Stephen Hanna 109097796

HW8

**Problem 1:**

**data** ProcessingTime;

Input ProcessingTime @@;

datalines;

28 25 27 31 10 26 30 15 55 12 24 32 28 42 38

;

**proc** **print** data = ProcessingTime;

title 'Processing Time'

run;

**proc** **univariate** data=ProcessingTime normal;

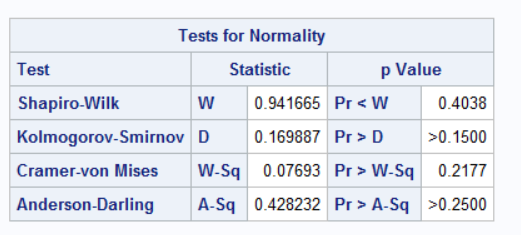
var ProcessingTime;

**run** ;

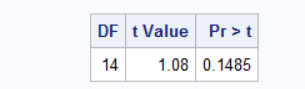
**proc** **ttest** h0=**25** data=ProcessingTime sides=u alpha=**.05**;

var ProcessingTime;

**run**;



p> .05 for normality testing, so the data is assumed to be normally distributed.



p>.05, so we do not reject the null hypothesis. The processing time is not significantly greater than 25 minutes.

**Problem 2:**

**data** Hypertension;

Input Temperature $ BloodPressure;

datalines;

26 152

26 157

26 179

26 182

26 176

26 149

5 384

5 369

5 354

5 375

5 366

5 423

;

**proc** **print** data = Hypertension;

title 'Temperature dependent influence on blood pressure'

run;

**proc** **univariate** data=Hypertension normal;

var BloodPressure;

by Temperature;

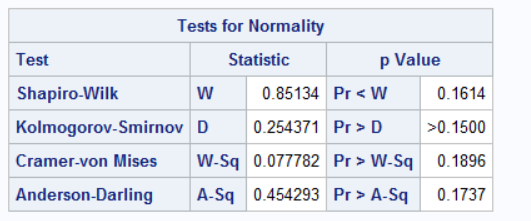
**run** ;

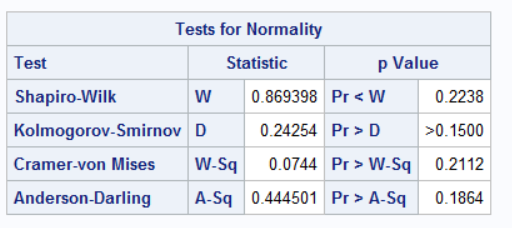
**proc** **ttest** data=Hypertension sides=**2** alpha=**0.05** h0=**0**;

Class Temperature;

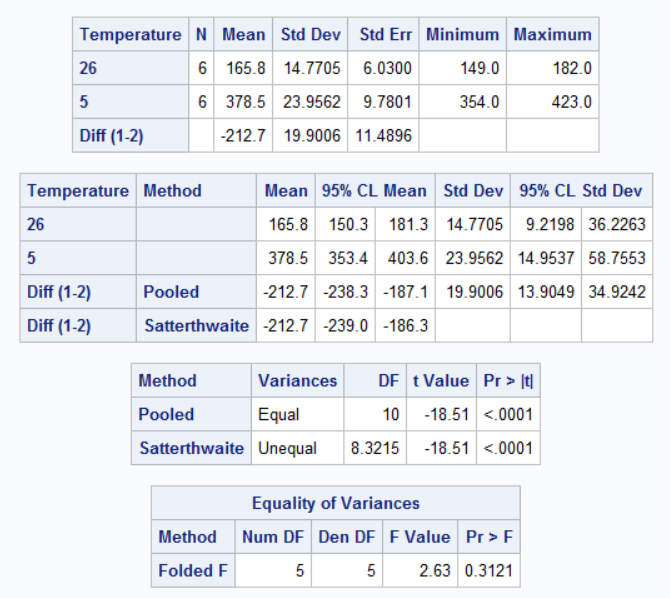
var BloodPressure;

**run**;

****

****

The shapiro test demonstrates that both sets of values for respective temperatures are normally distributed since p>.1 in both cases.

****

The p value for the equality of variances is greater than .05, so the two variances are assumed to not be statistically different. Thus, the pooled t test value will be used. The p value is less than .0001, so we can reject the null hypothesis and state that **the mean blood pressure at different temperatures is significantly different**

**Problem 3:**

**data** CornYield;

Input VarA $ VarB;

diff = varA - varB;

datalines;

48.2 41.5

44.6 40.1

49.7 44.0

40.5 41.2

54.6 49.8

47.1 41.7

51.4 46.8

;

**proc** **print** data = CornYield;

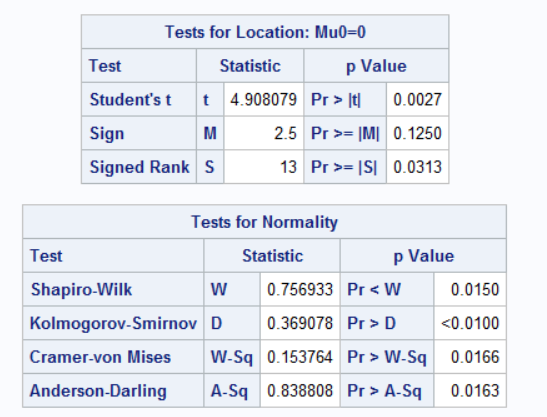
title 'Corn Yields for different varieties'

run;

**proc** **univariate** data=CornYield;

var diff;

**run**;



The data is paired. The data is not normal, since the p value is less than .05 for the shapiro- wilk test. Thus, we use the p value for the Wilcoxon signed rank test. The p value is less than .05, so we reject the null hypothesis and conclude, with over 95% confidence, that the means between the two yields are not the same.

**Problem 4:**

**data** SalesData;

Input Before $ After;

diff = after-before;

datalines;

12 18

18 24

25 24

9 14

14 19

16 20

;

**proc** **print** data = SalesData;

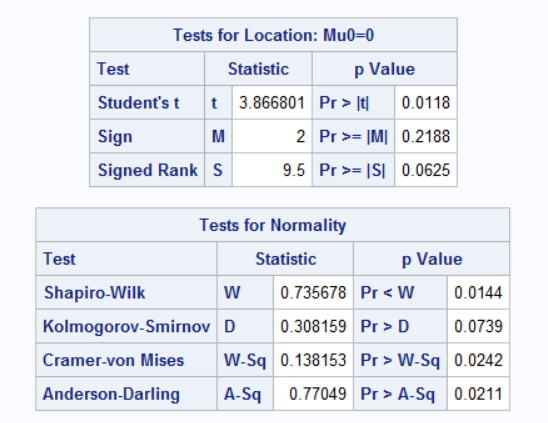
title 'Sales before and after training';

**run**;

**proc** **univariate** data=SalesData normal;

var diff;

**run**;



The data is paired. The data is not normal, since the p value is less than .05 for the shapiro- wilk test. Thus, we use the p value for the Wilcoxon signed rank test. However, the p value has to be halved when testing whether one data set is greater than the other, rather than one ranked data set simply being significantly different than the other. Hence, the p value used is .0625/2, or .03125. The p value is greater than .01, so we cannot reject the null hypothesis. **The sales before cannot be concluded, with 99% confidence, to be significantly less than the sales after.**