Statistics/Data Analysis

```
      name:  <unnamed>
       log:  C:\Users\Siebelm\Documents\4 FMG\Coding Presentation\Stata_WClog.smcl
  log type:  smcl
 opened on:   4 Jul 2018, 22:28:09

1 .
2 . ************************
3 . **Load data set
4 . *set working directory
5 . cd "C:\Users\Siebelm\Documents\4 FMG\Coding Presentation"
  C:\Users\Siebelm\Documents\4 FMG\Coding Presentation

6 .
7 . *Import CSV & drop variables
8 . import delimited using results.csv
  (9 vars, 39070 obs)

9 . save results.dta, replace
  file results.dta saved

10. drop city country

11.
12. *Check first few rows
13. list * in 1/10
```

	date	home_t~m	away_t~m	home_s~e	away_s~e	tourna~t	neutral
1.	1872-11-30	Scotland	England	0	0	Friendly	FALSE
2.	1873-03-08	England	Scotland	4	2	Friendly	FALSE
3.	1874-03-07	Scotland	England	2	1	Friendly	FALSE
4.	1875-03-06	England	Scotland	2	2	Friendly	FALSE
5.	1876-03-04	Scotland	England	3	0	Friendly	FALSE
6.	1876-03-25	Scotland	Wales	4	0	Friendly	FALSE
7.	1877-03-03	England	Scotland	1	3	Friendly	FALSE
8.	1877-03-05	Wales	Scotland	0	2	Friendly	FALSE
9.	1878-03-02	Scotland	England	7	2	Friendly	FALSE
10.	1878-03-23	Scotland	Wales	9	0	Friendly	FALSE

```
14.
15. *Check variables
16. describe

  Contains data from results.dta
    obs:        39,070
   vars:             7                          4 Jul 2018 22:28
   size:     4,649,330
```

variable name	storage type	display format	value label	variable label
date	str10	%10s		
home_team	str30	%30s		
away_team	str30	%30s		
home_score	byte	%8.0g		
away_score	byte	%8.0g		
tournament	str42	%42s		
neutral	str5	%9s		

```
  Sorted by:
       Note: Dataset has changed since last saved.
```

```
17.
18. *Check missings (dots or ".")
19. findit mdesc

20. mdesc
```

Variable	Missing	Total	Percent Missing
date	0	39,070	0.00
home_team	0	39,070	0.00
away_team	0	39,070	0.00
home_score	0	39,070	0.00
away_score	0	39,070	0.00
tournament	0	39,070	0.00
neutral	0	39,070	0.00

```
21.
22. *Subset dataset to only include Germany matches
23. keep if home_team=="Germany" | away_team=="Germany"
  (38,139 observations deleted)

24.
25. list * in 1/10
```

	date	home_team	away_team	home_s~e	away_s~e	tourna~t	neutral
1.	1908-04-05	Switzerland	Germany	5	3	Friendly	FALSE
2.	1908-06-07	Austria	Germany	3	2	Friendly	FALSE
3.	1909-04-04	Germany	Switzerland	1	0	Friendly	FALSE
4.	1909-04-04	Hungary	Germany	3	3	Friendly	FALSE
5.	1910-04-03	Switzerland	Germany	2	3	Friendly	FALSE
6.	1910-04-24	Netherlands	Germany	4	2	Friendly	FALSE
7.	1910-05-16	Germany	Belgium	0	3	Friendly	FALSE
8.	1910-10-16	Germany	Netherlands	1	2	Friendly	FALSE
9.	1911-03-26	Germany	Switzerland	6	2	Friendly	FALSE
10.	1911-04-23	Belgium	Germany	2	1	Friendly	FALSE

```
26.
27.
28. ************************
29. **Recodes and Data Manpulation
30. *Simple recode (each observation equals 1)
31. gen match = 1

32.
33. list * in 1/10
```

	date	home_team	away_team	home_s~e	away_s~e	tourna~t	neutral	match
1.	1908-04-05	Switzerland	Germany	5	3	Friendly	FALSE	1
2.	1908-06-07	Austria	Germany	3	2	Friendly	FALSE	1
3.	1909-04-04	Germany	Switzerland	1	0	Friendly	FALSE	1
4.	1909-04-04	Hungary	Germany	3	3	Friendly	FALSE	1
5.	1910-04-03	Switzerland	Germany	2	3	Friendly	FALSE	1
6.	1910-04-24	Netherlands	Germany	4	2	Friendly	FALSE	1
7.	1910-05-16	Germany	Belgium	0	3	Friendly	FALSE	

```
  >   1 |
    8. | 1910-10-16       Germany   Netherlands          1          2   Friendly     FALSE
  >   1 |
    9. | 1911-03-26       Germany   Switzerland          6          2   Friendly     FALSE
  >   1 |
   10. | 1911-04-23       Belgium       Germany          2          1   Friendly     FALSE
  >   1 |
      ____|
```

```
34.
35. *Conditional recode (Define Goals Scored)
36. gen goals = home_score if home_team=="Germany"
  (437 missing values generated)

37. replace goals = away_score if away_team=="Germany"
  (437 real changes made)

38.
39. *Conditional recode (Define Goals Conceded)
40. gen conceded = away_score if home_team=="Germany"
  (437 missing values generated)

41. replace conceded = home_score if away_team=="Germany"
  (437 real changes made)

42.
43. list home_team away_team home_score away_score goals conceded in 921/931
```

	home_team	away_team	home_s~e	away_s~e	goals	conceded
921.	Czech Republic	Germany	1	2	2	1
922.	Germany	Norway	6	0	6	0
923.	Northern Ireland	Germany	1	3	3	1
924.	Germany	Azerbaijan	5	1	5	1
925.	England	Germany	0	0	0	0
926.	Germany	France	2	2	2	2
927.	Germany	Spain	1	1	1	1
928.	Germany	Brazil	0	1	0	1
929.	Austria	Germany	2	1	1	2
930.	Germany	Saudi Arabia	2	1	2	1
931.	Germany	Mexico	0	1	0	1

```
44.
45. *Simple recode (Define Goal Differential)
46. gen goaldiff = goals - conceded

47.
48. list home_score away_score goals conceded goaldiff in 921/931
```

	home_s~e	away_s~e	goals	conceded	goaldiff
921.	1	2	2	1	1
922.	6	0	6	0	6
923.	1	3	3	1	2
924.	5	1	5	1	4
925.	0	0	0	0	0
926.	2	2	2	2	0
927.	1	1	1	1	0
928.	0	1	0	1	-1
929.	2	1	1	2	-1
930.	2	1	2	1	1
931.	0	1	0	1	-1

```
49.
50. *Dummy recode (Define Friendly v Competitive match)
51. gen friendly = tournament == "Friendly"

52. replace friendly = . if tournament == "."
  (0 real changes made)

53. label define frdlylabs 0 "FALSE" 1 "TRUE"

54. label values friendly frdlylabs

55.
56. list tournament friendly in 921/931
```

	tournament	friendly
921.	FIFA World Cup qualification	FALSE
922.	FIFA World Cup qualification	FALSE
923.	FIFA World Cup qualification	FALSE
924.	FIFA World Cup qualification	FALSE
925.	Friendly	TRUE
926.	Friendly	TRUE
927.	Friendly	TRUE
928.	Friendly	TRUE
929.	Friendly	TRUE
930.	Friendly	TRUE
931.	FIFA World Cup	FALSE

```
57.
58. *Categorical recode (home vs away vs neutral matches)
59. gen home = "home" if home_team=="Germany"
  (437 missing values generated)

60. replace home = "away" if away_team=="Germany"
  (437 real changes made)

61. replace home = "neutral" if neutral=="TRUE"
  variable home was str4 now str7
  (148 real changes made)

62. encode home, gen(home_num)

63. label list home_num
  home_num:
           1 away
           2 home
           3 neutral

64.
65. list home_team away_team neutral home home_num in 921/931
```

	home_team	away_team	neutral	home	home_num
921.	Czech Republic	Germany	FALSE	away	away
922.	Germany	Norway	FALSE	home	home
923.	Northern Ireland	Germany	FALSE	away	away
924.	Germany	Azerbaijan	FALSE	home	home
925.	England	Germany	FALSE	away	away
926.	Germany	France	FALSE	home	home
927.	Germany	Spain	FALSE	home	home
928.	Germany	Brazil	FALSE	home	home
929.	Austria	Germany	FALSE	away	away
930.	Germany	Saudi Arabia	FALSE	home	home
931.	Germany	Mexico	TRUE	neutral	neutral

```
66.
67. *Conditional recode (opponent)
68. gen opponent = away_team if home_team=="Germany"
  (437 missing values generated)

69. replace opponent = home_team if away_team=="Germany"
  (437 real changes made)

70.
71. list home_team away_team opponent in 921/931
```

	home_team	away_team	opponent
921.	Czech Republic	Germany	Czech Republic
922.	Germany	Norway	Norway
923.	Northern Ireland	Germany	Northern Ireland
924.	Germany	Azerbaijan	Azerbaijan
925.	England	Germany	England
926.	Germany	France	France
927.	Germany	Spain	Spain
928.	Germany	Brazil	Brazil
929.	Austria	Germany	Austria
930.	Germany	Saudi Arabia	Saudi Arabia
931.	Germany	Mexico	Mexico

```
72.
73. *String variables (Define year)
74. gen year = substr(date,1,4)

75. destring year, replace
  year has all characters numeric; replaced as int

76.
77. list date year in 921/931
```

	date	year
921.	2017-09-01	2017
922.	2017-09-04	2017
923.	2017-10-05	2017
924.	2017-10-08	2017
925.	2017-11-10	2017
926.	2017-11-14	2017
927.	2018-03-23	2018
928.	2018-03-27	2018
929.	2018-06-02	2018
930.	2018-06-08	2018
931.	2018-06-17	2018

```
78.
79. *Drop redundant variables
```

```
80. drop home_score away_score tournament neutral home_team away_team

81.
82. list * in 921/931
```

	date	match	goals	conceded	goaldiff	friendly	home	home_num	opponent	year
921.	2017-09-01	1	2	1	1	FALSE	away	away	Czech Republic	2017
922.	2017-09-04	1	6	0	6	FALSE	home	home	Norway	2017
923.	2017-10-05	1	3	1	2	FALSE	away	away	Northern Ireland	2017
924.	2017-10-08	1	5	1	4	FALSE	home	home	Azerbaijan	2017
925.	2017-11-10	1	0	0	0	TRUE	away	away	England	2017
926.	2017-11-14	1	2	2	0	TRUE	home	home	France	2017
927.	2018-03-23	1	1	1	0	TRUE	home	home	Spain	2018
928.	2018-03-27	1	0	1	-1	TRUE	home	home	Brazil	2018
929.	2018-06-02	1	1	2	-1	TRUE	away	away	Austria	2018
930.	2018-06-08	1	2	1	1	TRUE	home	home	Saudi Arabia	2018
931.	2018-06-17	1	0	1	-1	FALSE	neutral	neutral	Mexico	2018

```
83.
84.
85. ************************
86. **Descriptive Statistics
87. *Tabs
88. tab home friendly
```

home	friendly FALSE	TRUE	Total
away	108	271	379
home	119	285	404
neutral	139	9	148
Total	366	565	931

```
89.
90. *Unweighted proportions
91. prop home_num friendly

  Proportion estimation             Number of obs   =        931
```

	Proportion	Std. Err.	[95% Conf.	Interval]
home_num				
away	**.4070892**	**.0161101**	**.375897**	**.4390488**
home	**.433942**	**.0162519**	**.4023625**	**.4660671**
neutral	**.1589689**	**.01199**	**.1368244**	**.1839335**
friendly				
FALSE	**.3931257**	**.0160167**	**.3621732**	**.424961**
TRUE	**.6068743**	**.0160167**	**.575039**	**.6378268**

```
92.
93. *Goal difference table
94. table home friendly, contents(mean conceded)
```

home	friendly FALSE	TRUE
away	**.8055556**	**1.446494**
home	**.7226891**	**1.164912**
neutral	**1.172662**	**1.666667**

```
95. table home friendly, contents(mean goaldiff)
```

home	friendly FALSE	TRUE
away	**1.537037**	**.4464945**
home	**2.168067**	**1.25614**
neutral	**.7553957**	**.2222222**

```
96. table home friendly, contents(mean goals)
```

home	friendly FALSE	TRUE
away	**2.342592**	**1.892989**
home	**2.890756**	**2.421053**
neutral	**1.928058**	**1.888889**

```
97.         *or*
98. forval frdly=0/1 {
    2.          foreach hme in away home neutral {
    3.                  foreach var in conceded goaldiff goals {
    4.                          local slabel: value label friendly
    5.                          local vlabel: label `slabel' `frdly'
    6.                          di as res _n "`var' if home is `hme' and friendly is `vlabel'"
    7.                          mean `var' if home=="`hme'" & friendly==`frdly'
    8.                  }
    9.          }
   10. }
```

conceded if home is away and friendly is FALSE

Mean estimation Number of obs = **108**

	Mean	Std. Err.	[95% Conf.	Interval]
conceded	**.8055556**	**.0867257**	**.633632**	**.9774791**

goaldiff if home is away and friendly is FALSE

Mean estimation Number of obs = **108**

	Mean	Std. Err.	[95% Conf.	Interval]
goaldiff	**1.537037**	**.2060969**	**1.128474**	**1.9456**

goals if home is away and friendly is FALSE

Mean estimation Number of obs = **108**

	Mean	Std. Err.	[95% Conf.	Interval]
goals	**2.342593**	**.1943125**	**1.957391**	**2.727795**

conceded if home is home and friendly is FALSE

Mean estimation Number of obs = **119**

	Mean	Std. Err.	[95% Conf.	Interval]
conceded	**.7226891**	**.0851654**	**.5540383**	**.8913398**

goaldiff if home is home and friendly is FALSE

Mean estimation Number of obs = **119**

	Mean	Std. Err.	[95% Conf.	Interval]
goaldiff	**2.168067**	**.217232**	**1.737889**	**2.598246**

goals if home is home and friendly is FALSE

Mean estimation Number of obs = **119**

	Mean	Std. Err.	[95% Conf.	Interval]
goals	**2.890756**	**.1930976**	**2.50837**	**3.273142**

conceded if home is neutral and friendly is FALSE

Mean estimation Number of obs = **139**

	Mean	Std. Err.	[95% Conf.	Interval]
conceded	**1.172662**	**.1075446**	**.9600135**	**1.38531**

goaldiff if home is neutral and friendly is FALSE

Mean estimation Number of obs = **139**

	Mean	Std. Err.	[95% Conf.	Interval]
goaldiff	**.7553957**	**.1677346**	**.4237335**	**1.087058**

goals if home is neutral and friendly is FALSE

Mean estimation Number of obs = **139**

	Mean	Std. Err.	[95% Conf.	Interval]
goals	**1.928058**	**.1357103**	**1.659717**	**2.196398**

conceded if home is away and friendly is TRUE

Mean estimation Number of obs = **271**

	Mean	Std. Err.	[95% Conf.	Interval]
conceded	**1.446494**	**.0812448**	**1.286541**	**1.606448**

goaldiff if home is away and friendly is TRUE

Mean estimation Number of obs = **271**

	Mean	Std. Err.	[95% Conf.	Interval]
goaldiff	**.4464945**	**.1171955**	**.2157613**	**.6772277**

goals if home is away and friendly is TRUE

Mean estimation Number of obs = **271**

	Mean	Std. Err.	[95% Conf.	Interval]
goals	**1.892989**	**.0909982**	**1.713833**	**2.072145**

conceded if home is home and friendly is TRUE

Mean estimation Number of obs = **285**

	Mean	Std. Err.	[95% Conf.	Interval]
conceded	**1.164912**	**.0654996**	**1.035986**	**1.293839**

goaldiff if home is home and friendly is TRUE

Mean estimation Number of obs = **285**

	Mean	Std. Err.	[95% Conf.	Interval]
goaldiff	**1.25614**	**.1388426**	**.9828492**	**1.529431**

goals if home is home and friendly is TRUE

Mean estimation Number of obs = **285**

	Mean	Std. Err.	[95% Conf.	Interval]
goals	**2.421053**	**.1187548**	**2.187301**	**2.654804**

conceded if home is neutral and friendly is TRUE

Mean estimation Number of obs = **9**

	Mean	Std. Err.	[95% Conf.	Interval]
conceded	**1.666667**	**.372678**	**.8072697**	**2.526064**

goaldiff if home is neutral and friendly is TRUE

Mean estimation Number of obs = **9**

	Mean	Std. Err.	[95% Conf.	Interval]
goaldiff	**.2222222**	**.5719795**	**-1.096765**	**1.541209**

goals if home is neutral and friendly is TRUE

Mean estimation Number of obs = **9**

	Mean	Std. Err.	[95% Conf.	Interval]
goals	**1.888889**	**.4843221**	**.7720401**	**3.005738**

```
99.
100 *Opponent table
101 mean goaldiff if opponent=="Korea Republic"
```

Mean estimation Number of obs = **3**

	Mean	Std. Err.	[95% Conf.	Interval]
goaldiff	**0**	**1**	**-4.302653**	**4.302653**

```
102 tab match if opponent=="Korea Republic"
```

match	Freq.	Percent	Cum.
1	**3**	**100.00**	**100.00**
Total	**3**	**100.00**	

```
103
104
105 ************************
106 **Graphs
107 *Histogram
108 hist goaldiff, bin(15) freq ///
  >         title(Histogram of Goal Differences) ///
  >         ytitle(Count) xtitle(Goal Differential)
  (bin=15, start=-6, width=1.2666667)

109
110 *Line graphs
111 bysort year: egen m_goals = mean(goals)
```

```
112 twoway (line m_goals year, lcolor(green)), xlabel(1908(20)2018)

113
114 bysort year: egen m_conceded = mean(conceded)

115 replace m_conceded = m_conceded*-1
  (930 real changes made)

116 twoway (line m_conceded year, lcolor(red)), xlabel(1908(20)2018)

117
118 bysort year: egen m_goaldiff = mean(goaldiff)

119 twoway (line m_goaldiff year, lcolor(navy)), xlabel(1908(20)2018)

120
121 twoway (line m_goaldiff year, lcolor(navy)) ///
  >         (line m_goals year, lcolor(green) lwidth(thin)) ///
  >         (line m_conceded year, lcolor(red) lwidth(thin)), ///
  >         xlabel(1908(30)2018) yline(0, lcolor(grey) lwidth(thin))
  (note:  named style grey not found in class color, default attributes used)

122
123 translate Stata_WClog.smcl Stata_WClog.pdf, rmargin(.5)
  (file Stata_WClog.pdf written in PDF format)

124 translate Stata_WClog.smcl Stata_WClog.txt , replace
  (note: file Stata_WClog.txt not found)
  (file Stata_WClog.txt written in .txt format)

125
126
127
128
129
130
  end of do-file

131 exit, clear
```