Deliverable 1:

Choosing the right wine

Rebekah Sander

Research Question

Are the wines in this dataset significantly lower than the average of 15 points on the wine quality 20-point rating scale?

Research Variable

WHO	WHAT measurement is made on each		TYPE OF MEASURE
	Name of Variable	Question Asked	
Pinot Noir wine samples	Wine Quality	What is wine quality?	Quantitative Variable Unit: points
One Quantitative Variable being tested against the average wine quality rating of 15.			
1.One Mean t Test			
2.Sign Test			

Assessing Normality

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.984138	Pr < W	0.8559
Kolmogorov-Smirnov	D	0.054375	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.017034	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.140776	Pr > A-Sq	>0.2500

▶ **Normality Tests:** All 4 normality tests resulted in p-values > α=0.05, suggesting that we may assume normality.

▶ **QQ Plot:** The data shows minimal deviation from the agreement line, supporting normality. Further, this agrees with the normality tests that say that the data is relatively normal.
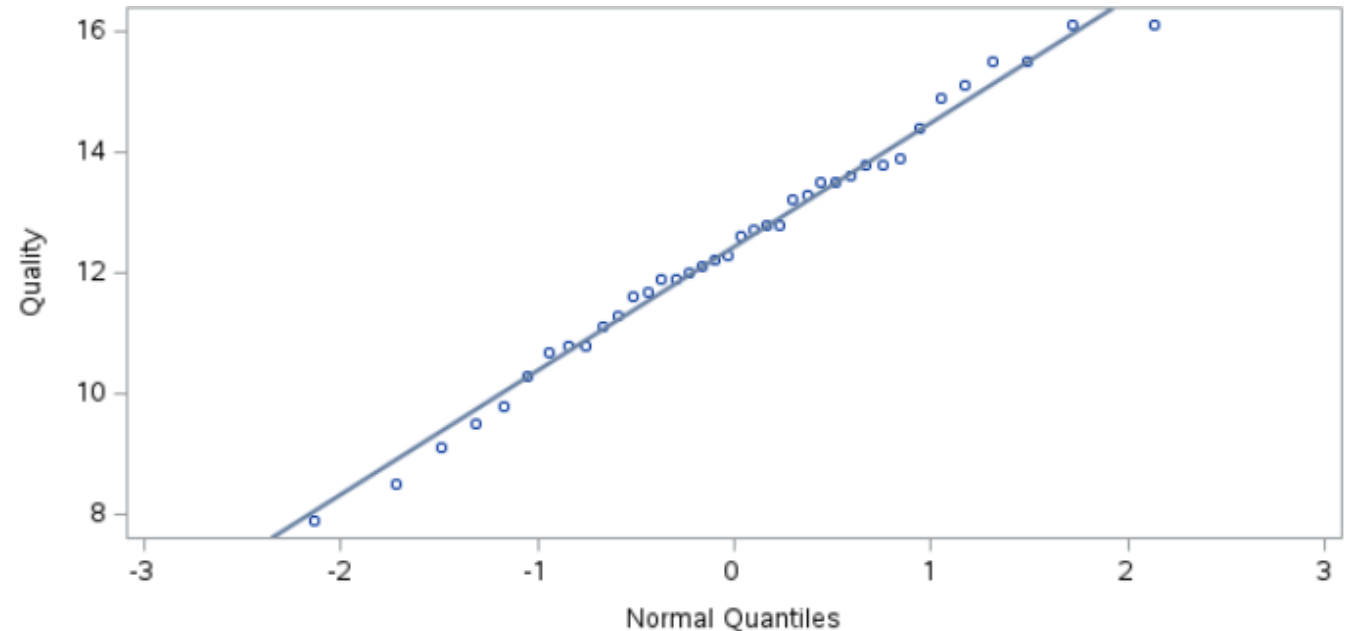

```
/* Checking normality: qq plot and normality tests */
title "Figure 1: QQ Plots";
proc univariate data=work.wine normaltest plots;
VAR 'Wine Quality'n;
run;
title;
```

3

Choosing Hypothesis Test: One Mean t-test

▶ Since the data is relatively normal, we will perform the one mean t-test.

$$H_0: \mu = 15$$
$$H_A: \mu < 15$$
$$\alpha = 0.05$$

▶ The null hypothesis is that the true mean wine quality of all pinot noir wine is 15 points.

▶ The alternative hypothesis is that the true mean wine quality of all pinot noir wine types is less than 15 points.

▶ The level of significance, $\alpha = 0.05$, tells us we have a 5% chance of concluding the alternative hypothesis is true when really the null hypothesis is true.

Performing One Mean t-test

► t-value: The sample average wine quality of 12.4368 points is 7.73 standard errors to the left of the hypothesized wine quality of 15 points.

► p-value: There is less than a 0.01% chance of observing a sample average wine quality of 38 pinot noir wine samples to be 12.4368 points or less when the true wine quality is 15 points.

► Conclusion: p-value = 0.001 < α = 0.05, we reject H_0. We are 95% confident that the true average pinot noir wine quality is less than 15 points. This is statistically significant.

The TTEST Procedure

Variable: Wine Quality (Quality)

N	Mean	Std Dev	Std Err	Minimum	Maximum
38	12.4368	2.0454	0.3318	7.9000	16.1000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
12.4368	-Infty	12.9966	2.0454	1.6675	2.6462

DF	t Value	Pr < t
37	-7.73	<.0001

```
/* Run the hypothesis test: one mean t-test */
Proc ttest data=work.wine plots=summary sides=L H0=15 alpha=0.05;
var 'Wine Quality'n;
run;
```

Post hoc test: Confidence Interval

▶ We are 95% confident that the true average wine quality for all pinot noir wine types is between 11.76 points and 13.11 points.

▶ We can conclude that 15 is too high for the true average wine quality from this confidence interval because 15 is above all of the values contained in the confidence interval.

▶ The confidence interval agrees with the one mean t-test.

Analysis Variable : Wine Quality Quality				
N	Mean	Std Dev	Lower 95% CL for Mean	Upper 95% CL for Mean
38	12.44	2.05	11.76	13.11

```
/*Post hoc test: confidence interval*/
Proc means data=work.wine MAXDEC=2 n mean stddev LCLM UCLM ;
    var 'Wine Quality'n;
run;
```

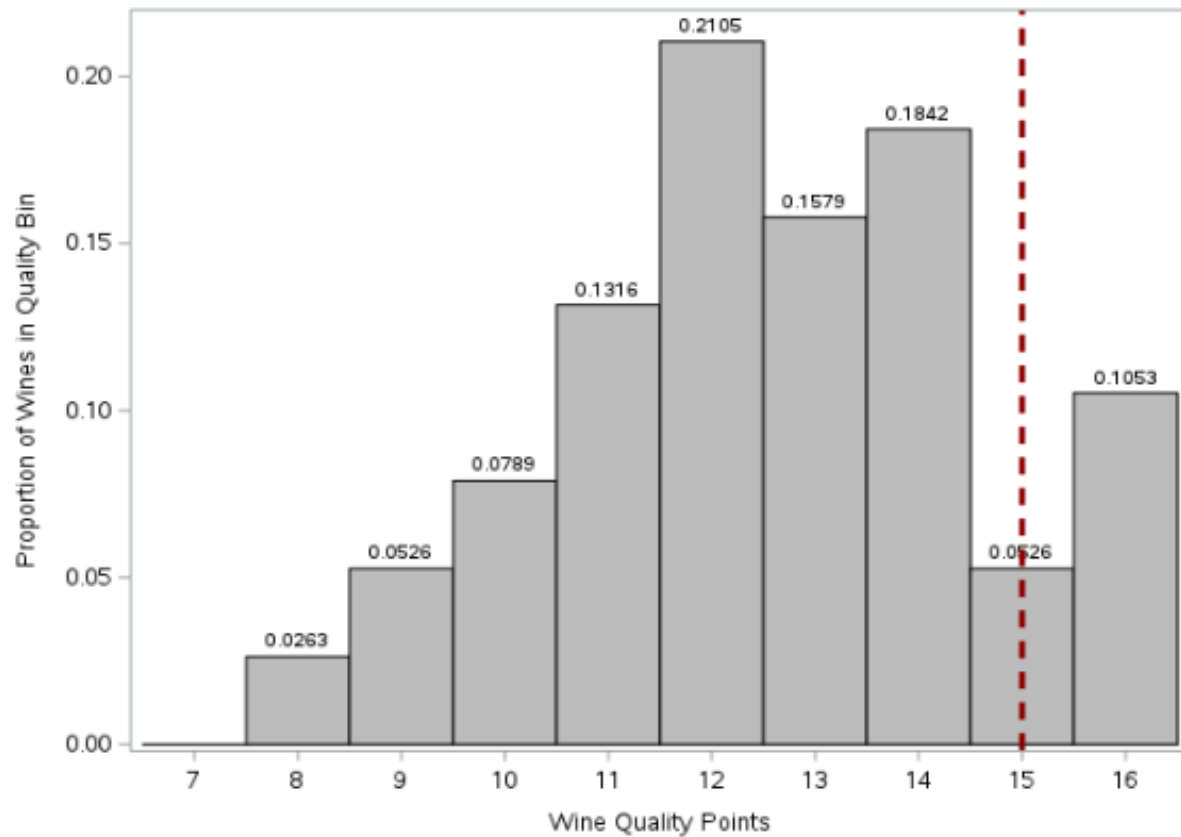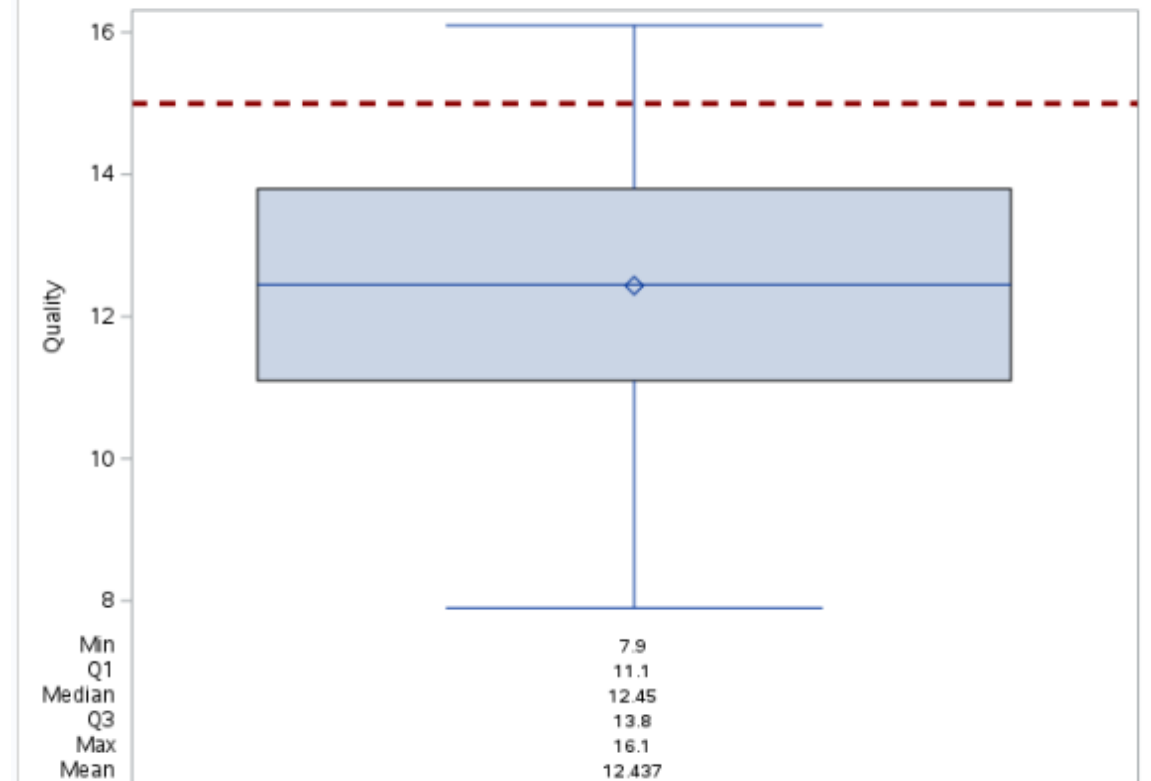# Supporting Graphics

Figure 1: Histogram of Wine Quality (n = 38)

Figure 2: Boxplot of Wine Quality

Min	7.9
Q1	11.1
Median	12.45
Q3	13.8
Max	16.1
Mean	12.437

Supporting Graphics: Boxplot

```
/* Histogram using SGPLOT */
TITLE1 "Figure 1: Histogram of Wine Quality (n = 38)";
Proc SGPLOT data=work.wine;
histogram 'Wine Quality'n/
    datalabel=proportion
    scale=proportion showbins
    binwidth=1 binstart=7
    fillattrs=(color=H000BB00)
    boundary=upper;
    refline 15 / axis=x lineattrs=(thickness=3 color=darkred pattern=dash);
    xaxis label='Wine Quality Points' labelattrs=(size=10);
    yaxis label='Proportion of Wines in Quality Bin' labelattrs=(size=10);
run;
TITLE;
```

```
/*Boxplot*/
PROC SGPLOT DATA = work.wine;
  VBOX 'Wine Quality'n/ boxwidth=1 displaystats=(MEAN MAX Q3 MEDIAN Q1 MIN);
  refline 15 / axis=y lineattrs=(thickness=3 color=darkred pattern=dash);
  TITLE1 "Figure 2:  Boxplot of Wine Quality";
RUN;
```
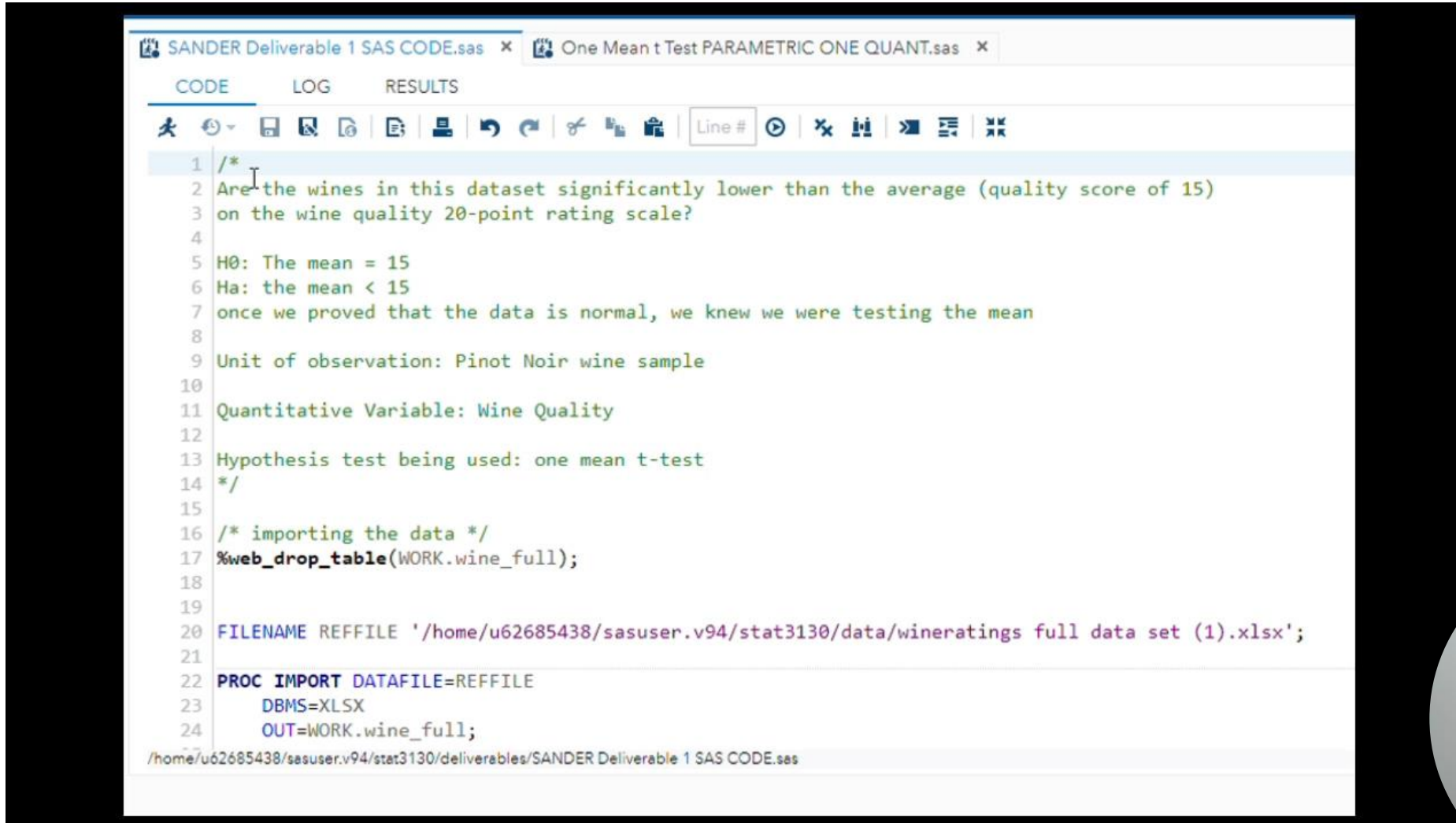
Taking Action

▶ From this analysis, we are able to gain insight when it comes to choosing a pinot noir wine.

▶ Suppose you are planning an important event where you wish to serve wine. If you are debating between different types of pinot noir, you can base the decision off of the analysis conclusion.

▶ The analysis tells us we are 95% confident that the true average wine quality is less than 15 points. Therefore 15 should not be the standard for choosing a pinot noir wine.

SAS Code Screen Recording