ANALİZ İLE İLGİLİ TEMEL KAVRAMLAR

Frekans/Sıklık (frequency): Bir değişkene ait veri setinde bir ölçümün tekrarlanma sayısını gösterir. Bu sayı yüzde olarak da ifade edilebilir. Frekans bir araştırmada toplanan verilerin genel olarak betimlenmesinde kullanılan bir niceliktir. Tablo veya grafiklerle gösterilebilir. Frekansı belirlenecek değişken süreksiz ise sütun (bar) veya pasta (pie) grafiği ile sürekli ise histogram grafiği ile gösterilebilir.

2.1. MERKEZİ EĞİLİM ÖLÇÜTLERİ

Aritmetik Ortalama/Ortalama (Mean): Sürekli değişken olan bir veri setindeki bir değişkene ait tüm değerlerin toplamının veri sayısına bölünmesi ile bulunur. Aritmetik ortalama evren için μ, örneklem için $\bar{X}$ ile sembolize edilir. Bir örnekleme ait $X_1, X_2, X_3,\ldots X_n$ gibi bir veri seti için aritmetik ortalama şu denklem ile hesaplanır:

$$\bar{X} = \frac{\sum X}{n}$$

Ortanca (Median): Küçükten büyüğe sıralanmış bir veri setinde tam ortada yer alan değerdir. Veri sayısı tek sayı ise $(n+1)/2$ sırasındaki veri değeri ortancadır. Veri sayısı çift sayı ise $n/2$ sırasındaki ve bir sonraki sıradaki sayıların toplamının 2'ye bölünmesi ile elde edilen sayı ortancadır.

Tepe Değeri (Mode): Bir veri setinde en çok çıkan yani frekansı en yüksek olan veri değeridir.

2.2. ORTALAMADAN SAPMA ÖLÇÜTLERİ

Varyans (Variance): Bir örneklem için varyans, verilerin ortalamadan sapmalarının kareleri toplamının, toplam veri sayısının 1 eksiğine bölünmesi ile bulunur. Varyans evren için σ^2, örneklem için S^2 ile sembolize edilir. Bir örnekleme ait $X_1, X_2, X_3,\ldots X_n$ gibi bir veri seti için için varyans şu denklem ile hesaplanır:

$$S^2 = \frac{\sum(X-\bar{X})^2}{n-1}$$

Standart Sapma (Standard deviation): Ölçümlerin ortalamadan ne kadar uzaklaştığını gösterir, varyansın pozitif karekökünе eşittir. Standart sapma evren için σ, örneklem için S ile sembolize edilir. Bir örnekleme ait $X_1, X_2, X_3,\ldots X_n$ gibi bir veri seti için için standart sapma şu denklem ile hesaplanır:

$$S = \sqrt{\frac{\sum(X-\bar{X})^2}{n-1}}$$

Dağılım Genişliği (Range): Bir veri setinde en büyük değer ile en küçük değer arasındaki farktır.

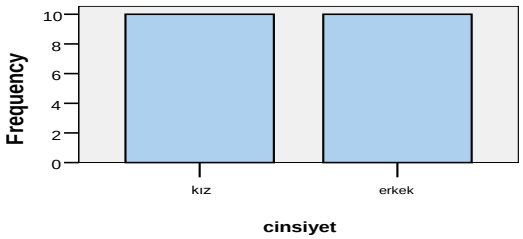
Merkezi eğilim ve ortalamadan sapma ölçüleri hesaplanacak değişkenin sürekli olması gerekmektedir.

Uygulama-3: Deneme-1.sav dosyasında araştırmaya katılan öğrencilerin cinsiyet ve öğrenim durumuna göre frekans dağılımlarını **(a)** tablo ile, **(b)** sütun grafiği ile ve **(c)** pasta grafiği ile gösterip, **(d)** bu öğrencilerin yaşlarının frekansa göre dağılımını da bir histogram grafiğini çiziniz.

(a) ve **(b)** Süreksiz değişken olan cinsiyet ve öğrenim durumuna göre frekans dağılımlarının tablo ve grafikleri şu adımlar takip edilerek aşağıdaki gibi elde edilir. (Bu çıktı dosyası Uygulama-3.spo olarak kaydedilebilir.)
Analyze → Descriptive Statistics → Frequencies (açılan pencerede cinsiyet ve öğrenim durumu değişkenleri sağ ekrana aktarılır.) → Charts (açılan pencerede Bar seçeneği işaretlenir.) → Continue → OK.

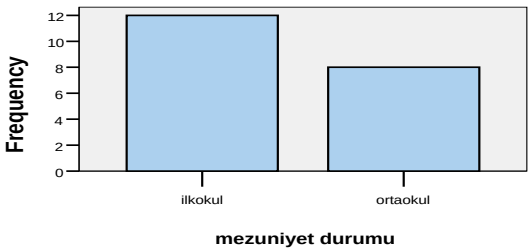
cinsiyet

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	kiz	10	50,0	50,0	50,0
	erkek	10	50,0	50,0	100,0
	Total	20	100,0	100,0	

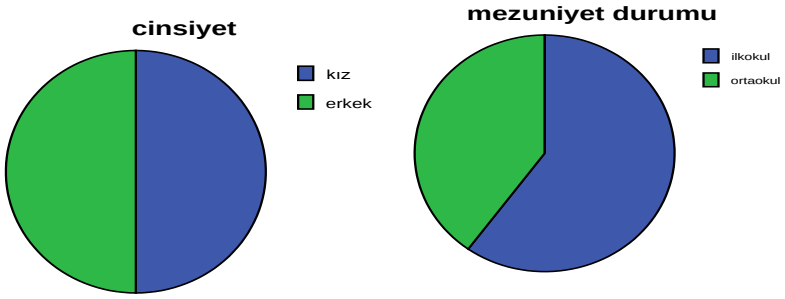


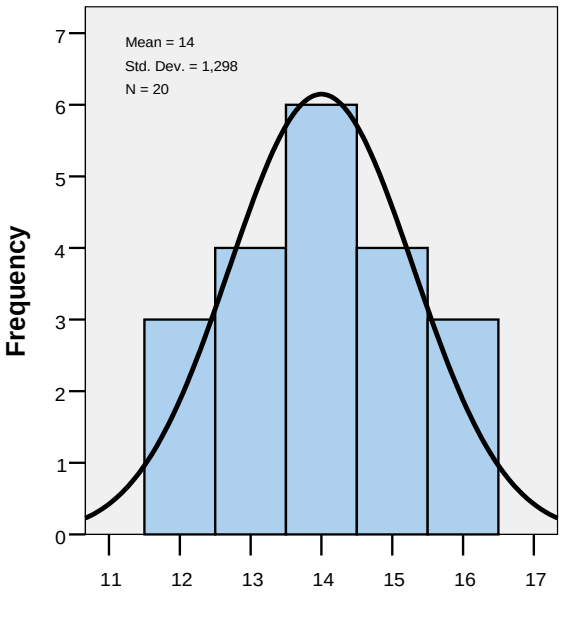
mezuniyet durumu

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ilkokul	12	60,0	60,0	60,0
	ortaokul	8	40,0	40,0	100,0
	Total	20	100,0	100,0	



(c) Pasta grafiği için şu adımlar takip edilerek, cinsiyet ve öğrenim durumu için pasta grafiklerinin çıktıları şekildeki gibi elde edilir:
Analyze → Descriptive Statistics → Frequencies (açılan pencerede cinsiyet ve öğrenim durumu değişkenleri sağ ekrana aktarılır.) → Charts (açılan pencerede Pie seçeneği işaretlenir.) → Continue → OK





(d) Sürekli değişken olan öğrenci yaşlarının histogram grafiği için şu adımlar takip edilir şekildeki gibi elde edilir:
Analyze → Descriptive Statistics → Frequencies (açılan pencerede yaş değişkenleri sağ ekrana aktarılır.) → Charts (açılan pencerede Histogram seçeneği "with normal curve" tıklanarak işaretlenir.) → Continue → OK.

Uygulama-4: Deneme-1.sav dosyasında öğrencilerin **(a)** cinsiyet ve eğitim durumuna (ikili) ve **(b)** cinsiyet, eğitim durumu ve yaşa (üçlü) göre frekans dağılımlarını çapraz tabloda gösteriniz. (Katılımcıların nominal/ordinal iki veya üç değişkene ait kişisel özellikleri betimlenmek istendiğinde çapraz tablo kullanılabilir)

(a) Cinsiyet ve eğitim durumuna (ikili) çapraz tablo için şu adımlar takip edilir:
Analyze → Descriptive Statistics → Crosstabs (açılan pencerede Row(s) kısmına öğrenim durumu, Column(s) kısmına da cinsiyet aktarılır.) → Cells (örneğin "Row" işaretlenir.) →Continue →OK.

mezuniyet durumu * cinsiyet Crosstabulation

			cinsiyet		
			kiz	erkek	Total
mezuniyet durumu	ilkokul	Count	5	7	12
		% within mezuniyet durumu	41,7%	58,3%	100,0%
	ortaokul	Count	5	3	8
		% within mezuniyet durumu	62,5%	37,5%	100,0%
Total		Count	10	10	20
		% within mezuniyet durumu	50,0%	50,0%	100,0%

(b) Cinsiyet, eğitim durumu ve yaşa göre üçlü çapraz tablo için öncelikle sürekli bir değişken olan yaşı süreksiz hale getirmeliyiz. Bunun için Recode (nto different variable) komutu kullanılarak (Uygulama-2'deki adımlar takip edilerek) reyas sütunu oluşturulur. Bu sütunda 14 yaş ve altında olanlar "1" ile 14 yaş üstü olanlar "2" olarak kodlanır. Daha sonra aşağıdaki adımlar takip edilerek üçlü çapraz tablo elde edilir.
Analyze → Descriptive Statistics → Crosstabs(açılan pencerede Row(s) kısmına öğrenim durumu, Column(s) kısmına cinsiyet ve Layer kısmına da reyas aktarılır.) → Cells (örneğin "Row" işaretlenir.) → Continue → OK.

mezuniyet durumu * cinsiyet * reyas Crosstabulation

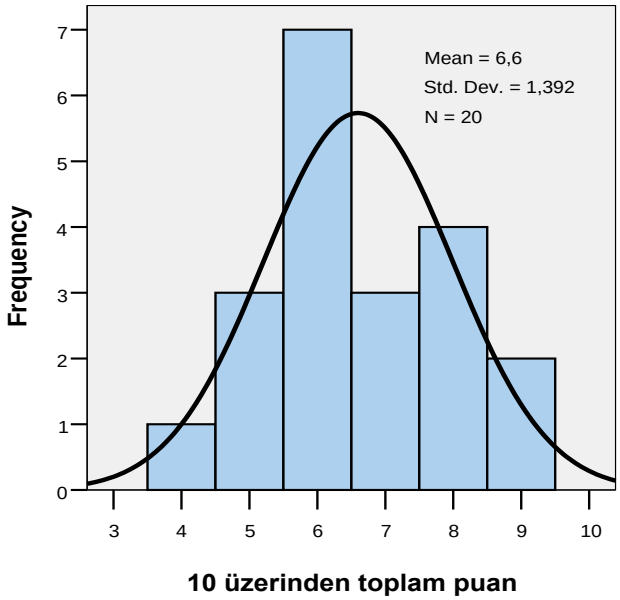
				cinsiyet		
reyas				kiz	erkek	Total
14 yas ve alti	mezuniyet durumu	ilkokul	Count	5	4	9
			% within mezuniyet durumu	55,6%	44,4%	100,0%
		ortaokul	Count	1	3	4
			% within mezuniyet durumu	25,0%	75,0%	100,0%
	Total		Count	6	7	13
			% within mezuniyet durumu	46,2%	53,8%	100,0%
14 yas üstü	mezuniyet durumu	ilkokul	Count	0	3	3
			% within mezuniyet durumu	,0%	100,0%	100,0%
		ortaokul	Count	4	0	4
			% within mezuniyet durumu	100,0%	,0%	100,0%
	Total		Count	4	3	7
			% within mezuniyet durumu	57,1%	42,9%	100,0%

Uygulama-5: Deneme-1.sav dosyasında araştırmaya katılan öğrenci puanlarının **(a)** frekansa göre dağılımını gösteren histogram grafiği ve **(b)** merkezi eğilim ve ortalamadan sapma ölçülerini gösteren tabloyu bulunuz.

Öğrenci puanlarının frekansa göre dağılımını gösteren histogram grafiği, merkezi eğilim ve ortalamadan sapma ölçülerini gösteren tabloyu elde etmek için şu adımlar takip edilir:
Analyze → Descriptive Statistics → Frequencies (açılan pencerede yaş değişkenleri sağ ekrana aktarılır.) → Charts (açılan pencerede Histogram seçeneği "with normal curve" tıklanarak işaretlenir.) Continue → Statistics (açılan pencerede Mean, Mode, Median, Sum, Std. Deviation, Variance, Range, Minimum, Maximum, S.E. Mean seçenekleri işaretlenir.) → Continue →OK.

10 üzerinden toplam puan

N	Valid	20
	Missing	0
Mean		6,60
Std. Error of Mean		,311
Median		6,00
Mode		6
Std. Deviation		1,392
Variance		1,937
Range		5
Minimum		4
Maximum		9
Sum		132



Merkezi eğilim ve ortalamadan sapma ölçütleri hesaplanacak değişkenin sürekli olması gerekmektedir. Sürekli bir değişkenden elde edilen veriler için bir diğer önemli ölçüt ise normallikten sapma ölçütleridir.

2.3. NORMALLİKTEN SAPMA ÖLÇÜLERİ

İstatistik analizlerde pek çok testin uygulanabilmesi için veri dağılımının normal veya normale yakın olması gerekir. Normal (Gauss) dağılım sürekli bir dağılımdır. Normal dağılım eğrisi simetriktir ve şekli çan eğrisine benzer. Bu eğri şu şekilde elde edilir: Bir veri setindeki değerlerin (x_1, x_2, x_3,….x_n) her biri için $f(x) = \frac{1}{S\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\bar{x}}{S}\right)^2}$ "olasılık yoğunluk fonksiyonu" denkleminden (S=standart sapma $\bar{x}$ ortalamadır) yararlanılarak $f(x)$ değerleri bulunur. x-$f(x)$ grafiği çizildiğinde normal dağılım eğrisi elde edilmiş olur. SPSS programı histogram grafiğinin üzerine oturtulmuş bir şekilde bu eğriyi otomatik olarak gösterir. Ortalama, ortanca, mod, çarpıklık (Skewnes) ve basıklık (Kurtosis) normal dağılımın değerlendirilebileceği ölçütlerdir. Ç***arpıklık (Skewnes)*** dağılımın ortalama etrafındaki simetriden ne kadar saptığını gösterir. ***Basıklık (Kurtosis)***, verilerin tepe noktalarının durumunun normalden sapması (basık veya sivri olması) hakkında bilgi verir. Bu ölçütler yorumlanarak bir veri seti dağılımının normal bir dağılım olup olmadığı hakkında bir fikir edinilebilir:

1.) Ortalama = Ortanca = Mod ise dağılım normaldir. Normalden sapma arttıkça bu değerler bir birinden uzaklaşır.
2.) Çarpıklık Katsayısı (ÇK) = **0** ise dağılım normaldir. ÇK ± 1 sınırları arasında ise dağılımın normal dağılımdan önemli ölçüde sapmadığı şeklinde yorumlanır. ÇK **negatif** ise **sola** çarpıklık vardır ve ortalama ortancadan küçüktür. ÇK **pozitif** ise **sağa** çarpıklık vardır ve ortalama ortancadan büyüktür.

3.) Basıklık katsayısı (B.K) = **0** ise dağılım normaldir. BK. **negatifse** eğri normale göre daha basıktır. BK. **pozitifse** eğri normale göre daha sivridir.

Histogram grafiği ve gövde-yaprak (Steam &Leaf) tablosu, normal Q-Q grafiği, detrended Q-Q grafiği ve kutu grafiği (box plot) veri dağılımının normalliği hakkında fikir verecek yöntemlerdir. Ancak verilerin normal dağılımı hakkında daha kesin hüküm verebilmek için iki türlü test yöntemi kullanılır:

1.) *Kolomogrow-Simirnov ve Shapiro-Wilk normallik testleri:* Veri sayısı 29'dan az olduğunda Shapiro-Wilk testi, 29'dan fazla olduğunda ise Kolomogrow-Simirnov testi kullanılır. Bu test SPSS programında tablo olarak elde edilir. Tablodaki Sig. Değeri 0.05'den büyük olduğu durumda %5 anlamlılık düzeyinde verilerin normal dağılıma sahip olduğu sonucuna ulaşılır.

2.) *Çarpıklık katsayısının standartlaştırılması:* Çarpıklık katsayısı standart hatasına oranlanarak elde edilen standart çarpıklık değeri(SÇD) bir kritik değerle karşılaştırılır. Eğer -1.96 < SÇD < 1.96 ise .05 düzeyinde dağılımın normal olduğu; eğer -2.58 < SÇD < 2.68 ise .01 düzeyinde dağılımın normal olduğu sonucuna ulaşılır. (Burası "3. Test İstatistiği" bölümünde daha anlaşılır hale gelecektir)

Uygulama-6: Yandaki tabloda 40 öğrencinin dakika cinsinden bir metni okuma hızı verilmiştir. Deneme-2.sav dosyasına bu verileri girip okuma hızı dağılımının normal bir dağılım olup olmadığını **a)** merkezi eğilim, ortalamadan ve normallikten sapma ölçütlerini gösteren tabloya **b)** histogram grafiği ve gövde-yaprak tablosuna **c)** Q-Q grafiklerine **d)** kutu grafiğine ve **e)** Kolomogrow-Simirnov/Shapiro-Wilk testine göre belirleyiniz.

Öğrenci	Okuma Hızı (dk)
1	2,135
2	1,766
3	1,696
4	2,078
5	2,095
6	2,068
7	3,391
8	4,054
9	4,854
10	4,569
11	4,450
12	5,305
13	4,518
14	5,571
15	6,010
16	7,959
17	14,709
18	16,215
19	15,446
20	15,420
21	18,706
22	15,706
23	13,966
24	13,370
25	12,632
26	10,850
27	13,038
28	8,248
29	8,937
30	10,185
31	8,292
32	7,523
33	11,867
34	10,380
35	10,704
36	9,415
37	9,379
38	7,126
39	9,349
40	11,134

a) Analyze → Descriptive Statistics → Explore (açılan pencerede "okuma hizi" değişkeni Depend kısmınaa aktarılır.) → Statistics (açılan pencerede Descriptives ve Outliers işaretlenir.) → Continue → Plot (açılan pencerede Factor Levels Together, Steam And Leaf ve Histogram işaretlenir)→ OK.

			Statistic	Std. Error
okuma hizi	Mean		8,6279	,74651
	95% Confidence Interval for Mean	Lower Bound	7,1179	
		Upper Bound	10,1379	
	5% Trimmed Mean		8,5204	
	Median		8,6145	
	Variance		22,291	
	Std. Deviation		4,72133	
	Minimum		1,70	
	Maximum		18,71	
	Range		17,01	
	Interquartile Range		7,91	
	Skewness		,204	,374
	Kurtosis		-,940	,733

-1 < ÇK < +1 olduğundan dağılım kabaca normal kabul edilebilir. Ancak ÇK değeri pozitif ve Ortalama > Ortanca olduğundan SAĞA doğru bir çarpıklık vardır. BK negatif olduğundan normal eğrisi normalden biraz basıktır. Daha kesin hüküm verebilmek için standart çarpıklık değeri hesaplanır: SÇD=.204/.374=.545 bulunur. -1.96 < .545 < 1.96 olduğu için verilerin normal dağılmıştır.

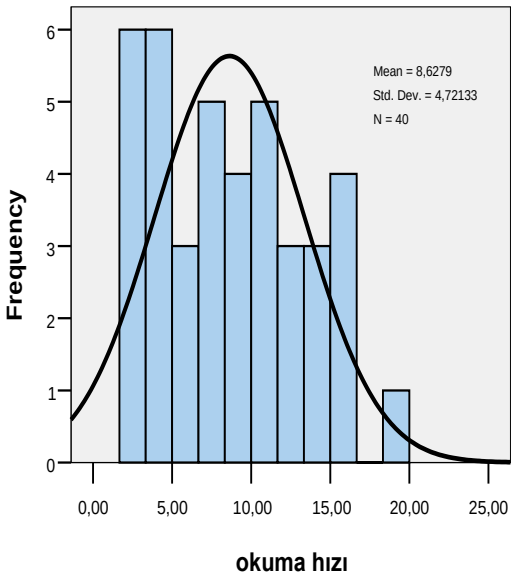
b) İlk şıktaki uygulamalar yapıldığında analiz sayfasında yukarıdaki tabloya ek olarak histogram grafiği, gövde yaprak diyagramı ve Q-Q grafkleri de vardır. Ancak histogramın normal eğrisi grafik üzerinde görülmediğinden Uygulama-1 (d)'deki gibi adımları takip ederek histogram

grafiğini normal eğrisi ile birlikte görebiliriz. Bu verilerin gövde-yaprak tablosu ve normal eğrili histogram grafiği aşağıda verilmiştir.

okuma hızıı Stem-and-Leaf Plot___
Frequency Stem & Leaf

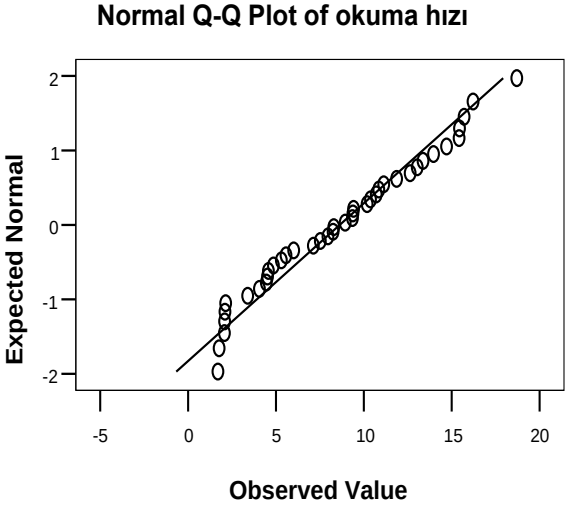
2,00	0 . 11	(2 adet 11'li veri var demek)
5,00	0 . 22223	(4 adet 2'li veri, 1 adet 3'lü veri var)
7,00	0 . 4444455	(5 adet 4'lü veri, 2 adet 5'li veri var)
4,00	0 . 6777	(1 adet 6'lı veri, 3 adet 7'li veri var)
6,00	0 . 888999	(3 adet 8'li veri, 3 adet 9'lu veri var)
6,00	1 . 000011	(4 adet 10'lu veri, 2 adet 11'li veri var)
4,00	1 . 2333	(1 adet 12'li veri, 3 adet 13'lü veri var)
4,00	1 . 4555	(1 adet 14'lü veri, 3 adet 15'li veri var)
1,00	1 . 6	(1 adet 16'lı veri var)
1,00	1 . 8	(1 adet 18'li veri var)

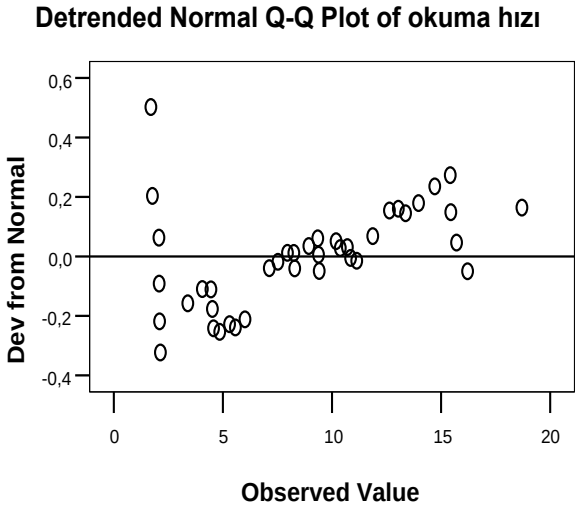
Stem width: 10,00
Each leaf: 1 case(s)



"Okuma hızı"na ait verilerin frekansı gövde yaprak tablosunda histograma göre daha detaylı bilgi veririr.

c) İlk şıktaki uygulamalar yapıldığında analiz sayfasında Normal ve detrended normal Q-Q grafkleri aşağıdaki gibidir:

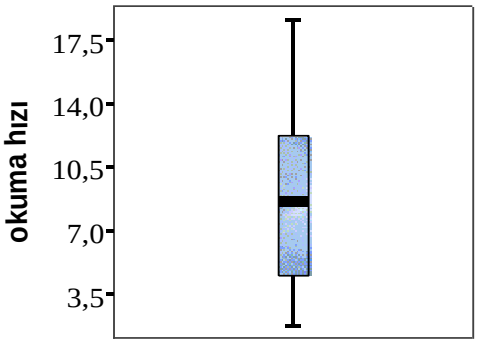




Eğer üzerinde çalışılan örneklem normal dağılım veya normale yakınsa Normal Q-Q grafiğinde değerlerin bir doğru üzerinde veya etrafında yakın olarak toplanması gerekir. Okuma hızına ilişkin verilerin normale yakın bir dağılım gösterdiği grafikten gözükmektedir. Pozitif çarpıklık katsayısı (sağa çarpıklık) durumunda normal Q-Q grafiğinde noktalar doğru üstünde yay şeklinde birikir, negatif çarpıklık katsayısı (sola çarpıklık) durumunda normal Q-Q grafiğinde noktalar doğru altında yay şeklinde birikir.

Detrended normal Q-Q grafiğinde normal dağılım şartı ise verilerin yatay çizgi etrafında bir fonksiyon biçimini oluşturmadan rastgele dağılmasıdır.

d) Graphs→ Interactive→Boxplot (açılan pencerede y eksenine molekül genişliğini aktar) → Tamam.
Kutu ortasındaki çizgi alt ve üst çizginin tam ortasında kalırsa normal dağılımdır, altında ise pozitif çarpık, üstünde ise negatif çarpıktır. Şekle bakıldığında kutunun aşağıda yer alması verinin sağa doğru pozitif bir çarpık olduğunu gösterir. Kutunun dışında her hangi bir değer olmaması uç değerlerin olmadığını gösterir.



e) İlk şıkta elde edilen veri dosayından normallik test sonucu aşağıdaki gibi elde edilir:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
okuma_hizi	,091	40	,200*	,958	40	,141

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Gözlem sayısı 29'dan az olduğunda Shapiro-Wilk testi, gözlem sayısı 29'dan fazla olduğunda ise Kolomogrow-Simirnov testi kullanılır. Örnekte gözlem sayısı 40 olduğu için Kolomogrow-Simirnov testi kullanılır. Sig. Değeri (0,2) 0.05'den büyük olduğu için %5 anlamlılık düzeyinde normal dağılım vardır.

2.3.1. Normal Olmayan Dağılım

Bir araştırmanın istatistik analizinde pek çok testin uygulanabilmesi için veri dağılımın normal veya normale yakın olması gerekir. Eğer veriler normal veya normale yakın bir dağılıma sahip değilse analiz sonuçları yanlış çıkar ve bu da sonuçların yanlış yorumlanmasına sebep olur. Normal dağılım göstermeyen veriler, uç değrler varsa uç değer analizi ile veya belirli dönüştürmelerle normal dağılıma dönüştürlebilir.

.

Uçdeğer analizi: Bir veri setinde uç değerler verilerin normal dağılımını önleyebilir. Bu nedenle uç değerler çıkarılarak veriler normal dağılıma dönüştürülebilir. Ancak tespit edilen uç değerlerin nedenleri açıklanmalıdır. U ç değerler ya hatalı veri girişi ve yanlış kodlama ya da nadir gözlenen bir durum nedeniyle ortaya çıkabilir. Araştırmanın önemine göre uç değerlerin çıkarılıp çıkarılmayacağına karar verilir.