Hacettepe Journal of Mathematics and Statistics
Volume 46 (6) (2017), 1199 – 1210

Autocorrelation corrected standard error for two sample t-test under serial dependence

Ayfer Ezgi YILMAZ[*†] and Serpil AKTAS[‡]

Abstract

The classical two-sample t-test assumes that observations are independent. A violation of this assumption could lead to inaccurate results and incorrectly analyzing data leads to erroneous statistical inferences. However, in real life applications, data are often recorded over time and serial correlation is unavoidable. In this study, two new autocorrelation corrected standard errors are proposed for independent and correlated samples. These standard errors are replaced by the classical standard error in the presence of serially correlated samples in two samples t-test. Results based upon the simulation show that the proposed standard errors gives higher empirical power than other approaches.

Keywords: Hypothesis testing, Two sample tests, t-test, Serial dependence, Autocorrelation.

2000 AMS Classification: 62F03

Received : 16.02.2015 *Accepted* : 05.02.2016 *Doi* : 10.15672/HJMS.201611515847

1. Introduction

Two-sample hypothesis testing is a classical statistical analysis designed in order to test whether there is a difference between two means drawn from two different populations.

Let $X_1 = (X_{1,1}, X_{1,2}, ..., X_{1,n_1})'$ and $X_2 = (X_{2,1}, X_{2,2}, ..., X_{2,n_2})'$ be random samples from two populations at consecutive time points $1, 2, ..., n_1$ and $1, 2, ..., n_2$, respectively. Let μ_1 and μ_2 be the means of these population. Then the hypothesis can be written as,

$$\begin{aligned} &H_0 : \mu_1 = \mu_2 \\ &H_1 : \mu_1 \neq \mu_2 \end{aligned} \tag{1.1}$$

The classical two-sample t-test assumes that the observations are independent. A violation of this assumption could lead to inaccurate results and incorrect conclusions. However, in some studies, recording data over time leads to the serial correlation. In

[*]Department of Statistics,Hacettepe University, Ankara–Turkey, Email: ezgiyilmaz@hacettepe.edu.tr
[†]Corresponding Author.
[‡]Department of Statistics,Hacettepe University, Ankara–Turkey, Email: spxl@hacettepe.edu.tr

such cases, the classical variance estimators are generally found smaller than the actual variance and that affects the absolute value of the observed t-test statistic. Several methods have been proposed in the literature for estimating standard error of the difference between the means for two autocorrelated data. Those are Wilks [7], Box-Hunter [1], Seitshiro [5], and Zimmerman [8] approaches.
In this study instead of using classical methods, alternative methods for the different variance estimators have been discussed via a simulation study. In section 2, Student's t-test which is one of the most frequently used test in statistics is introduced. The approaches which have been proposed in the literature and new approaches that are used to compare two autocorrelated means are introduced in Section 3. These approaches are illustrated by a numerical example in Section 4 and the simulation study results are discussed in Section 5.

2. Student's t-test

One of the most popular approach for equality of population means is Student's t-test. This approach requires the observations in both samples are independent and normally distributed [3].

Let $X_1 \sim (\mu_1, \sigma_1^2)$ and $X_2 \sim (\mu_2, \sigma_2^2)$ be normal distributed random variables, then the t-test statistic is defined as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \tag{2.1}$$

where the sample means are $\bar{X}_i = \sum_{j=1}^{n_i} X_{i,j}/n_i$ and the sample variances are $s_i^2 = \sum_{j=1}^{n_i} (X_{i,j} - \bar{X}_i)/(n_i - 1)$ for $i = 1, 2$. Under H_0, t follows approximately a t distribution with υ degrees of freedom. Under the assumption of unequal variances ($\sigma_1^2 \neq \sigma_2^2$), the υ is calculated as follows:

$$\upsilon = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}. \tag{2.2}$$

Under the assumption of equal variances ($\sigma_1^2 = \sigma_2^2$), t has a t-distribution with $\upsilon = n_1 + n_2 - 2$ degrees of freedom. t and the pooled variances can be calculated as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p\sqrt{1/n_1 + 1/n_2}}, \tag{2.3}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}. \tag{2.4}$$

Although Student's t-test is one of the most commonly used method for testing a hypothesis on the basis of a difference between sample means, this method is not proper for the autocorrelated data. In order to analyze the difference between two sample means, another approaches have been suggested for the autocorrelated data.

3. Two sample comparison of two autocorrelated means

The general form of t-test is

$$(3.1)\qquad t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{Var(\bar{X}_1 - \bar{X}_2)}}.$$

Several methods have been proposed in the literature for estimating standard error of the difference between means for two autocorrelated data. The Wilks,vBox-Hunter, and Seitshiro approaches are presented in the following sub-sections.

3.1. Wilks approach. Wilks approach estimates the standard error of the sampling distribution of the mean based on variance inflation factor is defined as follows:

$$(3.2)\qquad SE = \sqrt{V\frac{s_x^2}{n}}.$$

This approach is successful when the sample size n is sufficiently large. In the Equation (3.2), s_x^2 is the sample variance and V is the variance inflation factor which depends on the autocorrelation in the data. V can be calculated as:

$$(3.3)\qquad V = 1 + 2\sum_{k=1}^{n-1}\{1 - \frac{k}{n}\}r_k,$$

where r_k values are estimates of the autocorrelations at lags k [7].
In order to obtain more stable estimates for V, the time series model can be useful [4, 6]. When assuming an AR(1) model for the data, only the lag-1 autocorrelation needs to be directly estimated from the data [7],

$$(3.4)\qquad r_1 = \frac{\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}.$$

Because the estimates of V are substantially biased for samples that are not large, instead of using V, the adjusted variance inflation factor given in Equation (3.5) is suggested to use.

$$(3.5)\qquad V^{'} = Vexp\{\frac{2V}{n}\}.$$

Then, the standard error (SE) that is suggested by Wilks is [7],

$$(3.6)\qquad SE_W = \sqrt{V_1^{'}\frac{s_{x_1}^2}{n_1} + V_2^{'}\frac{s_{x_2}^2}{n_2}}.$$

The general form of t statistic to test whether the means are different can be calculated as follows:

$$(3.7)\qquad t = \frac{\bar{X}_1 - \bar{X}_2}{SE}.$$

3.2. Box-Hunter approach. Box *et al.* [1] presented a numerical example to discuss the serial dependence in the industrial data. In this study, two different methods are applied to data during the ongoing process [5]. The standard error is defined by taking the autocorrelation into consideration as,

$$SE_{BH} = \sqrt{\frac{2s^2}{n}[1 + \frac{2n-3}{n}r_1]}. \tag{3.8}$$

The t statistic can be calculated from Equation (3.7). Here the sample sizes are $n_1 = n_2 = n$. The sample mean is $\bar{X} = \sum_{i=1}^{2}\sum_{j=1}^{n} X_{i,j}/2n$ and the sample variance is $s^2 = \sum_{i=1}^{2}\sum_{j=1}^{n}(X_{i,j} - \bar{X})/(2n-1)$.

3.3. Seitshiro approach. Seitshiro approach is proposed based on the paired samples t-test. The hypothesis of no difference between the series X_1 and X_2 are formulated in terms of the differences, given by:

$$\begin{aligned} H_0 &: \mu_D = 0 \\ H_1 &: \mu_D \neq 0 \end{aligned} \tag{3.9}$$

The test statistic that tests this hypothesis is [5]

$$t_{dep} = \frac{\bar{D}}{\sqrt{\hat{\sigma}^2(\bar{D})}} \tag{3.10}$$

where $\hat{\sigma}^2(\bar{D})$ denotes the estimated variance of $\bar{D}$ and $\bar{D} = \bar{X}_1 - \bar{X}_2$.
The estimator for $\hat{\sigma}^2(\bar{D})$ is

$$\hat{\phi}_D = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{\sum_{i=1}^{n}(D_i - \bar{D})(D_{i+1} - \bar{D})/n}{\sum_{i=1}^{n}(D_i - \bar{D})^2/n} \tag{3.11}$$

$$\hat{\sigma}^2(\bar{D}) = \frac{\hat{\gamma}_0(1 + \hat{\phi}_D)}{n(1 - \hat{\phi}_D)}. \tag{3.12}$$

3.4. The proposed approaches. There are some disadvantages of Box-Hunter approach. Although Box-Hunter approach is useful for serially dependent data, this approach ignores that there are two groups and an overall variance is calculated instead of two different variances. The restriction of Box-Hunter's approach is that, the sample sizes of two groups should be equal. The approach also ignores the effects of sample autocorrelation and an overall value is calculated. Because of these disadvantages, Box-Hunter approach is extended to the approaches that allow the unequal sample sizes for independence and correlated samples. The effects of sample variances and autocorrelation are also considered.

The standard error of the difference for independent samples is

$$(3.13)\quad SE=\sqrt{\left|\frac{s_1^2}{2n_1}(1+\frac{2n_1-3}{n_1}r_1^X)+\frac{s_2^2}{2n_2}(1+\frac{2n_2-3}{n_2}r_1^Y)\right|}.$$

The standard error of the difference for correlated samples is

$$(3.14)\quad SE=\sqrt{\left|\frac{s_1^2}{2n_1}(1+\frac{2n_1-3}{n_1}r_1^X)+\frac{s_2^2}{2n_2}(1+\frac{2n_2-3}{n_2}r_1^Y)+2cov(X_1,X_2)\right|}.$$

These approaches will be illustrated on a numerical example. Then, they will be compared through the simulation study.

4. Numerical example

The data set given in Table 1 concerns the assessments of a modification in a manufacturing plant [1]. When the process continues, A method is applied to the first 10 observation, then B method is applied to the others.

Table 1. Yield data from an industrial experiment (plant trial)

Time Order	**Method**	**Yield**	**Time Order**	**Method**	**Yield**
1	A	89.7	11	B	84.7
2	A	81.4	12	B	86.1
3	A	84.5	13	B	83.2
4	A	84.8	14	B	91.9
5	A	87.3	15	B	86.3
6	A	79.7	16	B	79.3
7	A	85.1	17	B	82.6
8	A	81.7	18	B	89.1
9	A	83.7	19	B	83.7
10	A	84.5	20	B	88.5

The descriptive statistics of A and B methods are given in Table 2. Table 3 shows t-test results and standard errors of difference for the approaches that are given in Section 2.

Table 2. Descriptive statistics of yield data

Method	**A**	**B**	**A-B**
Mean	84.24	85.54	-1.30
St.D.	2.90	3.65	1.27
St.E.	0.92	1.15	1.47
r_1	-0.44	-0.17	-

After analyzing the data by five different methods, it can be seen that, independent samples t-test has the largest standard error. This is due to the serially dependence structure of the variables. Wilks approach has the smallest standard error. Seitshiro approach and proposed approach have similar results with Wilks'. The results in Table 3 show that, the hypotheses are not rejected for all approaches. Hence, there is not a statistically significance difference between the A and B methods.

Table 3. t-test results of yield data

Method	**t-value**	v	**P-value**	**St.E.**
Student t (Equal variances)	-0.882	18	0.390	1.474
Wilks	-1.962	18	0.065	**0.663**
Box-Hunter	-1.113	18	0.280	1.167
Seitshiro	-1.927	18	0.070	0.675
Proposed 1	-1.892	18	0.075	0.687

5. Simulation study

In this section, we performed a simulation study to compare the performance of five approaches with respect to their power values and Type I error probabilities. One of the time series models is the first-order autoregressive (AR(1)) process, defined as,

$$X_t = X_{t-1}r_1 + \varepsilon_t \tag{5.1}$$

where ε_t are independent and generated from normal distribution [7, 2]. In this study, two random AR[1] processes are generated. For the simplicity and also to compare the results of Box-Hunter approach, the sample sizes are assumed equal ($n_1 = n_2$) and considered as 10, 20, 30 and 50. All the results are based on 10000 replication of each sample. The r_1 values are taken as:
-0.9, -0.5, -0.3, 0.3, 0.5, 0.9

In the study, to generate a hypothesis test, it is assumed that the two samples came from different populations with $X \sim N(50, 10^2)$ and $Y \sim N(30, 10^2)$ for equal variances, and $X \sim N(50, 5^2)$ and $Y \sim N(30, 15^2)$ for unequal variances. In the second step, it is assumed that the two samples came from the same populations with $X \sim N(50, 10^2)$ and $Y \sim N(48, 10^2)$ for equal variances, and $X \sim N(50, 5^2)$ and $Y \sim N(48, 15^2)$ for unequal variances.

After setting the simulation parameters, five methods are applied to random samples and the null hypothesis of no difference is tested at the level of $\alpha = 0.05$. Table 4 shows the empirical power of the t-tests under equal and unequal variances for different sample sizes and the different values of autocorrelation. The values of autocorrelation for X and Y samplings are accepted as equal. Table 4 shows that, the powers of proposed methods are the highest in many cases. For instance for $r_1^X = r_1^Y = 0.9$; $n = 20$ and unequal variances and for $r_1^X = r_1^Y = 0.9$; $n = 20$ and unequal variances, power is the highest for the proposed method for independent samples. Proposed 1 and Proposed 2 methods give the highest power for $r_1^X = r_1^Y = 0.3$; $n = 10$ and equal variances.

Table 4. The empirical power of t-tests under equal and unequal variances for different sample sizes and autocorrelations

		Equal Variances				Unequal Variances			
		n				n			
$r_1^X = r_1^Y$	**Method**	**10**	**20**	**30**	**50**	**10**	**20**	**30**	**50**
-0.9	Student t	0.000	0.000	0.006	0.274	0.000	0.000	0.005	0.224
	Wilks	0.071	0.977	0.994	0.999	0.050	0.970	0.987	0.991
	Box-Hunter	0.091	0.076	0.241	0.703	0.010	0.079	0.228	0.615
	Seitshiro	0.911	0.998	1.000	1.000	0.853	0.994	1.000	1.000
	Proposed 1	0.077	0.247	0.602	0.968	0.006	0.149	0.505	0.930
	Proposed 2	0.002	0.014	0.083	0.527	0.000	0.006	0.050	0.480
-0.5	Student t	0.001	0.640	0.993	1.000	0.000	0.552	0.973	1.000
	Wilks	0.581	0.999	1.000	1.000	0.568	0.997	1.000	1.000
	Box-Hunter	0.356	0.996	1.000	1.000	0.325	0.990	1.000	1.000
	Seitshiro	0.956	1.000	1.000	1.000	0.914	0.999	1.000	1.000
	Proposed 1	0.871	1.000	1.000	1.000	0.924	1.000	1.000	1.000
	Proposed 2	0.103	0.929	1.000	1.000	0.054	0.938	1.000	1.000
-0.3	Student t	0.058	0.960	1.000	1.000	0.044	0.923	0.999	1.000
	Wilks	0.709	1.000	1.000	1.000	0.693	0.999	1.000	1.000
	Box-Hunter	0.063	0.990	1.000	1.000	0.072	0.975	1.000	1.000
	Seitshiro	0.954	1.000	1.000	1.000	0.916	0.999	1.000	1.000
	Proposed 1	0.854	1.000	1.000	1.000	0.858	1.000	1.000	1.000
	Proposed 2	0.568	0.999	1.000	1.000	0.587	0.999	1.000	1.000
0.3	Student t	0.973	1.000	1.000	1.000	0.961	1.000	1.000	1.000
	Wilks	0.968	1.000	1.000	1.000	0.952	1.000	1.000	1.000
	Box-Hunter	0.489	1.000	1.000	1.000	0.454	0.997	1.000	1.000
	Seitshiro	0.940	0.999	1.000	1.000	0.910	0.997	1.000	1.000
	Proposed 1	0.990	1.000	1.000	1.000	0.985	1.000	1.000	1.000
	Proposed 2	0.997	1.000	1.000	1.000	0.991	1.000	1.000	1.000
0.5	Student t	0.996	1.000	1.000	1.000	0.990	1.000	1.000	1.000
	Wilks	0.995	1.000	1.000	1.000	0.985	1.000	1.000	1.000
	Box-Hunter	0.877	1.000	1.000	1.000	0.814	0.999	1.000	1.000
	Seitshiro	0.918	0.998	1.000	1.000	0.892	0.994	1.000	1.000
	Proposed 1	0.998	1.000	1.000	1.000	0.992	1.000	1.000	1.000
	Proposed 2	0.997	1.000	1.000	1.000	0.990	1.000	1.000	1.000
0.9	Student t	0.532	0.997	1.000	1.000	0.528	0.986	1.000	1.000
	Wilks	0.007	0.137	0.306	0.770	0.029	0.179	0.324	0.740
	Box-Hunter	0.000	0.651	0.996	1.000	0.001	0.611	0.982	1.000
	Seitshiro	0.616	0.910	0.981	1.000	0.608	0.882	0.973	0.999
	Proposed 1	0.472	0.991	1.000	1.000	0.477	0.971	1.000	1.000
	Proposed 2	0.985	1.000	1.000	1.000	0.972	0.999	1.000	1.000

Table 5 and Table 6 show the means of t-values and their standard deviations under equal and unequal variances for different sample sizes and the different values of autocorrelation, respectively. The mean and standard deviations of the approaches with the mean and standard deviation of theoretical t distribution can be compared by means of Table 5 and Table 6. The deviations from the expected value and variance of t distribution occur in negative autocorrelated variables. The results are similar when the variances are not equal.

Table 7 shows the empirical power, the means of t-values, the standard deviations of t-values, and means of standard errors for t-tests under the different variances for different sample sizes. Here the sample sizes are $n_1 = n_2 = 50$ and sample autocorrelations are $r_1 = r_2 = 0.5$. Table 8 shows the means and standard deviations of t values, and standard errors for t-test under equal variances for different sample sizes. The values of autocorrelation for X and Y samplings are assumed as unequal. Here the sample sizes are $n_1 = n_2 = 50$.

Table 5. The empirical distributions of t-values under equal variances for different sample size and autocorrelations

	n	10		20		30		50	
r	Method	t-value	St.D.	t-value	St.D.	t-value	St.D	t-value	St.D.
-0.9	Student t	0.486	0.154	0.882	0.217	1.231	0.269	1.799	0.332
	Wilks	1.445	0.436	3.450	0.780	6.097	9.323	8.149	10.798
	Box-Hunter	0.835	0.748	1.300	0.843	1.734	1.037	2.428	1.086
	Seitshiro	3.802	1.375	5.711	1.449	7.187	1.474	9.542	1.496
	Proposed1	1.261	1.236	1.795	1.307	2.323	1.108	3.158	0.897
	Proposed2	0.498	0.288	0.943	0.403	1.393	0.854	2.199	1.025
-0.5	Student t	1.129	0.290	2.167	0.386	3.038	0.452	4.459	0.531
	Wilks	2.253	0.624	4.875	1.167	7.362	3.832	12.136	16.351
	Box-Hunter	2.259	4.173	3.575	2.879	4.369	2.081	5.659	1.031
	Seitshiro	4.254	1.482	6.203	1.598	7.655	1.626	9.952	1.628
	Proposed1	4.890	5.849	10.300	13.441	15.502	19.506	24.384	45.781
	Proposed2	1.491	0.653	3.181	1.555	4.974	3.233	8.947	7.775
-0.3	Student t	1.501	0.373	2.862	0.490	3.974	0.572	5.804	0.661
	Wilks	2.511	0.710	5.372	1.367	7.833	3.099	12.251	12.704
	Box-Hunter	1.669	0.328	2.765	0.303	3.563	0.300	4.813	0.282
	Seitshiro	4.320	1.511	6.287	1.687	7.695	1.735	9.987	1.744
	Proposed1	3.956	6.537	8.194	22.702	11.000	11.888	15.284	12.320
	Proposed2	2.635	2.528	6.321	9.063	10.866	16.969	21.660	28.082
0.3	Student t	3.980	0.958	6.858	1.205	9.053	1.340	12.395	1.490
	Wilks	4.590	1.617	8.547	2.902	11.728	4.855	17.406	70.774
	Box-Hunter	2.050	0.234	3.054	0.201	3.810	0.189	4.983	0.185
	Seitshiro	4.255	1.562	6.304	1.976	7.794	2.108	10.081	2.259
	Proposed1	6.060	5.417	8.738	2.624	11.021	2.432	14.535	2.415
	Proposed2	9.049	14.491	10.835	7.980	12.515	6.625	15.493	3.329
0.5	Student t	5.815	1.320	9.535	1.717	12.259	1.890	16.438	2.085
	Wilks	6.253	2.351	10.485	3.891	13.578	5.887	18.974	54.725
	Box-Hunter	2.289	0.183	3.269	0.160	4.017	0.156	5.200	0.155
	Seitshiro	3.969	1.483	6.067	2.024	7.608	2.264	10.007	2.536
	Proposed1	7.278	2.952	10.681	2.622	13.212	2.682	17.193	2.784
	Proposed2	5.983	2.138	9.665	2.530	12.335	2.580	16.533	2.886
0.9	Student t	2.145	0.566	3.903	0.696	5.618	0.804	9.052	1.011
	Wilks	1.259	0.336	1.684	0.310	1.867	0.290	2.225	0.319
	Box-Hunter	1.363	0.280	2.107	0.263	2.793	0.255	4.024	0.241
	Seitshiro	2.389	0.725	3.078	0.907	3.801	1.118	5.334	1.553
	Proposed1	2.063	0.548	3.476	0.625	4.892	0.708	7.766	0.881
	Proposed2	4.000	0.825	5.705	0.883	7.721	1.009	11.881	1.283

Table 6. The empirical distributions of t-values under unequal variances for different sample size and autocorrelations

	n	10		20		30		50	
r	Method	t-value	St.D.	t-value	St.D.	t-value	St.D.	t-value	St.D.
-0.9	Student t	0.476	0.149	0.865	0.230	1.188	0.286	1.715	0.370
	Wilks	1.412	0.412	3.364	0.790	5.984	11.605	7.432	9.907
	Box-Hunter	0.783	0.416	1.283	0.824	1.684	0.815	2.342	0.925
	Seitshiro	3.457	1.360	5.144	1.398	6.436	1.415	8.531	1.414
	Proposed1	1.007	0.357	1.561	0.460	2.059	0.539	2.887	0.681
	Proposed2	0.470	0.153	0.905	0.445	1.312	0.447	2.049	0.714
-0.5	Student t	1.103	0.300	2.086	0.409	4.169	0.564	4.169	0.564
	Wilks	2.217	0.626	4.818	1.228	12.416	15.886	12.416	15.886
	Box-Hunter	2.191	2.176	3.701	4.404	5.825	2.206	5.825	2.206
	Seitshiro	3.890	1.459	5.618	1.527	8.909	1.514	8.909	1.514
	Proposed1	5.654	20.322	12.284	33.876	26.241	32.758	26.241	32.758
	Proposed2	1.406	0.417	3.023	0.960	8.284	7.149	8.284	7.149
-0.3	Student t	1.460	0.379	2.739	0.507	3.771	0.599	5.404	0.706
	Wilks	2.486	0.729	5.300	1.461	7.912	4.438	13.171	18.575
	Box-Hunter	1.654	0.364	2.728	0.454	3.509	0.380	4.732	0.371
	Seitshiro	3.991	1.530	5.693	1.596	6.971	1.637	8.997	1.632
	Proposed1	3.199	1.377	6.526	4.515	9.070	5.005	13.052	7.077
	Proposed2	2.373	1.061	6.113	5.559	11.551	19.929	24.147	61.614
0.3	Student t	3.792	0.928	6.402	1.247	8.341	1.409	11.322	1.564
	Wilks	4.744	1.889	8.702	3.410	12.434	14.482	17.781	31.394
	Box-Hunter	2.026	0.261	2.992	0.248	3.720	0.245	4.862	0.245
	Seitshiro	4.060	1.590	5.831	1.938	7.146	2.054	9.144	2.160
	Proposed1	5.426	4.324	8.017	2.272	10.043	2.206	13.230	2.248
	Proposed2	7.463	6.854	9.282	4.866	10.917	3.206	13.847	2.666
0.5	Student t	5.496	1.461	8.870	1.934	11.272	2.105	14.932	2.268
	Wilks	6.117	2.384	10.183	3.997	13.296	8.213	18.640	22.671
	Box-Hunter	2.250	0.234	3.207	0.220	3.937	0.213	5.090	0.214
	Seitshiro	3.856	1.514	5.740	2.008	7.058	2.212	9.112	2.387
	Proposed1	6.305	1.890	9.608	2.337	11.907	2.446	15.501	2.610
	Proposed2	5.318	1.375	8.740	1.962	11.179	2.204	14.935	2.495
0.9	Student t	2.164	0.704	3.893	0.875	5.588	0.988	8.993	1.240
	Wilks	1.272	0.414	1.681	0.379	1.857	0.333	2.204	0.335
	Box-Hunter	1.358	0.340	2.091	0.331	2.773	0.316	4.000	0.299
	Seitshiro	2.407	0.803	3.108	1.022	3.822	1.254	5.340	1.656
	Proposed1	2.077	0.676	3.463	0.780	4.861	0.861	7.708	1.066
	Proposed2	3.883	0.848	5.589	0.963	7.570	1.065	11.658	1.299

Table 7. The empirical power and results of t-tests for different values of autocorrelations ($\alpha = 0.05$)

X	Y	**Method**	**E.Power**	**t-value**	**St.D**	**St.E**
$N(50, 5^2)$	$N(48, 10^2)$	Student t	0.512	2.010	1.739	1.830
		Wilks	0.460	2.301	4.153	2.115
		Box-Hunter	0.310	1.357	1.124	2.622
		Seitshiro	0.277	1.338	1.231	2.893
		Proposed1	0.526	2.084	1.812	1.772
		Proposed2	0.549	2.173	1.902	1.702
$N(50, 15^2)$	$N(48, 10^2)$	Student t	0.369	1.243	1.735	2.882
		Wilks	0.353	1.490	3.470	3.238
		Box-Hunter	0.184	0.852	1.175	4.061
		Seitshiro	0.157	0.812	1.179	4.654
		Proposed1	0.387	1.302	1.837	2.779
		Proposed2	0.398	1.341	1.906	2.727
$N(50, 25^2)$	$N(48, 10^2)$	Student t	0.319	0.855	1.790	4.256
		Wilks	0.336	1.123	3.512	4.666
		Box-Hunter	0.153	0.591	1.231	5.959
		Seitshiro	0.128	0.557	1.206	6.944
		Proposed1	0.337	0.900	1.902	4.103
		Proposed2	0.344	0.915	1.944	4.066
$N(50, 35^2)$	$N(48, 10^2)$	Student t	0.288	0.618	1.776	5.724
		Wilks	0.317	0.919	5.053	6.225
		Box-Hunter	0.132	0.428	1.231	7.983
		Seitshiro	0.105	0.399	1.185	9.356
		Proposed1	0.308	0.650	1.887	5.516
		Proposed2	0.312	0.658	1.914	5.491

Table 8. The results of t-tests under $X \sim N(50, 10^2)$ and $Y \sim N(30, 10^2)$ where $n_X = n_Y = 50$ for different autocorrelations

r_1^X	r_1^Y	**Method**	**t-value**	**St.D**	**St.E**
-0.3	-0.5	Student t	6.544	0.635	2.795
		Wilks	15.997	14.781	1.389
		Box-Hunter	5.362	0.269	3.405
		Seitshiro	13.000	2.000	1.427
		Proposed1	17.821	64.163	1.237
		Proposed2	10.099	2.728	1.854
-0.3	-0.9	Student t	4.938	0.894	4.715
		Wilks	19.322	27.782	1.698
		Box-Hunter	12.558	20.380	2.441
		Seitshiro	18.330	2.312	1.248
		Proposed1	13.163	15.390	2.091
		Proposed2	19.439	23.691	1.673
0.3	-0.3	Student t	19.538	1.460	2.448
		Wilks	35.284	42.452	1.723
		Box-Hunter	5.577	0.053	8.544
		Seitshiro	23.062	3.663	2.116
		Proposed1	22.904	3.052	2.1113
		Proposed2	24.417	9.243	2.010
0.3	-0.5	Student t	19.247	1.510	2.652
		Wilks	40.260	54.038	1.648
		Box-Hunter	5.650	0.050	8.995
		Seitshiro	26.834	3.906	1.932
		Proposed1	22.576	3.067	2.286
		Proposed2	23.826	5.118	2.191
0.3	-0.9	Student t	12.340	2.253	4.620
		Wilks	41.225	50.525	1.904
		Box-Hunter	6.192	0.411	8.940
		Seitshiro	34.897	4.198	1.602
		Proposed1	14.451	3.040	3.983
		Proposed2	14.736	3.392	3.930
0.3	0.5	Student t	4.846	1.564	2.219
		Wilks	6.183	6.938	2.332
		Box-Hunter	3.028	0.744	3.485
		Seitshiro	3.505	1.378	3.219
		Proposed1	5.689	1.970	1.914
		Proposed2	5.993	2.257	1.847
0.3	0.9	Student t	-22.383	2.210	8.362
		Wilks	-5.940	1.425	32.796
		Box-Hunter	-5.331	0.089	34.795
		Seitshiro	-6.373	1.351	30.394
		Proposed1	-26.177	3.630	7.212
		Proposed2	-25.840	3.672	7.313

The type I errors under equal and unequal variances for different sample sizes and different autocorrelation levels for $\alpha = 0.05$ are summarized in Table 9. The probabilities below 0.05 means that the null hypothesis is rejected. The deviation from nominal alpha is the highest for the Student t-test. Proposed 1 approach for $r_1^X = r_1^Y = 0.9$; $n = 50$ and unequal variances gives the most reasonable results. Box Hunter approach for $r_1^X = r_1^Y = -0.3$; $n = 20$ and unequal variances gives the perfect fit associated with the actual nominal alpha value.

6. Conclusions

In order to compare two autocorrelated data, the classical two-sample t-test cannot be used. Because its assumption is the independence of observations, these test cannot be used. In this study, suggested autocorrelation corrected standard errors for independent and correlated samples were introduced. The introduced methods were applied on plant trial data set and compared via a simulation study.

The results show that, the empirical power is higher when the variances are equal for all the combinations of autocorrelation. When the sample size increases, the empirical

Table 9. The type I errors for different sample sizes and autocorrelations

		Equal Variances				**Unequal Variances**			
		n				**n**			
$r_1^X = r_1^Y$	**Method**	**10**	**20**	**30**	**50**	**10**	**20**	**30**	**50**
-0.9	Student t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Wilks	0.000	0.007	0.075	0.100	0.000	0.009	0.073	0.100
	Box-Hunter	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Seitshiro	0.071	0.080	0.103	0.147	0.069	0.077	0.090	0.132
	Proposed1	0.007	0.002	0.001	0.001	0.000	0.000	0.000	0.000
	Proposed2	0.000	0.000	0.002	0.002	0.000	0.001	0.000	0.000
-0.5	Student t	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001
	Wilks	0.002	0.051	0.132	0.229	0.005	0.061	0.144	0.246
	Box-Hunter	0.064	0.155	0.238	0.380	0.068	0.158	0.221	0.343
	Seitshiro	0.093	0.097	0.120	0.163	0.095	0.094	0.111	0.143
	Proposed1	0.115	0.234	0.332	0.481	0.172	0.335	0.416	0.537
	Proposed2	0.001	0.015	0.042	0.147	0.000	0.008	0.036	0.144
-0.3	Student t	0.000	0.000	0.002	0.013	0.000	0.000	0.004	0.012
	Wilks	0.006	0.081	0.154	0.252	0.013	0.094	0.177	0.267
	Box-Hunter	0.013	0.038	0.062	0.132	0.016	0.050	0.082	0.141
	Seitshiro	0.104	0.107	0.120	0.165	0.107	0.103	0.114	0.145
	Proposed1	0.068	0.163	0.233	0.344	0.044	0.145	0.222	0.317
	Proposed2	0.026	0.133	0.256	0.462	0.017	0.155	0.320	0.527
0.3	Student t	0.119	0.182	0.227	0.304	0.125	0.177	0.209	0.280
	Wilks	0.257	0.333	0.372	0.392	0.278	0.347	0.373	0.393
	Box-Hunter	0.018	0.068	0.107	0.164	0.027	0.077	0.106	0.156
	Seitshiro	0.178	0.152	0.155	0.188	0.172	0.145	0.143	0.172
	Proposed1	0.270	0.290	0.319	0.384	0.257	0.274	0.295	0.358
	Proposed2	0.356	0.333	0.348	0.398	0.334	0.309	0.314	0.368
0.5	Student t	0.199	0.289	0.345	0.431	0.227	0.286	0.330	0.402
	Wilks	0.265	0.339	0.373	0.394	0.288	0.342	0.378	0.392
	Box-Hunter	0.030	0.110	0.160	0.242	0.039	0.120	0.155	0.221
	Seitshiro	0.222	0.185	0.187	0.215	0.230	0.178	0.162	0.187
	Proposed1	0.296	0.337	0.376	0.448	0.290	0.321	0.357	0.419
	Proposed2	0.410	0.390	0.398	0.463	0.368	0.354	0.374	0.427
0.9	Student t	0.000	0.002	0.007	0.056	0.001	0.006	0.021	0.085
	Wilks	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Box-Hunter	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
	Seitshiro	0.392	0.377	0.363	0.367	0.399	0.380	0.365	0.337
	Proposed1	0.000	0.000	0.002	0.024	0.000	0.003	0.009	0.044
	Proposed2	0.167	0.086	0.125	0.257	0.188	0.117	0.163	0.279

power also increases. Student's t-test does not have sufficient results when the autocorrelation is negative and the sample size is small. When the sample size increases or the autocorrelation is positive, empirical power increases.

If there is a negative and high autocorrelation, Seitshiro approach has the highest empirical power. In the case that the autocorrelation is $r_1 = r_2 = -0.5$ and -0.3, Seitshiro approach for $n = 10$, proposed approaches for $n \geqslant 20$ have the highest empirical powers. In the case that the autocorrelation is positive but not at high levels, proposed approaches have the highest empirical powers. If there is a positive and high autocorrelation, proposed approach for correlated samples gives better results. When $n \geqslant 20$ and the level of autocorrelation is low or moderate, the empirical powers of t-tests results are similar. In general, except presence of negative autocorrelations for $n = 10$ and $r_1 = r_2 = -0.9$, the proposed approaches have the highest empirical power.

The proposed approaches are extended from the Box-Hunter approach. The proposed approaches have higher empirical power than the Box-Hunter approach for all cases. Whether the variances of two groups are equal or unequal and for all values of autocorrelation.

When the values of autocorrelation are unequal and one of them is negative, Wilks and Seithiro approaches; when both of them are positive, the proposed approaches; when both of them are negative and $r_1 = 0.3, r_2 = -0.3$, the proposed approaches; and; when both of them are negative, the proposed and Seitshiro approaches have the lowest mean of standard errors. When the difference of the two sample variances increases, the empirical power of test decreases and the mean of standard errors increases.

References

[1] Box, G.E.P, Hunter, W.G and Hunter, J.S. *Statistics for experimenters: An introduction to design, data analysis, and model building* (John Wiley and Sons, 1978).

[2] Box, G.E.P and Jerkins, W.G. *Time series analysis: Forecasting and control* (San Francisco: Holden-Day, 1976).

[3] Chen, B. and Gel, Y.R. *A sieve boostrapt two-sampe t-test under serial correlation*, Journal of Biopharmaceutical Statistics **21**, 1100-1112, 2011.

[4] Katz, R.W. *Statistical evaluation of climate experiments with general circulation models: A parametric time series approach*, Journal of the Atmospherie Sciences **39**, 1446-1455, 1982.

[5] Seitshiro, M.B. *Two-sample comparisons for serially correlated data.* Dissertation Thesis for Master of Science in Statistics, School of Computer, Statistical and Mathematical Sciences, North-West University, Potchefstroom, South Africa, 2006.

[6] Thiébauz, H.J and Zwiers, F.W. *The interpretation and estimation of effective sample size*, Journal of Applied Meteorology and Climate **23**, 800-811, 1984.

[7] Wilks, D.S. *Resampling hypothesis tests for autocorrelated fields*, Journal of Climate **10**, 65-82, 1997.

[8] Zimmerman, D.W. *Correcting two-sample z and t tests for correlation: An alternative to one-sample tests on difference scores*, Psicológica **33**, 391-418, 2012.