27

Large-sample properties of variance stabilizing transformations

27.1 Existence of the variance stabilizing transformation

The following description of (asymptotic) univariate variance stabilizing transformations (*vst*'s) is taken from Holland (1973), which gives a nicely written account of the subject.

Let X_n be a real-valued random variable with distribution depending upon a real parameter $\theta \in D$, an open interval in $\mathbb{R}$. X_n may for example be an estimator based on a sample of size n. Suppose that for every $\theta \in D$ the quantity $\sqrt{n}(X_n - \theta) \rightarrow N(0, \sigma^2(\theta))$ in distribution. The asymptotic variance $\sigma^2(\theta) > 0$ is assumed to be continuous in D.

An asymptotic *vst* is a one-to-one, continuously differentiable mapping $f : D \rightarrow \mathbb{R}^1$ such that $\sqrt{n}(f(X_n) - f(\theta)) \rightarrow N(0, 1)$ in distribution. Since $X_n \rightarrow \theta$ in probability, $X_n \in D$ with probability as close to 1 as needed for n large enough. Therefore f is defined for the possible values of X_n with a probability approaching 1 as $n \rightarrow \infty$. This situation is, of course, not satisfactory in practice. As a remedy and in order to apply a *vst*, one may have to extend the definition of $f(\cdot)$.

Assume that f exists and has a differential at each $\theta \in D$, i.e. if $|x_n - \theta| = O(n^{-1/2})$ then $f(x_n) = f(\theta) + (x_n - \theta) f'(\theta) + o(n^{-1/2})$. For the random variable

Meta Analysis: A Guide to Calibrating and Combining Statistical Evidence Elena Kulinskaya, Stephan Morgenthaler, and Robert G. Staudte

X_n the same is true in probability, resulting in

$$\sqrt{n}(f(X_n) - f(\theta)) \to N(0, \sigma^2(\theta)(f'(\theta))^2)$$

in distribution. Since only continuously differentiable solutions of the differential equation $\sigma^2(\theta)(f'(\theta))^2 = 1$ are acceptable the sign should be 1 or -1 for all $\theta \in D$. In summary, the one-dimensional asymptotic *vst* problem always has a one-to-one continuously differentiable solution given by

$$f(\theta) = f(\theta_0) \pm \int_{\theta_0}^{\theta} (\sigma(t))^{-1} \mathrm{d}t. \tag{27.1}$$

The solution is unique up to an additive constant and the sign of its derivative. The only requirement is that $\sigma(\theta)$ is a continuous nonzero function of θ in D, it does not have to be one-to-one and may be constant. In the following, we will choose the positive sign in the defining equation for the *vst*. The additive constant allows us to fix the value of the *vst* at one point, for example, $f(\theta_0) = 0$.

So far we were looking at the asymptotic *vst* valid in the $n^{-1/2}$-vicinity of θ_0. But often the *vst* is defined on a much larger region, or even globally, as will be seen in the examples in the next section. Interestingly, in all these examples, and in the majority of variance functions for traditional exponential families, the variance $\sigma^2(\theta)$ is a first- or second-degree polynomial in θ. In such cases a global *vst* exists and is a rather simple transformation.

We also were working in a most simple case of $\sigma^2 = \sigma^2(\theta)$, but a more general case is $\sigma^2 = \sigma^2(\xi; \theta)$, where ξ is a nuisance parameter. An example is the *vst* for the Student t, where the variance is the nuisance parameter. The presence of nuisance parameters modifies the above asymptotic theory as follows. Equation (27.1) changes to

$$f(\theta|\xi) = f(\theta_0|\xi) + \int_{\theta_0}^{\theta} (\sigma(\xi; t))^{-1} \mathrm{d}t, \tag{27.2}$$

which means that the *vst* depends on ξ. Suppose $\hat{\xi}$ is asymptotically independent of $X_n = \hat{\theta}$, and $\hat{\xi} \to \xi$ in probability, then we may solve (27.2) with $\hat{\xi}$ substituted for ξ.

The asymptotic *vst* is based on the asymptotic variance $\sigma^2(\theta)$ and Equation (27.1). In a finite sample setting, approximate variance stabilization can be achieved by applying (27.1) to the actual variance $\sigma_n^2(\theta)$. We denote the finite sample *vst* by $f_n(\cdot)$. When $n \to \infty$, the *vst* has the effect of rendering the asymptotic variance equal to 1 for all θ. For finite n, this holds only approximately, but in practice often goes a long way towards this goal.

27.2 Tests and effect sizes

Let us now compare tests and effect sizes before and after the *vst*. In the previous section we considered an estimator $X_n = \hat{\theta}$. Its mean $\mathrm{E}(X_n) = \theta$ was the main parameter of interest. In a slightly more general setting we shall consider test statistics X_n of a null hypothesis involving a real-valued parameter ζ. Denote the expectation of the test statistic $\theta(\zeta) = \mathrm{E}_\zeta(X_n)$. Assume without loss of generality that under the null hypothesis $\zeta = 0$ and let $\theta_0 = \theta(0)$. The effect size associated with the test based on $Y_n = \sqrt{n}\,(X_n - \theta_0)$ is

$$\delta = (\theta - \theta_0)/\sigma(\theta) = (\theta - \theta_0) f'(\theta), \tag{27.3}$$

where $f(\cdot)$ is the asymptotic *vst* for X_n.

The Pitman efficacy of a test describes the behavior of the asymptotic power. The Pitman efficacy of the test Y_n is

$$e_Y = (d\theta(0)/d\zeta)\,\sigma(\theta_0)^{-1} = (d\theta(0)/d\zeta)\; f'(\theta_0).$$

This result holds if $\theta(\zeta)$ is differentiable in ζ at 0 with a positive derivative and σ is continuous at θ_0 and nonzero. The ARE of such tests is the ratio of squared efficacies, see Theorem 14.19 from van der Vaart (1998).

After application of the *vst* we obtain what we call evidence statistic throughout this book. This is another test, which has the form $T_n = \sqrt{n}\, f(X_n)$. For T_n we have weak convergence to a unit normal distribution $\sqrt{n}(f(X_n) - f(\theta)) \to N(0, 1)$. This new test statistic thus has an effect size of

$$f(\theta) - f(\theta_0). \tag{27.4}$$

It is easy to see that the Pitman efficacy of a test is not affected by the application of a *vst*. The two tests based on Y_n and T_n are asymptotically equivalent.

Lemma 27.1 *The Pitman efficacy of a test remains constant under the application of the variance stabilizing transformation. In this sense, the tests Y_n and T_n are equivalent.*

The proof is straightforward. The efficacy of T_n is computed with the help of (27.4) and equals $e_T = \mathrm{d} f(\theta)/\mathrm{d}\zeta$, where the derivative is evaluated at $\zeta = 0$. The chain rule then leads to $e_T = (\mathrm{d}\theta(0)/\mathrm{d}\zeta)\; f'(\theta_0) = e_Y$.

Comparing the original effect size δ and the effect size after variance stabilization $f(\theta) - f(\theta_0)$ on an interval (θ_0, θ) we obtain the following result.

Lemma 27.2 *Suppose the vst $f(\cdot)$ is twice continuously differentiable. It follows that the effect size of the transformed test T_n is larger than the effect size of the original test Y_n if and only if (iff) the vst is concave on* (θ_0, θ). *This holds iff* $\mathrm{d}\sigma/\mathrm{d}\theta > 0$, *which means that σ is an increasing function of the parameter θ.*

Proof: We expand $f(\theta_0)$ around the θ and obtain $f(\theta_0) = f(\theta) + f'(\theta)(\theta_0 - \theta) + \{f''(c)/2\}\,(\theta - \theta_0)^2$, for some c lying between θ_0 and θ. When applying this to the effect size for the test based on T_n we have

$$\begin{aligned} f(\theta) - f(\theta_0) &= (\theta - \theta_0) f'(\theta) - \frac{(\theta - \theta_0)^2}{2} f''(c) \\ &= \delta - \frac{(\theta - \theta_0)^2}{2} f''(c). \end{aligned}$$

The transformed effect size is thus larger than the original effect size iff the *vst* is concave on (θ_0, θ), i.e. iff $f''(\theta) < 0$.

Recall that $f'(\theta) = \sigma(\theta)^{-1}$. From this it follows that $f''(\theta) = -(\sigma(\theta))^{-2}\mathrm{d}\sigma/\mathrm{d}\theta$. This shows that the *vst* $f(\cdot)$ is concave on (θ_0, θ) iff $\mathrm{d}\sigma/\mathrm{d}\theta > 0$.

Example 1. Poisson counts

We observe a sample of counts, each having a Poisson distribution with expectation μ. The estimate for μ is the sample mean X_n, which satisfies $\sqrt{n}(X_n - \mu) \to N(0, \mu)$. To test $\mu = \mu_0$ versus $\mu > \mu_0$ we use $Y_n = \sqrt{n}(X_n - \mu_0)$. The asymptotic *vst* for X_n is up to an additive constant and a sign change equal to twice the square root, so that $T_n = \sqrt{n} f(X_n) = 2\sqrt{Y_n}$.

The effect size before applying the *vst* is $\delta = (\mu - \mu_0)/\sqrt{\mu}$. The transformed effect is $2(\sqrt{\mu} - \sqrt{\mu_0})$. The derivative of $\sigma(\mu)$ is positive for $\mu > 0$, therefore the *vst* increases the effect size.

Example 2. The *t*-test

The parameter of interest when using the t-statistic is the mean μ, but we are in the presence of a nuisance parameter, the variance σ^2. The test is constructed with the help of X_n and s_n, the sample mean and standard deviation, which are asymptotically independent. We reject $\mu = \mu_0$ in favor of $\mu > \mu_0$ for large values of $Y_n = \sqrt{n}(X_n - \mu_0)/s_n$. A relevant standardized effect is Cohen's $d = (\mu - \mu_0)/\sigma$, and $Y_n = \sqrt{n}\hat{d}_n$. The statistic Y_n is approximately normal with mean $\sqrt{n}d$ and variance $\sigma^2(d) = 1 + d^2/2$ and the corresponding finite sample *vst* is discussed in Chapter 20. One finds

$$f_n(\hat{d}_n) = \sqrt{2}\,\ln\left(\hat{d}_n/\sqrt{2} + \sqrt{1 + (\hat{d}_n/\sqrt{2})^2}\right).$$

The effect size before applying the *vst* is $\delta = d/\sqrt{1 + d^2/2}$. For the statistic $T_n = \sqrt{n} f_n(\hat{d}_n)$, the effect size is $f(d) = \sqrt{2}\ln(d/\sqrt{2} + \sqrt{1 + d^2/2})$. The two effect sizes are very close for small values of d. Say for $d = 0.05$, $\delta = 0.049967$ and $f(d) = 0.04999$. These functions grow very far apart for large values of d, with $\delta \to \sqrt{2}$ in the limit, whereas $f(\delta)$ is unlimited.

Example 3. Binomial proportions

Here $\sigma(p) = \sqrt{p(1 - p)}$, and $\delta = (p - p_0)/\sigma(p)$. The sign $\mathrm{sgn}(\mathrm{d}\sigma/\mathrm{d}\theta) = \mathrm{sgn}((1/2) - p)$. The sign is constant on (p_0, p) if p_0 and p are at the same side of 1/2. When $p_0 < p < 0.5$ the *vst* should result in increased effect size. When $0.5 < p_0 < p$ the transformed effect size should be smaller. The transformed effect size is

$$f(\delta) = \arcsin(1 - 2p_0) - \arcsin\left(\frac{1-2p_0-\delta\sqrt{1+\delta^2-(1-2p_0)^2}}{1+\delta^2}\right). \quad (27.5)$$

When $p_0 = 1/2$,

$$f(\delta) = \arcsin\left(\frac{\delta}{\sqrt{1+\delta^2}}\right) = \arctan(\delta).$$

This shows that $|f(\delta)| \leq |\delta|$ for $p_0 = \frac{1}{2}$.

Example 4. The sign test

Given n observations from a continuous distribution $F(\mu, \sigma)$ with unknown median μ and scale parameter σ, we wish to test $\mu = 0$ in favor of $\mu > 0$. The sign statistic $S_n = \sum_i I\{X_i > 0\} \sim$ Binomial (n, p_μ), with $p_\mu = 1 - F(-\mu/\sigma)$. This is exactly the previous example with $p_0 = 1/2$ and $p_\mu > 1/2$. The effect size δ is positive, and it is decreased by the *vst*.

27.3 Power and efficiency

The power of an asymptotically normal α-level test based on Y_n is approximately equal to $1 - \Phi(z_{1-\alpha} - \sqrt{n}\delta)$, where δ is given by (27.3). The sample size of a test with power $1 - \beta$ can be calculated from

$$\sqrt{n_Y} = \delta^{-1}(z_{1-\alpha} + z_{1-\beta}). \tag{27.6}$$

Similarly, after the *vst*, the test based on T_n has power $1 - \Phi(z_{1-\alpha} - \sqrt{n}(f(\theta) - f(\theta_0))$, and the sample size is calculated from

$$\sqrt{n_T} = (f(\theta) - f(\theta_0))^{-1}(z_{1-\alpha} + z_{1-\beta}). \tag{27.7}$$

Note that (27.6) implies that $\delta = (z_{1-\alpha} + z_{1-\beta})/\sqrt{n_Y}$, which is small for large sample sizes. Recall from Lemma 27.1 that for small values of δ we found $f(\theta) = f(\theta_0) + (\theta - \theta_0)\, f'(0) + o(\delta)$, which means that for large sample sizes (27.6) and (27.7) give approximately identical sample sizes, since the above implies that $f(\theta) - f(\theta_0) \approx \delta$. In many practical cases, though, this asymptotic equivalence is not sufficiently accurate and the two sample sizes n_Y and n_T may be quite different. The ratio of (nominal) sample sizes is

$$n_Y/n_T = ((f(\theta) - f(\theta_0))/\delta)^2. \tag{27.8}$$

The following result is a corollary of Lemma 27.2.

Corollary 27.3 *Asymptotic sample size calculation based on Y_n results in a larger/ smaller sample size than the one based on T_n (i.e. $n_Y > n_T$) iff the vst is concave ($\mathrm{d}\sigma/\mathrm{d}\theta > 0$)/convex ($\mathrm{d}\sigma/\mathrm{d}\theta < 0$).*

Example 5. The *t*-test (continued)

The effect size of the t-statistic is $\delta = d/\sqrt{1 + d^2/2}$ and becomes $\sqrt{2}\ln(d/\sqrt{2} + \sqrt{1 + d^2/2})$ after the stabilizing transformation, where $d = (\mu - \mu_0)/\sigma$ specifies the alternative. The ratio of the nominal sample sizes is

$$n_Y/n_T = (2+d^2)d^{-2}[\ln(d/\sqrt{2}+\sqrt{1+d^2/2})]^2.$$

Fixing the sample size n and the false positive and false negative error rates α and β, one can solve Equation (27.6) for d. Substituting the result in the above equation, we obtain an expression for the ratio of nominal sample sizes n_Y/n_T for a given sample size, level and power, denoted by $r(Y, T|n, \alpha, \beta)$. We leave it to the reader to check that this leads to

$$r(Y, T|n, \alpha, \beta) = \frac{2n}{(z_{1-\alpha}+z_{1-\beta})^2}\left[\ln\frac{(z_{1-\alpha}+z_{1-\beta})+\sqrt{2n}}{\sqrt{2n-(z_{1-\alpha}+z_{1-\beta})^2}}\right]^2. \qquad (27.9)$$

This ratio of sample sizes is a decreasing function of all three parameters, the sample size and the false positive and false negative error rates α and β. The limit of $r(Y, T|n, \alpha, \beta)$ when the sample size $n \to \infty$ is 1, but for moderate values of n (between 10 and 100) the ratio is considerably greater than 1. Some examples are given in Table 27.1.

Example 6. The sign test (continued)

Consider testing $H_0 : p_\mu = 1 - F(-\mu/\sigma) = 1/2$ versus $H_A : p_\mu > 1/2$. The effect size of the sign statistic is $\delta = (p_\mu - 1/2)/\sqrt{p_\mu(1-p_\mu)}$. After transformation to evidence via the *vst*, the effect size is $f(\delta) = \arctan(\delta)$. The ratio of the sample sizes is $n_Y/n_T = (\delta/\arctan(\delta))^2$. Substituting $\delta = (z_{1-\alpha}+z_{1-\beta})/\sqrt{n}$ in the above equation, we obtain an expression for the ratio $r(Y, T|n, \alpha, \beta)$ for a given triplet (n, α, β). This is an increasing function of the error rates α and β, and of the sample size n. The limit when the sample size $n \to \infty$ is 1, but for moderate values of n (between 10 and 100) the ratio $r(Y, T|n, \alpha, \beta)$ is considerably smaller than 1, i.e. the asymptotic

Table 27.1 Values of the ratio of nominal sample sizes $r(Y, T|n, \alpha, \beta)$ calculated from Equation (27.9) for *t*-test (columns 2–4) and the sign test (columns 5–7; see the text for explanation) for $\alpha = 0.05$, $\beta = 0.05, 0.10$ and 0.20 and for various sample sizes n.

	t-test			Sign test		
n	$\beta = 0.2$	$\beta = 0.1$	$\beta = 0.05$	$\beta = 0.2$	$\beta = 0.1$	$\beta = 0.05$
10	1.27	1.43	1.64	0.72	0.65	0.60
15	1.16	1.24	1.34	0.79	0.73	0.69
20	1.12	1.17	1.23	0.83	0.78	0.74
25	1.09	1.13	1.17	0.86	0.82	0.78
30	1.07	1.11	1.14	0.88	0.84	0.81
35	1.06	1.09	1.12	0.90	0.86	0.83
40	1.05	1.08	1.10	0.91	0.88	0.85
45	1.05	1.07	1.09	0.92	0.89	0.86
50	1.04	1.06	1.08	0.92	0.90	0.88
100	1.02	1.03	1.04	0.96	0.95	0.93

Table 27.2 Comparison of sample sizes for the t-test at $\alpha = 0.05$, $\beta = 0.05, 0.10$ and 0.20. The sample sizes were selected for the t-test (n_t) and calculated using Equation (27.7) for the evidence-based test (n_T). To do so, the value of $\delta_t = (\mu - \mu_0)/\sigma$ was calculated with the help of the NCSS-PASS (2005) software in order to match the chosen n_t, α and β.

n_t	β	δ_t	$f(\delta_t)$	n_T	n_t/n_T
10	0.05	1.131	1.036	10.09	0.99
15	0.05	0.894	0.843	15.22	0.99
20	0.05	0.764	0.731	20.25	0.99
25	0.05	0.677	0.653	25.34	0.99
30	0.05	0.615	0.597	30.35	0.99
35	0.05	0.568	0.554	35.29	0.99
40	0.05	0.529	0.517	40.43	0.99
45	0.05	0.498	0.488	45.40	0.99
50	0.05	0.472	0.464	50.34	0.99
100	0.05	0.331	0.328	100.56	0.99
10	0.1	1.005	0.935	9.79	1.02
15	0.1	0.795	0.758	14.90	1.01
20	0.1	0.679	0.655	19.94	1.00
25	0.1	0.603	0.586	24.93	1.00
30	0.1	0.547	0.534	30.01	1.00
35	0.1	0.505	0.495	34.97	1.00
40	0.1	0.471	0.463	40.00	1.00
45	0.1	0.443	0.436	45.04	1.00
50	0.1	0.420	0.414	49.95	1.00
100	0.1	0.295	0.293	99.82	1.00
10	0.2	0.853	0.808	9.46	1.06
15	0.2	0.675	0.652	14.56	1.03
20	0.2	0.577	0.562	19.57	1.02
25	0.2	0.512	0.501	24.59	1.02
30	0.2	0.465	0.457	29.60	1.01
35	0.2	0.429	0.423	34.61	1.01
40	0.2	0.400	0.395	39.66	1.01
45	0.2	0.376	0.372	44.75	1.01
50	0.2	0.357	0.353	49.53	1.01
100	0.2	0.250	0.249	99.94	1.00

sample size calculation based on the standard normal approximation to the sign test results in a considerably smaller sample size n_Y in comparison to the evidence-based sample size calculation n_T. Some examples are given in the last three columns of Table 27.1.

The results of these two examples are rather striking. The evidence-based sample size calculations for the t-test for sample sizes up to 100 give considerably smaller values of n, whereas for the sign test they result in considerably larger values of

Table 27.3 Comparison of sample sizes for the sign test at nominal level $\alpha = 0.05$ and for three values of the type II error β. The sample size n_S was chosen, whereas n_Y and n_T were calculated with the help of p_μ, the alternative corresponding to the triplet (n_S, α, β). These were computed by the program NCSS-PASS (2005).

n_S	p_μ	α	β	δ	n_T	n_Y	$n_T - n_S$	$n_S - n_Y$
10	0.963	0.011	0.05	2.460	11.08	2.57	1.08	7.43
15	0.903	0.018	0.05	1.365	15.98	7.56	0.98	7.44
20	0.860	0.021	0.05	1.040	20.95	12.55	0.95	7.45
25	0.830	0.022	0.05	0.877	25.92	17.47	0.92	7.53
30	0.779	0.049	0.05	0.672	31.02	24.05	1.02	5.95
35	0.764	0.045	0.05	0.623	35.99	28.78	0.99	6.22
40	0.753	0.040	0.05	0.586	40.97	33.53	0.97	6.47
45	0.743	0.036	0.05	0.556	45.93	38.25	0.93	6.75
50	0.735	0.032	0.05	0.532	50.95	43.03	0.95	6.97
100	0.664	0.044	0.05	0.346	100.98	93.59	0.98	6.41
10	0.946	0.011	0.1	1.963	10.61	3.33	0.61	6.67
15	0.878	0.018	0.1	1.156	15.60	8.58	0.60	6.42
20	0.834	0.021	0.1	0.898	20.60	13.67	0.60	6.33
25	0.804	0.022	0.1	0.765	25.58	18.64	0.58	6.36
30	0.752	0.049	0.1	0.585	30.71	25.15	0.71	4.85
35	0.739	0.045	0.1	0.544	35.71	29.94	0.71	5.06
40	0.729	0.040	0.1	0.514	40.67	34.70	0.67	5.30
45	0.720	0.036	0.1	0.490	45.64	39.46	0.64	5.54
50	0.713	0.032	0.1	0.470	50.61	44.22	0.61	5.78
100	0.647	0.044	0.1	0.306	100.72	94.83	0.72	5.17
10	0.917	0.011	0.2	1.508	10.16	4.34	0.16	5.66
15	0.843	0.018	0.2	0.942	15.23	9.80	0.23	5.20
20	0.799	0.021	0.2	0.745	20.25	14.96	0.25	5.04
25	0.770	0.022	0.2	0.641	25.25	19.97	0.25	5.03
30	0.718	0.049	0.2	0.486	30.41	26.35	0.41	3.65
35	0.707	0.045	0.2	0.455	35.39	31.18	0.39	3.82
40	0.698	0.040	0.2	0.432	40.37	35.98	0.37	4.02
45	0.691	0.036	0.2	0.413	45.34	40.78	0.34	4.22
50	0.685	0.032	0.2	0.398	50.34	45.60	0.34	4.40
100	0.626	0.044	0.2	0.260	100.41	96.12	0.41	3.88

n than the simple asymptotic approximation of the traditional test statistic would lead one to believe. How do the *vst* -based sample sizes compare with exact sample sizes obtained from the noncentral t or from the binomial distribution? The NCSS-PASS[1] (2005) software was used to obtain the values of $\delta_t = (\mu - \mu_0)/\sigma$

[1]Power analysis and sample size software produced by NCSS (http://www.ncss.com)

(effect size for t-test) for given values of n, α and β (see Table 27.2). Then the *vst* was applied to calculate the effect size of the evidence-based t-test, $f(\delta_t)$. Finally, the sample size n_T was calculated with (27.7). Surprisingly, the sample sizes agree down to $n = 10$, which shows that sample size computations based on the evidence are very accurate. Note that minor differences are explained by using exact and not rounded-up sample sizes for n_T.

For the sign test at nominal level $\alpha = 0.05$ and fixed β, the values of the actual levels α and P_μ (probability under alternative) were calculated by the NCSS-PASS (2005) for a given sample size n_S. Then effect size δ (effect size for sign test) and transformed effect size $f(\delta)$ were calculated as in Example 4, and were used to calculate approximate sample sizes n_Y and n_T using not the nominal, but the true α level. Table 27.3 contains the numerical values. The differences between the calculated sample sizes and the actual sample sizes derived from the program are given in last two columns. The evidence-based sample size is within 1 of the true sample size, whereas the classic asymptotic sample size calculation substantially underestimates the sample size needed.

27.4 Summary

In this chapter we have seen that under an assumption of asymptotic normality and some standard regularity conditions, the *vst* always exists. Evidence obtained via a *vst* is also asymptotically normal. Its ARE to the original test is 1. We have also demonstrated that the *vst* may both increase and decrease (positive) effect size, depending on the behavior of variance as the function of the distance from the null. This difference of the effect sizes may be very large, even unlimited. When the variance increases, the *vst* increases the effect size. When the opposite is true, the variance is the highest at the null (see Examples 3 and 4 above). In this case the effect size decreases when stabilizing the distribution. Finally, sample size calculations based on variance stabilizing transformations perform considerably better than standard asymptotic sample size calculations for sample sizes up to 100, as was shown for the t-test and the sign test.