Michele Scagliairini

A Sequential Hypothesis Testing Procedure for the Process Capability Index $C_{pk}$

Quaderni di Dipartimento

Serie Ricerche 2016, n. 5

ISSN 1973-9346

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche “Paolo Fortunati”
Abstract

In this study we propose a sequential procedure for hypothesis testing on the $C_{pk}$ process capability index. We compare the properties of the sequential test with the performances of non-sequential tests by performing an extensive simulation study. The results indicate that the proposed sequential procedure makes it possible to save a large amount of sample size, which can be translated into reduced costs, time and resources.

Keywords

Average Sample Size; Brownian Motion; Maximal Allowable Sample Size; Power Function; Simulation Studies.
1. Introduction

Process capability indices assess the relationship between the actual process performance and the manufacturing specifications, and are the tools most frequently used for measuring the capability of a manufacturing process. The analytical formulation of these indices is easy to understand and straightforward to apply. The process capability indices most widely used in industry today are (Montgomery, 2009):

\[ C_p = \frac{USL - LSL}{6\sigma} = \frac{d}{3\sigma} \]  \hspace{2cm} (1)

\[ C_{pk} = \frac{d - |\mu - \frac{1}{2}(USL - LSL)|}{3\sigma} = \frac{d - |\mu - m|}{3\sigma} \]  \hspace{2cm} (2)

\[ C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} = \frac{d}{3\sqrt{\sigma^2 + (\mu - T)^2}} \]  \hspace{2cm} (3)

where \( \mu \) is the process mean, \( \sigma \) is the process standard deviation, \( LSL \) and \( USL \) are the specification limits, \( d = (USL - LSL)/2 \) is the half-length of the specification interval, \( m = (USL + LSL)/2 \) is the midpoint of the specification interval and \( T \) is the target value of the process.


Often, as a part of contractual agreement, suppliers are required to provide evidence that their processes satisfy a minimum level of capability. Such decision-making problem of demonstrating whether the process capability exceeds a pre-set capability requirement can be approached in terms of hypothesis testing.

Literature concerning process capability hypothesis testing includes numerous interesting researches. Just to mention a few we quote: the pioneering work by Kane (1986); the tests on \( C_{pk} \) investigated by Pearn and Chen (1999), Perakis and Xekalaki (2003), Pearn and Lin (2004), Chen and Hsu (2004) and Lin (2006); the Bayesian approach proposed by Fan and Kao (2006); the hypothesis testing studies on \( C_{pwm} \) by Pearn and Lin (2002) and Pearn et al. (2005); the model free
approach testing procedure proposed by Vännman and Kulachi (2008); the recently unified and comprehensive analysis of hypothesis testing with process capability indices by Lepore and Palumbo (2015); and finally, the sequential procedure for testing the equality of two indices $C_{pm}$ by Hussein et al. (2012).

In general, sequential methods for hypothesis testing (Tartakovsky et al. 2014) are appealing since they make it possible to reach decisions much more quickly, on average, than non-sequential procedures with the same discriminating power. This property, in the framework of the manufacturing industry, means that sequential procedures can lead to saving sample size, time and cost with consequent economic benefits and without any loss in quality.

In this study, starting from some of the results obtained by Hussein et al. (2012) which provided a sequential approach for testing the equality of the indices $C_{pm}$ for two processes, we propose a sequential procedure for hypothesis testing on the index $C_{pk}$.

We compared the sequential test properties with the performances of two non-sequential tests by performing an extensive simulation study. The results indicate that the proposed sequential test makes it possible to save a large amount of sample size, which can be translated into reduced costs, time and resources.

The paper is organized as follows. In Section 2, we review two of the most used tests for assessing whether a process is capable or not based on the $C_{pk}$ process capability index. In Section 3 we present the general sequential test procedure proposed by Hussein et al. (2012) and Hussein (2005). In Section 4 we develop and propose a sequential method for testing hypotheses on $C_{pk}$. In Section 5 we study the performances of the proposed test by performing a set of simulation studies. Section 6 contains a discussion of the results and finally, our concluding remarks are reported in Section 7.

2. Hypothesis testing on $C_{pk}$

To demonstrate whether a process meets the capability requirements the hypotheses of interest are

\[ H_0 : C_{pk} \leq c_{pk,0} \quad \text{the process is not capable} \]  
\[ H_1 : C_{pk} > c_{pk,0} \quad \text{the process is capable} \]

where $c_{pk,0}$ is the minimal requirement for $C_{pk}$.

For testing $H_0$ versus $H_1$, Pearn and Chen (1999), assuming a normally distributed quality characteristic, $X \sim N(\mu, \sigma^2)$, proposed a statistical test (PC-test) based on the distribution of the estimator
\[ \hat{C}_{pk} = b_j \hat{C}_{pk} \]  
(6)

where

\[ \hat{C}_{pk} = \frac{d - (\bar{X} - m) I_A(\mu)}{3S} \]  
(7)

\[ b_j = \frac{\Gamma[(n-1)/2]}{\Gamma((n-2)/2)} \sqrt{\frac{2}{n-1}} \], \( n \) is the sample size, \( \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \), \( S = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} \), \( I_A(\mu) = 1 \) if \( \mu \in A \) and \( I_A(\mu) = -1 \) if \( \mu \notin A \) with \( A = \{\mu | |\mu - m| \leq \delta\} \).

Given the type I error probability \( \alpha \), the critical value of the test is

\[ C_0 = \frac{b_j t_{n-1, \alpha} (\delta_c)}{3\sqrt{n}} \]  
(8)

where \( t_{n-1, \alpha} (\delta_c) \) is the upper \( \alpha \) quantile of a non-central \( t \) with \( n-1 \) degrees of freedom and non-centrality parameter \( \delta_c = 3\sqrt{n}c_{pk,0} \).

The power of the PC-test can be computed as

\[ \pi_{PC}(c_{pk,1}) = \Pr \{ \hat{C}_{pk} > C_0 | C_{pk} = c_{pk,1} \} = \Pr \{ t_{n-1}(\delta) > \frac{3\sqrt{n}C_0}{b_j} \} \]  
(9)

where \( \delta = 3\sqrt{n}c_{pk,1} \).

Lepore and Palumbo (2015), for testing \( H_0: C_{pk} \leq c_{pk,0} \) versus \( H_1: C_{pk} > c_{pk,0} \), discussed a test (LP-test) based on the estimator (Vännman, 2006)

\[ \hat{C}_{pk} = \frac{1-|\hat{\delta}|}{3\hat{\gamma}} \]  
(10)

where \( \hat{\gamma} = \frac{\hat{\sigma}}{\hat{\gamma}} \), \( \hat{\delta} = \frac{\bar{X} - m}{d} \) and \( \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}} \).

Under the assumption of a normally distributed quality characteristic, the authors obtained the cumulative distribution function of \( \hat{C}_{pk} \) as

\[
F_{\hat{C}_{pk}}(x) = \begin{cases} 
1 - \Pr(t_{n-1}(\delta_1) \leq -t) + \Pr(t_{n-1}(\delta_2) \leq -t) & x \leq 0 \\
1 - Q_{n-1}(-t, \delta_1; 0, R) + Q_{n-1}(t, \delta_2; 0, R) & x > 0
\end{cases}
\]  
(11)
where \( t_{n-1}(\delta) \) and \( t_{n-1}(\delta) \) are non-central \( t \) variables with \( n-1 \) degrees of freedom and non-centrality parameters \( \delta_1 = -3\sqrt{n} \frac{(1-\delta)}{1-|\delta|} C_{pk} \) and \( \delta_2 = 3\sqrt{n} \frac{(1+\delta)}{(1+|\delta|)} C_{pk} \) respectively, \( \delta = \frac{X - \mu}{d} \),

\[
R = \sqrt{n-1}(\delta_2 - \delta_1)/2t
\]

and

\[
Q_f(t;\alpha;0,R) = \frac{\sqrt{2\pi}}{\Gamma(f/2)^{2(f-2)/2}} \int_0^R \Phi\left(\frac{tx}{\sqrt{f}} - \delta\right) x^{f-1}\phi(x) dx
\]

is the \( Q_f \)-function proposed by Owen (1965), where \( \Gamma \) is the gamma function, and \( \Phi \) and \( \phi \) are respectively the normal cumulative distribution function and probability density function.

Lepore and Palumbo (2015) obtained the critical value for the test as

\[
c_{pk,\alpha} = \frac{t_{n-1}(\delta_0)}{3\sqrt{n-1}}
\]  

where \( t_{n-1}(\delta_0) \) is the upper \( \alpha \) quantile of a non-central \( t \) distribution with \( n-1 \) degrees of freedom and non-centrality parameter \( \delta_0 = 3\sqrt{n}c_{pk,0} \).

The power of the LP-test is then

\[
\pi_{LP}(c_{pk,\alpha}) = \Pr\{C_{pk} > c_{pk,\alpha} | \hat{C}_{pk} = c_{pk,\alpha}\} = 1 - \hat{F}_{c_{pk,\alpha}}(c_{pk,\alpha})
\]  

given that \( C_{pk} = c_{pk,\alpha} > c_{pk,0} \).

3. A general sequential method.

To describe the general sequential testing procedure proposed by Hussein et al. (2012) and Hussein (2005), we used \( x_1, x_2, \ldots, x_k, \ldots, \) to denote a sequence of multivariate independent observations collected over time. We assumed that these data came from a common multivariate distribution with density function \( f(x;\theta) \) where the vector of parameters \( \theta \) is unknown.

We were interested in testing

\[
H_0 : h(\theta) = 0 \text{ versus } H_1 : h(\theta) \neq 0
\]  

where \( h(\theta) : \mathbb{R}^d \to \mathbb{R}^q \), with \( q \leq d \), is a function with first order derivative matrix denoted by \( H(\theta) \).
Let us assume that for $\theta \in \Theta \subset \mathbb{R}^d$ with $q \geq d$ the following regularity conditions hold:

C1. The distribution function $F(x; \theta)$ of the vector random variable $X$ is identifiable over $\Theta$.

C2. There exists an open subset, $\Theta_0 \subset \Theta$, containing the true value of the parameter under $H_0$, such that the partial derivatives
\[ \frac{\partial}{\partial \theta_i} \ln f(x; \theta), \quad \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x; \theta), \quad \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln f(x; \theta) \]
exist and are continuous for all $x \in \mathbb{R}^d$, $\theta \in \Theta_0$.

C3. For each $\theta \in \Theta_0$ and $k = 1, 2, 3, \ldots$, the score equation
\[ \sum_{i=1}^k \frac{\partial \ln f(x; \theta)}{\partial \theta_i} = 0 \]
has a unique solution.

C4. There are functions, $M_1(x)$ and $M_2(x)$, that have finite expectations under any of the parameter values, $\theta \in \Theta_0$, such that
\[ \left| \frac{\partial}{\partial \theta_i} \ln f(x; \theta) \right| \leq M_1(x), \quad \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x; \theta) \right| \leq M_2(x), \quad \left| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_k} \ln f(x; \theta) \right| \leq M_2(x) \]
for all $\theta \in \Theta_0$, $1 \leq i, j, k \leq d$.

C5. $E_\theta \left( \frac{\partial}{\partial \theta_i} \ln f(x; \theta) \right) = 0$, $1 \leq i \leq d$, $\theta \in \Theta_0$.

C6. $I^{-1}(\theta)$ and its elements $I_{ij}(\theta) = E_\theta \left[ \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x; \theta) \right) \right]$ exist and are continuous for all $\theta \in \Theta_0$, $1 \leq i, j \leq d$.

C7. $\text{Var}_\theta \left[ \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(x; \theta) \right) \right] < \infty$ for $1 \leq i, j \leq d$.

Let us further assume that:

C8. $E_\theta \left( \frac{\partial}{\partial \theta_i} \ln f(x; \theta) \right)^{2+\delta} < \infty$, $i = 1, 2, \ldots, d$, and for some $\delta > 0$.

C9. The function $h(\theta)$ is continuously differentiable over $\Theta_0$ and its first-order derivative matrix $H(\theta)$, is bounded and of rank $q$.

Let us now consider a fixed sample design with sample size equal to $k$ and let us consider the Wald’s statistic.
\[ W_k = kh\left(\hat{\theta}_k\right)\left[H'(\theta) I^{-1}(\theta) H(\theta)\right]^{-1} h\left(\hat{\theta}_k\right)' \]  

(15)

where \( \hat{\theta}_k \) is a consistent estimator of \( \theta \).

Hussein et al. (2012) in Theorem 1 showed that under \( H_0 \), and if conditions C1-C9 hold, there exists an independent Wiener process, \( B_j(t), \ j = 1,2,...,q \), such that for \( \alpha \leq \frac{1}{2} - 1/(2 + \delta) \) with \( \delta > 0 \),

\[ \sup_{t \leq \alpha} \left| W_{[k]}^{-1} - U_{[k]} \right| = O\left(k^{-\alpha} (\ln \ln k)^{1/2}\right) \]  

(16)

Where

\[ U(x) = \frac{1}{x} \sum_{j=1}^{q} B_j^2(x) \]  

(17)

\[ W_{[k]} = [kt]h\left(\hat{\theta}_{[k]}\right)\left[H'(\theta) I^{-1}(\theta) H(\theta)\right]^{-1} h\left(\hat{\theta}_{[k]}\right)' \]  

(18)

and \([.]\) denotes the integer part of its argument.

The statistic \( W_k \) can therefore be approximated by a functional of Brownian motions. Furthermore, the authors derived the limiting distribution of \( W_k \). In detail they showed that (Corollary 1):

- under the conditions of Theorem 1
  \[ \max_{1 \leq k \leq n} \left[ \frac{k}{n} W_k \right]^{1/2} \overset{D}{\to} \sup \left( \sum_{j=1}^{q} B_j^2(t) \right)^{1/2} \]  

(19)

where \( \overset{D}{\to} \) denotes convergence in distribution;

- replacing the unknown \( \theta \) in the term \( H'(\theta) I^{-1}(\theta) H(\theta) \) by any almost surely convergent estimator, Corollary 1 remains valid.

Therefore, Hussein et al. (2012) defined as test statistic

\[ W_k^* = kh\left(\hat{\theta}_k\right)\left[H'(\hat{\theta}_k) I^{-1}(\hat{\theta}_k) H(\hat{\theta}_k)\right]^{-1} \left(h(\hat{\theta}_k)\right)' \]  

(20)

where \( \hat{\theta}_k \) is the maximum likelihood estimator of \( \theta \), and proposed the following \( \alpha \)-level sequential test truncated at the maximal allowable sample size \( n_0 \).

The sequential test procedure is performed as follows:

- for \( k = 2,3,...,n_0 \) compute of the statistic
  \[ W_k^{(1)} = \sqrt{k/n_0} \sqrt{W_k^*} ; \]  

(21)
• the hypothesis $H_0$ is rejected the first time that $W_k^{\alpha(1)}$ exceeds the critical value $w_\alpha$;
• if $W_k^{\alpha(1)}$ does not exceed $w_\alpha$ by $n_0$ then do not reject $H_0$.

The maximal sample size $n_0$ can be decided on the basis of financial, ethical or statistical reasons as, for example, to achieve a desired power level.

Given the Type I error probability $\alpha$, the critical value $w_\alpha$ can be obtained from Borodin and Salminen (1996).

4. A sequential test for $C_{pk}$

Let us now consider the hypothesis

$$H_0 : C_{pk} = c_{pk,0}$$

versus

$$H_1 : C_{pk} \neq c_{pk,0}$$

and assume that the quality characteristic of interest $X$ is normally distributed: $X \sim N(\mu, \sigma^2)$.

For $C_{pk} \geq 0$, $H_0$ is equivalent to

$$H_0 : \ln\left(\left(C_{pk}\right)^2\right) - \ln\left(\left(c_{pk,0}\right)^2\right) = 0$$

and the alternative hypothesis is equivalent to

$$H_1 : \ln\left(\left(C_{pk}\right)^2\right) - \ln\left(\left(c_{pk,0}\right)^2\right) \neq 0$$

Let us define the function $h(\theta)$ as

$$h(\theta) = \ln\left(\left(C_{pk}\right)^2\right) - \ln\left(\left(c_{pk,0}\right)^2\right) = \ln\left[\frac{(d - |\mu - m|)}{9\sigma^2\left(c_{pk,0}\right)^2}\right]$$

where $\theta = (\mu, \sigma^2)$. Note that in this framework $h(\theta) : \mathbb{R}^d \to \mathbb{R}^q$ with $d=2$ and $q=1$.

In the case at hand, where $q=1$, the statistic $W_k^*$ can be written as

$$W_k^* = kh^2\left(\hat{\theta}_k\right)\left[H'\left(\hat{\theta}_k\right)\Gamma^{-1}\left(\hat{\theta}_k\right)H\left(\hat{\theta}_k\right)\right]^{-1}$$

(27)
where \( \hat{\theta}_k = (\bar{X}_k, S_k^2) \) with \( \bar{X}_k = \frac{\sum_{i=1}^{k} X_i}{k} \) and \( S_k^2 = \frac{\sum_{i=1}^{k} (X_i - \mu)^2}{k} \).

The function \( h(\hat{\theta}_k) \) is therefore given by

\[
h(\hat{\theta}_k) = \ln \left( \frac{(d - |\bar{X}_k - m|)^2}{9S_k^2 (c_{pk,0})^2} \right)
\]

and the partial derivative matrix of \( h(\theta) \) computed at \( \hat{\theta}_k \) can be written as

\[
H(\hat{\theta}_k) = \begin{bmatrix}
-2\text{signum}[\bar{X}_k - m] \\
\frac{d - |\bar{X}_k - m|}{1 + S_k^2} \\
\end{bmatrix}
\]

where \( \text{signum}[a] = \frac{a}{|a|} \) if \( a \neq 0 \); \( \text{signum}[a] = 0 \) if \( a = 0 \).

Since the Fisher information matrix for normally distributed data is \( I(\theta) = \begin{bmatrix} 1/\sigma^2 & 0 \\ 0 & 1/2\sigma^4 \end{bmatrix} \), it follows that the statistic \( W_k^* \) can be written as

\[
W_k^* = k \left[ \ln \left( \frac{(d - |\bar{X}_k - m|)^2}{9S_k^2 (c_{pk,0})^2} \right) \right]^2 \times \frac{4\text{*signum}[\bar{X}_k - m]^2 S_k^2}{[d - |\bar{X}_k - m|]^2 + 2}^{-1}
\]

Therefore, given the value of \( \alpha \) and the maximal allowable sample size \( n_0 \), the test is performed by computing, for \( k=2,3,\ldots, n_0 \), the statistic

\[
W_k^{\gamma(\alpha)} = \sqrt{\frac{k}{n_0}} W_k^* = \sqrt{\frac{k}{n_0}} k \left[ \ln \left( \frac{(d - |\bar{X}_k - m|)^2}{9S_k^2 (c_{pk,0})^2} \right) \right]^2 \times \frac{4\text{*signum}[\bar{X}_k - m]^2 S_k^2}{[d - |\bar{X}_k - m|]^2 + 2}^{-1}
\]
Let \( n_{\text{stop}} \) be the first integer \( k=2,3,\ldots, n_0 \) for which \( W_k^{(i)} > w_\alpha \):  

- we reject \( H_0 \) if \( W_{\text{stop}}^{(i)} > w_\alpha \): 
- we do not reject \( H_0 \) if \( W_k^{(i)} \) does not exceed \( w_\alpha \) by \( n_0 \).

In this framework \( n_{\text{stop}} \) is the stopping sample size of the test.

In the case at hand, where \( q=1 \), the critical value \( w_\alpha \) is obtained from the distribution of \( \sup_{0,1} |B(t)| \). In particular \( w_\alpha \) is such that (Feller 1970)

\[
1 - \alpha = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp \left( \frac{-\left(2k+1\right)^2 \pi^2}{8w_\alpha^2} \right)
\]

(32)

As an example, for \( \alpha=0.02, 0.1 \) and 0.2, the values of \( w_\alpha \) are 2.576, 1.96 and 1.645 respectively.

5. Simulation studies

In this section we study the properties of the sequential procedure by comparing its performances, under \( H_0 \) and \( H_1 \), with those of the LP and PC-tests. More precisely, we compare the tests in terms of the sample size required for achieving a given value of power.

Note that the sequential test is two sided with composite alternative hypothesis \( H_1 : C_{pk} \neq c_{pk,0} \), while the LP and PC-tests are unilateral. In order to correctly compare the performances of the tests, we considered cases under \( H_1 \) where \( C_{pk} = c_{pk,1} \) with \( c_{pk,1} > c_{pk,0} \). In this manner the sequential bilateral test with Type I error probability \( \alpha \) can be compared with the non-sequential unilateral tests with Type I error probability equal to \( \alpha_u = \alpha / 2 \).

To study the properties of the sequential procedure under \( H_1 \), we examined several scenarios (details on how the scenarios were built are reported in Appendix A), which are also discussed in Lepore and Palumbo (2015), where several values of \( C_{pk} \) under \( H_1 \ (c_{pk,1}) \) were considered for the unilateral test with \( \alpha_u =0.01, 0.05, 0.1 \) and \( c_{pk,0} =1.00, 1.33, 1.67 \).

For the LP and PC-tests we determined the minimum sample size, \( n_{\text{LP},0.80} \) and \( n_{\text{PC},0.80} \), using formulas (9) and (13) respectively to achieve a power at least equal to or greater than 0.80: i.e. \( \pi_{\text{LP}}(c_{pk,1}) \geq 0.80 \) and \( \pi_{\text{PC}}(c_{pk,1}) \geq 0.80 \).
As far as the sequential test is concerned we used a set of simulation studies. For each value of \( \alpha \), \( c_{pk,0} \) and \( c_{pk,1} \) we generated, using R (R core team 2013), \( 10^4 \) replicates from a normally distributed quality characteristic. The aim of these simulations was to determine the smallest maximal allowable sample size, \( n_{0,\hat{\pi}_{0.80}} \), which gives an empirical power \( \hat{\pi}_s \) greater than 0.80: i.e. \( \hat{\pi}_s > 0.80 \). The empirical power \( \hat{\pi}_s \) of the sequential test is estimated as the proportion of correctly rejected \( H_0 \).

In order to obtain \( n_{0,\hat{\pi}_{0.80}} \) we implemented an iterative search algorithm with an initial value for \( n_{start} = n_{PC,0.80} \). The algorithm works as follows:

1. With fixed \( n_{start} \) as the maximal allowable sample size of the sequential test, the empirical power of the test \( \hat{\pi}_s \) is estimated as the proportion of correctly rejected \( H_0 \) on \( m = 10^4 \) simulations.

2. If
   \[
   0 < \frac{\hat{\pi}_s - 0.80}{0.80} \leq 0.025
   \]
   then \( n_{0,\hat{\pi}_{0.80}} \) is set equal to \( n_{start} \) and the search algorithm stops. At the same time the average stopping sample size \( n_{avg} \) was empirically assessed as the average of the stopping sample sizes \( n_{stop} \) required by the sequential test to correctly reject \( H_0 \) when the maximal allowable sample size is equal to \( n_{0,\hat{\pi}_{0.80}} \).

3. Otherwise: if \( \hat{\pi}_s \leq 0.80 \), then \( n_{start} = n_{start} + 1 \); if \( \hat{\pi}_s > 0.80 \), then \( n_{start} = n_{start} - 1 \) and the algorithm starts other \( m \) simulations.

The simulation results are summarized in Tables 1-9 where, for each combination of \( \alpha \), \( c_{pk,0} \) and \( c_{pk,1} \), the following quantities are reported: \( n_{LP,0.80} \), the minimum sample size required by the LP-test for achieving a power level \( \geq 0.80 \); \( n_{PC,0.80} \), the minimum sample size required by the PC-test for achieving a power level \( \geq 0.80 \); \( n_{0,\hat{\pi}_{0.80}} \), the smallest maximal allowable sample size for the sequential test for achieving an empirical power \( \hat{\pi}_s > 0.80 \); \( n_{avg} \), the average of the stopping sample sizes \( n_{stop} \) required for the sequential test with maximal allowable sample size \( n_{0,\hat{\pi}_{0.80}} \) for concluding in favor of \( H_1 \); S.D.(\( n_{stop} \)) the standard deviation of the final sample sizes \( n_{stop} \); \( \hat{\pi}_s \), the estimated power of the sequential test.
<table>
<thead>
<tr>
<th>$c_{pk,1}$</th>
<th>$n_{LP,0.80}$</th>
<th>$n_{PC,0.80}$</th>
<th>$n_{0,\hat{z}_r&gt;0.80}$</th>
<th>$n_{avg}$</th>
<th>S.D.($n$ stop)</th>
<th>$\hat{s}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.20</td>
<td>&gt;200</td>
<td>195</td>
<td>182</td>
<td>123.6</td>
<td>31.5</td>
<td>0.807</td>
</tr>
<tr>
<td>1.30</td>
<td>107</td>
<td>94</td>
<td>88</td>
<td>59.4</td>
<td>15.7</td>
<td>0.817</td>
</tr>
<tr>
<td>1.40</td>
<td>66</td>
<td>59</td>
<td>51</td>
<td>34.3</td>
<td>9.7</td>
<td>0.806</td>
</tr>
<tr>
<td>1.50</td>
<td>46</td>
<td>41</td>
<td>35</td>
<td>23.2</td>
<td>7.0</td>
<td>0.810</td>
</tr>
<tr>
<td>1.60</td>
<td>35</td>
<td>32</td>
<td>25</td>
<td>16.4</td>
<td>5.3</td>
<td>0.805</td>
</tr>
<tr>
<td>1.70</td>
<td>28</td>
<td>26</td>
<td>19</td>
<td>12.3</td>
<td>4.3</td>
<td>0.802</td>
</tr>
<tr>
<td>1.80</td>
<td>23</td>
<td>21</td>
<td>16</td>
<td>10.1</td>
<td>3.8</td>
<td>0.820</td>
</tr>
<tr>
<td>1.90</td>
<td>20</td>
<td>18</td>
<td>13</td>
<td>8.2</td>
<td>3.1</td>
<td>0.817</td>
</tr>
<tr>
<td>2.00</td>
<td>17</td>
<td>16</td>
<td>11</td>
<td>6.9</td>
<td>2.7</td>
<td>0.816</td>
</tr>
</tbody>
</table>

Table 1: Simulation results under $H_1$ with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.00$ and $\alpha = 0.02$ ($w_{\alpha} = 2.576$).

<table>
<thead>
<tr>
<th>$c_{pk,1}$</th>
<th>$n_{LP,0.80}$</th>
<th>$n_{PC,0.80}$</th>
<th>$n_{0,\hat{z}_r&gt;0.80}$</th>
<th>$n_{avg}$</th>
<th>S.D.($n$ stop)</th>
<th>$\hat{s}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.20</td>
<td>140</td>
<td>116</td>
<td>116</td>
<td>71.2</td>
<td>23.3</td>
<td>0.804</td>
</tr>
<tr>
<td>1.30</td>
<td>68</td>
<td>57</td>
<td>56</td>
<td>33.6</td>
<td>12.2</td>
<td>0.820</td>
</tr>
<tr>
<td>1.40</td>
<td>42</td>
<td>36</td>
<td>32</td>
<td>18.5</td>
<td>7.6</td>
<td>0.813</td>
</tr>
<tr>
<td>1.50</td>
<td>29</td>
<td>25</td>
<td>21</td>
<td>11.8</td>
<td>5.3</td>
<td>0.816</td>
</tr>
<tr>
<td>1.60</td>
<td>22</td>
<td>20</td>
<td>15</td>
<td>8.2</td>
<td>3.9</td>
<td>0.806</td>
</tr>
<tr>
<td>1.70</td>
<td>18</td>
<td>16</td>
<td>11</td>
<td>6.0</td>
<td>2.9</td>
<td>0.815</td>
</tr>
<tr>
<td>1.80</td>
<td>15</td>
<td>13</td>
<td>9</td>
<td>4.9</td>
<td>2.3</td>
<td>0.812</td>
</tr>
<tr>
<td>1.90</td>
<td>13</td>
<td>12</td>
<td>7</td>
<td>4.0</td>
<td>1.7</td>
<td>0.801</td>
</tr>
<tr>
<td>2.00</td>
<td>11</td>
<td>10</td>
<td>6</td>
<td>3.5</td>
<td>1.4</td>
<td>0.808</td>
</tr>
</tbody>
</table>

Table 2: Simulation results under $H_1$ with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.00$ and $\alpha = 0.1$ ($w_{\alpha} = 1.96$).
<table>
<thead>
<tr>
<th>$c_{pk,1}$</th>
<th>$n_{LP,0.80}$</th>
<th>$n_{PC,0.80}$</th>
<th>$n_{0:z&gt;0.80}$</th>
<th>$n_{\text{avg}}$</th>
<th>S.D.$(n_{\text{stop}})$</th>
<th>$\hat{s}_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.20</td>
<td>104</td>
<td>84</td>
<td>83</td>
<td>46.7</td>
<td>18.6</td>
<td>0.803</td>
</tr>
<tr>
<td>1.30</td>
<td>51</td>
<td>42</td>
<td>40</td>
<td>21.3</td>
<td>9.9</td>
<td>0.811</td>
</tr>
<tr>
<td>1.40</td>
<td>31</td>
<td>26</td>
<td>22</td>
<td>11.2</td>
<td>5.8</td>
<td>0.806</td>
</tr>
<tr>
<td>1.50</td>
<td>22</td>
<td>19</td>
<td>14</td>
<td>7.0</td>
<td>3.7</td>
<td>0.804</td>
</tr>
<tr>
<td>1.60</td>
<td>17</td>
<td>14</td>
<td>10</td>
<td>5.0</td>
<td>2.6</td>
<td>0.810</td>
</tr>
<tr>
<td>1.70</td>
<td>13</td>
<td>12</td>
<td>8</td>
<td>4.1</td>
<td>2.0</td>
<td>0.819</td>
</tr>
<tr>
<td>1.80</td>
<td>11</td>
<td>10</td>
<td>6</td>
<td>3.3</td>
<td>1.4</td>
<td>0.807</td>
</tr>
<tr>
<td>1.90</td>
<td>10</td>
<td>9</td>
<td>5</td>
<td>2.9</td>
<td>1.1</td>
<td>0.800</td>
</tr>
<tr>
<td>2.00</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>2.6</td>
<td>0.8</td>
<td>0.801</td>
</tr>
</tbody>
</table>

Table 3: Simulation results under $H_1$ with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.00$ and $\alpha=0.2$ ($w_\alpha = 1.645$).

<table>
<thead>
<tr>
<th>$c_{pk,1}$</th>
<th>$n_{LP,0.80}$</th>
<th>$n_{PC,0.80}$</th>
<th>$n_{0:z&gt;0.80}$</th>
<th>$n_{\text{avg}}$</th>
<th>S.D.$(n_{\text{stop}})$</th>
<th>$\hat{s}_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.60</td>
<td>199</td>
<td>178</td>
<td>171</td>
<td>116.1</td>
<td>29.8</td>
<td>0.811</td>
</tr>
<tr>
<td>1.70</td>
<td>115</td>
<td>104</td>
<td>96</td>
<td>64.7</td>
<td>17.2</td>
<td>0.815</td>
</tr>
<tr>
<td>1.80</td>
<td>75</td>
<td>68</td>
<td>62</td>
<td>41.7</td>
<td>11.2</td>
<td>0.809</td>
</tr>
<tr>
<td>1.90</td>
<td>55</td>
<td>50</td>
<td>44</td>
<td>29.4</td>
<td>8.4</td>
<td>0.820</td>
</tr>
<tr>
<td>2.00</td>
<td>42</td>
<td>39</td>
<td>33</td>
<td>21.8</td>
<td>6.6</td>
<td>0.814</td>
</tr>
<tr>
<td>2.10</td>
<td>34</td>
<td>32</td>
<td>26</td>
<td>17.1</td>
<td>5.5</td>
<td>0.815</td>
</tr>
<tr>
<td>2.20</td>
<td>29</td>
<td>27</td>
<td>21</td>
<td>13.7</td>
<td>4.7</td>
<td>0.808</td>
</tr>
<tr>
<td>2.30</td>
<td>25</td>
<td>23</td>
<td>17</td>
<td>11.0</td>
<td>3.9</td>
<td>0.807</td>
</tr>
</tbody>
</table>

Table 4: Simulation results under $H_1$ with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.33$ and $\alpha=0.02$ ($w_\alpha = 2.576$).
Table 5: Simulation results under $H_1$ with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = \frac{1}{1.33}$ and $\alpha=0.1$ ($w_{\alpha} = 1.96$).

<table>
<thead>
<tr>
<th>$c_{pk,1}$</th>
<th>$n_{LP:0.80}$</th>
<th>$n_{PC:0.80}$</th>
<th>$n_{0:z_0&gt;0.80}$</th>
<th>$n_{avg}$</th>
<th>S.D.($n_{stop}$)</th>
<th>$\hat{\pi}_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.60</td>
<td>124</td>
<td>108</td>
<td>107</td>
<td>65.8</td>
<td>21.5</td>
<td>0.820</td>
</tr>
<tr>
<td>1.70</td>
<td>70</td>
<td>62</td>
<td>60</td>
<td>36.0</td>
<td>12.7</td>
<td>0.817</td>
</tr>
<tr>
<td>1.80</td>
<td>47</td>
<td>41</td>
<td>39</td>
<td>23.0</td>
<td>8.8</td>
<td>0.819</td>
</tr>
<tr>
<td>1.90</td>
<td>34</td>
<td>30</td>
<td>27</td>
<td>15.5</td>
<td>6.5</td>
<td>0.818</td>
</tr>
<tr>
<td>2.00</td>
<td>27</td>
<td>24</td>
<td>19</td>
<td>10.6</td>
<td>4.9</td>
<td>0.804</td>
</tr>
<tr>
<td>2.10</td>
<td>22</td>
<td>20</td>
<td>15</td>
<td>8.2</td>
<td>3.9</td>
<td>0.807</td>
</tr>
<tr>
<td>2.20</td>
<td>18</td>
<td>17</td>
<td>12</td>
<td>6.5</td>
<td>3.1</td>
<td>0.809</td>
</tr>
<tr>
<td>2.30</td>
<td>16</td>
<td>14</td>
<td>10</td>
<td>5.5</td>
<td>2.6</td>
<td>0.816</td>
</tr>
</tbody>
</table>

Table 6: Simulation results under $H_1$ with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = \frac{1}{1.33}$ and $\alpha=0.2$ ($w_{\alpha} = 1.645$).

<table>
<thead>
<tr>
<th>$c_{pk,1}$</th>
<th>$n_{LP:0.80}$</th>
<th>$n_{PC:0.80}$</th>
<th>$n_{0:z_0&gt;0.80}$</th>
<th>$n_{avg}$</th>
<th>S.D.($n_{stop}$)</th>
<th>$\hat{\pi}_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.60</td>
<td>92</td>
<td>77</td>
<td>77</td>
<td>43.2</td>
<td>17.3</td>
<td>0.808</td>
</tr>
<tr>
<td>1.70</td>
<td>52</td>
<td>44</td>
<td>44</td>
<td>23.7</td>
<td>10.6</td>
<td>0.816</td>
</tr>
<tr>
<td>1.80</td>
<td>35</td>
<td>30</td>
<td>27</td>
<td>14.1</td>
<td>7.0</td>
<td>0.811</td>
</tr>
<tr>
<td>1.90</td>
<td>25</td>
<td>22</td>
<td>19</td>
<td>9.7</td>
<td>5.1</td>
<td>0.816</td>
</tr>
<tr>
<td>2.00</td>
<td>20</td>
<td>18</td>
<td>13</td>
<td>6.5</td>
<td>3.5</td>
<td>0.802</td>
</tr>
<tr>
<td>2.10</td>
<td>16</td>
<td>14</td>
<td>10</td>
<td>5.0</td>
<td>2.6</td>
<td>0.809</td>
</tr>
<tr>
<td>2.20</td>
<td>14</td>
<td>12</td>
<td>8</td>
<td>4.1</td>
<td>2.0</td>
<td>0.804</td>
</tr>
<tr>
<td>2.30</td>
<td>12</td>
<td>11</td>
<td>7</td>
<td>3.7</td>
<td>1.7</td>
<td>0.817</td>
</tr>
<tr>
<td>$c_{pk,1}$</td>
<td>$n_{LP:0.80}$</td>
<td>$n_{PC:0.80}$</td>
<td>$n_{0;\hat{s}&gt;0.80}$</td>
<td>$n_{avg}$</td>
<td>S.D.($n_{stop}$)</td>
<td>$\hat{\pi}_s$</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>---------------</td>
<td>-----------------</td>
<td>----------</td>
<td>----------------</td>
<td>---------</td>
</tr>
<tr>
<td>1.90</td>
<td>&gt;200</td>
<td>&gt;200</td>
<td>&gt;200</td>
<td>&gt;200</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2.00</td>
<td>198</td>
<td>181</td>
<td>173</td>
<td>117.2</td>
<td>29.9</td>
<td>0.816</td>
</tr>
<tr>
<td>2.10</td>
<td>125</td>
<td>115</td>
<td>106</td>
<td>71.6</td>
<td>18.3</td>
<td>0.812</td>
</tr>
<tr>
<td>2.20</td>
<td>88</td>
<td>82</td>
<td>74</td>
<td>49.7</td>
<td>13.4</td>
<td>0.819</td>
</tr>
<tr>
<td>2.30</td>
<td>67</td>
<td>62</td>
<td>53</td>
<td>35.6</td>
<td>9.8</td>
<td>0.804</td>
</tr>
<tr>
<td>2.40</td>
<td>51</td>
<td>47</td>
<td>42</td>
<td>28.0</td>
<td>8.0</td>
<td>0.817</td>
</tr>
<tr>
<td>2.50</td>
<td>42</td>
<td>39</td>
<td>33</td>
<td>22.0</td>
<td>6.6</td>
<td>0.808</td>
</tr>
<tr>
<td>2.60</td>
<td>35</td>
<td>33</td>
<td>27</td>
<td>17.9</td>
<td>5.5</td>
<td>0.803</td>
</tr>
</tbody>
</table>

**Table 7:** Simulation results under $H_1$ with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.67$ and $\alpha = 0.02$ ($w_\alpha = 2.576$).

<table>
<thead>
<tr>
<th>$c_{pk,1}$</th>
<th>$n_{LP,0.80}$</th>
<th>$n_{PC,0.80}$</th>
<th>$n_{0;\hat{s}&gt;0.80}$</th>
<th>$n_{avg}$</th>
<th>S.D.($n_{stop}$)</th>
<th>$\hat{\pi}_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.90</td>
<td>&gt;200</td>
<td>&gt;200</td>
<td>208</td>
<td>129.3</td>
<td>40.7</td>
<td>0.802</td>
</tr>
<tr>
<td>2.00</td>
<td>123</td>
<td>109</td>
<td>106</td>
<td>65.2</td>
<td>21.4</td>
<td>0.809</td>
</tr>
<tr>
<td>2.10</td>
<td>78</td>
<td>69</td>
<td>65</td>
<td>39.4</td>
<td>13.6</td>
<td>0.800</td>
</tr>
<tr>
<td>2.20</td>
<td>53</td>
<td>48</td>
<td>45</td>
<td>27.0</td>
<td>10.0</td>
<td>0.813</td>
</tr>
<tr>
<td>2.30</td>
<td>40</td>
<td>36</td>
<td>32</td>
<td>18.7</td>
<td>7.6</td>
<td>0.806</td>
</tr>
<tr>
<td>2.40</td>
<td>32</td>
<td>29</td>
<td>25</td>
<td>14.4</td>
<td>6.1</td>
<td>0.807</td>
</tr>
<tr>
<td>2.50</td>
<td>26</td>
<td>24</td>
<td>19</td>
<td>10.7</td>
<td>4.9</td>
<td>0.806</td>
</tr>
<tr>
<td>2.60</td>
<td>22</td>
<td>20</td>
<td>16</td>
<td>8.8</td>
<td>4.1</td>
<td>0.812</td>
</tr>
</tbody>
</table>

**Table 8:** Simulation results under $H_1$ with $C_{pk} = c_{pk,1}$, when $c_{pk,0} = 1.67$ and $\alpha = 0.1$ ($w_\alpha = 1.96$).
To study the properties of the sequential procedure under $H_0 : C_{pk} = C_{pk,0}$, for each combination of $C_{pk,0} (1.00, 1.33, 1.67)$ and $\alpha (0.02, 0.1, 0.2)$, we generated $10^4$ replicates from a normally distributed process. The aim of these simulations was to determine the smallest maximal allowable sample size, $n_{0:1-\hat{\alpha} > 1-\alpha}$, which gives an empirical type I error probability $\hat{\alpha}$ smaller than the nominal $\alpha$ value: $\hat{\alpha} < \alpha$ or equivalently $1-\hat{\alpha} > 1-\alpha$.

In order to obtain $n_{0:1-\hat{\alpha} > 1-\alpha}$ we implemented an iterative search algorithm with initial value for $n_{0:1-\hat{\alpha} > 1-\alpha}$ given by $n_{\text{start}} = 75$. The algorithm works as follows:

1. With fixed $n_{\text{start}}$ as the maximal allowable sample size of the sequential test, the value of $1-\hat{\alpha}$ is estimated as the fraction of correctly accepted $H_0$ on $m = 10^4$ simulations.

2. If

$$0 < \frac{(1-\hat{\alpha})-(1-\alpha)}{(1-\alpha)} \leq 0.025$$

then $n_{0:1-\hat{\alpha} > 1-\alpha}$ is set equal to $n_{\text{start}}$ and the search algorithm stops. Contextually the average stopping sample size under $H_0$, $n_{H_0,\text{avg}}$, is empirically assessed as the average of the final sample sizes $n_{\text{stop}}$ required by the sequential test to correctly accept $H_0$ when the maximal allowable sample size is equal to $n_{0:1-\hat{\alpha} > 1-\alpha}$.

<table>
<thead>
<tr>
<th>$c_{pk,1}$</th>
<th>$n_{LP,0.80}$</th>
<th>$n_{PC,0.80}$</th>
<th>$n_{0:1-\hat{\alpha} &gt; 0.80}$</th>
<th>$n_{\text{avg}}$</th>
<th>S.D.($n_{\text{stop}}$)</th>
<th>$\hat{\alpha}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.90</td>
<td>175</td>
<td>151</td>
<td>154</td>
<td>89.3</td>
<td>32.8</td>
<td>0.806</td>
</tr>
<tr>
<td>2.00</td>
<td>91</td>
<td>79</td>
<td>79</td>
<td>44.5</td>
<td>17.6</td>
<td>0.812</td>
</tr>
<tr>
<td>2.10</td>
<td>56</td>
<td>49</td>
<td>49</td>
<td>26.8</td>
<td>11.6</td>
<td>0.814</td>
</tr>
<tr>
<td>2.20</td>
<td>39</td>
<td>35</td>
<td>32</td>
<td>16.8</td>
<td>8.2</td>
<td>0.811</td>
</tr>
<tr>
<td>2.30</td>
<td>30</td>
<td>26</td>
<td>23</td>
<td>11.9</td>
<td>6.1</td>
<td>0.813</td>
</tr>
<tr>
<td>2.40</td>
<td>24</td>
<td>21</td>
<td>17</td>
<td>8.6</td>
<td>4.6</td>
<td>0.812</td>
</tr>
<tr>
<td>2.50</td>
<td>19</td>
<td>17</td>
<td>14</td>
<td>7.0</td>
<td>3.7</td>
<td>0.818</td>
</tr>
<tr>
<td>2.60</td>
<td>16</td>
<td>15</td>
<td>11</td>
<td>5.5</td>
<td>2.9</td>
<td>0.817</td>
</tr>
</tbody>
</table>

**Table 9:** Simulation results under $H_1$ with $C_{pk} = C_{pk,1}$, when $c_{pk,0} = 1.67$ and $\alpha=0.2$ ($w_\alpha = 1.645$).
3. Otherwise: if \((1 - \hat{\alpha}) \leq (1 - \alpha)\), then \(n_{\text{start}} = n_{\text{start}} + 1\); if \((1 - \hat{\alpha}) > (1 - \alpha)\), then \(n_{\text{start}} = n_{\text{start}} - 1\) and the algorithm starts other \(m\) simulations.

The simulation results are summarized in Tables 10-12, where for each combination of \(\alpha\) and \(c_{pk,0}\) the following quantities are reported: \(n_{0\leq \hat{\alpha}>1-\alpha}\) the smallest maximal allowable sample size for the sequential test for achieving an empirical type I error probability \(\hat{\alpha}\) smaller than the nominal \(\alpha\) \((\hat{\alpha} < \alpha)\); \(n_{H_{\text{avg}}}\) the average of the final sample sizes \(n_{\text{stop}}\) required for the sequential test with maximal sample size \(n_{0\leq \hat{\alpha}>1-\alpha}\) for correctly concluding in favor of \(H_0\); \(\text{S.D.}(n_{\text{stop}})\) the standard deviation of the final sample sizes \(n_{\text{stop}}\); \(\hat{\alpha}\) the estimated type I error probability; 1 - \(\hat{\alpha}\).

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>(n_{0\leq \hat{\alpha}&gt;1-\alpha})</th>
<th>(n_{H_{\text{avg}}})</th>
<th>(\text{S.D.}(n_{\text{stop}}))</th>
<th>(\hat{\alpha})</th>
<th>1 - (\hat{\alpha})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>127</td>
<td>100.9</td>
<td>21.9</td>
<td>0.0191</td>
<td>0.9809</td>
</tr>
<tr>
<td>0.1</td>
<td>82</td>
<td>56.8</td>
<td>19.5</td>
<td>0.099</td>
<td>0.9010</td>
</tr>
<tr>
<td>0.2</td>
<td>62</td>
<td>37.7</td>
<td>16.5</td>
<td>0.1949</td>
<td>0.8051</td>
</tr>
</tbody>
</table>

**Table 10:** Simulation results under \(H_0: C_{pk} = c_{pk,0}\) with \(c_{pk,0} = 1.00\).

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>(n_{0\leq \hat{\alpha}&gt;1-\alpha})</th>
<th>(n_{H_{\text{avg}}})</th>
<th>(\text{S.D.}(n_{\text{stop}}))</th>
<th>(\hat{\alpha})</th>
<th>1 - (\hat{\alpha})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>110</td>
<td>86.8</td>
<td>19.3</td>
<td>0.0195</td>
<td>0.9805</td>
</tr>
<tr>
<td>0.1</td>
<td>67</td>
<td>44.6</td>
<td>16.6</td>
<td>0.0989</td>
<td>0.9011</td>
</tr>
<tr>
<td>0.2</td>
<td>69</td>
<td>43.0</td>
<td>18.1</td>
<td>0.1958</td>
<td>0.8042</td>
</tr>
</tbody>
</table>

**Table 11:** Simulation results under \(H_0: C_{pk} = c_{pk,0}\) with \(c_{pk,0} = 1.33\).

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>(n_{0\leq \hat{\alpha}&gt;1-\alpha})</th>
<th>(n_{H_{\text{avg}}})</th>
<th>(\text{S.D.}(n_{\text{stop}}))</th>
<th>(\hat{\alpha})</th>
<th>1 - (\hat{\alpha})</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>90</td>
<td>69.2</td>
<td>17.6</td>
<td>0.0198</td>
<td>0.9802</td>
</tr>
<tr>
<td>0.1</td>
<td>61</td>
<td>41.2</td>
<td>14.6</td>
<td>0.0982</td>
<td>0.9018</td>
</tr>
<tr>
<td>0.2</td>
<td>74</td>
<td>47.0</td>
<td>19.0</td>
<td>0.1978</td>
<td>0.8022</td>
</tr>
</tbody>
</table>

**Table 12:** Simulation results under \(H_0: C_{pk} = c_{pk,0}\) with \(c_{pk,0} = 1.67\).
6. Discussion

As far as the behavior of the test under $H_1$ is concerned, by examining the averages of the final sample sizes $n_{avg}$ the results show that the sequential test, with the same power of the LP and PC-tests, saves a lot of sample size. Furthermore, the maximum allowable sample size $n_{0.80}$ required to achieve the desired power is almost always smaller than $n_{LP,0.80}$ and $n_{PC,0.80}$. This indicates that even in the worst cases the sequential test needs a maximum allowable sample size not greater than the sample size of the non-sequential tests.

As an example, under $H_1: C_{pk} = c_{pk,1}$ with $c_{pk,1} = 1.30$, when $c_{pk,0} = 1.00$ and $\alpha=0.02$, we have $n_{LP,0.80} = 107$, $n_{PC,0.80} = 94$ (Table 1 and Figure 1), while with a maximum allowable sample size equal to $n_{0.80} = 88$ the power of the sequential test is $\hat{\pi} > 0.80$ with an $n_{avg} = 59.4(\approx 60)$.

![Power functions of the LP and PC tests as a function of the sample size for the case $c_{pk,0} = 1.00$, $\alpha = 0.02(\alpha_u = 0.01)$ and $c_{pk,1} = 1.30$](image)

**Figure 1**: Power functions of the LP and PC tests as a function of the sample size for the case $c_{pk,0} = 1.00$, $\alpha = 0.02(\alpha_u = 0.01)$ and $c_{pk,1} = 1.30$

In this case the sequential procedure saves, on average, 43.9% of the sample size as to the LP-test and 36.2% as to the PC-test.

18
Under $H_1: C_{pk} = c_{pk,1}$ with $c_{pk,1} = 1.60$, when $c_{pk,0} = 1.33$ and $\alpha = 0.1$, we have $n_{LP,0.80} = 124$ and $n_{PC,0.80} = 108$ (Figure 2 and Table 5) while with a maximum allowable sample size equal to $n_{0, \hat{\pi}, >0.80} = 107$ the power of the sequential test is $\hat{\pi}_s > 0.80$ with an $n_{avg} = 65.8 (\approx 66)$.

![Figure 2: Power functions of the LP and PC tests as a function of the sample size for the case $c_{pk,0} = 1.33, \alpha = 0.1 (\alpha_u = 0.05)$ and $c_{pk,1} = 1.60$](image)

In this case the sequential procedure therefore saves, on average, 46.8% of the sample size as to the LP test and 38.9% as to the PC test.

Finally, under $H_1: C_{pk} = c_{pk,1}$ with $c_{pk,1} = 1.90$, when $c_{pk,0} = 1.67$ and $\alpha = 0.2$, we have $n_{LP,0.80} = 175$ and $n_{PC,0.80} = 151$, (Figure 3 and Table 9) while with a maximum allowable sample size equal to $n_{0, \hat{\pi}, >0.80} = 154$ the power of the sequential test is $\hat{\pi}_s > 0.80$ with an $n_{avg} = 89.3 (\approx 90)$. 
Figure 3: Power functions of the LP and PC tests as a function of the sample size for the case $c_{pk,0} = 1.67$, $\alpha = 0.2 (\alpha_u = 0.1)$ and $c_{pk,1} = 1.90$.

In this case the sequential procedure saves, on average, 48.6% of the sample size as to the LP-test and 40.4% as to the PC-test.

Also under $H_0$ the proposed sequential test displays features that are very interesting. For each value of $\alpha$ and $c_{pk,0}$ the smallest maximal allowable sample sizes $n_{0,\hat{\alpha}>1-\alpha}$, required by the sequential test for ensuring an empirical type I error probability $\hat{\alpha}$ smaller than the nominal value $\alpha$ are always smaller than the values of $n_{0,\hat{\alpha}>0.80}$ corresponding to the worst scenario examined under $H_1$: i.e. the value of $c_{pk,1}$ closer to $c_{pk,0}$.

Let us consider the following cases:

1. when $H_0: C_{pk} = c_{pk,0}$ holds, with $c_{pk,0} = 1.00$ and $\alpha = 0.02$, we have that $n_{0,1-\hat{\alpha}>1-\alpha} = 127$ (Table 10) while for $c_{pk,1} = 1.20$ it results that $n_{0,\hat{\alpha}>0.80} = 182$ (Table 1).

2. when $H_0: C_{pk} = c_{pk,0}$ holds, with $c_{pk,0} = 1.33$ and $\alpha = 0.2$, we have that $n_{0,1-\hat{\alpha}>1-\alpha} = 69$ (Table 11) while for $c_{pk,1} = 1.60$ it results that $n_{0,\hat{\alpha}>0.80} = 77$ (Table 6).

3. when $H_0: C_{pk} = c_{pk,0}$ holds, with $c_{pk,0} = 1.67$ and $\alpha = 0.1$, we have that $n_{0,1-\hat{\alpha}>1-\alpha} = 61$ (Table 12) while for $c_{pk,1} = 1.90$ it results that $n_{0,\hat{\alpha}>0.80} = 208$ (Table 8).
These results ensure that for all the cases examined the empirical type I error probability does not exceed the nominal $\alpha$-level of the test. Furthermore, by examining $n_{H_0,avg}$, the averages of the final sample sizes $n$ required for the sequential test for correctly concluding in favor of $H_0$, we can assert that the sequential procedure allow early stopping sample sizes also in those cases where the process is not capable.

7. Conclusions

In this article we proposed a sequential procedure for hypothesis testing on the $C_{pk}$ index. We studied the statistical properties of the sequential test with an extensive simulation study with regard to the type I error, the average of the sample sizes for correctly deciding, for $H_0$ and $H_1$, the maximum allowable sample size required to achieve a pre-set power level and for ensuring that the empirical type I error probability does not exceed the nominal $\alpha$-level of the test. We compared the performances of the sequential procedure with two non-sequential tests. The results showed that the sequential test allows on average smaller stopping sample sizes as compared with the fixed sample size tests while maintaining the desired $\alpha$-level and power. Furthermore, the maximum allowable sample sizes required by the sequential test to achieve the desired power level are smaller than, or at most equal to, the sample sizes required by the non-sequential tests: this means that even in the worst cases the sequential procedure uses a sample size that does not exceed the sample size of the non-sequential tests with the same power level (under $H_1$) or without exceeding the type I error probability (under $H_0$).

Summarizing, the proposed sequential procedure has several interesting features: it offers a substantial decrease in sample size compared with the non-sequential tests, while type I and II error probabilities are correctly maintained at their desired values. We consider these results as valuable, because in a highly competitive context where both cost and quality are relevant, the availability of statistical methods which make it possible to save sampling size can be directly translated into saved resources and reduced costs. Furthermore, process capability analysis is increasingly used in healthcare-related studies (Chen et al. 2014, Liu et al. 2010) where, in addition to the economic matter, important ethical issues must be taken into account. It is worth noting that within this particular framework, methods capable of shortening the time span or reducing the sample size required for testing process capability can be of great value.
Appendix A

To explain how the scenarios for our study have been built let us consider, without loss of any generality, the case where $c_{pk,0} = 1.33$.

We considered, under $H_0$, a process where: $USL = 25$, $LSL = 15$, $\mu = 23$ and $\sigma = 0.501253$ in such a way that $C_{pk} = c_{pk,0} = 1.33$.

Under $H_1$ we examined cases where $c_{pk,1}$ ranged from 1.60 to 2.30 (see e.g. Table 4). Therefore the alternative hypothesis consists of scenarios where the capability was improving: these scenarios can be obtained by improving process centering or by reducing the process variability.

As an example, the case $c_{pk,1} = 1.60$ can be obtained with a mean level closer to the center, $\mu = 22.59398$, with fixed $\sigma = 0.501253$ or by reducing the variability, $\sigma = 0.416667$, with fixed $\mu = 23$.

For the sake of completeness in the simulations, we initially examined the scenarios for both the possible situations: an improvement in process centering and a reduction in process variability. However, the results are very similar therefore in the paper we reported only the results concerning the improvement in the process centering. As an example, in Table A1 for the case $c_{pk,0} = 1.33$ and $\alpha = 0.02$ (the same as Table 4), the following are reported: $n_{0,\hat{t}_{s}>0.80}$ the smallest maximal allowable sample size for the sequential test for achieving an empirical power $\hat{t}_{s} > 0.80$; $n_{\text{avg,}\mu}$ the average of the stopping sample sizes $n_{\text{stop,}\mu}$ required for the sequential test with maximal allowable sample size $n_{0,\hat{t}_{s}>0.80}$ for concluding in favor of $H_1$ when process centering is improved; $\text{S.D.}(n_{\text{stop,}\mu})$ the standard deviation of the final sample sizes $n_{\text{stop,}\mu}$; $\hat{t}_{s;\mu}$ the estimated power of the sequential test when process centering is improved; $n_{\text{avg,}\sigma}$ the average of the stopping sample sizes $n_{\text{stop,}\sigma}$ required for the sequential test with maximal allowable sample size $n_{0,\hat{t}_{s}>0.80}$ for concluding in favor of $H_1$ when the process variability is reduced; $\text{S.D.}(n_{\text{stop,}\sigma})$ the standard deviation of the final sample sizes $n_{\text{stop,}\sigma}$; $\hat{t}_{s;\sigma}$ the estimated power of the sequential test when the process variability is reduced $\hat{t}_{s;\sigma}$.
Table A1: Results of the simulations for the case $c_{pk,0} = 1.33$, $\alpha=0.02$ ($w_\alpha = 2.576$) obtained by improving process centering and with a reduction in process variability

It can be noted that the results of the two cases (improving the process centering vs reduction of the process variability) are very similar. For this reason in the paper we focused only on the scenarios concerning the improvement in process centering.

References


