# ID Variable

## Anomalies Detected

| **Obs** | **ID** | **ID** |
| --- | --- | --- |
| 1758 | **722** | 272 |
| 1759 | 272 | **722** |

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **1751** | 271 | 0 | 1 | Male | 1 | 1751 |
| **1752** | 271 | 1 | 1 | Male | 1 | 1752 |
| **1753** | 271 | 2 | 1 | Male | 0 | 1753 |
| **1754** | 271 | 3 | 1 | Male | 0 | 1754 |
| **1755** | 271 | 6 | 1 | Male | 0 | 1755 |
| **1756** | 272 | 0 | 1 | Female | 1 | 1756 |
| **1757** | 272 | 1 | 1 | Female | 1 | 1757 |
| **1758** | **722** | 2 | 1 | Female | 1 | 1758 |
| **1759** | 272 | 3 | 1 | Female | 1 | 1759 |
| **1760** | 272 | 6 | 1 | Female | 0 | 1760 |
| **1761** | 272 | 9 | 1 | Female | 0 | 1761 |
| **1762** | 272 | 12 | 1 | Female | 0 | 1762 |
| **1763** | 273 | 0 | 0 | Male | 1 | 1763 |

## Corrective Action

Probable typo identified in Obs=1758 (highlighted in red above): 722 should read 272.

**Note:** The output of the FREQ procedure that most of the patients have an observation for all 7 valid values of the TIME variable. However, many of the patients have fewer observations than that. (See the table at the end of page 2).

While we don’t know enough about the dataset in question to know if this means that this is indicative of missing observations for the latter group of patients, in the real world, this potential issue would warrant further investigation.

# TREAT Variable

## Anomalies Detected

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **1052** | 163 | 2 | **A** | Male | 1 | 1052 |

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **1043** | 162 | 0 | 0 | Female | 1 | 1043 |
| **1044** | 162 | 1 | 0 | Female | 0 | 1044 |
| **1045** | 162 | 2 | 0 | Female | 0 | 1045 |
| **1046** | 162 | 3 | 0 | Female | 0 | 1046 |
| **1047** | 162 | 6 | 0 | Female | 0 | 1047 |
| **1048** | 162 | 9 | 0 | Female | 0 | 1048 |
| **1049** | 162 | 12 | 0 | Female | 0 | 1049 |
| **1050** | 163 | 0 | 0 | Male | 1 | 1050 |
| **1051** | 163 | 1 | 0 | Male | 1 | 1051 |
| **1052** | 163 | 2 | **A** | Male | 1 | 1052 |
| **1053** | 163 | 6 | 0 | Male | 1 | 1053 |
| **1054** | 163 | 12 | 0 | Male | 1 | 1054 |
| **1055** | 164 | 0 | 1 | Female | 0 | 1055 |
| **1056** | 164 | 1 | 1 | Female | 0 | 1056 |
| **1057** | 164 | 2 | 1 | Female | 0 | 1057 |
| **1058** | 164 | 3 | 1 | Female | 0 | 1058 |
| **1059** | 164 | 6 | 1 | Female | 0 | 1059 |
| **1060** | 164 | 9 | 1 | Female | 0 | 1060 |

## Corrective Action

Probable typo identified in Obs=1052 (highlighted in red above): ‘A’ should read ‘0’.

* **B**ecause TREAT value is always the same for a particular patient throughout the dataset – as indicated by the extract from the relevant SQL query below.

| **ID** | **Treat** | **count** |
| --- | --- | --- |
| 160 | 0 | 7 |
| 161 | 1 | 7 |
| 162 | 0 | 7 |
| 163 | 0 | 4 |
| 163 | **A** | 1 |
| 164 | 1 | 7 |
| 165 | 1 | 7 |
| 166 | 0 | 7 |
| 167 | 1 | 6 |

# GENDER Variable

## Anomalies Detected

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **1121** | 174 | 2 | 1 | **A** | 0 | 1121 |
| **1626** | 252 | 0 | 0 | **A** | 0 | 1626 |
| **1627** | 252 | 1 | 0 | **A** | 0 | 1627 |
| **1628** | 252 | 2 | 0 | **A** | 1 | 1628 |
| **1629** | 252 | 3 | 0 | **A** | 1 | 1629 |
| **1630** | 252 | 6 | 0 | **A** | 0 | 1630 |
| **1631** | 252 | 9 | 0 | **A** | 0 | 1631 |
| **1632** | 252 | 12 | 0 | **A** | 0 | 1632 |

## Corrective Action

8 typos identified in (highlighted in red above): ‘A’ should read ‘Female’ for ID=174 (Obs=1121) – because all other records for ID=174 specify the ‘Female’ value – and ‘A’ should be set to ‘’ for ID=252 (Obs=1626-1632) because all records for ID=252 are impacted by this anomaly, as shown in the following table.

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **1110** | 172 | 9 | 1 | Female | 0 | 1110 |
| **1111** | 172 | 12 | 1 | Female | 0 | 1111 |
| **1112** | 173 | 0 | 0 | Male | 0 | 1112 |
| **1113** | 173 | 1 | 0 | Male | 0 | 1113 |
| **1114** | 173 | 2 | 0 | Male | 0 | 1114 |
| **1115** | 173 | 3 | 0 | Male | 0 | 1115 |
| **1116** | 173 | 6 | 0 | Male | 0 | 1116 |
| **1117** | 173 | 9 | 0 | Male | 0 | 1117 |
| **1118** | 173 | 12 | 0 | Male | 0 | 1118 |
| **1119** | 174 | 0 | 1 | Female | 0 | 1119 |
| **1120** | 174 | 1 | 1 | Female | 0 | 1120 |
| **1121** | 174 | 2 | 1 | **A** | 0 | 1121 |
| **1122** | 174 | 3 | 1 | Female | 0 | 1122 |
| **1123** | 174 | 6 | 1 | Female | 0 | 1123 |
| **1124** | 174 | 9 | 1 | Female | 0 | 1124 |
| **1125** | 174 | 12 | 1 | Female | 0 | 1125 |
| **1623** | 251 | 0 | 1 | Male | 0 | 1623 |
| **1624** | 251 | 1 | 1 | Male | 0 | 1624 |
| **1625** | 251 | 2 | 1 | Male | 0 | 1625 |
| **1626** | 252 | 0 | 0 | **A** | 0 | 1626 |
| **1627** | 252 | 1 | 0 | **A** | 0 | 1627 |
| **1628** | 252 | 2 | 0 | **A** | 1 | 1628 |
| **1629** | 252 | 3 | 0 | **A** | 1 | 1629 |
| **1630** | 252 | 6 | 0 | **A** | 0 | 1630 |
| **1631** | 252 | 9 | 0 | **A** | 0 | 1631 |
| **1632** | 252 | 12 | 0 | **A** | 0 | 1632 |
| **1633** | 253 | 0 | 1 | Male | 0 | 1633 |
| **1634** | 253 | 1 | 1 | Male | 0 | 1634 |
| **1635** | 253 | 2 | 1 | Male | 0 | 1635 |

# TIME Variable

## Anomalies Detected

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **329** | 55 | **13** | 1 | Male | 0 | 329 |

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **320** | 54 | 6 | 0 | Female | 0 | 320 |
| **321** | 54 | 9 | 0 | Female | 0 | 321 |
| **322** | 54 | 12 | 0 | Female | 0 | 322 |
| **323** | 55 | 0 | 1 | Male | 0 | 323 |
| **324** | 55 | 1 | 1 | Male | 0 | 324 |
| **325** | 55 | 2 | 1 | Male | 0 | 325 |
| **326** | 55 | 3 | 1 | Male | 0 | 326 |
| **327** | 55 | 6 | 1 | Male | 0 | 327 |
| **328** | 55 | 9 | 1 | Male | 0 | 328 |
| **329** | 55 | **13** | 1 | Male | 0 | 329 |
| **330** | 56 | 0 | 0 | Female | 0 | 330 |
| **331** | 56 | 1 | 0 | Female | 0 | 331 |
| **332** | 56 | 2 | 0 | Female | 0 | 332 |
| **333** | 56 | 3 | 0 | Female | 0 | 333 |
| **334** | 56 | 6 | 0 | Female | 0 | 334 |
| **335** | 56 | 9 | 0 | Female | 0 | 335 |
| **336** | 56 | 12 | 0 | Female | 0 | 336 |

## Corrective Action

Probable typo identified in Obs=329 (highlighted in red above): ‘13’ should read ‘12’.

* Because all valid values for the TIME variable – except 12 - are present in the records for the impacted patient (ID=55).

# Y Variable

## Anomalies Detected

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **412** | 67 | 12 | 0 | Male | **5** | 412 |
| **784** | 123 | 2 | 0 | Male | **4** | 784 |

| **Obs** | **ID** | **Time** | **Treat** | **Gender** | **y** | **Obs** |
| --- | --- | --- | --- | --- | --- | --- |
| **405** | 66 | 12 | 0 | Female | 0 | 405 |
| **406** | 67 | 0 | 0 | Male | 0 | 406 |
| **407** | 67 | 1 | 0 | Male | 0 | 407 |
| **408** | 67 | 2 | 0 | Male | 0 | 408 |
| **409** | 67 | 3 | 0 | Male | 0 | 409 |
| **410** | 67 | 6 | 0 | Male | 0 | 410 |
| **411** | 67 | 9 | 0 | Male | 0 | 411 |
| **412** | 67 | 12 | 0 | Male | **5** | 412 |
| **780** | 122 | 9 | 1 | Female | 0 | 780 |
| **781** | 122 | 12 | 1 | Female | 0 | 781 |
| **782** | 123 | 0 | 0 | Male | 1 | 782 |
| **783** | 123 | 1 | 0 | Male | 1 | 783 |
| **784** | 123 | 2 | 0 | Male | **4** | 784 |
| **785** | 123 | 3 | 0 | Male | 0 | 785 |
| **786** | 123 | 6 | 0 | Male | 0 | 786 |
| **787** | 123 | 9 | 0 | Male | 0 | 787 |
| **788** | 123 | 12 | 0 | Male | 0 | 788 |
| **789** | 124 | 0 | 0 | Female | 1 | 789 |

## Corrective Action

2 probable typos identified in Obs=412 and 784 (highlighted in red above): ‘5’ and ‘4’, respectively, should be replaced with ‘.’.

* Because analysis of whole dataset indicates that value of the Y variable can vary across the records of a particular patient and, therefore, its true value cannot be derived for the two impacted patients (ID=67 and 123, respectively).