

```

/* A dataset called Toenail.xlsx has been posted on Moodle. The dataset contains the results from */
an experiment comparing two methods for treating toenail dermatophyte onychomycosis (TDO).
TDO is a common toenail infection that can be difficult to treat.
There are five variables in the dataset:
• ID: a unique identifier for each patient.
• Time: the time point at which the response was recorded. This was either 0, 1, 2, 3, 6, 9 or
12 months.
• Treat: The treatments were coded 1 or 0.
• Gender: Male or Female.
• y: the response recorded was that the infection was present (1) or not (0). */
/*(a) Read the data into your ST662 SAS library */

```

```

PROC IMPORT OUT=ST662.pt1
DATAFILE='/home/u45187342/my_courses/rafaeldeandrade0/ST662_data/Toenail.xlsx'
DBMS=XLSX replace;
GETNAMES=YES;
RUN;

```

```

/* (b) Create a variable in your dataset that is unique for each row of data. Do this by using the
code:
data NewDatasetName;
set OldDatasetName;
Obs = _n_;
run;
This new variable may help with the screening process. */

```

```

data st662.pt1;
set st662.pt1;
obs=_n_;
run;
/* (c) Create SAS code to screen the data for any anomalies. */

```

```

data st662.pt1;
set st662.pt1;
Gender=upcase(Gender);
run;

```

```

proc freq data=st662.pt1;
tables gender/ nocum nopercent;
run;
proc print data=ST662.pt1;
where gender not in ('MALE','FEMALE','');
run;
/*Gender has 8 unusual observation that is A*/

```

```

proc freq data=st662.pt1;
tables treat /nocum nopercent;
run;
proc print data=ST662.pt1;
where TREAT not in ('0','1','');
run;
/*Treat has 1 unusual observation as A*/

```

```

proc freq data=st662.pt1;
tables y /nocum nopercent;
run;
proc print data=ST662.pt1;
where Y not in (0,1);
run;

```

```

/*Table |Y| has two unusual observation one observation as 4 and other observation as 5 */

```

```

proc freq data=st662.pt1;
tables time /nocum nopercent;
run;
proc print data=ST662.pt1;
where TIME not in (0,1,2,3,6,9,12);
run;

```

```
/* Table time has one unusual observation as 13 in given column*/
```

```
proc univariate data=st662.pt1;
```

```
var id;
```

```
run;
```

```
PROC SQL;
```

```
SELECT DISTINCT(ID)
```

```
FROM ST662.PT1;
```

```
/* Table id has one unusual observation as 722 in given column*/
```

```
/* (d) Create SAS code to deal with any anomalies that you have found. I.e. generate code to either  
change the observation (if there is an obvious error) or to mark the observation as missing. */
```

```
data st662.pt1;
```

```
set st662.pt1;
```

```
if Gender in ('MALE', 'FEMALE') then Gender = Gender;
```

```
else Gender = '';
```

```
if TREAT IN ('0', '1', '') then TREAT=TREAT;
```

```
ELSE TREAT='';
```

```
IF Y IN (0,1) THEN Y=Y;
```

```
ELSE Y =.;
```

```
IF TIME IN (0,1,2,3,6,9,12) THEN TIME=TIME;
```

```
ELSE TIME=.;
```

```
if id in (174) then gender='FEMALE';
```

```
else gender=gender;
```

```
if id in (722) then id=272;
```

```
else id=id;
```

```
if id in (163) then treat='0';
```

```
then id=id;
```

```
if id in(55) then time=12;
```

```
else time=time;
```

```
run;
```

```
/*Gender has 8 unusual observation that is A was replaced using blank space as  
this was not a valid data*/
```

```
/* Treat has 1 unusual observation as A that was also replaced with blank as  
this was not a valid data */
```

```
/* Table |Y| has two unusual observation one observation as 4 and other  
observation as 5 that was replaced with "." as this was not a valid data*/
```

```
/* Table time has one unusual observation as 13 in given column as this is not specific  
to the response that was recorded and was replaced with "." */
```

```
/* 174 column was converted to gender missing as female */
```

```
/* After contacting the Db manager the 722 column was converted to 272 */
```

```
/* After contacting the Db manager the id 55 had an unusual observation in  
time which was converted to 12 */
```