```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
68
69         /* importing data from Excel
70         /* setting column names to match Excel file */
71         data professionalSoccer;
72         infile '/home/u63996162/STAT468 Senior Research Project/Professional Soccer Injuries 2016-2021.csv' dlm = ',' firstobs=2;
73         input
74             start_year
75             season_days_injured
76             season_minutes_played
77             height_cm
78             weight_kg
79             position $
80             age
81             bmi
82             work_rate_numeric
83             position_numeric
84             significant_injury_prev_season
85             season_days_injured_prev_season
86             isForward
87             isMidfielder
88             isDefender;
89         run;

NOTE: The infile '/home/u63996162/STAT468 Senior Research Project/Professional Soccer Injuries 2016-2021.csv' is:
      Filename=/home/u63996162/STAT468 Senior Research Project/Professional Soccer Injuries 2016-2021.csv,
      Owner Name=u63996162,Group Name=oda,
      Access Permission=-rw-r--r--,
      Last Modified=15Dec2025:20:44:00,
      File Size (bytes)=86514

NOTE: 1301 records were read from the infile '/home/u63996162/STAT468 Senior Research Project/Professional Soccer Injuries
      2016-2021.csv'.
      The minimum record length was 53.
      The maximum record length was 80.
NOTE: SAS went to a new line when INPUT statement reached past the end of a line.
NOTE: The data set WORK.PROFESSIONALSOCCER has 1300 observations and 15 variables.
NOTE: DATA statement used (Total process time):
      real time           0.02 seconds
      user cpu time       0.01 seconds
      system cpu time     0.00 seconds
      memory              908.56k
      OS Memory           19876.00k
      Timestamp           12/19/2025 03:23:08 AM
      Step Count                        24  Switch Count  2
      Page Faults                       0
      Page Reclaims                     184
      Page Swaps                        0
      Voluntary Context Switches        21
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           520


90
91         /* creating regression and including diagnostic plots */
92         proc reg data=professionalSoccer plots=diagnostics;
93         model season_days_injured = age season_minutes_played bmi
94         season_days_injured_prev_season isForward isMidfielder isDefender;
95         output out=residuals predicted=predict residual=Residuals;
96         run;

97
98         /* normal probability plot of residuals */

NOTE: The data set WORK.RESIDUALS has 1300 observations and 17 variables.
NOTE: PROCEDURE REG used (Total process time):
      real time           2.74 seconds
      user cpu time       0.22 seconds
      system cpu time     0.04 seconds
      memory              15608.68k
      OS Memory           36164.00k
      Timestamp           12/19/2025 03:23:11 AM
      Step Count                        25  Switch Count  28
      Page Faults                       0
      Page Reclaims                     16300
      Page Swaps                        0
      Voluntary Context Switches        1317
      Involuntary Context Switches      9
      Block Input Operations            0
      Block Output Operations           4928
```

```
99         proc univariate data=residuals;
100            var Residuals;
101            probplot Residuals / normal (mu=est sigma=est);
102            title "Normal Probability Plot of Residuals";
103        run;

NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.15 seconds
      user cpu time       0.10 seconds
      system cpu time     0.01 seconds
      memory              8551.68k
      OS Memory           37632.00k
      Timestamp           12/19/2025 03:23:11 AM
      Step Count                        26  Switch Count  1
      Page Faults                       0
      Page Reclaims                     1655
      Page Swaps                        0
      Voluntary Context Switches        175
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           648


104
105        /* residuals versus predicted */
106        proc sgplot data=residuals;
107            scatter x=predict y=Residuals;
108            refline 0 / axis=y;
109            title "Plot of Predicted Values Versus Residuals";
110        run;

NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.07 seconds
      user cpu time       0.03 seconds
      system cpu time     0.00 seconds
      memory              1471.62k
      OS Memory           34092.00k
      Timestamp           12/19/2025 03:23:11 AM
      Step Count                        27  Switch Count  2
      Page Faults                       0
      Page Reclaims                     345
      Page Swaps                        0
      Voluntary Context Switches        140
      Involuntary Context Switches      1
      Block Input Operations            0
      Block Output Operations           544

NOTE: There were 1300 observations read from the data set WORK.RESIDUALS.

111
112
113        data residuals_time;
114           set residuals;
115           obs_number + 1;  /* creates a time-like variable starting at 1 */
116        run;

NOTE: There were 1300 observations read from the data set WORK.RESIDUALS.
NOTE: The data set WORK.RESIDUALS_TIME has 1300 observations and 18 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      user cpu time       0.00 seconds
      system cpu time     0.00 seconds
      memory              1209.96k
      OS Memory           33704.00k
      Timestamp           12/19/2025 03:23:11 AM
      Step Count                        28  Switch Count  2
      Page Faults                       0
      Page Reclaims                     95
      Page Swaps                        0
      Voluntary Context Switches        14
      Involuntary Context Switches      0
      Block Input Operations            0
      Block Output Operations           520


117
118        /* scatter plot of residuals versus the time that they come in */
119        proc sgplot data=residuals_time;
120            scatter x=obs_number y=Residuals;
121            refline 0 / axis=y;
122            title "Independence of Errors Test";
123        run;
```

```
NOTE: PROCEDURE SGPLOT used (Total process time):
      real time           0.08 seconds
      user cpu time       0.05 seconds
      system cpu time     0.01 seconds
      memory              2353.21k
      OS Memory           34348.00k
      Timestamp           12/19/2025 03:23:12 AM
      Step Count                        29  Switch Count  2
      Page Faults                       0
      Page Reclaims                     294
      Page Swaps                        0
      Voluntary Context Switches        142
      Involuntary Context Switches      3
      Block Input Operations            0
      Block Output Operations           536

NOTE: There were 1300 observations read from the data set WORK.RESIDUALS_TIME.

124
125
126
127        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
137
```