ON TESTING FOR SIGNIFICANT QUANTITATIVE TRAIT LOCI (QTL) EFFECTS WHEN VARIANCES ARE UNEQUAL

Pradeep Singh

Shesh N. Rai

On Testing for Significant Quantitative Trait Loci (QTL) Effects When Variances are Unequal

Pradeep Singh[1] and Shesh N. Rai[2]

[1]Department of Mathematics, Southeast Missouri State University, Cape Girardeau, MO 63701, psingh@semo.edu
[2]Department of Bioinformatics & Biostatistics , University of Louisville, Louisville, KY 40292, shesh.rai@louisville.edu

Abstract

The basic theory of QTL (Quantitative Trait Loci) mapping is to score a population for a quantitative trait according to the marker genotype, and then to use statistics to identify differences associated with the markers and the quantitative trait of interest. Permutation based methods have been used to estimate threshold values for quantitative mapping. The permutation test based on the Student *t*-test for equality of means does not control Type I error rate to its nominal value when variances are unequal. In this study we propose a modification of the Student *t*-test based on the jackknife estimator of population variance. Janssen [20] had proposed a permutation version of the Welch test to compare equality of means under heterogeneous error distributions. The Monte Carlo method is used to compare the type I error rate of the proposed jackknife test, Janssen's permutation test, and permutation test based on the Student *t*-test. The Monte Carlo study also compares the power of the proposed jackknife test, Janssen's permutation test, and permutation test based on the Student *t*-test. Also, the power for each test was calculated and compared after adjusting for Type I error rates.

Key Words: Quantitative Trait Loci, Jackknife Estimator, Student *t*-test, Permutation test, Power, Type I error rate

1. Introduction

The detection and location of genes that control quantitative characters has been a subject that has attracted the curiosity and attention of many researchers and scholars. Liu [26] defined quantitative traits, such as fruit size and plant weight, as traits that are typically known to follow a continuous distribution; and that are often controlled by many genes, each of which have a small effect on the trait. The loci containing these genes, generally stretches of DNA, are called quantitative trait loci, or QTLs.

Substantial gains have been made in last few years in the search for genes affecting the quantitative traits. The characterization of genes will help in the study of clinical diagnosis and enhance animal and plant breeding programs. Doerge *et al*. [9], Doerge [10], and Churchill and Doerge [5] described the many substantial statistical contributions that were carried out through the use of molecular markers and quantitative genetics for the purpose of completely understanding the relationship between genotype and phenotype in human, animal and plant populations.

A common issue encountered in all QTL mapping methods however, is that of determining appropriate threshold values for declaring significant QTL effects. This could be due to several reasons such as sample size, the genome size of the organism under study, and the proportion and pattern of missing data. Churchill and Doerge [4], presented a solution to this problem by providing a simple and statistically sound empirical method, based on the concept of a permutation test. The procedure they

proposed involved repeated "shuffling" of the phenotypic trait values and the generation of a random sample of the test statistic from an appropriate null distribution.

Huang *et. al.* [17] have shown that for two non – identical distributions (such as two normal distributions with unequal variances), using the permutation based method for testing the equality of means can produce inflated type I error rate. The permutation test is liberal in some situations and conservative in other situations.

Two common methods for estimating or approximating the sampling distribution of a test statistic and its characteristics are the jackknife and the bootstrap. Bootstrapping is the practice of estimating properties of an estimator (such as its variance) by measuring those properties when sampling from an approximating distribution. Bootstrapping results in an approximate null distribution. The bootstrap procedure has been used in a lot of studies to determine the location of a QTL. An example of this can be reviewed under the study performed by Manichaikul *et al.* [27] where the Monte Carlo method of simulation was performed to investigate the performance of bootstrap confidence intervals for QTL location. They concluded that in the case of a backcross design with a single segregating QTL, the bootstrap confidence intervals for QTL location varied greatly as a function of the location of the QTL relative to the available genetic markers, and hence was not reliable.

Janssen [20] showed that permutation tests based on studentized statistics are asymptotically of size α . A test was applied to generalized two-sample Behrens-Fisher problem. The test includes variance correction of the permutation distribution. The resulting critical values of the permutation distribution worked well for large samples.

The jackknife procedure, introduced by Quenouille [30] to estimate the bias of an estimator, has rarely been used in the study of QTL location. Typically, the jackknife method of estimation is carried out by deleting one observation each time from the original data set and recalculating the estimator based on the rest of the data. This method has become a more valuable tool since Tukey [40] found that the jackknife can also be used to construct variance estimators. A heuristic justification for using the jackknife in variance estimation is given by Tukey [40].

In this study, we propose a modified Student *t*-statistic computed by replacing the pooled sample variance by the jackknife estimator of variance. We compare the type I error rate of the following tests using the Monte Carlo method.

i. Permutation test using the Student *t* – statistic
ii. Jackknife *t* – test
iii. Janssen's studentized permutation test

We also compare the power and alpha-adjusted power of the above tests using Monte Carlo Method.

2. Single Marker Analysis in Backcross Progeny

2.1 Single Marker Single QTL Analysis

Sax [32], Thoday [39], Elston and Stewart [12], and Edwards *et al*. [11] well established the use of genetic markers to locate QTL. Considerable attention has been paid to the case of associations between a single marker and a quantitative trait. Single marker analysis is simple in terms of data analysis and implementation and can be performed using common statistical software such as SAS. Single marker analysis can be implemented as a simple *t*-test, an analysis of variance, a linear regression, a likelihood ratio test and maximum likelihood estimation. This technique is conducted by analyzing

one marker at a time. Liu [26] brings out the idea that a QTL is determined to be located near a marker if phenotypic values for the trait are significantly different among the marker genotypes.

2.1.1 . QTL Hypotheses

Testing for QTL effects is an attempt to detect or locate a single QTL that affects a quantitative trait. Genetic markers which get transmitted along with specific values of a quantitative trait have more chance of being close to the gene affecting that quantitative trait. For any statistical test, one must be able to state the appropriate hypotheses being tested. There are essentially three types of hypotheses involved when testing for QTL effects. Haley and Knott [15] proposed the two commonly used interpretations with reference to the null hypothesis. The first one rests on the simple idea that no QTL is present anywhere in the genome, which can simply be stated as $H_0^1 : \text{no QTL present}$. According to the second interpretation, there is a QTL present in the genome, but it is not linked to the position where the test is being made in the genome, i.e., $H_0^2 : \text{QTL present}$ and unlinked to the testing position. The alternative hypothesis, however, almost always used in the situation for testing a single QTL is the obvious conclusion, given by $H_A : \text{A QTL present}$ and is linked to the testing position. Doerge et al. [9] brought out the consequences of each null hypothesis in the form of the likelihood used to construct the test statistic. According to their research, under H_0^1, the distribution followed a single normal distribution, while under H_0^2, the distribution followed a mixture of normal distributions where the mixing proportions depended on the position in the genome relative to the genetic ordered genetic markers.

In most of the studies conducted, there has been a common problem of obtaining appropriate significance thresholds (critical values) for the tests being applied. These important issues were all addressed by Lander and Botstein [23, 24], Rebai *et al*. [31] and Churchill and Doerge [4]. According to Churchill and Doerge [6], the defining feature of a threshold value was that, under the assumption of no QTL effects, the value of the test statistic should exceed the threshold with probability not to exceed some nominal level α (e.g., $\alpha = 0.05$).

2.1.2 Single Marker Single QTL Model

Observations on marker genotype and trait value are taken in order to test the hypothesis that the marker is unlinked to the putative QTL. The hypothesis that quantitative trait and marker is unlinked is equivalent to the hypothesis that two markers classes have equal means. Rejection of this hypothesis has a dual implication. Not only does it confirm a genetic basis for the trait, but also it suggests that the trait is affected by a gene (QTL) that is close to the marker. Single-marker analysis is carried out by comparing means of marker classes using Student *t*–test or analysis of variance.

Early work on the association between the trait value and marker segregation patterns has been based on linear models. Liu [26] proposed a linear model for this analysis which can be given by

$$Y_{ijk} = \mu + \mu_j + g(\mu)_{k(j)} + e_{ijk}, \qquad 2.1$$

where Y_{ijk} is the phenotypic quantitative trait value for the k^{th} individual with QTL genotype j at locus i, μ is the population mean, μ_j is the effect of QTL genotype j,

$g(\mu)_{k(j)}$ is the genotypic effect within the QTL genotype *j*, and e_{ijk} is the error term. With QTL genotypes observed, for locus *i*, we test the hypothesis $H_{i0}: \mu_{i1} = \mu_{i2}$ using the Student *t*–test statistic given by

$$t_Q = \frac{\hat{\mu}_{i1} - \hat{\mu}_{i2}}{\sqrt{\hat{s}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \qquad 2.2$$

where $\hat{\mu}_{i1}$ and $\hat{\mu}_{i2}$ are the maximum likelihood estimator (mle) of μ_{i1} and μ_{i2} respectively. Sample sizes of the two classes of QTL genotypes are given by n_1 and n_2, where $n_1 + n_2 = n$ is the total sample size, and the pooled estimate of the variance within the two classes is $\hat{s}^2$; that is,

$$\hat{s}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \qquad 2.3$$

2.1.3 Permutation as a Method of Testing

Permutation test is a popular technique for testing a hypothesis of no effect, when the distribution of the test statistic was unknown. To test the equality of two means, a permutation test might use a test statistic that is the difference of the two sample means in the univariate case or one that is the maximum of the univariate test statistics in the multivariate case. A permutation test would then estimate the null distribution of the test statistic by permuting the observations between the two samples. For example, to test the equality of two means $H_0^{\mu}: \mu_1 = \mu_2$ one might use a test statistic which is the difference of the two sample means, and estimate its null distribution by permuting the observations in the combined Y_1 and Y_2 samples. The basis for permutation testing is if the Y_1 data are sampled from distribution P_{y_1} and the Y_2 data from distribution P_{y_2}, then under the null hypothesis of identical distributions $H_0^P: P_{y_1} = P_{y_2}$ all permutations of the observations are equally probable.

One important point is that the concept of permutation, as proposed by Fisher [13] and as applied to QTL mapping by Churchill and Doerge [4], relies on exchangeability. Churchill and Doerge [7] proposed that in simple experimental designs, such as an intercross or a backcross mapping population, the individual units could safely be assumed to be exchangeable, while such was not the case in more complex designs which works very well with our study as we will be focusing on single marker analysis in a backcross experimental population. An advantage of using permutation tests is that they require relatively few assumptions and can be applied to a wide variety of settings. Permutation based tests to test $H_0^{\mu}: \mu_1 = \mu_2$ are only appropriate when the only difference between the null distribution is the differences in the mean.

3. Proposed New Test

3.1 The Jackknife Estimator

A jackknife sample is a sample that leaves out one observation at a time from the original sample. To get a better idea of the jackknife sample, we consider an original sample of *n* objects. Let $R = (y_1, y_2, \ldots\ldots\ldots, y_n)$ denote this sample. Then, from the description we used above, the i^{th} jackknife sample will be given by $R_i^j = (y_1, y_2, \ldots., y_{i-1}, y_{i+1}, \ldots., y_n)$ where, $i = 1, 2, 3..., n$, and is the data set obtained by removing the i^{th} observation. Note that for an original sample of size *n*, the jackknife can be replicated *n* times where each of the *n* times is a jackknife replication.

Liu [26] explained how an estimator can be obtained from each of the *n* jackknife samples given by $R_1^J, R_2^J, ..., R_n^J$ and stated that if $\hat{\theta}_i^J = F\left(R_i^J\right)$ denotes the estimate for the i^{th} replication, then the jackknife mean, variance and estimate of bias, respectively, can be given by the following formulas:

$$\bar{\theta}^J = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_i^J \qquad 3.1$$

$$\hat{V}^J = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i^J - \bar{\theta}^J\right)^2 \qquad 3.2$$

$$\hat{B}ias^J = (n-1)\left(\bar{\theta}^J - \theta\right) \qquad 3.3$$

where θ is the true parameter. In practice, the true parameter is usually unknown and can be replaced by the estimate $\hat{\theta}$ from the original sample. The coefficients $\frac{n-1}{n}$ and $n-1$ in equations 3.2 and 3.3 for the jackknife variance and bias are chosen to satisfy asymptotic properties of the estimator. The rationale behind them is that the variation among the jackknife samples is smaller than among the bootstrap samples because the jackknife samples are *n* fixed data sets, which are more similar to the original samples than the bootstrap samples.

Sundrum [37] explained how the use of a statistic that is efficient in estimation does not imply that a more powerful test will be obtained compared to that given by a less efficient estimator. Singh, Saxena, and Srivastava [33] proposed a solution to the Behrens's Fisher problem, which by definition arises from testing the equality of two means from normal populations with unequal variances. In their study, they replaced the pooled variance of a regular Student *t*–test with the jackknife estimate of variance. To arrive at this result, they chose two random samples $(y_{11}, y_{12}, \ldots\ldots\ldots, y_{1n_1})$ and $(y_{21}, y_{22}, \ldots\ldots\ldots, y_{2n_2})$ from populations having probability density functions $f_1(y_1, \theta, \alpha)$ and $f_1(y_2, \theta, \beta)$, respectively; and supposed $\hat{\theta}_{n_1,n_2}$ to be the initial estimator of θ based on both samples, $\hat{\theta}^{i,.}_{n_1-1,n_2}$ represented the estimator of θ on deleting the i^{th} observation in the first sample and keeping the second sample intact, and $\hat{\theta}^{.,j}_{n_1,n_2-1}$ represented the estimator of θ obtained on deleting the j^{th} observation in the second

sample and keeping the first sample intact. The pseudo values with respect to the two samples were then obtained by using the same idea that Liu [26] proposed, namely:

$$J_{i.}(\hat{\theta}) = (n_1 + n_2)\hat{\theta}_{n_1,n_2} - (n_1 + n_2 - 1)\hat{\theta}^{i,\cdot}_{n_1-1,n_2}$$
$$J_{.j}(\hat{\theta}) = (n_1 + n_2)\hat{\theta}_{n_1,n_2} - (n_1 + n_2 - 1)\hat{\theta}^{\cdot,j}_{n_1,n_2-1} \qquad 3.4$$

The jackknife estimators, based on the first and second samples, were

$$J_1(\hat{\theta}) = \frac{1}{n_1}\sum_{i=1}^{n_1} J_{i.}(\theta) \text{ and } J_2(\hat{\theta}) = \frac{1}{n_2}\sum_{j=1}^{n_2} J_{.j}(\theta), \qquad 3.5$$

and the jackknife estimator on both samples was

$$J(\hat{\theta}) = \frac{1}{n_1 + n_2}(n_1 J^1(\theta) + n_2 J^2(\theta)) \quad . \qquad 3.6$$

Using 3.6, they obtained the jackknife estimator of common variance of two normal populations as:

$$J(\hat{\sigma}^2) = \frac{1}{n_1 + n_2}(n_1 s_1^2 + n_2 s_2^2) \qquad 3.7$$

where s_1^2 is the estimate of variance of the first sample $(y_{11}, y_{12}, \ldots\ldots\ldots, y_{1n_1})$, and s_2^2 is the estimate of variance with respect to the second sample $(y_{21}, y_{22}, \ldots\ldots\ldots, y_{2n_2})$. The new statistic, which we will refer to as the jackknife *t*-statistic, is

$$t^{**}_{jack} = \frac{\bar{y}_1 - \bar{y}_2}{\left(\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}\right)^{\frac{1}{2}}} \quad . \qquad 3.8$$

Singh, Saxena, and Srivastava [33] also approximated the denominator of 3.8 by a chi-square distribution.

3.2 The Proposed New Statistic

The hypothesis being tested is

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_1 : \mu_1 - \mu_2 \neq 0 \quad ,$$

which means the means of the marker genotypes are the same versus the alternative that they are not the same.

In terms of the null hypothesis provided above, equation 3.8 for the jackknife *t*-statistic can be redefined as

$$t^{**}_{jack} = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\left(\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}\right)^{\frac{1}{2}}} \quad , \qquad 3.9$$

where $\hat{\mu}_1 - \hat{\mu}_2$ refers to the difference in the marker genotypic means between the two marker genotypes, y_1 and y_2.

The critical values will be computed from the distribution obtained by using a chi -square approximation to the denominator of 3.9.

3.3. Chi-square Approximation

Singh, Saxena, and Srivastava [33] approximated of the denominator of 3.9 by a chi-square distribution using the method of moments. The statistic is given by

$$t^{**2} = \frac{\chi^2(1)/1}{\hat{\delta}\left(\chi^2(\nu)/\nu\right)}. \qquad 3.10$$

The distribution of t^{**} could be approximated by $\hat{\delta}^{-\frac{1}{2}} t(\hat{\nu})$, where $t(\hat{\nu})$ was the usual Student's *t*–distribution with $\hat{\nu}$ degrees of freedom and the values of the parameter estimates for δ and ν were given by

$$\hat{\nu} = \frac{\left(\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}\right)^2}{\frac{\left(\frac{s_1^2}{n_2}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_1}\right)^2}{n_2 - 1}}, \text{ and } \hat{\delta} = \frac{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}. \qquad 3.11$$

3.4. Janssen's studentized permutation test

Janssen [20] showed that permutation tests based on studentized statistics are exactly of size α under certain non–i.i.d. null hypotheses. Janssen [20] proposed a permutation version of the Welch test for testing equality of the means under heterogeneous error distribution. The studentized statistic for total sample size $n = n_1 + n_2$ is given by

$$\tilde{T}_n = \frac{T_n}{V_n^{1/2}}, \text{ where} \qquad 3.12$$

$$T_n = \left(\frac{n_1 n_2}{n}\right)^{1/2} (\mu_1 - \mu_2) \text{ and } V_n = \left(\frac{n_1 n_2}{n}\right)\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right). \qquad 3.13$$

It is very important that the permutation procedure takes into account the appropriate variance. To accomplish this, the data are pooled,

$$(Z_1, Z_2, ..., Z_n) := (Y_{11}, Y_{12}, ..., Y_{1n_1}, Y_{21}, Y_{22}, ..., Y_{2n_2}) \qquad 3.14$$

and for fixed outcomes of $(Z_1, \ldots\ldots, Z_n)$ the conditional distribution under random permutations $(\sigma(i))_{i \le n}$, which are independent of the data, is considered. It is given by

$$\sigma \mapsto \tilde{T}_n(Z_{\sigma(1)}, ..., Z_{\sigma(n)}). \qquad 3.15$$

The studentized permutation test is

$$\varphi_{n,perm} = I_{(c_n(1-\alpha))}(\tilde{T}_n), \qquad 3.16$$

which is (1-α) –quantile of the conditional distribution of 3.15 given the data 3.14.

4. Monte Carlo Study

4.1. Simulation

Simulation was performed using Statistical Analysis Software (SAS version 9.1). We generated two independent normal distributions of the quantitative trait, namely y_1 and y_2, for two classes of QTL-marker genotypes, AA and Aa. Since genotypes normally occur in different proportions at each marker, we used different sample sizes. To test the null hypothesis that the two-marker genotypic means are equal, $\mu_1 = \mu_2$, unequal standard deviations were also considered. Different combinations of sample sizes were used for the two normal distributions simultaneously.

The ratios of the two standard deviations used were $k = \frac{\sigma_1}{\sigma_2} = 1, 1.2, 1.5, 2, 3$. Each combination of n_1, n_2 and k was used to generate two normal populations. For every combination, the *t*–statistic and the jackknife *t*–statistic was computed. The critical values at α=.05 were obtained. Also, we generated uniform distribution of size $n = n_1 + n_2$ and ranked the n observations. The pooled data was ranked accordingly. First n_1 observations were used as sample 1 and the remaining n_2 observations were considered sample 2. The studentized statistic given in 3.12 was computed. This procedure was repeated 20000 times to obtain 20000 permutations for each set of combinations. The critical value was $(1-\alpha)*20000^{th}$ ordered value corresponding to a given value of α.

The jackknife *t*-statistic was compared with respect to the critical value obtained from 3.10. The studentized permutation statistic given in 3.12 was compared with critical value obtained from its permutation distribution. Type I error rates were computed by repeating the process 5000 times for each combination and calculating the proportion of rejections. The results of the comparison of type I error rates are given in tables 1-8.

The power comparisons were done for $\mu_1 - \mu_2 = 1, 2, 3, 4$. Alpha adjusted power was obtained by first finding the 'alpha-adjusted critical level' by obtaining the 5^{th} percentile of the 5000 p-values when $H_0 : \mu_1 - \mu_2 = 0$ was true. Estimated powers of the tests are given in tables 9-12. The following notations were used.

P_t: studentized permutation distribution
t: pernmutation test using Student *t* –test statistic
t_j: jackknife *t*-test

4.2. Results and Discussion

4.2.1 Comparison of Type I error

In tables 1-8, under the assumption of equality of variances (*k=1*), the type I error rate of each statistic measured up to its nominal value α. For instance, the type I error rates using the permutation distribution of the Student *t*-statistic, studentized permutation test, and the type I error rates of the jackknife *t*-statistic, all yielded values between 0.0450 and 0.0510, which is close to their nominal significance levels of α= 0.05. Since this trend was observed for all combinations of n_1 and n_2 when k = 1, it seems reasonable to conclude that when testing the equality of the means of two identical distributions (which in this case was two normal distributions with equal variances), the permutation test using the Student *t*-test, the jackknife *t*-statistic, and studentized permutation test are all effective in controlling the type I error rate.

However, this trend does not stay the same when the variances were unequal (i.e., when *k*>1). For instance, in tables 2 and 5, the type I error rates obtained from the permutation distribution of Student *t*-statistic showed an increase in value as the ratio of

the variances increased. It can be concluded that when testing the equality of the means of two non-identical distributions with different sample sizes and variances, the permutation test using the Student *t*-test failed to control type I error rate. The error rates were inflated when the larger variance was associated with the smaller sample size.

In tables 3 and 4, we notice the exact opposite behavior. That is, the type I error rates decreased as the ratio of the variances increased. We conclude that when testing the equality of the means of two non-identical distributions, the permutation test using the Student *t*-statistic, produced deflated type I error rates when larger variance was associated with the larger sample size.

However, in every combination discussed so far, the jackknife *t*-statistic approximated by the chi-square distribution consistently had type I error rates close to its nominal value of $\alpha = 0.05$. Also, the studentized permutation test was robust under violation of equality of variance assumption and controls the type I error rate to its nominal level. In particular, if larger variance was associated with smaller sample size, the jackknife *t*-test did a better job of controlling type I error rates (tables 2 and 5). But when larger variance was associated with larger sample size, the studentized permutation test performed better in controlling type I error rate (tables 4 and 5).

We also wanted to see if same trends are observed for sample size is ≥ 30. Table 7 give the empirical type I error rates for $n_1 = 45$ and $n_2 = 30$, for increasing values of *k*. This was the case when the larger variance was associated with the larger sample size. It followed the same trend as Tables 2 and 4. The type I error rates obtained from the permutation test of the Student *t*-statistic decreased as the ratio of the variances systematically increased, while the type I error rate of the jackknife *t*–statistic and the studentized permutation test produced results close to the nominal error rate of 0.05.

When $n_1 = 30$ and $n_2 = 45$, in table 8, the type I error rates obtained via the permutation distribution of the Student *t*–statistic increased as the ratio of the variances systematically increased, whereas the type I error rate of the jackknife *t*–statistic and the studentized permutation test stayed close to its nominal value.

All three tests behave the same way when sample sizes are the same as is evident from tables 1 and 6.

In summary, three types of trends were consistently observed when testing the equality of means of two non – identical distributions, namely (1) the type I error rate increased for combinations where the larger variance was associated with the smaller sample size; (2) the type I error rate decreased for combinations where the larger variance was associated with the larger sample size; and (3) the proposed test using the jackknife estimator and the studentized permutation test did an excellent job in controlling the type I error rate.

Table 1: Type I error rates when $n_1 = 10$, $n_2 = 10$, and $k = 1,1.5, 2, 3$

n_1	n_2	*k*	*P_t*	*t*	*t_j*
10	10	1	.0410	.0410	.0390
10	10	1.5	.0480	.0480	.0443
10	10	2.0	.0530	.0530	.0476
10	10	3.0	.0623	.0623	.0520

Table 2: Type I error rates when $n_1 = 10$, $n_2 = 15$, and $k = 1, 1.2, 1.5, 2, 3$

n_1	n_2	k	*P_t*	*t*	*t_j*
10	15	1	.0380	.0400	.0386
10	15	1.2	.0593	.0643	.0596
10	15	1.5	.0523	.0690	.052
10	15	2.0	.0533	.0783	.0506
10	15	3.0	.072	.1150	.065

Table 3: Type I error rates when $n_1 = 15$, $n_2 = 10$, and $k = 1, 1.2, 1.5, 2, 3$

n_1	n_2	k	*P_t*	*t*	*t_j*
15	10	1	.0476	.0490	.0493
15	10	1.2	.0450	.0363	.0446
15	10	1.5	.0416	.0303	.0410
15	10	2.0	.0440	.0296	.0430
15	10	3.0	.0476	.0230	.0450

Table 4: Type I error rates when $n_1 = 9$, $n_2 = 6$, and $k = 1, 1.2, 1.5, 2, 3$

n_1	n_2	k	*P_t*	*t*	*t_j*
9	6	1	.0502	.0494	.0516
9	6	1.2	.0474	.0450	.0472
9	6	1.5	.0450	.0322	.0444
9	6	2.0	.0478	.0342	.0450
9	6	3.0	.0512	.0300	.0462

Table 5: Type I error rates when $n_1 = 6$, $n_2 = 9$, and $k = 1, 1.2, 1.5, 2, 3$

n_1	n_2	k	*P_t*	*t*	*t_j*
6	9	1	.0464	.0466	.0466
6	9	1.2	.0510	.0596	.0526
6	9	1.5	.0574	.0710	.0568
6	9	2.0	.070	.0986	.0634
6	9	3.0	.076	.1182	.0640

Table 6: Type I error rates when $n_1 = 30$, $n_2 = 30$, and $k = 1, 1.5, 2, 3$

n_1	n_2	k	*P_t*	*t*	*t_j*
30	30	1	.0450	.0450	.0446
30	30	1.5	.0537	.0537	.0527
30	30	2.0	.0473	.0473	.0463
30	30	3.0	.0586	.0586	.0556

Table 7: Type I error rates when $n_1 = 45$, $n_2 = 30$, and $k = 1, 1.2, 1.5, 2, 3$

n_1	n_2	k	P_t	t	t_j
45	30	1	.0452	.0464	.0458
45	30	1.2	.0496	.0422	.0498
45	30	1.5	.0540	.0378	.0544
45	30	2.0	.0534	.0312	.0524
45	30	3.0	.0484	.0220	.0498

Table 8: Type I error rates when $n_1 = 30$, $n_2 = 45$, and $k = 1, 1.2, 1.5, 2, 3$

n_1	n_2	k	P_t	t	t_j
30	45	1	.0516	.0502	.0510
30	45	1.2	.0546	.0622	.0554
30	45	1.5	.0516	.0718	.0516
30	45	2.0	.0428	.0756	.0428
30	45	3.0	.0500	.0968	.0480

4.2.2 *Comparison of the Power of tests*

Tables 9 and 10 give the power functions of permutation tests using the Student *t*-statistic, jackknife *t*-statistic, and studentized permutation test for different parameter combinations. It is obvious from the tables that the power of the proposed jackknife test is comparable to Janssen's permutation test for the combinations where larger variance is associated with larger sample size.

For the parameter combinations where larger variance is associated with small sample size, the power of the proposed test is less than the Janssen's permutation test. The power of all the tests increases as the difference between means increases.

Tables 11 and 12 give alpha adjusted powers of all three tests. The jackknife *t*-statistic and Janssen's permutation test have comparable power when larger variance was associated with large sample size. But Janssen's permutation test had more power when smaller sample size was associated with larger variance.

Also, the permutation test based on the Student *t*-statistic had highest power when larger variance was associated with smaller sample size. It had the least power when larger variance was associated with larger sample size.

Table 9: Power for combination where $n_1 = 9$, $n_2 = 6$

n_1	n_2	k	$\mu_1 - \mu_2$	P_t	t	t_j
9	6	2	1	.0830	.0583	.081
9	6	2	2	.2023	.1563	.1973
9	6	2	3	.4100	.3303	.4033
9	6	2	4	.6420	.5636	.6360
9	6	3	1	.0886	.0540	.0817
9	6	3	2	.2197	.1490	.2067
9	6	3	3	.3996	.2970	.3743
9	6	3	4	.6213	.4886	.5980

Table 10: Power for combination where $n_1 = 6$, $n_2 = 9$

n_1	n_2	k	$\mu_1 - \mu_2$	P_t	t	t_j
6	9	2	1	.0900	.1296	.0823
6	9	2	2	.2023	.2650	.1827
6	9	2	3	.3690	.4600	.3430
6	9	2	4	.5743	.6700	.5343
6	9	3	1	.100	.1570	.0803
6	9	3	2	.1986	.2910	.1663
6	9	3	3	.6043	.7243	.5367
6	9	3	4	.8243	.9063	.7623

Table 11: Alpha-adjusted power for combination where $n_1 = 9$, $n_2 = 6$

n_1	n_2	k	$\mu_1 - \mu_2$	P_t	t	t_j
9	6	2	1	.0816	.0966	.0873
9	6	2	2	.2003	.2220	.2087
9	6	2	3	.4066	.4370	.4160
9	6	2	4	.6390	.6660	.6510
9	6	3	1	.0877	.0877	.0867
9	6	3	2	.2173	.2110	.2180
9	6	3	3	.3966	.3880	.3923
9	6	3	4	.6183	.6083	.6173

Table 12: Alpha-adjusted power for combination where $n_1 = 6$, $n_2 = 9$

n_1	n_2	k	$\mu_1 - \mu_2$	P_t	t	t_j
6	9	2	1	.0683	.0700	.0677
6	9	2	2	.1603	.1670	.1557
6	9	2	3	.3010	.3150	.2916
6	9	2	4	.5056	.5256	.4773
6	9	3	1	.0713	.0723	.0687
6	9	3	2	.1506	.1576	.1400
6	9	3	3	.5256	.5530	.4743
6	9	3	4	.7586	.7860	.715

5. Discussion

The permutation test is an exact test limited to significance testing rather than estimation of effects. An exact permutation test based on the Student *t*-test can be advantageous in that no knowledge of the distribution of the observations is required. However, its control of the type I error rate may not hold under every condition of non-identical distribution among groups to be compared. For two normal distributions with unequal variances, using the *t*-test based permutation method for testing the equality of means can produce inflated or deflated type I error rates.

Under the violation of the assumption of equality of variances for two normal populations, the proposed test using the jackknife estimator, does an excellent job of controlling type I error rate to its nominal value. The proposed test is powerful in detecting deviation from the null hypothesis. Also, the studentized permutation test of Janssen controls type I error rate and is powerful under violation of the homogeneity of variance assumption.

The present study considered normal distributions for the groups to be compared. For small samples, the studentized permutation test can be a preferred method of testing for significant difference between means. For even a moderately large sample size, the studentized permutation test becomes more computer-intensive. The test has to be repeated for a large number of marker loci making it even more computer-intensive. For the jackknife *t*-test this is not an issue. We recommend using the jackknife *t*-test for large samples under normal distribution assumption.

Because separate statistical test is performed at each locus, we have to use appropriate simultaneous testing procedures to control for the experimentwise type I error rate. For the combination where larger variance is associated with smaller sample size, the jackknife *t*-test is a better choice as it controls type I error rate to its nominal value. For the combination where larger variance is associated with larger sample size, the studentized permutation test should be preferred.

6. Acknowledgements

The authors wish to thank the referee(s) for their valuable comments and suggestions in improving the manuscript.

References

[1] B. S. B. Abler, M. D. Edwards, and C. W. Stuber, *Iso-enzymatic identification of quantitative trait loci in crosses of elite maize inbreds*, Crop Science, 31, 267-274, 1991.

[2] W. Barendse, S. M. Armitage, L. M. Kossarek, A. Shalom, B. W. Kirkpatrick, A. M. Ryan, D. Clayton, L. Li, H. L. Neibergs, N. Zhang, W. M. Grosse, J. Weiss, P. Creighton, F. Mccarthy, M. Ron, A. J. Teale, R. Fries, R. A. Mcgraw, S. S. Moore, M. Georges, M. Soller, J. E. Womack, and D. J. S. Hetzel, *A genetic linkage map of the bovine genom,* Nature Genetics, 6, 227-235, 1994.

[3] A. Blumenfeld, S. A. Slaugenhaupt, F. B. Axelrod, D. E. Lucente, C. Maayan, C. B. Lieberg, L. J. Ozelius, J. A. Trofatter, J. L. Haines, X. O. Breakefield, J. F. Gusella, *Localization of the gene for familial dysautomia on chromosome 9 and definition of DNA markers for genetic diagnosi,* Nature Genetics, 4, 160-163, 1993.

[4] G. A. Churchill, and R. W. Doerge, *Empirical threshold values for quantitative trait mapping,* Genetics, 138, 963-971, 1994.

[5] G. A. Churchill, and R. W. Doerge, *Permutation tests for multiple loci affecting a quantitative character,* Genetics, 178, 609-610, 2007.

[6] G. A. Churchill, and R. W. Doerge, *Naïve application of permutation testing leads to inflated type I error rates,* Genetics, 142, 285-294, 2008.

[7] G. A. Churchill, and R. W. Doerge, *Statistical issues in the search for genes affecting quantitative traits in experimental populations,* Statistical Science, 12, 195-219, 1997.

[8] N. G. Copeland, N. A. Jenkins, D. J. Gilbert, J. T. Eppig, L. J. Maltais, J. C. Miller, W. F. Dietrich, S.E. Lincoln, R. G. Steen, L. D. Sein, J. H. Nadeau, and E. S. Lander, *A genetic linkage map of the mouse: current applications and future prospects,* Science, 262, 57-66, 1993.

[9] R. W. Doerge, Z-B. Zeng and B. S. Weir, *Statistical issues in the search for genes affecting quantitative traits in experimental populations,* Statistical Science, 12, 195-219, 1997.

[10] R. W. Doerge, *Mapping and analysis of quantitative trait loci in experimental populations,* Genetics, 3, 43-52, 2002.

[11] M. D. Edwards, C.W. Stuber, and J. F. Wendel, *Molecular-marker-facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action,* Genetics, 116, 113-125, 1987.

[12] R. C. Elston, and J. Stewart, *The analysis of quantitative traits for simple genetic models from parental, F_1 and backcross data,* Genetics, 73, 695-711, 1973.

[13] R. A. Fisher, *The design of experiments,* Ed. 3, Oliver & Boyd, London, 1935.

[14] A. Graner, A. Jahoor, J. Schondelmaier, H. Seilder, K. Pillen, G. Fishbeck, G. Wenzel, and

R. G. Herrmann, *Construction of an RFLP map of barley,* Theoretical and Applied Genetics, 83, 250-256, 1991.

[15] C. S. Haley, and S. A. Knott, *A simple regression method for mapping quantitative trait loci in line crosses using flanking marker,* Heredity, 69, 315-324, 1992.

[16] C. S. Haley, S. A. Knott and J-M. Elsen, *Mapping quantitative trait loci in crosses between outbred lines using least squar,* Genetics, 136, 1195-1207, 1994.

[17] Y. Huang, H. Xu, V. Calian, and J. C. Hsu, *To permute or not to permute,* Bioinformatics, 22, 2244-2248, 2006.

[18] The Huntington's Disease Collaborative Research Group, *A novel gene containing a trinulcleotide repeat that is expanded and unstable on Huntington's disease chromosome,* Cell, 72, 971-983, 1993.

[19] R. C. Jansen, and P. Stam, *High resolution of quantitative traits into multiple loci via interval mapping,* Genetics, 136, 1447-1455, 1994.

[20] Arnold Janssen, *Studentized permutation tests for non-i.i.d. hypotheses and the generalized Behrens-Fisher problem,* Statistics and Probability Letters, 36, 9-21, 1997.

[21] Arnold Janssen and Thorsten Pauls, *A Monte Carlo comparison of studentized bootstrap and permutation tests for heteroscedastic two-sample problems,* Computational Statistics, 20, 369-383, 2005.

[22] S. A. Knott, and C. S. Haley, *Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses,* Genet. Res, 60, 139-151, 1992.

[23] E. S. Lander, and D. Botstein, *Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps,* Genetics, 121, 185-199, 1989.

[24] E. S. Lander, and D. Botstein, *Corrigendum,* Genetics 36, 705, 1994.

[25] E. C. Lehman, *Testing statistical hypotheses,* Ed. 2. John Wiley & Sons, New York, 1986.

[26] B. H. Liu, *Statistical Genomics: Linkage, Mapping, and QTL analysis,* CRC Press LLC, Boca Raton, Florida. 1997.

[27] A. Manichaikul, J. Dupis., S. Saunak., and K. W. Browman, *Poor performance of bootstrap confidence intervals for the location of a quantitative trait locu,* Genetics, 174, 481-489, 2006.

[28] J. Nienhus, T. Helentjaris, M. Slocum, B. Ruggero, and A. Schaefer, *Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato,* Crop Science, 27, 797-803, 1987.

[29] A. H. Paterson, E. S. Lander, J. D. Hewitt, D. Samir, H. D. Rabinowitch, S. E. Lincoln, E. S. Lander, and S. D. Tanksley, *Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments,* Genetics, 127, 181-197, 1991

[30] Quenouille, M., *Approximation tests of correlation in time series,* J. R. Statist. Soc. B, 11: 18-84, 1949.

[31] Rebai, B. Goffinet, and B. Mangin, *Approximate thresholds if interval mapping tests for QTL detection,* Genetics, 138, 235-240, 1994.

[32] K. Sax, *The association of size difference with seed – coat pattern and pigmentation in Phaseolus vulgaris,* Genetics, 8, 552-560, 1923.

[33] P. Singh, K. K. Saxena, and O. P. Srivastava, *Power comparisons of solutions to the Behrens-Fisher Problem,* American Journal of Mathematical and Management Sciences. 22, 233-250, 2002.

[34] S. Sen, and G. A. Churchill, *A statistical framework for quantitative trait mapping,* Genetics, 159, 371-387, 2001.

[35] M. K. Slocum, S. S. Figdore, W. C. Kennard, J. Y. Suzuki, and T. C. Osborne, *Linkage arrangement of restriction fragment length polymorphism loci in Brassica oleracea,* Theoretical and Applied Genetics, 80, 3103-3106, 1990.

[36] C. W. Stuber, M. D. Edwards, and J. F. Wendel, *Molecular –marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component trait,.* Crop Science, 27, 639-648, 1987.

[37] R. M. Sundrum, *On the relation between estimating efficiency and the power of the tests,* Biometrika, 41, 542-544, 1964.

[38] S. D. Tanksley, M. W. Ganal, J. P. Prince, M. C. De Vicente, M. W. Bonierbale, P. Broun, T. M. Fulton, J. J. Giovannoni, S. Grandillo, G. B. Martin, R. Messeguer, J. C. Miller, L. Miller, A. H. Paterson, O. Pineda, M. S. Roder, R. A. Wing, W. Wu, and N. D. Young, *High density molecular linkage maps of tomato and potato genome,* Genetics, 132, 1141-1160, 1990.

[39] Thoday, J. M, *Location of polygenes,* Nature, 191, 368-370, 1961.

[40] J. Tukey, *Bias and confidence in not quite large samples,* Ann. Math. Statist., 29, 614, 1958

[41] J. I. Weller, *Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers,* Biometrics, 42, 627-640, 1986.

[42] Z. -B. Zeng, *Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci,* Proc. Natl. Acad. Sci. USA. 90, 10972-10976, 1993.

[43] Z. -B. Zeng, *Precision mapping of quantitative trait loci,* Genetics, 136, 1457-1468, 1994.