Testing homogeneity of variance or expected value in sequence of independent normal distributions

Janusz L. Wywiał

Department of Statistics
Katowice University of Economics

Wrocławskie Obchody 100-lecia Istnienia
Polskiego Towarzystwa Statystycznego
18-19 X Wrocław

Problem:

Let: $\mathbf{Z} = [Z_1, ..., Z_r]^T$ has spherical normal distribution.

Hypothesis:

$$H_0 : \quad \mathbf{Z}^T \sim N\left(\mathbf{0}_m, \sigma^2 \mathbf{I}_m\right),$$

$$H_1 : \quad \mathbf{Z}^T \sim N\left(\boldsymbol{\mu}, \sigma^2 \text{diag}\boldsymbol{\delta}\right), \qquad \boldsymbol{\mu} \neq \mathbf{0}_m \quad \text{or} \quad \boldsymbol{\delta} \neq \mathbf{J}_m.$$

Particularly:

$$H_1 : \quad \mathbf{Z}^T \sim N\left(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_m\right), \quad \text{else} \quad H_1 : \quad \mathbf{Z}^T \sim N\left(\mathbf{0}_m, \sigma^2 \text{diag}\boldsymbol{\delta}\right).$$

Test statistic:

$$L_m = -\frac{1}{m}\sum_{i=1}^{m} ln\left(Z_i^2\right) + ln\left(\sum_{i=1}^{m} Z_i^2\right). \quad (1)$$

The similar construction to the Bartlett (1937) test statistic. The critical region is right side.

If $m \longrightarrow \infty$, then $Q_m = \frac{L_m - E(L_m)}{D(L_m)} \longrightarrow U \sim N(0, 1)$.

The test L_m is asymptotically unbiased for H_0 when $m \longrightarrow \infty$.

Moments I:

$$E(L_m) = \begin{cases} 2\sum_{r=1}^{m'} \frac{1}{2r-1}, & \text{if} \quad m = 3, 5, ..., \\ 2ln(2) + \sum_{r=1}^{m''} \frac{1}{r}, & \text{if} \quad m = 2, 4, ... \end{cases} \tag{2}$$

where $m' = \frac{m-1}{2}, \qquad m'' = \frac{m-2}{2},$

$$D^2(L_m) = \begin{cases} 4\sum_{r=1}^{m'} \frac{1}{(2r-1)^2} - \frac{m-1}{2m}\pi^2, & \text{if} \quad m = 3, 5, ..., \\ \sum_{r=1}^{m''} \frac{1}{r^2} - \frac{m-3}{6m}\pi^2, & \text{if} \quad m = 2, 4, ..., \end{cases} \tag{3}$$

$$E(L_m) \approx ln(2m) + 0,5 + \frac{1}{m} = o(m), \tag{4}$$

$$D^2(L_m) \approx \frac{2,935}{m} - \frac{2}{m^2} = O(m^{-1}), \tag{5}$$

Moments II:

$$\eta_3(L_m) = \begin{cases} 16\sum_{r=1}^{m'} \frac{1}{(2r-1)^3} - 14\zeta(3)\frac{m^2-1}{m^2}, & \text{if} \quad m = 3, 5, ..., \\ 2\sum_{r=1}^{m''} \frac{1}{r^3} - 2\zeta(3)\frac{m^2-7}{m^2}, & \text{if} \quad m = 2, 4, ... \end{cases} \tag{6}$$

where $\zeta(3) \approx 1,202$,

$$\Delta(L_M) = \begin{cases} 96\sum_{r=1}^{m'} \frac{1}{(2r-1)^4} - \pi^4\frac{m^3-1}{m^3}, & \text{if} \quad m = 3, 5, ..., \\ 6\sum_{r=1}^{m''} \frac{1}{r^4} - \frac{\pi^4}{15}\frac{m^3-15}{m^3}, & \text{if} \quad m = 2, 4, ... \end{cases} \tag{7}$$

where: $\Delta(L_M) = \eta_4(L_m) - 3D^4(L_m)$,

$$\eta_3(L_m) \approx \frac{12,828}{m^2} - \frac{8}{m^3} = O(m^{-2}), \tag{8}$$

$$\Delta(L_m) \approx \frac{81,409}{m^3} - \frac{48}{m^4} = O(m^{-3}). \tag{9}$$

Cornish-Fisher'a approximation

of the p-th quantile q_p of the test statistic Q_m.

$Q = \frac{L_m - E(L_m)}{D(L_m)}$

$$q_p = z_p + \frac{1}{6}\beta_1\left(z_p^2 - 1\right) + \frac{1}{24}(\beta_2 - 3)\left(z_p^3 - 3z_p\right)$$

where: z_p be a $p - th$ quantile of order p of the standard normal variable and

$$\beta_1 = \frac{\eta_3(L_m)}{D^3(L_m)}, \qquad \beta_2 = \frac{\eta_4(L_m)}{D^4(L_m)}.$$

Simulated critical values of the test

evaluated on the basis 10000 replican of the normal samples.

Tablica: The critical values of the test L_m.

m	$\alpha = .1$	$\alpha = .05$	m	$\alpha = .1$	$\alpha = .05$	m	$\alpha = .1$	$\alpha = .05$
3	3.138	3.687	9	4.086	4.371	15	4.489	4.699
4	3.425	3.898	10	4.169	4.430	16	4.533	4.736
5	3.607	4.018	11	4.235	4.492	17	4.587	4.781
6	3.762	4.123	12	4.311	4.544	18	4.634	4.824
7	3.894	4.227	13	4.373	4.592	19	4.675	4.854
8	3.999	4.301	14	4.442	4.651	20	4.713	4.890

Let $Y^{(1)}, Y^{(2)}, ..., Y^{(m)}$ be such a possible order of the sequence of the random variables $Y_1, ..., Y_m$ that

$$|Y^{(1)} - \bar{Y}| \leq |Y^{(2)} - \bar{Y}| \leq .... \leq |Y^{(m)} - \bar{Y}|.$$

The $T - M$ statistics (1972):

$$T_k = \frac{1}{nS^2} \sum_{i=1}^{m-k} \left(Y^{(i)} - \bar{Y}_k \right)^2$$

where

$$\bar{Y}_k = \frac{1}{m-k} \sum_{i=1}^{m} Y^{(i)}, \qquad S^2 = \frac{1}{m} \sum_{i=1}^{m} (Y_i - \bar{Y})^2, \qquad \bar{Y} = \frac{1}{m} \sum_{i=1}^{m} Y_i.$$

The left side critical region.
The $T - M$ test become the Grubbs (1950) test, when $k = 1$.
Tables with the critical values of $T - M$ test: Smaliak and Titarienko (1980) for $k = 1(1)10$ and $m = 3(1)20(5)50$.

- $Z_i \sim N(\mu_i; 1), \quad i = 1, ..., m,$
- $H_0: \quad \mu_i = 0, \quad i = 1, ..., m, \quad \mu_i = E(Z_i), \quad \sim N(\mu_i; 1)$
- $H_1: \quad \mu_i = 0, \quad i = 1, ..., m - k, \quad \mu_i \neq 0, \quad i = 1, ..., k$

There is k outliers.

Tablica: The power (%) of the tests for one or two outliers.

μ :	2		3		4		-2, 2		-3, 3		-4, 4	
n	T_1	L_n	T_1	L_n	T_1	L_n	T_2	L_n	T_2	L_n	T_2	L_n
6	8	5	17	3	30	9	8	12	16	23	22	42
8	10	5	23	6	49	19	10	13	18	25	43	47
10	10	6	27	9	56	26	13	13	26	27	57	49
15	11	6	29	13	59	35	13	13	39	30	68	49
20	12	6	30	14	65	41	15	14	43	30	75	55
30	13	5	32	16	68	48	16	12	48	29	82	53
40	12	5	33	18	68	50	16	9	46	22	85	48
50	12	4	30	14	68	48	15	9	45	22	85	45

Tablica: The power (%) of the tests for six outliers.

μ :	-2,-2,-2,2,2,2		-3,-3,-3,3,3,3		-2,-3,-4,2,3,4	
n	T_6	L_n	T_6	L_n	T_6	L_n
12	4	16	6	36	11	27
15	8	16	14	46	24	35
20	8	21	28	54	42	50
30	15	26	49	55	66	52
40	20	23	64	57	79	62
50	19	22	70	59	82	60

Tablica: The power (%) of the tests for outliers μ_1 and μ_2. Each of them share forth part of the sample size.

μ :	$\mu_1 = -2$, $\mu_2 = 2$		$\mu_1 = -3$, $\mu_2 = 3$		$\mu_1 = -4$, $\mu_2 = 4$	
n	T_k	L_n	T_k	L_n	T_k	L_n
4	11	9	11	19	14	28
8	3	10	6	24	13	40
12	4	16	6	36	12	63
16	3	17	8	41	18	73
20	4	21	8	52	15	78

Transformations of hypotheses I

Case of the simple random sample

- H_0 : $\mathbf{Y}_n^T \sim N\left(\mu \mathbf{J}_n, \sigma^2 \mathbf{I}_n\right)$
- H_1 : $\mathbf{Y}_n^T \sim N\left(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n\right)$, $\quad \boldsymbol{\mu} \neq \mu \mathbf{J}_n$

can be transformed into

- H_0' : $\mathbf{Z}_{n-1}^T \sim N\left(\mathbf{0}_{n-1}, \sigma^2 \mathbf{I}_n\right)$
- H_1' : $\mathbf{Z}_{n-1}^T \sim N\left(\boldsymbol{\mu}_z, \sigma^2 \mathbf{I}_n\right)$, $\quad \boldsymbol{\mu}_z \neq \mathbf{0}_{n-1}$

by means of the transf., see e.g. Csorgo and Seshardi (1970):

$$Z_t = \frac{\sum_{j=1}^{t} Y_j - tY_{t+1}}{\sqrt{t(t+1)}} \qquad t = 1, ..., n, \; m = n - 1.$$

The matrix form:

$$\mathbf{Z} = \mathbf{A}^T \mathbf{Y}$$

where: $\mathbf{A} = [\mathbf{A}_1 \; \mathbf{A}_2 ... \mathbf{A}_{n-1}]$, $\mathbf{A}_t^T = \frac{1}{\sqrt{t(t+1)}} [\mathbf{J}_t^T \; t \; \mathbf{0}_{n-t-1}^T]$, $\mathbf{0}_k$.

Transformations of hypotheses II

Case of testing the linearity of trend

- $H_0: \quad Y_t \sim N\left(\alpha_0 + \alpha_1 t, \quad t = 1, ..., n.\sigma^2\right)$
  against, particularly:

$$H_1: \begin{cases} Y_t \sim N\left(\alpha_{0,t} + \alpha_{1,t} t, \sigma^2\right) & \text{where} \\ \alpha_{0,t} = \alpha_0^{(1)} \quad \text{and} \quad \alpha_{1,t} = \alpha_1^{(1)} & \text{for} \quad t = 1, ..., n_1 \\ \alpha_{0,t} = \alpha_0^{(2)} \quad \text{and} \quad \alpha_{1,t} = \alpha_1^{(2)} & \text{for} \quad t = n_1 + 1, ..., n. \end{cases}$$

are transf. by means of $\mathbf{Z}_{n-2} = \mathbf{A}_{(n-2)\times n}\mathbf{Y}_n$, Wywiał(2010), into

- $H_0': \quad \mathbf{Z}_{n-1}^T \sim N\left(\mathbf{0}_{n-2}, \sigma^2 \mathbf{I}_n\right)$

$$H_1': \begin{cases} E(Z_t) = 0 & \text{for} \quad t = 1, ..., n_1 - 2, \\ E(Z_t) = g_t & \text{for} \quad t = n_1 - 1, ..., n - 2 \end{cases}$$

$$g_t = \frac{1}{3} n_1 c_t((n_1 + 1)(t - 2n_1 + 2)d_1 + (2t - 3n_1 + 3)d_0),$$

$$d_1 = \alpha_1^{(1)} - \alpha_1^{(2)},\ d_0 = \alpha_0^{(1)} - \alpha_0^{(2)},\ c_t = \frac{3}{\sqrt{t(t+1)(t+2)(t+3)}}.$$

Let $\mathbf{Y} = \mathbf{X}\beta + \xi$,
$\mathbf{X}$ is of full column rank non-random matrix of dimensions $n \times k$, $n > k$.

- $H_0 : \quad \xi \sim N\left(\mathbf{0}_n, \sigma^2 \mathbf{I}_n\right)$
- $H_1 : \quad \xi \sim N\left(\mathbf{\Delta}, \sigma^2 diag \delta\right), \qquad \mathbf{\Delta} \neq \mathbf{0}_n, \quad \delta > \mathbf{0}_n$

are transformed into

- $H_0' : \quad \mathbf{Z}_{n-k} \sim N\left(\mathbf{0}_{n-k}, \sigma^2 \mathbf{I}_{n-k}\right)$
- $H_1' : \quad \mathbf{Z}_{n-k} \sim N\left(\mathbf{\Delta}', \sigma^2 diag \delta'\right), \mathbf{\Delta}' \neq \mathbf{0}_{n-k}, \delta' > \mathbf{0}_{n-k}$

where $\mathbf{Z}_{n-k} = \mathbf{AY}$ is the residual vector of type BLUS, see Theil (1980).

Bibliography

- Bartlett M.S. (1937): Properties of sufficiency and statistical tests. *Proceedings of the Royal Society*, series A.
- Csorgo M., Seshardi V. (1970): On the problem of replacing hypotheses by equivalent simple ones. *Revue de l'Institute International de Statistique*, 333-339.
- Grubbs F. (1950): Sample criteria for testing autlying observations. *Ann. Math. Stat.,* vol. 21, 27-58.
- Smaliak S.A., Titarienko B.P. (1980): *Robust methods of estimation* (in Russian). Statistika, Moscow.
- Theil H. (1971): *Principles of Econometrics.* John Wiley and Sons, New York, London, Sydney, Toronto.
- Tietjen G., Moore H. (1972): Some Grubb's type stat. for detection of sev. outliers. *Technometrics,* vol. 14,583-597.
- Wywiał J. L. (2010): **Testing homogeneity of variance or expected value in sequence of independent normal distributions. *Acta Applicandae Mathematicae* vol. 110, Issue 2, 687-704**.