

## Index

- 21 word solution, Simmons, Nelson, and Simonsohn, 237
- Achinstein, P., 68–9, 296–7, 369n10
- ad hoc hypothesis, 90, 107
- adequacy for a problem, 297
- adjustments for selection, 19, 106, 276
- appeal procedure in, 279
- when not needed, 281
- Adler, A., 76, 96
- affirming the consequent
- invalid, 63
- probabilistic, 62–3, 67
- severe tester rejects, 71
- Akaike, H., 318
- information criteria (AIC), 318–19
- score (AIC), 318n5
- all flesh is grass, 222–3, 296
- Altman, D., 264
- American Statistical Association
- Six Principles, 215n4
- statement on *P*-values (ASA 2016 Guide), 17, 215–16, 272, 395
- ancillary statistic, 200
- Anderson, D., 159, 162, 318–19
- apophenia, 19, 301
- approximate vs. probable model, 296–7
- argument, 60
- convergent, 15, 107, 320
- linked vs. convergent, 15
- soundness/unsoundness, 61
- validity/invalidity, 60
- argument from error and coincidence, 14–16, 85, 210, 285
- and stringent tests, 237
- argument from intentions, *see* intentions, argument from
- Armitage, P., 46–50, 303, 398, 430
- array of models, 86–7
- in 1919 eclipse experiments, 121–2
- hierarchy, 109, 237, 300
- auditing, 20, 122, 267
- an inquiry (three questions), 122
- of *P*-value, 43, 94, 154
- tasks, 269, 308
- autoregressive model, 316
- Bacchus, F., 415
- background information
- and auditing, 127
- and causal attribution, 233–5
- Cox on, 233
- in FIRST interpretations and Cox's taxonomy, 150–8
- use of in frequentist inference, 86, 94, 150, 158, 317
- Gelman on Cox, 234
- bad evidence, no test (BENT), 5–6, 20, 78, 149, 201, 221, 223, 232, 235, 247, 277, 301, 362
- Baggerly, K., 6, 18
- Bailer, J., 281n7
- Bakan, D., 239–40
- Baker, M., 286
- Banerjee, A., 290–2
- Barnard, G., 13, 31, 46, 48, 126–7, 139, 156n8, 181–2, 324, 388, 420–1
- Bartlett, M., 47, 389
- batch effects, 293–4, 361
- Bayarri, M., 175, 179, 183, 305, 338, 434
- Bayes B-boost (incremental), 63, 66–73, 95
- vs. absolute, 67; *see also* affirming the consequent (probabilistic)
- Bayes factor, 37, 184, 253–5, 261–4, 305, 320, 335–6
- aka post-experimental rejection ratio, 338
- Bayes' Theorem, 24, 61
- and the Likelihood Principle, 45–6, 398
- Bayes/Fisher disagreement, *see* Jeffreys–Lindley paradox
- Bayesian foundations
- Bayes/frequentist, 24–6
- classical subjective Bayesians and criticisms, 397–400
- current state of play, 23, 395–7, 400

## 472 Index

- Bayesian foundations (cont.)  
 Dutch book, 415  
 error statistical basis, 27  
 need new, 432  
 and wash-out theorems, 231–2
- Bayesian incoherence and violating Bayes' Rule, 411, 415, 419, 421  
 betting incoherency, 417  
 empirical studies of, 422  
 replying to axiomatic proofs, 421  
 temporal, 423
- Bayesian inference  
 as conditional on specific data, 51, 183, 188, 431  
 role of probability flipped, 207, 405–7
- Bayesian model checking, 27–8, 304–5, 432–6
- Bayesian priors  
 and background information, 413  
 changing priors, 417–19  
 data generating vs. state of knowledge, 403  
 vs. degree of plausibility/belief, 406  
 default/non-subjective, 25, 400  
 empirical (frequentist) Bayes, 185, 404  
 gallimaufry, 402–3  
 hierarchy of, 433  
 and objectivity, 232–3  
 subjective elicitations, 410–11
- Bayesian probabilists  
 casual, 412  
 classical subjective, and criticisms of, 228, 397–400  
 default/non-subjective Bayesians, 25, 51, 64, 230, 262, 417  
 vs. error statistical/falsificationist Bayesian, 432–5  
 family feuds, 25, 409  
 grace and amen, 413  
 ironic/bad faith, 412  
 pragmatic, 424–6
- Bayesian statistics  
 betting, 257, 397, 405–8, 415, 421  
 Bayes hypothesis test, 183, 261  
 and known data, 417  
 notions of probability, 396, 402
- Bayesian vs. frequentist, 24–6  
 empathy with frequentists, 401  
 orthogonal, 400  
 unifications, 25–8, 183–7, 396, 409–15
- Beall, G., 299  
 begging the question, 61–2, 184, 247, 399n2, 422
- Bem, D., 283–4
- Benjamin, D., 264, 266, 337–8
- Benjamini, Y., 277–8, 362n7
- Berger, J., 23, 26, 49, 54, 74, 144, 175–6, 183–7, 199, 230–1, 248–52, 258–60, 305, 337–8, 400, 409–15, 417–18, 426, 430–1, 434
- Berger, R., 162n11, 248, 251–2, 259, 264, 303–4
- Bernardo, J., 198, 255–6, 259, 400, 402–3, 411, 431n5
- Bernoulli model, 17
- Bernoulli trials, 31, 111, 147, 186, 254, 275, 298, 303
- Berry, S., 287, 289
- Bertrand–Borel debate, 372–82
- Bertrand, J., 372–4
- Best Bet and Breakfast, 350
- Beyond the Standard Model (BSM) physics, 212–14
- bias, 20–1
- biasing selection effects, *see* selection effects, biasing
- Big Data/machine learning, 123, 229, 267, 294, 395, 397  
 and experimental design, 293  
 and positivism, 229
- big picture inference, 127, 226, 279, 423
- Binomial distribution, 111–14, 139, 147, 285, 298, 320, 386
- Binomial vs. negative Binomial, 303, 411n6
- Birnbaum, A., 38, 53, 55, 170, 172–3
- Black, L., 294–295
- Bogen, J., 121
- Bonferroni correction, 275, 294  
 and false discovery rates, 277–8
- Boole, G., 387
- bootstrap resampling, 305–7
- Borel, É., 131, 372–5, 377
- Bowley, A., 189–90, 387
- Box, G., 27, 297, 301–4, 313, 405n5, 421, 433
- Box, J. (Fisher), 120
- Braithwaite, R., 65
- Brans–Dicke theory, 161  
 adjustable parameters, 74
- Breiman, L., 414
- Bristol-Roach, M., 16–18, 31
- Brown, E., 7–8, 381, 427

- Buchen, L., 124n2  
 Buehler, R., 390n8  
 Burgman, M., 358  
 Burnham, K., 159, 162, 318–19  
 Byrd, J., 358
- Calibration (lounge), 350  
 calibration, 54, 396  
   and pragmatic Bayesians, 305, 424; *see also*  
   error probabilities  
 cancer clusters, power lines, 157–8  
 capability  
   and severity, 189, 193–5  
   and confidence intervals 191–2  
   of methods 300  
 capitalizing on chance, 274–5  
 Carlin, B., 43  
 Carlin, J., 264, 289, 359–61  
 Carnap, R., 59, 63–4, 67, 108, 110–11, 114, 147,  
   341, 400, 416  
 Cartlidge, E., 213  
 Cartwright, N., 291–2, 294  
 Casella, G., 248, 251–2, 259, 264, 303–4  
 Castelvocchi, D., 214  
 catchall, hypothesis ( $\sim H$ ) and factor ( $\Pr(x|\sim H)$ ),  
   68, 84, 213, 302  
   and open-endedness, 399, 420–1, 443  
 Central Limit Theorem (CLT), 186n7, 298  
   and water plant accident, 142  
   and relative frequencies, 299  
 CERN, 202  
 Center for Open Science, 97  
 Chalmers, A., 129  
 Chalmers, T., 339n4  
 Chandra Sekar, C., 269, 376  
 Charpentier, E., 229  
 Cherkassky, V., 79  
 chestnuts and howlers  
   of CIs, 198–200  
   Jeffreys' tail area, 168–70  
   and large  $n$ , 242–3  
   of power, 325  
   of selection effects, 276–7, 282  
   of tests, 165–72  
 chi square, 139  
 chutzpah, 12, 221  
 clinically relevant/irrelevant difference, 326–7  
 Cobb, A., 235–6  
 Cochrane collaboration, 292  
 Cohen, J., 323–4, 338–9, 340, 356  
 Colquhoun, D., 277n4, 289, 365–6  
 comparativism/comparative accounts, 13, 30,  
   33, 36, 82, 261, 332, 334–6, 441  
   and falsification, 82, 318  
   vs. significance test, 243, 268, 284; *see also*  
   Likelihood, Law of  
 COMpare Team, 40  
 conditional (if–then) claim, 150, 167  
 conditional probability, 24, 35, 37, 46, 405  
   vs. an error probability, 205  
   illicit transposing, 208, 331, 363  
 conditioning (D. R. Cox)  
   to achieve relevance, 173, 200  
   to separate from nuisance parameters, 385  
 confidence concept (Conf)  
   Birnbaum, 55, 170  
   Mayo extension, 55  
 Confidence Court Inn, 350  
 confidence distribution, 195, 358, 442  
   as fiducial, 382, 391  
   and unifying Bayesian, fiducial, frequentist  
   (BFF), 391, 435n8  
 confidence intervals (CIs), 18, 153, 189, 356,  
   427–8  
   and Bayes–Frequentist agreement, 425  
   duality with tests, 190–3, 198, 244, 357, 442  
   and fiducial, 390  
   generic vs. specific (particular), 191, 246  
   in the Higgs, 211  
   performance construal, 244, 384  
   reforming, 193, 244–6, 358  
   severity interpretation, 198, 346, 429, 442  
   vacuous (chestnut), 198–9  
   warranted by severity analysis, 358  
 confirmation theory, 59  
   Carnapian, 59, 63–4  
   C-function ( $c^\dagger$ ), 63  
   incremental vs. absolute, 66–7  
   inductive intuitions, 64  
   as fit measures, 72–3  
   logical probabilities, 63  
   paradoxes of, 73  
   positive instance, 60  
   Popper against, 68; *see also* Carnap  
 Consolidated Standards of Reporting Trials  
   (CONSORT), 40, 49  
 Coombes, K., 6, 18  
 counterfactual reasoning, 52, 110, 178  
   in array of models, 300  
   and possible worlds, 429n2  
   and randomization, 287  
   and severity reasoning, 195–6, 245, 342, 429

**474 Index**

- Cousins, R., 203, 210–12, 216
- Cox, D., 27, 45, 47, 53–4, 59, 93–4, 132, 147–54, 158–9, 162, 164–5, 171–5, 180, 194–5, 198–200, 221, 231, 233–4, 237, 248, 250–3, 281–2, 288, 296, 298, 303n1, 343, 352n10, 371, 382–6, 396, 398, 401, 403, 408–9, 412, 418, 428–9, 440
- Cox's taxonomy (of test hypotheses), 150, 310–12
- dividing nulls, 154, 253
  - fully embedded null, 152–3
  - nested alternative, 153, 312
  - omnibus vs. focused test, 154, 310
  - substantively based hypotheses, 157–8
- credible intervals, 426, 428
- CRISPR, 229
- critical region (rejection region), 134, 140, 153n7
- and Bayesian tests, 262
  - relevant for inference, 169–70
  - similar regions, 386n6
- crud factor, 367
- Crupi, V., 72
- Cumming, G., 244–6, 293n12, 354
- $c_\alpha$  (critical value) for standard normal, 191
- use of  $>$  or  $\geq$ , 138, 143, 192, 197
- Dawid, A. P., 282, 363n8, 399, 412–13, 419
- de Finetti, B., 66, 146–7, 223, 227, 414
- Deaton, A., 291
- default/non-subjective priors, 25, 184–7, 402, 431
- non-informative priors, 400
  - relative to parameters of interest 411–12, 415
  - violate Likelihood Principle, 431
- deep learning, 79
- Delampady, M., 51, 124, 252n4, 305n2, 402–3, 405, 431, 434
- Demarcation Problem (between science and pseudoscience), 59, 75–6, 90, 106, 235
- and GTR, 128
  - and Popper, 75–8
  - severe tester and, 88–9, 222
- Diaconis, P., 235
- diagnostic screening (DS) view of tests, 185, 361–3, 408
- and Bayes' rule, 363
  - bias adjustments in, 364
  - dangers of, 369–70
  - false finding rate (FFR), 362, 370
  - point against point, 366
  - positive predictive value (PPV), 363
  - and probabilistic instantiation fallacy, 367
  - sensitivity (SENS) in, 363
  - specificity (SPEC) in, 364
- Dicke, R., 161
- Dienes, Z., 319
- dirty hands argument, 224
- discrepancy, hypothesis as, 130, 143, 151, 241
- our convention on, 240, 263
- DNA match, 281–2
- Dominus, S., 104
- double counting, 92, 269
- Doudna, J., 229
- Dr. Hack, court case, 267–8, 271–2
- Draper, D., 405
- Duflo, E., 290–2
- Duhem, P., 83–5
- Duhem's Problem, 83–9, 107, 154, 311, 385, 435
- Dupré, J., 78
- Durante, K., 105
- Durbin–Watson tests, 311, 314
- Dyson, E., 120, 156
- Earman, J., 69, 74, 128, 156
- eclecticism in statistics, 27, 397, 424
- eclipse tests, 1919, 119, 121–5
- Barnard on, 126, 139
  - Einstein effect, 122, 125
  - H. Jeffreys on, 124, 127
  - mirror distortion controversy, 125, 156–7
  - Newton saving hypotheses, 127–8
  - Sobral and Principe results, 123–4; *see also* General Theory of Relativity (GTR)
- ecumenism in statistics, 27, 301
- Eddington, A., 119–22, 124, 126, 156, 226, 369
- Edwards, A., 32
- Edwards, Lindman, and Savage (E, L, & S), *see also* individual authors
- Edwards, W., 41–3, 45, 49–50, 248, 252, 256, 260, 269
- effect size (ES), population (discrepancy) and observed, 340
- efficient tests of statistical hypotheses, 374–8, 421
- and power, 377–8
- Efron, B., 6, 24, 85, 279n6, 298, 305–6, 391n9, 395, 397, 399–400, 413
- Einstein, A., 119–24, 127–8, 224, 226, 229
- Einstein's Café, 154
- Eisenhart, C., 404
- either or question/horn, 270, 282

- Ellis, P., 356
- empirical Bayes, 404  
 Lindley on, 405; *see also* Robbins, H.  
 Efron, B.  
*en quelque sorte remarquable*, 373–5
- Englert, F., 202
- enumerative induction (EI)/(straight rule of induction), 61  
 Popper on, 75  
 Carnap's, 111  
 and Bayes Theorem (B-boost), 63
- epistemic probability, 195
- epistemology, 4; *see also* normative epistemology
- equivalence principle (weak), 161  
 Einstein (self-gravitating bodies), 161
- equivalence testing, 160n10
- error probabilities, xii, 9, 20, 26  
 for discrediting cherry-picking, 283  
 epistemic interpretation, 26, 429  
 inferential use of, 14, 49, 164  
 and Likelihoodism, 41, 48  
 violate Likelihood Principle, 49–51, 164, 270, 333, 431  
 meaning vs. application, 147, 194  
 performance construal, 13, 38, 140, 429  
 and preregistration, 286, 320, 439  
*P*-value as, 175, 183  
 rubbing off construal, 194, 390, 429  
 and stringency, 382  
 for solving induction, 114–15  
 Type I and Type II, 9, 137–40  
 understanding, 174
- error probability<sub>1</sub> vs. error probability<sub>2</sub>, 183–7, 231, 338
- error statistics, xii, 9  
 error probes, 17  
 foundations for Bayesian tools, 27–8  
 vs. logics of statistical inference, 32, 65, 438  
 severe testing as a proper subset of, 55  
 blog, 418; *see also* Likelihood, Law of ESP, 31, 283–4  
 degenerating program/falsified, 235
- explaining a known effect  
 lost key, 281–2  
 in eclipse results, 127, 157  
 in testing assumptions, 320
- fallacies of rejection  
 spurious *P*-value, isolated result, magnitude error, 3, 22, 94  
 high power = high hurdle, 332; *see also* mountains out of molehills  
 fallacy of non-rejection/insignificance, 3, 152, 339, 353; *see also* SIN
- false discovery rate (FDR) (Benjamini and Hochberg), 277–8, 363n7
- falsification  
 of alternatives in significance tests, 158–62  
 and anomaly, 83  
 asymmetry w/confirmation, 60, 81, 125  
 of Bayesian priors, 417–20  
 and Big Data, 294–5  
 of central dogma of biology, 81–2, 85  
 confusion about, 125  
 and eclipse tests, 125–7  
 error statistics as direct, 436, 443  
 falsifying hypotheses, 83, 121  
 and Fisher, 132  
 in the Higgs, 212  
 and *i*-assumptions, 298  
 of inquiry, 101–2  
 methodological, 82–3  
 and non-replication in GWAS, 294  
 and Popper, 75–89  
 vs. probabilism, 27, 82  
 as provisional, 235  
 of replication, 99, 104  
 statistical, 152; *see also* Duhem's Problem
- family-wise error rate (FWER), 275, 278–9
- Feddersen, A., 390n8
- Feyerabend, P., 228
- Feynman, R., 3, 10, 23, 89
- fiducial  
*c* percent limit, 382  
 frequency distribution, 383  
 inference, 382–4, 389  
 vs. inverse inference, 383  
 Island(s), 371, 382, 391; *see also* Fisher
- file drawer, 98, 176, 212, 292
- FIRST (Fairly Intimately Related to the Statistical Test) interpretations, 150–1, 174n2, 234  
 and Cox's Taxonomy, 150–8
- Fisher and Neyman wars, 141, 165, 386–8  
 contrast to early compatibility, 140, 384, 387  
 deconstructing, 371  
 pathologies of, 390–1  
 similarly behavioristic, 176–7; *see also* incompatibilism
- Fisher, R. A., 4–5, 8, 16, 26, 46, 59, 75, 93, 95, 120, 126, 130–1, 134–7, 147, 151, 159, 161,

## 476 Index

- 166–70, 173–190, 200, 250, 266, 281, 286,  
 290, 303, 323–5, 331, 335, 339, 368, 371,  
 374–91, 398, 409, 424, 440–1, 444  
 and autopsy, 294  
 on background information, 233–4  
 challenged the old guard, 386–7  
 criticized by Neyman, 341, 377  
 relations with Neyman, 139, 146, 165, 181,  
 190, 386–7; *see also* Neyman, feud with  
 Fisher  
 fiducial probability and Neyman’s  
 performance, 382–4, 384n4, 389–91  
 and fraud busting, 54, 284  
 and induction, 66, 86  
 against isolated results, 4, 22, 83, 204, 362,  
 438  
 and making theories elaborate, 237  
 political principle, 5  
 on priors, 232, 383, 402, 404–5  
 shifting views on behavioristic performance,  
 176, 182, 390  
 statistical model, 132  
 tribe, 146–47, 187  
 on UMP tests, 386, 390  
 Fisher’s testing principle (against isolated  
 results), 4, 204  
 Fisherian test statistic criteria, 132, 167, 384  
 Fitelson, B., 68–71  
 flexibility, rewards, and bias hypothesis, 98  
 Folks, J., 325  
 forking paths, 105  
 formal epistemology, 59  
 and intuitive principles of evidence, 73  
 Forster, M., 318  
 Foucault, M., 228  
 foundations, decoupling from traditional, 164,  
 401–402; *see also* Bayesian foundations  
 Fraser, D., 23, 391, 392n10, 405, 428, 430  
 fraud busting, 21, 54, 284–5  
 role of error probabilities, 285  
 Frayn, M., 375  
 and *Copenhagen*, 375  
 Freedman, D., 272, 305  
 Freiman, J., 339n4  
 frequentist matching priors, 254–6, 264, 400,  
 411, 428–30  
 frequentist evidence principle (FEV), 148–50,  
 212  
 applications to Cox’s taxonomy, 152–8  
 FEV(i): evidence of discrepancy, 149,  
 152–3  
 FEV(ii): evidence for the absence of a  
 discrepancy, 150, 152–3, 157, 343  
 and severity (FEV/SEV), 153, 159, 352n10  
 frequentist statistics, 13, 53–5, 110–11, 407  
 as pessimistic, 298  
 as piecemeal, 85, 392; *see also* error statistics  
 frequentist theory of induction, 111, 115  
 Freud, S., 76, 92, 96, 280–1  
 Freudian metaphor for significance tests  
 (Gigerenzer), 178–80  
 Fukushima, 69  
  
 Gaifman, H., 419  
 Gaito, J., 239  
 Garcia, V., 79  
 Geller, U., 31, 235  
 Gellerized hypothesis, 31–4, 38, 235  
 Gelman, A., 27, 83, 94n4, 105, 228, 234, 256,  
 289, 304–5, 359, 361, 401, 403, 407, 417,  
 419, 432–5  
 General Theory of Relativity (GTR), 69,  
 119–29, 156, 160–2, 217, 226, 238, 300; *see  
 also* eclipse tests  
 getting beyond I’m rubber and you’re glue, 247,  
 440–1  
 getting beyond statistics wars, xii, 9, 11–12,  
 178, 283, 436, 444  
 Ghosh, J., 51, 124, 305n2, 402–3, 405, 431, 434  
 Gianotti, F., 210  
 Gibbons, J., 175, 339  
 Gibney, E., 214  
 Giere, R., 285, 422  
 Gigerenzer, G., 22, 93, 96, 136, 178–80, 182,  
 224, 422  
 Gill, R., 284  
 Gillies, D., 299, 421  
 Glymour, C., 51, 69, 70, 73, 156, 227  
 Goldacre, B., 20, 40, 276  
 Goldman, A., 235  
 Good, I. J., 31, 46–7, 223, 282, 389, 418  
 Goodman, S., 264, 269, 332–7, 365–6  
 Gopnik, A., 79  
 Gorroochurn, P., 390n7  
 Gosset, W. (aka Student), 132–3, 141, 384  
 Greenland, S., 26, 252, 256–7, 264, 353–4, 356,  
 365, 406  
 Grünwald, P., 253  
 Guggenheim Museum, 59  
 Gurney, J., 158  
 GWAS, genome-wide association studies, 277,  
 294

- Hacking, I, 30, 32, 37, 45, 64–5, 136, 222–3, 378–80, 417
- Haidt, J., 7
- Haig, B., 264
- Hand, D., 166
- Hannig, J., 392n10
- Hardie, J., 291–2, 294
- Harlow, H., 100
- Hawking, S., 157
- Hawthorne, J., 69–70
- Heisey, D., 355–8
- Hempel, C., 90
- Hendry, D., 279
- Henkel, R., 239, 274, 280
- Hennig, C., 228, 407, 419, 435
- heuristics and biases literature, 422
- Higgs discovery, 10, 189, 202–7, 210–15, 217, 331, 405  
 ATLAS, 203, 206, 209, 217  
 ATLAS and CMS, 202, 211, 213  
 null hypotheses in, 203, 211  
 severity principle, implicit, 208–10  
 statistical controversies in, 202–17  
 subjective Bayesians on, 202–3, 213; *see also* statistical fluctuations
- Higgs, P., 202
- Hinkley, D., 45, 53, 153n7, 172, 174–5, 194, 198, 250, 252, 440
- Hitchcock, C., 92n3
- HIV/AIDS, 292
- Hjort, N., 195, 392n10
- Hochberg, Y., 277–8, 362n7
- Hoenig, J., 355–358
- honest hunter, 275
- hormone replacement therapy (HRT), 20–1, 230
- hot air balloon, 8, 23, 29, 395, 397
- Howson, C., 48, 51n5, 240, 242, 289, 368, 379, 415–17, 419
- Hsu, J., 162n11
- Hubbard, R., 175, 179, 183
- Huber, P., 302
- Huff, D., 3
- Hume, D., 61, 75
- Hurlbert, S., 264
- IID/NIID, 81, 141, 184, 191, 248, 299, 303, 309  
 and bootstrap resampling, 305–7  
 figure of typical realization, 309  
 and GTR, 129
- M-S testing of, 309–13  
 and omnibus tests, 154  
 and runs test, 310
- implicationary assumption (i-assumption), 109–10, 167, 187, 276, 298, 310, 319, 434
- improper posteriors, 412–3
- incompatibilism (between Fisher and N-P), 173, 175, 178–9, 216n4, 439  
 getting beyond, 178–82
- incomplete statistic, 200
- induction  
 faulty analogy with deduction, 64–6, 380  
 logical problem, 60–2  
 Peirce's rationale, 267  
 solving it now, 107–15, 296
- inductive behavior, 110, 146, 176–7, 181, 390–1  
 and de Finetti, 146; *see also* performance
- inductive inference, 8, 59, 162  
 deductively invalid, 61  
 as ampliative, 64, 307, 399, 409, 427  
 frequentist statistics as a theory of, 147  
 vs. inductive behavior, 110  
 as severe testing, 92, 108–9, 209, 302  
 as warranted, 10, 50, 65, 87, 301, 443; *see also* induction
- inductive logic, 32, 379  
 evidential relationship, 275  
 Fisherian leanings, 391; *see also* confirmation theory
- infant training study, 280–1
- inference, defined, 65
- insignificant results, 149, 338–9, 346  
 in cancer clusters, 158  
 in developmental economics, 292  
 and Fisher, 176–7  
 fortifying inferences: DNA, 282  
 in GTR, 125–6  
 in Higgs, 211–13  
 and infant training, 280–1  
 and power, 342–3, 434
- instrumentalism, 79, 297
- intentions, argument from, 48–50, 268–70, 283, 286, 320  
 data other than observed, 273  
 and stopping rules, 273, 431, 438, *see also* Likelihood Principle
- International Society of Bayesian Analysis (ISBA), 202
- inverse inference, 24, 383, 386–7, 404, 409
- Ioannidis, J., 4, 20n2, 92n2, 293n12, 361–2, 364–6, 368–370

## 478 Index

- irrelevant conjunctions, paradox of (tacking paradox), 69–70  
 and severity, 71–2  
 Isaac, story of, 368–9  
 Iyer, R., 7
- Jastrow, J., 288n9  
 Jaynes, E., 403n4, 419n9  
 Jefferys, W., 74  
 Jeffreys, H., 26, 64, 124, 127, 165, 167–9, 183–4, 186–7, 248–9, 252–5, 259n8, 400, 403n4, 409, 424  
 Jeffreys–Lindley paradox (Bayes/Fisher or Jeffreys/Fisher Disagreement), 239, 241–2, 250–1, 253–5, 256n7, 284  
 Jeffreys’ tail area criticism, 168–70, 206, 332  
 Jenkins, G., 47, 302, 313  
 Johnson, V., 261–4, 266, 271, 333
- Kadane, J., 25, 50–1, 166–7, 180, 207, 231, 240, 287, 289, 400  
 Kahneman, D., 97, 99, 422  
 Kaku, M., 119  
 Kalbfleisch, J., 180  
 Kass, R., 7–8, 11, 28, 381, 400–1, 411n6, 424–30  
 Kaye, D., 272  
 Kempthorne, O., 174, 325  
 Kennefick, D.: re-analysis Sobral, 157  
 Kerridge, D., 47n4  
 Keynes, J., 90  
 Kheifets, L., 158  
 Kish, L., 280  
 Klioner, S., 157n9  
 known/old evidence problem, 51, 405  
 Kruschke, J., 273  
 Kruse, M., 47–8  
 Kuebler, R., 339n4  
 Kuhn, T, xii, 88n1  
 kuru, 81–2, 109, 238  
 Kyburg, H., 64, 232, 415
- Lad, F., 227  
 lady tasting tea, 16–19, 31, 169  
 Lakatos, I., 83, 90, 92–3, 235  
 Lakens, D., 162n11, 264  
 Lambert, C., 293–5  
 Lambert, P., 414  
 large  $n$  problem, 240–5  
 in confidence intervals, 245  
 and  $P$ -values, 240  
 and severe tester, 241; *see also* mountains out of molehills  
 Laudan, L., 77, 88, 129, 225n1  
 Law of Large Numbers (LLN), 112, 298, 426  
 Binomial, 112  
 empirical law, 112–13  
 theoretical (mathematical) law, 112  
 Lazar, N., 17, 215–16, 395  
 Lebesgue, H., 131  
 Leek, J., 7  
 Leeper, E., 158  
 Lehmann, E., 120, 131, 137, 139, 146–7, 172–3, 175–7, 182, 191, 368, 372–5, 385n5, 388, 440  
 Letzter, R., 106  
 Levitt Committee (on Diederik Stapel), 78  
 Levi, I., 379  
 Liddell, T., 273  
 lift-off vs. drag-down, 15, 66, 86, 217, 307, 380  
 light deflection, three predictions, 120, *see also* eclipse tests  
 likelihood, 10, 30–31  
 Likelihood, Law of (LL), 30–38, 71  
 vs. error statistical, 41  
 excludes compound hypotheses, 35  
 vs. minimal severity requirement, 30  
 compared to significance test, 34, 242  
 Likelihood Principle (LP), 30, 41, 44–6, 172, 438–9  
 and Bayesian inference, 398, 431  
 and Birnbaum’s “proof”, 173  
 and irrelevance of predesignation, 269  
 and logical empiricism, 90  
 and  $P$ -values, 50, 164, 268, 319  
 post-LP inference, 173  
 violated in bootstrap resampling, 305  
 violated in model testing, 303–4  
 weak, 148  
 and weak conditionality principle (WCP), 172  
 likelihood ratio (LR), 30, 37, 303, 332  
 and Bayes factors, 184, 247  
 exhaustive [LR], 68, 70  
 hypothesis most generous to the alternative,  $H_{\max}$ , 252, 254, 260  
 maximizing data, 260  
 with the Normal, 53n6  
 using  $\alpha/(1-\beta)$ , 337, 366  
 likelihood ratio tests (lambda criterion), 133–5, 139n4  
 limb-sawing logic, 167

- Linda paradox, 422
- Lindemann, F., 128
- Lindley, D., 47, 202, 204, 213–14, 228, 239, 241–2, 250, 253–6, 259, 284, 288–9, 304–5, 397–8, 400–1, 405, 410, 413, 417–18
- Lindley's "Philosophy of Statistics", 397  
criticism of, 398–400
- Lindman, H., 41–3, 45, 49–50, 248, 252, 256, 260, 269
- linear regression model (LRM), 309, 312  
dynamic (DLRM), 316
- Little, R., 431
- Liu, C., 392n10
- live exhibits  
drill prompt (how tail areas exaggerate), 335  
final, 444  
Macho Men, 104–5  
Revisiting Popper's Demarcation of Science, 88  
severity when incalculable, 200
- Lodge, O., 127, 128, 230
- logical empiricism/positivism, 10, 64, 90  
and Big Data, 229  
naïve, 79  
and probabilism, 65  
course of, 224–5, 227  
and verificationism, 229
- Loken, E., 105
- Lombardi, C., 264
- Longino, H., 236
- look elsewhere effect (LEE), 210–11  
and Bayesian analysis, 54
- Louis, T., 43
- Lykken, D., 367
- Machery, E., 156
- Madigan, D., 305, 405
- Maher, P., 71
- Marcus, G., 79
- Marewski, J., 136, 182, 224
- Markov dependency, 314, 316
- Martin, R., 392n10
- Marx, K., 76–7, 125
- Mason, J., 239, 250
- maximum likelihood  
estimator, 133  
hypothesis, 271
- Mayo, D., 26–7, 46–8, 59–60, 74, 82, 84, 86, 88n1, 91, 92n3, 95, 114, 128, 132, 146n5, 147–50, 152–4, 159, 160n10, 162, 164, 171–3, 180, 183, 199–200, 221, 233–4, 237, 239, 270, 273, 281, 286, 298, 303n1, 305, 308, 339, 352n10, 363, 367, 369n10, 385, 399n2, 409, 416, 428, 432–3
- McCloskey, D., 229, 230, 272n2, 330–1
- mean heterogeneity, 313
- measurement, operationalist, 100, 182  
in psychology, 103
- Meehl, P., 92–5, 125, 294, 367–8
- Meng, X. L., 433, 435n8
- meta-analysis, 279, 292, 365
- meta-methodology, 9–10, 54
- meta-research, 9, 98, 106–07  
Meta-Research Innovative Center at Stanford (METRICS), 269n1
- methodological falsification, 82–3, 95, 377;  
*see also* Popper, K.
- methodological probability, 80, 124, 383–4  
and error probability, 79–80, 170, 216  
in Higgs episode, 207
- metric theory of gravity, 161
- Michell, J., 100–1
- microarray, 53, 277, 293
- Mignard, F., 157n9
- Mill, J. S., 89, 296
- Miller, J., 236
- miserable passages  
statistical theater ("Les Miserables Citations"), 372–5  
inferential construal, 375–8  
interpreted by contemporaries, 378–81
- misspecification test/model testing (M-S)  
A. Spanos' approach, 308–20; *see also* auditing  
error fixing in, 311  
in error statistics, 308  
independent of primary question, 319  
irreplication and violated assumption, 320  
and model building, 319  
vs. model selection, 317  
non-exhaustive, 154, 312  
ordering in, 319  
and Peirce, 307  
and predesignation, 319  
residuals as key, 303, 310, 314  
role of significance tests, 298, 301, 309, 433, 441; *see also* probabilistic reduction
- models, all are false, 4, 296–7
- modus tollens*, 84  
unsound example, 126  
statistical, 94, 351n9
- Molière, 208
- Morey, R., 363
- Morrison, D., 239, 274, 280

## 480 Index

- mountains out of molehills fallacy  
 1st form (large  $n$  problem), 240, 243n2  
 2nd form, 144, 266, 326, 359, 366
- multiverse analysis, 105
- Munafò, M., 106–7
- Museums  
 of Statistics, 30, 46, 397  
 Statistical Science and Philosophy of  
 Science, 59, 119, 131, 156, 189  
 on Power Peninsula, 323–4
- Musgrave, A., 90, 129
- National Women’s Health Network, 21
- Nelder, J., 399, 421
- Nelson, L., 43, 237, 270
- Nevins, J., 6, 13, 18, 230
- new experimentalism, 85, 100n7
- Newton, I., 91, 120–4, 126–8, 156–7, 160, 224,  
 229–30, 296, 369
- Neyman, J., 8, 26, 37, 50, 55, 59, 64–5, 83, 86,  
 93, 95, 121, 131–3, 135–7, 144, 146–7, 151,  
 164, 169, 172–4, 183, 185n7, 186–7, 226,  
 239, 285, 346, 357, 384–7, 403–4, 409, 424  
 1933 paper with Pearson, 371–8  
 and behavioral performance, 126–7, 146,  
 165, 173, 176–80, 379–81  
 on CIs and Fisher’s fiducial intervals, 189–  
 90, 382–4, 389–91  
 criticizes Fisher, 341–2, 389–90  
 justification of statistical models, 111–13,  
 298–300  
 equates statistical tests and tests of  
 significance, 174n1, 176, 386  
 feud with R. A. Fisher, 139–41, 165, 179,  
 181–2, 325, 382, 387–91, 440  
 modeling pest control, 299–300, 421  
 power analysis, 339–42, 355–6  
 in prison, 120  
 “The Problem of Inductive Inference”  
 (1955), 110, 341  
 quarrel with Carnap, 108, 110–11, 114, 147,  
 341  
 view of inference, 88, 181, 390–1; *see also*  
 Neyman and Pearson
- Neyman, O. (Lola), 131, 136, 146n6
- Neyman and Pearson  
 1933 paper, 372  
 begin collaboration, 121, 371  
 and Birnbaum, 55  
 importance of error control, 37, 50  
 rescue Fisher, 386–7  
 revolutionized statistics, 139  
 and stopping rules, 50; *see also* miserable  
 passages
- Neyman and Pearson (N-P) tests, 132  
 3 steps in, 131  
 and balancing errors, 341  
 behavioral vs. evidential interpretations, 65,  
 181, 378  
 behavioristic, as verbal preference, 390  
 criticisms of, 144, 164, 170, 324  
 378–81  
 development, 132–7  
 and Fisher dovetail, 137, 169, 384–6  
 gives rigor to Fisher, 132  
 generic form of null and alternative  
 hypotheses, 133  
 Hacking on, 64–5, 378  
 ingredients, 129–30  
 lambda criterion, 133  
 lemma, 139n4  
 meaning vs. application, 147, 194  
 and N-P-Wald behavioral-decision  
 approach, 139, 146, 386, 390  
 and Popper’s methodological falsification, 83  
*P*-values in, 138, 175, 180  
 as severe tests, 142–6  
 usual formulation, 137  
 worse than useless, 136; *see also* statistical  
 tests
- Neyman–Fisher break-up, 1935, 387; *see also*  
 Fisher and Neyman wars
- Neyman–Pearson disagreements, 180, 390–1
- Neyman–Pearson–Fisher battle 1955–6  
 “triad”, 388–90
- Neymanian interpretation of Fisher’s fiducial  
 distribution, 391
- nonsense and ludicrous, Senn on, 326–7, 336,  
 366
- Nordtvedt, K., 161
- Nordtvedt effect and Einstein equivalence  
 principle, 160–2
- Normal distribution, 32–3, 53n6, 123, 129, 136,  
 139, 141, 169, 298–9, 309  
 CI estimation of mean of, 191, 196, 248, 382  
 mixture of, 171  
 standard normal, 143, 241, 326, 348, 357, 378  
 one-sided test of mean of, 142–5, 323, 348,  
 357  
 two-sided tests of mean of, 42, 248, 257, 430
- normative epistemology, 54, 422  
 normative statistical requirements, 437, 441

- Nosek, B., 106  
 novelty requirement, 90–2  
   and severity, 92  
   and eclipse tests, 119n1  
   temporal, 90–1  
   theoretical, 90–1, 96, 119n1  
   use, 90–1  
 Novick, M., 288–9  
 nuisance parameters, 385, 392, 411, 433  
   replaced by sufficient statistic, 385  
 null hypothesis significance testing (NHST), 94  
   illicit animal, 179, 438
- O'Hagan, T., 202, 213–14, 412  
 O'Neil, C., 229  
 objectivity, 4  
   three requirements, 223  
   and Bayesian priors, 232–4  
   as embracing subjectivity, 232  
   in epistemology, 235–6  
   and equipoise priors, 231  
   idols of, 224  
   and journalism, 232  
   and model checking, 298  
   and observation, 225  
   repertoire of mistakes, 235  
   and sampling distribution, 231  
   in statistics, 221  
   and transparency, 236–7  
   and triangulation, 237, 423  
   washout theorems, 231–2; *see also* default/  
     non-subjective priors
- Oishi, S., 101  
 optional stopping, *see* stopping rules  
 outcome-switching, 40, 439  
 Overbye, D., 210–11
- P*-curves, 285  
*P*-value distribution, 151, 325  
   as measuring sensitivity (D. Cox), 151  
*P*-values, 4  
   actual (computed) vs. reported (nominal),  
     17, 43, 179, 274, 303  
   invalidated by selection effects, 17, 285  
   and N-P tests, 138, 175; *see also* N-P tests  
   Bayesian *P*-values, 305, 433  
   can't be trusted except when used to show  
     can't be trusted, 284–5  
   and error probabilities, 173–6, 440  
   exaggerate evidence, 246–53, 260–4, 332,  
     411, 440–1  
   police, 204  
   precise vs imprecise, 333–4, 366; *see also*  
     significance tests  
 paradox of replication, 270–1, 441  
 Parameterized Post Newtonian (PPN)  
   framework, 160–1  
 Pascal's wager, 381  
 Pearson, E., 8, 37, 47, 50, 55, 59, 64–5, 83, 86,  
   88, 93, 95, 121, 132–3, 135–7, 144, 146–7,  
   151, 164–5, 172–4, 182, 189, 239, 269, 285,  
   341, 371, 379–81, 384–8, 392, 403–4, 421  
   1933 paper with Neyman, 371–8  
   answering Fisher criticism, 140, 177, 388  
   armour-piercing, 180–1  
   on Bayesian priors, 226, 404  
   and inferential construal, 180–1, 380–1,  
     391–2  
   in love with woman his cousin was to marry,  
     137  
   on power (post data), 324  
   rejects behaviorism, 127, 180, 381–2, 391  
   and tail areas, 169  
   and three steps in test construction, 131, 178,  
     386; *see also* Neyman and Pearson  
 Pearson, K., 120–1, 131–2, 140, 146, 189, 386,  
   404  
 Peirce, C. S., 18, 86  
   on faulty analogy of induction and  
     deduction, 64–6, 380  
   on inverse chance, 408–9  
   on justifying induction, 113–14, 267, 307  
   randomization and predesignation, 89, 267,  
     288n9  
   and testing assumptions, 307  
 Perez, B., 6, 18  
 performance construal of N-P, 174–8  
   vs. severity, 139–40  
   and Fisher's fiducial probability, 382, 390;  
     *see also* N-P tests  
 pest control, 299–300  
 phenomenon vs. data, 121  
 philosophy in statistical methodology, xii, 4–5,  
   8–14, 28, 49, 73, 114, 432  
   and cheating, 46, 270, 332  
   in identifying good science, 77  
   severe testing philosophy, 23, 195, 437, 444  
   Bayesian philosophy, 24, 26, 396  
 Pickrite method, 19, 30, 51, 276

## 482 Index

- piecemeal testing, 162, 308, 380, 400, 443  
 division of labor, 85, 392, 423
- Pigliucci, M., 78
- Playfair, L., 17–18
- Pleiades, 373
- Poole, C., 26, 256–7, 264, 406
- Popper, K., 8–9, 27, 40, 59, 66–8, 72–3, 75–80, 82–93, 95–6, 114, 119, 125–6, 159, 195, 209, 227, 229, 237, 259, 294, 390, 433; *see also* falsification
- positive predictive value (PPV), *see* diagnostic screening
- posterior predictive distribution, 433–4  
 Duhem’s Problem in, 435; *see also* M-S tests
- Potti, A., 6, 13, 18, 97, 230
- power, 135–9  
 3 roles accorded by N-P, 324  
 and the clinically relevant/irrelevant difference, 326–7  
 Cohen’s snafu, 324  
 detailed discussion, 323–41  
 fallacious transposition in, 331  
 and Fisher, 325  
 how to increase, 325  
 incomplete concept, 353  
 low power and violated assumptions, 361  
 predata and postdata, 323–5  
 retrospective (post hoc), 353–9, 359–61  
 and severity, 323–32  
 and Type II error probability, 138; *see also* power attained
- power analysis, 323  
 and CIs, 356–8  
 fallacy of, 353–6  
 Jacob Cohen on, 324, 338  
 and Neyman, 339  
 ordinary, 340  
 ordinary vs. shpower, 355  
 vs. severity, 338, 343, 350  
 and significance test reasoning, 339
- power attained (att power), and attained sensitivity  $\Pi(\gamma)$ , 151, 164, 196, 324, 342–3, 355n2, 358
- Power Peninsula, 323, 353–4, 382
- Pratt, J., 44, 175, 240, 248, 252, 254, 339
- precautionary principle, 341
- prespecification/predesignation, 40, 106, 269, 373–7, 438  
 and error probabilities, 286, 320; *see also* novelty requirement
- principle of indifference, 386, 391, 400
- prions, 81–2, 85, 88, 109–10, 238  
 prion protein (PrP), 109  
 protein folding (pN), 109  
 protein misfolding (pD), 110
- probabilistic instantiation fallacy, 367
- probabilistic reduction (Spanos), 312  
 dememorized data, 316  
 detrended data, 307, 316  
 lags, 314, 316  
 menu of assumptions, 312  
 reparameterizations, 320n5  
 respecification, 313; *see also* M-S tests
- probability, roles of  
 probabilism, performance, and  
 probativeness, 13–14, 24–7, 33, 77, 436–7  
 avoiding the need for different, 54–5, 429  
 events vs. hypotheses, 407  
 formal vs. informal meanings, 10, 194, 214, 427  
 performance vs. severity, 15, 26–7, 50, 54, 162  
 probabilism and performance, 428  
 probativism vs. probabilism, 127, 226  
 variability vs. belief, xi, 54, 80, 428; *see also* methodological probability
- probable errors, 124
- probare*, 10, 226, 423
- protein misfolding cyclical amplification (PMCA), 110
- Prusiner, S., 81–2, 109–10, 238, 369
- pseudoscience, *see* Demarcation Problem
- questionable research practices (QRPs), 20, 78, 98, 267, 271, 292, 439
- quicksand, 183, 187–8, 367, 402
- radical skepticism, 229
- Raftery, A., 305
- Raiffa, H., 44
- randomization, 286–9  
 possible Bayesian home for, 288–9  
 and cloud seeding, 126  
 and deliberation, 292, 294  
 in GWAS, 293  
 and C. S. Peirce, 18, 267, 288n9  
 and the philosophers, 289–90  
 Poverty Action Lab (MIT), 290–1  
 randomized controlled trials (RCTs), 88  
 RCT4D, 290–2
- rational reconstruction, 8, 73, 85, 162
- Ratliff, K., 101
- real random experiments, 111, 298–9
- realism vs. antirealism, 79

- severe tester agnostic on, 297  
theoretical mistakes, 297
- Reich, E., 210
- Reid, C., 120–1, 137, 139, 141, 146, 189–90, 372, 387–8, 404
- Reid, N., 54, 186, 392, 396, 429
- rejection ratio, 337–8
- repertoire of errors, 89, 234, 308, 400, 414, 442  
in selection effects, 279
- replicability/reproducibility, 6, 20, 28  
ASA definition, 97  
and diagnostic testing, 368  
equivocation in, 246  
and predesignation, 270, 320  
and Popper, 82–3
- replication (crisis), 59, 89, 156, 221, 361  
in GWAS, 295  
in psychology, 78, 97–107; *see also* paradox of replication
- residuals, 298, 303, 310–311, 317  
small residuals vs. adequacy, 318
- rigged hypothesis, 108
- Robbins, H., 390, 404
- Robert, C., 401, 402, 406, 407, 413, 428
- Romano, J., 172, 175, 191
- Rosenkrantz, R., 40, 69, 269, 319–20, 419n9
- Rosenthal, R., 239
- Rothman, K., 264, 276, 272n3
- Royall, R., 33–9, 41, 44, 50, 52, 68, 70n5, 82, 212, 225, 243, 283, 319, 332, 421
- rubbing-off construal, 65, 194, 244, 391, 429
- Rubin, D., 47, 433
- Salmon, W., 64, 95, 114, 310
- Samanta, T., 51, 124, 305n2, 402–3, 405, 431, 434
- sampling distribution, 32, 142  
and error probabilities, 130, 173, 428, 438  
and bootstrapping, 306  
and frequentist objectivity, 231  
relevant, 199  
as testable meeting ground, 178
- sampling plan, freedom from, *see* Likelihood Principle  
stopping rules
- sampling theory/philosophy, 55, 172
- Sanna, L., 284
- Savage Forum 1959, 46–8, 420–1, 430
- Savage, L., 8, 41–3, 44–50, 173, 214, 228, 230n3, 248, 252, 256, 260, 269, 287, 302, 397, 401, 417, 420, 424, 430, 432
- Schachtman, N., 272n3
- schizophrenia and split personality, 409, 414, 424, 436, 434
- Schlaifer, R., 44
- Schnall, S., 100
- Schweder, T., 195, 392n10
- SCOTUS, 272n3
- Sebastiani, P., 293
- Seidenfeld, T., 411n6
- selection effects, biasing, 3, 19, 21n2, 40–1, 78  
defn., 92, 285, 437; *see also* stopping rules  
adjusting for, 154, 268, 275–7, 364–5, 418  
and auditing, 234, 267, 269  
NHST, 95  
preregistration, 106, 266, 275, 286; *see also* Bonferroni correction
- self-correcting, 20, 162  
induction as, 114, 307; *see also* arguments from error and coincidence
- self-sealing fallacy, 103
- Sellke, T., 175–6, 184, 248–52, 258–60, 338
- Selvin, H., 274, 279
- semantic entailment: severity version, 65
- Senn, S., 151, 162n11, 247, 251–3, 259n8, 264–5, 266, 287–8, 290, 293n12, 326–7, 336, 345–6, 365n9, 366, 413–14, 417–19
- sensitivity, achieved or attained, 151  
function,  $\Pi(y)$ , 151–2  
and severity, 152; *see also* power attained
- sequential trials, 47; *see also* stopping rules
- severe tester, tribal features, 9, 27, 114, 437  
on comparativism, 79, 421, 441  
on the demarcation of science, 88–9  
and Duhem's problem, 85–6  
on improving confidence intervals, 194, 244–5, 358, 429, 442  
interpretation of probable flukes, 217  
on large-scale theories, 129  
on Likelihoodism and the LP, 39–41, 48–50, 72  
new name, 55  
vs. the N-P behavioristic prison, 140  
vs. Popperian severity, 83  
on the revolution in psychology, 100, 103–4, 107, 370  
solving the problem of induction, 107–14  
on statistical falsification, 235  
on statistical objectivity, 320  
and translation guide, 52
- severity, 5  
applied, water plant accident, 143–5

## 484 Index

- severity (cont.)  
 attained power, 342–3  
 and confidence levels, 193  
 and difference between two means, 345–6  
 disobeys the probability axioms, 423  
 and explanatory content/informativeness, 79–80, 237  
 and Fisherian tribes, 146  
 function (SEV), 143  
 and large-scale theories, 128, 162  
 in meta-methodology, 9, 32  
 and Popperian corroboration, 72, 75, 87  
 vs. power analysis, 343  
 and replicability, 370  
 and sensitivity, 152  
 when not calculable, 200  
 severity curves, 348–9, 360n4  
 severity interpretation of negative results (SIN), 143–5, 152, 212, 343, 346–7, 351;  
*see also* severity  
 severity interpretation of rejection (SIR), 143, 265–6, 351  
 severity requirement/principle, 5, 92, 125, 258  
 and biasing selection effects, 92  
 and error control, 269  
 and failed replication, 158, 266  
 to block fallacies of rejection, 144, 357  
 as heuristic tool, 12, 264  
 informal, 109  
 from low *P*-value, 209  
 as minimal principle of evidence, 5, 396  
 for non-significance (Higgs), 212  
 in terms of solving a problem, 300  
 vs. fit measures, 72  
 weak and strong, 22, 108  
 strong, 14  
 Sewell, W., 280  
 sexy science: severe testing in large-scale theories, 121, 163, 300  
 Shaffer, J., 275  
 Shalizi, C., 27, 432, 434  
 Sharpe, G., 137  
 shpower (retrospective power analysis), 354–6  
 howlers of, 355–6  
 vs. severity, 356; *see also* power analysis.  
 significance levels  
 as predesignated, 137  
 attained vs. predesignated, 173–5, 177;  
*see also* *P*-values  
 significance tests  
 vs. comparativism, 35  
 Cox definition of, 93  
 criticisms of, 93–5, 438; *see also* chestnuts and howlers of tests  
 fallacies of rejection/non-rejection  
 falsifying alternatives in, 159  
 Fisherian (pure/simple), 132, 150  
 in Higgs, 202  
 roles for model testing and discovery, 298–304; *see also* M-S tests  
 simple or point hypotheses, 33  
 test T+, 144; *see also* Neyman and Pearson (N-P) Tests  
 Silberstein, L., 127  
 Silver, N., 232–3  
 similar tests, 385, 386n6  
 Simmons, J., 43, 237, 270  
 Simonsohn, U., 43, 237, 270, 284, 285  
 Singh, K., 391  
 skin off your nose, 273  
 Skyrms, B., 62, 73  
 Slovic, P., 422  
 Smeesters, D., 284  
 Smith, C., 47  
 Smith, H., 339n4  
 Sober, E., 35–6, 47, 92n3, 242, 317–18, 380–1  
 Spanos, A., 120, 133–4, 139, 146n5, 200, 254–5, 305, 308, 312–13, 317–19, 331, 352n10, 355, 367, 387, 426  
 Spiegelhalter, D., 204–5, 401, 404  
 spike and smear priors, 239, 248, 250–1, 259, 336  
 Bayesian justification for, 251  
 coffee shop, 257  
 criticisms of, 252n4, 256, 259, 406, 440  
 cult of the holy, 252  
 severe tester on, 257–8  
 spongiform diseases, 81, *see also* kuru  
 Sprenger, J., 307  
 Sprott, D., 180, 399, 421  
 spurious associations, 3  
 batch effects, 293  
 in longevity study, 293, 362  
 population and number of shoes, 308, 317  
 sea level and price of bread, 317  
 Staley, K., 203, 235–6  
 standard model (SM) physics, 203, 206, 214, 215  
 Stapel, D., 78, 97, 100, 276  
 statistical battles, current state of play, 11–12, 23–8, 395–7, 400–2, 444  
 proxy, 437; *see also* getting beyond statistics wars

- statistical fluctuations/flukes and Higgs, 202–5,  
     210–12  
     interpreting, 214–15  
 statistical inference, 7, 20, 42, 65–6, 174  
     and scientific theory appraisal, 119, 202; *see*  
     *also* inductive inference  
*Statistical Methods for Research Workers*  
     (SMRW), 387  
     Fisher backtracks, 182  
*Statistical Power Analysis for the Behavioral*  
     *Sciences* (Cohen), 324  
 statistical tests, elements of, 129–30  
     test hypothesis, 31, 109, 130, 133, 341  
     test rule, 130  
     test statistic, 34, 94, 129, 132, 167  
     test statistic, pivotal, 378  
 statistical tests, properties of  
     consistent, 136  
     monotonicity, 134  
     powerful, 135–6  
     unbiased, 136, 141  
 Steegen, S., 105  
 Stern, H., 433  
 Stevens, S., 100  
 Stigler, S., 288n9  
 Stone, M., 254, 413  
 stopping rules/optional stopping, 42–53,  
     170, 270  
     and Bayesian intervals, 430  
     principle, 43, 54, 431  
     proper, 42  
     statisticians on: Armitage, 47  
     G. Barnard, 48  
     J. Berger /Wolpert, 49, 187, 430–1  
     G. Box, 303  
     D. Cox and Hinkley, 45  
     Savage (E, L, & S), 43, 46; *see also* intentions  
 Strassler, M., 212  
 Strawderman, W., 391  
 Student's t, 141, 345, 378  
 suddenly smitten, 131, 137, 388  
 sufficient statistic, 147–8, 200, 298, 385–6,  
     303n1, 383  
     complete, 385–6  
 Sugden, R., 21  
 Suppes, P., 86  
  
 tacking paradox, *see* irrelevant conjunctions,  
     paradox of  
 Talbott, W., 416  
 Taleb, N., 78, 267  
  
 Tentori, K., 72  
 testing, proving (in the biblical sense), 302  
 Texas marksman (sharpshooter), 19, 31, 158,  
     276  
 Thaler, R., 101  
 thalidomide, 282  
 Thalos, M., 415  
*The Significance Test Controversy* (Morrison  
     and Henkel), 239, 274, 280  
 Tonelli, G., 210  
 tradeoffs and benchmarks, 328–9  
 Translation Guide (severe tester's), 52, 170,  
     204, 335  
 transparency, 3, 237  
 TSA screening, 363  
 Tuerlinckx, F., 105  
 Tversky, A., 422  
 two measuring instruments of different  
     precisions (Cox 1958), 170–2, 180,  
     199; *see also* weak conditionality  
     principle  
 Type I, Type 2 error probabilities, 9, 137–40;  
     *see also* error probabilities  
 Type II rationality, 418  
  
 underdetermination, 376  
     weak severity blocks, 108  
 unifications (reconciliations), *see* Bayesian vs.  
     frequentist  
 uniformity of nature, 62  
 uniformly most powerful (UMP) test, 35, 136,  
     139, 139n4, 141, 262  
     UMP unbiased, examples, 141–2  
     and uniformly most accurate CIs, 191  
     Fisher on, 386, 390  
 uninformative priors, do not exist, 231, 400  
 universal Bayesianism, 417  
 unobservables, 297  
 Urbach, P., 48, 240, 242, 289, 379  
  
 van Belle, G., 364  
 Vanpaemel, W., 105  
 Vasudevan, A., 419  
 Venn, J., 387  
 Vigen, T., 317  
 von Mises, R., 113  
  
 Wagenmakers, E.-J., 253, 269, 283, 284  
 Wald, A., 139–40, 146–7, 386, 390  
 Waller, N., 95  
 Wang, J., 6, 18

## 486 Index

---

- Wasserman, L., 24, 204, 287, 289, 392, 400, 411n6, 429, 433  
 Wasserstein, R., 17, 215–16, 395  
 warranted inference *see* inductive inference  
 water plant accident, 142–5, 153, 186, 249–51, 259, 306–7, 348, 353–4, 359–60  
 weak conditionality principle (WCP), 172–3  
   claim it leads to LP, 172–3; *see also* Likelihood Principle  
 Weak Law of Large Numbers, 408  
 weak repeated sampling principle, 45, 303, 396  
 Wellek, S., 162n11  
 Wertheimer, N., 158  
 Westfall, P., 279, 281n7  
 Wilks, S., 134, 139  
 Will, C., 124, 160–1  
 Williamson, J., 416–17  
 Wilson, R., 21  
 Wolpert, R., 49, 199, 430–1  
 Woodward, J., 121, 235  
 Worrall, J., 91, 92n3, 290  
 Wright, F., 59  
  
 Xie, M., 391  
 Xu, F., 79  
  
 Yates, F., 390  
 Young, S., 279, 281n7, 293  
  
 Zabell, S., 390  
 Zidek, J., 413  
 Ziliak, S., 229–30, 272n2, 330–1