DAĞILIŞ ÖLÇÜLERİ

Bundan önceki bölümde yer alan yer ölçülerinin verilerin merkezini belirlediği belirtilmişti. Ancak, yer ölçüleri verilerin dağılışı ve şeklini belirtmede yetersiz kalmaktadır.

A sınıfından rastgele seçilen 25 erkek öğrencinin ağırlık ortalamasının 50 kg ve B sınıfından rastgele seçilen 30 erkek öğrencinin ağırlık ortalamasının da tesadüfen de olsa 50 kg olduğunu varsayalım. Burada A sınıfındaki varyasyon mu (farklılık) daha fazladır, B sınıfındaki varyasyon mu daha fazladır,şeklinde bir soru yöneltilse hesaplanan ortalama (yer ölçüsü) buna cevap veremez. Halbuki bu sorunun cevabını ancak dağılış ölçüleri ile alabiliriz.

Dağılış ölçüleri, bir veri grubundaki varyasyonu belirlemede veya iki veri setini gösterdikleri varyasyon bakımından karşılaştırmada kullanılan istatistiklerdir.

DAĞILIŞ ÖLÇÜLERİ

Değişim Genişliği (Range)

Değişim genişliği, bir veri setinde en yüksek değerle en küçük değer arasındaki farka denir.

$$D.G. = X_{maks} - X_{\min}$$

Örnek: X_i = {2, 5, 1, 6, 7, 13, 9, 4} ise değişim genişliğini hesaplayınız.

D.G =13 - 1 = 12 olarak elde edilir.

DAĞILIŞ ÖLÇÜLERİ

Değişim genişliği, en basit ve en klasik bir dağılış ölçüsüdür. Buna karşılık değişim genişliğinin bazı dezavantajlı yönleri de vardır. Bunlar:

a) İki grup varyasyon bakımından karşılaştırılmak istense tesadüfen de olsa aynı değişim genişliğine sahip olabilirler. Bu da grupların varyasyon bakımından karşılaştırılmasında yanıltıcı sonuç verir.

b) Yine grupların varyasyon bakımından karşılaştırılmasında eğer gruplarda eşit sayıda gözlem değeri yoksa hesaplanan değişim genişlikleri bizi yanıltabilir.

c) Değişim genişliği hesaplanırken tüm veriler göz önünde tutulmamaktadır. Bu da veri kaybına neden olur.

DAĞILIŞ ÖLÇÜLERİ

Ortalama Mutlak Sapma

n sayıdaki gözlem değerin ortalama mutlak sapması, bu gözlem değerlerin her birinin ortalama sapmalarının mutlak değerinin ortalaması alınarak hesaplanır.

$$O.M.S = \sum_{i=1}^{n} \left| X_i - \overline{X} \right| / n$$

Eşitlikte;

O.M.S : Ortalama mutlak sapmayı,

X_i : X; değişkenine ait i. gözlem değerini,

$\overline{X}$: X değişkenine ait aritmetik ortalamayı,

n : Toplam gözlem sayısını göstermektedir.

DAĞILIŞ ÖLÇÜLERİ

Örnek: X_i = {3, 6, 2, 4, 1, 8} değerlerine ait ortalama mutlak sapmayı hesaplayınız

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{3 + \ldots + 8}{6} = 4$$

$$O.M.S = \frac{|3-4| + \ldots + |8-4|}{6} = \frac{12}{6} = 2$$

Ortalama mutlak sapmanın birimi, gözlem değerlerin birimidir. Yani gözlem değerlerin birimi kg ise ortalama mutlak sapmanın birimi de kg' dır, cm ise bu istatistiğin birimi de cm' dir.

DAĞILIŞ ÖLÇÜLERİ

Frekans tablosundan **ortalama mutlak sapma** hesaplanmak istense, aşağıdaki formül kullanılır.

$$O.M.S = \sum_{j=1}^{k} f_j \left| X_j - \overline{X} \right| \Big/ n$$

Eşitlikte;

O.M.S : Ortalama mutlak sapmayı,

k: Sınıf sayısını,

f_j : j. sınıfın frekansını,

X_j : j. sınıfın sınıf değerini,

$\overline{X}$: X değişkenine ait aritmetik ortalamayı,

n : Toplam gözlem sayısını veya toplam frekansı göstermektedir.

DAĞILIŞ ÖLÇÜLERİ

Örnek: Çizelge 3.1. deki frekans tablosuna ait ortalama mutlak sapmayı hesaplayınız.

$\overline{X} = 49.75$ olarak hesaplanmıştı. Buna göre;

$$O.M.S = \frac{3|27.5 - 49.75| + \ldots + 4|75.5 - 49.75|}{24} = \frac{335}{24} = 13.958$$

DAĞILIŞ ÖLÇÜLERİ

Varyans ve Standart Sapma

Gözlemlerin ortalamadan sapmalarının karelerinin ortalamasına varyans denir. Varyansın birimi yoktur. Örneğin gözlemlerin birimi kg ise varyansın birimi olarak kg^2 dediğimizde bir anlam ifade etmez.

Populasyon varyansı σ^2 , örnek varyansı ise, s^2 ile gösterilir. Buna göre;

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$$

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$$

DAĞILIŞ ÖLÇÜLERİ

Uygulamada çok fazla gözlem değeri olduğunda mevcut formülü kullanmak zor olduğu için bu formül aşağıdaki şekilde kısaltılarak kullanılabilir.

$$\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2)$$

$$=\sum_{i=1}^{n}X_i^2-2\overline{X}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\overline{X}^2$$

$$=\sum_{i=1}^{n}X_i^2-2\frac{\sum X_i}{n}\sum X_i+n\overline{X}^2$$

$$=\sum_{i=1}^{n}X_i^2-2\frac{\left(\sum X_i\right)^2}{n}+n\frac{\left(\sum X\right)^2}{n^2}$$

$$=\sum_{i=1}^{n}X_i^2-\frac{\left(\sum X_i\right)^2}{n}$$

DAĞILIŞ ÖLÇÜLERİ

O halde örnek varyansı kısaca aşağıdaki gibi yazılabilir.

$$S^2 = \frac{1}{n-1}\left[\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right]$$

Örnek: $X_i = \{4, 5, 6, 2, 14\}$ ise örnek varyansını hesaplayınız.

	X_i	X^2_i
	4	16
	5	25
	6	36
	2	2
	14	196
Σ	31	277

$$S^2 = \frac{1}{5-1}\left[277 - \frac{(31)^2}{5}\right] = 21.2$$

DAĞILIŞ ÖLÇÜLERİ

Frekans tablosundan varyans hesaplamak istendiğinde aşağıdaki formül kullanılır.

$$S^2 = \frac{\sum_{j=1}^{n} f_j (X_j - \overline{X})^2}{\sum_{j=1}^{n} f_j - 1} \quad \text{veya} \quad S^2 = \frac{1}{\sum_{j=1}^{k} f_j - 1}\left[\sum_{j=1}^{k} f_j X_j^2 - \frac{\left(\sum_{j=1}^{k} f_j X_j\right)^2}{\sum_{j=1}^{k} f_j}\right]$$

Eşitlikte;

S^2: Örnek varyansını,

f_j : j. sınıfın frekansını,

k : sınıf sayısını,

X_j: j. sınıfın sınıf değerini göstermektedir

DAĞILIŞ ÖLÇÜLERİ

Örnek: Aşağıda verilen frekans tablosuna ait varyansı hesaplayınız.

X_i	f_j	$f_j X_j$	$f_j X_j^2$
27.5	3	82.5	2268.75
33.5	3	100.5	3366.75
39.5	3	.	.
45.5	5	.	.
51.5	1	.	.
57.5	2	.	.
63.5	2	.	.
69.5	1	.	.
75.5	4	.	.
Σ	24	1194	65628

$$S^2 = \frac{1}{24-1}\left[65628 - \frac{1194^2}{24}\right]$$

$$S^2 = 270.717$$

DAĞILIŞ ÖLÇÜLERİ

Standart sapma, varyansın kareköküdür. Varyans, tek başına bir anlam ifade etmediği için ve birimi olmadığı için standart sapma hesaplanır. Standart sapmanın birimi vardır. Gözlem değerlerinin birimi neyse standart sapmanın birimi de odur. Standart sapma, populasyon için s ile örnek için S ile gösterilir.

$$\sigma = \sqrt{\sigma^2}$$

$$S = \sqrt{S^2}$$

DAĞILIŞ ÖLÇÜLERİ

Örnek : $X_i = \{4, 5, 6, 2, 14\}$ değerlerine ait standart sapmayı hesaplayınız.

Bu örneğe ait varyans (S^2) daha önceden 21.2 olarak hesaplanmıştı. Buna göre standart sapma,

$$S = \sqrt{21.2} = 4.604$$ olur.

Varyansın Özellikleri

1) Tek bir gözlem değerine ait varyans hesaplanamaz.

2) Eğer gözlem değerlerinin hepsi birbirinin aynı ise aralarında bir varyasyon (farklılık) olmadığı için varyans yine sıfır olur.

DAĞILIŞ ÖLÇÜLERİ

3) Gözlem değerlerinin hepsine a gibi sabit bir sayı eklenirse varyansın değeri değişmez, eskisi ile aynı kalır.

$$Y_i = X_i + a$$ ise

$$S_y^2 = S_x^2$$ olur.

DAĞILIŞ ÖLÇÜLERİ

Örnek: $X_i = \{4, 5, 6, 2, 14\}$ ise bu değerlerin tamamına 3 gibi sabit bir sayı eklediğinizde elde edilen Y değişkenine ait varyans ne olur?

Daha önceden, X değişkenine ait varyans 21.2 olarak hesaplanmıştı. O halde, Y değişkeninin varyansı da 21.2 olur.

	Y_i	Y^2
	7	49
	8	64
	9	81
	5	25
	17	289
Σ	46	508

$$S^2 = \frac{1}{5-1}\left[508 - \frac{46^2}{5}\right] = 21.2$$

DAĞILIŞ ÖLÇÜLERİ

4) Gözlem değerlerinin hepsinden a gibi sabit bir sayı çıkartılırsa, yeni elde edilen değişkenin varyansı değişmez, eskisi ile aynı kalır.

$$Y_i = X_i - a \quad \textit{ise,}$$

$$S_y^2 = S_x^2$$

Örnek: $X_i = \{4, 5, 6, 2, 14\}$ ise bu değerlerden 2 gibi sabit bir sayı çıkartıldığında elde edilen Y değişkenine ait varyans ne olur?

$$S_y^2 = S_x^2 \qquad S_y^2 = 21.2$$

DAĞILIŞ ÖLÇÜLERİ

5) Gözlem değerlerinin hepsi a gibi sabit bir sayı ile çarpılırsa, yeni elde edilen değişkenin varyansı eskisinden a^2 değeri ile çarpımı kadar değişir.

$$Y_i = a.X_i$$ ise

$$S_y^2 = a^2.S_x^2$$

Örnek: $X_i = \{4, 5, 6, 2, 14\}$ ise bu değerlerin tamamı 2 gibi bir değerle çarpıldığında elde edilen Y değişkeninin varyansı ne olur?

$$S_y^2 = a^2.S_x^2 \qquad S_y^2 = 4x21.2 = 84.8$$

DAĞILIŞ ÖLÇÜLERİ

6) Gözlem değerlerinin hepsi a gibi sabit bir sayıya bölündüğü zaman, elde edilen değişkenin varyansı eskisinden $1/a^2$ kadar değişir.

$Y_i = 1/a.X_i$ ise

$S_y^2 = 1/a^2.S_x^2$ olur.

Örnek: $X_i = \{4, 5, 6, 2, 14\}$ ise bu değerlerin tamamı 2 gibi sabit bir sayıya bölündüğünde elde edilen Y değişkeninin varyansı ne olur?

$Y_i = 1/a.X_i$ idi

$S_y^2 = 1/4.21.2 = 5.3$

DAĞILIŞ ÖLÇÜLERİ

Standart Hata

İstatistikte hesaplanan her istatistik değerin mutlaka hatası da hesaplanmalıdır. Çünkü hesaplanan istatistikler, tahmini bir değer olduğu için mutlaka hataları da vardır.

Standart hatalar, ileride 8. ve 9. bölümlerde anlatılacak olan gerek güven aralıklarının oluşturulmasında gerekse hipotez testlerinin yapılmasında kullanılacaktır.

Standart hata, populasyon ve örnek için sırası ile aşağıdaki formüllerden hesaplanır.

$$\sigma_{\bar{x}} = \sqrt{\sigma^2 / N}$$

$$S_{\bar{x}} = \sqrt{S^2 / n}$$

DAĞILIŞ ÖLÇÜLERİ

Örnek: $X_i = \{3, 5, 6, 4, 2\}$

değerlerinin standart hatasını hesaplayınız.

$$S^2 = \frac{1}{n-1}\left[\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right]$$

$$S^2 = \frac{1}{5-1}\left[90 - \frac{20^2}{5}\right] = 2.5$$

$$S_{\bar{x}} = 0.707$$

DAĞILIŞ ÖLÇÜLERİ

Varyasyon Katsayısı

Şu ana kadar anlatılan dağılış ölçüleri farklı populasyonlara ait varyasyonu (farklılığı) karşılaştırmada yetersiz kalmaktadır. Örneğin sığırlarda ve koyunlarda canlı ağırlık bakımından varyasyon karşılaştırılmak istense, ayrıca dağılış ölçüsü olarak da standart sapma kullanılsa, sığırların standart sapması koyunlarınkinden her zaman büyük çıkacaktır. Çünkü sığırlardan elde edilen canlı ağırlık değerleri daima cebirsel olarak koyunlarınkinden büyüktür. Bu yüzden bu durumda varyasyon katsayısı kullanılmalıdır.

DAĞILIŞ ÖLÇÜLERİ

Varyasyon katsayısı aşağıdaki durumlar için kullanılır.

1) Farklı populasyonlarda aynı özellikler varyasyon bakımından karşılaştırılacağı zaman kullanılır (yukarıda verilen örnekte olduğu gibi).

2) Aynı populasyonda farklı özellikler varyasyon bakımından karşılaştırılacağı zaman kullanılır. Örneğin, aynı sınıfta istatistik ve kimya notları varyasyon bakımından karşılaştırılmak istendiğinde bu istatistik kullanılır.

3) Varyasyon katsayısı, bir araştırmanın güvenilirliğini kontrol etmek istendiğinde kullanılır. Genellikle varyasyon katsayısı %30'un üzerinde olan araştırma neticelerine güvenilmez.

DAĞILIŞ ÖLÇÜLERİ

Varyasyon katsayısı % olarak ifade edilip, aşağıdaki formülden hesaplanır.

$$V.K(\%) = \frac{S}{\overline{X}}.100$$

Eşitlikte;

$V.K(\%)$: % Varyasyon katsayısını,

S: Standart sapmayı,

$\overline{X}$: Ortalamayı göstermektedir.

DAĞILIŞ ÖLÇÜLERİ

Örnek: $X_i = \{12, 4, 6, 2, 13, 5\}$ değerlerine ait varyasyon katsayısını hesaplayınız.

$$S = \sqrt{S^2}$$

$$S^2 = \frac{1}{n-1}\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]$$

$$S^2 = \frac{1}{6-1}\left[394 - \frac{42^2}{6}\right] = 20$$

$$S = \sqrt{20} = 4.472$$

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{12 + \ldots + 5}{6} = 7$$

$$V.K(\%) = \frac{S}{\overline{X}}.100 = \frac{4.472}{7}.100 = 63.89$$

DAĞILIŞ ÖLÇÜLERİ

Örnek: A sınıfından rastgele seçilen 7 öğrencinin matematik notları 3, 7, 9, 5, 5, 8, 2 ve B sınıfından rastgele seçilen 10 öğrencinin matematik notları 9, 7, 1, 6, 2, 5, 8, 8, 3, 4, olarak tespit edilmiştir. Buna göre matematik notları bakımından A sınıfındaki varyasyon mu B sınıfındaki varyasyon mu daha fazladır, niçin?

$$S_A^2 = \frac{1}{7-1}\left[257 - \frac{39^2}{7}\right] = 6.619$$

$$S_A = \sqrt{6.619} = 2.573$$

$$\overline{X}_A = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{3 + \ldots + 2}{7} = 5.571$$

DAĞILIŞ ÖLÇÜLERİ

$$V.K(\%) = \frac{S}{\overline{X}}.100 = \frac{2.573}{5.571}.100 = 46.19$$

$$S_B = \sqrt{7.567} = 2.751$$

$$\overline{X}_B = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{9 + ... + 4}{10} = 5.3$$

$$V.K(\%) = \frac{S}{\overline{X}}.100 = \frac{2.751}{5.3}.100 = 51.91$$

B sınıfındaki varyasyon A sınıfına göre daha fazladır. Çünkü bu sınıftaki varyasyon katsayısı diğerinden daha büyüktür.