A review and comparison of continuity correction rules; the normal approximation to the binomial distribution

Presenter: Yu-Ting Liao (廖昱婷)

Advisor: Takeshi Emura (江村剛志) 博士

June 27, 2016

Graduate Institute of Statistics, NCU

Outline

- Introduction
- Normal approximation to binomial
- Application to statistical process control
- Numerical studies
- Conclusion

Introduction

- Continuity correction
  - Yates (1934) → adding 0.5
  - Cressie (1978) → improve

compare

- Choose the optimal correction value by Taylor expansions
  - Cressie (1978)
  - Fuh et al. (2015)

Basic probability theory

$$X \sim Bin(n, p)$$

n : number of trails.

p : probability of the event.

- pmf

$$p_X(k) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- cdf

$$F_X(k) = \Pr(X \le k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$

k : succesess in n trials

Normal approximate to binomial

$$X = \sum_{j=1}^{n} X_j, X_j \overset{iid}{\sim} Ber(p)$$

- mgf 's of X_j

$$M_{X_j}(t) = E(e^{tX_j}) = \sum_{x=0}^{1} e^{tx} p^x (1-p)^{1-x} = pe^t + (1-p).$$

By the Central Limit Theorem (CLT)

- cdf $F_X(k) = \Pr(X \le k)$

$$= \Pr\left(\frac{X - np}{\sqrt{np(1-p)}} \le \frac{k - np}{\sqrt{np(1-p)}} \right) \approx \Phi\left(\frac{k - np}{\sqrt{np(1-p)}} \right),$$

$$\text{where } \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.$$

Continuity correction

- The continuity correction

$$F_X(k) \approx \Phi\left(\frac{k - np + d}{\sqrt{np(1-p)}}\right), \quad k \in \{0, 1, \cdots, n\}.$$

- The absolute error

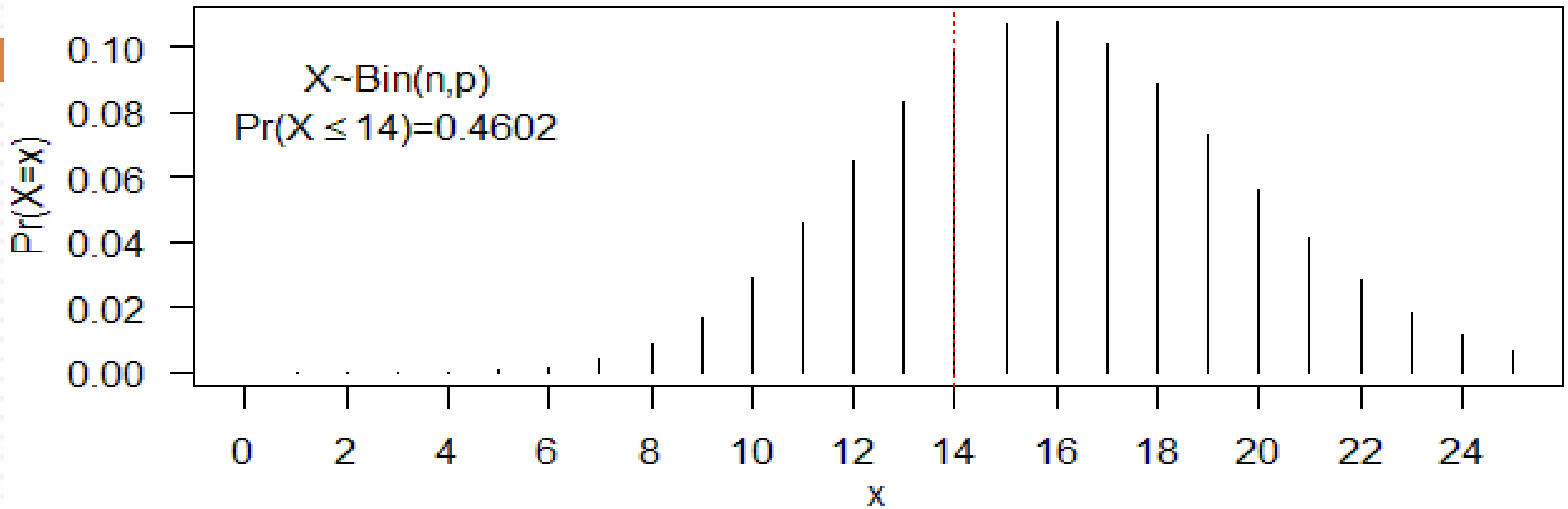
$$Err(n, p, k) = \left| F_X(k) - \Phi\left(\frac{k - np + d}{\sqrt{np(1-p)}}\right) \right|$$

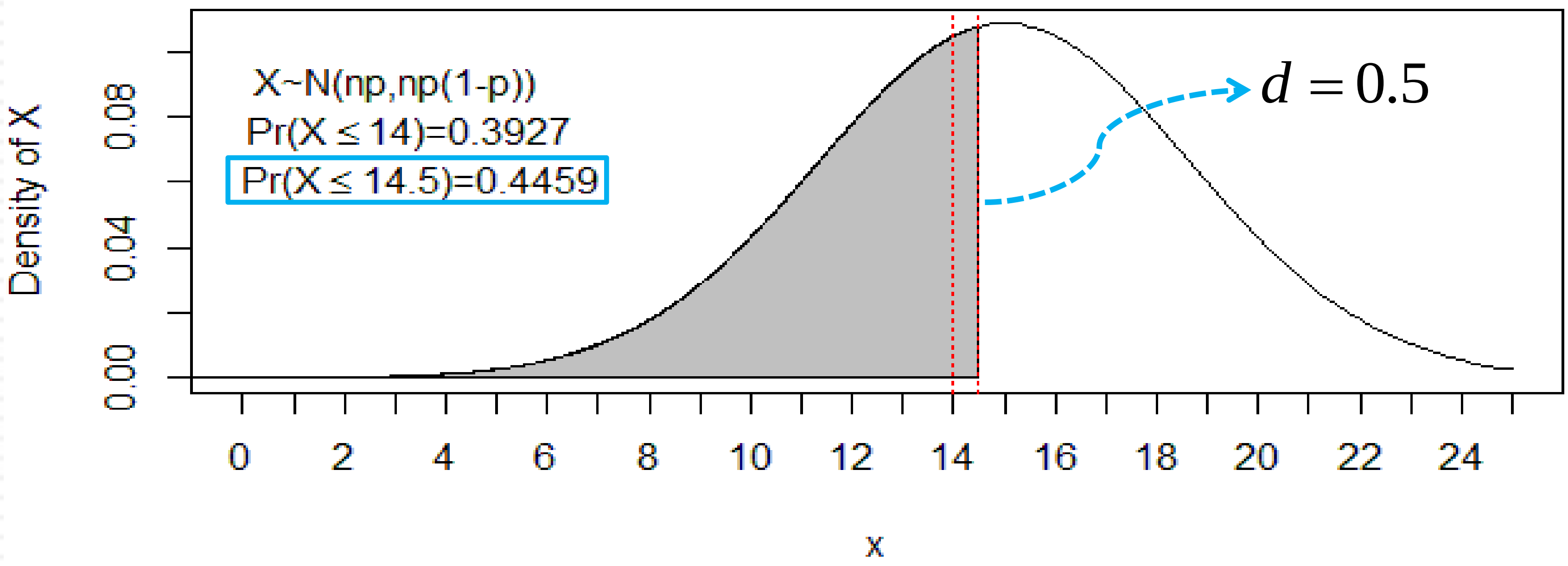
  - Yates (1934): $d = 0.5$
  - Cressie (1978): $d(k, p) = 0.5 - (q - p)(\delta^2_{k+0.5} - 1)/6,$

    where $\delta_k = (k - np)/\sqrt{np(1-p)}.$
  - $d = 0.3$

$X \sim Bin(150, 0.1)$ and $k = 14$

Success in *n* trails





np -control chart

$X \sim Bin(n, p),$

X : number of nonconforming items

p : fraction nonconforming

n : sample size

- Center line $= np,$

- $UCL(\text{upper control limit}) = np + 3\sqrt{np(1-p)}.$

np -control chart

- The in-control probability

$$P^* = \Pr(X \le [UCL]) = \sum_{x \le [UCL]} \binom{n}{x} p^x (1-p)^{n-x},$$

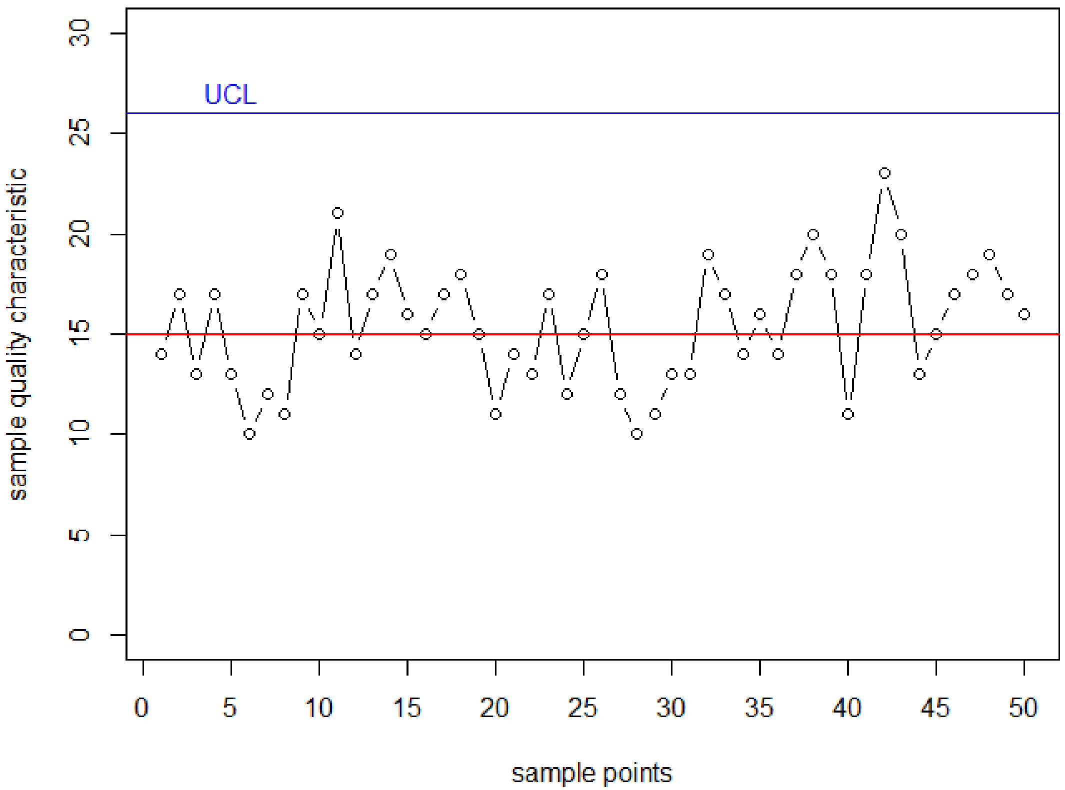
- Approximate the in-control probability

$$P^* = \Pr\left(\frac{X - np}{\sqrt{np(1-p)}} \le \frac{[UCL] - np}{\sqrt{np(1-p)}} \right) \approx \Phi\left(\frac{[UCL] - np}{\sqrt{np(1-p)}} \right).$$

where $[UCL]$ is the largest integer not greater than UCL.

Example for *np* -control chart

$$X \sim Bin(150, 0.1)$$



Process is in-control

$n = 150, p = 0.1$

Figure 2 *np*-control chart based on 50 replications under

and $P^* = 0.9981, \; UCL = 26.0227, \; [UCL] = 26$ and center line =15

Application of continuity correction

- The absolute error of the in-control probability

$$Err(n, p, [UCL]) = \left| P^* - \Phi\left(\frac{[UCL] - np + d}{\sqrt{np(1-p)}} \right) \right|$$

- Yates (1934): $d = 0.5$
- Cressie (1978): $d(k, p) = 0.5 - (q - p)(\delta^2_{k+0.5} - 1)/6,$

where $\delta_k = (k - np)/\sqrt{np(1-p)}.$

- $d = 0.3$

Example

$X \sim Bin(150, 0.1)$

	Example ($n=150$, $p=0.1$)
np	15
$[UCL]$	26
P^*	0.9981
$d=0.5$	$Err(n, p, [UCL]) = 0.0010$
$d=0.3$	$Err(n, p, [UCL]) = 0.0009$
$d=d(p,[UCL])$	$Err(n, p, [UCL]) = 0.0001$

Note: $d=0.5$ (Yates,1934; Cox, 1970), $d=d(p,[UCL])$ (Cressie, 1978), where

$d=0.5-(q-p)(\delta^2_{[UCL]+0.5}-1)/6$ and $q=1-p$.

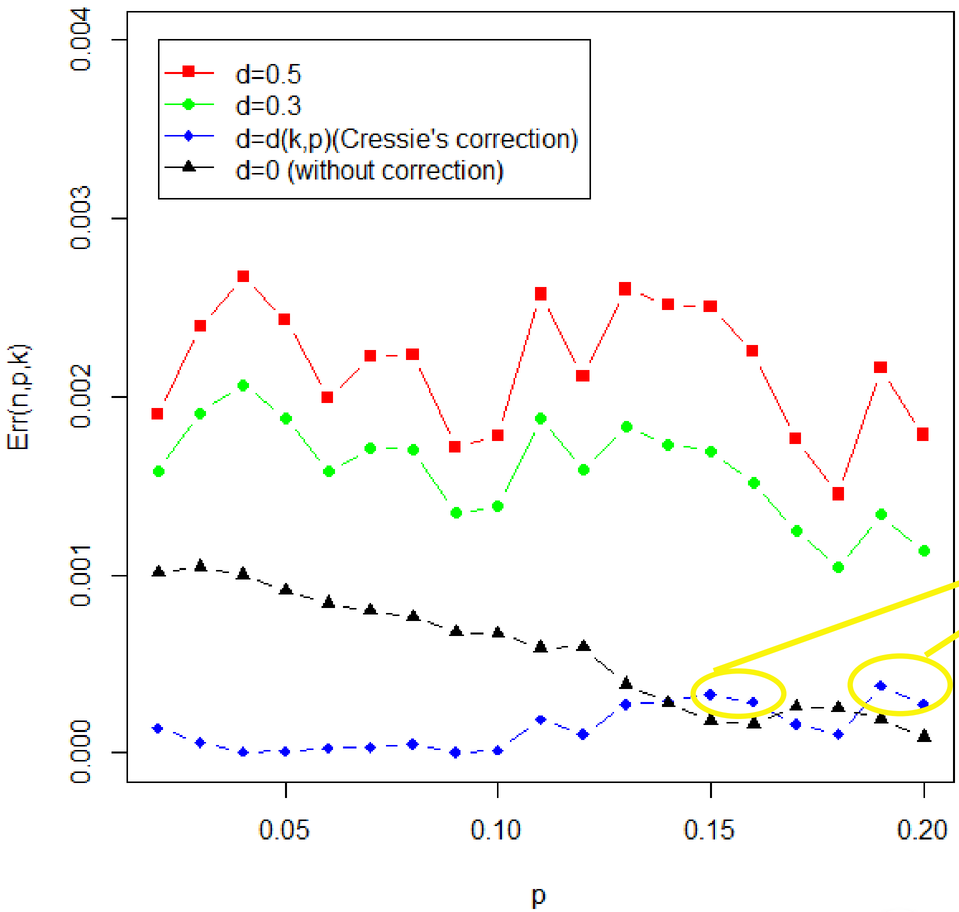
Choose 19 pairs of n, p and k

We choose 19 different pairs

$$(n,p) = (800,0.02), (550,0.03), (400,0.04), (350,0.05), (300,0.06),$$
$$(300,0.06), (250,0.07), (200,0.08), (200,0.09), (150,0.1),$$
$$(100,0.11), (90,0.12), (80,0.13), (80,0.14), (70,0.15),$$
$$(70,0.16), (60,0.17), (60,0.18), (60,0.19), (60,0.20).$$

- $np > 15$ or $np > 10$ and $p \geq 0.1$ (Emura and Lin 2015).
- $k = [\, np + \Phi^{-1}(\alpha)\sqrt{np(1-p)}\,]$
  - $\alpha = 0.9973, 0.0027, 0.95$ and 0.05.

Comparison absolute error of normal approximation

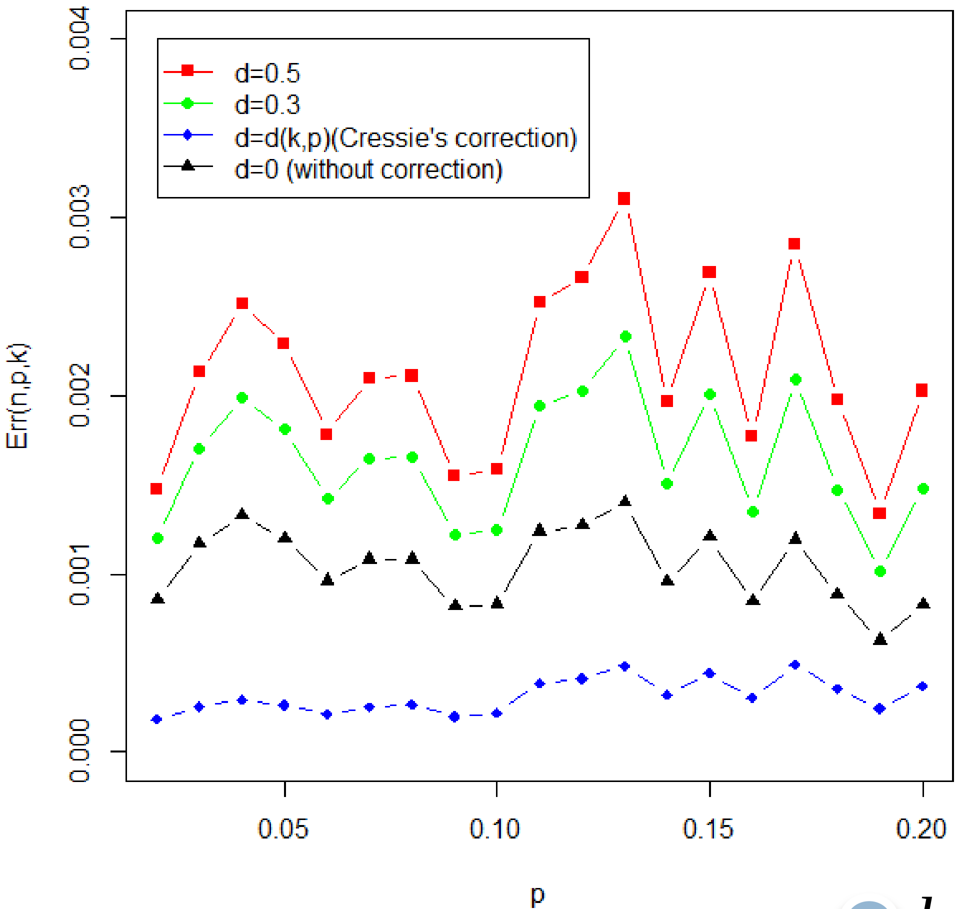


$(n, p) = (70, 0.15), (70, 0.16),$

$(60, 0.19), (60, 0.2).$

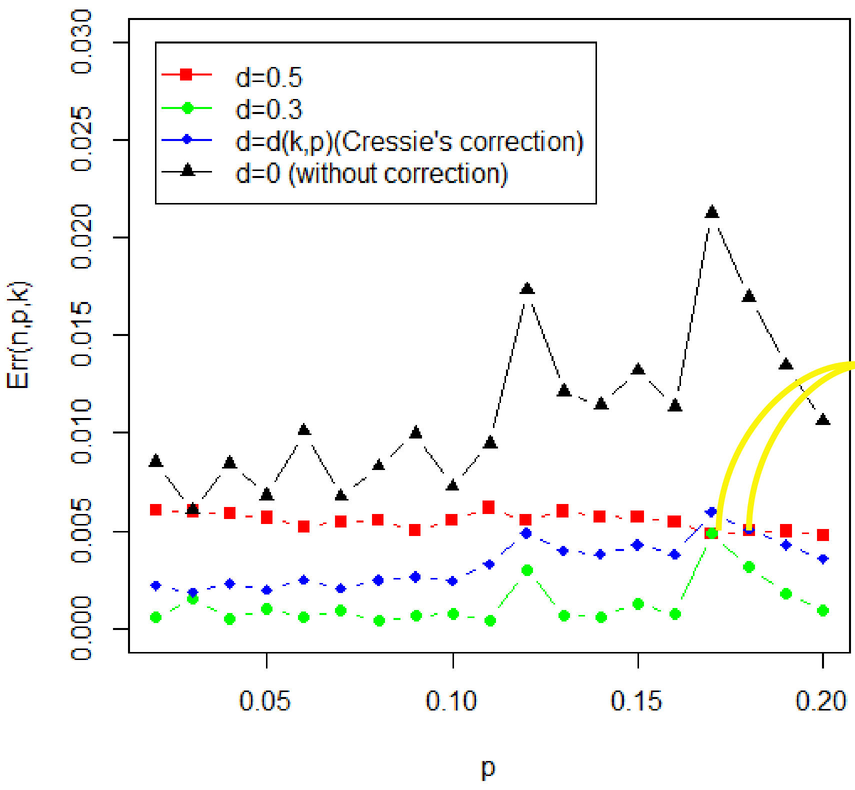
- $k = [np + \Phi^{-1}(0.9973)\sqrt{np(1-p)}]$

Comparison absolute error of normal approximation



- $k = [np + \Phi^{-1}(0.0027)\sqrt{np(1-p)}]$

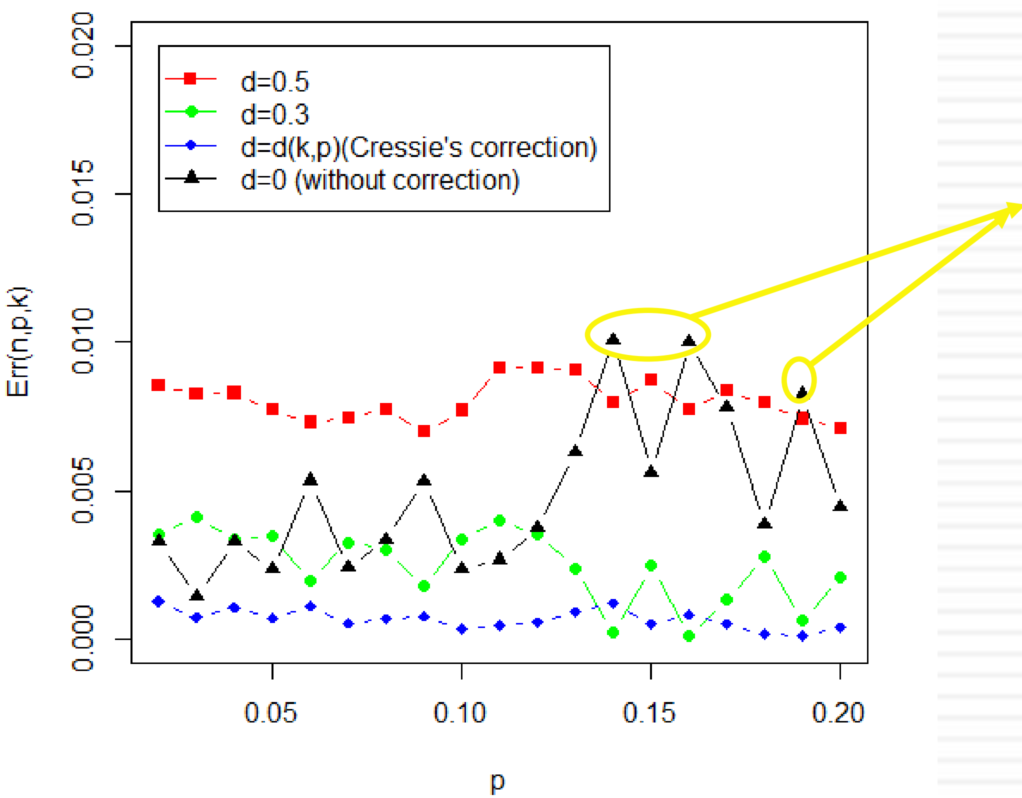
Comparison absolute error of normal approximation



$(n, p) = (60, 0.17), (60, 0.18)$

- $k = [np + \Phi^{-1}(0.95)\sqrt{np(1-p)}]$

Comparison absolute error of normal approximation



$(n, p) = (80, 0.14), (70, 0.16),$
$(60, 0.19)$

- $k = [np + \Phi^{-1}(0.05)\sqrt{np(1-p)}]$

Comparison absolute error of in-control probability

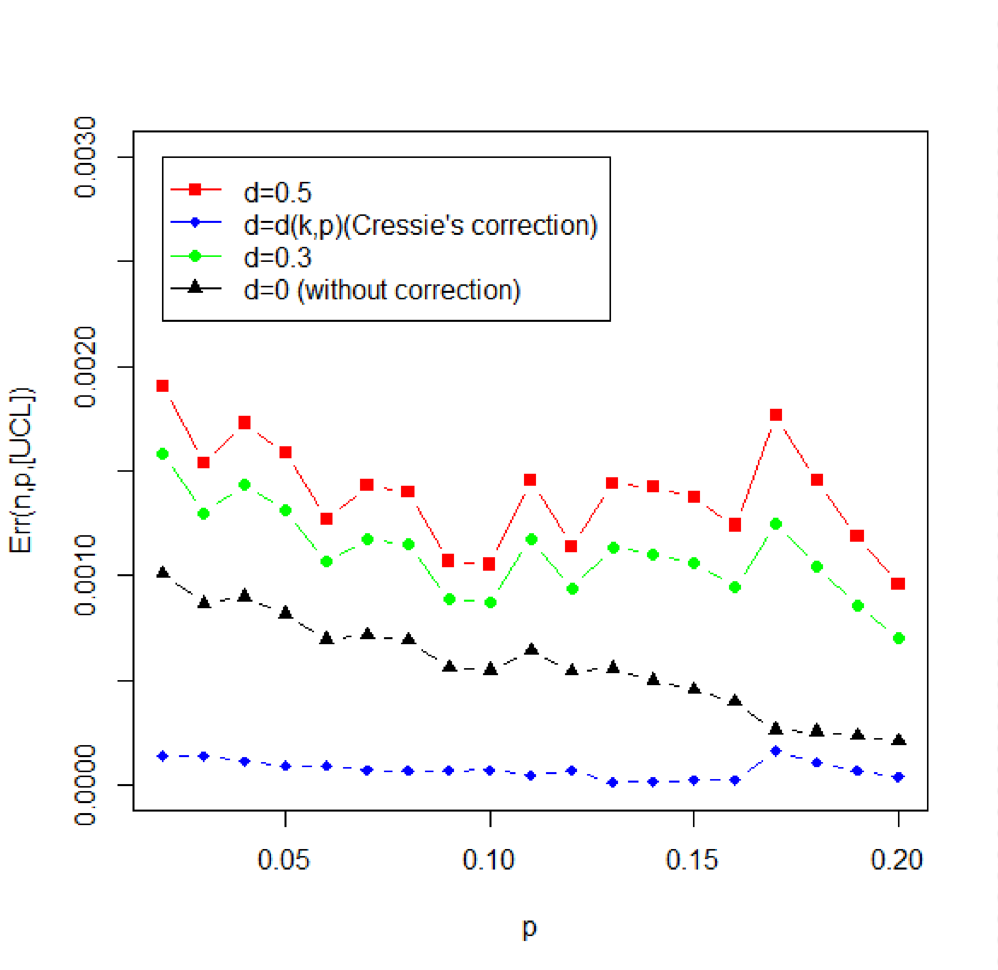


Figure 3 Comparison of the absolute error of continuity corrections

Data example

Table 1 The analysis of two datasets in Chapter7 of Montgomery (2009).

	Data 1 Example 7.1 ($n = 30,\ p = 0.2313$)	Data 2 Exercise 7.2 ($n = 150,\ p = 0.0025$)
np	6.9400	0.375
$[UCL]$	13	2
P^*	0.9961	0.9935
$d = 0.5$	$Err(n, p, [UCL]) = 0.0017$	$Err(n, p, [UCL]) = 0.0062$
$d = 0.3$	$Err(n, p, [UCL]) = 0.0010$	$Err(n, p, [UCL]) = 0.0057$
$d = d(p, [UCL])$	$Err(n, p, [UCL]) = 0.0001$	$Err(n, p, [UCL]) = 0.0042$

Note: $d = 0.5$ (Yates,1934; Cox, 1970), $d = d(p,[UCL])$ (Cressie, 1978), where

$d = 0.5 - (q - p)(\delta^2_{[UCL]+1/2} - 1)/6$ and $q = 1 - p$.

Conclusion

- Cressie's finely continuity correction performs best among all other corrections.

- Cressis's finely tuned continuity correction can be reliably applied for the problem of statistical process control.

Thank you for your listening.