Adjusting the α-level for Sample Size

Scott M. Berry

Department of Statistics, Texas A&M University, College Station, TX 77843, U.S.A.

Kert Viele

Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

Summary

We review hypothesis testing results of Schervish, Seidenfeld, and Kadane (1990), Jeffreys (1961), Lindley (1972) and others that show using a constant α-level across different sample sizes results in preferring an inadmissible test. We derive the rates (for testing normal means) that α must decrease with n to avoid preferring an inadmissible test. We show that adjusting α according to these rates corresponds to performing hypothesis tests based on Bayes Factors, providing a resolution to Jeffreys's paradox.

Some key words: α-level, Bayes Factor, Hypothesis Testing, Jeffreys's Paradox.

1 Introduction

One of the most widely used (and misused) statistical techniques is the Neyman-Pearson (NP) hypothesis test. Much has been written about the failings of the NP approach (Jeffreys (1961); Edwards, Lindman, and Savage (1963); Berger and Delampady (1987); Berger and Berry (1988); and many others). One shortcoming of NP is that it is not well suited to a decision-making framework. The approach is intended to be "objective", without explicit use of utilities or prior opinions. Any choice of utilities should enter the test through choosing the probability of type I error, usually referred to as the α-level of the test. The way this should be done is not clear. This paper examines the relationship between the sample size, n, and the α-level. We show that the standard approach of using a common α-level for different sample sizes leads to some inconsistencies and how to change the α-level with n to avoid these inconsistencies.

In section 2 we review results of Jeffreys (1961,pg.247), Lindley (1972, pg. 14) and Schervish, Seidenfeld, and Kadane (1990) that show keeping the α-level constant with n is irrational in that it results in preferring an inadmissible test. These results also shed light on how to avoid preferring an inadmissible test. In sections 3.1 and 3.2 we derive (for testing normal means) how the α-level should vary with n to maintain rationality. If both hypotheses are simple α should decrease exponentially. Suppose that data are distributed $N(\mu, 1)$ and one is testing $H_0 : \mu = 0$ against $H_A : \mu = 1$. A consequence of our result is that if one is willing to accept $\alpha = 0.05$ for $n = 30$, then one should use $\alpha = 0.0000054$ when $n = 100$. This extremely small α-level indicates that those willing to accept $\alpha = 0.05$ when $n = 30$ are "wasting" (relative to their implied loss function) much of the information contained in the observations when n

is moderately large. This result also applies in the composite alternative case when the decision-maker is willing to "ignore" some region around the null hypothesis. In the general composite alternative case we use a weighted type II error function to find the rate that α should decrease. The rate is slower for a general composite alternative, but still appreciable. In section 3.3 we show that the rates given in sections 3.1 and 3.2 correspond to performing a test based on Bayes Factors.

In section 4 we provide some discussion of the theoretical results. We show that changing the α-level with n provides a resolution to Jeffreys's paradox. This paradox states that if the p-value of a test is constant with n, then the posterior probability of the null hypothesis tends to 1. By appropriately decreasing the α-level with n, a fixed p-value will eventually result in accepting the null hypothesis as n increases.

2 Indifference Curves Argument

In this section we show that keeping the α-level constant with n results in preferring an inadmissible test. We show this by reviewing results of Schervish, Seidenfeld, and Kadane (1990) and Lindley (1972) that show nonparallel indifference curves in (Type I Error, Type II Error) result in preferring an inadmissible test. We provide an example (a test for a normal mean) that shows maintaining a constant α levels results in nonparallel indifference curves.

Consider testing two simple hypotheses. We observe $\bar{X}_n \sim N(\mu, \sigma^2/n)$ with σ^2 known and wish to test $H_0 : \mu = 0$ versus $H_1 : \mu = a$ where $a > 0$. For each n we choose a constant α_n and use the uniformly most powerful level α_n test. We denote this test by $T_n(\alpha_n)$. This test rejects H_0 if $\bar{X}_n > \Phi^{-1}(1 - \alpha_n)\sigma/n^{1/2}$ where Φ is the standard normal distribution function. We let $\delta_T(x)$ be the decision function for the test T, so $\delta_T(\bar{X}_n) = 1$ if T rejects H_0 and $\delta_T(\bar{X}_n) = 0$ if T accepts H_0. The probability of type I error (rejecting the null hypothesis when H_0 is true) for $T_n(\alpha_n)$ is α_n by construction. The type II error function (the probability of accepting H_0 when H_1 is true) depends on n, the α-level, and the value of μ. We denote the type II error function of $T_n(\alpha_n)$ as $\beta_n(\alpha_n, a) = \mathrm{pr}(\delta_{T_n(\alpha_n)} = 0 | H_1, \mu = a)$. A test T dominates a test T' when both errors for T are no greater than those for T' and at least one of the errors for T is strictly less than that for T'.

This section discusses choices for the sequence α_n. We show that it is irrational to keep α_n constant and to maintain rationality one must keep $d/d\alpha_n\ \beta_n(\alpha_n, a)$ constant. Intuitively, the problem with keeping α_n constant with n is that one is not taking advantage of the tradeoff between α_n and $\beta_n(\alpha_n, a)$. In the simple-simple hypothesis case (the composite case is analogous but more cumbersome) as n becomes large a decision-maker should be able to tell which hypothesis is true with extreme accuracy since the sample mean converges to the value of μ. It is therefore possible to make both α_n and $\beta_n(\alpha_n, a)$ small for large values of n.

By keeping the same α-level for different sample sizes the statistician is essentially specifying two different loss structures. When the sample size is n_1 the statistician uses a test with type I and type II errors, α, and $\beta_{n_1}(\alpha, a)$, respectively. If the statistician receives more data, creating a sample size of $n_2 > n_1$, and uses the same α level, then the statistician is accepting errors of α and $\beta_{n_2}(\alpha, a) < \beta_{n_1}(\alpha, a)$. By imposing different loss structures for the same decision problem the statistician is behaving irrationally. The statistician should be using some of the increased power to decrease α.

2.1 *Inadmissibility*

Seidenfeld, Schervish, and Kadane(1990) show that $d/d\alpha_n\ \beta_n(\alpha_n, z)$ must be constant with n for the decision-maker to be rational. In other words, the indifference curves across (Type I Error, Type II error) space must be parallel.

** Figure 1 about here **

Figure 1 shows the curves $\beta_5(\alpha, 1)$ and $\beta_{10}(\alpha, 1)$ for testing $H_0 : \mu = 0$ against $H_A : \mu = 1$ with $\sigma = 1$. The dotted lines are the tangents to the curves at $\alpha = 0.05$. Note that they are not parallel, and that the tangent line to β_5 is steeper than the tangent line to β_{10}. Suppose a decision-maker prefers to use test $T_5(0.05)$ with errors (0.05,0.2772) when $n = 5$ and test $T_{10}(0.05)$ with errors (0.05,0.0646) when $n = 10$. Because the chosen test is preferred to all other tests with the same sample size, this implies that $T_5(0.05)$ is preferred to $T_5(0.06)$ (with errors (0.06,0.2478)) and that $T_{10}(0.05)$ is preferred to $T_{10}(0.04)$ (with errors (0.04,0.0790). Together these preferences imply that the randomized test $T = 0.5T_5(0.05)+0.5T_{10}(0.05)$ is preferred to the randomized test $T' = 0.5T_5(0.06) + 0.5T_{10}(0.04)$ (consider someone flipping a fair coin to determine n and then the decision-maker chooses which test to use). However, T' dominates T because T' has errors $(0.05, 0.1634)$ while T has errors $(0.05, 0.1709)$. The decision-maker therefore prefers to use an inadmissible test. The differing slopes have allowed us to use the geometry of the type II error functions to construct a dominating test.

2.2 *Expected Utility Approach*

It is also possible to show that keeping α_n constant with the sample size leads to inconsistencies using an expected utility approach (see Lindley (1972) Savage (1972), DeGroot (1970)). Suppose the loss structure for this decision problem is

	H_0	H_1
reject H_0	a_{11}	a_{12}
accept H_0	a_{21}	a_{22}

where outcomes a_{12} and a_{21} are each preferred to either outcome a_{11} or a_{22}. By assuming act-state independence the loss matrix can be rewritten as

	H_0	H_1
reject H_0	c	0
accept H_0	0	1

The expected loss from $T_n(\alpha_n)$ is $\pi_0\alpha_n c+(1-\pi_0)\beta_n(\alpha_n, a)$, where π_0 is the probability of H_0. Minimizing the expected loss with respect to α_n requires

$$\frac{d}{d\alpha_n}\left[\pi_0\alpha_n c + (1-\pi_0)\beta_n(\alpha_n, a)\right] = \pi_0 c + (1-\pi_0)\frac{d}{d\alpha_n}\beta_n(\alpha_n, a) = 0$$

so $d/d\alpha_n\ \beta_n(\alpha_n, a) = -c\pi_0/(1-\pi_0)$. This shows that keeping $d/d\alpha_n\ \beta_n(\alpha_n, a)$ fixed results in the optimal choice of α_n. In order to maintain rationality, the indifference curves across type I error and type II error must be parallel lines.

3 Theoretical Results

In this section we derive the rates at which α_n should change with n to give parallel indifference curves. Each section assumes that data $X = (X_1, \ldots, X_n)$ are distributed normally with unknown mean μ and known variance σ^2. Section 3.1 shows the rate for a simple alternative while section 3.2 shows the rate for a composite alternative. In section 3.3 we show that the α_n sequences given correspond to performing a hypothesis test based on Bayes Factors.

3.1 *Simple vs Simple Test*

Theorem 1: If we observe data $X = (X_1, \ldots, X_n)$ where X is a random sample from a $N(\mu, \sigma^2)$ population with μ unknown and σ^2 known and we are testing the hypothesis $H_0 : \mu = 0$ against the simple alternative $H_A : \mu = a$ $(a > 0)$, then $d/d\alpha_n\ \beta_n(\alpha_n, a)$ is constant if

$$\alpha_n = 1 - \Phi\left(\frac{2\sigma^2 \log(-C) + na^2}{2\sigma a n^{1/2}}\right).$$

This implies that

$$\alpha_n = O\left(n^{-\frac{1}{2}} \exp\left(-\frac{a^2 n}{8\sigma^2}\right)\right).$$

Proof: $T_n(\alpha_n)$ will reject H_0 when $\bar{X}_n > \Phi^{-1}(1-\alpha_n)\sigma/n^{1/2}$ where Φ is the standard normal distribution function. Thus

$$\beta_n(\alpha_n, a) = \Phi\left(\Phi^{-1}(1-\alpha_n) - \frac{an^{1/2}}{\sigma}\right)$$

We denote the standard normal density by ψ. Taking the derivative of $\beta_n(\alpha_n, a)$ and setting it to a constant we get

$$C = d/d\alpha_n\ \beta_n(\alpha_n, a) = \psi\left(\Phi^{-1}(1-\alpha_n) - \frac{an^{1/2}}{\sigma}\right)\frac{-1}{\psi(\Phi^{-1}(1-\alpha_n))}$$

which simplifies to

$$C = -\exp\left\{-\frac{1}{2}\left[\frac{a^2 n}{\sigma^2} - \frac{2an^{1/2}}{\sigma}\Phi^{-1}(1-\alpha_n)\right]\right\}.$$

Solving for α_n,

$$\alpha_n = 1 - \Phi\left(\frac{2\sigma^2 \log(-C) + na^2}{2\sigma a n^{1/2}}\right). \tag{1}$$

This choice of α_n is intuitive. One rejects the null hypothesis if

$$\bar{X}_n > \frac{\sigma}{n^{1/2}}\left[\frac{2\sigma^2 \log(-C) + na^2}{2\sigma a n^{1/2}}\right],$$

which implies reject H_0 if

$$\bar{X}_n > \frac{2\sigma^2 \log(-C) + na^2}{2an}, \tag{2}$$

or if

$$\bar{X}_n > \frac{\sigma^2 \log(-C)}{an} + \frac{a}{2}.$$

As n tends toward infinity, one rejects the null hypothesis if $\bar{X}_n$ is greater than $a/2$. One chooses the hypothesis closest to the sample mean.

This result also implies that α_n must decrease exponentially fast. By Mill's ratio (Read (1985)) equation (1) implies

$$\alpha_n \asymp \left(\frac{2\sigma a n^{1/2}}{2\sigma^2 \log(-C) + na^2} \right) \exp\left\{ -\frac{1}{2} \left[\frac{2\sigma^2 \log(-C) + na^2}{2\sigma a n^{1/2}} \right]^2 \right\},$$

so

$$\alpha_n = O\left(n^{-1/2} \exp\left\{ -\frac{a^2 n}{8\sigma^2} \right) \right\}.$$

$\square$.

3.2 *Composite Alternatives*

This section deals with composite alternative hypotheses. Here we are testing $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$ using the most powerful level α_n test, $T_n(\alpha_n)$. We only deal with the one-sided test because the details are simpler (the two-sided case differs by a factor or 2).

In this case there are uncountably many values in the alternative hypothesis and the probability of a type II error is defined for one value of the alternative. We account for this by using a weighted type II error function

$$\beta_n(\alpha) = \int_{H_1} \beta_n(\alpha_n, z) w(z) dz.$$

The weight function, w, describes which values in the alternative are included in the weighted type II error function. If w is an indicator function, $I_{[z=a]}$, then this reduces to the simple-simple case covered in section 3.1. To compute the sequence α_n we note that if the weight function integrates to one then the weight function may be considered a prior distribution, allowing us to use the expected utility argument from section 2.2 to motivate keeping the derivative of the type II error function constant. We derive a sequence of α_n that keeps the derivative of the type II error curve asymptotically constant.

This does not imply that one must be Bayesian to use this argument. We show the result holds for an extremely general class of weight functions (all continuous weight functions with positive mass near 0) indicating that as long as there is some belief that the true value of the mean is close to 0 (this should be true if one is conducting the hypothesis test) the results of the theorem apply.

Theorem 2: If w is integrable, continuous at 0, and $w(0) > 0$, then in order for $d/d\alpha_n\ \beta_n(\alpha_n)$ to remain constant as n increases, $\alpha_n = O\left((n \ln n)^{-1/2}\right)$.

Proof

$$\frac{d}{d\alpha_n}\beta_n(\alpha_n) = \frac{d}{d\alpha_n}\int_0^\infty \Phi\left(\Phi^{-1}(1-\alpha_n) - \frac{zn^{1/2}}{\sigma}\right) w(z)dz$$

By exchanging the derivative and the integral,

$$\frac{d}{d\alpha_n}\beta_n(\alpha_n) = \int_0^\infty \psi\left[\Phi^{-1}(1-\alpha_n) - \frac{z\sqrt{n}}{\sigma}\right]\left[\frac{-1}{\psi(\Phi^{-1}(1-\alpha_n))}\right] w(z)dz.$$

$$= -\int_0^\infty \exp\left\{-\frac{n}{2\sigma^2}\left[z - \frac{\sigma\Phi^{-1}(1-\alpha_n)}{n^{1/2}}\right]^2\right\}\exp\left\{\frac{n}{2\sigma^2}\left[\frac{\sigma\Phi^{-1}(1-\alpha_n)}{n^{1/2}}\right]^2\right\} w(z)dz. \tag{3}$$

If we let f_n be the density of a normal distribution with mean $\Phi^{-1}(1-\alpha_n)\sigma/n^{1/2}$ and variance σ^2/n, then equation (3) can be rewritten as

$$\frac{d}{d\alpha_n}\beta_n(\alpha_n) = -\sigma(2\pi/n)^{1/2}\exp\left\{\frac{n}{2\sigma^2}\left[\frac{\sigma\Phi^{-1}(1-\alpha_n)}{n^{1/2}}\right]^2\right\}\int_0^\infty f_n(z)w(z)dz.$$

If w is continuous and positive at 0, then as n tends to infinity the integral converges to $\frac{1}{2}w(0)$. In order for the derivative to remain constant

$$C = -\sigma(2\pi/n)^{1/2}\exp\left\{\frac{n}{2\sigma^2}\left[\frac{\sigma\Phi^{-1}(1-\alpha_n)}{n^{1/2}}\right]^2\right\}\frac{1}{2}w(0),$$

which implies that

$$\alpha_n = 1 - \Phi\left[\left[2\log\left(\frac{-2Cn^{1/2}}{w(0)\sigma(2\pi)^{1/2}}\right)\right]^{\frac{1}{2}}\right]. \tag{4}$$

Applying Mills' ratio yields

$$\alpha_n \asymp \frac{w(0)\sigma}{-2Cn^{1/2}}\left[\log\left(\frac{4nC^2}{w(0)^2\sigma^2 2\pi}\right)\right]^{-\frac{1}{2}},$$

so

$$\alpha_n = O\left((n\log n)^{-1/2}\right).$$

□

3.3 *Agreement with Bayesian Methods*

In this section we show that the asymptotic rates shown in sections 3.1 and 3.2 are equivalent to performing a test based on Bayes Factors. Suppose one chooses to construct a hypothesis test by rejecting the null hypothesis if the observed value of the Bayes Factor is greater than some reference value. In this section we show that if the alternative hypothesis is simple, rejecting based on the Bayes Factor is equivalent to using the α_n sequence shown in section 3.1. We show that correspondence is approximate in the composite case using the Schwarz (1978) approximation to the log of the Bayes Factor.

For a simple alternative, let BF_0 denote a specific value of the Bayes factor for testing $H_0 : \mu = 0$ against $H_A : \mu = a$. After observing data $X = (X_1, ...X_n)$, the Bayes factor against H_0 is

$$BF = \frac{\mathrm{pr}(X|\mu = a)}{\mathrm{pr}(X|\mu = 0)} = \exp\left\{\frac{-1}{2\sigma^2}\sum\left[(X_i - a)^2 - (X_i - 0)^2\right]\right\},$$

which simplifies to

$$BF = \exp\left\{\frac{-na^2}{2\sigma^2} + \frac{2a}{2\sigma^2}\sum X_i\right\}.$$

If one chooses to reject the null hypothesis if BF is greater than the reference value BF_0, then one rejects if

$$\exp\left\{\frac{-na^2}{2\sigma^2} + \frac{2a}{2\sigma^2}\sum X_i\right\} > BF_0,$$

which simplifies to

$$\bar{X}_n > \frac{2\sigma^2 \ln BF_0 + na^2}{2an}.$$

This is the same rejection region as shown in equation (2).

For the composite case we use the Schwarz approximation to the log of the Bayes Factor. For testing the hypotheses $H_0 : \mu = 0$ against $H_A : \mu \neq 0$, the Schwarz approximation to the log of the Bayes factor based on data $X = (X_1, \ldots, X_n)$ is

$$\log BF = \frac{-1}{2\sigma^2}\sum(X_i - \bar{X}_n)^2 + \frac{1}{2\sigma^2}\sum(X_i - 0)^2 - \frac{\log n}{2} + O(1).$$

This simplifies to

$$\log BF = \frac{n(\bar{X}_n)^2}{2\sigma^2} - \frac{\log n}{2} + O(1)$$

Choosing a reference value of the Bayes factor for performing the test, BF_0, we reject H_0 if

$$\frac{n(\bar{X}_n)^2}{2\sigma^2} - \frac{\log n}{2} > \log BF_0.$$

This simplifies to rejecting H_0 if

$$\bar{X}_n > \left(\frac{\sigma^2 \log n + 2\sigma^2 \log BF_0}{n}\right)^{1/2}$$

This rejection region corresponds to

$$\alpha_n = 1 - \Phi\left[\left(\frac{n}{\sigma^2}\frac{\sigma^2 \log n + 2\sigma^2 \log BF_0}{n}\right)^{\frac{1}{2}}\right] = 1 - \Phi\left[\left(2\log(BF_0 n^{1/2})\right)^{\frac{1}{2}}\right]. \qquad (5)$$

The α_n sequence corresponding to this rejection region is not the same as the α_n sequence shown in equation (4). The asymptotic rate is the same, however. Applying Mill's ratio to equation (5),

$$\alpha_n \asymp \frac{1}{BF_0 n^{1/2}}\left[\log(BF_0)^2 n\right]^{-1/2},$$

so

$$\alpha_n = O((n \ln n)^{-1/2}).$$

which is the same rate shown in section 3.2.

4 Discussion

The results in sections 2 and 3 show that adjusting the α-level for sample size removes an important objection from using hypothesis testing as a decision-making tool. As we showed in section 2, keeping the α-level constant with n causes the resulting test to be irrational across different sample sizes.

Intuitively, this irrationality follows from the fact that the data provide a lot of information about the hypotheses. As n increases, the probability of type II error quickly decreases to 0. For large sample sizes a decision-maker has an extremely high probability of choosing the alternative hypothesis if it is true. Keeping the α-level constant with n forces the decision-maker to accept an unnecessarily high probability of making a type I error. By trading some of the increased power to lower the α-level the decision-maker is able to attain small probabilities for both types of errors. Theorem 1 indicates that in the simple-simple case this tradeoff occurs at an exponentially fast rate. If one is testing $H_0 : \mu = 0$ against $H_A : \mu = 1$ when $\sigma = 1$, then if one is willing to use an α-level of 0.05 for $n = 30$, they should use an α-level of 0.0000054 for $n = 100$. This is a vast difference. In the general composite case the rate is slower, but still appreciable. If one is testing $H_0 : \mu \leq 0$ against $H_A : \mu > 0$ when $\sigma = 1$, then if one is willing to use an α-level of 0.05 for $n = 30$, they should use an α-level of .024 for $n = 100$.

We note that in the composite case the rate depends on the w function around 0. It is common to consider the difference between a "statistically significant" difference and a "substantively meaningful" difference. In situations where one can place a region around the null consisting of "meaningless differences" by letting $w(z)$ be zero in some region around 0, then the rates should be at least as fast as in the simple-simple case, regardless of the size of the band around 0. If $w(z)$ is 0 for $z < b$, then the most difficult situation the decision-maker has to deal with is the simple-simple case where $\mu = b$.

The results in section 3.3 show that decreasing the α-level with n also removes some of the inconsistencies between frequentist and Bayesian hypothesis testing procedures. We show that the α_n sequence shown in sections 3.1 and 3.2 correspond to accepting or rejecting the null hypothesis based on a Bayes factor. The null hypothesis is rejected if the observed Bayes Factor is greater than some reference value. The sequences of tests resulting from decreasing α therefore has many properties of Bayes factors. In particular, the tests are consistent. If ones observes $X = (X_1, X_2, \ldots)$ and lets $\delta_n(X)$ be the decision rule based on the first n observations with α decreasing, then the δ_n sequence converges almost surely to the correct hypothesis (Doob (1949)). This is in contrast to leaving α constant with n. A consequence of the law of the iterated logarithm is that the decision rules based on a constant α-level will reject the null hypothesis infinity often, even if the null hypothesis is true.

In the introduction we described Jeffreys's paradox. The paradox shows that a constant p-value results in the posterior probability of the null hypothesis tending to 1 as the sample size increases. Although we have considered only the decision concerning accepting or rejecting the null hypothesis, the adjustment of the α-level also avoids Jeffreys's paradox in interpreting the p-value. For any fixed p-value, as n increases the p-value will eventually be contained in the acceptance region.

References

Berger, J.O. and Berry, D.A. (1988). The Relevance of Stopping Rules in Statistical Inference. In *Statistical Decision Theory and Related Topics IV*, Editors S.Gupta and J. Berger), **1**, 29-47.

Berger, J.O. and Delampady, M. (1987). Testing Precise Hypotheses. *Statistical Science*, **2**, 317-352.

Degroot, M.H. (1970). *Optimal Statistical Decisions.* McGraw-Hill Book Company.

Doob, J. L. (1949). Application of the Theory of Martingales. In *Le Calcul des*

Probabilités et ses Applications, Colloques Internationaux du Centre National de la Recherche Scientifique. 23-27.

EDWARDS, W., LINDMAN, H., and SAVAGE, L.J. (1963). Bayesian Statistical Inference for Psychological Research. *Psychol. Rev.*, **70** 193-242.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Oxford University Press.

LINDLEY, D.V. (1972). *Bayesian Statistics, A Review.* SIAM.

READ, C.B. (1985) Mill's Ratio. In *Encyclopedia of Statistics*, Editors S. Kotz, N.L. Johnson, and C.B. Read, **5**, 504-506.

SAVAGE, L.J. (1972). *The Foundations of Statistics.* Dover Publications, Inc.

SCHWARZ, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics* **6**, 461-464.

SEIDENFELD, T., SCHERVISH, M.J., and KADANE, J.B. (1990). Decisions Without Ordering. In *Acting and Reflecting*, Editor W. Sieg, 143-170.

Figure 1: The inadmissibility of choosing tests with common α-levels