

```

* Programmed by Charles Hwang *
* Coded in SAS OnDemand      *
* Tuesday, February 19, 2019 *
* Course: STAT 303           *
* Title: Homework 2          *;

/* a */ Proc Import Out=pitchers datafile="/home/chwang10/sasuser.v94/PitcherData.xlsx" DBMS=xlsx replace;

/* b */ Data work.pitcherstats; /* Creating new variables for ERA and WHIP via combining */
Set pitchers;
ERA=9*ER/IP;
ERA=round(ERA,.01);
WHIP=(BB+H)/IP;
WHIP=round(WHIP,.001);

/* c */ Proc Sort;
By ERA;
Proc Print data=work.pitcherstats (obs=10);
Var Player ERA WHIP;

/* d */ Proc Means mean stddev skewness kurtosis;
Var HR SO WHIP;
* The distributions of HR and WHIP appear to be nearly symmetric. The
distribution for SO does have a noticeable skewness value of 1.000 and
kurtosis of 0.819 which may indicate that the distribution has some skew. ;

/* e */ Proc Means CLM;
Var ERA;
* We are 95% confident that the average ERA of the top 57 MLB
pitchers for a particular season is between 3.46 and 3.90. ;

/* f */ Proc Sort data=work.pitcherstats;
By Player;
Proc Sort out=playerbyteam data="/home/chwang10/sasuser.v94/playerbyteam.sas7bdat";
By Player;
Data PitchersTeam (drop=D E); /* Removing extraneous variables from .txt file */
Merge work.pitcherstats playerbyteam;
By Player;
Proc Print data=PitchersTeam;

/* g */ Proc Means data=PitchersTeam median Qrange;
Var ERA WHIP HR BB;
Class League;
* Both leagues' top pitchers have roughly the same ERA and WHIP. The medians of HR and BB indicate
that the AL's top pitchers appear to allow more home runs than the NL's top pitchers (25 vs. 20.5)
and that the NL's top pitchers appear to allow more walks than the AL's top pitchers (56 vs. 51). ;

/* h */ Proc Sort;
By League;
Proc Boxplot;
Plot SO*League;
* The mean number of strikeouts and range of the boxplots is roughly the same for both leagues.
The IQR is significantly larger in the AL boxplot than the NL boxplot, suggesting a
greater variance in number of strikeouts among top pitchers in the AL than in the NL. ;

/* i */ Proc Univariate noprint;
By League;
Histogram SO /normal (noprint); /* Suppressing UNIVARIATE tables */
Inset mean="Mean" (6.2) var="Variance" (7.2) skew="Skewness" (4.2) /Pos=NW;
Run;
* The distribution appears to be approximately normal in the histogram of strikeouts
by top AL pitchers. The distribution does not appear to be normal in the histogram
of strikeouts by top NL pitchers. The histogram is noticeably right-skewed. ;

/* j */ ODS select TestsForNormality;

```

```
Proc Univariate normal;
```

```
Var S0;
```

```
* Shapiro-Wilk Test for Normality (on Strikeouts)
```

```
  H0: Strikeouts are normally distributed
```

```
  HA: Strikeouts are not normally distributed
```

```
   $\alpha = .01$ 
```

```
  t = 0.9289, p = 0.0024
```

```
  We reject H0 at  $\alpha = .01$ . There is sufficient evidence  
  that strikeouts are not normally distributed. ;
```

```
Run;
```