

Contents

1	Introduction	1
1.1	The Fiducial Bayesian Inference	2
1.2	The Stranglehold of Significance Tests	3
1.3	Beyond the Significance Test Controversy	4
1.4	The Feasibility of Fiducial Bayesian Methods	5
1.5	Plan of the Book	5
	References	6
2	Preamble—Frequentist and Bayesian Inference	9
2.1	Two Different Approaches to Statistical Inference	9
2.1.1	A Simple Illustrative Example	10
2.2	The Frequentist Approach: From Unknown to Known	10
2.2.1	Sampling Probabilities	11
2.2.2	Null Hypothesis Significance Testing in Practice	11
2.2.3	Confidence Interval	12
2.3	The Bayesian Approach: From Known to Unknown	13
2.3.1	The Likelihood Function and the Bayesian Probabilities	13
2.3.2	An Opinion-Based Analysis	14
2.3.3	A “No Information Initially” Analysis	16
	References	19
3	The Fisher, Neyman–Pearson and Jeffreys Views of Statistical Tests	21
3.1	The Fisher Test of Significance	21
3.1.1	An Objective Method for Reporting Experimental Results	21
3.1.2	The Null Hypothesis	22
3.1.3	The Outcome of the Test of Significance	22
3.1.4	The Test Statistic and the Level of Significance p	22
3.1.5	How to Evaluate the Smallness of p	22

3.2	The Neyman–Pearson Hypothesis Test	23
3.2.1	Rational Decision Rules.	23
3.2.2	The Hypothesis to be Tested and Alternative Hypotheses	23
3.2.3	The Outcome of the Hypothesis Test.	24
3.2.4	A Long-Run Control	24
3.2.5	Two Types of Errors and Their Long-Run Frequencies	24
3.2.6	Power of the Test and Best Critical Region	25
3.3	The Jeffreys Bayesian Approach to Testing	25
3.3.1	The Jeffreys Rule	26
3.3.2	The Function of Significance Tests	26
3.3.3	A Specific Prior for Testing Precise Hypothesis	27
3.3.4	A Measure of Evidence Against the Null Hypothesis.	27
3.3.5	An Averaged Risk of Error	28
3.4	Different Views of Statistical Inference	29
3.4.1	Different Scopes of Applications: The Aim of Statistical Inference	29
3.4.2	The Role of Bayesian Probabilities	31
3.4.3	Statistical Tests: Judgment, Action or Decision?	33
3.5	Is It Possible to Unify the Fisher and Neyman–Pearson Approaches?	34
3.5.1	Demonstrating Equivalence	34
3.5.2	Neyman–Pearson’s Criterion Leads to Incoherent and Inadmissible Procedures.	35
3.5.3	Theoretical Debates: Counterintuition or Good Sense?	35
3.6	Concluding Remarks	36
	References.	36
4	GHOST: An Officially Recommended Practice	39
4.1	Null Hypothesis Significance Testing	39
4.1.1	An Amalgam	39
4.1.2	Misuses and Abuses	40
4.2	What About the Researcher’s Point of View?	42
4.2.1	A Cognitive Filing Cabinet	42
4.2.2	It Is the Norm	42
4.3	An Official Good Statistical Practice	43
4.3.1	Guidelined Hypotheses Official Significance Testing	43
4.3.2	A Hybrid Practice.	45
	References.	46

5	The Significance Test Controversy Revisited	49
5.1	Significance Tests Versus Pure Estimation	49
5.1.1	The Meehl Paradox	49
5.2	The Null Hypothesis: A Straw Man	50
5.3	Usual Two-Sided Tests Do Not Tell the Direction	51
5.3.1	Two-Sided Versus One-Sided Tests and Their Shortcomings	51
5.3.2	Jones and Tukey's Three-Alternative Conclusion Procedure	51
5.4	Determining Sample Size	52
5.5	Critique of P-values: A Need to Rethink	53
5.5.1	Jeffreys' Answer to the Problem of Pure Estimation	53
5.5.2	The Bayesian Interpretation of the P-value	54
5.5.3	Student's Conception	54
5.5.4	Jaynes' Bayesian Test	54
5.5.5	The Methodological Shortcomings of NHST Clearly Pointed Out	55
5.5.6	The Bayesian Interpretation of the Two-Sided P-value	55
5.5.7	Killeen's P_{rep}	56
5.6	Decision and Estimation	57
5.6.1	The Decision-Making Viewpoint: A Very Controversial Issue	57
5.6.2	Jeffreys' Bayesian Methodology	58
5.7	The Role of Previous Information and the Sample Size	58
5.8	The Limited Role of Significance Problems	59
5.9	Other Issues	59
5.9.1	Noninferiority and Equivalence Questions	59
5.9.2	Stopping Rules and the Likelihood Principle	60
	References	61
6	Reporting Effect Sizes: The New Star System	63
6.1	What Is an Effect Size?	63
6.1.1	A Definition Restricted to Standardized Measures	64
6.2	Abuses and Misuses Continue	64
6.2.1	A Psychological Example	64
6.2.2	An ES Indicator that Does Not Tell the Direction	65
6.2.3	Disregarding the Robust Beauty of Simple Effect Sizes	65
6.2.4	Heuristic Benchmarks: A New Star System	66
6.2.5	Observed ES Indicators Can Be Misleading	66
6.2.6	A Good Adaptive Practice Is Not a Good Statistical Practice	67

6.2.7	The Need for a More Appropriate Sample Size.	67
6.2.8	The Shortcomings of the Phi Coefficient	68
6.3	When Things Get Worse	69
6.3.1	A Lot of Choices for a Standardized Difference	69
6.3.2	A Plethora of ES Indicators	70
6.3.3	Don't Confuse a Statistic with a Parameter	71
6.4	Two Lessons	73
6.4.1	The New Star System	73
6.4.2	Should Standardized Effect Sizes Ever be Used?	74
	References.	74
7	Reporting Confidence Intervals: A Paradoxical Situation	77
7.1	Three Views of Interval Estimates.	77
7.1.1	The Bayesian Approach (Laplace, Jeffreys)	77
7.1.2	Fisher's Fiducial Inference	79
7.1.3	Neyman's Frequentist Confidence Interval	80
7.2	What Is a Good Interval Estimate?	82
7.2.1	Conventional Frequentist Properties.	82
7.2.2	The Fatal Disadvantage of "Shortest Intervals"	82
7.2.3	One-Sided Probabilities Are Needed	83
7.2.4	The Jeffreys Credible Interval is a Great Frequentist Procedure	83
7.3	Neyman-Pearson's Criterion Questioned	83
7.3.1	The Inconsistencies of Noncentral F Based Confidence Intervals for ANOVA Effect Sizes	84
7.3.2	The Official Procedure for Demonstrating Equivalence	86
7.4	Isn't Everyone a Bayesian?	87
7.4.1	The Ambivalence of Statistical Instructors	87
	References.	89
8	Basic Fiducial Bayesian Procedures for Inference	
	About Means	91
8.1	Fiducial Bayesian Methods for an Unstandardized Contrast . . .	92
8.1.1	The Student Pharmaceutical Example	92
8.1.2	Specific Inference	92
8.2	Fiducial Bayesian Methods for a Standardized Contrast	94
8.2.1	A Conceptually Straightforward Generalization.	94
8.2.2	Inference About the Proportion of Population Differences.	97
8.3	Inference About Pearson's Correlation Coefficient.	97

8.4	A Coherent Bayesian Alternative to GHOST	98
8.4.1	NHST: The Fiducial Bayesian Interpretation of the p -Value	98
8.4.2	Interval Estimates: The Fiducial Bayesian Interpretation of the Usual CI.	99
8.4.3	Effect Sizes: Straight Bayesian Answers	99
8.4.4	Making Predictions	100
8.4.5	Power and Sample Size: Bayesian Data Planning and Monitoring	102
8.5	Our Guidelines	102
	References.	102
9	Generalizations and Methodological Considerations for ANOVA	105
9.1	From F Tests to Fiducial Bayesian Methods for ANOVA Effect Sizes	106
9.1.1	The Traditional Approach	106
9.1.2	Fiducial Bayesian Procedures	108
9.1.3	Some Conceptual and Methodological Considerations	112
9.2	Alternatives to the Inference About ANOVA ES	114
9.2.1	The Scheffé Simultaneous Interval Estimate and Its Bayesian Justification	114
9.2.2	Contrast Analysis	117
9.3	An Illustrative Example: Evaluation of the “0.05 Cliff Effect”.	118
9.3.1	Numerical Results.	118
9.3.2	A Cliff Effect Indicator	119
9.3.3	An Overall Analysis Is Not Sufficient	122
9.3.4	A Simultaneous Inference About All Contrasts.	123
9.3.5	An Adequate Analysis.	124
9.3.6	What About Standardized Effects?	124
9.4	Our Guidelines for ANOVA.	125
	References.	125
10	Conclusion	127
	References.	129
	Index	131