TAHMİN

Örneklem istatistikleri yardımıyla anakütle parametrelerinin tahmin edilmesi, nokta tahmini ve aralık tahmini olmak üzere iki şekilde gerçekleştirilir.

1. Nokta tahmini

Bir anakütle parametresinin tahmininde kullanılan istatistik değeri, *nokta tahmini* olarak adlandırılır.

$$\text{Ö}rneklem\ ortalaması\ (\bar{x}), \qquad Kitle\ ortalamasının\ (\mu\ )$$

$$\text{Ö}rneklem\ varyansı\ (s^2), \qquad Kitle\ varyansının\ (\sigma^2\ )$$

$$\text{Ö}rneklem\ oranı\ (p), \qquad Kitle\ oranının\ (\pi\ )$$

nokta tahminleridir.

Örnek: Bir tavuk çiftliği sahibi çiftlikte üretilen yumurtaların ortalama ağırlığını, varyansını ve beyaz renkte üretilen yumurta oranını tahmin etmek istemektedir. Bu amaçla, üretilen yumurtalardan 20 tanesi rasgele seçilmiş ve aşağıdaki bilgiler kaydedilmiştir.

Yumurta	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Ağırlık(gr)	42	39	36	44	41	38	36	48	43	39	40	36	42	37	35	44	42	39	45	40
Renk	S	B	B	S	B	S	S	B	S	B	B	B	S	B	S	B	B	S	B	B

Çiftlikte üretilen tüm yumurtalar için ortalama ağırlığı, varyansı ve beyaz yumurta oranını tahmin ediniz.

$n = 20$

$X_i : Tavuk\ çiftliğinde\ \text{ü}retilen\ yumurtanın\ ağırlığı\ (gr)$

$Y_i : Tavuk\ çiftliğinde\ \text{ü}retilen\ yumurtanın\ rengi\ (beyaz, sarı)$

$$\bar{x} = \frac{\sum X_i}{n} = \frac{806}{20} = 40.3\ gr$$

Çiftlikte üretilen tüm yumurtaların ortalama ağırlığı tahmini olarak(yaklaşık) 40.3 gr. dır.

$$s^2 = \frac{\sum(X_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{1}{20-1}[(42-40.3)^2 + (39-40.3)^2 + \cdots + (45-40.3)^2 + (40-40.3)^2] = 12.11$$

Çiftlikte üretilen tüm yumurtaların varyansı tahmini olarak(yaklaşık) 12.11 dir.

$$Beyaz\ yumurta\ oranı = p = \frac{12}{20} = 0.6$$

Çiftlikte üretilen yumurtaların yaklaşık %60 ‘ı beyaz renktedir.

Nokta tahminlerinin (tahmin edicilerin) özellikleri

Tahmin ediciler,

a) Sapmasızlık(Yansızlık)
b) Tutarlılık
c) Etkinlik(En küçük varyans)
d) Yeterlilik.

özelliklerine sahiptir.

2. Aralık tahmini (Güven Aralığı)

Kitle parametresini, belli bir hata payı ile içinde bulunduracak sınırların belirlenmesine *aralık tahmini* denir.

Parametrelerin içinde bulunacağı aralığa da *güven aral*ığı denir.

Güven aralığının belirlenmesinde, ilgilenilen parametreye ait kitle varyansının bilinip bilinmemesi ve örneklem büyüklüğüne bağlı olarak z dağılışı, t dağılışı ve $Ki - kare$ (χ^2) dağılışı kullanılmaktadır.

- **Kitle varyansı** $(\boldsymbol{\sigma^2})$ **biliniyorsa;** örneklem büyüklüğüne bakılmaksızın

standart normal z dağılışı kullanılır.

- **Kitle varyansı (σ^2) bilinmiyorsa;** kitle varyansı yerine örneklem varyansı kullanılır ve

$$n \geq 30 \; ise \;\; z \; dağılışı$$

$$n < 30 \; ise \;\; t \; dağılışı$$

kullanılır.

Ayrıca oranlar ile ilgili testler ve tahminlerde z dağılışı, varyans ile ilgili problemlerde $Ki - kare$ dağılışı kullanılır.

ANAKÜTLE ORTALAMASI (μ) İÇİN GÜVEN ARALIĞI

$$P(\bar{x} - z_{\alpha/2}\sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$$

formülü ile elde edilir. Formülde,

$\bar{x}$: $Aritmetik\ ortalama$

$z_{\alpha/2}$: $z\ tablo\ değeri$

$1 - \alpha$: $Güven\ katsayısı(güven\ düzeyi)$

α : $Anlamlılık$ düzeyi $(hata\ payı)$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$$

anlamındadır.

Örnek: Bir bölgeden alınan 150 kişilik örneklemde ortalama yaşam süresi 65 yıl bulunmuştur. Kitle varyansı 200 olduğuna göre kitle ortalamasının %95 lik güven sınırlarını bulunuz.

$n = 150$, $\bar{x} = 65$, $\sigma^2 = 200$, $1 - \alpha = 0{,}95$ verilmiş.

$\alpha = 0{,}05$ olur.

$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{200}{150}} = 1{,}15$ dir.

$z_{\alpha/2} = z_{0.05/2} = z_{0.025} = 1{,}96$

[0.5 – 0,025 = 0,475 değerini z tablosunun gövdesinden bul. Sola ve yukarı gidip sayıları birleştirince 1,96 bulunur.]

Yukarıdaki bilgileri güven aralığı formülünde yerine yazarsak;

$$P(\bar{x} - z_{\alpha/2}\sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_{\alpha/2}\sigma_{\bar{x}}) = 1 - \alpha$$

$$P(65 - (1{,}96)(1{,}15) \leq \mu \leq 65 + (1{,}96)(1{,}15)) = 0.95$$

$$P(62{,}746 \leq \mu \leq 67{,}254) = 0.95$$

“ % 95 güven seviyesinde kitle ortalaması 62,746 ile 67,254 değerleri arasındadır.”

“ Kitle ortalamasının 62,746 ile 67,254 değerleri arasında bulunma olasılığı % 95 tir.”

“ Hesaplanan güven aralığının μ parametresini içeren aralıklardan biri olma olasılığı % 95 tir.”

Örnek: Rasgele seçilen 25 öğrencinin not ortalaması 65 ve varyansı 121 olarak hesaplanmıştır. Kitle ortalaması için %95 'lik güven sınırlarını bulunuz.

$n = 25$, $\bar{x} = 65$, $s^2 = 121$, $1 - \alpha = 0{,}95$ verilmiş.

$\alpha = 0{,}05$ olur.

Kitle varyansı bilinmiyor.

Kitle varyansı yerine örneklem varyansı kullanılır ve $n = 25 < 30$ olduğundan t dağılışı kullanılır.

Bu durumda güven aralığı formülü aşağıdaki biçimde olur.

$$P(\bar{x} - (t_{n-1,\ \frac{\alpha}{2}})s_{\bar{x}} \leq \mu \leq \bar{x} + (t_{n-1,\ \frac{\alpha}{2}})s_{\bar{x}}) = 1 - \alpha$$

$s_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{121}{25}} = 2{,}2$ dir.

$t_{n-1\ ,\ \alpha/2} = t_{25-1\ ,\ 0.05/2} = t_{24\ ,\ 0.025} = 2{,}06$ dır.

[t ablosunun ilk sütunundan serbestlik derecesine bak, ilk satırından $\alpha/2$ ye bak ve birleştir]

Yukarıdaki bilgileri güven aralığı formülünde yerine yazarsak;

$$P(65 - (2{,}06)(2{,}2) \leq \mu \leq 65 + (2{,}06)(2{,}2)) = 0{,}95$$

$$P(60{,}468 \leq \mu \leq 69{,}532) = 0{,}95$$

“ Hesaplanan güven aralığının μ parametresini içeren aralıklardan biri olma olasılığı % 95 tir.”

“ % 95 güven seviyesinde kitle ortalaması 60,468 ile 69,532 değerleri arasındadır.”

“ Kitle ortalamasının 60,468 ile 69,532 değerleri arasında bulunma olasılığı % 95 tir.”

Örnek: 200 hastaya ait hemoglobin değerlerinin ortalaması 11,8 ve varyansı 265,8 olarak hesaplanmıştır. Hastalara ait gerçek hemoglobin ortalamasının %99'luk güven sınırlarını tahmin ediniz.

$n = 200$, $\bar{x} = 11{,}8$, $s^2 = 265{,}8$, $1 - \alpha = 0{,}99$ verilmiş.

$\alpha = 0{,}01$ olur.

Kitle varyansı bilinmiyor.

Kitle varyansı yerine örneklem varyansı kullanılır ve $n = 200 > 30$ olduğundan z dağılışı kullanılır.

$$P(\bar{x} - z_{\alpha/2} s_{\bar{x}} \leq \mu \leq \bar{x} + z_{\alpha/2} s_{\bar{x}}) = 1 - \alpha$$

$s_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{265{,}8}{200}} = 1{,}15$ dir.

$z_{\alpha/2} = z_{0.01/2} = z_{0.005} = 2{,}57$

$$P(11{,}8 - (2{,}57)(1{,}15) \leq \mu \leq 11{,}8 + (2{,}57)(1{,}15)) = 0.99$$

$$P(8{,}84 \leq \mu \leq 14{,}75) = 0.99$$

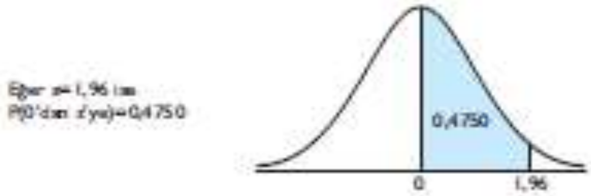
" Hesaplanan güven aralığının μ parametresini içeren aralıklardan biri olma olasılığı % 99 dur."

" % 99 güven seviyesinde kitle ortalaması 8,84 ile 14,75 değerleri arasındadır."

" Kitle ortalamasının 8,84 ile 14,75 değerleri arasında bulunma olasılığı % 99 dur."

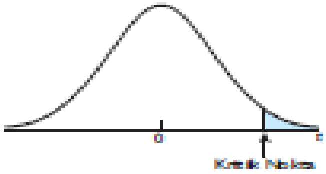
z tablosu (s.229)

Ek-1: Standart Normal Dağılım Altında Kalan Alan (z) Tablosu



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,10	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,20	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,30	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,40	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,50	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,60	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,70	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,80	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,90	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,00	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,10	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,20	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,30	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,40	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,50	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,60	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,70	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,80	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,90	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,00	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,10	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,20	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,30	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,40	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,50	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,60	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,70	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,80	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,90	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,00	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

t tablosu (s.230)



s.d.	α 0,25	0,10	0,05	0,025	0,010	0,005	0,0025	0,0010	0,0005
1	1,00	3,08	6,31	12,71	31,82	63,66	127,32	318,31	636,62
2	0,82	1,89	2,92	4,30	6,96	9,92	14,09	22,33	31,60
3	0,76	1,64	2,35	3,18	4,54	5,84	7,45	10,21	12,92
4	0,74	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
5	0,73	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87
6	0,72	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
7	0,71	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
8	0,71	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
9	0,70	1,38	1,83	2,26	2,82	3,25	3,69	4,30	4,78
10	0,70	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
11	0,70	1,36	1,80	2,20	2,72	3,11	3,50	4,02	4,44
12	0,70	1,36	1,78	2,18	2,68	3,05	3,43	3,93	4,32
13	0,69	1,35	1,77	2,16	2,65	3,01	3,37	3,85	4,22
14	0,69	1,35	1,76	2,14	2,62	2,98	3,33	3,79	4,14
15	0,69	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
16	0,69	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,01
17	0,69	1,33	1,74	2,11	2,57	2,90	3,22	3,65	3,97
18	0,69	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
19	0,69	1,33	1,73	2,09	2,54	2,86	3,17	3,58	3,88
20	0,69	1,33	1,72	2,09	2,53	2,85	3,15	3,55	3,85
21	0,69	1,32	1,72	2,08	2,52	2,83	3,14	3,53	3,82
22	0,69	1,32	1,72	2,07	2,51	2,82	3,12	3,50	3,79
23	0,69	1,32	1,71	2,07	2,50	2,81	3,10	3,48	3,77
24	0,68	1,32	1,71	2,06	2,49	2,80	3,09	3,47	3,75
25	0,68	1,32	1,71	2,06	2,49	2,79	3,08	3,45	3,73
26	0,68	1,31	1,71	2,06	2,48	2,78	3,07	3,43	3,71
27	0,68	1,31	1,70	2,05	2,47	2,77	3,06	3,42	3,69
28	0,68	1,31	1,70	2,05	2,47	2,76	3,05	3,41	3,67
29	0,68	1,31	1,70	2,05	2,46	2,76	3,04	3,40	3,66
30	0,68	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
40	0,68	1,30	1,68	2,02	2,42	2,70	2,97	3,31	3,55
60	0,68	1,30	1,67	2,00	2,39	2,66	2,91	3,23	3,46
120	0,68	1,29	1,66	1,98	2,36	2,62	2,86	3,16	3,37
∞	0,67	1,28	1,64	1,96	2,33	2,58	2,81	3,09	3,29

χ^2 (Kİ-KARE) TESTLERİ

İkisi de sınıflama düzeyinde ya da biri sınıflama diğeri sıralama düzeyinde ölçülmüş değişkenlere ait veri, çapraz tabloda özetlendiğinde hesaplanabilecek ilişki katsayılarından çoğu ki-kare ölçüsüne bağlıdır.

Ki-kare değeri,

$$\boldsymbol{\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(f_{g_{ij}} - f_{b_{ij}})^2}{f_{b_{ij}}}}$$

formülünden hesaplanır. Formülde,

$f_{g_{ij}}$: Gözlenen frekans

$f_{b_{ij}}$: Beklenen frekans

anlamındadır.

χ^2 **ölçüsüne bağlı ilişki katsayılarından** bazıları aşağıda verilmiştir:

i) $\phi = \sqrt{\frac{\chi^2}{N}}$

ii) $c = \sqrt{\frac{\chi^2}{\chi^2+N}}$

iii) $v = \sqrt{\frac{\chi^2}{N[min\{(I-1),(J-1)\}]}}$; $I : satır\ sayısı$, $J : sütun\ sayısı$, $0 \leq v \leq 1$ dir.

2x4 boyutlu çapraz tabloya bir örnek aşağıda verilmiştir.

EİTVPT / Cinsiyet	Spor	Haber	Magazin	Diğer	TOPLAM
Bay	140	120	20	20	300
Bayan	20	80	80	20	200
TOPLAM	160	200	100	40	500

SORU: Yukarıda verilen çapraz tablodaki değişkenler arasında ilişki var mıdır? Varsa ne kadarlık, nasıl bir ilişki vardır?

EİTVPT / Cinsiyet	Spor		Haber		Magazin		Diğer		TOPLAM
Bay	140	96	120	120	20	60	20	24	300
Bayan	20	64	80	80	80	40	20	16	200
TOPLAM	160		200		100		40		500

Sarı renkli olanlar, beklenen frekanslardır.

"[(Satır toplamı)x(sütun toplamı)] / genel toplam" 'dan hesaplanmıştır.

Değişkenler ve ölçme düzeyleri:

……

…….

Değişkenler arasındaki ilişkiyi

$$v = \sqrt{\frac{\chi^2}{N[min\{(I-1),(J-1)\}]}}$$

ile hesaplayabiliriz.

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(f_{g_{ij}} - f_{b_{ij}})^2}{f_{b_{ij}}}$$

$$= \frac{(140-96)^2}{96} + \frac{(120-120)^2}{120} + \frac{(20-60)^2}{60} + \frac{(20-24)^2}{24} + \frac{(20-64)^2}{64} + \frac{(80-80)^2}{80}\frac{(80-40)^2}{40} + \frac{(20-16)^2}{16} = 118{,}7$$

olup,

$$v = \sqrt{\frac{118{,}7}{500[min\{(2-1),(4-1)\}]}} = 0{,}48$$

bulunur. Değişkenler arasında %48 'lik orta düzeyde bir ilişki vardır.