7. DERS

İSTATİSTİKSEL TAHMİNLEME

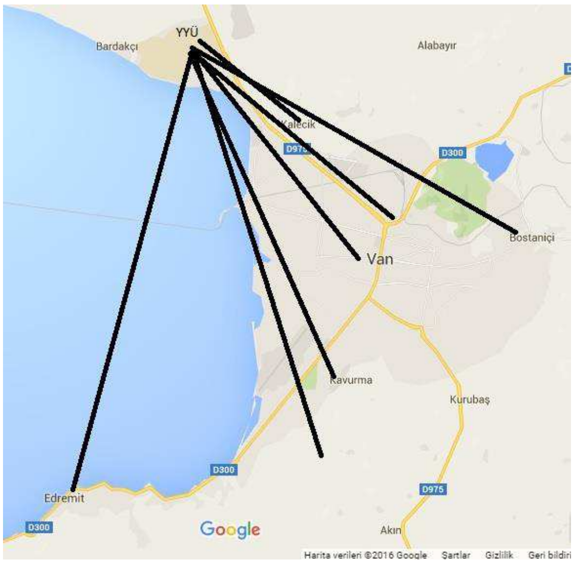
Örneklem istatistiklerinden hareketle ana kütle parametreleri hakkında genelleme yapmaya istatistiksel tahminleme denir.

ŞEKİL: Evren uzay-örneklem uzay

İstatistiksel tahmin iki aşamada gerçekleşir.

1. Bilinmeyen parametre değerinin tahminlenmesi,
2. Parametre değeri hakkında karar verme.

Üniversite çalışanlarının servis problemiyle uğraşmaktadır. Çalışanların ikametgahıyla kampüs arasındaki ortalama uzaklık bilgisine ihtiyaç vardır.



Soru: “Ne yapılabilir?”

İlk akla gelen şudur. Tüm servis kullanıcıların ikametgah-kampüs uzaklıkları belirlenir. Bunlar toplanır ve kişi sayısına bölünür. Ancak bu pratik ve bazen mümkün olmayan bir yoldur.

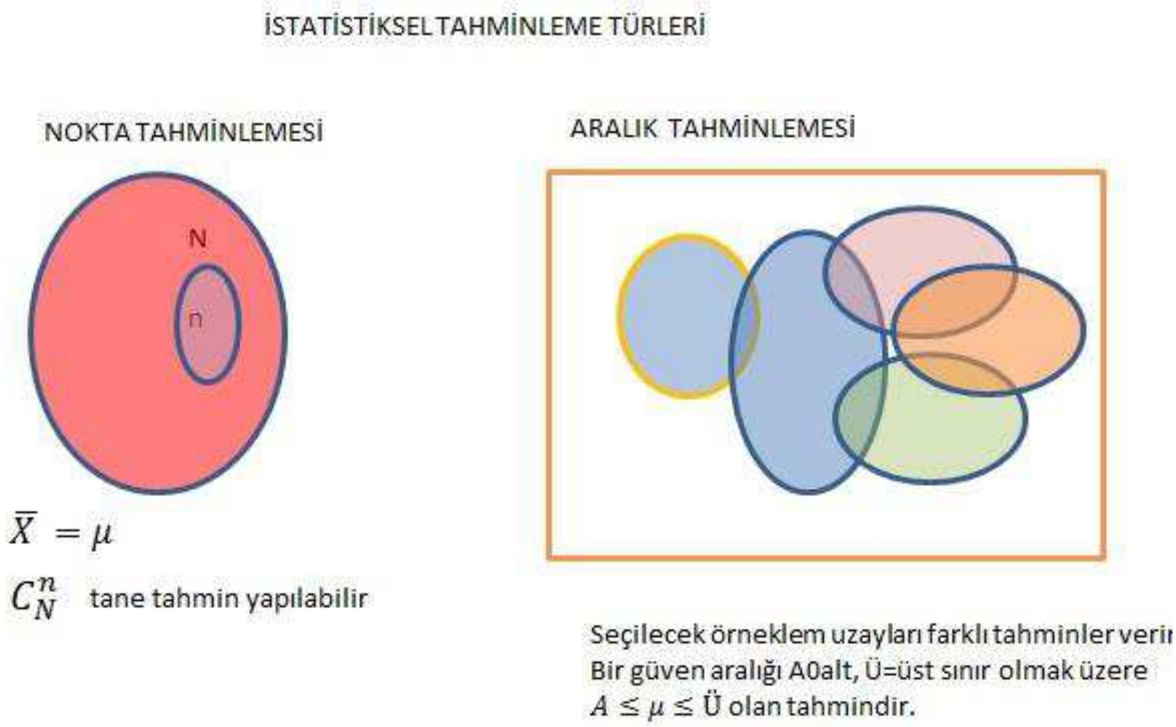
Diğer bir yol (ki bunu tercih edeceğiz), örneklem alt uzay seçilir ve hesaplama buna göre yapılır. Algoritma şöyledir.

1. Ana kütleden n-elemanlı (büyüklü-hacimli) bir rassal örneklem seçilir.
2. Gözlem değerleri kullanılarak tahmin edilerek parametre için istatistikler hesaplanır.
3. Parametre değeri tahmin edilir.

X_i ler i yinci değer olmak üzere;

1. i. ler (örneklem rassal alt uzay) belirlenir.
2. X_i ler hesaplanır.
3. Hangi istatistik formülünün kullanılacağı belirlenir (Mesela $\frac{\sum X_i}{n}$ Bu tahminleyicidir).
4. Tahmin yapılır.(Eğer tahminleyici olarak $\bar{X}$ seçildiyse $\bar{X} = \frac{\sum X_i}{n}$ hesaplanır.

Genel olarak istatistik tahminleme iki türlüdür. Nokta tahminlemesi ve aralık tahminlemesi.



NOKTA TAHMİNLEMESİ

a) A telefon firmasının 2015 yılı müşterilerinin ortalama aylık fatura ödemeleri $\bar{X} = 50TL$ ise, nokta tahminlemesi μ=50TL dir.

b) Bir dizi izleyicileri arasında seçilen 1000 kişiden, 715 i diziyi sevmiştir. Dizinin sevilme nokta tahminlemesi

$$\mu = \frac{r}{n} = \frac{715}{1000} = 0{,}715$$

μ =%71,5 tir.

ARALIK TAHMİNLEMESİ

Standard hatanın küçüklüğü tahminin güvenirliği anlamına gelir. Güvenilir tahmin, tanımlanan ana kütleden seçilen aynı hacimli farklı örneklemlerde büyük ölçüde fark göstermeyen tahmindir. Nokta tahminlemesi, tahminin ne kadar güvenli olduğu hakkında bilgi vermez. Dolayısıyla nokta tahminleme sınırlı tahminlemedir. Örneklem istatistiklerinin değerleri ve standard hataları örneklemin seçilişine bağlı olarak değişir. Böylece güven aralığının sınırları da değişir.

Bir istatistik tahminlemeye eşlik eden iki kavram vardır. Bunlar: güven düzeyi ve anlamlılık düzeyidir.

Güven düzeyi +Anlamlılık düzeyi =1 dir. α anlamlılık düzeyini göstermek üzere Güven düzeyi (GD), $1-\alpha$ ile gösterilir. Mesela güven düzeyi %95 olsun isteniyor ise, α anlamlılık düzeyi,

$$\alpha = 1 - 0{,}95 = 0{,}05$$

olarak belirlenir. İyi bir istatistik tahminleme için güven düzeyi(GD) %95 veya %99 olarak seçilmelidir.

ARİTMETİK ORTALAMA ARALIK TAHMİNLEME ALGORİTMASI

Ana kütle aritmetik ortalaması, μ, için $1-\alpha$ güven sınırlarının ya da güven aralığının belirleme sürecine μ nün aralık tahminlemesi denir. İşlem algoritması şöyledir.

1. $GD = 1-\alpha$ dolayısıyla α belirlenir (bu bir tercihtir).
2. n −hacimli bir rassal örneklem seçilir.
3. $\bar{X}$ nin standard hatası $\sigma_{\bar{X}}$ hesaplanır.
4. $\bar{X}_A \leq \mu \leq \bar{X}_Ü$ güven aralığı tahmin edilir.

Büyük örneklem hacmi, $(n \geq 30)$ için esas olarak alırsak, alt ve üst sınırlar,

$$\bar{X}_A = \mu - Z_{\frac{\alpha}{2}}.\sigma_{\bar{X}}$$

$$\bar{X}_Ü = \mu + Z_{\frac{\alpha}{2}}.\sigma_{\bar{X}}$$

olarak bellidir. $Z_{\frac{\alpha}{2}}$ için hesaplanmış değerler tablosu şöyledir. 4

GD:$1-\alpha$	α	$Z_{\frac{\alpha}{2}}$
0.99	0,01	2.576
0.95	0,05	1.960
0.90	0,1	1.645
0.80	0,2	1.282
0.60	0.4	0.842
0.50	0,5	0.674

$\sigma_{\bar{X}}$ in bilinmediği, ki gerçek hayatta genellikle bilinmez, durumlarda σ yerine S standard sapması kullanılarak hesap yapılır ve bu $S_{\bar{X}}$ ile gösterilir.

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}$$

ÖRNEK: Bir firmanın ürettiği ekmeklerin 100 tanesi rassal örneklem olarak seçilmiştir. Ortalama ağırlık $385gr$ ve standard sapma $S = 12gr$ olarak hesaplanmıştır. Üretilen ekmeklerin ortalama ağırlığını %95 güven düzeyi ile belirleyiniz.

Çözüm:

Veriler

$n = 100$

$\bar{X} = 385$
$S = 12$
$GD = 1 - 0{,}95 = 0{,}05$
$\sigma = 0{,}05$
$Z_{\frac{\alpha}{2}} = 1{,}96$

Formül:

$$\bar{X} - Z_{\frac{\alpha}{2}}.S_{\bar{X}} < \mu < \bar{X} + Z_{\frac{\alpha}{2}}.S_{\bar{X}}$$

$S_{\bar{X}} = \frac{12}{\sqrt{100}} = 1{,}2$ değerler yerine konursa,

385-1,96.1,2 < μ< 385+1,96.1,2

382,648 < μ< 387,352

Bu sonucun yorumu şudur: Seçilen rassal bir ekmeğin gramının 382,648 < μ< 387,352

Arasında olmasına %95 güvenebiliriz. Başka bir ifade ile rastgele seçilen bir ekmeğin ağırlığının 382,648 < μ< 387,352 arasında olma olasılığı %95 dir.

Küçük Hacimli Örneklemlerde μ nün Aralık Tahminlemesi

Örneklem hacmi küçük olduğu zaman , $n < 30$, güven aralığı

$$\bar{X} - t.S_{\bar{X}} < \mu < \bar{X} + t.S_{\bar{X}}$$

den hesaplanır. Burada;

$\bar{X}$:ortalama

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

$$S_{\bar{X}} = \frac{S}{\sqrt{n-1}}$$

S: standard sapma

t: α ve γ ye bağlı tablo değeri

γ=$n - 1$: serbestlik derecesi

ÖRNEK:

Lastik üreticisi bir firma %95 güven düzeyi ile ürettiği lastiklerin km/ömrünü hesaplayacaktır. Yapılan ölçümlerde lastiklerin ortalama ömrünün 38000 km, standard sapmanın 1400 km olduğunu tespit etmiştir. 26 denek kullanılmıştır.

ÇÖZÜM:

Verilenler

$\bar{X}$: 38000km

S: 1400km

GD=0,95

$\alpha = 0{,}05$

$\gamma = n - 1 = 25$

$t(\alpha,\gamma) = 0{,}684$ (tablodan)

(t tablosu sonda verilecektir)

$$\bar{X} - t.S_{\bar{X}} < \mu < \bar{X} + t.S_{\bar{X}}$$

Formülünde;

S=1400, $S_{\bar{X}} = \frac{S}{\sqrt{n-1}} = \frac{1400}{25} = 280$, t=0,684 değerleri yerine yazılırsa, 6

38.000-0,684.280< μ <28.000+0,684.280

37.808,48< μ <38.191,15

elde edilir. Yani, firmanın ürettiği lastiklerin km ömrü %95 olasılıkla, 37.808 km ile 28.191 km arasındadır.

ANA KÜTLE ORANININ ARALIK TAHMİNLEMESİ

Örneklem hacmi büyük $n \geq 0$ veya $\frac{n}{N} \leq 0{,}05$ olduğu durumlarda güven aralığı şöyle bulunur.

$$P - Z_{\frac{\alpha}{2}}.S_P < \pi < P + Z_{\frac{\alpha}{2}}.S_P$$

Burada;

$$P = \frac{r}{n}$$

$$S_P = \sqrt{\frac{P(1-P)}{n}}$$

$Z_{\frac{\alpha}{2}}$: $(1-\alpha)$ ya bağlı tablo değeri

ÖRNEK:

2000 kişilik sigara içen bir örneklem grupta, 225 kişi kanserdir. Bu bölgedeki kanser olma oranını %95 güven düzeyiyle tahmin edelim.

ÇÖZÜM:

Verilenler

n=2000

r=225

$\alpha = 1 - 0,95 = 0,05$

$P = \frac{r}{n} = \frac{225}{2000} = 0,1125$

$Z_{\frac{\alpha}{2}} = 1,960$

$S_P = \sqrt{\frac{0,1125(1-0,1125)}{2000}}$=0,007

$$P - Z_{\frac{\alpha}{2}}.S_P < \pi < P + Z_{\frac{\alpha}{2}}.S_P$$
$$0,1125 - 0,0137 < \pi < 0,1125 + 0,0137$$
$$0,09878 < \pi < 0,1262$$

t - TABLOSU

	Tek Yönlü Testte, α				
	0.25	0.05	0.025	0.01	0.005
Serbestlik	Çift Yönlü Testte, α				
Derecesi	0.50	0.10	0.05	0.02	0.01
1	1.000	6.34	12.71	31.82	63.66
2	0.816	2.92	4.30	6.96	9.92
3	0.765	2.35	3.18	4.54	5.84
4	0.741	2.13	2.78	3.75	4.60
5	0.727	2.02	2.57	3.36	4.03
6	0.718	1.94	2.45	3.14	3.71
7	0.711	1.90	2.36	3.00	3.50
8	0.706	1.86	2.31	2.90	3.36
9	0.703	1.83	2.26	2.82	3.25
10	0.700	1.81	2.23	2.76	3.17
11	0.697	1.80	2.20	2.72	3.11
12	0.695	1.78	2.18	2.68	3.06
13	0.694	1.77	2.16	2.65	3.01
14	0.692	1.76	2.14	2.62	2.98
15	0.691	1.75	2.13	2.60	2.95
16	0.690	1.75	2.12	2.58	2.92
17	0.689	1.74	2.11	2.57	2.90
18	0.688	1.73	2.10	2.55	2.88
19	0.688	1.73	2.09	2.54	2.86
20	0.687	1.72	2.09	2.53	2.84
21	0.686	1.72	2.08	2.52	2.93
22	0.686	1.72	2.07	2.51	2.82
23	0.685	1.71	2.07	2.50	2.81
24	0.685	1.71	2.06	2.49	2.80
25	0.684	1.71	2.06	2.48	2.79
26	0.684	1.71	2.06	2.48	2.78
27	0.684	1.70	2.05	2.47	2.77
28	0.683	1.70	2.05	2.47	2.76
29	0.683	1.70	2.04	2.46	2.76
30	0.683	1.70	2.04	2.46	2.75
35	0.682	1.69	2.03	2.44	2.72
40	0.681	1.68	2.02	2.42	2.71
45	0.680	1.68	2.02	2.41	2.69
50	0.679	1.68	2.01	2.40	2.68
60	0.678	1.67	2.00	2.39	2.66
70	0.678	1.67	2.00	2.38	2.65
80	0.677	1.66	1.99	2.38	2.64
90	0.677	1.66	1.99	2.37	2.63
100	0.677	1.66	1.98	2.36	2.63
125	0.676	1.66	1.98	2.36	2.62
150	0.676	1.66	1.98	2.35	2.61
200	0.675	1.65	1.97	2.35	2.60
300	0.675	1.65	1.97	2.34	2.59
400	0.675	1.65	1.97	2.34	2.59
500	0.674	1.65	1.96	2.33	2.59
1000	0.674	1.65	1.96	2.33	2.58
∞	0.674	1.64	1.96	2.33	2.58

Kaynaklar

1. Probability and Statistical Inference, Robert V. Hogg, elliot A. Tanis, Macmillian Publ. Co. London
2. D.Huff, 1993 How to lie with statistics, W.W.Norton & co
3. Wilk,Samuel S. (1962), Mathematical Statistics, John Wiley, Bol.8.1
4. http://tr.wikipedia.org/wiki/%C3%96rnekleme
5. http://w3.gazi.edu.tr/~hasanbal/istnedir.html
6. http://www.cellotin.com/forum/istatistik/istatistigin_tanimi_ve_konusu-t1306.0.html
7. İstatistik, Doç. Dr. Ziya Gökalp GÖKTOLGA, Seçkin Yy, 12416, Ankara, 2013
8. BK