REVIEW FOR Midterm!

HIGH SCHOOL & Beyond

providing a lifetime of insight

HS&B began in 1980 as a nationally representative sample of 30,030 sophomores and 28,240 seniors in 1,015 public and private high schools in the United States. From the initial sample of 58,270 public and private high school students, 14,825 sophomores and 11,995 seniors were selected to be re-interviewed over their early adult years. Each school contained a representative sample of 36 sophomores and 36 seniors, making possible inferences about each school and its student body. The student questionnaires in 1980 gathered important information about educational experiences, cognitive skills

HIGH SCHOOL & Beyond

providing a lifetime of insight

id	female	race	ses	schtyp	prog	read	write	math	science	socst
70	0	4	1	1	1	57	52	41	47	57
121	1	4	2	1	3	68	59	53	63	61
86	0	4	3	1	1	44	33	54	58	31
141	0	4	3	1	3	63	44	47	53	56
172	0	4	2	1	2	47	52	57	53	61
113	0	4	2	1	2	44	52	51	63	61
50	0	3	2	1	1	50	59	42	53	61
11	0	1	2	1	2	34	46	45	39	36
84	0	4	2	1	1	63	57	54	58	51
48	0	3	2	1	2	57	55	52	50	51
75	0	4	2	1	3	60	46	51	53	61
60	0	4	2	1	2	57	65	51	63	61
95	0	4	3	1	2	73	60	71	61	71
104	0	4	3	1	2	54	63	57	55	46
38	0	3	1	1	2	45	57	50	31	56
115	0	4	1	1	1	42	49	43	50	56
76	0	4	3	1	2	47	52	51	50	56
195	0	4	2	2	1	57	57	60	58	56
114	0	4	3	1	2	68	65	62	55	61

HSB2 Data Set:

This data file contains 200 observations from a sample of high school students with demographic information about the students, such as their gender (**female**), socio-economic status (**ses**) and ethnic background (**race**). It also contains a number of scores on standardized tests, including tests of reading (**read**), writing (**write**), mathematics (**math**) and social studies (**socst**).

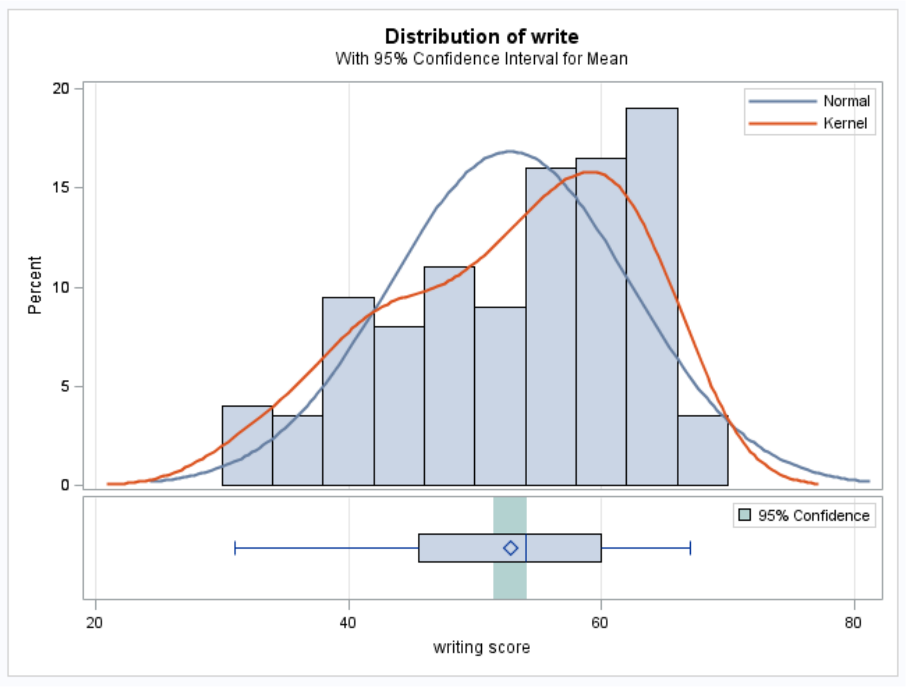
Review!!!!

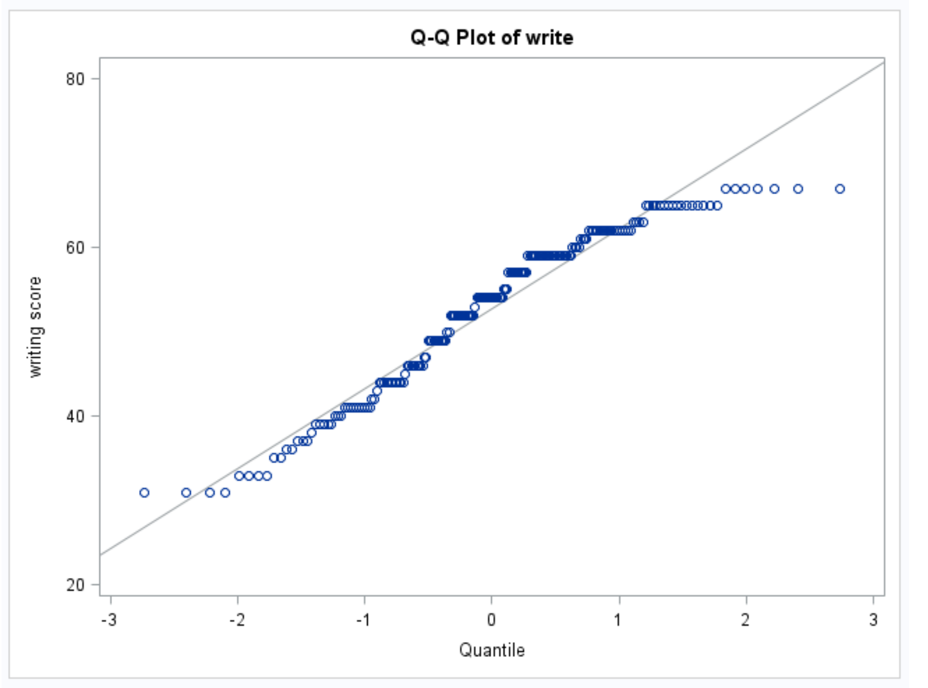
1. Download the HSB2 Dataset and import into a SAS session.
2. Test the claim that the mean writing score is equal to 50. Write all 6 steps.
3. Test the claim that the mean writing score is different for males and females. Include all 6 steps.
4. There are 4 unique race categories. Test to see if the 1st race has a different mean writing score than the 4th race.
5. Use proc power to find the power of a test to detect a difference of 3 in the means if the two groups have standard deviations of 5 and 8, respectively and sample sizes of 20 and 40, respectively. Assume we want to use a Satterthwaite approximation.

2. Test the claim that the mean writing score is equal to 50. Write all steps.

Problem Statement: Test the claim that the mean writing score is significantly different from 50.

Assumptions:





The histogram and q-q plot provide strong evidence of a left skew, although the sample size of 200 should ensure that the sample mean will be normally distributed (central limit theorem). We will assume the scores are independent of one another and proceed with a one-sample t-test.

*A transformation is another option if it improves normality, but the inference will be on the median.

2. Test the claim that the mean writing score is equal to 50. Write all 6 steps of the hypothesis test.

1. $H_0: \mu_{writing} = 50$

   $H_A: \mu_{writing} \neq 50$

```
data quantile;
myquant = quantile('t', 0.975, 199);
run;

proc print data = quantile;
run;
```

Obs	myquant
1	1.97196

```
proc ttest data = TMP1.hsb2 H0 = 50;
var write;
run;
```

DF	t Value	Pr > |t|
199	4.14	<.0001

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
52.7750	51.4533	54.0967	9.4786	8.6318	10.5110

2. $Critical\ Value: \pm t_{.975, 200-1} = \pm 1.97$

3. $Test\ Statistic: 4.14$

4. $P - value < .0001$

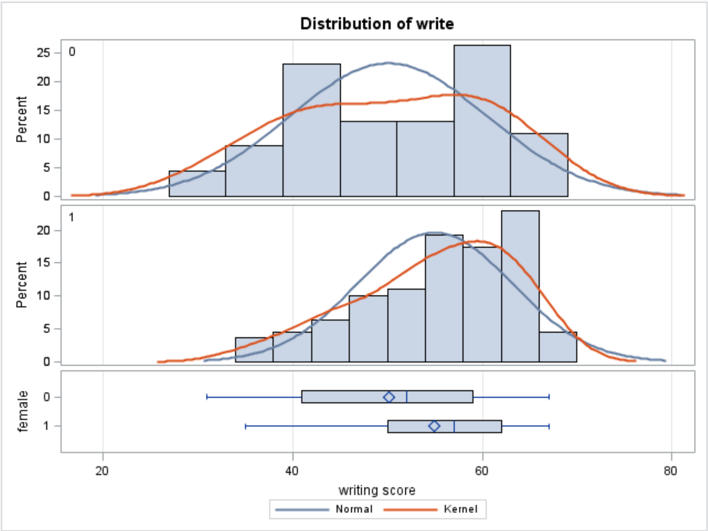
5. $Reject\ H_0$

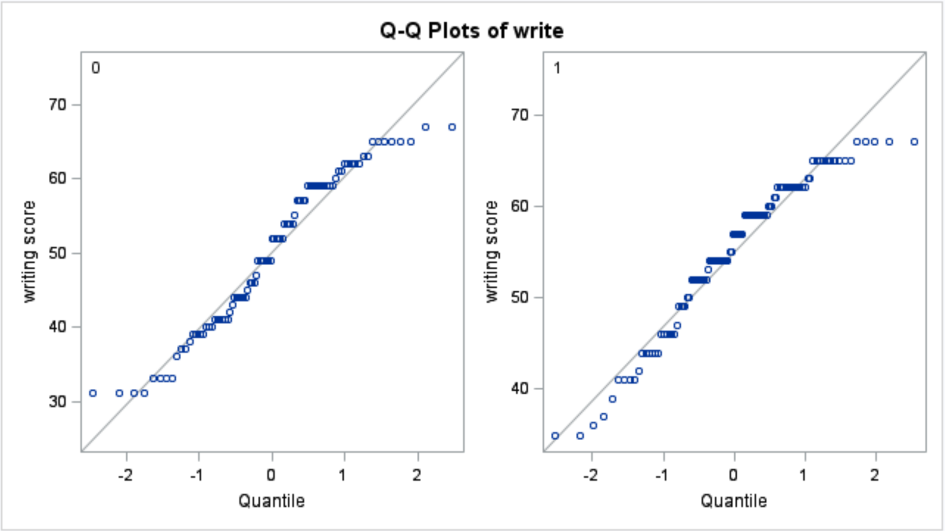
6. There is strong evidence to suggest at the alpha = .05 level of significance (p-value < .0001) that the mean writing score is different than 50 points. A 95% confidence interval for the true mean writing score is (51.5 points, 54.1 points). We can infer that the mean is not equal to 50 for the entire population of interest as the data was a random sample.

3. Test the claim that the mean writing score is different for males and females. Include all steps.

Problem statement: Test the claim that the mean writing score is different for males and females.

Assumptions:





While there is not significant evidence that the male writing scores (female = 0) are not normally distributed, the histograms and q-q plots provide evidence of a left skewed distribution of writing scores for females (female = 1). However, since there are 91 in the male group and 109 in the female group, the Central Limit Theorem will ensure that sample means from these distributions are normally distributed, thus making the t-test robust to the normality assumption.

We will assume the scores are independent of one another both within and between groups.

3. Test the claim that the mean writing score is different for males and females. Include all steps.

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	90	108	1.61	0.0187

Brown and Forsythe's Test for Homogeneity of write Variance ANOVA of Absolute Deviations from Group Medians					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
female	1	262.6	262.6	9.62	0.0022
Error	198	5404.9	27.2976		

The histograms on the last slide were somewhat inconclusive on the question of equality of variances. For this reason, we seek secondary evidence in the form of a formal hypothesis test. Since there is evidence that the writing scores are not normally distributed, the Brown-Forsythe test of equality of variance should be used instead of the F-Test. There is significant evidence at the alpha = .05 level of significance (p-value = .0022) to suggest that the variance of the male writing scores is different from that of the females.

There is some evidence that the standard deviations are different; therefore, since the Welch's t-test is nearly as powerful as the Student t-test even when the standard deviations are the same, we will proceed with the Welch's test of the difference of means.

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	198	-3.73	0.0002
Satterthwaite	Unequal	169.71	-3.66	0.0003

*A transformation is another option if it improves normality and makes variances more equal.

3. Test the claim that the mean writing score is different for males and females. Include all 6 steps of the hypothesis test.

1. H_0: $\mu_{female} = \mu_{male}$
   H_A: $\mu_{female} \neq \mu_{male}$

```
proc ttest data = hsb sides = 2;
class female;
var write;
run;
```

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	198	-3.73	0.0002
Satterthwaite	Unequal	169.71	-3.66	0.0003

```
data quantile;
thisquant = quantile('t', 0.975, 169.71);
run;

proc print data = quantile;
run;
```

Obs	thisquant
1	1.97404

female	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		50.1209	47.9747	52.2670	10.3052	8.9947	12.0662
1		54.9908	53.4466	56.5351	8.1337	7.1786	9.3843
Diff (1-2)	Pooled	-4.8699	-7.4418	-2.2981	9.1846	8.3622	10.1878
Diff (1-2)	Satterthwaite	-4.8699	-7.4992	-2.2407			

2. $Critical\ Value$: (run t-test to get Satterthwaite DF)

$\pm t_{.975,169.71} = \pm 1.97$

3. $Test\ Statistic$: -3.66

4. $P - value = .0003$

5. $Reject\ H_0$

6. There is strong evidence to suggest at the alpha = .05 level of significance (p-value = .0003) that the mean writing score of female high school students in the U.S. is different than the mean writing score of males. (Note the inference to the broader population.) This was an observational study; thus, no causal inference can be deduced. A 95% confidence interval for this difference is: (2.2 points, 7.5 points)-the positive difference in favor of females.

3. Test the claim that the mean writing score is different for males and females.

Welch's t-test vs. alternative solution: Nonparametric options

Variable: write (writing score)

female	N	Mean	Std Dev	Std Err	Minimum	Maximum
0	91	50.1209	10.3052	1.0803	31.0000	67.0000
1	109	54.9908	8.1337	0.7791	35.0000	67.0000
Diff (1-2)		-4.8699	9.1846	1.3042		

female	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		50.1209	47.9747	52.2670	10.3052	8.9947	12.0662
1		54.9908	53.4466	56.5351	8.1337	7.1786	9.3843
Diff (1-2)	Pooled	-4.8699	-7.4418	-2.2981	9.1846	8.3622	10.1878
Diff (1-2)	Satterthwaite	-4.8699	-7.4992	-2.2407			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	198	-3.73	0.0002
Satterthwaite	Unequal	169.71	-3.66	0.0003

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	90	108	1.61	0.0187

```
proc npar1way data =hsb Wilcoxon;
class female;
var write;
run;
```

Wilcoxon Scores (Rank Sums) for Variable write Classified by Variable female					
female	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
0	91	7792.0	9145.50	406.559086	85.626374
1	109	12308.0	10954.50	406.559086	112.917431
Average scores were used for ties.					

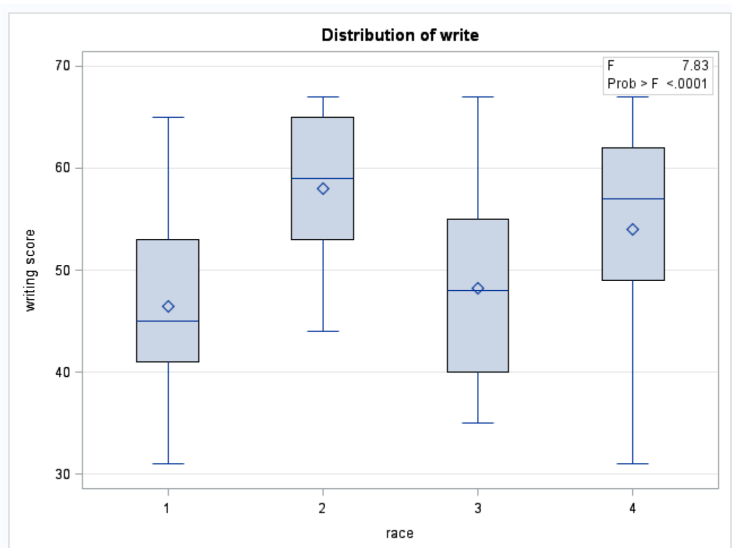
Wilcoxon Two-Sample Test	
Statistic	7792.0000
Normal Approximation	
Z	-3.3279
One-Sided Pr < Z	0.0004
Two-Sided Pr > |Z|	0.0009
t Approximation	
One-Sided Pr < Z	0.0005
Two-Sided Pr > |Z|	0.0010
Z includes a continuity correction of 0.5.	

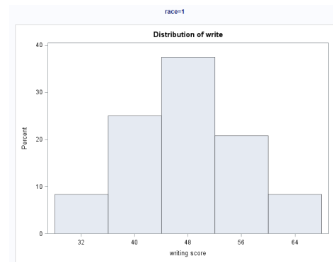
Kruskal-Wallis Test	
Chi-Square	11.0833
DF	1
Pr > Chi-Square	0.0009

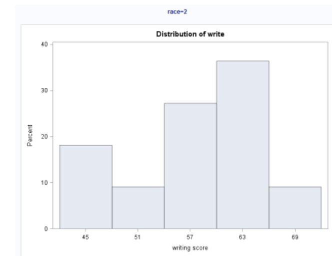
4. There are 4 unique race categories. Test to see if the 1st race has a different mean writing score than the 4th race.

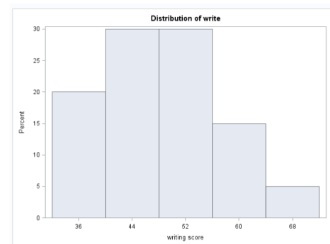
Problem Statement: Test the claim that the writing scores of the first and fourth race significantly different. First, we will test the claim that any of the means/medians are different.

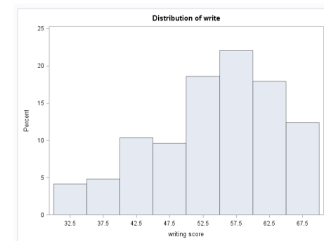
Assumptions:









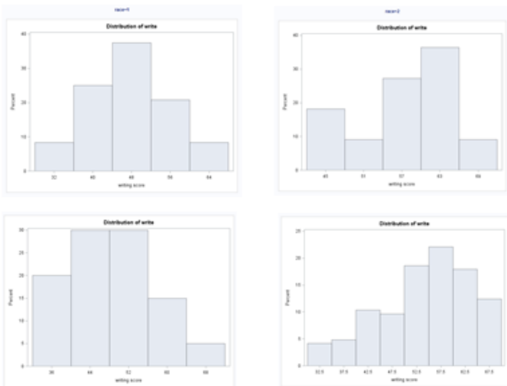


Level of race	N	write	
		Mean	Std Dev
1	24	46.4583333	8.27242232
2	11	58.0000000	7.89936706
3	20	48.2000000	9.32229924
4	145	54.0551724	9.17255819

While the only histogram that provides strong evidence (n = 145 and a considerable left skew) against normality is race = 4, that is the group with the largest sample size. Remember that with small sample sizes, it is difficult to ascertain the true shape of the underlying distribution. There is no reason to believe the shape of the other races should be any different than the shape of the distribution of writing scores of race = 4. We will assume that the shapes (although maybe not the locations) of the distributions of the underlying populations of the first three groups are the same as that of the last (skewed) group. Some smaller sample sizes make it unclear that the CLT will kick in, calling the normality of some of the sample means into question. For this reason, we will conduct a Kruskal-Wallis Test and make our inference about the medians.

*A transformation is another option if it improves normality and makes variances more equal.

4. There are 4 unique race categories. Test to see if the 1st race has a different mean writing score than the 4th race.



Again, given that the other groups have small sample sizes and there is no reason to believe the shape of the other races should be any different than the shape of the distribution of writing scores of race = 4, we will assume that the shapes (although maybe not the locations) of the distributions are the same. For this reason we will conduct a Kruskal-Wallis Test and make our inference about the medians.

```
proc npar1way data =hsb;
class race;
var write;
run;
```

Kruskal-Wallis Test	
Chi-Square	22.1942
DF	3
Pr > Chi-Square	<.0001

1. $H_0: Median_1 = Median_2 = Median_3 = Median_4$
   $H_A: Median_i \neq Median_j \; for \; some \; i, j$

2. $Critical\ Value: can\ skip$

3. $Test\ Statistic: \chi^2 = 22.1942$

4. $P - value < .0001$

5. $Reject\ H_0$

There is sufficient evidence at the alpha = .05 level of significance (p-value < .0001) that at least one of the medians is different among the race groups.

Next, we shall test to see if the 1st race has a different median than the 4th race. A multiple comparison adjustment is not needed here since our hypothesis was formulated before we looked at the data and we are only examining one comparison.

4. There are 4 unique race categories. Test to see if the 1st race has a different mean writing score than the 4th race.

A Wilcoxon Rank Sum test will be performed between the two groups.

1. $H_0: Median_1 = Median_4$
   $H_A: Median_1 \neq Median_4$

2. $Critical\ Value: z_{.975} \cong \pm 1.96$

```
data quantile;
aquant = quantile('NORMAL', 0.975);
run;
proc print data = quantile;
run;
```

Obs	aquant
1	1.95996

3. $Test\ Statistic: z = -3.72$

4. $P - value = .0002$

5. $Reject\ H_0$

6. There is strong evidence to suggest at the alpha = .05 level of significance (p-value = .0002 from the Rank Sum Test) that the median writing score of U.S. high school students with race = 1 is different than the median writing score of those of race = 4. This was an observational study; thus, no causal inference can be deduced. The best estimate of the difference in locations is 8, in favor of race 4 scores, with a Hodges-Lehmann 95% confidence interval of the difference of (5 points, 13 points).

*Note that the difference in medians is actually 12. That is because the HL confidence interval is not exactly for the medians when the data is not symmetric.

Hodges-Lehmann Estimation

Location Shift (1 - 4) -8.0000

95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-13.0000	-5.0000	-9.0000	2.0409

Wilcoxon Two-Sample Test

Statistic	1215.0000
Normal Approximation	
Z	-3.7240
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	0.0002
t Approximation	
One-Sided Pr < Z	0.0001
Two-Sided Pr > |Z|	0.0003
Z includes a continuity correction of 0.5.	

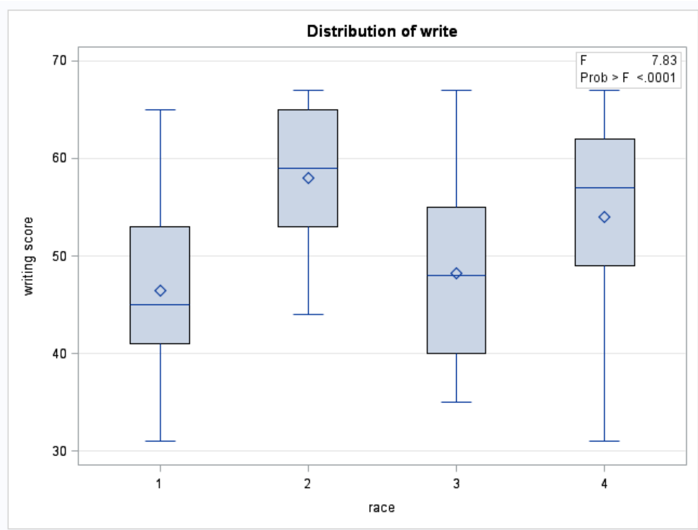
```
proc means data = hsb median;
class race;
var write;
run;
```

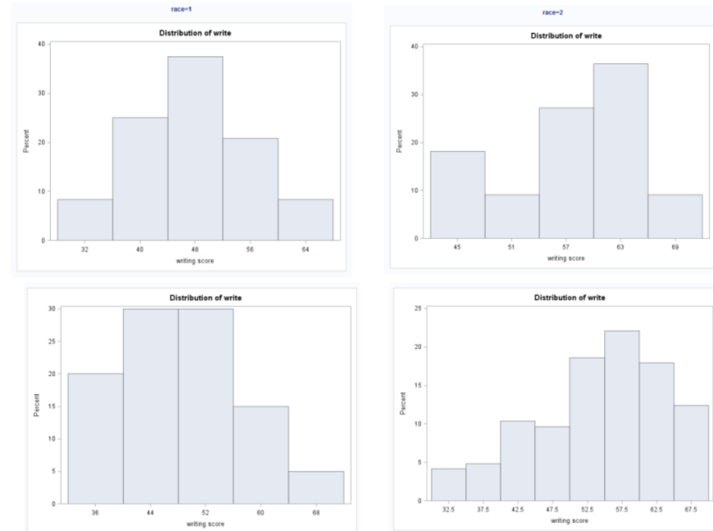
The MEANS Procedure

Analysis Variable : write

race	N Obs	Median
1	24	45.0000000
2	11	59.0000000
3	20	48.0000000
4	145	57.0000000

4. There are 4 unique race categories. Test to see if the 1st race has a different mean writing score than the 4th race.
ALTERNATIVE SOLUTION: CONTRASTS





Level of race	N	write	
		Mean	Std Dev
1	24	46.4583333	8.27242232
2	11	58.0000000	7.89936706
3	20	48.2000000	9.32229924
4	145	54.0551724	9.17255819

The race = 4 histogram provides strong evidence (n = 145 and a considerable left skew) against normality. While the other histograms do not provide much evidence against normality, they are based on small sample sizes and may indeed be left skewed as the evidence suggests race = 4 is. However, it can be argued that even the smallest group n = 11 has a sufficient sample size to ensure the central limit theorem will provide normally distributed sample means. For this reason, we will proceed with an ANOVA and t-tests for the planned analysis. In addition, we will assume the standard deviations are equal although this is a risky assumption given the data has very unequal sample sizes.

4. There are 4 unique race categories. Test to see if the 1st race has a different mean writing score than the 4th race. ALTERNATIVE SOLUTION: CONTRASTS (first ANOVA)

1. $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

   $H_A: \mu_i \neq \mu_j \;for\; some\; i, j$

2. $Critical\ Value: can\ skip$

```
proc glm data = hsb;
class race;
model write = race;
run;
```

The GLM Procedure
Dependent Variable: write

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

3. $Test\ Statistic: F = 7.83$

4. $P - value < .0001$

5. $Reject\ H_0$

6. There is sufficient evidence at the alpha = .05 level of significance (p-value < .0001) that at least one of the means is different between the race groups.

Next, we shall test to see if the 1st race has a different mean than the 4th race. A multiple comparison adjustment is not needed here since our hypothesis was formulated before we looked at the data.

4. There are 4 unique race categories. Test to see if the 1st race has a different mean writing score than the 4th race.

ALTERNATIVE SOLUTION: CONTRASTS

1. $H_0\!: \mu_1 = \mu_4$

   $H_A\!: \mu_1 \neq \mu_4$

```
proc glm data = hsb;
class race;
model write = race;
contrast 'Contrast Race 1 v. Race 4' race 1 0 0 -1 ;
estimate 'Est Race 1 v. Race 4' race 1 0 0 -1;
run;
```

2. *Critical Value: can skip if only using contrast*

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Est Race 1 v. Race 4	-7.59683908	[illegible]	-3.82	0.0002

3. *Test Statistic*: $F = 14.59$ or $t = -3.82$

4. $P - value = .0002$

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Contrast Race 1 v. Race 4	1	1188.388371	1188.388371	14.59	0.0002

5. *Reject* H_0

6. There is strong evidence at the alpha = .05 level of significance (p-value = .0002 from a two sample t-test) to suggest that the mean writing score of race = 1 U.S. high school students is different than that of students who are race = 4.

A 95% confidence interval for the difference in means is:

$$point\ estimate \pm multiplier * standard\ error$$

$$-7.60 \pm 1.97 * 1.99$$

$$-7.60 \pm 3.92$$

$$(-11.52, -3.68)$$

Obs	myquant
1	1.97214

```
data quantile;
myquant = quantile('t', 0.975,200-4);
run;
proc print data = quantile;
run;
```

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced Model:

Full model:

Source	DF	SS	MS	F	Pr > F
Model					
Error					
Total					

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model:
Full model:

Reduced model:
Full Model:

Reduced Model: $\boldsymbol{\mu_o}$ $\boldsymbol{\mu_2}$ $\boldsymbol{\mu_3}$ $\boldsymbol{\mu_o}$

Full model: $\boldsymbol{\mu_1}$ $\boldsymbol{\mu_2}$ $\boldsymbol{\mu_3}$ $\boldsymbol{\mu_4}$

Source	DF	SS	MS	F	Pr > F
Model					
Error					
Total					

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model:

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model:

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Source	DF	SS	MS	F	Pr > F
Model					
Error					
Total					

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\boxed{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model:

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Source	DF	SS	MS	F	Pr > F
Model					
Error					
Total					

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\boxed{\mu_o\ \mu_2\ \mu_3\ \mu_o}$ *recode

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Source	DF	SS	MS	F	Pr > F
Model					
Error					
Total					

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Source	DF	SS	MS	F	Pr > F
Model					
Error					
Total					

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$

Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	SS	MS	F	Pr > F
Model					
Error	196	15964.71695			
Total					

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Subtract to get the df (degrees of freedom) and SS (sum of squares) for top row.

Source	DF	SS	MS	F	Pr > F
Model					
Error	196	15964.71695			
Total	197	17153.10533			

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Subtract to get the df (degrees of freedom) and SS (sum of squares) for top row.

Source	DF	SS	MS	F	Pr > F
Model	1	1188.3883			
Error	196	15964.71695			
Total	197	17153.10533			

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Divide to get the MS column (SS/df) for each row.
You may omit the last row MS calculation.

Source	DF	SS	MS	F	Pr > F
Model	1	1188.3883			
Error	196	15964.71695			
Total	197	17153.10533			

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Divide to get the MS column (SS/df) for each row.
You may omit the last row MS calculation.

Source	DF	SS	MS	F	Pr > F
Model	1	1188.3883	1188.3883		
Error	196	15964.71695	81.45264		
Total	197	17153.10533			

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu \; \mu \; \mu \; \mu$
Full model: $\mu_1 \; \mu_2 \; \mu_3 \; \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu \; \mu \; \mu \; \mu$
Full Model: $\mu_o \; \mu_2 \; \mu_3 \; \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\boldsymbol{\mu_o \; \mu_2 \; \mu_3 \; \mu_o}$

Full model: $\boldsymbol{\mu_1 \; \mu_2 \; \mu_3 \; \mu_4}$

Divide to get the F statistic (MS_{Model}/MS_{Error}).

Source	DF	SS	MS	F	Pr > F
Model	1	1188.3883	1188.3883		
Error	196	15964.71695	81.45264		
Total	197	17153.10533			

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\boldsymbol{\mu_o\ \mu_2\ \mu_3\ \mu_o}$

Full model: $\boldsymbol{\mu_1\ \mu_2\ \mu_3\ \mu_4}$

Divide to get the F statistic (MS_{Model}/MS_{Error}).

Source	DF	SS	MS	F	Pr > F
Model	1	1188.3883	1188.3883	14.59	
Error	196	15964.71695	81.45264		
Total	197	17153.10533			

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$

Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Use software to get the p-value.

```
data quantile;
myquant = 1-CDF('F', 14.59, 1, 196);
run;
proc print data = quantile;
run;
```

Obs	myquant
1	.000179262

Source	DF	SS	MS	F	Pr > F
Model	1	1188.3883	1188.3883	14.59	
Error	196	15964.71695	81.45264		
Total	197	17153.10533			

Compare the contrast (race 1 vs. 4) to building your own F-table (using plain ANOVAs).

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	1914.15805	638.05268	7.83	<.0001
Error	196	15964.71695	81.45264		
Corrected Total	199	17878.87500			

Reduced model: $\mu\ \mu\ \mu\ \mu$
Full Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$ *recode

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	725.76967	362.88484	4.17	0.0169
Error	197	17153.10533	87.07160		
Corrected Total	199	17878.87500			

Reduced Model: $\mu_o\ \mu_2\ \mu_3\ \mu_o$

Full model: $\mu_1\ \mu_2\ \mu_3\ \mu_4$

Use software to get the p-value.

```
data quantile;
myquant = 1-CDF('F', 14.59,1, 196);
run;
proc print data = quantile;
run;
```

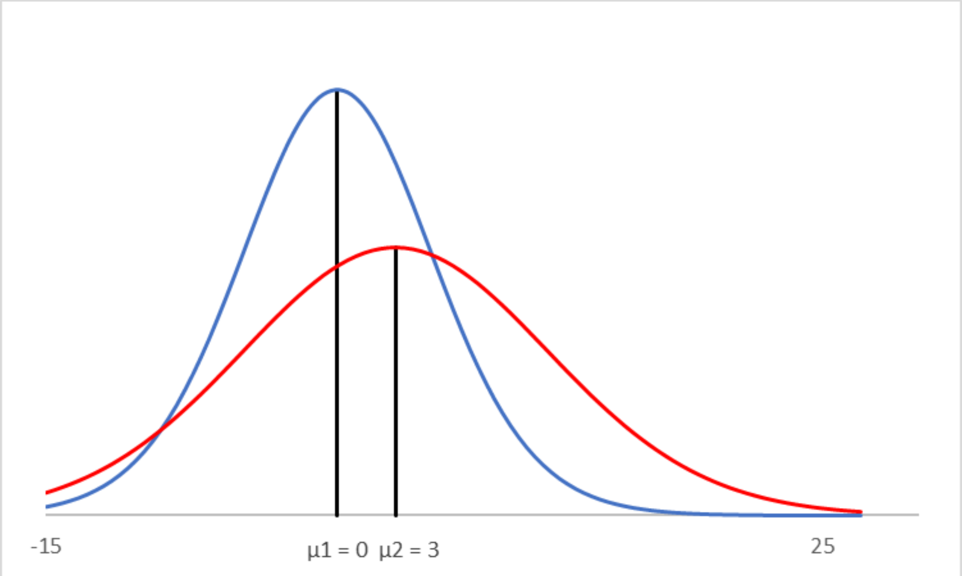
Obs	myquant
1	.000179262

Source	DF	SS	MS	F	Pr > F
Model	1	1188.3883	1188.3883	14.59	.0002
Error	196	15964.71695	81.45264		
Total	197	17153.10533			

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Race 1 v. Race 4	1	1188.388371	1188.388371	14.59	0.0002

5. Use proc power to find the power of a test to detect a difference of 3 in the means if the two groups have standard deviations of 5 and 8, respectively, and sample sizes 20 and 40, respectively. Assume we want to use a Satterthwaite approximation (Welch's test of difference of means assuming different standard deviations).

```
proc power;
      twosamplemeans test=diff_satt
         meandiff = 3
         groupstddevs = 5 | 8
         groupweights = (1 2)
         ntotal = 60
         power = .;
   run;
```



Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Mean Difference	3
Group 1 Standard Deviation	5
Group 2 Standard Deviation	8
Group 1 Weight	1
Group 2 Weight	2
Total Sample Size	60
Number of Sides	2
Null Difference	0
Nominal Alpha	0.05

Computed Power	
Actual Alpha	Power
0.0498	0.415

Beer Prices!!!

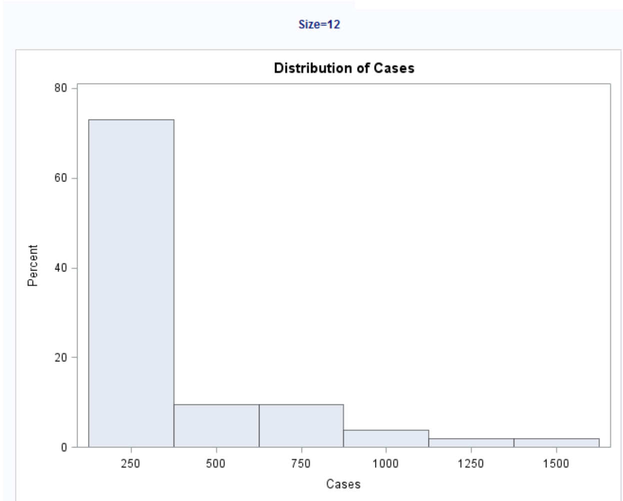
PRICE 18PK	CASES 18PK
14.10	439
18.65	98
18.65	70
18.65	52
18.65	64
18.65	72
18.65	47
18.73	85
18.75	59
18.75	63
18.75	57
18.75	54
13.87	404
14.27	380
18.76	65
18.77	40
13.87	456
14.14	176
18.76	61
18.72	91
18.76	59
18.76	83
18.74	41
18.75	47
18.75	84
18.75	85
18.75	116
13.79	544

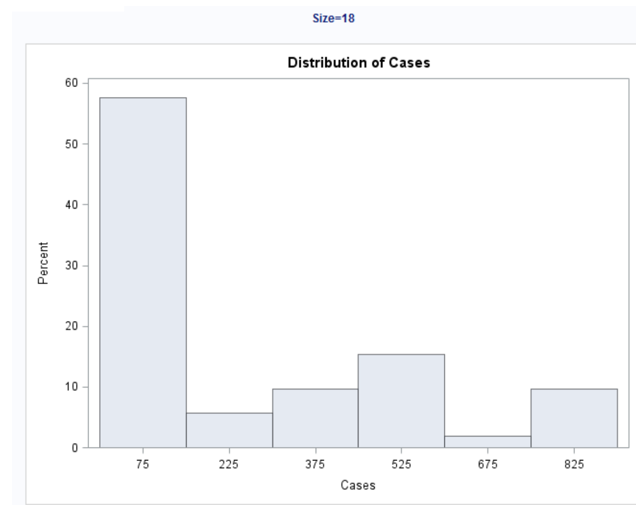
The data file contains 52 weeks of average-price and total-sales records for three different carton sizes: 12-packs, 18-packs, and 30-packs. (This is real data, apart from some very minor adjustments for the 30-packs.) We would like to perform an analysis to detect any differences in the mean or median sales between the 3 sizes of cases.

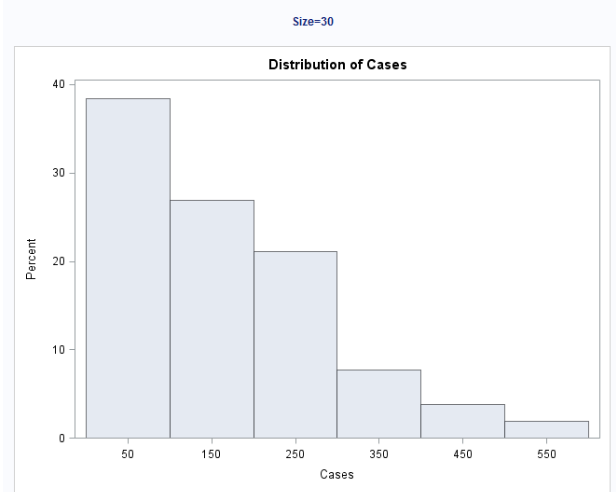
*The data are likely not independent. Sales from week to week are probably related and sales for each size are likely related by week. However, for our purposes, we will assume independence.

Beer Prices!!!


```
proc univariate data = beer;
by type;
histogram cases;
run;
```







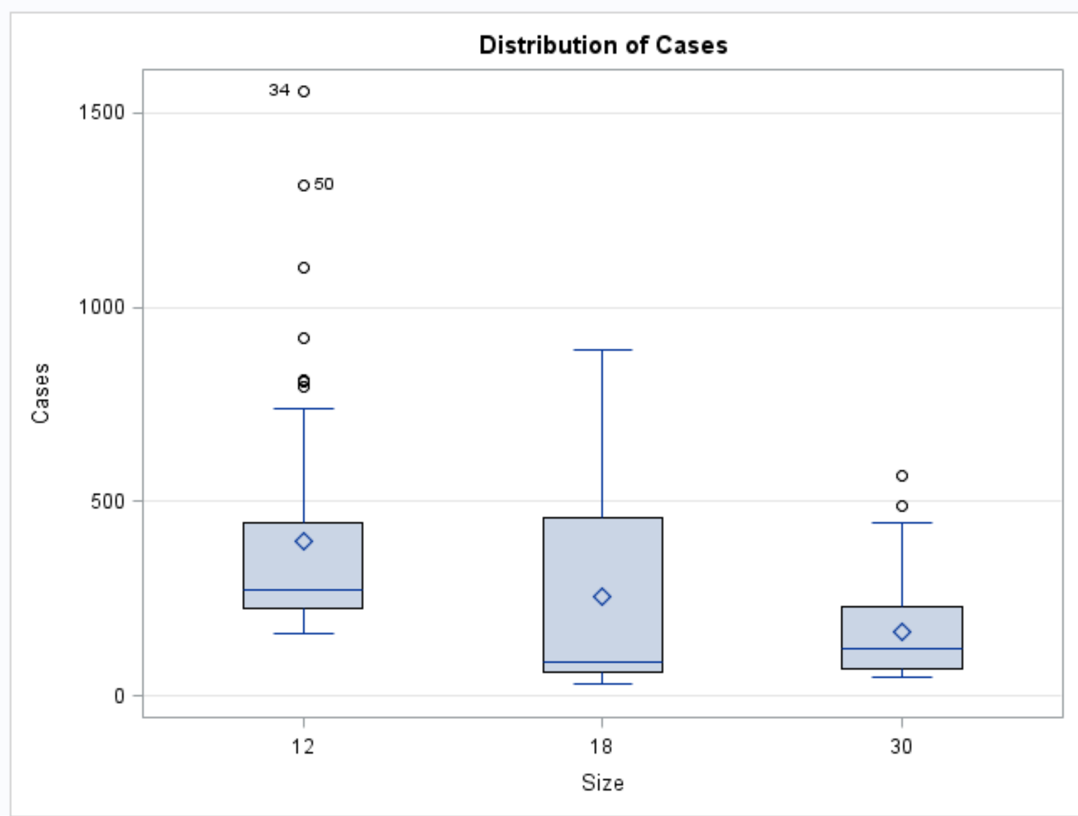
Clearly, there is strong visual evidence that the distribution of the amount of each size case sold is heavily right skewed. Looking closely at the units of the x-axes, we can see that there is strong visual evidence that the standard deviations are quite different.

Beer Prices!!!!

Visual evidence of the right skew and heteroscedasticity is also shown in the box plot:

```
proc glm data = beer;
class type;
model cases = type;
run;
```

A transformation might be able to fix the normality and unequal standard deviations problems.



Analysis

As noted above, there is strong evidence against normality of number of cases sold from the three case sizes and strong evidence that the standard deviations are different as well. However, we will argue that the sample sizes are large enough per group to make Welch's ANOVA and Student t-tests robust to the normality assumption (CLT), and we know that Welch's ANOVA and Student t-tests are robust to the equal standard deviation assumption, especially when sample sizes are similar.* Therefore, we will use these tests to test for differences in means.

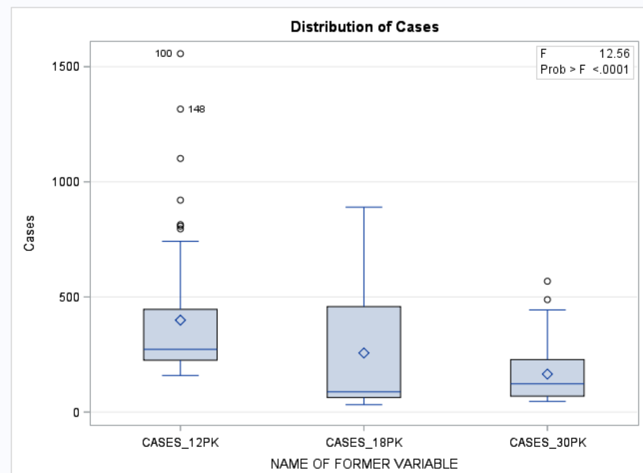
*Even though traditional t-tools are robust to some level of different standard deviations when the sample sizes are similar, traditional ANOVA (with more than two groups) is not nearly as robust to different standard deviations, even when the samples sizes are similar (although equal sample sizes help). Therefore, we will use Welch's ANOVA and traditional t-tests.

Test for any difference? Welch's ANOVA

```
/*Make Long Form */
proc transpose data=beer out=beerlong;
   by f;
run;

data beerwide;
set beer;
keep cases_30pk cases_12pk cases_18pk f
;

/*Make Long Form */
proc transpose data=beerwide out=beerlong;
   by f;
run;

data beerlong;
set beerlong;
rename _NAME_ = Size;
rename Col1 = Cases;
run;

proc glm data = beerlong;
class size;
model Cases = size;
run;
```

Welch's ANOVA for Cases

Source	DF	F Value	Pr > F
Size	2.0000	14.53	<.0001
Error	86.6145		



There is overwhelming evidence that at least 1 pair of case sizes have different mean sales (p-value < .0001). Note that this is NOT a complete analysis.

Test for the Differences: Bonferroni

```
proc glm data = beerlong;
class size;
model Cases = size;
means Size / bon cldiff;
run;
```

There is overwhelming evidence that the 12 packs have greater mean sales than both 18 packs and 30 packs. A 95% confidence interval for the difference is (28.55, 256.44) cases for the 18 packs and (120.18, 348.07) cases for the 30 packs. Note that this is not a complete 6 step analysis.

Bonferroni (Dunn) t Tests for Cases

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than Tukey's for all pairwise comparisons.

Alpha	0.05
Error Degrees of Freedom	153
Error Mean Square	57613.37
Critical Value of t	2.42059
Minimum Significant Difference	113.95

Comparisons significant at the 0.05 level are indicated by *.**

Size Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
CASES_12PK - CASES_18PK	142.49	28.55	256.44	***
CASES_12PK - CASES_30PK	234.12	120.18	348.07	***
CASES_18PK - CASES_12PK	-142.49	-256.44	-28.55	***
CASES_18PK - CASES_30PK	91.63	-22.32	205.57	
CASES_30PK - CASES_12PK	-234.12	-348.07	-120.18	***
CASES_30PK - CASES_18PK	-91.63	-205.57	22.32	

Alternative solution

We know that equality among variances is a crucial assumption of ANOVA; so, we could be conservative with our assumptions and seek a non-parametric solution: Kruskal-Wallis. Given the difference in standard deviations, the shapes cannot be said to be the same, so we will have to limit our tests to differences of distribution.

```
proc npar1way data =beer;
class type;
var cases;
run;
```

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable Cases Classified by Variable Size

Size	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
12	52	5530.50	4082.0	265.994971	106.355769
18	52	3579.50	4082.0	265.994971	68.836538
30	52	3136.00	4082.0	265.994971	60.307692

Average scores were used for ties.

Kruskal-Wallis Test	
Chi-Square	30.5811
DF	2
Pr > Chi-Square	<.0001

1. $H_0: F_{twelve} = F_{eighteen} = F_{thirty}$
   $H_A: F_i \neq F_j$ for some i, j
2. $Critical\ Value: can\ skip$
3. $Test\ Statistic: \chi^2 = 30.5811$
4. $P - value < .0001$
5. $Reject\ H_0$

6. There is strong evidence at the alpha = .05 level of significance (p-value < .0001 from the Kruskal-Wallis test) that at least one of the distributions is different from the other two. Three separate rank sum tests will be conducted with a Bonferroni multiple comparison adjustment in order to test for these differences. Could also do three separate signed-rank tests or sign tests and pair the data by week.

```
proc npar1way data =beer wilcoxon;
where type = 12|type = 18;
class type;
var cases;
run;
```

Beer Prices!!!

Wilcoxon Scores (Rank Sums) for Variable Cases Classified by Variable Size

Size	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
18	52	2765.0	2730.0	153.814382	53.173077
30	52	2695.0	2730.0	153.814382	51.826923

Average scores were used for ties.

Wilcoxon Two-Sample Test

Statistic	2765.0000
Normal Approximation	
Z	0.2243
One-Sided Pr > Z	0.4113
Two-Sided Pr > |Z|	0.8225
t Approximation	
One-Sided Pr > Z	0.4115
Two-Sided Pr > |Z|	0.8230

Z includes a continuity correction of 0.5.

Wilcoxon Scores (Rank Sums) for Variable Cases Classified by Variable Size

Size	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
12	52	3267.50	2730.0	153.814792	62.836538
18	52	2192.50	2730.0	153.814792	42.163462

Average scores were used for ties.

Wilcoxon Two-Sample Test

Statistic	3267.5000
Normal Approximation	
Z	3.4912
One-Sided Pr > Z	0.0002
Two-Sided Pr > |Z|	0.0005
t Approximation	
One-Sided Pr > Z	0.0004
Two-Sided Pr > |Z|	0.0007

Z includes a continuity correction of 0.5.

Wilcoxon Scores (Rank Sums) for Variable Cases Classified by Variable Size

Size	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
12	52	3641.0	2730.0	153.816844	70.019231
30	52	1819.0	2730.0	153.816844	34.980769

Average scores were used for ties.

Wilcoxon Two-Sample Test

Statistic	3641.0000
Normal Approximation	
Z	5.9194
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001
t Approximation	
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001

Z includes a continuity correction of 0.5.

Since there are three simultaneous tests, we will adjust the alpha level down to .05/3 =0.0167. OR, we can adjust the p-values using a Bonferroni adjustment: .8225 * 3 > 1, .0005* 3=.0015, and .0001 * 3 = .0003. Therefore at the alpha = .05 family wise significance level, there is strong evidence that the distribution of 12 pack cases sold is different than the distribution of both 18 and 30 packs sold (Bonferroni adjusted p-values = .0015 and < .0003, respectively.) Note that this page is NOT a complete analysis.