A Closer Look at Assumptions

ONE LAST TIME: CONFIDENCE INTERVAL/HYPOTHESIS TEST DUALITY

ROBUSTNESS OF T-TOOLS

Examining Confidence Intervals

To the “Rossman/Chance” applet:

http://www.rossmanchance.com/applets/ConfSim.html

Confidence Intervals & Hypothesis Tests

H_0: $\mu_I - \mu_E = 0$

We formed a 95% confidence interval that doesn't contain zero

We did a (two-sided) hypothesis test for the difference in means equaling zero:
0.0054 = p-value < α = 0.05

```
PROC TTEST DATA=creativity ORDER=DATA SIDES=2;
    CLASS intrinsic;
    VAR SCORE;
RUN;
```

The TTEST Procedure
Variable: score

intrinsic	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	24	19.8833	4.4395	0.9062	12.0000	29.7000
0	23	15.7391	5.2526	1.0952	5.0000	24.0000
Diff (1-2)		4.1442	4.8541	1.4164		

intrinsic	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		19.8833	18.0087	21.7580	4.4395	3.4504	6.2276
0		15.7391	13.4677	18.0105	5.2526	4.0623	7.4343
Diff (1-2)	Pooled	4.1442	1.2914	6.9970	4.8541	4.0261	6.1138
Diff (1-2)	Satterthwaite	4.1442	1.2776	7.0108			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	45	2.93	0.0054
Satterthwaite	Unequal	43.108	2.92	0.0056

QUESTION:
Is this a coincidence?

Answer: Not At All!

Using these ingredients…

Let's get a more precise p-value

The TTEST Procedure
Variable: score

intrinsic	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	24	19.8833	4.4395	0.9062	12.0000	29.7000
0	23	15.7391	5.2526	1.0952	5.0000	24.0000
Diff (1-2)		$\boxed{4.1442}$	4.8541	$\boxed{1.4164}$		

intrinsic	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		19.8833	18.0087	21.7580	4.4395	3.4504	6.2276
0		15.7391	13.4677	18.0105	5.2526	4.0623	7.4343
Diff (1-2)	Pooled	4.1442	1.2914	6.9970	4.8541	4.0261	6.1138
Diff (1-2)	Satterthwaite	4.1442	1.2776	7.0108			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	45	$\boxed{2.93}$	0.0054
Satterthwaite	Unequal	43.108	2.92	0.0056

```
DATA prob;
    pval = 2*(1-CDF('T',4.1442/1.4164,45));
RUN;
```

Obs	pval
1	.005366587

Confidence Intervals & Hypothesis Tests

H_0: $\mu_I - \mu_E = 0$

Obs	pval
1	.005366587

```
PROC TTEST DATA=creativity ORDER=DATA SIDES=2 ALPHA=0.005366587;
    CLASS intrinsic;
    VAR SCORE;
RUN;
```

We set α = 0.0053…
To form a 100(1- α)% Confidence Interval

It barely excludes 0!

The TTEST Procedure
Variable: score

intrinsic	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	24	19.8833	4.4395	0.9062	12.0000	29.7000
0	23	15.7391	5.2526	1.0952	5.0000	24.0000
Diff (1-2)		4.1442	4.8541	1.4164		

intrinsic	Method	Mean	99.46334% CL Mean		Std Dev	99.46334% CL Std Dev	
1		19.8833	17.0975	22.6691	4.4395	3.1264	7.2908
0		15.7391	12.3564	19.1219	5.2526	3.6742	8.7436
Diff (1-2)	Pooled	4.1442	0.000011	8.2884	4.8541	3.7372	6.7843
Diff (1-2)	Satterthwaite	4.1442	-0.0243	8.3127			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	45	2.93	0.0054
Satterthwaite	Unequal	43.108	2.92	0.0056

The Take Away

100(1-α)% Confidence Intervals are

EQUIVALENT

to Hypothesis Tests with p-value = α

(Note: we are referencing two-sided version of both here)

EQUIVALENT: The confidence interval will contain the null hypothesis value if and only if $\alpha <$ p-value

EXAMPLE:

A 95% confidence interval for the mean is equivalent to a hypothesis test with p-value = 0.05

Assumptions for T-Tools: Normality

Assumptions of the One Sample T-Tools

1. Samples are drawn from a Normally distributed population.
2. The observations in the sample are independent of one another.

(CENTRAL LIMIT THEOREM)

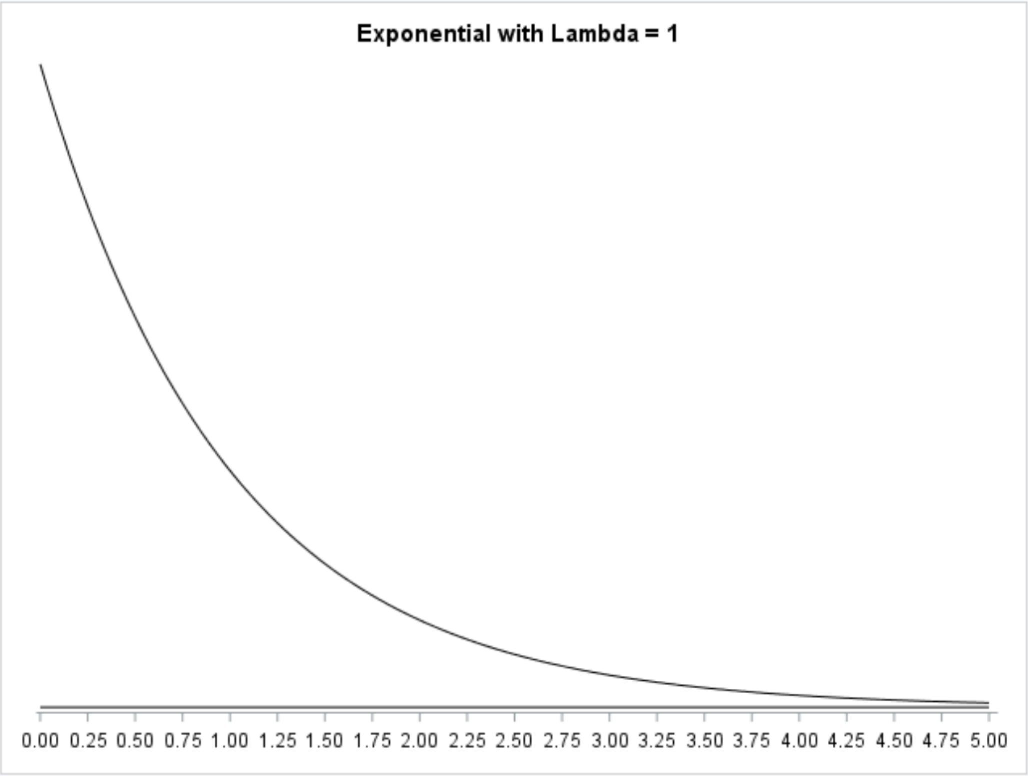
When the original (population) distribution is not normal, the one sample t-test is still valid with a large enough sample size.

That is, the one sample t-test is **ROBUST** to the normality assumption when the sample size is large enough.

ROBUST means that even though the assumption isn't necessarily met, the procedure can (but not always) work fine

Robustness of T-tools

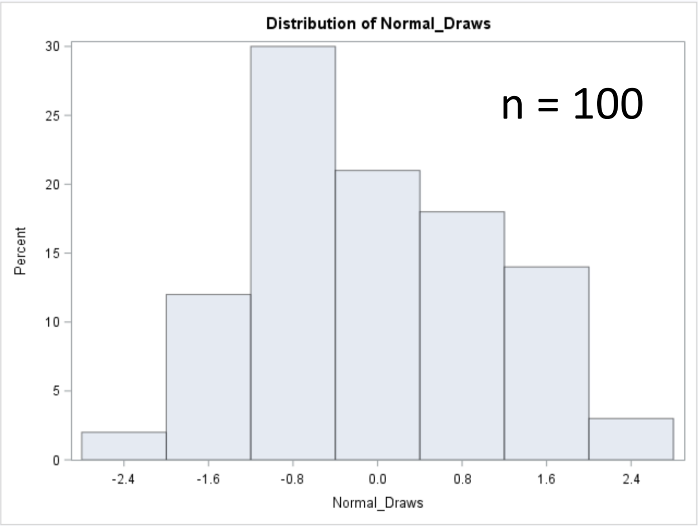
$$PDF\left('EXPO', x, \lambda\right) = \begin{cases} 0 & x < 0 \\ \frac{1}{\lambda} exp\left(-\frac{x}{\lambda}\right) & x \geq 0 \end{cases}$$



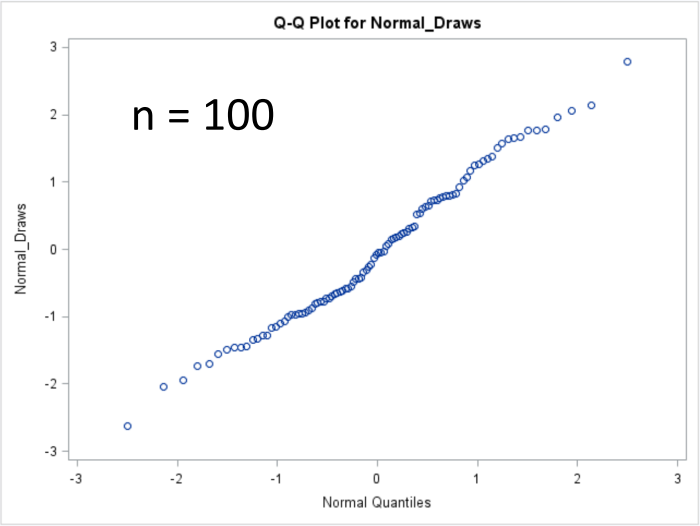
Back To the “Rossman/Chance” applet:

http://www.rossmanchance.com/applets/ConfSim.html

Given Data, How Do We Check the Normality Assumption?



Histogram



QQ Plot

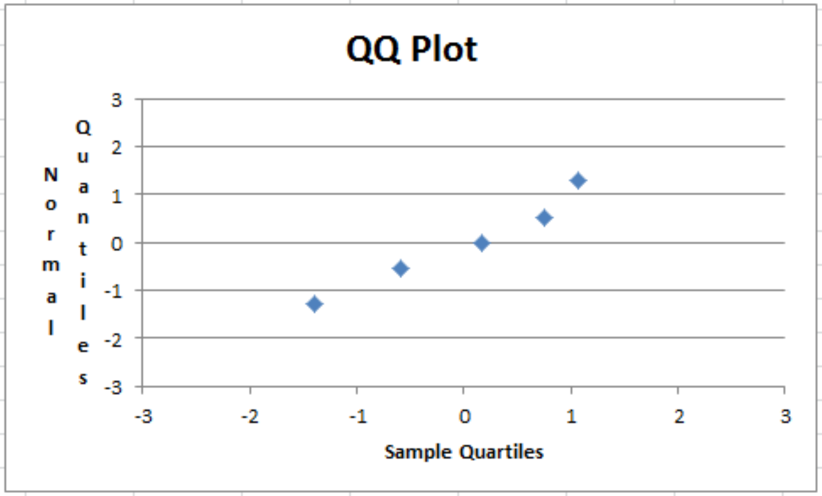
```
/* Generate Normal Random Draws */
data Normal(keep = Normal_Draws);
call streaminit(14);
do i = 1 to 100;
Normal_Draws = rand("Normal");
output;
end;
run;
```

```
proc univariate data = Normal;
var Normal_Draws;
histogram Normal_Draws;
qqplot Normal_Draws;
run;
```

Creating QQ plots

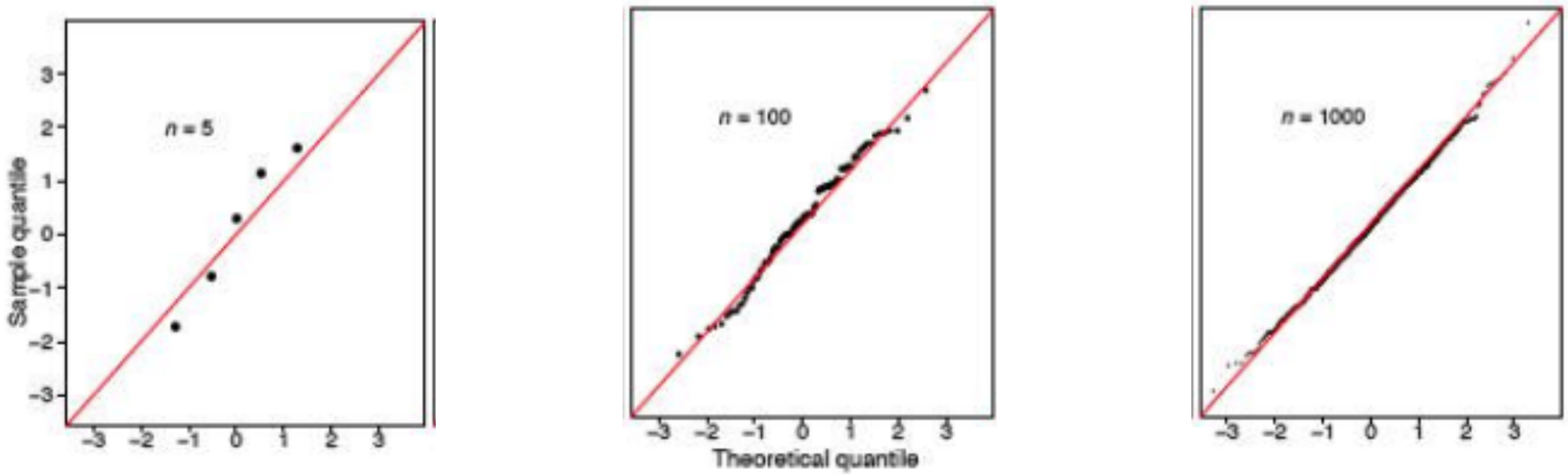
DATA
41.2
76.6
109.3
134.5
148.6

	Data z	Rank	Middle	Normal z	
41.2	-1.39367	1	0.1	-1.281551566	46.0944
76.6	-0.58276	2	0.3	-0.524400513	79.14751
109.3	0.166306	3	0.5	0	102.04
134.5	0.743564	4	0.7	0.524400513	124.9325
148.6	1.066555	5	0.9	1.281551566	157.9856

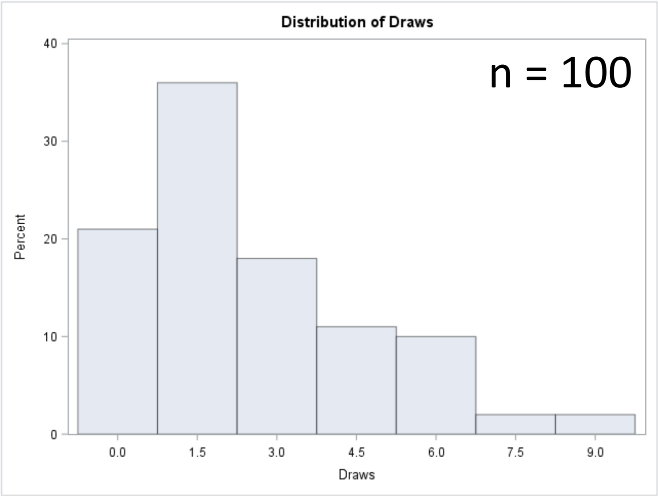


Normal QQ Plot

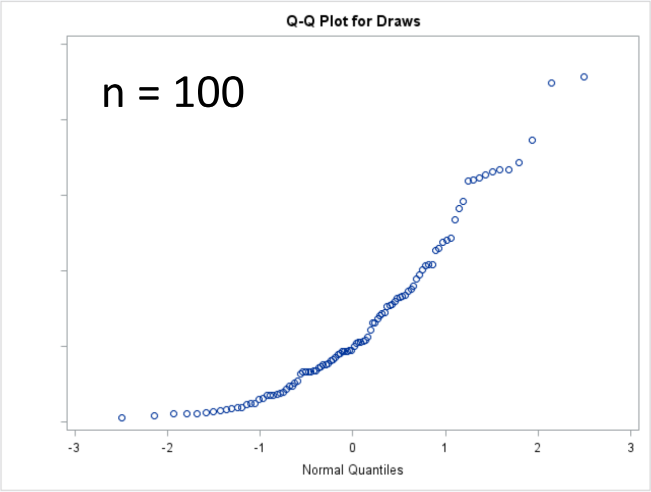
If the sample comes from a normal distribution, it will have similar quantiles as the normal



Given Data, How Do We Check the Normality Assumption?



Histogram

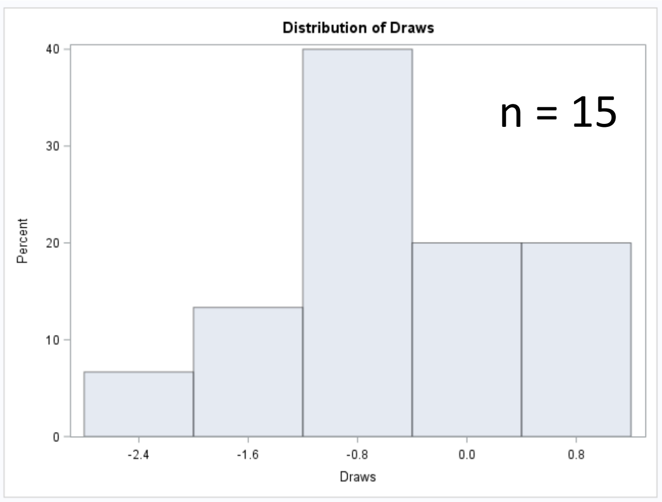


QQ Plot

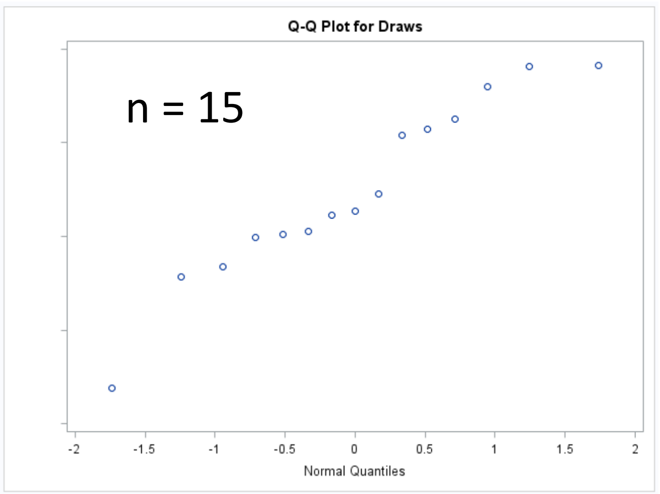
What about in this case?

Exponential distribution

Given Data, How Do We Check the Normality Assumption?



Histogram



QQ Plot

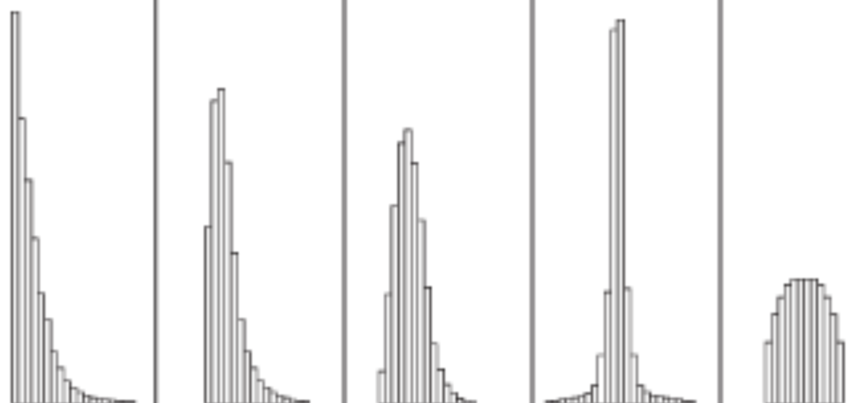
What about in this case?

```
/* Generate Normal Random Draws */
data Normal(keep = Draws);
call streaminit(14);
do i = 1 to 15;
Draws = rand("NORMAL");
output;
end;
run;
```

A Way to Decide:

	Small Sample Size	Large Sample Size
Little to no Evidence Against Normality	No Problem if you feel Normality is a safe assumption … run the T-Test. (You may want to be “conservative” here and run a test with less assumptions.	No Problem! Run the T-Test
Significant Evidence Against Normality	Assumptions are not met and test is not robust here … Do not run the T-Test and proceed to a test with less / different assumptions.	No Problem .. You have the CLT. Run the T-Test.

DISPLAY 3.4 Percentage of 95% confidence intervals that are successful when the two populations are non-normal (but with same shape and SD, and equal sample sizes) (each percentage is based on 1,000 computer simulations)



Sample size	Strongly skewed	Moderately skewed	Mildly skewed	Long-tailed	Short-tailed
5	95.5	95.4	95.2	98.3	94.5
10	95.5	95.4	95.2	98.3	94.6
25	95.3	95.3	95.1	98.2	94.9
50	95.1	95.3	95.1	98.1	95.2
100	94.8	95.3	95.0	98.0	95.6

Assumptions for T-Tools: Equal Variances

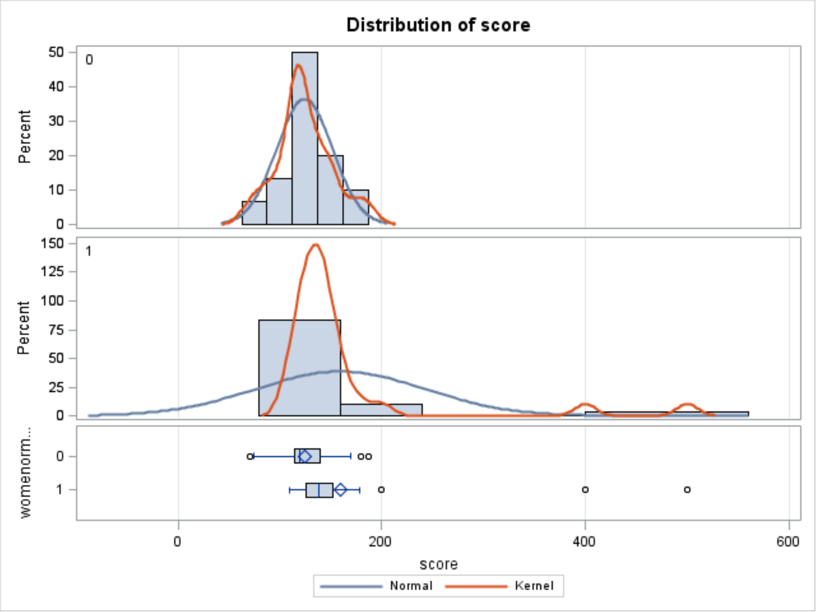
Assumptions of two sample T-Tools

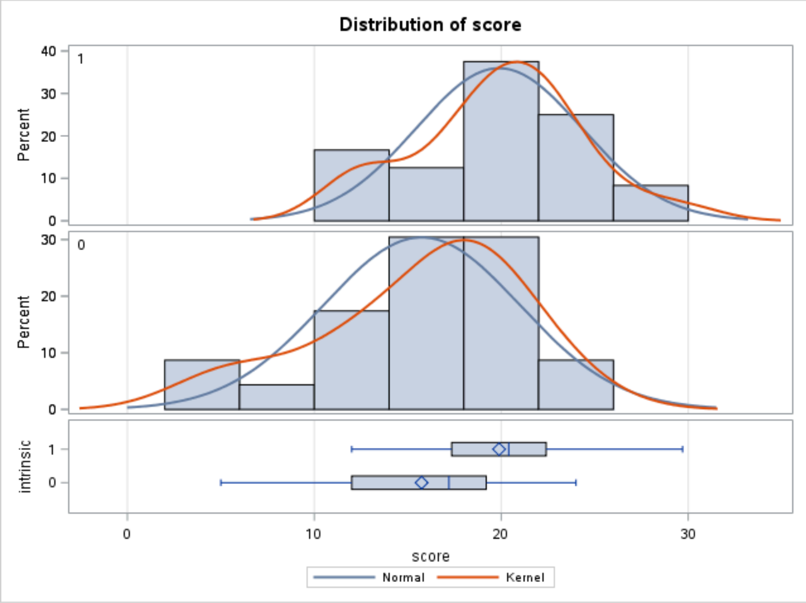
1. Samples are drawn from a normally distributed population
2. The observations in the sample are independent of one another
3. If it is a two sample test, both populations are assumed to have the same standard deviation

There are two standard diagnoses:

- Look at histograms of the groups
- Do a hypothesis test for equal variances

Looking at Histograms





Test for Equal Variances

H_0: population variances are equal

H_A: population variances are not equal

The TTEST Procedure
Variable: score

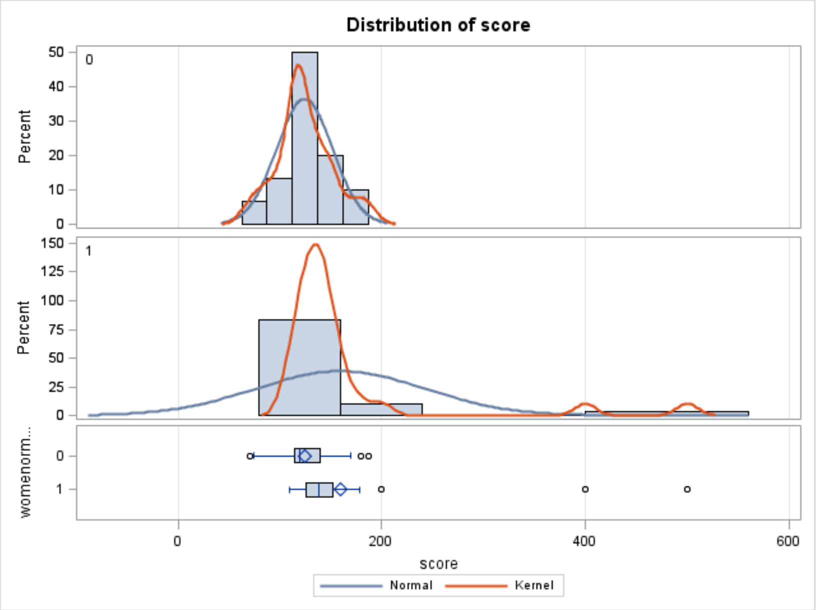
intrinsic	N	Mean	Std Dev	Std Err	Minimum	Maximum
1	24	19.8833	4.4395	0.9062	12.0000	29.7000
0	23	15.7391	5.2526	1.0952	5.0000	24.0000
Diff (1-2)		4.1442	4.8541	1.4164		

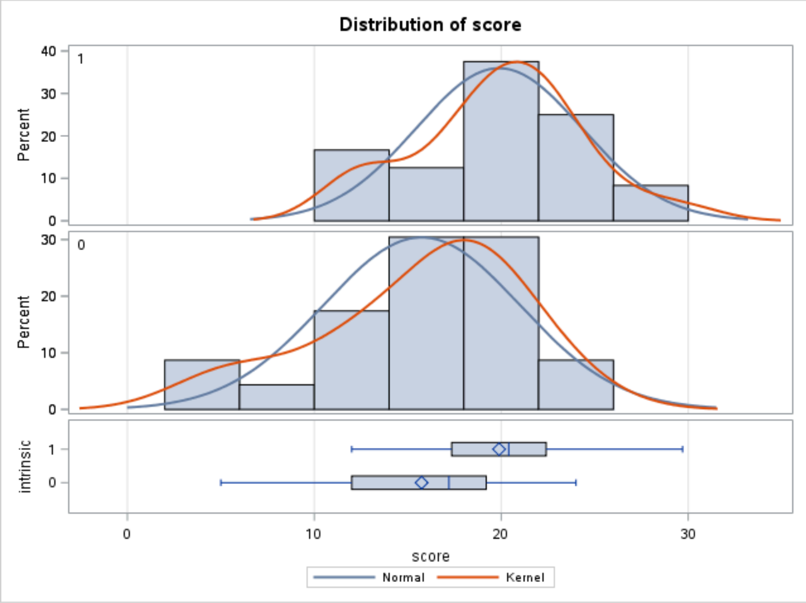
intrinsic	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		19.8833	18.0087	21.7580	4.4395	3.4504	6.2276
0		15.7391	13.4677	18.0105	5.2526	4.0623	7.4343
Diff (1-2)	Pooled	4.1442	1.2914	6.9970	4.8541	4.0261	6.1138
Diff (1-2)	Satterthwaite	4.1442	1.2776	7.0108			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	45	2.93	0.0054
Satterthwaite	Unequal	43.108	2.92	0.0056

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	22	23	1.40	0.4289

Looking at Histograms & Test for Equal Variances





Equality of Variances				
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Folded F	29	29	9.08	<.0001

Equality of Variances				
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Folded F	22	23	1.40	0.4289

What happens if we detect unequal variances?

It no longer makes sense to "pool" the sample standard deviation estimators together

We need to make some sort of change

This could mean a transformation or a different method