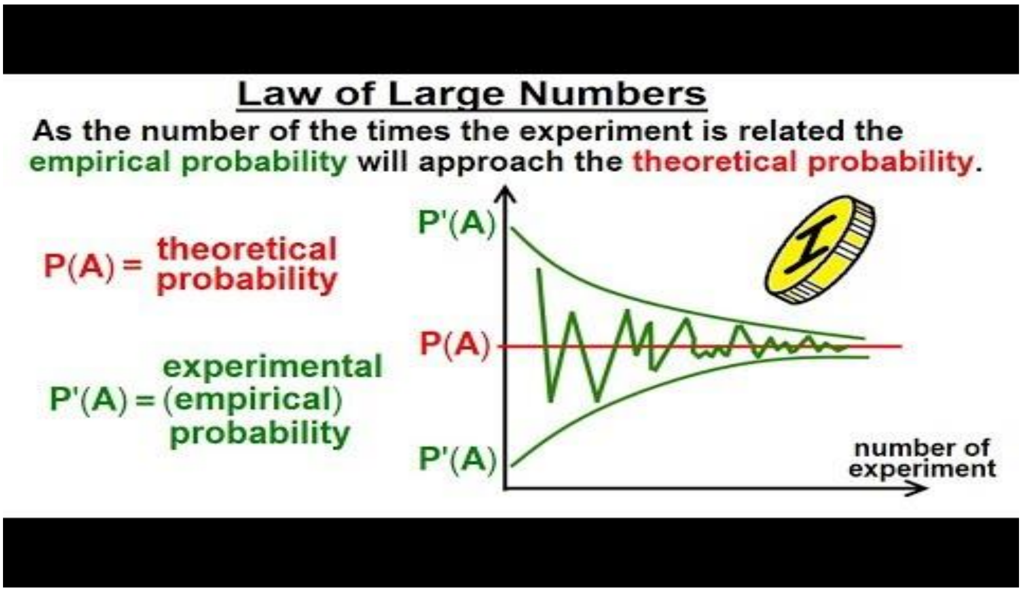
Marmara Üniversitesi

İstatistik Bölümü

Örnekleme Teorisi (2)

Doç. Dr. Atıf Evren

Büyük Sayılar Yasası



$X_1,\cdots,X_n$ bir bağımsız ve özdeş dağılan ve $E(X_i)=\mu \quad (i=1,2,\cdots,n)$ sonlu beklenen değerine sahip rastlantı değişkenleri dizisi olsun. Bu durumda herhangi bir ε>0 olmak üzere

$$n\to\infty \quad \Rightarrow P\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n}-\mu\right|\geq\varepsilon\right\}\to 0$$

olur. Bu yasaya "Büyük Sayılar Yasasının Zayıf Formu" adı verilir.

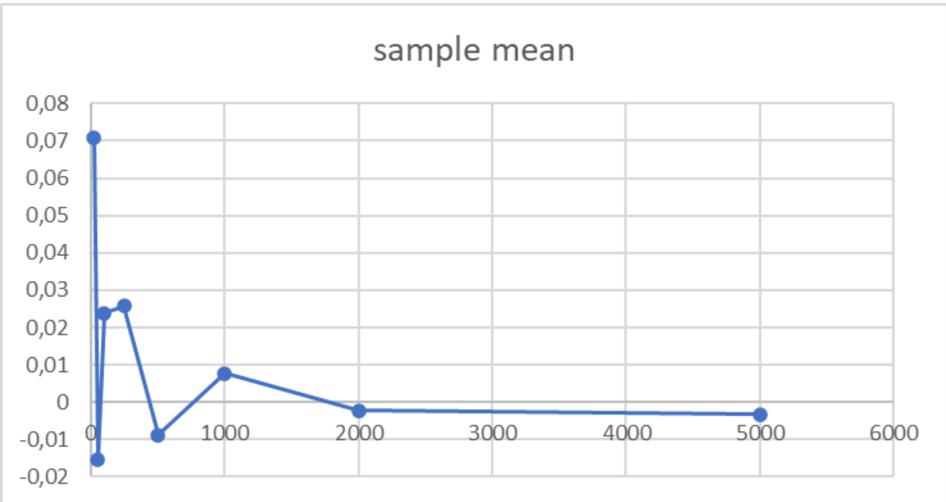
Simülasyonla gösterim:

25, 50, 100,250,500,1000,2000 ve 5000 birimlik rassal örnekler EXCEL yardımı ile standart normal dağılımdan elde edilmiş ve aritmetik ortalamalar hesaplanmıştır. Örneklem hacmine karşılık elde edilen ortalamalar aşağıdaki tabloda verilmektedir:

n	Aritmetik ortalama
25	0,070710666
50	-0,015220962
100	0,023795099
250	0,02575106
500	-0,008693985
1000	0,007762106
2000	-0,002275343
5000	-0,003289503

Tablo: Simülasyonla elde edilen değerlerin ortalamaları

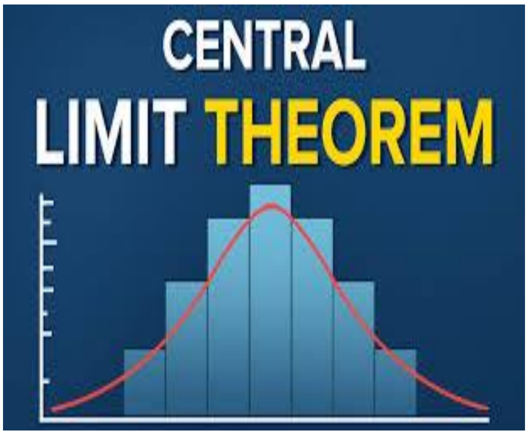
Aşağıdaki şekilde örnek (simülasyon) sayısı arttıkça örnek ortalamalarının sistematik olarak sıfıra yakınsadığı görülmektedir.

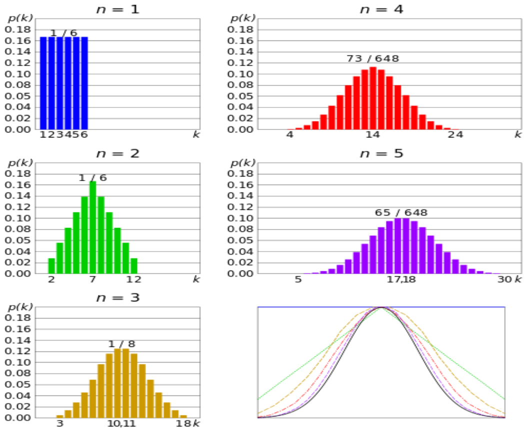


Şekil **2: Dikey eksende ortalama değerler, yatay eksende simülasyon sayıları gösterilmektedir.**

n arttıkça örnek aritmetik ortalamalarının Büyük Sayılar Yasasına uygun olarak gerçek ana kütle ortalaması olan $\mu = 0$ etrafında stabilize olduğunu görmek mümkündür.

<u>Merkezi Limit Teoremi</u>

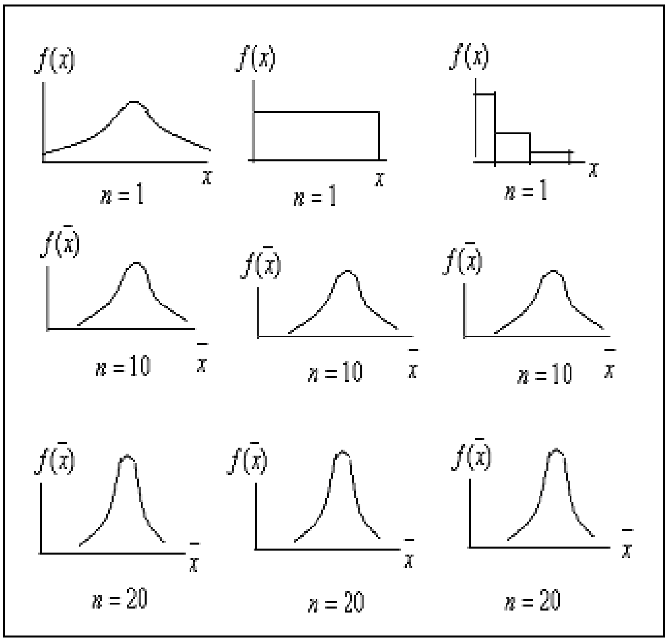




$X_1, \cdots, X_n$; $E(X_i) = \mu$ $\quad Var(X_i) = \sigma^2$; $\quad (i = 1, 2, \cdots, n)$ olmak üzere bağımsız ve özdeş dağılmış bir dizi rastlantı değişkeni olsunlar. Bu durumda

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \sim \text{Normal}(0,1) \quad \text{olur.}$$

Merkezi Limit Teoremini bağımsız gözlem değerlerinin toplamları cinsinden $n \to \infty$ için $\sum_{i=1}^{n} X_i \sim Normal(n\mu, n\sigma^2)$ şeklinde formüle etmek mümkündür. Aynı teorem örnek ortalamaları cinsinden $n \to \infty$ için $\bar{X} \sim Normal\left(\mu, \frac{\sigma^2}{n}\right)$ biçiminde de dile getirilebilir.



Şekil 1: Merkezi Limit Teoremi'nin şekille anlatılması: Örnek büyüklüğü arttıkça örnek ortalamalarının olasılık dağılımı normal dağılıma yakınsamaktadır.

Örnek: Sürekli rastlantı değişkeni $\lambda = 0,02$ parametreli bir üstel dağılıma uymaktadır. Bu dağılımdan n=150 büyüklüğünde bir örnek çekilmektedir. $\bar{X}$ örnek ortalamasını oluşturduğuna göre $P(40 < \bar{X} < 55)$ olasılığını bulunuz.

Çözüm: Bu kez de $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ ~Normal(0,1) eşitliğinden yararlanalım.

$f(x) = \lambda e^{-\lambda x} = 0,02e^{-0,02x}$ olur.

Öte yandan $E(X) = \mu = \dfrac{1}{\lambda} = 50$;

$Var(X) = \dfrac{1}{\lambda^2} = 2500$ bulunur.

n=150 sayısının yeterince büyük olması nedeniyle $\bar{X} \sim Normal(\mu, \dfrac{\sigma^2}{n})$; $\bar{X} \sim Normal(50, \dfrac{2500}{150})$ olur.

$$P(40 < \bar{X} < 55) = P\left(\frac{40-50}{\sqrt{2500/150}} < \frac{\bar{X}-50}{\sqrt{2500/50}} < \frac{55-50}{\sqrt{2500/150}}\right)$$
$$= P(-2,45 < Z < 1,225) = 0,8827$$

Örnekleme Dağılımları

İstatistik temel olarak rastlantısal sonuçlardan genelleme yapma ve istatistiksel deney sonuçlarına ilişkin a priori (deney öncesi) öngörüde bulunma işi ile ilgilenir. Rastlantısal olayların sonuçları istatistikçiler için birer örnek (ya da örneklem) oluşturur. Bu örnekten yola çıkarak örneğin içinden geldiği ana kütle hakkında genelleme yapılmaya çalışılır. İçinden örnek çekilen ana kütle bazı hallerde sonsuz ana kütle olarak nitelendirilir. Buradaki "sonsuzluk" kavramı çekilen gözlemlerin ana kütlenin olasılık dağılımını etkilememesi (bağımsızlık) anlamındadır. Yoksa "sonsuz" sıfatı zorunlu olarak ana kütleyi oluşturan birimlerin sonsuzluğu anlamına gelmeyebilir.

Tanım: $X_1, X_2, \cdots, X_n$ rastlantı değişkenleri bağımsız ve aynı dağılımdan çekilen (independently and identically distributed) rastlantı değişkenleri olsunlar. Bu durumda bu rastlantı değişkenlerinin aynı sonsuz büyüklükteki ana kütleden çekilmiş bağımsız bir örnek oluşturdukları söylenir.

Eğer $X_1, X_2, \cdots, X_n$ bağımsız ise bunların birleşik olasılık fonksiyonu marjinal fonksiyonlarının çarpımına eşit olur.

$$f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$$

Burada $f(x_i)$ $X = x_i$ noktasında olasılık fonksiyonunun değeridir. İstatistiksel kestirimler örneği oluşturan $X_1, X_2, \cdots, X_n$'in fonksiyonlarından yararlanılarak gerçekleştirilir. Tipik olarak i istatistik dendiğinde ilk akla gelen örnek ortalaması, örnek varyansı gibi istatistikler olmaktadır.

Tanım:
$X_1, X_2, \cdots, X_n$ bir tesadüfi örnek oluştursun. Bu durumda

$\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ örnek ortalaması ve $S^2 = \dfrac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$

örnek varyans olarak adlandırılır.

Örnek Ortalamasının Olasılık Dağılımı

$X_1, X_2, \cdots, X_n$ örneği rastlantısal olduğuna göre bu değerlere bağlı olarak hesaplanan örnek ortalaması $\overline{X}$ da rastlantısal olmalıdır. Bu durumda $\overline{X}$ 'in olasılık dağılımından söz etmek anlamlı olmaktadır. Örnek ortalamasının örnekten örneğe bağlı olarak alacağı bütün değerlerden oluşan olasılık dağılımına; örnek ortalamasının örnekleme dağılımı adı verilmektedir. Burada

$$E(\overline{X}) = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{\sum_{i=1}^{n} E(X_i)}{n}$$ olur.

$X_1, X_2, \cdots, X_n$ aynı dağılımdan gelen bağımsız ve birbiri ile özdeş rastlantı değişkenleri olduklarına ve $E(X_i) = \mu \quad i = 1, 2, \cdots, n$

olduğuna göre $E(\overline{X}) = \dfrac{\sum_{i=1}^{n} \mu}{n} = \mu$ olur.

Yine $Var(X_i) = \sigma^2 \quad i = 1, 2, \cdots, n$ olduğuna göre

$$Var(\overline{X}) = E\left[\overline{X} - E(\overline{X})\right]^2$$ ya da

$$Var(\overline{X}) = Var\left[\frac{\sum_{i=1}^{n} X_i}{n}\right]$$

$$Var(\overline{X}) = \frac{\sigma^2}{n}$$

$\sigma_{\overline{X}}$ (ortalamanın örnekleme dağılımının standart sapması), ortalamanın standart hatası olarak da adlandırılır ve örnekleme dağılımının varyansının kareköküne eşittir:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Dolayısıyla örnek büyüklüğü arttıkça standart hatanın düşeceğine dikkat edilmelidir. n=1 için $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{1}} = \sigma$. n'in sonsuz arttırıldığı durumda ise $\lim_{n\to\infty}\sigma_{\bar{X}} = \lim_{n\to\infty}\frac{\sigma}{\sqrt{n}} = 0$ olacağından örnek ortalamaları ile ilgili dağılımın varyansının (ve standart sapmasının) sıfıra eşit olacağını söyleyebiliriz.

Örnek: $X_1, X_2, \cdots, X_n$ ortalaması μ ve varyansı σ^2 olan sonsuz bir ana kütleden çekilen n=50 birimlik tesadüfi bir örnek olsun. Örnek ortalaması ile ana kütle ortalaması arasındaki mutlak farkın standart hatanın 1,33 katından küçük olma olasılığını Merkezi Limit Teoremine dayanarak bulunuz.

Çözüm: $\bar{X}$ örnek ortalaması olsun. Buna göre

$$P\left(\left|\bar{X}-\mu\right| \le 1,33\sigma_{\bar{X}}\right) = P\left(-1,33\sigma_{\bar{X}} \le \bar{X}-\mu \le 1,33\sigma_{\bar{X}}\right)$$

$$= P\left(\frac{-1,33\sigma_{\bar{X}}}{\sigma_{\bar{X}}} \le \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \le \frac{1,33\sigma_{\bar{X}}}{\sigma_{\bar{X}}}\right)$$

$$= P\left(-1,33 \le Z \le 1,33\right) = 2.P\left(0 \le Z \le 1,33\right)$$

$$= 2.(0,4082) = 0,8164$$

Örnek: $\bar{X}$ ortalaması $\mu = 14$ ve varyansı $\sigma^2 = 9$ olan normal bir dağılımdan çekilen n=16 birimlik bir örneğin ortalaması olsun. Bu durumda $P(12 \le \bar{X} \le 18)$ olasılığını bulunuz.

Çözüm:

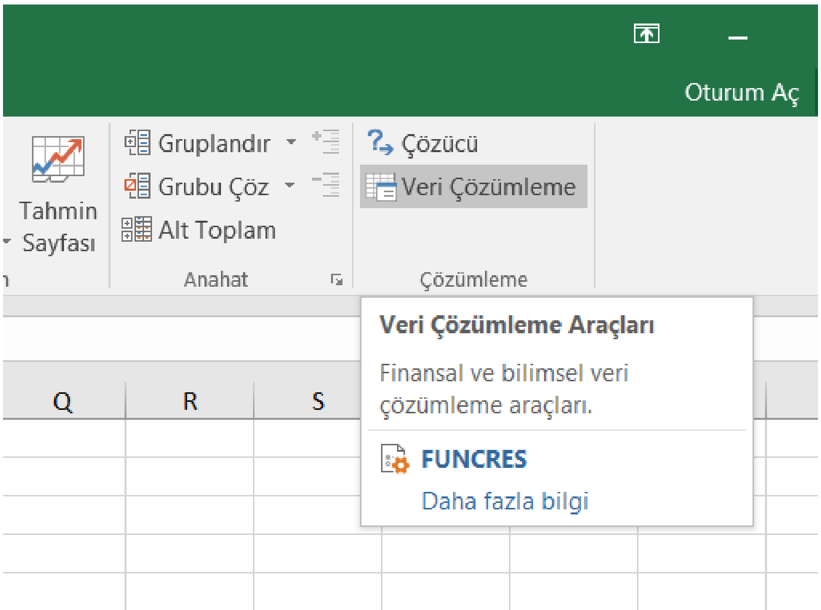
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{4}$$

$$P(12 \le \bar{X} \le 18) = P(\frac{12-\mu}{\sigma_{\bar{X}}} \le \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \le \frac{18-\mu}{\sigma_{\bar{X}}}) = P(\frac{12-14}{0,75} \le \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \le \frac{18-14}{0,75})$$

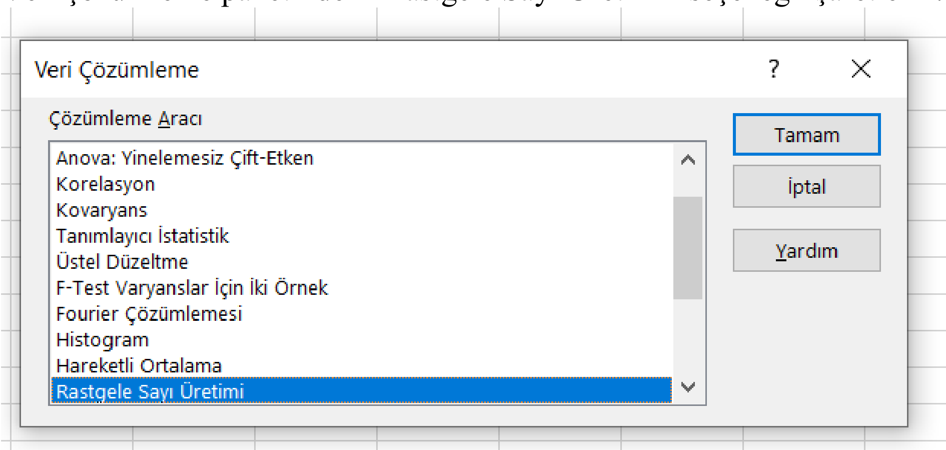
$$= P(-2,67 \le Z \le 5,33) = P(-2,67 \le Z \le 0) + P\left(0 \le Z \le 5,33\right) = 0,4962 + 0,5 = 0,9962$$

Simülasyonla tahmin:

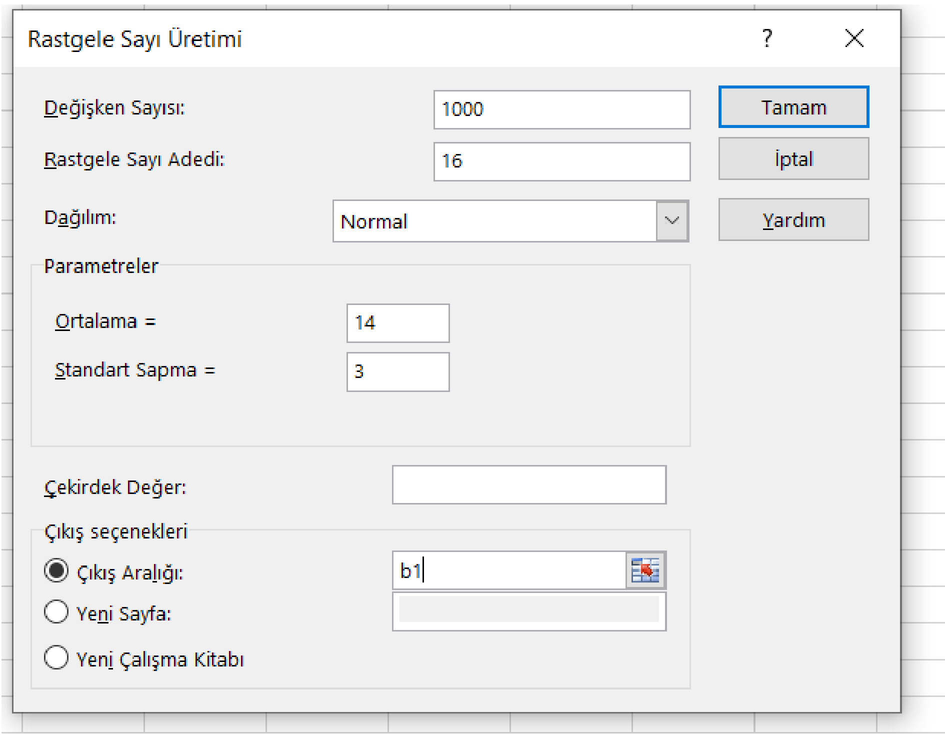
Aynı soruyu EXCEL Veri Çözümleme Paketi ile cevaplandıralım.



Veri çözümleme paketinden "Rastgele Sayı Üretimi" seçeneği işaretlenir.



Açılan pencere aşağıdaki gibi doldurulur.



Elde edilen simülasyon sonuçları aşağıdaki gibi görüntülenir.

B1 | 10,5131814860797

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		10,51318	10,41809	16,95257	16,68424	11,66863	11,77458	12,80676	14,79702	12,80626	8,674477	12,47998	15,5323	16,0332	5,506775
2		15,25334	14,83155	12,1064	10,23147	16,36566	16,39712	11,44611	14,47113	15,07076	14,25217	9,055332	15,44276	13,18037	13,55046
3		10,95054	16,79817	13,24261	16,1054	13,26013	17,61679	19,83235	11,89313	17,98094	16,72178	18,15532	19,13965	17,10304	14,50276
4		12,44727	14,15854	13,08441	10,83477	15,20691	15,54172	13,08321	17,24441	13,98015	16,96003	10,19379	16,78155	14,02123	12,3424
5		14,01733	21,74562	17,11128	15,74873	25,00474	13,29699	11,44545	8,358941	15,6187	12,56857	14,74626	16,81096	11,12322	16,95854
6		12,73061	14,87672	12,96464	16,9481	12,49536	14,53677	11,90749	17,55352	13,17298	18,05465	15,70782	11,2181	16,2759	15,79271
7		15,42707	13,8626	13,29227	15,53439	9,113326	14,56898	18,94288	12,61086	11,65588	14,01756	11,75337	12,65541	15,2817	17,15443
8		13,6789	11,62462	8,324971	13,26226	10,13612	11,89724	14,58535	13,30665	14,06782	14,74815	10,73409	15,19597	14,12661	15,88716
9		19,80973	17,22798	10,5918	11,11157	10,76628	14,36569	8,656632	17,23249	15,9821	17,46884	9,800522	12,29299	17,83098	16,9765
10		18,48877	18,33452	12,71402	14,88486	15,94117	13,83227	12,90033	15,37613	15,33901	16,95891	12,70269	8,227237	13,59683	9,444755
11		8,621517	14,35528	14,25931	11,79749	7,476358	20,78672	16,0928	17,91225	8,712866	14,74224	13,71765	13,59428	15,22685	12,73462
12		11,74181	16,16096	7,554666	18,28175	18,33452	10,36558	12,9378	14,021	19,37049	16,42755	14,7517	9,066901	12,65719	12,69841
13		14,32757	18,58896	11,82981	15,72025	11,62304	11,27683	14,20613	12,91333	18,19155	14,83393	12,6422	14,82201	14,65146	12,64093
14		14,9387	12,2352	11,59593	14,53141	18,84732	11,1417	16,5562	13,30359	12,83255	19,35801	11,47206	14,69264	14,65945	11,02162
15		16,53514	14,11328	14,37887	12,56265	15,5195	11,137	14,07195	17,75142	17,90314	16,14226	17,81495	15,93185	16,54007	12,84541
16		17,20068	15,34941	9,848871	15,80755	12,65947	20,76031	18,90493	15,97297	12,46744	12,6911	20,57335	16,06974	18,79474	12,40857

Daha sonra B17 hücresine “”=ORTALAMA(B1:B16)”   formülü yazılır.

B17 | =ORTALAMA(B1:B16)

Ad Kutusu

	A	B	C	D	E	F
1		10,51318	10,41809	16,95257	16,68424	11,66863
2		15,25334	14,83155	12,1064	10,23147	16,36566
3		10,95054	16,79817	13,24261	16,1054	13,26013
4		12,44727	14,15854	13,08441	10,83477	15,20691
5		14,01733	21,74562	17,11128	15,74873	25,00474
6		12,73061	14,87672	12,96464	16,9481	12,49536
7		15,42707	13,8626	13,29227	15,53439	9,113326
8		13,6789	11,62462	8,324971	13,26226	10,13612
9		19,80973	17,22798	10,5918	11,11157	10,76628
10		18,48877	18,33452	12,71402	14,88486	15,94117
11		8,621517	14,35528	14,25931	11,79749	7,476358
12		11,74181	16,16096	7,554666	18,28175	18,33452
13		14,32757	18,58896	11,82981	15,72025	11,62304
14		14,9387	12,2352	11,59593	14,53141	18,84732
15		16,53514	14,11328	14,37887	12,56265	15,5195
16		17,20068	15,34941	9,848871	15,80755	12,65947
17	ortalama	14,16764				

Bu formül aşağıda görüldüğü gibi C17,D17,… hücrelerine kopyalanarak 1000 tane simülasyon sonucundan elde edilen örnek ortalamaları kaydedilir.

A17 | ortalama

	A	B	C	D	E	F	G	H	I	J	K
1		10,51318	10,41809	16,95257	16,68424	11,66863	11,77458	12,80676	14,79702	12,80626	8,674477
2		15,25334	14,83155	12,1064	10,23147	16,36566	16,39712	11,44611	14,47113	15,07076	14,25217
3		10,95054	16,79817	13,24261	16,1054	13,26013	17,61679	19,83235	11,89313	17,98094	16,72178
4		12,44727	14,15854	13,08441	10,83477	15,20691	15,54172	13,08321	17,24441	13,98015	16,96003
5		14,01733	21,74562	17,11128	15,74873	25,00474	13,29699	11,44545	8,358941	15,6187	12,56857
6		12,73061	14,87672	12,96464	16,9481	12,49536	14,53677	11,90749	17,55352	13,17298	18,05465
7		15,42707	13,8626	13,29227	15,53439	9,113326	14,56898	18,94288	12,61086	11,65588	14,01756
8		13,6789	11,62462	8,324971	13,26226	10,13612	11,89724	14,58535	13,30665	14,06782	14,74815
9		19,80973	17,22798	10,5918	11,11157	10,76628	14,36569	8,656632	17,23249	15,9821	17,46884
10		18,48877	18,33452	12,71402	14,88486	15,94117	13,83227	12,90033	15,37613	15,33901	16,95891
11		8,621517	14,35528	14,25931	11,79749	7,476358	20,78672	16,0928	17,91225	8,712866	14,74224
12		11,74181	16,16096	7,554666	18,28175	18,33452	10,36558	12,9378	14,021	19,37049	16,42755
13		14,32757	18,58896	11,82981	15,72025	11,62304	11,27683	14,20613	12,91333	18,19155	14,83393
14		14,9387	12,2352	11,59593	14,53141	18,84732	11,1417	16,5562	13,30359	12,83255	19,35801
15		16,53514	14,11328	14,37887	12,56265	15,5195	11,137	14,07195	17,75142	17,90314	16,14226
16		17,20068	15,34941	9,848871	15,80755	12,65947	20,76031	18,90493	15,97297	12,46744	12,6911
17	ortalama	14,16764	15,29259	12,49078	14,37793	14,02616	14,33102	14,27352	14,66993	14,69704	15,28876
18											
19											

Elde edilen bu ortalamalar başka bir sayfaya aşağıdaki gibi kopyalanır:

	A	B	C	D	E
1	ortalama				
2	14,16764				
3	15,29259				
4	12,49078				
5	14,37793				
6	14,02616				
7	14,33102				
8	14,27352				
9	14,66993				
10	14,69704				
11	15,28876				
12	13,51882				
13	14,21715				
14	15,06893				
15	13,27913				
16	13,50687				
17	15,24191				
18	14,22277				
19	14,5919				
20	13,58149				
21	15,08415				
22	13,94088				
23	14,68338				

Daha sonra değeri 12 ile 18 arasında olan hücreleri belirlemek için B2 hücresine aşağıdaki görselde formül çubuğunda yazılı olan formül girilir. Bu formül B2:B1001 adresinde bulunan bütün diğer hücrelere de kopyalanır.

B2 =EĞER(VE(A2>12;A2<18);1;0)

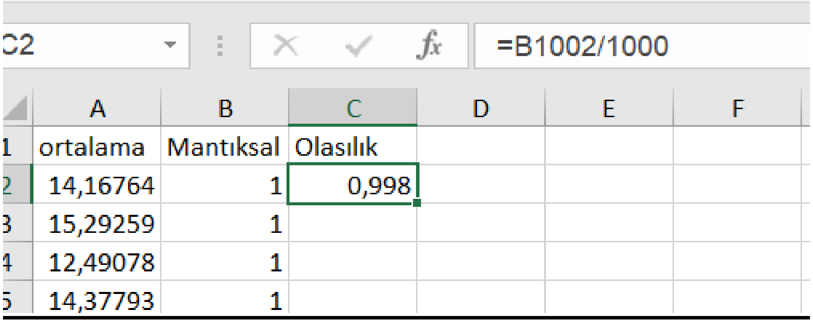
	A	B	C	D	E	F	G
1	ortalama	Mantıksal					
2	14,16764	1					
3	15,29259						
4	12,49078						
5	14,37793						
6	14,02616						
7	14,33102						
8	14,27352						
9	14,66993						
10	14,69704						
11	15,28876						

	A	B	C	D
1	ortalama	Mantıksal		
2	14,16764	1		
3	15,29259	1		
4	12,49078	1		
5	14,37793	1		
6	14,02616	1		
7	14,33102	1		
8	14,27352	1		
9	14,66993	1		
10	14,69704	1		
11	15,28876	1		
12	13,51882	1		
13	14,21715	1		
14	15,06893	1		
15	13,27913	1		
16	13,50687	1		
17	15,24191	1		
18	14,22277	1		
19	14,5919	1		
20	13,58149	1		

Daha sonra B1002 hücresine "=TOPLA(B2:B1001) formülü yazılır.

B1002 =TOPLA(B2:B1001)

Son olarak C2 hücresine "=B1002/1000" yazılır. Bu oran sorulan olasılığın tahmini değeri olacaktır.



C2 =B1002/1000

	A	B	C	D	E	F
1	ortalama	Mantıksal	Olasılık			
2	14,16764	1	0,998			
3	15,29259	1				
4	12,49078	1				
5	14,37793	1				

Bu olasılık rastgele sayı kullanılarak yaklaşık oarak 0,998 olarak bulunmuştur. Bu olasılığın gerçek değerinin 0,9962 olduğu hatırlanmalıdır.