**Potential Panel Data and Duplicates**

**First look – July 24, 2016**

This short document is intended to detail the potential for multiple observations in the production data.

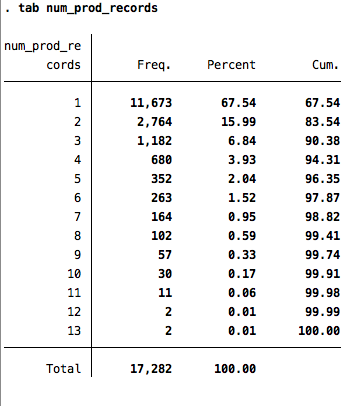
First, there are a number of duplicates for the COD\_LOTE variable even within each data file. See table below – some files have as many as 5 entries for a single COD\_LOTE number. It’s typical to have ~200 entries repeated 2x, and 10-50 repeated 3x.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Year** | **Time** | **1 entry** | **2 entries** | **3 entries** | **4 entries** | **5 entries** |
| **2001** | 1 | 1588 | 98 | 15 | 5 | 0 |
| **2001** | 2 | 1564 | 84 | 27 | 0 | 0 |
| **2002** | 1 | 1266 | 214 | 30 | 0 | 0 |
| **2002** | 2 | 1170 | 210 | 27 | 4 | 0 |
| **2003** | 1 | 1185 | 190 | 36 | 8 | 0 |
| **2003** | 2 | 1230 | 150 | 24 | 12 | 0 |
| **2004** | 1 | 1123 | 180 | 18 | 0 | 0 |
| **2004** | 2 | 1211 | 190 | 36 | 0 | 0 |
| **2005** | 1 | 1168 | 140 | 36 | 8 | 5 |
| **2005** | 2 | 681 | 76 | 3 | 0 | 0 |
| **2006** | 1 | 1540 | 172 | 45 | 8 | 0 |
| **2006** | 2 | 1621 | 244 | 27 | 0 | 0 |
| **2007** | 1 | 1570 | 192 | 18 | 16 | 0 |
| **2007** | 2 | 1662 | 204 | 51 | 4 | 0 |
| **2008** | 1 | 1515 | 150 | 24 | 4 | 0 |
| **2008** | 2 | 1624 | 216 | 39 | 0 | 0 |
| **2009** | 1 | 1576 | 172 | 21 | 0 | 0 |
| **2009** | 2 | 1589 | 194 | 30 | 0 | 0 |
| **2010** | 1 | 1463 | 150 | 9 | 0 | 0 |
| **2010** | 2 | 1511 | 178 | 9 | 4 | 0 |

Now, I tried to figure out if there are any panel data observations which would give us multiple years of data at a single field/farm. So I dropped any duplicates from the data (as reported above) and merged all of the COD\_LOTE into a single file. Also, I dropped any COD\_LOTE observations below 1000, because I think those are just too short.

Each data file of production data got a unique dummy variable, so that if the COD\_LOTE was in the first file, x2001\_1 was set to one, etc. Then, you can generate a new rowtotal variable that adds one for each year in which that COD\_LOTE appears.

See screenshot:



So, obviously there is potential for finding farms with multiple years of production data, but we’ll need to tread cautiously given all of the duplicates that are known to exist.