

DESCRIPTION OF CT COUNT ANALYSIS

Thomas House, University of Manchester

December 16, 2020

These outputs have been provided as management information for operational planning purposes. They are provisional estimates circulated before public release. This management information should not be shared widely and should only be used for operational planning purposes. This information is not to be used publicly and ahead of being released publicly by the ONS.

1 General approach

Suppose we have a set of n individuals in households, indexed $i, j, \dots \in [n]$. Let X_i be a random variable for the type of individual i taking values with generic labels A, B, \dots . In this case we are interested in CT count. Write the m households as a set of subsets of individuals such that each individual is in exactly one household, i.e.

$$H_a \subseteq [n], \forall a \in [m], \quad H_a \cap H_b = \emptyset, \forall a \in [m], b \in [m] \setminus \{a\}, \quad \bigcup_{a=1}^m H_a = [n].$$

The size of the a -th household is then $n_a = |H_a|$. We are then interested in the table of pairs of individuals in households with certain properties,

$$Y_{AB} = \sum_{a \in [m], i \in H_a, j \in H_a \setminus \{i\}} \mathbf{1}_{\{X_i=A\}} \mathbf{1}_{\{X_j=B\}},$$

where $\mathbf{1}$ stands for the indicator function. Next, we want to think about what the expected value for such a quantity is under the assumption that each individual picks its state with independent probabilities given by the vector $\boldsymbol{\pi} = (\pi_A)$. The standard maximum likelihood estimator for such probabilities is

$$\hat{\pi}_A = \frac{1}{n} \sum_{i \in [n]} \mathbf{1}_{\{X_i=A\}}.$$

The expected table under the null hypothesis of independence will then be

$$M_{AB} = \pi_A \pi_B \sum_{a \in [m]} n_a (n_a - 1).$$

The Pearson residual associated with the (A, B) -th table entry is then

$$R_{AB} = \frac{Y_{AB} - M_{AB}}{\sqrt{M_{AB}}}.$$

Given certain conditions these residuals are asymptotically standard normal under the null [1]. (Note – a ‘to do’ is checking these conditions more carefully here, but the residuals remain a measure of departure from independence regardless.)

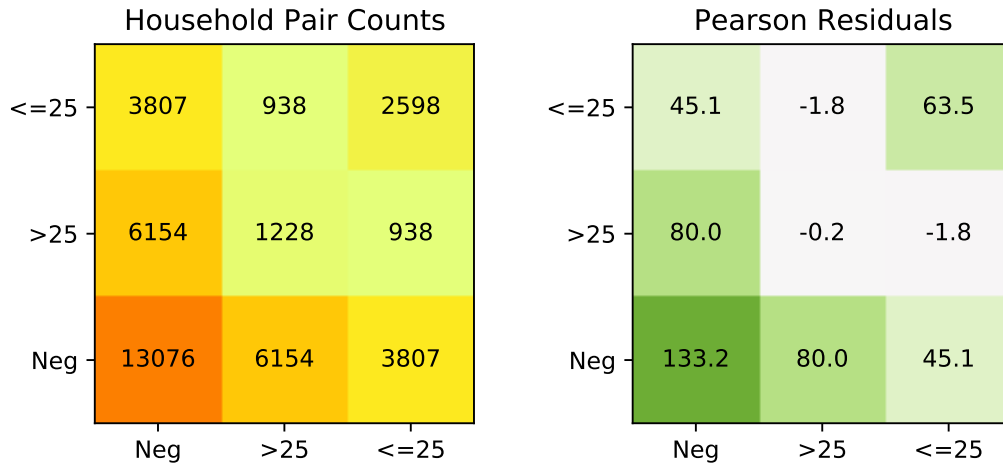


Figure 1: Left: Counts of pairs of individuals in households with at least one member ever positive by each member of the pairs lowest recorded CT count. Right: Pearson residuals compared to the null hypothesis of independence. Results larger than 2 are indicative of a significant positive association (and less than -2 of a significant negative association).

2 Applied to CT counts in households

Data up to 7 December were taken from the ONS CIS. The households with at least one positive test were taken, and pairs of individuals were tabulated by the minimum observed CT count for each member of the pair. Results are shown in Figure 1.

These results show that low CT counts are clustered with low CT counts, and Negatives with Negatives; there is also clustering of low CT counts and Negatives as would be expected since secondary attack rates are not infinite.

In contrast, the high CT counts do not cluster with other high CTs or low CTs – they only cluster with Negatives. This implies that they are not associated with transmission-type patterns. Caution is required in interpretation – because we tended to see the higher CT counts over summer, when other factors may have reduced within-household transmission, this cannot be interpreted causally at this stage.

References

- [1] Y. Bishop, S. E. Fienberg, and P. W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. Massachusetts Institute of Technology Press, Cambridge, 1975.