Colloquium Biometricum 42
2012, 121–126

# THE APPLICATION OF A SEQUENTIAL TEST IN THE STUDY ON THE INFLUENCE OF HUMIDITY LEVELS ON THE SELECTED GEOMETRIC FEATURES OF CEREAL GRAINS

**Dorota Domagała[1], Tomasz Guz[2]**

[1]Department of Applied Mathematics and Computer Science
[2]Department of Engineering and Food Machines
University of Life Sciences in Lublin
Akademicka 13, 20-950 Lublin, Poland
e-mail: dorota.domagala@up.lublin.pl; tomek.guz@up.lublin.pl

**Summary**

This paper was an attempt to examine the hypothesis that grains humidity levels have no effect on the chosen geometric features of five cereal species. For this purpose the sequential Rushton test was applied to allow the controlling of the type I and II errors simultaneously. The results are presented for the selected values of $\alpha$ and $\beta$.

**Keywords and phrases**: testing hypothesis, sequential Rushton test

**Classification AMS 2010**: 62F03, 62F99

## 1. Introduction

Sequential analysis was devised by Abraham Wald in 1943 for use in some problems of war research (Ghosh and Sen 1991, Marek and Noworol 1987). In sequential analysis the final number of observations are not predetermined but depend on the data themselves as they become available (Ghosh 1970). In traditional hypothesis testing, two possible action is taken: there are no reasons for rejecting the null hypothesis, or rejecting the null hypothesis in favour of the alternative hypothesis. Sequential tests give us the third possibility – to take more observations. Observations can be collected one by one or as a whole group (Denne and Jennison 2000, Jennison and Turnbull 1990). And each such observation, or group of observations, can result in accepting or rejecting of the verified hypothesis, or continuing of the study (Ghosh and Sen 1991, Marek and Noworol 1987).

## 2. Data

Grains of rye of Slavic variety, oats of Sławko variety, wheat of Tonacja variety, triticale of Pawo variety and barley of Stratus variety were humidified to the humidity of 14 and 18%. Basic geometric features of grains were measured using computer image analysis. In this study the length and width of grains were chosen.

A sequential Rushton test was used for testing the hypothesis that the grains humidity level has no effect on the chosen geometric features: length and width for each cereal species.

## 3. Background

Let $f(X,\theta)$ means a distribution of a random variable $X$ depends on unknown parameter $\theta$. One is interested in discriminating between two simple hypotheses

$$H_0;\theta=\theta_0 \quad H_1;\theta=\theta_1 \quad \theta_0\neq\theta_1 \tag{3.1}$$

The distribution of a random variable X is $f(X,\theta_0)$ when the null hypothesis is true or $f(X,\theta_1)$ when the alternative hypothesis is true.

A sequential probability ratio test (SPRT) for (3.1), that was proposed by Wald (Ghosh and Sen 1991, Ghosh 1970, Marek and Noworol 1987, Siegmund 1985), is defined by:

observe values of $X$, $x_i$, $i=1,2\ldots$ successively and independently. Let $I_n=\prod_{i=1}^{n}\frac{f(x_i,\theta_1)}{f(x_i,\theta_0)}$ be the observed likelihood ratio at a stage $n$:

- accept $H_0$ if $I_n\leq B$,
- accept $H_1$ if $I_n\geq A$,

- continue to stage $n+1$ if $B < I_n < A$,

where the stopping bounds, $-\infty < B < A < \infty$, are two real numbers determined by the type I error ($\alpha$) and type II error ($\beta$) as follows:

$$B = \frac{\beta}{1-\alpha}, \qquad A = \frac{1-\beta}{\alpha}.$$

In contrast to traditional hypothesis testing, in sequential hypothesis testing both the type I and type II error are predermined.

In practice, $i_n = \ln I_n = \sum_{i=1}^{n} \ln \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}$ is easier to use, taking

$$a = \ln A = \ln(1-\beta) - \ln\alpha \text{ and } b = \ln B = \ln\beta - \ln(1-\alpha).$$

Then the testing hypothesis procedure consists in comparing, at each stage, value of $i_n$ with $a$ and $b$.

In traditional hypothesis testing only the type I error can be controlled, whereas the use of the sequential tests enables to control both the type I and II errors. Moreover, for sequential tests fewer observations are necessary in contrast to traditional analysis where there is a risk of making decision based on insufficient number of observations.

## 4. Testing hypotheses about two means equality

Let us assume that $X_1 \sim N(\mu_1, \sigma)$ and $X_2 \sim N(\mu_2, \sigma)$ and $\sigma$ is unknown. We take two sample both of size $n$ for each random variable $X_1$ and $X_2$.

Consider now the problem of discriminating between

$$H_0: \mu_1 - \mu_2 = 0 \text{ and } H_1: \mu_1 - \mu_2 = \pm\psi\sigma \tag{4.1}$$

where $\psi$ is a positive constant. The alternative hypothesis are formulated in terms of standard deviation.

A sequential probability ratio test for (4.1) has the following form (Marek and Noworol 1987)

$$I_n = exp(-n\psi^2) M\left(\frac{2n-1}{2}, \frac{1}{2}, \frac{(2n-1)\psi^2 t^2}{2n-2+t^2}\right)$$

where $M$ is the confluent hypergeometric function, and $t^2 = \frac{n(\bar{x}_1 - \bar{x}_2)^2}{2s^2}$,
$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij}$, $i = 1, 2$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^{2}\sum_{j=1}^{n}\frac{(x_{ij}-\bar{x}_i)^2}{2}$.
From practical point of view, using the Rushton approximation of $M$

$$M\left(x, \frac{1}{2}, y\right) \cong \frac{1}{2} exp\left(\frac{1}{2}y + 2\sqrt{xy}\right)$$

is convenient (Marek and Noworol 1987, Tweel and all. 1996). Then the logarithm of $I_n$ is as follows

$$i_n \cong \frac{2(2n-1)\psi t}{\sqrt{2(2n-2+t^2)}} + \frac{1}{2}\frac{(2n-1)\psi^2 t^2}{2n-2+t^2} - n\psi^2 - \ln 2$$

## 5. Results

The hypotheses about means equality for grains at the humidity levels of 14% and 18% were tested for $\beta = 0.05, 0.1$ and $\alpha = 0.01, 0.05$ and for $\psi = 0.25, 0.5, 0.75, 1$. The analyses were both carried out on each individual cereal species and on the two studied features. At the first stage two samples, both of size 2, were taken from these two populations and the value of $i_n$ was calculated. For each of the samples one additional observation had been taken until $i_n < b$ or $i_n > a$. The results: which hypothesis, null or alternative, should be accepted and the number of observations $n$ allowing to make this decision, are presented in Table 1 for the grain length and in Table 2 for grain width.

A large number of observations were necessary to decide on accepting the null hypothesis both for the grain length and grain width for $\alpha = 0.01$, $\beta = 0.05$, $\psi = 0.25$, relatively (40-98). For $\psi = 0.25$ and the values of $\alpha$ and $\beta$ the alternative hypothesis was accepted and fewer observations were necessary (2-23). For the selected values of $\alpha$, $\beta$, $\psi$ the null hypothesis was accepted slightly more frequently. For $\psi = 1$ the alternative hypothesis was accepted almost every time (except $\beta = 0.05$ and $\alpha = 0.01, 0.05$ for the length of a rye grain).

**Table 1.** The grain length: a decision about accepting hypothesis $H_0$ or $H_1$ for number of observations *n*.

| $\alpha$ | $\beta$ | $\psi$ | barley | oats | wheat | triticale | rye |
|---|---|---|---|---|---|---|---|
| 0.01 | 0.05 | 0.25 | $H_0$ (48) | $H_0$ (43) | $H_0$ (86) | $H_0$ (95) | $H_0$ (88) |
| | | 0.5 | $H_0$ (12) | $H_0$ (14) | $H_0$ (14) | $H_0$ (15) | $H_0$ (28) |
| | | 0.75 | $H_0$ (9) | $H_0$ (6) | $H_0$ (5) | $H_0$ (5) | $H_0$ (18) |
| | | 1 | $H_0$ (9) | $H_0$ (4) | $H_0$ (5) | $H_0$ (5) | $H_1$ (10) |
| | 0.1 | 0,25 | $H_1$ (6) | $H_1$ (15) | $H_1$ (23) | $H_1$ (16) | $H_1$ (4) |
| | | 0.5 | $H_1$ (7) | $H_1$ (16) | $H_0$ (8) | $H_1$ (20) | $H_1$ (7) |
| | | 0.75 | $H_0$ (9) | $H_0$ (4) | $H_0$ (5) | $H_0$ (5) | $H_1$ (9) |
| | | 1 | $H_0$ (9) | $H_0$ (4) | $H_0$ (4) | $H_0$ (5) | $H_0$ (5) |
| 0.05 | 0.05 | 0.25 | $H_1$ (5) | $H_1$ (15) | $H_1$ (23) | $H_1$ (4) | $H_1$ (3) |
| | | 0.5 | $H_1$ (6) | $H_1$ (15) | $H_0$ (16) | $H_1$ (16) | $H_1$ (3) |
| | | 0.75 | $H_0$ (10) | $H_0$ (6) | $H_0$ (5) | $H_0$ (6) | $H_1$ (4) |
| | | 1 | $H_0$ (9) | $H_0$ (4) | $H_0$ (5) | $H_0$ (5) | $H_1$ (9) |
| | 0.1 | 0.25 | $H_1$ (5) | $H_1$ (15) | $H_1$ (23) | $H_1$ (4) | $H_1$ (3) |
| | | 0.5 | $H_1$ (6) | $H_1$ (15) | $H_0$ (8) | $H_1$ (16) | $H_1$ (3) |
| | | 0.75 | $H_0$ (9) | $H_0$ (4) | $H_0$ (5) | $H_0$ (5) | $H_1$ (4) |
| | | 1 | $H_0$ (9) | $H_0$ (4) | $H_0$ (5) | $H_0$ (5) | $H_0$ (5) |

**Table 2.** The grain width: a decision about accepting hypothesis $H_0$ or $H_1$ for number of observations *n*.

| $\alpha$ | $\beta$ | $\psi$ | barley | oats | wheat | triticale | rye |
|---|---|---|---|---|---|---|---|
| 0.01 | 0.05 | 0.25 | $H_0$ (51) | $H_0$ (98) | $H_0$ (40) | $H_0$ (86) | $H_0$ (56) |
| | | 0.5 | $H_0$ (14) | $H_0$ (38) | $H_0$ (20) | $H_0$ (15) | $H_0$ (15) |
| | | 0.75 | $H_0$ (7) | $H_0$ (9) | $H_0$ (13) | $H_0$ (6) | $H_0$ (13) |
| | | 1 | $H_0$ (6) | $H_0$ (5) | $H_0$ (10) | $H_0$ (5) | $H_0$ (5) |
| | 0.1 | 0.25 | $H_1$ (3) | $H_1$ (7) | $H_1$ (6) | $H_1$ (10) | $H_1$ (10) |
| | | 0.5 | $H_0$ (16) | $H_1$ (7) | $H_1$ (7) | $H_1$ (12) | $H_0$ (16) |
| | | 0.75 | $H_0$ (7) | $H_1$ (19) | $H_1$ (8) | $H_0$ (6) | $H_0$ (13) |
| | | 1 | $H_0$ (6) | $H_0$ (5) | $H_0$ (10) | $H_0$ (5) | $H_0$ (5) |
| 0.05 | 0.05 | 0.25 | $H_1$ (2) | $H_1$ (2) | $H_1$ (5) | $H_1$ (9) | $H_1$ (7) |
| | | 0.5 | $H_1$ (3) | $H_1$ (2) | $H_1$ (5) | $H_1$ (10) | $H_1$ (10) |
| | | 0.75 | $H_1$ (3) | $H_1$ (2) | $H_1$ (8) | $H_0$ (6) | $H_0$ (13) |
| | | 1 | $H_0$ (6) | $H_0$ (5) | $H_0$ (10) | $H_0$ (5) | $H_0$ (5) |
| | 0.1 | 0.25 | $H_1$ (2) | $H_1$ (2) | $H_1$ (5) | $H_1$ (7) | $H_1$ (7) |
| | | 0.5 | $H_1$ (2) | $H_1$ (2) | $H_1$ (5) | $H_1$ (10) | $H_1$ (10) |
| | | 0.75 | $H_1$ (3) | $H_1$ (2) | $H_1$ (8) | $H_0$ (6) | $H_0$ (13) |
| | | 1 | $H_0$ (6) | $H_0$ (5) | $H_0$ (10) | $H_0$ (5) | $H_0$ (5) |

## References

Denne J. S., Jennison C. (2000). A group sequential *t*-test with updating of sample size. *Biometrika* 87, 125–134,

Ghosh B.K. (1970). *Sequential test of statistical hypotheses*. Addison-Wesley Publishing Company, Inc.

Ghosh B.K., Sen P.K. (1991). *Handbook of sequential analys*is. Marcel Dekker New York.

Jennison C., Turnbull B.W., (1990). *Exact calculation for sequential t,* $\chi^2$ *and F tests*. Technical Report No. 888, Cornell University, Ithaca.

Marek T., Noworol C. (1987). *Analiza sekwencyjna w badaniach empirycznych*. PWN Warszawa.

Siegmund D. (1985). *Sequential analysis: tests and confidence intervals*. Springer-Verlag New York.

Tweel van der I., Kaaks R., Noord van P.A.H. (1996). Comparison of one-sample two-sided sequential *t*–tests for application in epidemiological studies. *Statistics in Medicine* 15, 2781–2795.