Exercise 1a

Variable: length

Moments			
N	210	**Sum Weights**	210
Mean	5.62853333	**Sum Observations**	1181.992
Std Deviation	0.44306348	**Variance**	0.19630525
Skewness	0.52548156	**Kurtosis**	-0.7856445
Uncorrected SS	6693.90917	**Corrected SS**	41.0277963
Coeff Variation	7.87173943	**Std Error Mean**	0.03057428

Basic Statistical Measures			
Location		**Variability**	
Mean	5.628533	**Std Deviation**	0.44306
Median	5.523500	**Variance**	0.19631
Mode	5.236000	**Range**	1.77600
		Interquartile Range	0.71800

Note: The mode displayed is the smallest of 2 modes with a count of 3.

Seed lengths center around 5.6 (mean = 5.629; median = 5.524), with spread (SD = 0.443; CV ≈ 7.9%). The distribution shows a slight right skew (skewness = +0.525; mean > median) and kurtosis = –0.786. The middle 50% of lengths lie between 5.26 and 5.98 (IQR = 0.718), and overall values range from about 4.90 to 6.68 (range = 1.776).

Exercise 1b

Basic descriptives for LENGTH by VARIETY

Variable: length
variety = Canadian

Moments			
N	70	**Sum Weights**	70
Mean	5.22951429	**Sum Observations**	366.066
Std Deviation	0.1380152	**Variance**	0.0190482
Skewness	0.05119272	**Kurtosis**	-0.4784569
Uncorrected SS	1915.6617	**Corrected SS**	1.31432549
Coeff Variation	2.63915906	**Std Error Mean**	0.01649597

Basic Statistical Measures			
Location		**Variability**	
Mean	5.229514	**Std Deviation**	0.13802
Median	5.224000	**Variance**	0.01905
Mode	5.236000	**Range**	0.64200
		Interquartile Range	0.18900

Basic descriptives for LENGTH by VARIETY

Variable: length
variety = Kama

Moments			
N	70	**Sum Weights**	70
Mean	5.50805714	**Sum Observations**	385.564
Std Deviation	0.23150803	**Variance**	0.05359597
Skewness	-0.3473746	**Kurtosis**	-0.0908666
Uncorrected SS	2127.40667	**Corrected SS**	3.69812177
Coeff Variation	4.20307966	**Std Error Mean**	0.0276705

Basic Statistical Measures			
Location		**Variability**	
Mean	5.508057	**Std Deviation**	0.23151
Median	5.534000	**Variance**	0.05360
Mode	5.395000	**Range**	1.15100
		Interquartile Range	0.29400

Basic descriptives for LENGTH by VARIETY

Variable: length
variety = Rosa

Moments			
N	70	**Sum Weights**	70
Mean	6.14802857	**Sum Observations**	430.362
Std Deviation	0.26819115	**Variance**	0.07192649
Skewness	-0.2537146	**Kurtosis**	0.26275797
Uncorrected SS	2650.8408	**Corrected SS**	4.96292794
Coeff Variation	4.36223002	**Std Error Mean**	0.03205497

Basic Statistical Measures			
Location		**Variability**	
Mean	6.148029	**Std Deviation**	0.26819
Median	6.148500	**Variance**	0.07193
Mode	5.979000	**Range**	1.31200
		Interquartile Range	0.33600

Note: The mode displayed is the smallest of 3 modes with a count of 2.

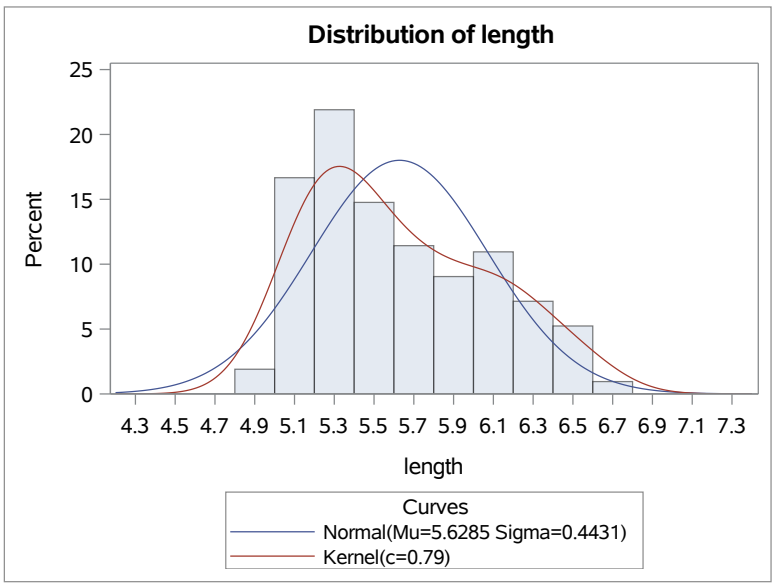
Canadian seeds are the shortest (mean ≈ 5.23, median ≈ 5.224), with the smallest spread (SD ≈ 0.138, IQR ≈ 0.189, range ≈ 0.64) and no skew (skew ≈ +0.05). Kama seeds are mid-length (mean ≈ 5.51, median ≈ 5.534) with some variability (SD ≈ 0.232, IQR ≈ 0.294, range ≈ 1.15) and a slight left skew (skew ≈ −0.35). Rosa seeds are the longest and most variable (mean ≈ 6.15, median ≈ 6.149; SD ≈ 0.268, IQR ≈ 0.336, range ≈ 1.31) with a mild left skew (skew ≈ −0.25). Canadian seeds have the smallest variantion and Rosa seeds have the widest spread.
Exercise 2a

Normality checks for LENGTH

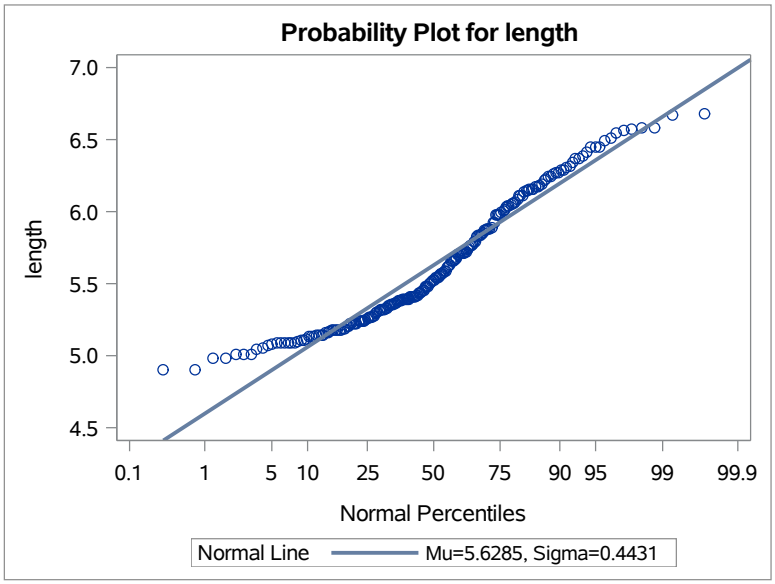
Variable: length

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.943799	**Pr < W**	<0.0001
Kolmogorov-Smirnov	**D**	0.119154	**Pr > D**	<0.0100
Cramer-von Mises	**W-Sq**	0.655958	**Pr > W-Sq**	<0.0050
Anderson-Darling	**A-Sq**	3.891567	**Pr > A-Sq**	<0.0050

Normality checks for LENGTH



Normality checks for LENGTH



Visually, the histogram of length is roughly bell-shaped and shows a right tail, and the Q-Q plot has a curvature away from the straight line. Quantitatively, every formal normality test strongly rejects normality: Shapiro-Wilk $W = 0.9438$, $p < 0.0001$; Kolmogorov-Smirnov $D = 0.1192$, $p < 0.01$; Cramér-von Mises $W^2 = 0.6560$, $p < 0.005$; Anderson–Darling $A^2 = 3.8916$, $p < 0.005$. Therefore, the normality assumption is not reasonable for length.

Exercise 2b

Normality checks for LENGTH by VARIETY

Variable: length
variety = Canadian

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.990179	**Pr < W**	0.8631
Kolmogorov-Smirnov	**D**	0.066585	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.040448	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.253382	**Pr > A-Sq**	>0.2500

Normality checks for LENGTH by VARIETY

Variable: length
variety = Kama

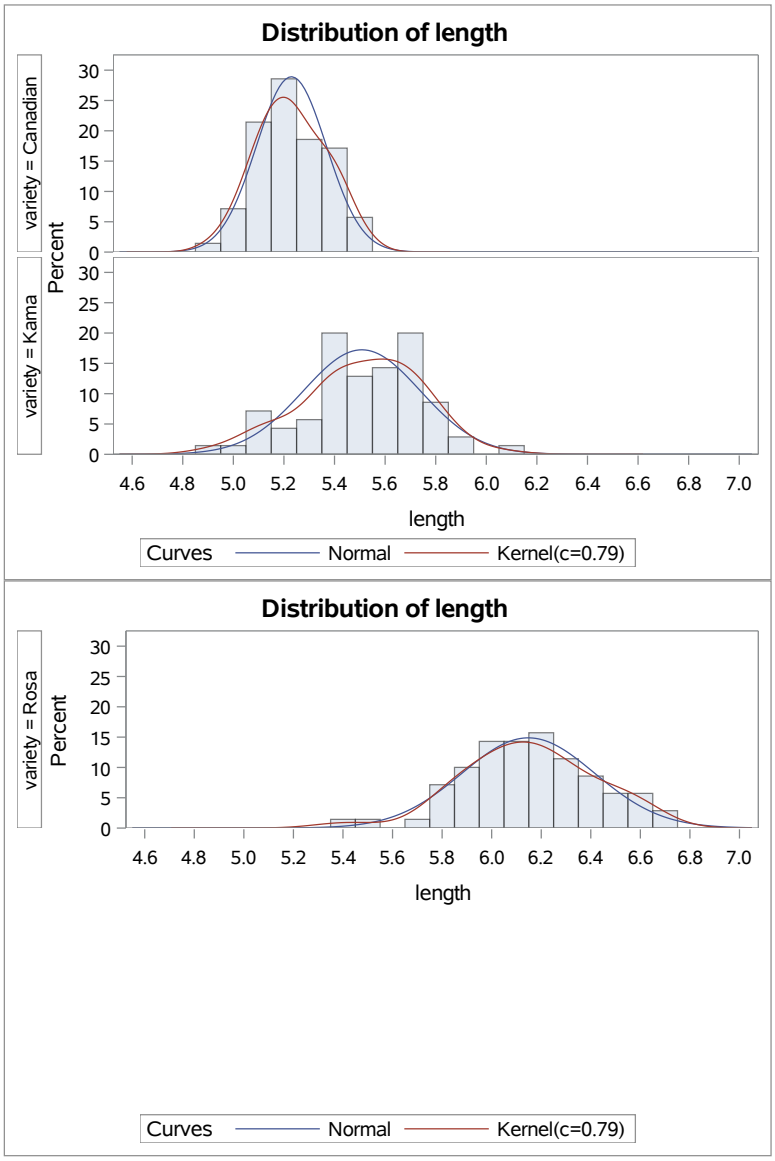
Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.984427	**Pr < W**	0.5367
Kolmogorov-Smirnov	**D**	0.067173	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.06146	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.408654	**Pr > A-Sq**	>0.2500

Normality checks for LENGTH by VARIETY

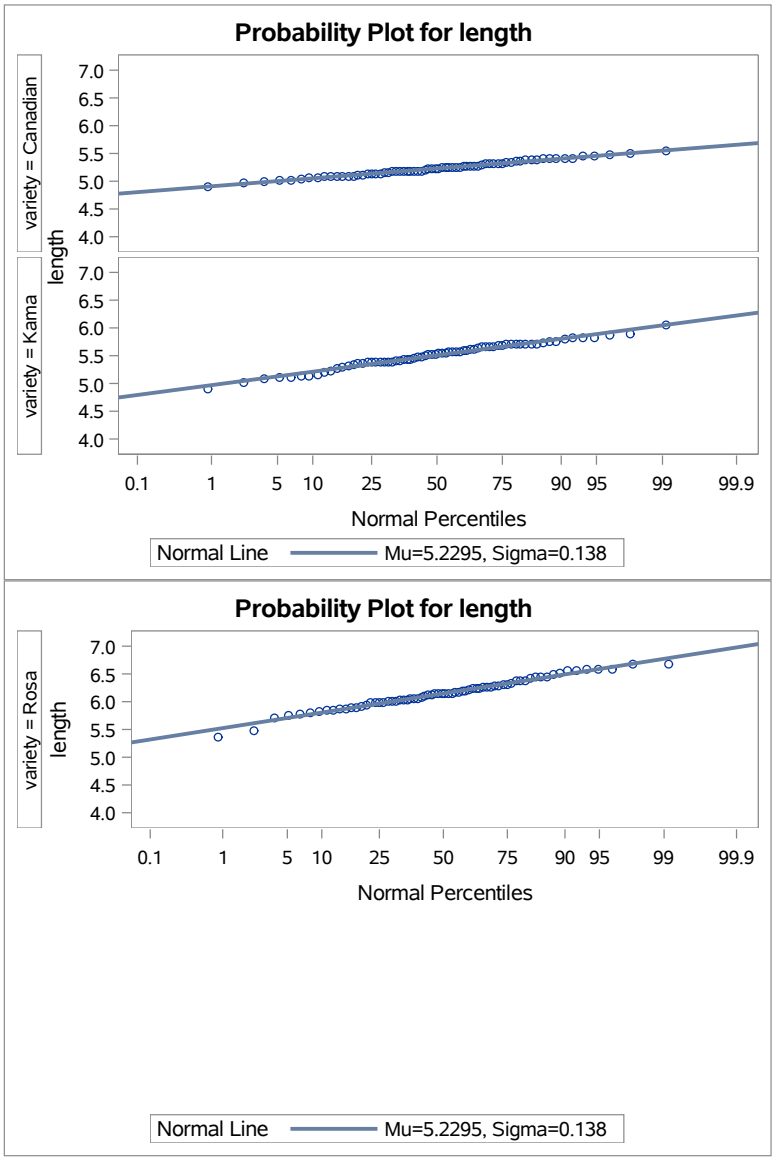
Variable: length
variety = Rosa

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.984142	**Pr < W**	0.5210
Kolmogorov-Smirnov	**D**	0.04876	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.01902	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.20045	**Pr > A-Sq**	>0.2500

Normality checks for LENGTH by VARIETY



Normality checks for LENGTH by VARIETY



None of the three groups show a statistically significant deviation from normality. For Canadian seeds, all tests are nonsignificant-Shapiro-Wilk p=0.863, K-S p>0.15, CvM/AD p>0.25 and a straight Q-Q line. Kama seeds have nonsignificant tests-Shapiro-Wilk p=0.537, K-S p>0.15, CvM/AD p>0.25 and the histogram/Q-Q look bell-shaped. Rosa seeds also have nonsignificant tests—Shapiro-Wilk p=0.521, K-S p>0.15, CvM/AD p>0.25 and a linear Q-Q plot. Each variety's seed lengths are consistent with normality.

Exercise 3a

One-sample location test for LENGTH

Variable: length

Tests for Location: Mu0=5.5				
Test	**Statistic**		**p Value**	
Student's t	**t**	4.203969	**Pr > |t|**	<.0001
Sign	**M**	3	**Pr >= |M|**	0.7302
Signed Rank	**S**	2691.5	**Pr >= |S|**	0.0021

Since the distribution is a bit symmetric, we will use Wilcoxon signed-rank. The p-value = 0.0021 which is << 0.05. Therefore, the tests are highly significant and the median value differs from the hypothesized value. Since the test statistic is positive, the median is slightly greater than 5.5.
Exercise 3b

Test to check if Rosa is shorter than Kama

Variable: length

variety	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Kama		5.5081	5.4529	5.5633	0.2315	0.1985	0.2778
Rosa		6.1480	6.0841	6.2120	0.2682	0.2300	0.3218
Diff (1-2)	**Pooled**	-0.6400	-0.7101	Infty	0.2505	0.2241	0.2840
Diff (1-2)	**Satterthwaite**	-0.6400	-0.7101	Infty			

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	138	-15.11	1.0000
Satterthwaite	Unequal	135.12	-15.11	1.0000

Equality of Variances				
Method	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Folded F	69	69	1.34	0.2243

The lengths are roughly normal within each variety. The two-sample t-test shows Rosa seeds are longer on average: Rosa=6.15 and Kama=5.51. The test gives t= -15.11 with a one-sided p-value of 1.000. Therefore, we reject the claim and conclude Rosa seeds are longer than Kama seeds.
Exercise 4a

Correlation analysis for the area, compactness, and length variables for all of the data

3 Variables:	area compactness length

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**
area	210	14.84752	2.90970	14.35500	10.59000	21.18000
compactness	210	0.87100	0.02363	0.87345	0.80810	0.91830
length	210	5.62853	0.44306	5.52350	4.89900	6.67500

Pearson Correlation Coefficients, N = 210 Prob > |r| under H0: Rho=0			
	area	**compactness**	**length**
area	1.00000	0.60829 <.0001	0.94999 <.0001
compactness	0.60829 <.0001	1.00000	0.36792 <.0001
length	0.94999 <.0001	0.36792 <.0001	1.00000

Spearman Correlation Coefficients, N = 210 Prob > |r| under H0: Rho=0			
	area	**compactness**	**length**
area	1.00000	0.63850 <.0001	0.93092 <.0001
compactness	0.63850 <.0001	1.00000	0.37808 <.0001
length	0.93092 <.0001	0.37808 <.0001	1.00000

Area and length are perfectly correlated (Pearson r≈0.95, Spearman r≈0.93; p<.0001), longer seeds have larger area. Area and compactness show a moderate positive link (r≈0.61), and length and compactness have a weak positive link (r≈0.37). As seed length goes up, area goes up a lot, and compactness tends to increase a bit
Exercise 4b

Correlation analysis by variety

variety=Canadian

3 Variables:	area compactness length

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**
area	70	11.87386	0.72300	11.83500	10.59000	13.37000
compactness	70	0.84941	0.02176	0.84935	0.80810	0.89770
length	70	5.22951	0.13802	5.22400	4.89900	5.54100

Pearson Correlation Coefficients, N = 70 Prob > |r| under H0: Rho=0			
	area	**compactness**	**length**
area	1.00000	0.54676 <.0001	0.51660 <.0001
compactness	0.54676 <.0001	1.00000	-0.37000 0.0016
length	0.51660 <.0001	-0.37000 0.0016	1.00000

Spearman Correlation Coefficients, N = 70 Prob > |r| under H0: Rho=0			
	area	**compactness**	**length**
area	1.00000	0.53005 <.0001	0.47883 <.0001
compactness	0.53005 <.0001	1.00000	-0.36143 0.0021
length	0.47883 <.0001	-0.36143 0.0021	1.00000

Correlation analysis by variety

variety=Kama

3 Variables:	area compactness length

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**
area	70	14.33443	1.21570	14.35500	11.23000	17.08000
compactness	70	0.88007	0.01619	0.88050	0.83920	0.91830
length	70	5.50806	0.23151	5.53400	4.90200	6.05300

Pearson Correlation Coefficients, N = 70 Prob > |r| under H0: Rho=0			
	area	**compactness**	**length**
area	1.00000	0.37104 0.0016	0.83478 <.0001
compactness	0.37104 0.0016	1.00000	-0.14630 0.2268
length	0.83478 <.0001	-0.14630 0.2268	1.00000

Spearman Correlation Coefficients, N = 70 Prob > |r| under H0: Rho=0			
	area	**compactness**	**length**
area	1.00000	0.34061 0.0039	0.79428 <.0001
compactness	0.34061 0.0039	1.00000	-0.18310 0.1292
length	0.79428 <.0001	-0.18310 0.1292	1.00000

Correlation analysis by variety

variety=Rosa

3 Variables:	area compactness length

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Median**	**Minimum**	**Maximum**
area	70	18.33429	1.43950	18.72000	15.38000	21.18000
compactness	70	0.88352	0.01550	0.88260	0.84520	0.91080
length	70	6.14803	0.26819	6.14850	5.36300	6.67500

Pearson Correlation Coefficients, N = 70 Prob > |r| under H0: Rho=0			
	area	**compactness**	**length**
area	1.00000	0.27263 0.0224	0.82643 <.0001
compactness	0.27263 0.0224	1.00000	-0.21695 0.0712
length	0.82643 <.0001	-0.21695 0.0712	1.00000

Spearman Correlation Coefficients, N = 70 Prob > |r| under H0: Rho=0			
	area	**compactness**	**length**
area	1.00000	0.19515 0.1055	0.85062 <.0001
compactness	0.19515 0.1055	1.00000	-0.17674 0.1433
length	0.85062 <.0001	-0.17674 0.1433	1.00000

Within each variety, area and length move together strongly (bigger seeds -> larger area), just like in the full data. Area vs. compactness is a weak–moderate positive (strongest in Canadian, weakest in Rosa). Length vs. compactness is not positive inside varieties-it's near zero for Kama/Rosa and negative for Canadian.