The index of coincidence is defined as

$$I_c = \frac{\text{number of pairs of equal letters in ciphertext}}{\text{the total number of pairs of letters}}$$

That is if we set

- N_α = the number of occurrences of the letter α in the cyphertext
- $$D_c = \sum_{\alpha=A}^{Z} \binom{N_\alpha}{2}$$

  D_c represents the number of pairs of equal letters in the cyphertext.
- then $I_c = \frac{D_c}{\binom{N}{2}}$
- where N = the number of letters in the cyphertext

The index of coincidence is invariant under monoalphabetic cyphers and we estimate under this condition that $N_\alpha = N * p_{\sigma(\alpha)}$ for some permutation of the alphabet σ and so

$$\begin{aligned} I_c &= \frac{\sum_{\alpha=A}^{Z}(N_\alpha^2 - N_\alpha)}{N(N-1)} \\ &\approx \frac{N^2(\sum_{\alpha=A}^{Z} p_\alpha^2) - N}{N(N-1)} \\ &= \frac{N(.065) - 1}{N-1} \\ &\approx .065 \end{aligned}$$

If the cyphertext was obtained from a polyalphabetic cipher then the index of coincidence can also be used to estimate the period of the cipher.
Let p be the period of the cyphertext and place the letters of the cyphertext into groups of p so that the letters in the i^{th} position of the groups are all encrypted with the same key.

- Let $M_\alpha^{(i)}$ equal the number of occurrences of the letter α that appears in the i^{th} positions in the groups.
- If there are M groups of p, then $\sum_{\alpha=A}^{Z} M_\alpha^{(i)} = M$
- We also have $N = Mp$
- Also we can estimate that $M_\alpha^{(i)} \approx Mp_{\sigma(\alpha)}$ (again for some permutation for the alphabet σ)

Now, we calculate that

$$
\begin{aligned}
2D_c &= \sum_{i=1}^{p} \sum_{\alpha=A}^{Z} M_\alpha^{(i)} (M_\alpha^{(i)} - 1) + \sum_{i=1}^{p} \sum_{j=i+1}^{p} \sum_{\alpha=A}^{Z} M_\alpha^{(i)} M_\alpha^{(j)} \\
&\approx M^2 p(.065) - pM + M^2(.038)p(p-1) \\
&= \frac{N^2}{p}(.027) - N + N^2(.038)
\end{aligned}
$$

Note that because $I_c = \frac{D_c}{\binom{N}{2}}$, we have that

$$2D_c = N(N-1)I_c.$$

And we just derived that

$$2D_c \approx \frac{N^2}{p}(.027) - N + N^2(.038)$$

Therefore,

$$N(N-1)I_c \approx \frac{N^2}{p}(.027) - N + N^2(.038)$$

$$(N-1)I_c \approx \frac{N}{p}(.027) - 1 + N(.038)$$

$$(N-1)I_c + 1 \approx \frac{N}{p}(.027) + N(.038)$$

$$(N-1)I_c + 1 - N(.038) \approx \frac{N}{p}(.027)$$

$$p((N-1)I_c + 1 - N(.038)) \approx N(.027)$$

$$p \approx \frac{N(.027)}{(N-1)I_c + 1 - N(.038)}$$

Lets see how accurate this is (it gives an approximation to the period, not the actual period) with text that contains about 21K letters. We use the same text and vigenere cipher with period 3 through 7.

`indcoin < plaintext`
Index of coincidence : 0.063616
Estimate of the period : 1.052158

- `indcoin < cyphertextvig3`
  Index of coincidence : 0.044720
  Estimate of the period : 3.990527
- `indcoin < cyphertextvig4`
  Index of coincidence : 0.042903
  Estimate of the period : 5.455495
- `indcoin < cyphertextvig5`
  Index of coincidence : 0.042236
  Estimate of the period : 6.304608
- `indcoin < cyphertextvig6`
  Index of coincidence : 0.041899
  Estimate of the period : 6.842702

Lets do another experiment with less letters (precisely 3183 letters).

`indcoin < plaintext`
Index of coincidence : 0.069377
Estimate of the period : 0.852563

- `indcoin < cyphertextvig3`
  Index of coincidence : 0.045386
  Estimate of the period : 3.512710
- `indcoin < cyphertextvig4`
  Index of coincidence : 0.045457
  Estimate of the period : 3.480884
- `indcoin < cyphertextvig5`
  Index of coincidence : 0.045034
  Estimate of the period : 3.681678
- `indcoin < cyphertextvig6`
  Index of coincidence : 0.043903
  Estimate of the period : 4.352677

Lets do another experiment with less letters (precisely 14590 letters).

`indcoin < plaintext`
Index of coincidence : 0.064586
Estimate of the period : 1.013137

- `indcoin < cyphertextvig3`
  Index of coincidence : 0.045976
  Estimate of the period : 3.357689
- `indcoin < cyphertextvig4`
  Index of coincidence : 0.042790
  Estimate of the period : 5.560689
- `indcoin < cyphertextvig5`
  Index of coincidence : 0.041953
  Estimate of the period : 6.718174
- `indcoin < cyphertextvig6`
  Index of coincidence : 0.041019
  Estimate of the period : 8.752510