İST254: Mühendisler İçin İstatistik

Ders 9: Kitle Ortalaması ve Varyansı için Tahmin

19.04.2013

Ders İçeriği

- Kitle ve Örneklem
- Örneklem Dağılımı
- Nokta Tahmini
- Tahmin Edicilerin Özellikleri
- Kitle ortalaması için Aralık Tahmini
- Kitle Standart Sapması için Aralık Tahmini
- İki Kitlenin Ortalamalarının Farkı için Aralık Tahmini
- Ölçüm Çiftleri için Aralık Tahmini

Kitle ve Örneklem

- Üzerinde çalışılan tüm gruba, ya da elde edilebilecek tüm sonuçlara **kitle** denir.
- Kitlenin her bir üyesi için ölçülebilir olan değişkenlere **karakteristik** denir.
- Kitleyi tanımlayan sayılsal ölçütlere **parametre** denir.

Kitle ve Örneklem

- Kitlenin belli bir veya birden fazla özelliğini incelemek üzere seçilen bireyler topluluğuna **örneklem** denir
- Örneklemin tanımlayıcı istatistiklerine **örneklem istatistiği** denir.
- Bir istatistiğin dağılımına **örneklem dağılımı** denir.

Örneklem Ortalaması ve Dağılımı

- Örneklem ortalaması $\bar{X}$ ile gösterilir.
- $E(\bar{X}) = E\frac{X_1+X_2+\cdots+X_n}{n} = \frac{1}{n}[E(X_1)+E(X_2)+\cdots+E(X_n)] = \mu_x$
- $Var(\bar{X}) = Var\left[\frac{1}{n}(X_1+X_2+\cdots+X_n)\right] = \frac{1}{n^2}[Var(X_1)+Var(X_2)+\cdots+Var(X_n)] = \frac{1}{n}{\sigma_x}^2$
- $\sigma_{\bar{X}} = \frac{1}{\sqrt{n}}\sigma_x$

Merkezi Limit Teoremi

$X_1, X_2, \ldots X_n$; μ ortalamalı ve σ^2 varyanslı bağımsız rastgele değişkenler olsunlar; n yeterince büyükse bunların ortalaması olan $\bar{X}$, μ ortalamalı ve $\frac{\sigma^2}{n}$ varyanslı normal dağılım gösterir

Nokta Tahmini

- Bir parametre için tek değerden oluşan tahmine **nokta tahmini** denir.
- Kitle ortalaması için örneklem ortalaması, kitle varyansı için örneklem standart sapması yaygın olarak kullanılan **tahmin edicilerdir**.
- Tahmin ile parametre arasındaki farka **hata** denir.

Tahmin Edicinin Özellikleri

- **Yansızlık (Sapmasızlık):** Tahmin edicinin beklenen değerinin tahmin edilen parametre ile aynı olmasıdır $E(\theta) = \theta$
- **Etkinlik:** Tahmin edicinin varyansının az olmasıdır. İki tahmin edici arasında varyansı daha az olan, fazla olandan daha etkindir
- **Tutarlılık:** Örneklem boyutu büyüdükçe tahminin hatasının sıfıra yaklaşmasıdır.
- **Yeterlilik:** Örneklemden alınan tüm bilgiyi kullanan tahmin edicilere yeterli denir.

Aralık Tahmini

- Nokta tahminleri, parametrenin tahmine ne kadar yakın olduğu konusunda bilgi vermez.
- Kitle parametresini belirli bir olasılıkla kapsayan aralığa **Güven Aralığı** denir.
- %95, %99 ve %90 en sık kullanılan güven aralıklarıdır.
- **Yanılgı payı** olarak adlandırılan α, parametrenin güven aralığı dışında kalma olasılığını verir.

Kitle Ortalaması için Aralık Tahmini: Bilinen Varyans

- Varyans biliniyorsa; örneklem ortalaması $\bar{X}$'ın, kitle ortalaması μ ortalama ve $\frac{\sigma}{\sqrt{n}}$ standart sapmaya sahip normal dağılıma sahip rastgele değişkendir.
- $P\left[\mu - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$
- Kitle ortalaması μ için $1-\alpha$ güven aralığı

$$\bar{X} \pm Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

şeklindedir

Alıştırmalar

Bir LPG tesisindeki dolum prosesinin standart sapmasının 0,15kg olduğu bilinmektedir. Rastgele seçilen 25 tüpün ortalama ağırlıkları 12,05kg olarak ölçülmüştür. Buna göre, tesiste doldurulan tüplerin ortalama ağırlığı için %95 güven aralığı nedir?

Alıştırmalar

Bir LPG tesisindeki dolum prosesinin standart sapmasının 0,15kg olduğu bilinmektedir. Rastgele seçilen 25 tüpün ortalama ağırlıkları 12,05kg olarak ölçülmüştür. Buna göre, tesiste doldurulan tüplerin ortalama ağırlığı için %95 güven aralığı nedir?

$$\bar{X} \pm Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 12{,}05 \pm 1{,}96 \times \frac{0{,}15}{\sqrt{25}}$$
$$= 12{,}05 \pm 0{,}0588kg$$

Örneklem Büyüklüğünün Belirlenmesi

- Örneklemin büyük olması güven aralığını daraltır ancak bu ek maliyet getirir.
- α olasılık ile gerçekleşecek maksimum tahmin hatası e verildiyse, bunu sağlayacak en küçük örneklem boyutu

$$n = \left[Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{e}\right]^2$$

şeklinde hesaplanır

Alıştırmalar

Bir gazlı içecek dolum tesisinde her bir şişeye konulan içecek miktarının standart sapmasının 2ml olduğu bilinmektedir. Şişelere doldurulan ortalama içecek miktarının en fazla 0,5 ml hata ve %99 güven ile hesaplanabilmesi için en az kaç şişe ölçülmelidir?

Alıştırmalar

Bir gazlı içecek dolum tesisinde her bir şişeye konulan içecek miktarının standart sapmasının 2ml olduğu bilinmektedir. Şişelere doldurulan ortalama içecek miktarının en fazla 0,5 ml hata ve %99 güven ile hesaplanabilmesi için en az kaç şişe ölçülmelidir?

$$n = \left[Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{e}\right]^2 = \left[2{,}58 \times \frac{2ml}{0{,}05ml}\right]^2 \cong 106{,}5$$

En az 107 şişe ölçülmelidir.

χ^2 ve t dağılımları

- **χ^2 dağılımı:** $X_i \sim N(\mu, \sigma^2)$ ve X_i bağımsız olsun $U = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$ rastgele değişkenine χ^2 rastgele değişken denir.
- **t dağılımı:** $X_i \sim N(\mu, \sigma^2)$ ve X_i bağımsız olsun $V = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$ t dağılımı gösterecektir.
- Bağımsız gözlem sayısı ile tahmin edilecek parametre sayısı arasındaki farka **serbestlik derecesi** denir

Kitle Ortalaması için Aralık Tahmini: Bilinmeyen Varyans

- Kitle varyansının bilinmediği durumda, kitle varyansı yerine örneklem varyansı kullanılır.
- Kitle varyansı ile örneklem varyansı arasındaki örneklem standart sapmasına göre normalize edilmiş fark t dağılımı gösterecektir.
- Kitle ortalaması için $1-\alpha$ güven aralığı:

$$\bar{X} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

Alıştırmalar

Çevre kirliliği üzerine yapılan bir araştırmada, bir akarsu üzerinde rastgele seçilen 36 noktada pH ölçümü yapıltır. Ölçümlerin ortalaması 5,4; varyansı 0,35 olduğuna göre akarsuyun ortalama pH değeri için %99 güven aralığı nedir?

Alıştırmalar

Çevre kirliliği üzerine yapılan bir araştırmada, bir akarsu üzerinde rastgele seçilen 36 noktada pH ölçümü yapıltır. Ölçümlerin ortalaması 5,4; varyansı 0,35 olduğuna göre akarsuyun ortalama pH değeri için %99 güven aralığı nedir?

$$\bar{X} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}} = 5{,}4 \pm 2{,}7239 \times \frac{0{,}35}{\sqrt{36}} \cong 5{,}4 \pm 0{,}159$$

Kitle Varyansı için Aralık Tahmini

- Normal dağılımli bir kitleden alınan örneklemin varyansı χ^2 dağılımı gösterecektir.
- Normal dağılımlı kitlenin varyansı için $1-\alpha$ güven aralığı:

$$\left[\frac{(n-1)s^2}{{\chi^2}_{1-\frac{\alpha}{2}}};\frac{(n-1)s^2}{{\chi^2}_{\frac{\alpha}{2}}}\right]$$

şeklindedir

Alıştırmalar

Bir fabrikada üretilmiş kiremitlerden rastgele 12'si seçilerek genişlikleri ölçülmüş, örneklemin varyansının 1,4mm olduğu hesaplanmıştır. Kitle varyansı için %90 güven aralığını bulunuz.

Alıştırmalar

Bir fabrikada üretilmiş kiremitlerden rastgele 12'si seçilerek genişlikleri ölçülmüş, örneklemin varyansının 1,4mm^2 olduğu hesaplanmıştır. Kitle varyansı için %90 güven aralığını bulunuz.

$$\left[\frac{(n-1)s^2}{{\chi^2}_{1-\frac{\alpha}{2}}};\frac{(n-1)s^2}{{\chi^2}_{\frac{\alpha}{2}}}\right]$$

$$=\left[\frac{11\times 1{,}4mm^2}{19{,}675};\frac{11\times 1{,}4mm^2}{4{,}575}\right]$$

$$=[0{,}783mm^2;3{,}366mm^2]$$

İki Kitlenin Ortalamalarının Farkı için Aralık Tahmini: Bilinen Varyans

- Birinci kitleden alınan n örneğin ortalaması $\bar{X}_1$, ikinci kitleden alınan m örneğin ortalaması $\bar{X}_2$ olsun.
- $var(\bar{X}_1 - \bar{X}_2) = \frac{{\sigma_1}^2}{n} + \frac{{\sigma_2}^2}{m}$ olacaktır.
- $\mu_1 - \mu_2$ için $1 - \alpha$ güven aralığı

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{{\sigma_1}^2}{n} + \frac{{\sigma_2}^2}{m}}$$

şeklindedir.

Alıştırmalar

Bir fabrikada iki farklı tankta etil alkol üretilmektedir. Farklı günlerde 15'er kez her iki tanktaki üretimin saflığı ölçülmüş, ve aşağıdaki değerler elde edilmiştir, üretim prosesinin standart sapması %0,8 olduğuna göre farklı tanklarda üretilen etanolün saflıkları arasındaki fark için %90 güven aralığını bulunuz.

Tank 1			Tank 2		
95,24	94,81	93,60	95,40	95,10	95,21
94,94	93,80	94,63	95,73	95,38	95,49
95,30	93,91	95,28	94,23	94,15	96,60
94,48	93,49	94,17	94,91	95,02	96,04
95,33	96,07	94,95	96,36	95,29	94,71

Alıştırmalar

Tank 1			Tank 2		
95,24	94,81	93,60	95,40	95,10	95,21
94,94	93,80	94,63	95,73	95,38	95,49
95,30	93,91	95,28	94,23	94,15	96,60
94,48	93,49	94,17	94,91	95,02	96,04
95,33	96,07	94,95	96,36	95,29	94,71

$\bar{X}_1 = 94{,}67 \qquad \bar{X}_2 = 95{,}31$

$\mu_1 - \mu_2$ için %90 güven aralığı:

$$(94{,}67 - 95{,}31) \pm 1{,}65\sqrt{\frac{0{,}8^2}{15} + \frac{0{,}8^2}{15}}$$

$$(-0{,}64 \pm 0{,}482\ )\%$$

İki Kitlenin Ortalamalarının Farkı için Aralık Tahmini: Bilinmeyen Varyans

- Varyansları bilinmeyen ancak normal dağılımlı ve eşit olduğu kabul edilen iki kitle için ortak varyansı:
- ${s_p}^2 = \frac{(n-1){s_1}^2+(m-1){s_2}^2}{n+m-2}$ şeklinde hesaplanır
- $\mu_1 - \mu_2$ için $1-\alpha$ güven aralığı

$$(\bar{X}_1 - \bar{X}_2) \pm t_{n+m-2,1-\frac{\alpha}{2}} \times s_p \times \sqrt{\frac{1}{n} + \frac{1}{m}}$$

şeklindedir.

Alıştırmalar

Hem Elektronik hem de Bilgisayar Mühendisliği bölümleri için zorunlu olan bir dersin sınavında 25 elektronik bölümü öğrencisinin ortalaması 76, standart sapması 18; 49 bilgisayar mühendisliği öğrencisinin ise ortalaması 74, standart sapması 16'dır. İki bölümün öğrencilerinin notlarının eşit varyanslı ve normal dağıldığı kabul edilmektedir. Buna göre iki bölümün öğrencilerinin notlarının beklenen değerleri arasındaki fark için %99 güven aralığı nedir?

Alıştırmalar

	Elektronik	Bilgisayar
Öğrenci sayısı	25	49
Ortalama	76	74
Standart sapma	18	16

$$s_p{}^2 = \frac{(n-1)s_1{}^2+(m-1)s_2{}^2}{n+m-2} = \frac{24\times18^2+48\times16^2}{25+49-2} \cong 278{,}7$$

Fark için %99 güven aralığı:

$$(76-74) \pm t_{72;0,995} \times 16{,}7 \times \sqrt{\frac{1}{25}+\frac{1}{49}}$$

$$2 \pm 10{,}9$$

Ölçüm Çiftleri için Aralık Tahmini

- Bazen gözlemler çiftler halinde yapılır (ör. aynı kişinin tansiyonu bir ilacı almadan önce ve aldıktan sonra ölçüldüyse); kitle ortalamalarının farkı, farkların ortalamasından daha anlamlıdır.
- Farklar $D_i = X_{1i} - X_{2i}$ şeklinde hesaplanır.
- μ_D için $1 - \alpha$ güven aralığı:

$$\bar{D} \pm t_{n-1,1-\frac{\alpha}{2}} \frac{S_D}{\sqrt{n}}$$

Alıştırmalar

Yeni geliştirilmiş bir yakıt katkısının yakıt tüketimine etkisi üzerinde yapılan bir araştırmada 15 farklı model otomobil benzer şartlar altında katkılı ve katkısız yakıt ile sürülmüştür. 100 km'deki ortalama yakıt tüketimleri aşağıda verilmiştir. Buna göre katkının yakıt tüketimine etkisi için %90 güven aralığını hesaplayınız.

Araçlar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Katkılı	8,31	7,33	8,66	6,59	6,68	6,43	6,42	7,28	7,4	7,61	7,09	6,52	8,23	6,58	7,36
Katkısız	9,48	7,7	9,57	7,82	7,8	7,24	6,2	7,33	7,73	7,96	7,41	6,56	8,03	6,35	7,58

Alıştırmalar

Araçlar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Katkılı	8,31	7,33	8,66	6,59	6,68	6,43	6,42	7,28	7,4	7,61	7,09	6,52	8,23	6,58	7,36
Katkısız	9,48	7,7	9,57	7,82	7,8	7,24	6,2	7,33	7,73	7,96	7,41	6,56	8,03	6,35	7,58
Fark	1,17	0,37	0,91	1,23	1,12	0,81	-0,22	0,05	0,33	0,35	0,32	0,04	-0,2	-0,23	0,22

$\overline{D} = 0{,}418$

$S_D = 0{,}512$

μ_D için %90 güven aralığı:

$$0{,}418 \pm t_{14;0{,}90} \frac{0{,}512}{\sqrt{15}}$$

$$0{,}418 \pm 0{,}178$$