399
Social Stratification in Eastern Europe after 1989 (Treiman and Szelényi), 327, 337; *See also* Czech Republic foreign-language competence
Socio-economic status (SES); of Chinese adults by size of place of residence, 373*t*; creating a scale for, 247; education as indicator of, 304
Sociological Abstracts (journal), 295
Sociological Methodology (journal), 133
Sociological Methods and Research (journal), 133
Software. *See* Statistical package programs
Soopramanien, D., 361
Sorokin, P. A., 149
Sousa-Poza, A., 392
South African Blacks; binary outcome of effect of migration on school enrollment among, 377–380; occupational status by race, 39–42; OLS and FE estimates for effect of migration on school enrollment, 379*t*; *See also* Racial differences
South, S. J., 361
Spearman rank order correlation, 6
Specification error, 106
Specifications, 26
SPSS; data file organization in, 67–70; origins and development of, 66
SPSS-X, 66
Spurious associations, 22–23*fig*
Sribney, W. M., 207
Standard deviations; calculating, 94; included in model of educational attainment by race, 176*t*
Standard error of estimate (root MSE), 112–113
Standard errors, 238–240
Standardized coefficients, 107–110
Standardized variables, 94
Stanley, J. C., 43
"Started logs," 143
Stata; commands to create probit model in, 334; for comparing goodness-of-fit of regression models, 125; converting household samples to person samples in, 205; cross-tabulation using, 69–70; data file organization in, 67–70; decomposition in, 175; description of, 67; documenting your work using, 403–404; dummy variable parameterization in, 271; estimating generalized ordered logit models with, 349; frequency distributions command in, 401; getting *p*-values via, 125; getting predicted percentages, 348; Heckman selection model command, 391; how to downweight sample size in, 219; incorporating side computations using, 405; IV estimation command in, 390; log-linear analysis using, 272–273; "moving average" function in, 325; multiple-imputation procedures written for, 186; particularly useful commands in, 84–86; survey estimation procedures in, 207, 214–215, 216–217; time-series data features in, 384; tips on doing analysis using, 80–84; weighting data in, 213; *See also* Statistical package programs
Stata -do- files; binomial logistic regression analysis using, 319–320; calculating mean difference using, 42; definition of, 30; on evolution beliefs, 31–32t; on occupational status by race in South Africa, 39–42; tips on using, 80–84
Stata -log- files; definition of, 30; on evolution beliefs, 31–32*t*; on occupational status by race in South Africa, 39–42
Statistical controls vs. experiments, 43–45
Statistical package programs; AMOS (software), 396; command syntax differences of, 66; description of, 66; EQS, 396; GLIM, 294; historical development of, 66–67; how data files are organized using, 67*e*–72; LISREL (Linear Structural Relations), 396; MPlus, 396; transforming data using, 72–79; *See also* Stata
Status attainment regression diagnostics case; added-variable plots, 232–234; background information on, 229–230; leverage, outliers, and influence statistics used in, 231–232; plots for assessing influence, 232*fig*, 233*fig*; preliminaries of, 230; residual-versus-fitted plots and formal tests for patterns in the data, 233–237
Steele, C., 98
Steiger, J., 396
Stephan, P. E., 355
Stewart, C., III, 387
Stoltzenberg, R., 147, 182
Stone, R. A., 393
Stouffer, S. A., 6
Stratification levels; Blau and Duncan's basic model of, 393–394*fig*; distribution by political

integration/technology level, 268*t*; models of relationship between technology, political integration, and, 269*t*–271; political integration, level of technology, and expected, 272*t*
Stratified probability samples; construction of Chinese, 212; description of, 196, 206–207
Stratifying samples, 209–212
Studentized residual, 231
Sudman, S., 197
Superpopulation concept, 267, 398–400
Suppressor effect, 21
Suppressor variables, 25–26*fig*
Survey of Economic Opportunity and Achievement in South Africa (Treiman, Lewin, and Lu), 39
Survey estimation procedures; alternative to, 219; on Chinese literacy rates, 215–219; Stata, 207, 214–215, 216–217; *See also* Questionnaires
Survey Sampling (Kish), 208
Surveys; mail, 199; multistage probability samples, 196, 197–202; simple random samples for, 196–197; stratified probability samples, 196, 202–207; telephone, 198; web, 199; *See also* Questionnaires
Szelényi, I., 327, 337

T
Tavits, M., 400
Telephone surveys, 198
Terrell, K., 257
Theory; importance of using, 45; model selection and role of, 271–277
Theory-based model selection; description of, 271; effect parameters testing as part of, 271–272, 295–297
Thomas, D., 388
Tienda, M., 392
Time-series procedures, 384
Tobin, J., 353, 354
Tobit model; description of, 355–356; frequency of sex example of, 356–361
Tolerance of atheists coefficients, 165*t*
Tomaskovic-Devey**, 342
"Too many degrees of freedom" problem, 385
Topological (or levels) models, 281–283
Treiman, D. J., 39, 40, 42, 112, 113, 146, 147, 158, 159, 179, 185, 187, 188, 202, 204, 208, 229, 234, 235, 257, 258, 287, 292, 314, 327, 328, 337, 377, 380, 383, 385, 387, 388, 390, 402
Trend analysis, 149–152
Treno, A. J., 355
Trimming the regression equation, 107
Trivedi, P. K., 361
Truncated variables. *See* Censored (or truncated) dependent variables
Truncation; correlations reduced by, 97*fig*; description of, 96–97; "real data" example of effect of, 97
Tsai, S.-L., 395
Tukey, J. W., 143

U
Udry, J. R., 75
Uniform association models, 286–287
Upton, G.J.G., 294
U.S. Census Bureau, 184, 185, 202
U.S. Census pretest (1970), 78
U.S. panel studies, 368–369

V
Validity; balancing sampling with, 35; definition of, 242; multiple-item scales to measure, 242–243
Value labels, 69
Van Buuren, S., 186, 190, 191
Variable labels, 69
Variable names, 69
Variable relationships; additive and interaction effects, 26–28; between standardized and unstandardized, 94; correlation ratios of, 99–102; direct standardization for establishing, 28–43; linear, 92*fig*; logic of elaboration, 22–25; regression analysis of, 89*fig*–91; spurious associations, 22–23*fig*; suppressor effects, 21; zero-order association, 24, 26*fig*; *See also* Correlation analysis
Variables; categorical, 335; control, 12–15, 16, 43–45; definition of, 4; descriptive statistics for, 114; dichotomous, 110, 121; dummy, 120–123, 166–172, 183, 271; holding constant, 45; interval, 6; intervening, 22, 24*fig*–25*fig*; nominal, 4; ordinal, 5–6, 335, 342–353; "passively imputed," 191; ratio, 6; recoding, 73–75; SAMPCODE, 218–219, 220; standardizing, 94; "started logs," 143; suppressor, 25–26*fig*; truncated (or censored) dependent, 335; unstandardized, 94; variance of dichotomous, 110; *See also* Dependent variables
Varimax rotation, 248
Vermunt, J., 294
Vocabulary knowledge; coefficients for model of determinants of, 169*t*–170*t*; contrast coding used for, 166, 170–171; design matrices for coding categorical variables of, 168*t*; effect coding used for, 166, 167–170; GSS (2004) data on, 166–167; sequential coefficients used for, 166, 171–172
von Bortkewitsch**, 361
Voting determinants, 278*t*

W
Wahba, S., 393
Wald tests, 309–310

Walder, A. G., 42, 385
Walton, A. G., 355
Weakliem, D. L., 135
Web surveys, 199
Weighted casewise deletion, 183–184
Weighting data, 212–214
Weitoff, G. R., 361
Weldon, W.F.R., 93
White, M. J., 361
Willis, R. J., 392
Wilson, F. F., 392
Winsborough, H. H., 179
Winship, C., 135, 342, 389, 391, 392, 393
Wise, D. A., 391
Witte, A., 355
Wolff, F.-C., 392
Wood, C. H., 355
Wooldridge, J. M., 334, 363, 370, 380, 384, 390
Wright, E. O., 243
Wright, S., 396
Wu, X., 287, 292, 390

X

Xiao, M., 380
Xie, Y., 285, 294, 263, 297, 300, 304, 334, 342, 361

Y

Yamaguchi, K., 294, 314; Yanovitzky, I., 361; Years of schooling. *See* Educational attainment
Yip, K.-B., 229, 234, 235, 388

Z

Zatz, M. S., 391
Zeisel, H., 22, 26, 27
Zellner, R. A., 243
Zellner's seemingly unrelated regression procedure, 257
Zero-order association; definition of, 24; suppressor variable effect on, 26*fig*
Zhang, J., 342
Zone punches, 71