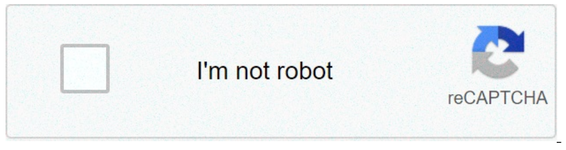

Continue

Psychological testing principles applications and issues pdf

Psychological testing principles applications and issues 9th edition pdf free. Psychological testing principles applications and issues 9th edition citation. Psychological testing principles applications and issues 7th edition. Psychological testing principles applications and issues (9th ed.). Psychological testing principles applications and issues 8th edition. Psychological testing principles applications and issues 9th edition. Psychological testing principles applications and issues 8th edition pdf free. Psychological testing principles applications and issues 9th edition ebook.

PDF 2018 â€ "Cengage â€" ISBN: 1337098132 â€ "Psychological tests: principles, applications and problems â by Robert M. Kaplanâ # 25269 9a English edition | 2018 |. | 738 pages | PDF | 17.44 MB Â Psychological tests: principles, applications and problems, the ninth edition explains the fundamentals of psychological tests, their important applications and disputes emerging from such applications in clinical, educational, industrial and legal contexts. Kaplan and Saccuzzo's immersive and complete text demonstrates the way psychological tests are built and used, both in a professional context and in everyday lives. Part focuses on the basic concepts that influence the evaluation of all tests. Part II discusses the main types of psychological tests, while Part III looks at current issues concerning tests such as stereotype threat, prejudice, laws and ethics. A multitude of test profiles and examples illustrate how psychological tests are used and reported. Real-life case studies demonstrate the uses and infactuals of psychological tests, while â€ "Technical Example" Boxes help to understand complex statistical concepts. Kaplan, Robert M. (Robert Malcolm), 1947-. Psychological tests: principles, applications and problems. Pacific Grove, CA: Brooks / Cole Pub. Co., 1997. Kaplan, Robert M. (Robert Malcolm), 1947-. (1997). Psychological tests: principles, applications and problems. Pacific Grove, CA: Brooks / Cole Pub. Co., Kaplan, Robert M. (Robert Malcolm), 1947-. Psychological tests: principles, applications and problems. Pacific Grove, CA: Brooks / Cole Pub. Co., 1997. Note: These quotes are generated software and may contain errors. To check accuracy, check the appropriate style guide. Export Export to Citation Manager (RIS) Back to article University of Wisconsin-Madison Upload File please Wait ... Seventh edition Psychological Principles, Applications and Problems This page intentionally left the seventh edition of psychological testing principles, applications and problems Robert M. Kaplan University of California, Los Angeles Dennis P. Saccuzzo San Diego University State Australia â€ ¢ Brazil â â€ ¢ Japan â€ ¢ Korea â€ ¢ Mexico â€ ¢ Singapore â€ ¢ Spain â€ ¢ United Kingdom â€ ¢ United States psychological tests: Principles, applications and problems, seventh edition © 2009, 2005 Wadsworth, Censage Learning Robert M. Kaplan, Dennis P. Saccuzzo All rights reserved. No part of this work covered by the manager of this here can be reproduced, transmitted, stored or used in any form or editor: Jaime Perkins with any graphic, electronic or mechanical medium, including but not editorial assistant: Wilson Co Limited to Fotocopied, Registration, Scan, Digitization, Recording, Project Manager Web Technology: Distribution of Amy Cohen, information networks or information storage and recovery systems, except for permission in Section 107 or 108 of Marketing Manager: Kim Russell Marketing Assistant: Molly Felz 1976 United States Copyright Act, without the written permission of the publisher. Marketing communication manager: Talia Wise Project Manager, Editorial Production: Charlene M. Carpenter for product information Tech support, contact us at Cengage Learning Customer & Sales Support, 1-800-354-9706. Creative Director: ROB HUGEL for permission to use the material from this text or product, send all online requests to cengage.com/permissions. Additional permissions Questions may be sent by e-mail to [E-mail protected] Article Manager: Vernon Banes Printing Actory: Linda HSU Editor: Bob Kauser Production Service: Newgenâ€ "Austin Text Designer: Lisa Henry Photo Ricercatore: Darren Wright Library of CongressNumber: 2008927883 Student Edition: ISBN-13: 978-0-495-09555-2 ISBN-10: 0-495-09555-9 Copy editor: Mary Ann Grobbel Cover Designer: Larry Didona Cover Image: â€ œgeometrics below the clouds € © Pete Turner / The Image Bank / Getty Images Wadsworth 10 Davis Drive Belmont, CA 94002-3098 94002-3098Composite: Newgen Cengage Learning is a leading provider of personalized learning solutions with destinations around the world, including Singapore, UK, Australia, Mexico, Brazil and Japan. Locate your local OÃ¯â¬CE on international.cengage.com/region. Cengage learning products are represented in Canada by Nelson Education, Ltd. for the solutions of your course and learning, visit academic.cengage.com. Buy one of our products at your local college or at our favorite online store www.ichepters.com. Printed in the United States of America 1 2 3 4 5 6 7 12 11 10 09 08 Short content PA RTI 1 2 3 4 5 6 7 PA RTII PR I NCI PLES Introduction 1 Basic standards and statistics for testing 25 Correlation and regression 65 reliability 101 validity 133 writing and evaluation Test articles 157 Test Administration 185 Actions Application 8 Interview techniques 201 9 Intelligence theories and Binet Scales 229 10 The Wechsler Intelligence Biles: WAIS-III, WISC-IV, 11 12 13 14 15 16 17 18 PA RTIII and WPPSI-III 249 Other individual ability tests in education and special education 273 Standardized evidence in education, civil service and military applications in clinical and consulting settings 333 projectives Personality tests 373 Computer and basic psychological science in the test 403 Test in Counseling Psychology 433 Test in health psychology and health care 451 tests in industrial and business settings 483 303 problems 19 test bias 511 20 Tes You Law 545 21 ethics and the future of the psychological test 585 V PA content rtiprinciple 1 Introduction 1 Basic concepts 6 What test is 6 types of test 7 Overview of the book 9 Principles of psychological tests 10 applications of psychological tests 10 test problems Psychological 11 Historical perspective 11 Early antecedents 11 Charles Darwin and individual DiÃ¯â¬ â,¬ Erences 12 Experimental psychology and psychophysical measurement 12 The evolution of intelligence and standardized test tests 14 Personality tests: 1920 - 1940 17 The emergence New Approaches to the Test of Personnel 18 The period of rapid changes in the test status 20 The current environment 21 Summary 22 2 Basic standards and statistics for test 25 Why do we need statistics 26 Measurement scales 27 VI Summary properties Scales 27 Types of scales 29 Allowed operations 30 Frequency distributions 31 Rankings percentile 34 percentile 38 Describing distributions 34 media 39 deviation and standard 40 z 42 s score tandard normal distribution 45 mccallÃ, â "¢ s t 50 quartiles and decil 51 standards 53 rules relating to age 54 tracking 55 criterion reference test Summary 3 60 63 correlation and regression 65 scatter diagram 66 correlation 68 regression 69 The regression line 69 The best assembly line 71 Test of the statistical meaning of a CoriÃ¯æ'cient 76 correlation How to interpret a regression plot 78 Other COEÃ¯ correlations 82 Terms and problems in the use of correlation 84 Standard Standard Error 84 Estimate 85 CoeÃ¯â¬ æ'Cient of determination 85 CoeÃ¯æ'cient of alienation 85 narrowing 86 Croce validation 86 Correlation problem-causation problem third Explanation Variable 88 Limited range 88 Multivariate analysis (optional) 87 88 General approach 89 An example using the discriminating multiple regression analysis 91 factor analysis 91 90 VII VIII Summary summary 94 Appendix 3.1: Calco The regression equation and a COEÃ¯Â¬æ'Cient 95 correlation calculation of a regressionequal equation (data from T-skilled 3.5) 95 Calculation of a COEÃ¯Â¬Cient correlation (data from Table 3.5) 98 4 Reliability 101 History and theory of reliability 102 Conceptualization of error 102 Spearman primitive studies 103 Basic notions of the test score theory 103 The element of the sampling model of the Response theory 107 reliability models 108 105 error sources 109 time sampling: the Test-Retest method 109 sampling of the object: forms parallel method 111 method half method 111 kr 20 formula 113 Coeï¬cient alpha alphaReliability of an Erence Score 116 Reliability in Observation behaviour Studies 120 connecting error sources with reliability Evaluation method 121 Use Reliability Information 124 Standard measuring errors and rubber meter How reliable is it? 125 What to do for low reliability 125 124 Summary 129 Appendix 4.1: Using Alpha Coeï¬ to estimate split-half Reliability when variances for the two half of the test are unequal 130 Appendix 4.2: the calculation of reliability Using KR20 130 5 validity 133 Deï¬ ning validity 135 Aspects of validity 135 Façade validity 135-corrected Efficacy tests 136 Evidence Criterion-related for validity 137 Evidence Construct-related for validity 147 relationship between reliability and validity 154 Summary 155 Summary 6 of writing and evaluation of test articles 157 Writing Article 158 Formats Item 159 168 Other possibilities Article Analysis 170 Article Diï¬ coltà 171 discriminabilità 172 Element images Features 174 Non-common link Measures 180 articles in test on reference criteria 181 Limitations Analyse object 182 Summary 7 183 Administrative test 185 I Examiner and Subject 186 report between Examiner and test Taker the run of the tester 187 Test language T Aker 188 Interview Surgery 295 Object Variables 196 Behavioral Assessment Methodology 197 Reactivity 197 Drift 198 planned for 198 Deception 198 Statistic Review of Error 199 Summary PA II 8 200 APPLICATION to IONI interview techniques 201 The problem of Deï¬ ning Intelligence 230 Bineta s Principles of Construction Test 232 Principle 1: Age Diï¬ erentiation 232 Principle 2: General Capacity Mental 233 Spearmanâ s Model of General Capacity Mentale 233 Implications of General Mentale Intelligence (g) g f-gc intelligence theory 234 The Early Binet Scales 235 1905 Binet-Simon Scale 235 1908 Scale 236 Termanâ s Stanford-Binet intelligence Scale 238 of 1916 Stanford-Binet Intelligence Scale 238 The intelligence quotient (QI) 238 1937 Scale 240 1960 Stanford-Binet Review and deviation IQ (SB-LM) 241 The Modern Binet Scale 242 Model for the fourth and fifth editions of the Binet Scale Features of the 1986 Revision 243 Features of the fifth edition 245 psychometric properties of the fifth edition 2003 247III Validity median 248 The minutes under test 254 first scores, climbing scores, and VIQ 258 Performance under test 259 INDEX Representation QI 261 FSIQs 261 Scores 261 interpretative characteristics of the Wechsler Prove 262 verbal-performance IQ compare pattern Analysis Studies 262 hypothetical case 263 262 psychometric properties of the 265 Wechsler Adult Scala 265 Standardization 265270 Summary 11 271 Other Individual Tests of IV-WISS IV-WISC Survey 376 survey 3 the 414 Behavior Scale 414 Irrational Credentials 415 Cognitive Functional Analysis 415 Psychological Procedures 417 Physiological Variables with Implications for Treatment Assessment of Psychological Techniques 418 417 Computers and Psychological Test 419 Psychological Interview 419 MAIN PROVENTION 419 Computer administered Test 420 Computer diagnosis, Score and Reporting of Results 421 The strong emptiness professional interest 435 Campbell 436 Inventory of Interests Investigation on interest and expertise Campbell 438 The emergency of the strong inventory of interests 439 the survey of interest of work of Kuder 440 Career assessment inventory 445 Auto-direct research 445 Elimination of gender Bias in Inter Measure EST 446 Attitudes and interests 447 Measurement Personal characteristics for job placement 447 Approach factor: Are the professional sizes of Osipow there stable personality traits? 477 Summary 17 448 449 Health test Psychology and health care 451 Neuropsychology 452 Clinical neuropsychology 452 Clinical neuropsychology 457 Neuropsychology development 457 Neuropsychology adult 460 California Minute learning test (CVLT) 466 Automated neuropsychological test 469 Anxiety and stress Assessment 470 Stress and Ansia 470 What is the quality of life related to health? 478 Common Methods for Life Quality Measurement 479 Summary 18 481 Testing in Industrial and Business Settings 483 Personnel Psychology - Employee Selection 484 Employment Interview 484 Basic rates and credit rates 485 Taylor-Russell Table 489 Utility Theory and Decision Analysis Increase Validity 496 495 Employee Perspective Perspective Psychology: assembly of people at Jobs 498 The Myers-Briggs 498 Test type indicator for use in industry: personnel test (WPT) 499 Work setting measuring characteristics 500 Social-ecology approach 501 Environmental classification 502 Work analysis 504 Individual measurement - Situation Interaction 507 Summary 509 PA RtiiIsuesues 19 Test Bias 511 Why is the Bias test controversial? 512 Equity Test and Law 513 The traditional defense of testing 517 content-related evidence for validity 517 sources related to the validity criterion 520 other approaches to testing members of the minority group 524 Ignorance against stupidity 524 Chitling test 526 chitling test 526 black intelligence test of cultural homogeneity 527 The multicultural pluralistic evaluation system 528 XV XVI Content Tips for solutions 532 Ethics and deï¬- nition of tests Bias 532 Thinking Diï¬ Erentemente: Find new interpretations of data Develop Diï¬ 537 535 Erent Criteria 540 Summary 542 20 Test and Law 545 Read that govern the use of tests 546 Federal Authorities 546 Specific laws 555 No child left behind (NCLB) 557 Main causes that have Aï¬ psychological test € 558 Cases of desegregation in advance 558 Stel VelUniversity of California v. Bakke 571 Golden Reganisa Insurance Company et al. v. Washburn et al. 571 Adarand Costructors, Inc. v. Pena, secretary of transport et al. 572 Actions AÃ¯â¬ "in higher education 572 of silence v. Bollinger and Gratz v. Bollinger 574 parents v. Seattle 574 Meredith v. JeÃ¯â¬ â,¬ Erson County Coard of Education 576 Personal staff 576 cases covered by the Americans with Disabilities Act (ADA) 581 A critical look at Lawsuits Summary 582 21 582 Ethics and future of Psychological Testing 585 Issues Shaping the field of experimentation 586 Professional Issues 586 Moral Issues Social Issues 591 595 597 The current trends proliferation of new tests 597 Higher Standards, improved technology, and to increase public awareness and objectivity 598 599 CONTENTS InÃ¯Â¬ influence the computerization of trial tests on the Internet xvii 600 600 601 future Trends outlook for tests are promising 601 proliferation of new and improved test'll continue 602 controversy, disagreement, and change will continue 603 integration of Cognitive Science and Computer Science will bring a number of innovations in the test 603 Summary 603 APPENDIX 1 Areas of a standard normal distribution 605 APPENDIX 2 r critical values for D = D = .05 and .01 (two test code) 609 APPENDIX 3 valor the critical t * 610 APPENDIX 4 code of fair practices Testing in education 612 Glossary 617 Bibliography 622 Index name 689 subjec T IND EX 703 List of test sample profiles Figure 9.7 cover of the Stanford-Binet Scale Intelligence 244 Figure 12.1 Example of a score report for the Stanford Achievement Test 311 FIGURE 12.2 a sample student profile from the verbal sample capacity voices ACT FIGURE 12.3 GRE 319 FIGURE 12.4 GRE quantitative elements of the ability sample 320 FIGURE elements of the sample 12.5 MAT 323 Figure 13.2 a MMPI sheet profile 341 FIGURE 13.3 A MMPI-2 sheet profile 348 FIGURE 13.4 Jackson personality inventory profile sheet 359 FIGURE 13.5 NEO profile sheet Personality inventory 363 TA BLE 14.1 MIRATA EXAMPLE 14.2 317 Synthesis of Rorschach scoring 382 The danger of basing Rorschach interpretations of tests insufficient 388A 389 activities to complete the sentence TA BLE 398 FIGURE 17.4 17.4 Prof. ilo of a patient tested with the battery Luria-Nebraska 465 Some of the questions used in the test anxiety Questionnaire 473 FI GURE 18.2 Examples of questions from Wonderlic 500 FIGURE 19.8 profile Sompa Example 531 TA BLE 20.1 Examples of items to a minimum competency test 570 PR eFace P sychology is a large, exciting Ã¯Â¬ field. Psychologists work in environments ranging from schools and clinics to laboratories, basic pharmaceutical Ã¯Â¬ rms, and international private companies. Despite this diversity, all psychologists have at least two things in common: all study behavior, and all depend to some extent on its measurement. This book relates to a particular type of measurement, psychological tests, such as measuring characteristics relating to all aspects of the behavior of humans. Psychological Testing is the result of a longstanding collaboration between the authors. As active participants in the development and use of psychological tests, we have become discouraged because © too many college students graduating display psychological test courses as boring and unrelated to their objectives or career interests. On the contrary, we see psychological tests as an exciting Ã¯Â¬ field. It has a solid place in the history of psychology, but it's constantly Ã¯Â¬ UX due to challenges, new developments and controversies. A book about the tests should encourage, not dampen, the interest of a Studenta ¢ s. So, we provide an overview of the many facets of psychological test and measurement principles in a way that caters to the contemporary university student. To understand the applications and psychological testing issues, the student must learn some basics, which requires some knowledge of introductory statistics. As a result, some review and a careful reading of Part I will open the way for the understanding of test applications discussed in Part II. III examines the problems now shape the future of the test. These issues include test anxiety, trial prejudices, and the interface between the test and the law. The future of applied psychology can depend on the ability of address these challenging issues. Throughout the book, we present a series of discussions focused and focused examples. These sections illustrate the material in the book through examples or xix xx PREFACE provide a more detailed discussion of a particular problem. We also use the characteristics of the designated box â Psychological Testing in Everyday lifeâ to demonstrate materials such statistical calculations. EMPHASIS ON APPLICATION INCRESA Students today often favor informal and relevant personal examples discussions. As a result, we decided to use models from various fields and to write in an informal style. However, since the © test is a serious and complicated field in which there are also major disagreements between scholars and experts, we have dealt with the controversial aspects of the test with a more formal discussion and a detailed arbitration. The first edition of Psychology Testing: Principles, Applications and Issues was published in 1982. In over a quarter of a century since the text was introduced, the world has changed in many ways. For instance, personal computers are new in 1982. The majority of students and professors had never heard of e-mail or the Internet. There were well under psychological testing applications that are there today. On the other hand, the principles of psychological testing have remained relatively constant. So, the new editions have included improvements and enhancements in the Principles chapters. The chapters on the applications and the problems have evolved considerably. Not only is the field of psychological tests has changed, but also the authors. One of us (RMK) has spent most of his career as a professor at a medical school and is now a public health school. The other (DPS) has completed law school and works as a professor of psychology and adjunct professor of law. While maintaining our core identity as psychologists, we also had the opportunity to explore the cutting-edge practice in medicine, public health, education and law. The seventh edition goes beyond any previous edition in shaping psychological test applications in a wide range of applied fields. In the development of the seventh edition, we organized around the topics application areas. Chapter 11 considers the psychological tests in education and special education. Chapter 12 examines the use of standardized tests in education, the civil service and the army. Chapters 13:14 consider the use of psychological testing in clinical settings and counseling. The age of computers has revolutionized the psychological tests. We deal with some of these issues in the Principles chapters discussing computer adaptation and testing of response theory of the elements. In Chapter 15, we discuss applications of psychological science in the age of computers. Chapter 16 discusses the use of psychological tests in psychology consulting and focuses primarily on the interests of inventors. Chapter 17 explores the rapidly developing fields of psychological assessment in health psychology, medicine and health care. Chapter 18 examines the psychological tests in industry and in corporate settings. The final chapters on psychological testing issues have been extensively updated to reflect new developments in social justice, law and ethics. The first edition of Psychological Testing has been produced on typewriters before word processors were commonly used. At the time, a few professors and students had access to private computers. The first editions of the book have offered instructions PREFACE xxi preparing the presentation of analysis within the mainframe computer statistics. As recently the production of the third edition, the Internet has been largely untapped by students Today, almost all students have ready access to the Internet and World Wide Web, and now we usually provide references to websites. We also provide a greater discussion of computer-administrated tests. ORGANIZATION OF THE EDITION OF SEVENTS: WARNING WARNINGPER PIANIFICATION Producing seven editions of the Psychological test over more than 25 years has been stimulating and rewarding. We are honored that hundreds of professors have adopted our text and which is now used in hundreds of colleges and universities worldwide. However, some professors suggested that we reorganized the book to facilitate their approach to the class. To host the great variety of approaches, we tried to keep chapters fairly independent for professors to teach them in any order they chose. For example, an approach to the course is to go systematically through the chapter sequence. The professors who want to emphasize psychometric problems, however, could assign chapters from 1 to 7, followed by chapters 19 and 20. Then, some chapters may return from the Applications section. On campuses that require a strong course statistics as a prerequisite, chapters 2 and 3 can be eliminated. Professors who emphasize applications could assign chapters from 1 to 5 and then proceed directly to Part II, with some professors assigning only some of its chapters. Although Chapters 9 to 13 are the most likely to be used in a basic course, we have found a sufficient interest in chapters 14 to 18 to hold them. Chapters 17 and 18 are more recent areas, where psychological testing is expanding. Finally, chapters 19 and 20 were written so that they could be assigned either at the end of the course or near the beginning. For example, some professors prefer chapters to assign 19 and 20 after Chapter 5. Beyond Compare As with previous editions, a student workbook is available. The professors have access to the manual an instructorâ s and a bank of electronic test elements. Book Companion Website Website The Web contains several components that will be valuable to instructors. First, a set of data consisting of 25 exams on different sizes can be downloaded and used with accompaniment of reliability and validity exercises. Secondly, many additional assignments including a report on a psychological test battery, an evaluation of a simulation test manual, and a critique of associated testing and classification rubrics will be published on the website. The additional allocation of ï¬ les and classification headings are amended, which allows for modifications so that it improves your specific objectives. xxii PREFATION Student Workbook (ISBN 0-495-59774-0) More than a traditional study guide, the Workbook student written by Katherine Nicolai of Rockhurst University really helps students understand the connections between abstract measurement concepts and development, evaluation, selection and use of psychological tests in the real world. The Student Workbook contains interesting hands-on exercises and tasks, including case studies for criticism, test proï¬ les to interpret, and studies on psychometric properties of tests to evaluate. Of course, the Student Workbook also contains traditional features, such as chapter contours and multi-choice quiz practice. Better than all, the workbook is presented in a three-ring collector where students can keep other course notes and dispenses. Students will discover that the Student Workbook will help them organize their study of the text of Kaplan and Saccuzzoâ s and Excel on the course exams, assignments and projects! Instructorâ s Resource Manual test Bank (ISBN: 0-495-50914-0) / The Instructorâ s Resource Manual (IRM) was written by Katherine Nicolai of Rockhurst University, and the Test Bank of Ira Bernstein and Yasushi Kyutoku of the University of Texas in Arlington. In an easy-to-use three-ring ligand, the IRMA stuo of resources, including course design guides, the use of psychological tests in the classroom, the use of measuring student test data teach, suggested class class use, and manifestations, activities, and lessons For activities. IRM provides a description of supplementary supplementary assignments On the Companion of the book website and gives the unique projective fantastic instructors and more. The test database contains more than 750 multiple choice questions as well as many applications for many rates Ã ¢ â ¬. "Acknowledgments We are highly indebted to the many reviewers and professors who provided feedback on the sixth or the seventh edition revised. SPECIAL THANKS GO TO THIS EDITION CONDITIONS, including: Virginia Allen, Idaho State University, David Bush, Utah State University, Ira Bernstein, University of Texas, Arlington; JeÃ¯â¬ â ¬ Conte, San Diego State University; Imogen Hall, University of Windsor; Maureen Hannah, Siena College, Ronald McLaughlin, Juniata College, Michael Mills, Loyola Marymount University; Philip Moberg, University of Akron; MJ Monnot, Central Michigan University; Jennifer NEEMANN, University of Baltimore; Karen Obremski Brandon, University of South Florida; Frederick Oswald, Michigan State University; St. Mark Pancer, Wilfrid Laurier University; Christopher Ralston, Iowa State University; Sharon Rostosky, University of Kentucky; Stefan Schulenberg, University of Mississippi; Theresa Sparks, Clayton State University; Chockalingam Viswesvaran, International University of Florida; Mark Wagner, Wagner College; and Nancy Zook, purchase of sunshine. The seven editions of this book have been developed under six editors events diÃ¯â¬ â ¬ Wadsworth. Previous editions benefit of the patient and inspired the introduction of the XXIII preface supervision of Todd Lueders, C. Deborah Laughton, Phil Curson, Marianne TaÃ¯Â¬, Inger and Jim Brace-Thompson. We are most appreciative of the support we have received from our new publisher, Jaime Perkins. He was patient, helpful and very well organized in directing the development of the current edition. Each of our editors arrived to the task with a personality of eressiva Diaa and a set of in-depth insights. We have learned immensely from each other and the seventh edition is a collection of what we have gained from advice and consultations for many years. We also want to thank Charlene Carpentier, Production Project Manager, Vernon Boes for the coordination of the cover and Wilson Co., Editorial Assistant for the Coordination of supplements. We want to give special thanks to Kate Nicolai to create the exciting new workbook for students and manual greatly expanded instructor resources. Special Thanks Go to Danielle Zuest and Nancy E. Johnson. Danielle has conducted much of the research for the renovation of about half of the chapters and has participated in numerous details. Nancy is also assisted in many ways, including research, editing and localization of diÃ¯Ã¯Â¬cult-to-Ã¯Â¬ sources. Without these two individuals, the publication of this edition would have been much delayed. Dr. Kaplan is also grateful to the Rockefeller Foundation, which provided a quiet work space with a view of Lake Como in Italy to allow completion of Ã¯Â¬ali phases of this project. Robert M. Kaplan Dennis P. Saccuzzo April 2008 This page intentionally left blank on the authors is Robert M. Kaplan Fred W. and Pamela K. Wasserman, Professor and Chair of the Department of health services at UCLA and a professor of medicine at UCLA DAVID GeÃ¯Â¬ â ¬ It School of Medicine. From 1997 to 2004, he was professor and chairman of the Department of Family and Preventive Medicine at the University of California, San Diego. He is a former president of several organizations, including the American Psychological Association Division of Health Psychology, Section J of the American Association for the Advancement of Science (PaciÃ¯Â¬ c), the International Society for the quality of research on life, society for behavioral Medicine and behavior research Academy on behavioral medicine. It's a Council of the Behavioral Science of the American Toracic Society. Dr. Kaplan is currently editor-in-head of health psychology and is the former head editor of the annals of behavioral medicine. He served as an associate editor of the American psychologist, American, advisory editor of four other academic journals. Selected additional honors include APA Division of Health Psychology Annual Award for Best Scientiï¬ c Contribute (for juniors scholar 1987 and again for an elderly scholar 2001), SDSU Distinguished Research Doctrine, 1988, and the Health Net Distinguished Lecturer Award in 1991, University of California 125 Anniversary Award for Most illustrious Quality Alnus Medical School In 2006, he received the Distinguished Mentor Research Award from the Society of Behavioral Medicine. Its public service contributions include various NIH, AHRQ, and VA relay groups, local American Lung Association (ALA) Board of Directors, and the regional research committee for the American Heart Association. He served as co-chairman of the XXV xxvi Behavioral Medicine Task Force and the Institute of Medicine (IOM) National Academy of Sciences Committee on Health and Behavioral Behavioral Medicine. He also served as part of the panel of experts for the public action plan advisory policy CDC-NIH to prevent heart disease and stroke. Kaplan is currently part of the National Advisory Committee for the Decade of Conduct. In addition, he is the chairman of the committee responsible for the NHLBI National Enfisema Treatment Trial (net). Dr. Kaplan is author or co-author of more than 15 books and over 400 articles or chapters. The ISI includes it in the list of the most mentioned authors in the world (deed as above 99.5th percentile). In 2005, he was elected to the National Academies of Sciences Medical Institute. DENNIS P. Saccuzzo is a professor of psychology at San Diego State University, and a professor of law at Thomas Jeï¬ Erson School of Law. He has been a scholar and practitioner of psychological testing for over 32 years and has numerous peer-reviewed publications and professional presentations in the field ï¬. Dr. Saccuzzo's research was supported by the National Science Foundation, the National Institutes of Mental Health, the National Institutes of Health, the US Department of Education, the Scottish Rite Foundation, and the US services. He is also a licensed California psychologist and a California lawyer. He is a certain board and in clinical psychology by the Board of Professional Psychology (ABPP). In addition, he is a Diplomat of the Board of Assessment Psychology (ABAP) and the American Legal Medicine Council. He is a member of the American Psychological Association, American Psychological Society and Western Psychological Association for outstanding and unusual contributions to the ï¬ field of psychology. Dr. Saccuzzo is author or co-author of over 300 publications and peerreviewed publications, including eight textbooks and 20 law manuals. He is the president and co-founder of applications of law psychology, Inc., an educational institution that applies psychological science to legal aspects and law study. Seventh psychological edition Test Principles, Applications and Problems Page intentionally left blank CHAPTER 1 Introduction AR NINGOBJ CE TYPE S After completing this chapter, you should be able to: to define the basic terms of psychological and educational tests is to distinguish between a single test and a group test to define the terms realization, attitude and intelligence and identify a concept that can understand all three termsdistinguish between skill tests and personality testdefine the term structured personality test ■ explain how structured personality tests differ from design personality tests ■ explain which is a normative sample or standardization and why this sample is important ■ identify the main developments in psychological testing history ■ explain the relevance of psychological tests in contemporary society 1 2 cappter 1 ■ introduction y u are seated at a table. You just got a fingerprint and showed an identity document. Look around and see 40 nervous people. a powerful-looking test proctor with a chronometer comes out of the booklets. you are warned not to open the book until he said to do it; face possible disciplinary action if disobey. is not a nightmare or a futuristic fantasy, this is real. Finally, after what looks like an eternity, you are told to open your booklet on page 3 and start working. your mouth is dry; Your palms are wet. open to page 3. you have 10 minutes to solve a problem of five parts based on the following information.1 a car guide in the central ring of a circus and exactly eight clowns—Q , r, s, t, v, w, y and Z— ocire from the car, a clown at a time. the order in which the clowns leave the car is consistent with the following conditions: v comes out at a certain time before it is y that q. q comes out at a moment after z. t comes out at a certain time before v, but at a certain time after r. s comes out at some time after v. r comes out at some time before question 1. if q is the clown fi fth to dare from the car, then each of the following could be true except: z is the first clown to dare from the car. t is the second clown to dare from the car. v is the third clown to dare from the car. w is the fourth clown to dare from the car. y is the sixth clown ocire from the car. not quite sure how to proceed, you look at the next question. Question 2. If r is the second clown to dare from the car, what of the following should be true? he comes out of the car at a time before t. t. t. comes out of the car at a certain time before w. w. comes out of the car at a certain time before he does. You get out of the car at a certain time before you do. z comes out of the car at a certain time before w. your heart beats a little faster and your mind begins to freeze as a computer overloaded with too little work memory. look at your watch and note that 2 minutes have passed and still do not have your bearings. The person sitting next to you seems a little weak. another three files on someone storms up to the test proctor and complains frantically that you can not do this type of problem. While the proctor struggles to calm this person, another makes a crazy dash for the bathroom. Welcome to the world of "high pole", standardized psychological tests in the xxi century. the questions you just faced were real problems from 1 used per permit by law school admission test, October 2002. (d) the answer to question 2 is e. Chapter 1 ■ Introduction 3 a past version of the LSAT — the admission test of the school of jurisprudence. if a student is admitted or not in the law school in the United States is almost entirely determined by the score of that person on the average of the degree and lsat point. Thus, the future can depend enormously on a single score from a single test given in one morning or in the afternoon in tension. Despite efforts to improve tests such as lsat to increase diversity (pashley, thornton, & duff y, 2005,) standardized tests tend to disadvantage women and ethnic minorities (sackett, schmitt, ellingson, & kabin, 2001.) similar problems appear on the gre — the graduated exam record, a test that plays an important role in determining who arrives to study at the degree level inAll right. (Later in this book we talk about how to prepare for such tests and what their meaning, or predictive validity, is.) tests such as lsat and gre areDiï¬cult modern psychological tests. The scenes we described are real; Some careers lead on a single test. Maybe you already got the GRE or LSAT. Or maybe you haven't graduated yet, but you're thinking about applying for an advanced diploma or a professional program and will soon face the GRE, LSAT or MCat (Medical College Admission Test). Clearly, it will help you to have a basic understanding of the multitude of psychological tests that people are invited to take throughout their lives. From our birth, the tests have an important inï¬ on our lives. When the pediatrician hits the palms of our hands and the soles of our feet, he or she is running a test. When we get into school, the tests decide whether we pass or not pass lessons. Tests can determine whether we need special education. In the United States and many industrialized countries test skills determine whether students will graduate at high school (Carnoy, 2005; Hursh, 2005). Other tests determine which college we could participate. And, of course, when we get into college, we're even more testing. After graduation, those who choose to avoid tests like the GRE may have to take tests to determine where it will work. In the modern world, most of the life and success of all depends on the results of the tests. In fact, the tests also have an international construction site. For example, 15-year-olds in 32 countries have received problems as follows from the organization for economic and development cooperation (OECD) and the International Student Assessment Programme (PISA) (Schleicher & Tamassia, 2000): A result of global warming is that ice from some glaciers is melting. Twelve years after the ice disappears, small plants, called Lichen, begin to grow on the rocks. Each lichen grows roughly in the shape of a circle. The relationship between the diameter of the circles and the age of the lichen can be approximated with the formula: D = 7,0 â- the square root of (t Â "12) for any T not exceeding 12, where D represents The diameter of the lichen in millimeters and t represents the number of years after the ice disappeared. Calculate the diameter of the lichen 16 years after the ice disappeared. The complete and correct answer is: D = 7.0 Â- the square root of (16 â '12 mm) D = 7.0 Â- the square root of 4 mm D = 14 mm 4 chapter 1 â- introduction International Mathematical literacy scores Brazil Mexico Luxembourg Greece Portugal Italy Latvia Poland Spain Russia Hungary Germany USA Rep. Norway Ireland Sweden Liechtenstein Iceland Denmark Austria France Belgium Uk Switzerland Canada Australia Finland New Zealand Japan 300 400 450 45 (Statistics used by OECD and PISA permission. Figure courtesy of W. J. Koen.) Eighteen countries classified above the United States in the percentage of 15 years which had mastered such concepts (see Figure 1.1). The results were similar for a literacy test of OECD science (see Figure 1.2), which had questions as follows: a bus moves along a straight stretch of road. The bus driver, the name of the bus, has a cup of water resting in a stand on the dashboard. All of a sudden, Ray has to hit the brakes. What is more likely to happen to the water in the cup immediately after Ray Slams on the brakes? A. The water will remain horizontal. B. Water will pour sideways 1. Chapter 1 Â- 5 Introduction International Scientific literacy scores Brazil Mexico Luxembourg Portugal Latvia Russia Greece Liechtenstein Italy Denmark Poland Germany Spain Switzerland Belgium Iceland Hungary USA Norway Czech France. REP. Sweden Ireland Austria Australia New Zealand Canada U.K.Japan Korea 300 350 400 450 450 500 550 600 points F I G u R and 1.2 Approximate average scores of 15-year students on the OECD scientific literacy test. (Statistics used by OECD and PISA permit. Figure figure by W. J. Koen.) C. The water will pour on side 2. D. The water will be poured but it cannot be said if it is poured on side 1 or side 2. The correct answer is C. How useful are the tests like these? Do they measure anything significant? How accurate are they? These issues concern not only every citizen of the United States, but also all members of the highly competitive international community. To respond to them, you need to understand the principles of psychological testing that you are going to learn. To answer test questions, you need to understand the concepts presented in this book, such as reliability, validity, article analysis and test construction. A full 6 CHAPTER 1 ■ Understanding these concepts will require careful study and knowledge of basic statistics, but your efforts will be rewarded richly. When you finish this book, you'll be a better test consumer. BASE CONCETTI You probably already know some basic concepts of psychological testing. For clarity, however, we will begin with the definitions of the most fundamental terms so as to know how they are used in this textbook. What a test is Everyone has experienced with tests. A test is a measuring device or a technique used to quantify behavior or help in understanding and predicting behavior. An orthography test, for example, measures how well someone writes or the extent to which someone learned to write a specific list of words. At some time in the coming weeks, your instructor will probably want to measure how well you have learned the material in this book. To achieve this, your instructor can give you a test. As you well know, the test your instructor gives can not measure your full understanding of the material. This is because a test measures only a sample of behavior, and the error is always associated with a sampling process. Test scores are not perfect measurements of behavior or a feature, but significantly add to the forecasting process, as you will see. An article is a specific stimulus to which a person responds excessively; This answer can be marked or evaluated (e.g., classified, scaled, or counted). Since psychological and educational tests are made up of elements, the data they produce is explicit and therefore subject to scientific investigation. In simple terms, the elements are the specific questions or problems that make up a test. The problems presented at the beginning of this chapter are examples of evidence. The excessive response would be of fil in or blacken one of the spaces: A B C D E F G A psychological test or an educational test is a set of elements that are designed to measure the characteristics of human beings concerning behavior. There are many types of behavior. Excessive behavior is the observable activity of an individual. Some psychological tests attempt to measure the extent to which someone might engage or "put" a particular overt behavior. Other tests measure how much a person has previously engaged in some excessive behavior. Behaviour can also be covered — that is, it happens within an individual and cannot be observed directly. For example, your feelings and thoughts are types of hidden behavior. Some tests attempt to measure such behavior. Psychological and educational tests thus measure past or current behavior. Some also try to predict future behavior, such as success in college or an advanced degree program. What does it mean when someone gets 75 correct articles on a 100-item test? One thing that means, of course, is that 75% of the elements have been answered correctly. CHAPTER 1 ■ Introduction 7 In manyHowever, knowing the percentage of correct elements a person obtained can be misleading. Consider two extreme examples. In one case, out of 100 students who took the exam, 99 had 90% correct or higher, and 1 had 75% correct. In another case, 99 of the 100 students had a score of 25% or less, while 1 had 75% correct. The meaning of scoresradically change, depending on how a sample Ned well deï¬ of individuals scores on a test. In the case before ï¬, a score of 75% is poor because it is in the lower part of the distribution; In the second case, 75% is actually a higher score. In order to cope with these interpretation problems, psychologists make use of scales, which refer crude scores on evidence elements in a certain theoretical or empirical distribution of Ned. Later in the book you will learn about such distributions. Test scores can be related to strokes, which are bearing features or tendency to respond in a certain way. Determination A, Sometimes as stubborn A, A is an example of a stroke; Â shynessâ is another. Test scores can also be related to the state or condition c speciï¬ or the state of an individual. An individual determined after many obstacles can, for example, be in a state of weakness and therefore be less prone than usual for manifest determination. Tests measure many types of behavior. Types of tests As well as there are many types of behavior, so there are many types of tests. Those who can be given to one person at a time are known as single tests (Figure 1.3). The examiner or test administrator (the person who gives the test) gives the test to one person at a time, so that psychotherapists see only one person at a time. A group test, on the contrary, can be administered to more than one person at a time by a single examiner, for example when an instructor gives everyone in class a test at the same time. You can also classify evidence according to the type of behavior they measure. skill tests contain elements that can be marked in terms of speed, accuracy, or both. On a skill test, faster or more accurate your answers, the best your scores on a particular feature. The most algebra problems can be solved correctly in a given period of time, the high is the score in the ability to solve such problems. Historically, experts have distinguished between realization, attitude and intelligence as different types of capacity. Realization concerns previous learning. A test that measures or evaluates as a number of words you can write correctly is called a test of spelling realization. Attitude, however, refers to the potential for learning or acquiring a specific skill. A test measures aptitude spelling as many words you could be able to specify given a certain amount of training, education and experience. Your musical attitude refers in part to the way you could be able to learn how to play a musical instrument given a number of lessons. Traditionally distinct from achievement and attitude, intelligence refers to a general potential character to solve problems, adapt to changing circumstances, abstract thinking, and experience. When you say a person is a smart, we usually refer to intelligence. When a father scolds his daughter because he did not do as well at school as she can, he most likely believes that she did not use her intelligence (general potential) to get (get new knowledge). The distinctions between realization, attitude, and intelligence are not always so cut-and-dry, because all three are highly interconnected. Attempts to separate previous learning from learning potential, for example, did not succeed. In view of 8 Chapter 1 an introduction Figure 1.3 A single test administration. (Ann Chwatsky / Jeroboam.) The considerable overlap of realization, attitude, and intelligence tests, all three concepts were understood by the human term ability.a clear distinction between skill tests and personality tests. While skill tests are related to capacity or potential, personality tests are related to the obvious and secret provisions of individualâ for example, a person's tendency to show a particular behavior or response in a given situation. Staying isolated from others, for example, requires no special skillability, but some people generally prefer or tend to remain so isolated. of personality tests measure typical behavior. There are different types of personality tests. In Chapter 13, you will learn about structured, or objective, of personality tests. of structured personality tests provide a statement, usually self-reportÃ type ¢ Ã ¢, and require the subject to choose between two or more alternative responses as TrueÃ ¢ Ã ¢ or Ã ¢ ¢ FalseÃ (see Figure 1.4). In contrast with the structured personality tests, personality of projective tests they are structured. In a personality projective test, both the stimulus (test materials) or responseÃ ¢ required or both,Äidon't are ambiguous. For example, in the highly controversial Rorschach test, the stimulus is an ink stain. Also, instead of being asked to choose between alternative answers, as in the structured personality tests, the individual is asked to provide a spontaneous response. The ink stain is presented to the subject, who is asked, what would it be? It projective tests assume that a character's interpretation of an ambiguous stimulus will reÃ¯Â¬ ect of her or her unique characteristics (see Chapter 14). CHAPTER 1 introduction 1. True False 9 I like heavy metal music. 2. I believe that honesty is the best policy. 3. I am in good health. 4. I am easily fatigued. 5. I sleep well at night. FIGURE TA B L E evidence 1.4 Self-Report. 1.1 Types of skill tests Tests .: measurement skills in terms of speed, accuracy, or both. A. Implementation: Learning from previous measures. B. Aptitude: potential measures for the acquisition of a speciÃ¯â¬ c skills. C. Intelligence: Possible measures to solve problems, adapt to changing circumstances, and t proÃ¯Â¬ experience. II. of personality tests: the traits of typical measurement behaviorÃ ¢, temperaments, and dispositions. A. Structured (objective): Provides a self-report statement in which the person responds TrueÃ ¢ Ã ¢ or False ¢ Ã, Ã Ã Yesa or No.A B. Projective: provides an unambiguous test stimulus; Response requirements are not clear. See Table 1.1 for a brief overview of the skills and personality tests. psychological testing refers to all possible uses, applications and basic concepts of psychological and educational testing. The main use of these tests, however, is to measure the individual erence diÃ¯Â¬ or variations among individuals. Such tests measure individual erence diÃ¯Â¬ of ability and personality and assume that the diÃ¯Â¬ erence displayed on trial reÃ¯Â¬ ect diÃ¯Â¬ actual erence between individuals. For example, individuals who score high on intelligence tests are assumed to have a higher degree of intelligence of those who get low scores. Therefore, the main purpose of the test is diÃ¯Â¬ erentiate among those taking the tests. We will discuss the idea of individual diÃ¯Â¬ ering later in this chapter. SUMMARY OF THE BOOK This book is divided into three parts: Principles, Applications and problems. Together, these parts cover psychological tests from the most basic to the most complex ideas. Basic ideas and events are introduced early and stressed during 10 CHAPTER 1 introduction to strengthen what you just learned. In covering the principles, applications and themes, we aim to provide not only whoa S of psychological tests, but also the Howa whyâ ¢ s and s of important developments in the field Ã¯Â¬. We also address a major concern of many studentsÃ relevanceÃ ¢ ¢ examining different test uses and to data. psychological principles for the testing according to principles of psychological tests means the basic concepts and fundamental ideas that underlie all the psychological and educational testing. Sections 2 and 3 present statistical concepts that provide the foundation for understanding test. Chapters 4 and 5 cover two of the most basic test concepts: reliability and validity. Reliability is precision, reliability, consistency, or repeatability of test results. In more technical terms, reliability refers todegree to which the test scores are without measuring errors. As you learn, there are many ways a test can be reliable. For example, test results can be reliable over time, which means that when the same test is given twice within a given time interval, results tend to be the same or very similar. Validity refers to the meaning and utility of test results. More specifically, validity refers to the degree in which a certain inference or interpretation based on a test is appropriate. When the question arises, "What does this psychological test measure do?" he essentially asks, "Why is this test valid for inference?" Another principle of psychological testing concerns how a test is created or built. In chapter 6, we present the principles of testing construction. The act of giving a test is known as test administration, which is the main theme of chapter 7. Although some tests are easy to administer, others should be administered in a highly specific way. The final chapter of Part I covers the foundations of the administration of a psychological test. Part II Psychological Test Applications, on Applications, provides a detailed analysis of many of the most popular tests and how they are used or applied. Start with an overview of the essential terms and concepts that refer to the test application. Chapter 8 talks about interview techniques. An interview is a method to collect information through verbal interaction, such as direct questions. Not only the interview traditionally served as an important technique of collecting psychological information in general, but also the data from interviews provide an important complement to test results. Chapters 9 and 10 cover individual human capacity tests. In these chapters, you will learn not only about the tests but also about the theories of intelligence that submit them. In Chapter 11, we take care of educational tests with the emphasis on special education. In chapter 12, we present group tests of human abilities. Chapter 13 covers structured personality tests and Chapter 14 covers projective personality tests. In chapter 15, we talk about the important role of computers in the field of testing. We also consider the influence of cognitive psychology, which today is the most important of the various schools of thought within psychology (Kellogg, 2003; Leahy & Dowd, 2002; Weinstein & Way, 2003). These chapters not only provide descriptive information, but also deepen the ideas below the various tests. Chapter 16 examines the test of interest, which Chapter 1 ■ Introduction 11 measures the behaviour of such factors as occupational preferences. Chapter 17 examines the relatively new area of medical tests for brain damage and health. It also covers important recent advances in neuropsychology of development. Finally, Chapter 18 covers tests for industrial and organizational psychology and business. Numerous social and theoretical issues, such as the controversial theme of racial differences in capacity, accompany the tests. Part III covers many of these problems. As a compromise between width and depth of coverage, we focus on a complete discussion of those issues that are of particular importance in the current professional, social and political environment. Chapter 19 examines test data, one of the most volatile issues in the industry (Geisinger, 2003; Reynolds & Ramsay, 2003; Ryan & DeMark, 2002). Since psychological tests have been accused of being discriminatory or prejudice against certain groups, this chapter carefully examines both sides of the subject. Due to the accusations of prejudice and other problems, psychological tests are increasingly under controlLaw (Phillips, 2002; saccuzzo, 1999). Chapter 20 examines test issues related to legal issues and discusses tests and law. Chapter 21 presents a general overview of other great issues that currently define the future of psychology psychology In the United States with an emphasis on ethics. From our problems review, speculate even on what the future holds for psychological tests. Historical perspectives now we briefly provide the historical context of psychological tests. This discussion touches some of the materials presented previously in this chapter. The first antecedents most of the main developments in the test occurred during the last century, many of them in the United States. The origins of the tests, however, are not recent nor-Americans. The tests suggest that the Chinese had a relatively sophisticated test program for civil services more than 4000 years ago (Dubois, 1970, 1972). Every third year in China, oral exams were given to help determine job evaluations and promotion decisions. With the Han Dynasty (206 B.C.E. at 220 C.E.), the use of test batteries (two or more tests used in combination) was quite common. These first tests related to topics such as civil law, military business, agriculture, revenue and geography. The tests had become quite well developed by the Ming Dynasty (1368-1644 C.E.). During this time, a national multistage test program involved local and regional test centers equipped with special test booths. Those who have done well the tests at the local level went to provincial capital for extended essays. After this second test, those with the highest test scores went to the nation's capital for a final round. Only those who have passed this third test set have been suitable for the public office. The Western world probably learned to test programs through Chinese. British missionaries and diplomatic relationships encouraged English 12 Chapter 1 - Introduction East India Company in 1832 to copy the Chinese system as a method of selection of employees for overseas duty. As the test programs worked well for the company, the British government adopted such a test system for its civil service in 1855. After the British approval of a civil service system, the French governments and Germans have followed the cause. In 1883, the US government founded the American Civil Service Commission, which developed and administered competitive exams for some government work. The impetus of the test movement in the western world has rapidly grown at that time (Wiggins, 1973). Charles Darwin and individual differences Perhaps the most fundamental basic concept of psychological and educational tests concerns individual differences. There are no two identical snowflakes, not two equal footprints. Similarly, no two people are exactly the same in capacity and typical behavior. As we have noticed, the tests are specifically designed to measure these individual differences in capacity and personnel among people. Although humans have realized a long time ago that individuals differ, the development of tools to measure these differences was not easy. To develop a measurement device, we need to understand what we want to measure. An important step towards understanding individual differences came with the publication of the highly influential book by Charles Darwin, the origin of the species, in 1859. According to Darwin's theory, the highest forms of living have evolved partly to Cause of differences between individual life forms within a species. Given that individual members of a species differ, some possess characteristics that are more adaptive or successful in a given environment that are those of other members. Darwin also believed that those with the best or more adaptive features survive at the expense of those who are less suitable and that survivors spend their characteristics to the next generation. The He claimed, life has evolved at its levels currently complex and intelligent. Sir Francis Galton, relative of Darwin, soon started to apply Darwin's theories to the study of human beings (see figure 1.5). Considering the survival concepts of the appropriate suitable Single Erenists DiÃ¯â¬, Galton specified to show that some people possessed characteristics that made them more tough of others, a theory is articulated in its genius hereditary book, published in 1869. Galton (1883) later has A series of experimental studies started to document the validity of its position. He focused on demonstrating that individuals of ï¯â¬ exist in sensory and humane motory, such as reaction time, visual acuity, and physical strength. In this way, Galton has launched a research of knowledge regarding the individual human differences, which is today one of the most important domains of scientiÃ¯â¬ C psychology. GaltonÃ ¢ s work has been extended by the US psychologist James McKeen Catattell, who coined the term mental test (cattell, 1890). Doctoral thesis cattelÃ ¢ s He was based on GaltonÃ ¢ s work on individual emergence of ï¯â¬ in reaction time. As such, a perpetual cattell and stimulated the forces that eventually led to the development of modern tests. Experimental and psychophysical psychology of measure A second important test base can be found in experimental psychology and early attempts to unlock the mysteries of human consciousness through the C method scientiÃ¯â¬. Before psychology was practiced as a science, mathematical models Chapter 1 Figure 1.5 An introduction 13 Sir Francis Galton. (From the National Library of Medicine.) Of the mind were developed, in particular those of J. E. Herbart. Herbart eventually used these models as a base for educational theories that have strongly influenced 19/19th century educational practices. Following Herbart, E. H. Weber tried to demonstrate the existence of a psychological threshold, the minimum stimulus necessary to activate a sensory system. Then, after Weber, G. T. Fechner designed the law that the strength of a feeling grows as the logarithm of the stimulus intensity. Wilhelm Wundt, who established a laboratory at the University of Leipzig in 1879, is accredited with the foundation of the science of psychology, following the tradition of Weber and Fechner (Hearst, 1979). Wundt happened E. B. Titchner, whose students, G. Whipple, recruited L. L. Thurstone. Whipple provided the base for immense changes in the field of testing conducting a seminar at the Carnegie Institute in 1919 Thursdone participated in 1919, and Forte, and other first first-first US psychologists. From this seminar it has arrived the Carnegie Inventory interest and later the strong professional blank interest. Later in this book we will discuss more in detail the work of these pioneers and controls that have helped to develop. So, psychological tests developed by at least two survey lines: a work-based workout, galton, and cattell on the measurement of individual emrenists ofÃ¯â¬, and the other (more important and probably stronger) on the basis of Work by the German Germans Herbart, Weber, Fechner and Wundt. Experimental psychology, developed by the latter. From this work there was also the idea that the test, as an experiment, requires rigorous experimental control. This control, as you will see, is from the administration of tests under highly standardized conditions. The Orts EÃ¯â¬ of these researchers, however necessary, do not have to lead to the creation of modern psychological tests. These tests also presented in response to important needs such as classifying and identifying the mentally and emotionally handicap. One of the first tests that resemble current procedures, the Seguin Test Board module (Seguin, 1866/1907), was developed in an Ort EÃ¯â¬ to educate and 14 chapter 1 an introduction to evaluate mental disabled people. Similarly, Kraepelin (1912) developed a series of exams to evaluate emotionally compromised people. An important step forward in creating modern tests came to of the 20th century. The French Minister for Education appointed a committee to study the ways of identifying intellectually subnormal persons in order tothem with appropriate educational experiences. A member of that commission was Alfred Binet. Working in collaboration with the French doctor T. Simon, Binet developed the first major general intelligence test. Binet's initial effort launched the first systematic attempt to assess individual differences in human intelligence (see Chapter 9). The evolution of intelligence and standardized achievement tests The history and evolution of Binet's intelligence test are educational. The first version of the test, known as Binet-Simon Scale, was published in 1905. This tool contained 30 elements of increasing difficulty and was designed to identify intellectually subnormal individuals. Like all well-structured tests, the Binet-Simon Scale of 1905 was increased by a comparison or standardization sample. Binet standardization sample consisted of 50 children who had received the test in standard conditions, i.e. with the same instructions and format. In obtaining this standardization sample, Binet test authors had rules to compare the results of any new subject. Without such rules, the meaning of scores would have been difficult, if not impossible, to assess. However, knowing such things as the average number of correct answers found in the standardization sample, you could at least declare whether a new subject was below or above it. It is easy to understand the importance of a standardization sample. However, the importance of obtaining a standardization sample that represents the population for which a test will be used has sometimes been ignored or neglected by test users. For example, if a standardization sample consists of 50 white men from wealthy families, then you can not easily or enough evaluate the score of an African American girl from a poor family. However, comparisons of this type are sometimes made. Clearly, it is not appropriate to compare an individual with a group that does not have the same characteristics as the individual. Binet was aware of the importance of a standardization sample. A further development of the Binet test resulted in attempts to increase the size and representativeness of the standardization sample. A representative sample is one that includes individuals similar to those for which the test must be used. When the test is used for the general population, a representative sample must reflect all segments of the population in proportion to their actual number. In 1908, the Binet-Simon Scale had been substantially improved. It was revised to include almost double of all elements of the 1905 scale. Even more significantly, the size of the standardization sample has been increased to more than 200. The 1908 Binet-Simon Scale also determined the mental age of a child, thus introducing a historically significant concept. In simplified terms, one could think of the mental age as a measure of a child's performance on the test compared to other children CAPPTER 1 ■ Introduction 15 of that particular age group. If the test performance of a child is equal to those of the 8-year-old child, for example, then his mental age is 8. In other words, in terms of capacity measured by the test, this child can be seen as having a similar skill level as the 8-year-old child. The chronological age of the child can be 4 or 12, but in terms of test performance, the child works at the same level as the average of 8 years. The concept of mental age was one of the most important contributions of the 1908 Binet-Simon scale. In 1911 the Binet-Simon Scale received a minor revision. At this time, the idea of intelligence testing was swept all over the world. In 1916, L. M.Stanford University had revised the binet test for use in the United States. Terman review, known as Stanford-Binet Intelligence Scale (Terman, 1916), was the only American version of the Binet test that prospered. And also One of the most important trends of the test - the guide to the best test. The revision of the 1916 Terman of the Binet-Simon scale contained many improvements. The standardization sample was increased to include 1,000 people, the original elements have been revised and many new features were added. The 1916 Stanford-Binet Intelligence Scale added respectability and momentum to the development of new testing movement. The World War I the testing movement has grown tremendously in the United States because of the demand for a quick and eÃ¯Â¬cient to evaluate the emotional and intellectual functioning of thousands of military recruits in World War I. The war He has created a demand for largescale Group Test © because relatively few members of the trained personnel could consider the high inclusion of military recruits. However, the test was BINET an individual test. Shortly after the United States became actively involved in World War I, the army asked the assistance of Robert Yerkes, who was then president of the American Psychological Association (see Yerkes, 1921). Yerkes led a panel of distinguished psychologists who soon developed two structured group test of human abilities: the army alpha and beta army. The Alpha Army has requested the reading ability, while the army Beta measured the intelligence of illiterate adults. The First World War has fueled the widespread development of the test group. About this time, although the test is carried is also expanded to include tests of achievement, aptitude, interest and personality. © Since the realization of tests, aptitude and intelligence overlap considerably, the distinctions have proved most illusory of reality. Even so, the scale of 1916 Stanford-Binet Intelligence had appeared at a time of strong demand and great optimism for the potential of measuring human behavior through testing. World War I and the creation of test group were then added momentum to test movement. Shortly after the appearance of the scale of the 1916 Stanford-Binet Intelligence and Army Alpha test, schools, universities and the industry began to use the test. It seemed that a lot of this new phenomenon, the psychological test, held the key to solving the emerging problems from rapid population growth and technology. The achievement tests among the most important developments after World War I was the development of standardized test results. In contrast to the assay test, the realization of standardized tests provide multiple choice questions on a standardized 16 Chapter 1 Ã - Introduction Large sample to produce standards against which the results of the new tested can be compared. The realization of standardized tests were quickly captured because of the relative ease of administration and the score and the lack of subjectivity or favoritism that may occur in the essay or other written tests. In school settings, the results of standardized tests have allowed us to maintain identical test

conditions and scoring rules for a large number of children. These tests have also allowed a wider coverage of the content and were less expensive and more eÃ¯Â¬cient of essays. In 1923, the development of tests of standardized results culminated in the publication of the results of Stanford T. L. Kelley test, G. M. Ruch, and M. L. M. Terman. In the 30s, it was widely considered that the objectivity and reliability of these new standardized tests have made them superior to the essay test. Their use widely proliferated. Interestingly, since © will discuss later in the book, that today's teachers seem to have come full circle. Currently, many people favor the written tests and the work (ports) on standardized results tests as the best way to assess children (Boerum, 2000; Harris, 2002). Increase to the challenge for each movement is a countermovement, and the test movement in the United States in the 1930s was no exception. Critics soon became vocal enough to damp enthusiasm and to make even lawyers more optimistic than defensive tests. Researchers, who asked nothing short ofThe standards, have observed the limits and weaknesses of existing tests. Not even the Stanford-Binet, a point of reference in the field of tests, was safe from criticism. Although tests were used between the two world wars and many new tests have been developed, their accuracy and utility remained under heavy. Towards the end of the 1930s, the developers began to re-establish the respective testing. New improved tests have reflected the knowledge and experience of the previous two decades. In 1937, Stanford-Binet had been magazine again. Among the many improvements has been the inclusion of a standardization sample of over 3000 individuals. Just 2 years after the audit of Stanford-Binet 1937, David Wechsler published the first version of the Wechsler intelligence scales (see chapter 10), the Wechsler-Bellevue Intelligence Scale (W-B) (Wechsler, 1939). The Wechsler-Bellevue scale contained several interesting innovations in the intelligence test. Unlike the Stanford-Binet test, which produced only a single score (the so-called IQ quotient, or intelligence), the Wechsler test has provided several scores, allowing an analysis of an individual or combination of ability. Among the various scores produced by the Wechsler test was the performance IQ. Performance tests do not require a verbal response; You can use to evaluate intelligence in people who have few verbal or linguistic skills. The Stanford-Binet test had long been criticized due to its emphasis on linguistic and verbal skills, making it inappropriate for many individuals, such as those who can't speak or who cannot read. Furthermore, few people believed that linguistic or verbal skills play an exclusive role in human intelligence. The inclusion of Wechsler of a non-verbal scale has so helped to overcome some of the practical and theoretical weaknesses of the Binet test. In 1986, the Binet test was drastically revised to include performance subtest. More recently, it was rehabilitated in 2003, as we will see in Chapter 9. (Other Chapter 1 - Introduction 17 Important concepts in the intelligence test will be formally defined in chapter 10, which covers the various forms of intelligence wechsler stairs .) Personality test: 1920-1940 just before and after the Second World War, the personality tests began to flourish. While intelligence tests measured capacity or potential, personality tests measured presumably stable characteristics or traits that theoretically submit behaviors. The traits are relatively lasting provisions (kept to act, think or feel in a certain way in a certain circumstance) that distinguish an individual on the other. For example, we say that some people are optimistic and pessimistic. Optimistic people tend to remain so regardless of whether things are going well. A pessimist, on the contrary, tends to look at the negative side of things. Optimism and pessimism can therefore be considered as traits. One of the fundamental objectives of traditional personality tests is to measure the traits. As you learn, however, the notion of traits has important limitations. The first personality tests were structured paper and pencil group tests. These tests provided multiple-choices and true-false questions that could be administered to a large group. Because it provides a high degree of structure, that is a specific stimulation and specific alternative answers that can be marked unequivocally, this type of test is a type of structured personality test. The first structured personality test, the Woodworth Personal Data Sheet, was developed during the First World War and was published in a final form immediately after the (See figure 1.6). As indicated above, the motivation underlying the development of the first personality test was the need to project military recruits. History indicates that tests such as Binet and Woodworth were created by the need to deal with unique challenges. As the first skill tests, however, the first structured structured personality test simple by today's standards. The interpretation of the test depended Woodworth hypothesis discredited UN that the content of an article could be accepted at face value. If the person has marked â Falseâ to the statement â I wet the bed, â then it is assumed that he or she does not wet the â Beda. As logical as this assumption Yes No 1. I wet the bed. 2. Drink a quarter of whiskey every day. 3. I'm afraid of enclosed spaces. 4. I was being followed. 5. People are out to get me. 6. Sometimes I see or hear things that other people do not hear or see. F I G U R E 1.6 About Woodworth Sheet has been an attempt to standardize the psychiatric interview. It contains questions like the ones shown here. 18 CHAPTER 1 â The introduction seems, experience has shown that it is often false. In addition to being dishonest, the person answering the question can not interpret the meaning of "wet the bed" as does the test administrator. (Other problems with tests such as Woodworth are discussed in Chapter 13.) The introduction of Woodworth test was followed enthusiastically by creating a variety of structured personality tests, which have all assumed that the response of a subject may be taken at face value. However, the researchers examined, analyzed and criticized the first structured personality tests, just as they had done with the skill test. In fact, the critical tests which were based on face value alone Ì It became so intense that the structured personality tests were nearly hunted out of existence. They followed the development of new tests based on the most modern concepts, revitalizing the use of structured personality test. So, after an initial surge of interest and optimism during most of the twenties, the structured personality tests have declined since the late '30s and early' 40s. After World War II, however, we have been introduced personality tests based on different assumptions or less, so recovering the structured personality tests. During the brief but dramatic rise and fall of the first structured personality tests, interest in projective tests began to grow. In contrast with the structured personality tests, which generally provide a test stimulus relatively unambiguous and specific alternative answers, the projective personality tests provide an ambiguous stimulus and response requirements unclear. In addition, the projective test score is often subjective. Unlike the first structured personality tests, interest in the inkblot Rorschach projective test grew slowly (see Figure 1.7). The Rorschach test was first published by Hermann Rorschach in Switzerland in 1921. However, several years passed before the Rorschach arrived in the United States, where David Levy introduced him. The first doctoral thesis Rorschach written in the US a university was not completed until 1932, when Sam Beck, Levy student, decided to investigate the properties of Rorschach test scientifically. Although the initial interest in the Rorschach test has been lukewarm at best, its popularity is growing rapidly after Beck's work, despite the suspicion, doubt and criticism from the scientific community. Today, however, the Rorschach is under a dark cloud (cfr. Chapter 14). Adding to the time of Henry Murray and Christina Morgan acceptance and use of projective tests has been the development of the Thematic Apperception Test (TAT) in 1935. While the Rorschach inkblot contained completely ambiguous stimuli, the TAT was more structured . His ambiguous stimuli consisted of images depicting a variety of scenes and situations, like a guy sitting at a table with a violin on it. Unlike the Rorschach test, which the subject to explain what the ink might be, the TAT required the subject to invent a story on the ambiguous scene. The TAT claimed to measure human needs and therefore to ascertain individual differences in motivation. The emergence of new approaches to personality testing Theof the two most important projective personality tests, the Rorschach and TAT, grew rapidly in the late 1930s and early 1940s, perhaps due to the disillusionment with structured personality tests (Dahlstrom, 1969a). However, as we shall see in Chapter 14, the objective tests, in particular the Rorschach, do not have CHAPTER 1 ■ Introduction 19 F I G U R E 1.7 Card 1 of the Rorschach Ink Test, a projective personality test. Such tests provide an ambiguous stimulus to which a subject is asked to make some response. resisted a vigorous examination of their psychometric properties (Wood, Nezworski, Lilienfeld, & Garb, 2003). In 1943, the Minnesota Multiphasic Personality Inventory (MMPI) began a new era for structured personality testing. The idea behind MMPI – to use empirical methods to determine the meaning of a test response – helped revolutionize structured personality tests. The problem with the first structured personality tests such as Woodworth was that they made too many assumptions that subsequent scientific investigations could not justify. The authors of the MMPI, on the contrary, argued that the meaning of a test response could only be determined by empirical research. The MMPI, together with its updated partner MMPI-2 (Macellaio, 1989, 1990), is currently the most used and reference personality test. Its emphasis on the need for empirical data has stimulated the development of tens of thousands of studies. Just when the MMPI appeared, personality tests based on statistical procedure called factor analysis began to emerge. Factor analysis is a method to find the minimum number of dimensions (characteristics, attributes), factors called, to take into account a large number of variables. We can say that a person is coming out, he is gregarious, he searches for the company, he is talking, and he enjoys relationships with others. However, these descriptions contain a certain amount of redundancy. An analysis of factors can identify how overlapping they are and whether everyone can be considered or surprised under a single dimension (or factor) such as estrogen. In the early 1940s, J. R Guilford made the first serious attempt to use factor analytical techniques in developing a structured personality test. At the end of that decade, R. B. Cattell introduced the Sixteen Personality Factor Questionnaire (16PF); despite its downward popularity, it remains one of the most well structured personality tests and an important example of a test developed with the help of factor analysis. Today, factor analysis is a tool used in design 20 CHAPTER 1 ■ Introduction TA B L E 1.2 Summary of personality tests Woodworth Personal Data Sheet: An early structured personality test that assumed that a test response can be taken at nominal value. The Rorschach Inkblot Test: A highly controversial projector test that provided an ambiguous stimulus (an inkblot) and asked the subject what it might be. Acceptance Thematic Test (TAT): A projective test that provided ambiguous images and asked subjects to invent a story. The Minnesota Multiphasic Personality Inventory (MMPI): A structured personality test that did not hypothesize about the meaning of a test response. This meant to be determined by empirical research. The Psychological Inventory of California (CPI): A structured personality test developed according to the same principles as MMPI. The Sixteen Personality Factor Questionnaire (16PF): A structured personality test based on statistical factor analysis procedure. or validation of almost all main tests. (Analytical personality tests will be discussed in Chapter 13.) See Table 1.2 for a briefpersonality tests. The period of rapid changes in the test state The 1940s saw not only the emergence of a new technology in the psychological test, but also the growth of the applied aspects of psychology. The role and significance of the tests used in the First World War have been reaffirmed inWar II. At this point, the US government began to encourage the continued development of applied psychological technology. As a result, a significant federal funding Purcha © paid, the supervised training for oriented clinical psychologists. In 1949, college formal training standards were developed and accepted, and clinical psychology was born. Other branches of applied psychologyÂ ¢ such as industrial, consulting, education and psychologyÂ ¢ school soon began to flourish. One of the main functions of the applied psychologist provided psychological tests. The Shakow, Hilgard, Kelly, Sanford, and ShaÃ¯Â¬ er (1947) report, which was the foundation of education standards in clinical psychology, speciÃ¯Â¬ ed psychological tests that had a unique function of the clinical psychologist and recommended that test methods be taught doctoral psychology students. An outlet of the American Psychological Association's position has published seven years later (APA, 1954) aÃ¯Â¬ confirmed that the domain of clinical psychologist including test. And 'formally it declared, however, that the psychologist would conduct psychotherapy only in collaboration trueÃ ¢ Ã ¢ with doctors. So, the psychologists could lead test independently, but not psychotherapy. In fact, until psychologists assumed the role of tester, they have played a complementary role, but often secondary in Ã -vis doctors. Although the medical profession could have prevented the emergence of clinical psychology, he did not, because as a tester © psychologist helped the doctor. Therefore, in the late 1940s and early 1950s, the test was the main function of the clinical psychologist (ShaÃ¯Â¬ er, 1953). A CHAPTER 1 Introduction 21 For better or worse, depending on oneâ ¢ s perspective, the governmentâ ¢ s eÃ¯Â¬ Orts to stimulate the development of the applied aspects of psychology, clinical psychology in particular, were very successful. Hundreds of young talented and creative people are attracted to areas of application and other clinical psychology. These individuals, who would use tests and other psychological techniques to solve practical human problems, have been uniquely trained as practitioners of principles, empirical foundations, and the applications of the science of psychology. Armed with powerful knowledge from scientiÃ¯Â¬ c psychology, many of these early clinical professionals must have felt frustrated by their relationship with physicians (see Saccuzzo & Kaplan, 1984). Unable to engage independently in the practice of psychotherapy, some psychologists felt like the technicians who serve the medical profession. The extremely talented group of psychologists World War II quickly began to refuse this role. In addition, because many tests © psychologists associated with this sub report, refused the test (Lewandowski and Saccuzzo, 1976). At the same time, the potentially intrusive nature of trials and fears of abuse began to create public suspicion, distrust and contempt for the test. The tests were attacks from inside and outside the profession. These intensiÃ¯Â¬ attacks and multiplied so fast that many psychologists thrown all ties with traditional tests developed during the semester Ã¯Â¬ before the 20th century. Test then suffered another sharp drop was in the late 1950s that persisted in 1970 (see Holt, 1967). The current environment in the course of 1980, 1990 and 2000 several important branches of applied psychology emerged and bloomed Ã¯Â¬: neuropsychology, health psychology, forensic psychology and child psychology. © Since each of these major areas of psychology makes extensive use of psychological tests, psychological tests again grew in the state and use. Neuropsychologists use tests in hospitals and other clinical settings to evaluate the brain injury. Health psychologists use tests and In a variety of medical environments. Forensic psychologists use tests in the legal system to evaluate the mental state as they are It refers to a defense of insanity, competence to stand trial or to be executed, and emotional damage. The child psychologists use tests to assess childhood disorders. The tests are currently in use in developed countries around the world (Marsh, Hau, Artelt, Baumet, & Peschar, 2006; Black & William, 2007). As in the past, the psychological test remains one of the most important but controversial issues in psychology. As a student, no matter what your profession or career goal, you will find the material in this invaluable text. If you are among those who are interested to use psychological techniques in an applied environment, then this information will be particularly significant. From the roots of psychology at present, the psychological tests were among the most important tools of the psychologist in general and of those who apply psychology in particular. The test is in fact one of the essential elements of psychology. Although not all psychologists use tests and some psychologists are opposed to them, all the areas of psychology depend on the knowledge gained in research studies that are based on measurements. The significance and the reliability of these measurements are essential for psychological research. To effectively investigate any area of human behavior, one must understand the basic principles of measurement. 22 â CHAPTER 1 Introduction In the complex society of today, the importance of the principles, applications, and of the psychological test questions extends well beyond the field of psychology. Even if you have no intention of becoming a psychologist, you will likely encounter psychological tests. Lawyers, doctors, social workers, business executives, educators and many other professionals often have to deal with on these tests based relationships. Even as a parent, it is likely to meet test (taken from your children). To properly interpret this information, it is necessary that the information presented in this book. The more you know about psychological tests, the more you can be confident in your meetings with them. Given the testing attack and threats to prohibit or greatly restrict their use, you have a responsibility to yourself and to society to know how much you can about psychological testing. The Test future may depend on you and people like you. A thorough understanding of testing will allow you to base your decisions on facts and to ensure that tests are used for more profitable and constructive purposes. The tests probably have never been so important as they are today. For example, consider a single type of test: academic aptitude. Every year more than 2.5 million students take tests that are designed to measure academic progress or fitness, and the testing process starts early in the lives of students. Presecondarie Some schools require some tests, and thousands of children take them every year. When these students reach their teens and want to get into college preparatory schools, tens of thousands will take a screening test. Few students who want to go to a 4 year college can avoid taking a college entrance exam. The Reason The SAT test alone is given to about 2 million high school students every year. Another 100,000 high school seniors take other tests to get advanced placement in college. These figures do not include the 75,000 people who take a special test for admission to business school or the 148,000 who take an admissions test for law school - or test for graduate school, medical school, dental school, l ' military, professional licenses, and others. In fact, the testing service for education alone administers PIA 1 of 11 million tests annually in 181 countries (Gonzalez, 2001). Nor © include the millions of tests provided worldwide for research and evaluation (Black & William, 2007; Marsh et al., 2006). As sources of information on human characteristics, the results of these tests affect critical life decisions. SUMMARY The history of psychological testingStates were brief but intense. Although these types of tests have long been available, psychological tests is a lot of a product of modern society with its unprecedented technology and population growth and unique problems. On the contrary, helping to solve the challenges posed by modern developments, the tests played an important role in the recent history of the United States and the world. It should be realized, however, that despite progress in theory and psychological testing technique, many unresolved technical problems and strongly discussed political and economic issues remain. However, the prevalence of tests despite the strong opposition indicates that, even if they are far from the perfect, psychological tests must satisfy some important needs in decision-making processes that permeate all aspects of society. Since the decisions must be taken, these tests will probably flow until a better or more objective way of making decisions will emerge. Chapter 1 - Introduction 23 Modern history shows that psychological tests have evolved into a complicated environment in which hostile and friendly forces have produced a balance characterized by innovation and continuous search for better methods. An interesting thing about testing is that people never seem to remain neutral about them. If you are not in favor of testing, then we ask that you keep a flexible and open mind while you study them. Our goal is to give you sufficient information to evaluate psychological tests intelligently throughout your life. The page intentionally left an empty chapter 2 basic standards and statistics for testing the AR Ningobj Ec Tivi s when you have completed this chapter, you should be able to: â- discuss three proprieti of measurement scales Â- determine why the properties Stairs are important in the measurement field Â- Identify methods for displaying scores distributions â- calculates the average and the standard deviation for a whole say how the rules are created â- relaxing the notion of tracking to the definition of standards 25 26 Chapter 2 â- basic standards and statistics for W test and everyone uses numbers as a basic way to communicate: our money system requires understanding and manipulating numbers, we will estimate how long it will take to do things, we count, we express Evaluations on stairs, and so on. Think how many times you use the numbers on a medium day. There is no way to avoid them. An advantage of numerical systems is that they allow us to manipulate information. Through the well-defined rules series, we can use the numbers to learn more about the world. The tests are devices used to translate observations into numbers. Because the outcome of a test is almost always represented as a score, most of this book is on what the scores mean. Th is the chapter examines some of the basic rules used to evaluate number systems. These number of rules and systems are the psychologist's partners in the learning of human behavior. If you have had a psychological statistical course, this chapter will strengthen the basic concepts you've already learned. If you need an additional review, read your book of introductory statistics. Most of these books cover the information in this chapter. If you have not had a statistical course, this chapter will provide some of the information you need to understand other chapters in this book. Because we need statistics through its commitment to the scientific method, modern psychology has advanced more centuries of speculation on human nature. The scientific study requires systematic observations and an estimate of the measure in which the observations could have been influenced only by chance (Salkind, 2007). Methods need two important purposes in researching scientific methodsFirst of all, statistics are used for the purpose of the description. The numbers provide comfortable reepylogues and allow us to evaluate some observations related to others (Cohen & Lea, 2004; Pagano, 2004; Thompson, 2006). For example, if you get a 54 score on a psychology exam, you probably want to know what the 54 means. Is it lower than the average score, or is it the same? Knowing the answer can make the feedback you get from your exam more significant. If you find out that 54 puts you at the top of 5% of the class, then you may assume that you have a good chance for an A. If it puts you down 5%, then you'll be feeling upright. Secondly, we can use statistics to make inferences, which are logical deductions on events that cannot be observed directly. For example, you don't know how many people have watched a particular TV movie unless you ask everyone. However, using Scientiï¬ C sample surveys, it is possible to deduct the percentage of people who have seen ï¬ lm. Data collection and analysis could be considered analogous in criminal investigation and judicial process (Cox, 2006; Regenwetter, 2006; TUKEY, 1977). The first comes the investigative work of collecting and displaying clues, or what statistics John Tukey calls analysis of exploratory data. Then a period of analysis of the connected data arrives, when the clues are evaluated by strict statistical rules. This last stage is like the work done by judges and juries. Some students have an aversion to numbers and anything mathematical. If you're in their midst, you're not alone. Not only students but also professional psychologists can feel uncomfortable on statistics. However, statistics and basic chapter 2 Â- Standards and basic statistics to test 27 measurement principles at the centre of modern psychology science. The scientific statements are usually based on a careful study, and this systematic study requires a numerical analysis. This chapter is both descriptive and inferential statistics. Descriptive statistics are methods used to provide a concise description of a collection of quantitative information. Inferential statistics are methods used to make inferences from the observations of a small group of people known as sample to a larger group of individuals known as population. In general, the psychologist wants to make statements on the larger group but cannot create all the necessary observations. Instead, he or she observes a relatively small group of subjects (campion) and uses inferential statistics to estimate the characteristics of the larger group (Salkind, 2007). Measurement scales The measurement can deï¬ NE as the application of the rules to assign numbers to objects. The rules are the specific procedures used to transform the attributes in numbers (Camilli, Cizek, & Lugg, 2001; Nunnally & Bernstein, 1994; Yanai, 2003). For example, in order to assess the quality of wines, wine ganopertes must use a set of specific rules. They could evaluate the wine on a 10-point scale where 1 means extremely bad and 10 means extremely good. For a murderer to assign numbers, the rules system must be clearly deed The basic feature of these types of systems is the measurement scale. For example, to measure the height of your classmates, you can use the thumb scale; To measure their weight, you could use the pounds scale. There are numerous systems with which we enlist numbers in psychology. In fact, the study of measuring systems is what this book is about. Before considering any specific measurement scale, however, we should consider the general properties of scalesmeasurement. Properties of scales Three important properties make measurement scales of€ Erent on the other: size, equal intervals and absolute magnitude 0. Magnitudo is the property of â€ œmores.Â € A scale has the property of magnitude if we can say that a particular instance of the attribute represents another, less or equalof the quantity given compared to another instance (ARON & ARON, 2003; Hurlburt, 2003; McCall, 2001; Howell, 2008). On a height scale, for example, if we can say that John is taller than Fred, then the scale has the property of magnitude. A scale that does not have this property arises, for example, when a gym trainer assigns identity numbers to teams in a league (team 1, team 2, and so on). Because numbers label only teams, they do not have the property of magnitude. If the coach were to classify the teams for the number of games they won, then the new numbering system (the winning games) would have the property of magnitude. Chapter 2 Â- Basic standards and statistics for testing 10 Evaluations of works of art 28 8 6 4 2 0 2 4 6 8 Manual test DESTRITY 10 F I G U R E 2.1 Hypothetic relationship between evaluations of works of art and manual dexterity. In some scale ranges, the relationship is more direct than in others. Equal intervals The concept of equal intervals is a little more complex than that of greatness. A scale has the property of equal intervals if the difference between two points in any place on the scale has the same meaning as the difference between two other points which are of ... € for the same number of units of scale. For example, the issue of Dieï¬ € between inch 2 and inch 4 on a ruler represents the same amount of the difference between inch 10 and inch 12: exactly 2 inches. Simple as this concept seems, a psychological test rarely has the property of equal intervals. For example, the difference between Intelligence Quotant (IQS) of 45 and 50 does not mean the same as the difference between IQ from 105 and 110. Although each of these heirs is 5 points (50 Â «or 45 Â«1⁄2 5 and 110 Â «or 105 Â«1⁄2 5), the 5 points at level ï¬ do not mean the same thing as 5 points per second. We know that IQ provides performance in class. However, the eruption of diï¬€ in the performance of the class associated with diïa-€ erections between the IQ scores of 45 and 50 are not the same as the class performance issues associated with the IQ score Erenze di 105 e 110. The next chapters will discuss this issue in more detail. When a scale has the properties of the same ranges, the relationship between the measured units and some results can be described by a straight line or by a linear equation in the Y module "1⁄2 A« 1 bx. This equation shows that an increase in equal shares on a given scale by relucting the same increases in significant unit correlations. For example, figure 2.1 shows the hypothetical relationship between scores on a manual dexterity test and art rating. Note that the report is not a straight line. If you look at the points on ï¬ Gure, you can see that the report is almost linear: increases in manual dexterity are associated with increases in art ratings. So the relationship becomes non-linear. The ï¬ Gure shows that after a manual score of the dexterity of about 5, the dexterity increases produce relatively smaller increases in the quality of the work of art. CHAPTER 2 TA BLE Â Standards and Base Statistics for testing 29 2.1 measurement scales and their properties Type of properties of magnitude equal to absolute 0 rated ranges No No Order No Yes No Yes No Yes No Interval Report Yes Yes Yes Absolute 0 0 0 An absolute is achieved when there is nothing of the measured property. For example, if you are measuring your heart rate and observe that your patient has a 0 rate and is dead, then you would have concluded that there is no heart rate at all. For many psychological qualities, it is extremely diïa, if not impossible, to give an absolute 0 point. For example, if you measure shyness on a scale from 0 to 10, then it is difficult to deï¬ that means that for a person you have absolutely no shyness2001). Types of stairs Table 2.1 DeaÃ¯â¬s four measurement scales based on the properties we have just discussed. You can see a nominal nominal scale do not have the property of magnitude, equal intervals, or an absolute 0. The nominal scales are not at all stairs; their only purpose is to appoint objects. For example, the numbers on the shoulders of the uniforms of football players are nominal. The nominal scales are used when the information is qualitative rather than quantitative. Social science researchers commonly label groups in sample surveys with numbers (such as 1 di African American, 2 di white, and 3). Mexican American). When these numbers have been attached to categories, most statistical procedures are not significant. For example, on the scale of ethnic groups, which means a means of 1.87? It cannot be said that the sophisticated statistical analysis of nominal data is impossible. In fact, several new and exciting developments in data analysis allow a detailed and detailed use of nominal data (Chen, 2002; Miller, Scurfield, Drga, Galvin, & Whitmore, 2002; Stout, 2002). A scale with the magnitude property but not of range or 0 absolute is an ordinal scale. This scale allows you to classify individuals or objects, but not to say anything about the meaning of differences between ranks. If you were to classify the members of your class by height, then you will have an ordinary scale. For example, if Fred was the highest, Susan the second highest, and George the third highest, assign them respectively ranks 1, 2 and 3. You wouldn't mind Fred being 8 inches taller than Susan, but Susan is only 2 inches taller than George. For most problems in psychology, accuracy to measure the exact differences between intervals does not exist. Therefore, most of the time you have to use order scales of 30 CHAPTER 2 ■ Basic standards and statistics for testing measurement. For example, IQ tests do not have the same or 0 absolute range properties, but have the magnitude property. If they had the same range ownership, then the difference between a 70 and one 90 IQ should have the same meaning as the difference between a 125 and 145 IQ. Because it does not, the scale can only be considered ordinary. In addition, there is no point on the scale which does not represent intelligence at all — that is, the scale does not have an absolute 0. When a scale has the properties of magnitude and equal but unabsolute ranges 0, we refer to it as a range scale. The most common example of a range scale is the temperature measurement in Fahrenheit degrees. This temperature scale clearly has the magnitude property, because 35°F is warmer than 32°F, 65°F is warmer than 64°F, and so on. In addition, the difference between 90°F and 80°F is equal to a similar difference of 10° at any point on the scale. However, on the Fahrenheit scale, the temperature does not have the absolute property 0. If he did, then point 0 would be more significant. As it is, 0 on the Fahrenheit scale has no particular meaning. Water freezes at 32°F and bubbles at 212°F. Since the scale does not have an absolute 0, we cannot make statements in terms of relationships. A temperature of 22°F is not twice as hot as 11°F, and 70°F is not twice as hot as 35°F. The Celsius temperature scale is also a range rather than a ratio scale. Although 0 represents freezing on the Celsius scale, it is not an absolute 0. Remember that an absolute 0 is a point where there is nothing of the measured property. Also on the Celsius temperature scale, there is still a lot of room on the thermometer below 0. When the temperature goes under freezing, some aspects of the heat is still being measured. A scale that has all three properties (magnitude, equal intervals and 0 absolutes) is called a scale of relationship. To continue our example, a temperature ratio scale would have the properties of the Fahrenheit stairsCelsius, but also include a significant point 0. There is a point where all molecular activities cease, a point point Absolute 0 on a temperature scale. Because the kelvin scale is based on the absolute point 0, it is a scale ratio: 22 Â ° k is two more cold as 44th Â ° K. Examples of stairs report also appear the numbers we see on a regular basis. For example, consume the number of yards obtained by performing teams. Zero yards means actually that the player has earned any construction sites. If a player earned 1000 yards and another earned only 500, so we can say that the athlete is prime earned twice as many as the second. Another example is running speed. For example, 0 miles per hour (MPH) is the point where there is no speed at all. If you drive on a 30 mph road and increase speed at 60 when you join, then you have doubled the speed. Admissible measurement level operations It is important because DeÃ¯â¬ NES as mathematical operations we can apply to numerical data. For nominal data, each observation can be placed in one mutually exclusive category. For example, you are a member of one kind. Nominal data can be used to create frequency distributions (see the next section), but no mathematical manipulation of the data are allowed. Ordinal measurements can be manipulated through arithmetic; However, the result is chapter 2 to 31 rules and basic statistics for the worship test often of ï¯â¬ to interpret why reÃ¯â¬ ects nor the sizes of the observations manipulated nor the real values of the property that are States measured. For example, if the heights of 15 children are ordered rank, knowing rank a given child ¢ s does not reveal how high he or she is. Medium in these ranks are equally uninformative about it. With interval data, you can apply any arithmetic operation to erensions ofÃ¯â¬ between columns. The results can be interpreted in relation to the quantities of the structure below. However, the interval data cannot be used to make statements about reports. For example, if IQ is measured on an interval scale, you can't say that a 160 Qi is twice superior as an qi of 80. This mathematical operation is reserved ratio scales, so that any math operation is Admissible. Frequency Distribution A single simple test score if it refers to other test scores. A scoring distribution summarizes scores for a group of individuals. In tests, there are many ways to record a scoring distribution. The scoring frequency distribution displays on a variable or a measure for reÃ¯â¬ ECT frequency has been obtained each value. With a frequency distribution, one deÃ¯â¬ does all the possible scores and determines the number of people who get each of these scores. Usually, scores are arranged on the highest higher axis. The vertical reÃ¯â¬ axis ects how many times each of the values on the horizontal axis was observed. For most test scores, the frequency distribution is, with the utmost frequency of scores to the distribution center and decreasing scores the values become greater or less than the value in the center of the bell distribution. Figure 2.2 shows a frequency distribution of 1000 observations that assumes values between 61 and 90. Note that most frequent observations fall towards occurrence frequency 100 80 60 40 20 0 62 64 66 68 70 72 74 76 78 80 82 Score 84 86 Figure 2.2 Frequency distribution approximate a normal distribution of 1000 observations. 88 90 32 Chapter 2, Statistical standards and base to test the distribution center, about 75 and 76. As you look towards the extremes of the distribution, you¡˜th to find a systematic decline of the frequency with which the scores occur. For example, the Of 71 he is less frequently observed by 72, which is less frequently observed 73, and so on. Similarly, 78 is more frequently observed than 79, which is noticed more often than 80, and so on. Although this clean clean symmetrical relationship They do not characterize all scoring sets, occurs quite frequently in practice for us to pay particular attention to it. In the normal distribution section, we explain this concept in more detail. Table 2.2 lists the amounts of precipitation in San Diego, California, between 1964 and 2007. Figure 2.3 is a host based histogram. The distribution is TA BLE 2.2 Inches of Rainfall in San Diego, 1964-2007 Rain Year (Inches) Rain Year (Inches) 1964 5.15 1988 12.44 1965 8.81 1989 5,88 1966 14.76 1990 7.62 1967 10.86 1991 12.31 1968 7.86 1992 12.48 1969 11.48 1993 18.26 1970 6.23 1994 8.03 1995 17.13 . Standard Deviation 3.62 454.8 10.34 4.71 Chapter 2 Â- Norms and Basic Statistics for Test 33 12 Frequency 10 8 6 4 2 0 0-3 Figure 2.3 3.1â € "6 6.1â €" 9 9.1â € "12.1â € "15 15.1â €" 18 18.1â € "21 21.1â €" 24.1â € "27 Inches of Rainfall Frequency Poligon for Precipitation in San Diego, 1964â €" 2007. Slightly SKEWED, or asymmetrical. We say that figure 2.3 has a positive sched because the queue goes towards the upper or positive side of the X axis. There is a slight decline in Figures 2.3 and 2.4, but asymmetry in these figures is relatively difficult To be detected. Figure 2.5 gives an example of a distribution that is clearly SKEWED. The data summarizes the annual income of families in the United States in 2007. Very few people make high incomes, while most of the population has been grouped towards the low end of the income distribution. Of particular interest is that this figure includes only families income than $ 100,000. For families income than $ 100,000, the government only reports income using class intervals of $ 50,000. In 2007, about 16% of US families had incomes over $ 100,000. Because some families have extremely high incomes, you can imagine that the queue of this distribution would go far to the right. So, income is an example of a variable that has a positive SKEW. You may also present this same set of data as a frequency polygon (see figure 2.4). $ 50 000 to $ 50 000, $ 50 to 50 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 49 This is a positive sched example. (Data of the Department of Statistics of the United States and Census Office. Http: //ferret.bls. Census .gov / macro / 032003 / hhhinc / new06_000.htm.) Frequencies with which each interval occurs. The lines are then drawn to connect these points. Each time you draw a frequency distribution or a frequency polygon, you need to decide on the width of the class interval. The class interval for inches of precipitation is the unit on the horizontal axis. For example, in Figures 2.3 and 2.4, the class interval is 3 inches â € œIs, the demarcations along the X axis increase in 3-inch intervals. This interval is used here for convenience; of 3 inches is otherwise arbitrary. PERCENTIAL BANKS The percentage degrees replace the simple ranks when we want to adjust for the number of scores in a group. A rank per centile answers the question "What percentage of scores fall under a particular score (Xi)?" To calculate a percentage degree, you only need to follow these simple steps: (1) determine how many cases fall under the interest score, (2) determine how many cases are in the group, (3) divide the number of cases under the interest score (Step 1) by the total number of cases in the group (Step 2), and (4) multiply the result of Step 3 by 100. CHAPTER 2 ■ Basic standards and statistics for testing 35 The formula is Pr di B di100 di X i N grade percentage where Pr = Xi degree percentage = interest score B = the number of scores under Xi N = the total number of scores This means that a ratio of the number of cases is formed under the interest score and the total number of scores. Because there will always be the same or less cases in the numberer (higher) of the equation than there are in the denominator (lower half), this ratio will always be less than or equal to 1. To get rid of the decimal points, it multiplies by 100. For example, consider the runner ending 62 ° on 63 riders in a gym class. To obtain the percentage degree, divide 1 (the number of people who end up behind the person of interest) of 63 (the number of scores in the group). This gives you 1/63, or .016. Then multiply this result of 100 to get rank per centile, which is 1.6. The degree says the runner is below 2nd percentile. Now consider the Bay to Breakers race, which attracts 50,000 runners in San Francisco. If I had finished 62 out of 50,000, then the number of people behind you would be 49,938. Divide this from the number of participants gives .9988. When you multiply by 100, you get a degree per centile of 99.88. This tells you that the end 62nd in the Bay to Breakers race is exceptionally good because it places you in 99.88th percentile. The Everyday Life 2.1 Psychological Test presents the calculation of the per centimeter degrees of child mortality rates of selected countries as reported by the World Health Organization in 2007. Infant mortality is defined as the number of children out of 1000 who are born alive but die before their first birthday. Before proceeding, we must emphasize that the meaning of this calculation depends on which countries are used in comparison. In this example, the calculation of the percentage degree is divided into five steps and uses the raw data in the table. In phase 1, we organize data points in increasing order. Singapore has the lowest infant mortality rate (2.3), Japan is next (2.8), and Afghanistan has the highest rate (157.0). In phase 2, the number of cases with rates worse than the case of interest is determined. In this example, the case of interest is the United States. Therefore, we count the number of cases with a worse rate than the United States. Eleven countries — Israel, Saudi Arabia, Colombia, China, Turkey, Morocco, Bolivia, Laos, Ethiopia, Mozambique and Afghanistan — have infant mortality rates greater than 6.4. 36 CHAPTER 2 ■ Standards and basic statistics for the PSYCHOLOGICAL TESTING IN E V ERYDAY LIFE 2.1 Infant mortality in selected countries, 2007 Bolivia Infant mortality for 1000 live births. China 23.0 Colombia 19.1 Ethiopia 86.9 France 3.4 Israel 6.8 Italy 5.7 Japan 2.8 Laos 51.4 Morocco 30.6 Mozambique 95.9 Saudi Arabia 18.8 Singapore 2.3 Spain 4.3 Turkey 2.3 27.5 United States 6.4 Mean To calculate the percentage degree of child mortality in the United States than in the selected countries, use the following formula: dove where B dove100 N Pr di the percentage level B di the number of cases with worse interest rates N totale the total number of cases CHAPTER 2 Country ■Basic statistics for testing child mortality for 1000 Live births Singapore 2.3 Japan 2.8 France 3.4 Spain 4.3 Australia 4.6 Italy 5.7 United States 6.4 Israel 6.8 Saudi Arabia 18.8 Colombia 19.1 China 23.0 Turkey 27.5 Morocco 30.6 Bolivia 45.6 Laos 51.4 Ethiopia 86.9 Mozambique 95.9 Afghanistan 157.0 STEPS 1. Organize data in increasing order - i.e., the lowest score before, the second lowest score second, and so on. N = 18, average = 32.9, standard deviation = 41.9 2. Determine the number of cases with worse interest score rates. There are 11 countries in this sample with infant mortality rates greater than in the United States. 3. Determine the number of cases in the sample (18.) 4. Share the worst scores in the score score (Step 2) for the total number of scores (Step 3): 11 61 .61 18 5. Multiply from 100: .61 × 100 = 61st percentile rank 37 38 CAPPTER 2 ■ Norms and Basic Statistics for Testing In Step 3, we determine the total number of cases (18.) In Stage 4, we share the number of scores worse than the score of interest for the total number of scores: 11 . .61 18 Technically, the percentage degree is a percentage. Step 4 gives a proportion. Therefore, in step 5 it turns this into a whole number multiplying by 100: .61 × 100 = 61 So, the United States is in the 61st percentile. The percentage degree depends on the cases used for comparison. In this example, he calculated that the United States is in 61 percentile for infant mortality within this group of countries. If all countries of the world had been included, then the US ranking could have been different. Using this procedure, try to calculate the percentage degree for Bolivia. The calculation is the same, except that there are four countries with worse rates than Bolivia (as opposed to 11 worst in the United States). Thus, the degree per centile to Bolivia is 4 . .22 . 100 22 18 or 22 ° per centile. Now try France. You should have a percentage of 83. PERFORMANCE The percentages are the specific scores or points within a distribution. Percents share the total frequency for a set of observations in cents. Instead of indicating which percentage of scores fall under a particular score, such as ranks per centile do, percentiles indicate the particular score, under which a definite percentage of scores falls. Try to calculate the degree per centile and per centile for some of the data in Psychological Tests in Everyday Life 2.1 For example, look at Italy. The rate of child mortality in Italy is 5.72/1000 When you calculate the percentage degree, you exclude the interest score and count those below (in other words, Italy is not included in the count.) There are 12 countries in this sample with child mortality rates worse than the Italian one. To calculate the percentage degree, divide this number of countries by the total number of cases and multiply by 100: Pr B 12 . 100 . . Thus, Italy is in 67th grade per centile, or 67th percentile in this example is 5.72/1000 or 5.72 deaths per 1000 live births. Now take the example of Israel. The calculation of the percentage degree requires to look at the number of cases under the case of interest. In this example, 10 countries CHAPTER 2 ■ Basic standards and statistics for testing 39 in this group have child mortality rates worse than Israel. Thus, the degree per centile to Israel is 10/18 × 100 = 56. The 56th percentile corresponds to the point or score of 6.75 (6.75/1000 live births).) In summary, the percentage and rank per centile are similar. Per centile gives the point in a distribution under which a certain percentage of cases fall (6.75/1000 for Israel). The percentile is in units ofRaw. The rank percentile gives the percentage of cases below per centile; In this example, the rank percentile is 56. When reporting percentailes and degrees percentile, you need to carefully specify carefully The population is working. Remember that a rank percentile is a measure of relative performance. In interpreting a rank percentile, should you always ask a question about what? Ã, suppose, for example, that was finished in 17th percentile in a swimming competition (or Ã¯Â¬ fifth in a warmth of six competitors). Does this mean that you are a good slow swimmer? Not necessarily. It can be that this was a warmth in the Olympic Games, and the participants were the fastest swimmers in the world. Olympic swimmer A competing against a random champion of all people in the world would probably Nish Ã¯â¬ in 99.99th percentile. The example for infant mortality rates depends on which countries of the world for comparison were selected. The United States actually does not enough badly if compared with European countries and advanced economies in Asia (Singapore and Japan). However, the glances of US infant mortality rates much better than in the developing countries world.1 Describe statistical distribution that are used to summarize the data. If you consider a series of scores, the mass of information can be too much to interpret everything at once. This is why we need numerical convenience for help summarize information. An example of a series of scores that can be summarized is shown in Table 2.2 (see page 32), quantity of rain in San Diego. We mean the variable like X. A variable is a score that can have external values. The quantity of rain is a variable because of eating quantities of rain fallen in eating years. The arithmetic average in a distribution is called the average. To calculate the average, we have total scores and divide the sum to the number of cases, or N. the capital Greek letter Sigma (£ â â â £) means summation. So, the media formula, which we indicate as x, is xÃ ¢ Â «1 Â½ Ã ¢ âºx n we used an example similar to the last edition based on data 2003. In 2007, there were improvements elevation signa ¯â¬ in the rates of infant mortality in developing countries. Th Rate and Mozambique fell from 148.6 to 95.9 per thousand live births. Ethiopia reduced the rate of infant mortality 142.6-86.9. However, rates deteriorated slightly for different developed countries, including Israel, Italy, and Spain. 40 Chapter 2, Basic Statistics and Statistics for Word Test, This formula tells the total scores and divide the sum for the number of cases. Using information in Table 2.2, we Ã¯â¬ we discover the average following these steps: 1. 2. Get £ x, or the sum of scores: 5.15 + 8.81 + 14.76 + 10.86 + 7.86 + ... + 3.62 = 454.80 Find N, or the number of scores: n = 44 3. Divide £ x from N: 454.80 / 44 = 10.34 Psychological testing in everyday life 2.2 summarizes common symbols to be used in basic statistics. Standard deviation The standard deviation is an approximation of the average deviation around the average. The standard deviation for the quantity of rain in San Diego is 4.71. To understand precipitation in San Diego, it is necessary to consider at least two dimensions: Ã¯â¬ first, the amount of rain falling in a particular year; Second, the degree of change of year in the year in the quantity of rain falling. The calculation suggests that, on average, the variation around the average is about 4.71 inches. However information, knowing the average of a group of scores does not give you much information. As an example, observe the following series of numbers. Set 1 4 4 4 4 4 4 Sep 2 5 5 4 4 3 3 Sep 3 8 8 6 2 0 0 Psychological In Test and V Eryday Life Symbols 2. 2 Municipalities needed to understand and recognize the symbols used in this book. X is the average; It is pronounced a x bar.Ã ¢ â Â £ is the sign of summation. It means sum, or add, the scores together and is the Greek letter Sigma. X is a variable that takes exterior values ofÃ¯â¬. Every value of XI represents a rough score, also called a score obtained. Chapter 2 Ã, standards and statistical base for testing 41 Calculate the average of the first set. You should get 4. what is average in the second set? If you calculate correctly, you should get back 4. Finding the means to Set 3. It also 4. The three distributions of scores look very different but have the same meaning, so it is important to consider other characteristics of the distribution of scores over average. The difference between the three sets to variability in. There's variation in set 1, a small amount in Set 2 and Set 3. Measure lot in this variation is akin to finding the average deviation around the middle. One way to measure the variability is to subtract the means by any score (X â X) and then totalized deviations. The statistics often indicate it with a lowercase x, as in x = (x â X). Try this for the data in Table 2.2. You've got 0? You should have, and this is not an unusual example. In fact, the sum of the deviations around the means will always be equal to 0. However, you have a choice: You can place all the deviations around the means to get rid of any negative sign. So it is possible to obtain the square average deviation around the middle, known as the variation. The formula for the variation is '2 â â "â º ½ (X â« º X) 2 where N (X â X) is the deviation of a score from the mean. The s mark is the tiny greek sigma; s 2 is used as the variance of the standard description. Even if the variation is a useful statistic commonly used in the analysis of data, showing the variable in the squared deviations around the middle rather than in deviations around the middle. In other words, the variation is the average squared deviation around the mean. To bring it back in the unit that will make sense, we have to take the square root of the variance. The square root of the variation is the standard deviation (s), and is represented by the following formula â'A '½ â º (X â «º X) 2 N The standard deviation is then the square root of the deviation squared average around the mean. Even if the standard deviation is not an average deviation, it gives a useful approximation of what a typical score is higher or lower than average score. Because of their mathematical properties, the variation and the standard deviation have many advantages. For example, knowing the standard deviation of a batch of data is normally distributed allows us to make precise statements about the distribution. The formulas are presented just to calculate the variation and the standard deviation of a population. © That's why we use the tiny greek sigma (s and s2). Psychological Test in Everyday Life 2.3 summarizes when you should use Greek and Roman letters. The more often we use the standard deviation for a sample to estimate the standard deviation for a population. When we speak of a sample, replace the greek with a Roman letter S. In addition, we divide 42 CHAPTER 2 â Standards and Basic Statistics for testing â 1 N N rather than to recognize that a sample is S only an estimate of population change. Sâ "â º ½ (X â 'º X) 2 N â' º1 When calculating the standard deviation, is often easier to use the equivalent formula of raw score, which is Sâ" ½ (â ºX) 2 NN â "â º1 ºX 2 â« º This calculation can also be done automatically by some minicomputers. In the formula reading, you may be confused by a few points. In particular, be careful not to confuse Î £ X 2 and (Î £ X) 2. To get Î £ X 2, each score is squared and the values are summed. For scores of 3, 5, 7, and 8, Î £ X 2 would be 32 + 52 + 72 + 82 = 9 + 25 + 49 + 64 = 147. To get (Î £ X) 2, the scores are first summarized and the total is square. Using the example, (Î £ X) = 2 (3 + 5 + 7 + 8) 2 = 232 = 529. A problem with means and standard deviations is that not convey enough information for us to make meaningful or accurate interpretations assessments some data. Other metrics are designed to more accurate interpretations. The Z score TESTING IN PSYCHOLOGICAL AND V LIFE 2. 3 Terms and symbols used to describe populations and samples Population Definition of sample sampleElements with the same deÃ¯â¬ nizione a subset of the population, usually attracted to represent it in a descriptive characteristics Impartial fashion statistical parameters Symbols used to describe Greek Roman symbol for the average M x Symbol for the SD S S deviation chapter 2 to standards and Basic test statistics 43 transforms data into standardized units that are easier to interpret. The AZ results is the erence ofÃ¯â¬ between a score and the average, divided for the standard deviation â € œ Â½ xi Ã ¢ â â âº XS in other words, a score z is the deviation of a score XI From the X Media to a standard unit deviation. If the score is equal to the average, then his z score is 0. For example, suppose the score and average are both 6; Then 6 to 6 = 0. Zero divided by nothing is still 0. If the score is higher than average, the Z score is positive; If the score is less than average, the Z score is negative. LETA s look for an example. Suppose that XI = 6, the average x = 3, and the standard deviation s = 3. By inserting these values in the formula, we get â € œ Â½ â½ â € â â âº3 3 Ã, Â½ Ã,Â â € Â½1 3 3 Let s Test another example. Suppose XI = 4, x = 5.75, and s = 2.11. What is the Z score? This is a .83 â â â½ 4 Ã ¢ â â â â € â € â € ¢ â â â âº1.74 â â â â½ Ã â â â â â½aâ â â â â â â € â € â â â € â € ™ 83 2.11 2.11 This means that the score we observed (4) .83 Standard deviation under the average, or that the score is lower than the average but its Erence diÃ¯â¬ from the average is slightly lower than the average deviation. Example of depression in medical students: center for epidemiological studies depression stairs (CES-D) CES-D is a general measure of depression that has been widely used in epidemiological studies. The scale includes 20 articles and taps depressed mood size, despair, loss of appetite, sleep disorders, and the level of energy. Every year, the students of the University of California, San Diego, medicine faculty are required to report as often they have experienced a particular symptom during the week Ã¯â¬ before school on a 4-point scale that recently mine or nobody Time [0 to 1 Day (0)] For most or all the time [5 to 7 days (3)]. Articles 4, 8, 12 and 16 on the CES-D are contrary marked. For these elements, 0 is marked by 3, 1 is marked by 2, 2 as 1, and goal of 3 to 0. The CES-D is obtained by adding the circled numbers. The scores on the CES-D range from 0 to 60, with scores greater than 16 indicate clinically detective levels of depressive symptoms in adults. Feel free to take the CES-D measure alone. Calculate your score by adding the numbers you are looking for. However, it is necessary first Ã¯â¬ reverse scores to points 4, 8, 12 and 16. As you will see in Chapter 5, the CES-D has not raised validity for clinical depression determination. If the score is less than 16, the evidence suggests that it is not clinically depressed. If the score is high, it raises suspicions on depressionÃ ¢ even if this doesn't mean you have a problem. (Of course, we recommend talking to your college consultant if you feel depressed.) 44 Chapter 2, Standards and statistical base for testing center for epidemiological studies Depression Scale (CES-D) Instructions: Circle The number for each education Which best describes how often he felt or behaved in this way during last week. RR rarely or noddom (less than 1 day) Some or some time (1st 2 days) Occasionally or a moderate amount of time (3a 4 days) most or all the time (5th 7 days) 1. I was disturbed by things that usually Dona t worry me. . . . . 0 .... 1 .... 2 .... 3 2. I didn't feel like eating. . . . . 0 1 2 .... 3 3. I felt that I could not shake off the blues, even with the help of Family or friends. . . . . 0 .... 1 .... 2 .... 3 4. I heard that I was just as good as other people. . . . . 0 .... 1 .... 2 .... 3 5. I had trouble keeping . . . . 6. I felt depressed. . . . . 0 .... 1 .... 2 .... 3 7. I heard that everything I did was an effort. 0 .... 1 .... 2 .... 3 8. I felt confident about the future. . . . . 0 .... 1 .... 2 .... 3 9. I thought my life had been a failure. . 0 ..... 1 .... 2 ..... 3 10. I felt feared. . . . . I've been happy. . . . . 0 ..... 1 .... 2 .... 3 13. I spoke less than usual. . . . . I had fun in life. . . . . I had crying spells. . . . . . I heard people despised me. . . . . In creating frequency distribution for CES-D scores of medical students we used an arbitrary class range of 5. CHAPTER 2 TA B L E ■ Basic standards and statistics for testing 45 2.3 The average calculation, standard deviation and Z scores for CES-D Test score (X) John Carla Fred Monica Eng Fritz Mary Susan Debbie Elizabeth Sarah Marcel Robin Mike Carl Phyllis 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 58 Now we consider standard normal distribution because it is central to psychological statistics and tests. First, however, you should participate in a short exercise. Take any coin and fl ip 10 times. Now repeat this exercise of 10 coins fl ips 25 times. Record the number of heads you see in each group of 10 laps. When it comes to Standards and Basic Statistics for Test 6 Frequency of Occurrence CAPPTER 2 5 4 3 2 1 0 FIGURA 0 2.6 Frequency of Occurrence 46 1 2 3 4 5 6 Number of heads 7 8 9 10 Frequency distribution of the number of heads in 25 set of 10 heads. .3413 .00135 .00135 .132 -3130214 The theoretical distribution of the number of heads in an infinite number of tokens. done, make a frequency distribution showing how many times you have observed 1 head in 10 fl ips, 2 heads, 3 heads, and so on. Frequency distribution may seem like the example shown in Figure 2.6. The most frequently observed events are approximately equal numbers of heads and tails. Towards the extremes of 10 heads and 0 tails or 10 tails and 0 heads, the events are observed with decreasing frequency. For example, there were no occasions when less than 2 heads were observed and only an opportunity where8 heads were observed. This is what we expect from the laws of probability. On average, we would expect half of the fl ips to show heads and half to show queues if heads and tails are similarly probable events. Although observing a long string of heads or tails is possible, it is unlikely. In other words, sometimes we see that the coin rises in 9 out of 10 fl ips. The probability of this happening, however, is quite small. Figure 2.7 shows the theoretical distribution of the heads in an infinite number of fl ips of the coin. This figure might seem a bit like distribution from your exercise coincidefl ipping or distribution shown in Figure 2.6. Actually, this is a normal distribution, or what is known as a symmetric distribution of binomial probability. CHAPTER 2 ■ Basic standards and statistics for testing 47 On most occasions, we refer to units on the X (or horizontal) axis of normal distribution in Z scoring units. Any variable turned into Z score unit assumes special properties. First, Z scores have an average of 0 and a standard deviation of 1.0. If you think about this for a minute, you should be able to understand why it is true. We remember that the sum of deviations around the medium is always equal to 0. The Z score equation numberer is the deviation around the middle, while the denominator is constant. Thus, the average of Z scores can be expressed as .(X i X)/S N or .Z N Why Σ(Xi − X) will always be equal 0, the medium of Z scores will always be 0. In figure 2.7, the standardized score or Z, the units are marked on the X axis. The numbers under the curve are the proportions of the cases (in decimal form) that we expect to observe in each area. By multiplying these proportions of 100 percentage yields. For example, we see that 34.13% or .3413 of cases fall between the average and a standard deviation above the average. Do not forget that 50% of the cases fall below average. Putting together these two bits of information, we can conclude that if a score is a standard deviation over the middle, then it is about the 84th percentile grade (50 + 34.13 = 84.13 to be exact). A score that is a standard deviation below the middle would be about the 16th percentile grade (50 − 34.13 = 15.87). So, you can use what you have learned on the means, standard deviations, Z scores, and the normal curve to transform the raw scores, which have little meaning, in scores per centile, which are easier to interpret. These methods can only be used when the score distribution is normal or roughly normal. Methods for non-normal distributions are discussed in most statistical books under "non-parametric stations". Percentages and Z Scores These grades per centile are the percentage of scores falling under the observed Z score. For example, the Z −1.6 score is associated with the 5.48 percent degree. The Z 1.0 score (third column) is associated with the rank per centile of 84.13. Part I of Appendix 1 is a simplified version of Part II, which is necessary for more advanced use of Z scores. The second part gives the areas between the average and various Z scores. The standard scored values are listed in the "Z" column. To find the percentage of the distribution between the means of the distribution and a given Z score, you need to locate the item indicated by a specific Z score. Z scores are taken to a second decimal point in the columns that cross the table. First, consider the second column of the table because it is similar to Part I of Appendix 1. Take the Z score of 1.0. The second column is labeled .00, which means that the second decimal place is also 0. The number shown in the table is .3413. Because this is apositive, it's above average. Since the area below the medium is .5, the total area under a Z score of 1.0 is .5 + .3413 = .8413. To make this in a per centile (as shown in Part I of the appendix), multiply by 100 to get 84.13. Now try the example of a Z score of 1.64. 1.64.locate this value, find 1.6 in the first column. Then move your hand through the line until you reach the number 04 under the heading The number is 4495 Once again, this is a positive z score, so you need to add the percentage observed at 48 .5 CHAPTER 2 â Standards and basic statistics proof that falls below the average. The proportion is less than 1.64 9495 Stated another way, 94.95% of the cases fall under a Z score of 1.64 Now try to find the degree percentualele of cases that fall under a Z score of 1.10 if you use the table properly, you should get 86.43 Now try â © .75 Since this is a negative Z-score, the percentage of cases that fall below average should be less than 50. But there are no negative values in part II of Appendix 1. For a Z score negative, there are several ways to get the appropriate area under the curve. The tables in Appendix 1 damage from the media area to score Z. For a Z score of â .75, the area between the mean and the Z score is 2734 You can find this by placing the table in the row .7 labeled and then moving through the line until it reaches the figure in that row under the heading 05 Here the number 2734 We know that .5 cases fall under the average. So, for a negative Z-score, we can get the percentage of cases that fall under the score subtracting .2734, the value listed in Appendix signposted, 5 In this case, the result is .5 = .2734 â © .2266 Since finding the ranks to centoile associated with negative Z scores can be difficult, you might want to use part I of Appendix 1 to see if it is the right balpark. This table gives both Z negative points than positive, but does not give the detail associated with the second decimal place. Watch on â .7 in Part I. The percentualele degree is 24.20 Now consider a Z score of â .8 That is able to centole 21:19 So, you know that a Z score of â .75 should it is associated with a degree percentualele between 21:19 and 24.20 actually, we calculated that the current level is 22.66 per centoile Practice with Appendix 1 to when you are sure you understand how it works. Do not hesitate to ask your professor or teaching assistant if you are confused. The thing is an important concept that you will need for the rest of the book. After learning to use the tables in Appendix 1, you could try a nifty website (http /: davidmlane.com/hyperstat/z_table.html) that can fi nd the odds for you. Look at another example from Table 2.2 (in fall in San Diego, page 32). California has had a dry year in 1999 and 2007. In both years, the newscasters have often commented that this was highly unusual. He described it as â lâ The Ninaa effect, and some even claimed that global warming. The question is whether the amount of precipitation received in 1999 and in 2007 has been unusual given what we know about the rainfall generally. To assess this, calculate the Z score for precipitation. According to Table 2.2, in 1999 there were 6.51 inches and 3.62 inches of precipitation in 2007. The average for precipitation is 10.33 inches and the standard deviation is 4.71 So, the Z score for 1999 is to 6:51 â 'â º10.33 "½A" º.81 4.71 next determine where a Z score of .81 â falls into Z. distribution According to Appendix 1, a â Z score .81 is equal to 20.9 Â ° centoile (50 â 29.1) So, the year low rainfall in 1999 was unusual, as every year, was about 21 Â ° centoile. However, it was not so unusual. It can be estimated that there would be less CHAPTER 2 â Standards and Basic Statistics for testing 49 rainfall in about 17% of all years. 2007 was a different case. The Z score for 2007 was â 1:43 Rainfall in 2007 was in 7,64esimo percentile. (Using the Appendix 1, you can search for Z score of −1.43 and find an area below average of 0.4236 Then you can estimate per centile as 50 − 42.36 = 7.64) You can also turn the process. Instead of using Z scores to find theRankings, you can use ranks Percentile for Ã¯â¬ nd the corresponding z scores. To do this, look at Part II of Appendix 1 under percentili and Ã¯â¬ nd the corresponding Z score.

For example, suppose you want the Z score associated with 90 ° percentile. When entering Part II of Appendix 1, look for the value closer to 90 ° percentile. This can be a bit complicated due to the way the table is structured. Because the 90 ° percentile is associated with a positive z score, you are actually looking for the area over 50 ° percentile. So you should look for the closer input to .4000 (.5000 + .4000 = .9000). The value closer to .4000 is .3997, which is located in the labeled line 1.2 and the labeled column .08. This tells you that a person who gets a z of 1.28 score is about 90 Â ° percentile in distribution. Now come back to the example of CES-D scores for medical students (Table 2.3). Monica had a Z score on the CES-D of ã, '44. Using Appendix 1, you can see that she was in the 33rd percentile (obtained as .50 Ã '.1700 = .33 Ã-100 = 33). Marcel, with his Z score of .14, was in 56 Â ° percentile; And Jennie, with a 1.70 z score, was 96% percentile. You may have a few concerns about Monica and Marcel. However, it seems that Jennie is more depressed than 96% of her classmates and you may need to talk to someone. An example near home one of the diÃˆâ¬culties in classification students is that performance is usually evaluated in terms of rough scores, such as the number of elements that a person responds correctly to an exam. You probably have familiar with the experience of having a test brought you back with a number that has little meaning for you. For example, the professor enters the classroom and delivers your test with a 72 on it. You must then wait patiently while he or she draws the distribution on the board and try to put your 72 in some category you understand, like B +. An alternative way to do things would give you a z score as feedback on your performance. To do this, your professor would have subtracted the average score (media) from your score and divide from the standard deviation. If your Z score was positive, you would know immediately that your score was above average; If it were negative, you would know that your performance was lower than average. Suppose your professor tells you in advance that you will be classified on a curve according to the following rigid criteria. If you are in the top of 15% of the class, you will get a A (85 Â ° percentile or above); between 60 Â ° and 84 Â ° percentili, a b; between 20 and 59 ° percentili, to C; between 6 and nineteenth percentile, a D; And in the fifth percentile or lower, a f. using Appendix 1, you should be able to Ãˆâ¬ nd the scores of the Z associated with each of these cuteÃˆâ¬ â,¬ points for the normal scores distributions. Try it alone and then see Table 2.4 to see if you are correct. Looking at the table 2.4, you should be able to determine what your vote would be in this class based on your Z score. If your Z score is 1.04 or higher, you would receive a A; If it were greater than .25 but less than 1.04, 50 Chapter 2 Ã - Standards and basic statistics for the test TA BLE 2.4 Z Score Scoreffs for an evaluation system VENDITILE VENSILE A SYSTEM 85-100 Z Score CUTOÃˆ â ¬ â ¬ 1.04 B 60 "C 20 84 .25 Ã â ¬" 59 Ã Â «Âº.84 D 6 Ã ¢ â ¬" 19A "Âº1.56 f 0 Ã ¢ ¬ â "5 0 Ã ¢ ¬" 5