Marmara Üniversitesi

İstatistik Bölümü

Örnekleme Teorisi (3)

Doç. Dr. Atıf Evren

Sonlu Ana Kütleler İçin Örnek Ortalamasının Dağılımı

Teorem: Eğer $X_1, X_2, \cdots, X_n$ N büyüklüğünde bir ana kütleden sırası ile çekilen birinci gözlem, ikinci gözlem;…; n. gözlem olsun. Bu durumda bu n rastlantı değişkeninin birleşik olasılık fonksiyonu

$f(x_1, \cdots, x_n) = \dfrac{1}{N(N-1)\cdots(N-n+1)}$, $f(x_r)$ marjinal olasılığı da

$$f_{X_r}(x_r) = \frac{1}{N} \quad x_r = c_1, c_2, \cdots, c_N \quad r = 1, 2, \cdots, n$$

Burada sonlu $\{c_1, c_2, \cdots, c_N\}$ ana kütlenin beklenen değeri ve varyansı

$\mu = \sum_{i=1}^{N} c_i . \dfrac{1}{N}$ ve $\sigma^2 = \sum_{i=1}^{N} \left(c_i - \mu\right)^2 . \dfrac{1}{N}$ şeklindedir.

Son olarak $X_1, X_2, \cdots, X_n$ değişkenleri içerisinden herhangi ikisinin birleşik marjinal olasılık fonksiyonu da sonlu ana kütlenin herhangi sıralı ikilisi için

$f_{X_r, X_s}(x_r, x_s) = \dfrac{1}{N(N-1)}$ şeklindedir.

Teorem: X_r *ve* X_s sonlu $\{c_1, c_2, \cdots, c_N\}$ ana kütlesinden çekilen n birimlik bir örneğin r. ve s. rastlantı değişkenleri olsun. Bu durumda

$Cov(X_r, X_s) = -\dfrac{\sigma^2}{N-1}$ olur.

İspat: Kovaryansın tanım formülüne göre

$$Cov(X_r, X_s) = \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{N(N-1)}(c_i-\mu)(c_j-\mu) \quad i \neq j$$

$$= \frac{1}{N(N-1)}\sum_{i=1}^{N}(c_i-\mu)\left[\sum_{j=1}^{N}(c_j-\mu)\right] \quad i \neq j \quad \text{olur.}$$

$i = j$ durumu söz konusu olamayacağına göre

$$\sum_{j=1}^{N}(c_j-\mu) = \sum_{j=1}^{N}(c_j-\mu)-(c_i-\mu) = -(c_i-\mu)$$

$$Cov(X_r, X_s) = -\frac{1}{N(N-1)}\sum_{i=1}^{N}(c_i-\mu)^2 = -\frac{1}{(N-1)}\sigma^2 \quad \text{bulunmuş olur.}$$

<u>Teorem:</u> Ortalaması ve varyansı μ *ve* σ^2 olan **N** büyüklüğünde bir

ana kütleden çekilen n birimlik bir örneğin ortalaması $\overline{X}$ olsun.

$E(\overline{X}) = \mu$ ve $Var(\overline{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1}$ olur.

<u>İspat:</u> $Var(X_i) = \sigma^2$ ve $Cov(X_i, X_j) = -\frac{\sigma^2}{N-1}$

değerleri beklenen değerin ve varyansın tanım formüllerinde yerine

konulacak olursa $E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \mu$ bulunur.

$$Var(\overline{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^{n}X_i\right) = \frac{1}{n^2}\left\{\sum_{i=1}^{n}Var(X_i)+2\sum\sum_{i<j}Cov(X_i, X_j)\right\}$$

$$\frac{1}{n^2}\left\{\sum_{i=1}^{n}Var(X_i)+2\sum\sum_{i<j}Cov(X_i, X_j)\right\}$$

$$= \frac{1}{n^2}\left\{n\sigma^2 - \frac{2\sigma^2}{N-1}\sum_{j=1}^{n}\sum_{i=1}^{j-1}1\right\}$$

$$= \frac{\sigma^2}{n^2}\left\{n - \frac{2}{N-1}\sum_{j=1}^{n}(j-1)\right\}$$

$$= \frac{\sigma^2}{n^2}\left\{n - \frac{2}{N-1}\left\{\frac{n(n+1)}{2}-n\right\}\right\}$$

$$=\frac{\sigma^2}{n}\frac{N-n}{N-1}$$

Bu şekilde ispat sağlanmış olur.

Örnek: Ortalaması ve varyansı $\mu = 20 \quad ve \quad \sigma^2 = 16$ olan N=100 büyüklüğünde bir ana kütleden çekilen n=25 birimlik bir örneğin ortalaması $\bar{X}$ olsun. $\bar{X}$ 'in beklenen değerini ve varyansını bulunuz.

Çözüm: $E(\bar{X}) = \mu$ =20;

$Var(\bar{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{16}{25}\frac{(100-25)}{(100-1)} = \frac{16}{33}$ bulunur.

Teorem: $X_{11}, X_{12}, \cdots, X_{1n_1}$ ve $X_{21}, X_{22}, \cdots, X_{2n_2}$ sırasıyla μ_1, σ_1^2 ve μ_2, σ_2^2 beklenen değer ve varyanslarına sahip iki sonsuz ana kütleden çekilmiş n_1 ve n_2 birimden oluşan iki bağımsız örnek olsun. Yine $\bar{X}_1$ ve $\bar{X}_2$ bu iki örneğin ortalaması olsun.

$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$ ve

$Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ olur.

Örnek: $\mu_1 = 25$, $\sigma_1^2 = 16$ ve $\mu_2 = 27$, $\sigma_2^2 = 9$ parametreli iki sonsuz kabul edilen ana kütleden $n_1 = 54$ ve $n_2 = 45$ birimden oluşan iki bağımsız örneğin ortalamaları $\bar{X}_1$ ve $\bar{X}_2$ olsun. Merkezi Limit Teoremine göre $P\left(\left|\bar{X}_1 - \bar{X}_2\right| \le 0,5\right)$ olasılığını hesaplayınız.

Çözüm:

$$E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2 = -2 \ , \quad Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{16}{54} + \frac{27}{45}$$

$$Var(\bar{X}_1 - \bar{X}_2) = 0,897 \ , \quad \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{0,897} = 0,947$$

$$P\left(\left|\bar{X}_1 - \bar{X}_2\right| \le 0,5\right) = P\left(-0,5 \le \bar{X}_1 - \bar{X}_2 \le 0,5\right)$$

$$= P\left(\frac{-0,5-(-2)}{0,947} \leq \frac{\left(\bar{X}_1-\bar{X}_2\right)-E\left(\bar{X}_1-\bar{X}_2\right)}{\sigma_{\bar{X}_1-\bar{X}_2}} \leq \frac{0,5-(-2)}{0,947} \right)$$

$$= P\left(\frac{1,5}{0,947} \leq \frac{\left(\bar{X}_1-\bar{X}_2\right)-\left(\mu_1-\mu_2\right)}{\sigma_{\bar{X}_1-\bar{X}_2}} \leq \frac{2,5}{0,947} \right) = P\left(1,58 \leq Z \leq 2,64\right) = 0,4959-0,4429 = 0,053$$

Teorem: $X_{11}, X_{12}, \cdots, X_{1n_1}$ ve $X_{21}, X_{22}, \cdots, X_{2n_2}$ sırasıyla θ_1 ve θ_2 parametreli Bernoulli dağılımlarına uyan iki sonsuz ana kütleden çekilmiş (ve n_1 ve n_2 birimden oluşan) iki bağımsız örnek olsun. Yine $\hat{\theta}_1$ ve $\hat{\theta}_2$ bu iki örneğin oranları olsun.

$E\left(\hat{\theta}_1-\hat{\theta}_2\right)=\theta_1-\theta_2$ ve

$Var(\hat{\theta}_1-\hat{\theta}_2)=\frac{\theta_1(1-\theta_1)}{n_1}+\frac{\theta_2(1-\theta_2)}{n_2}$ olur.

Örnek: $\theta_1=0,54$ ve $\theta_2=0,49$ parametreli Bernoulli dağılımlarına uyan iki sonsuz ana kütleden çekilen , $n_1=100$ ve $n_2=64$ birimden oluşan iki bağımsız örneğin oranları $\hat{\theta}_1$ ve $\hat{\theta}_2$ olsun. Merkezi Limit Teoremi'nden yararlanarak $P\left\{\left(\hat{\theta}_1-\hat{\theta}_2\right)\geq 0\right\}$ olasılığını hesaplayınız.

Çözüm $E(\hat{\theta}_1-\hat{\theta}_2)=\theta_1-\theta_2=0,05$ ve

$$Var(\hat{\theta}_1-\hat{\theta}_2)=\frac{\theta_1(1-\theta_1)}{n_1}+\frac{\theta_2(1-\theta_2)}{n_2}$$

$$Var(\hat{\theta}_1-\hat{\theta}_2)=\frac{0,54*0,46}{100}+\frac{0,49*0,51}{64}$$

$$Var(\hat{\theta}_1-\hat{\theta}_2)\cong 0,0064$$

$$\sigma_{\hat{\theta}_1-\hat{\theta}_2}=\sqrt{0,0064}=0,08$$

$$P\left\{\left(\frac{\hat{\theta}_1-\hat{\theta}_2-E\left(\hat{\theta}_1-\hat{\theta}_2\right)}{\sigma_{\hat{\theta}_1-\hat{\theta}_2}}\right)\geq \frac{0-0,05}{0,08}\right\}=P(Z\geq -0,625)$$

$P(Z\geq -0,625)\cong 0,7357$ bulunmuş olur.

Normallik Varsayımından Türeyen Dağılımlar

1) Ki-Kare Dağılımı

Gama dağılımında $\alpha = \frac{\nu}{2}$ ve $\lambda = \frac{1}{2}$ denecek olursa istatistikte sıkça kullanılan Ki-Kare (χ^2) dağılımı elde edilecektir. Gama dağılımının olasılık yoğunluk ilgili parametre değerleri yerlerine konacak olursa

$$f(x) = \frac{e^{-x/2} x^{\frac{\nu-2}{2}}}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} \quad x \geq 0$$

Ki-Kare dağılımının olasılık yoğunluk fonksiyonu elde edilir.

Ki-Kare Dağılımının Beklenen Değeri ve Varyansı

Yine α ve λ parametreli bir gama dağılımında $\alpha = \frac{\nu}{2}$ ve $\lambda = \frac{1}{2}$ değerleri yerine konacak olursa ki-kare dağılımı için $E(\chi^2) = \nu$ ve $Var(\chi^2) = 2\nu$ bulunur.

Burada ν**;** χ^2 dağılımının tek parametresidir ve serbestlik derecesi olarak adlandırılmaktadır.

Teorem: Eğer $\overline{X}$ ve S^2 ortalaması **μ** ve varyansı σ^2 olan normal bir dağılımdan çekilen n birimlik bir örneğin ortalaması ve varyansı ise

1) $\overline{X}$ ve S^2 bağımsızdır.

2) $\frac{(n-1)S^2}{\sigma^2}$ rastlantı değişkeni n-1 serbestlik dereceli bir **Ki-kare** dağılımına uyar

Örnek: n=100 birimlik bir örnek normal dağıldığı bilinen bir ana kütleden çekiliyor. S^2 ve σ^2 sırası ile örnek ve ana kütle varyansları olduklarına göre $P(S^2 \leq 0,75\sigma^2)$ olasılığını hesaplayınız.

Çözüm:

$$P(S^2 \leq 0,75\sigma^2) = P\left(\frac{(n-1)S^2}{\sigma^2} \leq \frac{(100-1).0,75.\sigma^2}{\sigma^2}\right) = P\left(\frac{(n-1)S^2}{\sigma^2} \leq 74,25\right)$$

Öte yandan $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$ olacağına göre n-1=99 serbestlik dereceli bir Ki-Kare dağılımından (Ms-Excel kullanarak) $P\left(\chi^2(99) \leq 74,25\right) = 0,03$ bulunmuş olur.

2) Student-t Dağılımı

İstatistik'te önemli bir konumu bulunan olasılık dağılımlarından bir tanesi de bu dağılımı ortaya atan William S. Gosset'in kullandığı takma ad nedeniyle (student) literatüre "Student-t" adı ile girmiş bulunan dağılımdır. Student- t dağılımının olasılık yoğunluk fonksiyonu

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \qquad -\infty < t < \infty$$

olur. Burada ν dağılımın tek parametresidir ve serbestlik derecesi olarak adlandırılmaktadır.

Student-t Dağılımının Beklenen Değeri ve Varyansı

$$E(T) = 0$$

$$Var(T) = \frac{\nu}{\nu-2}; \quad \nu > 2$$

Teorem: Eğer $\bar{X}$ ve S^2 ortalaması μ ve varyansı σ^2 olan normal bir dağılımdan çekilen n birimlik bir örneğin ortalaması ve varyansı ise $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ rastlantı değişkeni n-1 serbestlik dereceli bir **t** dağılımına uyar.

Örnek: Normal dağıldığı bilinen bir ana kütleden **n=65** birimlik bir örnek çekiliyor. $S^2 = 400$ bulunduğuna göre $P\left\{\left(\bar{X}-\mu\right) \leq 2\right\}$ olasılığını bulunuz.

Çözüm:

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1} \qquad P\left\{\frac{\left(\bar{X}-\mu\right)}{S/\sqrt{n}} \leq \frac{2}{20/\sqrt{65}}\right\} = P\left(t_{n-1} \leq 0,806\right)$$

Ms-Excel kullanarak $P\left(t_{n-1} \leq 0,806\right) = 0,788$ bulunmuş olur.

3) F- Dağılımı

Normal dağılımlardan örnekleme sırasında karşımıza çıkan önemli bir dağılım da Snedecor F dağılımıdır**.** F dağılımı İki bağımsız gelen χ^2 değişkenlerinin (her biri kendi serbestlik derecesine bölünmüş halde) oranlarının olasılık dağılımı olarak düşünülebilir. F-Dağılımının olasılık yoğunluk fonksiyonu

$$g(f) = \frac{\Gamma\left(\frac{v_1+v_2}{2}\right)}{\Gamma\left(\frac{v_1}{2}\right)\Gamma\left(\frac{v_2}{2}\right)}\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}} f^{\frac{v_1}{2}-1}\left(1+\frac{v_1}{v_2}f\right)^{-\frac{1}{2}(v_1+v_2)}$$

$$f > 0$$

Burada v_1 ve v_2 sırasıyla payın ve paydanın serbestlik derecesi olarak adlandırılmaktadır.

F Dağılımının Beklenen Değeri

$$E(F) = \frac{v_2}{v_2 - 2} \quad v_2 > 2$$

F-Dağılımının Varyansı

$$Var(F) = \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)} \quad v_2 > 4$$

Teorem: S_1^2 ve S_2^2 varyansları σ_1^2 ve σ_2^2 olan iki normal ana kütleden çekilen n_1 ve n_2 birimlik iki bağımsız örneğin varyansı ise $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$ rastlantı değişkeni $n_1 - 1$ ve $n_2 - 1$ serbestlik dereceli bir F dağılımına uymaktadır.

Örnek: Varyanslarının birbirine eşit olduğu bilinen ve normal dağılan iki ana kütleden $n_1 = 33$ ve $n_2 = 27$ birimlik iki örnek çekiliyor. S_1^2 ve S_2^2 bu örneklerin varyansları olduklarına göre $P(\frac{S_1^2}{S_2^2} \leq 0,7)$ olasılığını bulunuz.

Çözüm: $\dfrac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} \sim F_{n_1-1,n_2-1}$ ve $\sigma_1^2=\sigma_2^2$ olarak verildiğine göre $\dfrac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$ olur. $P(\frac{S_1^2}{S_2^2} \le 0,7) = P\left(F_{32,26} \le 0,7\right)$ Ms-Excel'deki FDAĞ fonksiyonu yardımı ile $P\left(F_{32,26} \le 0,7\right) \cong 0,17$ bulunmuş olur.