NEW EDITION

100 Statistical Tests

Gopal K. Kanji

100 STATISTICAL TESTS

New Edition

Gopal K. Kanji

SAGE Publications
London • Thousand Oaks • New Delhi

First edition published 1993, reprinted 1993
Reprinted with corrections 1994
Reprinted 1995, 1997
New edition published 1999

SAGE Publications Ltd
6 Bonhill Street
London EC2A 4PU

SAGE Publications Inc
2455 Teller Road
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd
32, M-Block Market
Greater Kailash – I
New Delhi 110 048

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0 7619 6152 6
ISBN 0 7619 6151 8 (Pbk)

Library of Congress catalog card number 98-61738

Typeset by the Alden Group, Oxford.
Printed in Great Britain by The Cromwell Press Ltd, Trowbridge, Wiltshire.

ACKNOWLEDGEMENTS

The author and publishers wish to thank the following for permission to use copyright material:

The American Statistical Association for Table 16 adapted from Massey, F.J. Jr (1951) 'The Kolmogorov–Smirnov test for goodness of fit', *Journal of the American Statistical Association*, 4(6). Copyright © 1951 by the American Statistical Association; the Biometrika Trustees for Table 33 from Durbin, J. and Watson, G.S. (1951) 'Testing for serial correlation in least squares regression II', *Biometrika*, 38, pp. 173–5; for Table 36 from Stephens, M.A. (1964) 'The distribution of the goodness of fit statistic, U_n^2 II', *Biometrika*, 51, pp. 393–7; for Table 3 from Pearson, E.S. and Hartley, H.O. (1970) *Biometrika Tables for Statisticians*, Vol. I, Cambridge University Press; for Table 12 from Merrington, M. and Thompson, C.M. (1946) 'Tables for testing the homogeneity of a set of estimated variances', *Biometrika*, 33, pp. 296–304; and for Table 7 from Geary, R.E. and Pearson, E.S. (n.d.) 'Tests of normality'; Harcourt Brace Jovanovich Ltd for Tables 38 and 39 from Mardia, K.V. (1972) *Statistics of Directional Data*, Academic Press; and Tables 35, 36 and 37 from Batschelet, E. (1981) *Circular Statistics in Biology*, Academic Press; the Institute of Mathematical Statistics for Table 28 from Hart, B.I. (1942) 'Significance levels for the ratio of the mean square successive difference to the variance', *Annals of Mathematical Statistics*, 13, pp. 445–7; and for Table 29 from Anderson, R.L. (1942) 'Distribution of the serial correlation coefficient', *Annals of Mathematical Statistics*, 13, pp. 1–13; Longman Group UK Ltd on behalf of the Literary Executor of the late Sir Ronald A. Fisher, FRS and Dr Frank Yates FRS for Table 2 from *Statistical Tables for Biological, Agricultural and Medical Research* (6th edition, 1974) Table IV; McGraw-Hill, Inc. for Tables 8, 15, 18 and 31 from Dixon, W.J. and Massey, F.J. Jr (1957) *Introduction to Statistical Analysis*; Macmillan Publishing Company for Table 1(a) from Walpole, R.E. and Myers, R.H. (1989) *Probability and Statistics for Engineers and Scientists*, 4th edition, Table A.3. Copyright © 1989 by Macmillan Publishing Company; Routledge for Tables 4 and 22 from Neave, H.R. (1978) *Statistical Tables*, Allen & Unwin; Springer-Verlag GmbH & Co. KG for Tables 9, 10, 14, 19, 23, 26 and 32 from Sachs, L. (1972) *Statistiche Auswertungsmethoden*, 3rd edition; TNO Institute of Preventive Health Care, Leiden, for Tables 6, 11, 13, 25, 27 and 30 from De Jonge, H. (1963–4) *Inleiding tot de Medische Statistiek*, 2 vols, 3rd edition, TNO Health Research.

Every effort has been made to trace all the copyright holders, but if any have been inadvertently overlooked the publishers will be pleased to make the necessary arrangement at the first opportunity.

PREFACE

Some twenty years ago, it was only necessary to know about a dozen statistical tests in order to be a practising statistician, and these were all available in the few statistical textbooks that existed at that time. In recent years the number of tests has grown tremendously and, while modern books carry the more common tests, it is often quite difficult for a practising statistician quickly to turn up a reference to some of the less used but none the less important tests which are now in the literature. Accordingly, we have attempted to collect together information on most commonly used tests which are currently available and present it, together with a guide to further reading, to make a useful reference book for both the applied statistician and the everyday user of statistics. Naturally, any such compilation must omit some tests through oversight, and the author would be very pleased to hear from any reader about tests which they feel ought to have been included.

The work is divided into several sections. In the first we define a number of terms used in carrying out statistical tests, we define the thinking behind statistical testing and indicate how some of the tests can be linked together in an investigation. In the second section we give examples of test procedures and in the third we provide a list of all the 100 statistical tests. The fourth section classifies the tests under a variety of headings. This became necessary when we tried to arrange the tests in some logical sequence. Many such logical sequences are available and, to meet the possible needs of the reader, these cross-reference lists have been provided. The main part of the work describes most commonly used tests currently available to the working statistician. No attempts at proof are given, but an elementary knowledge of statistics should be sufficient to allow the reader to carry out the test. In every case the appropriate formulae are given and where possible we have used schematic diagrams to preclude any ambiguities in notation. Where there has been a conflict of notation between existing textbooks, we have endeavoured to use the most commonly accepted symbols. The next section provides a list of the statistical tables required for the tests followed by the tables themselves, and the last section provides references for further information.

Because we have brought together material which is spread over a large number of sources, we feel that this work will provide a handy reference source, not only for practising statisticians but also for teachers and students of statistics. We feel that no one can remember details of all the tests described here. We have tried to provide not only a memory jogger but also a first reference point for anyone coming across a particular test with which he or she is unfamiliar.

Lucidity of style and simplicity of expression have been our twin objectives, and every effort has been made to avoid errors. Constructive criticism and suggestions will help us in improving the book.

COMMON SYMBOLS

Each test or method may have its own terminology and symbols but the following are commonly used by all statisticians.

n number of observations (sample size)
K number of samples (each having n elements)

α level of significance
ν degrees of freedom
σ standard deviation (population)
s standard deviation (sample)

μ population mean
$\bar{x}$ sample mean
ρ population correlation coefficient
r sample correlation coefficient

Z standard normal deviate

CONTENTS

INTRODUCTION TO STATISTICAL TESTING

Having collected together a number of tests, it is necessary to consider what can be tested, and we include here some very general remarks about the general problem of hypothesis testing. Students regard this topic as one full of pitfalls for the unwary, and even teachers and experienced statisticians have been known to misinterpret the conclusions of their analysis.

Broadly speaking there are two basic concepts to grasp before commencing. First, the tests are designed neither to prove nor to disprove hypotheses. We never set out to prove anything; our aim is to show that an idea is untenable as it leads to an unsatisfactorily small probability. Secondly, the hypothesis we are trying to disprove is always chosen to be the one in which there is no change; for example, there is no difference between the two population means, between the two samples, etc. This is why it is usually referred to as the null hypothesis, H_0. If these concepts were firmly held in mind, we believe that the subject of hypothesis testing would lose a lot of its mystique. (However, it is only fair to point out that some hypotheses are not concerned with such matters.)

To describe the process of hypothesis testing we feel that we cannot do better than follow the five-step method introduced by Neave (1976a):

Step 1 Formulate the practical problem in terms of hypotheses. This can be difficult in some cases. We should first concentrate on what is called the alternative hypothesis, H_1, since this is the more important from the practical point of view. This should express the range of situations that we wish the test to be able to diagnose. In this sense, a positive test can indicate that we should take action of some kind. In fact, a better name for the alternative hypothesis would be the action hypothesis. Once this is fixed it should be obvious whether we carry out a one- or two-tailed test.

The null hypothesis needs to be very simple and represents the status quo, i.e. there is no difference between the processes being tested. It is basically a standard or control with which the evidence pointing to the alternative can be compared.

Step 2 Calculate a statistic (T), a function purely of the data. All good test statistics should have two properties: (a) they should tend to behave differently when H_0 is true from when H_1 is true; and (b) their probability distribution should be calculable under the assumption that H_0 is true. It is also desirable that tables of this probability distribution should exist.

Step 3 Choose a critical region. We must be able to decide on the kind of values of T which will most strongly point to H_1 being true rather than H_0 being true. Critical regions can be of three types: right-sided, so that we reject H_0 if the test statistic is greater than or equal to some (right) critical value; left-sided, so that we reject H_0 if the test statistic is less than or equal to some (left) critical value; both-sided, so that we reject H_0 if the test statistic is *either* greater than or equal to the right critical value *or* less than or equal to the left critical value. A value of T lying in a suitably defined critical region will lead us to reject H_0 in favour of H_1; if T lies outside the critical region we do not reject H_0. We should never conclude by accepting H_0.

Step 4 Decide the size of the critical region. This involves specifying how great a risk we are prepared to run of coming to an incorrect conclusion. We define the significance level or size of the test, which we denote by α, as the risk we are prepared to take in rejecting H_0 when it is in fact true. We refer to this as an error of the first type or a Type I error. We usually set α to between 1 and 10 per cent, depending on the severity of the consequences of making such an error.

We also have to contend with the possibility of not rejecting H_0 when it is in fact false and H_1 is true. This is an error of the second type or Type II error, and the probability of this occurring is denoted by β.

Thus in testing any statistical hypothesis, there are four possible situations which determine whether our decision is correct or in error. These situations are illustrated as follows:

		Situation	
		H_0 is true	H_0 is false
Conclusion	H_0 is not rejected	Correct decision	Type II error
	H_0 is rejected	Type I error	Correct decision

Step 5 Many textbooks stop after step 4, but it is instructive to consider just where in the critical region the calculated value of T lies. If it lies close to the boundary of the critical region we may say that there is some evidence that H_0 should be rejected, whereas if it is at the other end of the region we would conclude there was considerable evidence. In other words, the actual significance level of T can provide useful information beyond the fact that T lies in the critical region.

In general, the statistical test provides information from which we can judge the significance of the increase (or decrease) in any result. If our conclusion shows that the increase is not significant then it will be necessary to confirm that the experiment had a fair chance of establishing an increase had there been one present to establish.

In order to do this we generally turn to the power function of the test, which is usually computed before the experiment is performed, so that if it is insufficiently powerful then the design can be changed. The power function is the probability of detecting a genuine increase underlying the observed increase in the result, plotted as a function of the genuine increase, and therefore the experimental design must be chosen so that the probability of detecting the increase is high. Also the choice among several possible designs should be made in favour of the experiment with the highest power. For a given experiment testing a specific hypothesis, the power of the test is given by $1 - \beta$.

Having discussed the importance of the power function in statistical tests we would now like to introduce the concept of robustness. The term 'robust' was first introduced in 1953 to denote a statistical procedure which is insensitive to departures from the assumptions underlying the model on which it is based. Such procedures are in common use, and several studies of robustness have been carried out in the field of 'analysis of variance'. The assumptions usually associated with analysis of variance are that the errors in the measurements (a) are normally distributed, (b) are statistically independent and (c) have equal variances.

Most of the parametric tests considered in this book have made the assumption that the populations involved have normal distributions. Therefore a test should only be carried out when the normality assumption is not violated. It is also a necessary part of the test to check the effect of applying these tests when the assumption of normality is violated.

In parametric tests the probability distribution of the test statistic under the null hypothesis can only be calculated by an additional assumption on the frequency distribution of the population. If this assumption is not true then the test loses its validity. However, in some cases the deviation of the assumption has only a minor influence on the statistical test, indicating a robust procedure. A parametric test also offers greater discrimination than the corresponding distribution-free test.

For the non-parametric test no assumption has to be made regarding the frequency distribution and therefore one can use estimates for the probability that any observation is greater than a predetermined value.

Neave (1976b) points out that it was the second constraint in step 2, namely that the probability distribution of the test statistic should be calculable, which led to the growth of the number of non-parametric tests. An inappropriate assumption of normality had often to be built into the tests. In fact, when comparing two samples, we need only look at the relative ranking of the sample members. In this way under H_0 all the rank sequences are equally likely to occur, and so it became possible to generate any required significance level comparatively easily.

Two simple tests based on this procedure are the Wald–Wolfowitz number of runs test and the median test proposed by Mood, but these are both low in power. The Kolmogorov–Smirnov test has higher power but is more difficult to execute. A test which is extremely powerful and yet still comparatively easy to use is the Wilcoxon–Mann–Whitney test. Many others are described in later pages of this book.

EXAMPLES OF TEST PROCEDURES

Test 1 *Z*-test for a population mean (variance known)

Hypotheses and alternatives

1. H_0: $\mu = \mu_0$
   H_1: $\mu \neq \mu_0$
2. H_0: $\mu = \mu_0$
   H_1: $\mu > \mu_0$

Test statistics

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

n is sample size
$\bar{x}$ is sample mean
σ is population standard deviation

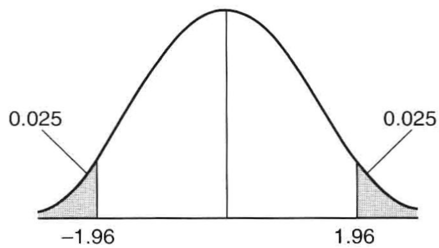
When used

When the population variance σ^2 is known and the population distribution is normal.

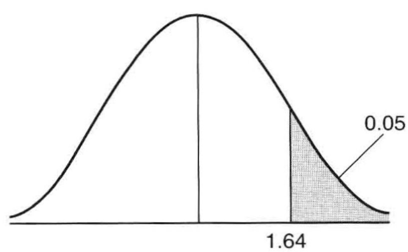
Critical region

Using $\alpha = 0.05$ [see Table 1]

1.



2.



Data

H_0: $\mu_0 = 4.0$
$n = 9$, $\bar{x} = 4.6$
$\sigma = 1.0$
$\therefore Z = 1.8$

Conclusion

1. Do not reject H_0 [see Table 1].
2. Reject H_0.

Test 3 *Z*-test for two population means (variances known and unequal)

Hypotheses and alternatives

1. H_0: $\mu_1 - \mu_2 = \mu_0$
   H_1: $\mu_1 - \mu_2 \neq \mu_0$
2. H_0: $\mu_1 - \mu_2 = \mu_0$
   H_1: $\mu_1 - \mu_2 > \mu_0$

Test statistics

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{\frac{1}{2}}}$$

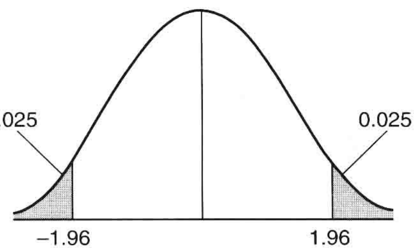
When used

When the variances of both populations, σ_1^2 and σ_2^2, are known. Populations are normally distributed.

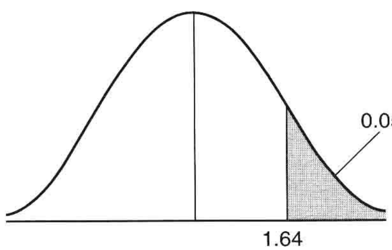
Critical region

Using $\alpha = 0.05$ [see Table 1]

1.



2.



Data

H_0: $\mu_1 - \mu_2 = 0$
$n_1 = 9$, $n_2 = 16$
$\bar{x}_1 = 1.2$, $\bar{x}_2 = 1.7$
$\sigma_1^2 = 1$, $\sigma_2^2 = 4$
$\therefore Z = -0.832$

Conclusion

1. Do not reject H_0.
2. Do not reject H_0.

Test 7 *t*-test for a population mean (variance unknown)

Hypotheses and alternatives

1. $H_0\colon \mu = \mu_0$
   $H_1\colon \mu \neq \mu_0$
2. $H_0\colon \mu = \mu_0$
   $H_1\colon \mu > \mu_0$

Test statistics

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

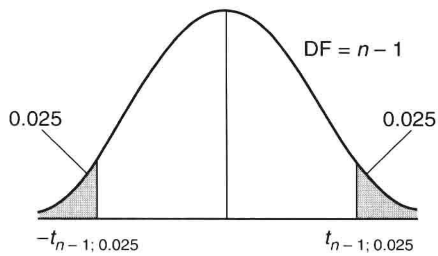
where

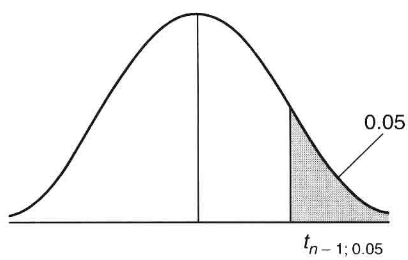
$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}.$$

When used

If σ^2 is not known and the estimate s^2 of σ^2 is based on a small sample (i.e. $n < 20$) and a normal population.

Critical region and degrees of freedom

1.


2.


Data

$H_0\colon \mu_0 = 4.0$
$n = 9,\ \bar{x} = 3.1$
$s = 1.0$
$\therefore\ t = -2.7$

Conclusion

1. $t_{8;0.025} = \pm 2.306$ [see Table 2].
   Reject H_0.
2. $t_{8;0.05} = -1.860$ (left-hand side) [see Table 2].
   Reject H_0.

Test 8 *t*-test for two population means (variance unknown but equal)

Hypotheses and alternatives

1. H_0: $\mu_1 - \mu_2 = \mu_0$
   H_1: $\mu_1 - \mu_2 \neq \mu_0$
2. H_0: $\mu_1 - \mu_2 = \mu_0$
   H_1: $\mu_1 - \mu_2 > \mu_0$

Test statistics

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}}}$$

where

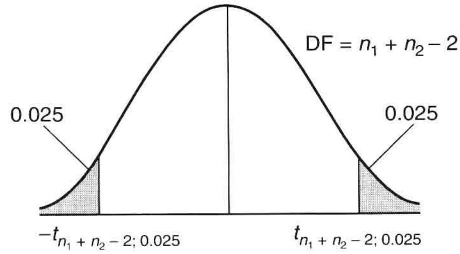
$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

When used

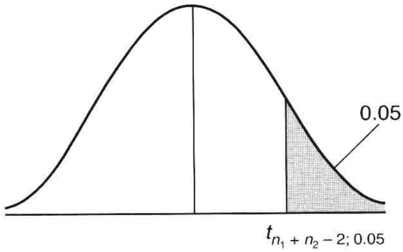
Given two samples from normal populations with equal variances σ^2.

Critical region and degrees of freedom

1.



2.



Data

H_0: $\mu_1 - \mu_2 = 0$
$n_1 = 16$, $n_2 = 16$
$\bar{x}_1 = 5.0$, $\bar{x}_2 = 4$
$s = 2.0$
$\therefore t = 1.414$

Conclusion

1. $t_{30;\, 0.025} = \pm 2.042$ [see Table 2].
   Do not reject H_0.
2. $t_{30;\, 0.05} = 1.697$ [see Table 2].
   Do not reject H_0.

Test 10 Method of paired comparisons

Hypotheses and alternatives

1. H_0: $\mu_d = 0$
   H_1: $\mu_d \neq 0$
2. H_0: $\mu_d = 0$
   H_1: $\mu_d > 0$

Test statistics

$$t = \frac{\bar{d} - \mu_d}{s/\sqrt{n}}$$

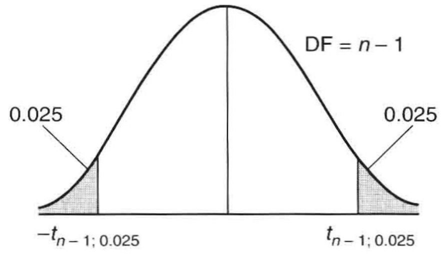
where $d_i = x_i - y_i$, the difference in the n paired observations.

When used

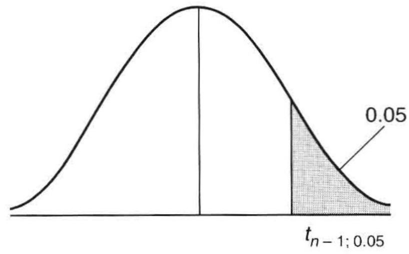
When an experiment is arranged so that each observation in one sample can be 'paired' with a value from the second sample and the populations are normally distributed.

Critical region and degrees of freedom

1.



2.



Data

$n = 16$, $\bar{d} = 1.0$
$s = 1.0$
$\therefore t = 4.0$

Conclusion

1. $t_{15;0.025} = \pm 2.131$ [see Table 2].
   Reject H_0.
2. $t_{15;0.05} = 1.753$ [see Table 2].
   Reject H_0.

Test 15 χ^2-test for a population variance

Hypothesis and alternatives

1. H_0: $\sigma^2 = \sigma_0^2$
   H_1: $\sigma^2 \neq \sigma_0^2$
2. H_0: $\sigma^2 = \sigma_0^2$
   H_1: $\sigma^2 > \sigma_0^2$

Test statistics

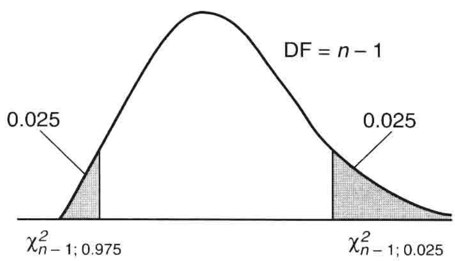
$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

When used

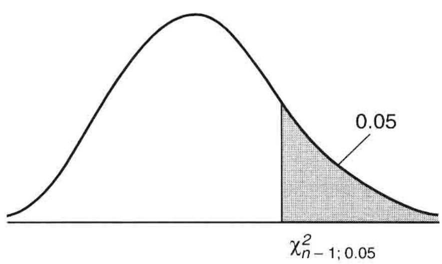
Given a sample from a normal population with unknown variance.

Critical region and degrees of freedom

1.



2.



Data

H_0: $\sigma^2 = 4.0$
$n = 17$, $s^2 = 7.0$
$\therefore \chi^2 = 28.0$

Conclusion

1. $\chi^2_{16;0.025} = 28.85$ [see Table 5].
   $\therefore$ Do not reject H_0.
2. $\chi^2_{16;0.05} = 26.30$ [see Table 5].
   $\therefore$ Reject H_0.