HİDROLİK ÇALIŞMALARDA İSTATİSTİKSEL YÖNTEMLERİN KULLANIMI

Giriş

İstatistiksel Maddelerin Önemi ve Sınıflandırılması

Hidrolojik büyüklüklerin birçoğu fizik yasalarıyla tam olarak açıklanamayan rastgele değişken niteliği taşırlar. Bunların en önemli nedeni yağışın rastgele karakteridir; bu nedenle yağışla ilişkili olan akım değişkenlerinde de rastgelelik görülür. Hidrolik sistemin rastgele karakteri; hidrolik verilerdeki örnekleme hataları ve hidrolik süreç için kabul edilen modeldeki hatalar hidrolik değişkenlerin rastgele nitelik taşımasına neden olur. Bir hidrolik büyüklüğüm rastgele değişkenliği önemli değilse bu yanı ihmal edilip ortalama değeri çalışılarak olay deterministik bir yaklaşımla incelenebilir. Ancak bazı büyüklükler için (taşkın debisi gibi) böyle bir yaklaşım anlamlı olmaz, bu durumda olasılık teorisi ve istatistik birimlerine dayanan, olasılıkların işin içine girdiği modeller kullanmak gerekir.

Hidrolik istatistik modellerin kullanıldığı yerler şu şekilde sınıflandırılabilir.

1) Hidrolik verilerin istatistik analizi: Frekans analizi, parametrelerin analizi ve güven aralıklarının belirlenmesi, dağılım fonksiyonunun belirlenmesi.

2) Taşkın debisi dağılım modülleri: Hidrolik tasarımda özel bir önem taşıyan taşkın debileri için uygun dağılım fonksiyonlarının belirlenmesi ve bunlarla proje dönüş aralığına karşı gelen taşkın debisinin tahmini için geliştirilen modeller.

3) Korelasyon ve regresyon modelleri: İki yada daha fazla hidrolik değişken arasında istatistik ilişkiyi belirleyen bir model yardımıyla değişkenlerden biri için eksik verilerin tamamlanması yada kısa verilerin uzatılması.

4) Hipotez test modelleri: Hidrolik değişkenlerin parametreleri ve dağılıln fonksiyonları için yapılan kabullerin uygunluğunun gözlemlerle karşılaştırılarak kontrolü.

5) Hidrolik süreç (zaman serisi) modelleri: Zaman içinde değişken bir hidrolik büyüklüğün (akımın) stokastik yapısının modellenmesi, kurulan maddelerin simülasyon ve akım tahminlerinde kullanılması.

1. HİDROLİK VERİLERİN İSTATİSTİK ANALİZİ

1.1) Frekans Dağılımı

Bir rastgele değişkenin toplumunun tümünü gözlemek mümkün olmadığından eldeki örneğin analiziyle elde edilen frekans dağılımına eşdeğer olduğu kabul edilir. Frekans dağılımını belirlemek amacıyla örneğin analizi, söz konusu rastgele değişkenin tipine göre aşağıdaki yöntemlerle yapılır.

a) Kesikli Değişkenlerin Frekans Analizi

Elde kesikli bir değişkene ait *N* elemanlı bir örnek bulunduğunu düşünelim.

Bu örnekte X=x olayı Mi defa görülüyorsa bu olayın frekansı $fi = \frac{Mi}{N}$

Eklenik frekans dağılımı ise; $Fi = \frac{\sum_{f=1}^{i} Mj}{N} = \sum_{f=1}^{i} fj$

b) Sürekli Değişkenlerin Frekans Analizi

1. Büyük örneklerin frekans analizi:

Rastgele değişkenin değişme bölgesi uygun genişlikte sınıf aralıklarına ayrılır. İ'inci sınıf aralığı içine düşen gözlemlerin sayısı n_i ise bu sınıf aralığının frekansı; $fi = \frac{ni}{N}$

2. Küçük örnekleri frekans analizi:

$$F(X_m) = \frac{M}{N+1}$$

1.2) Parametrelerin Tahmini

Parametrelerin tahmininde aranması gereken ilk şart tahminin tarafsız olmasıdır. Tarafsız tahmin birden fazla şekilde yapılabilir. Eldeki bir örnekte yapılacak tarafsızlığı ve etkinliği tahmin yöntemine bağlıdır. Tahmin yöntemlerinin başlıcaları şunlardır;

a) Grafik Yöntem

b) En Küçük Kareler Yöntemi

$$y = p(x;\alpha,\beta,......) \qquad S = \sum_{i=1}^{N}(yi - y)^2 = \sum_{i=1}^{N}\left[yi - p(xi;\alpha,\beta,.....)\right]^2$$

c) Momentler Yöntemi

Söz konusu olasılık dağılın fonksiyonunun α,β,..... parametreleri ile μ1, μ2,.... parametreleri arasındaki bağıntılar elde edilir.

$$\alpha = f1(\mu 1, \mu 2,.....) \quad , \quad \beta = f2(\mu 1, \mu 2,.....),.......... \qquad (1)$$

önce momentlerin değerleri eldeki örnekten tarafsız tahminleri veren ifadelerle hesaplandıktan sonra (1) denkleminden parametrelerin a,b,....... tahminleri çözülür. α,β,..... parametrelerinin a,b,...... tahminleri şu denklemleri çözerek bulunur;

$$\frac{dS}{da} = 0 \quad , \quad \frac{dS}{db} = 0 \quad ,................$$

d) Maksimum Olabilirlik Yöntemi:

$$L = \prod_{i=1}^{N} p)xi;\alpha,\beta,.........)$$

Bu yöntemle L yerine *l*nL i_n kullanmak uygundur.

$$\ell nL = \ell n\left[\prod_{i=1}^{N} p(xi;\alpha,\beta,......)\right] = \sum_{i=1}^{N}\ell np(xi;\alpha,\beta,......)$$

a, b,............. etkin tahminleri şu denklem takımı çözülerek bulunur.

$$\frac{d\ell nL}{da} = 0 \quad , \quad \frac{d\ell nL}{db} = 0 \qquad (2)$$

Bu yöntem her durumda etkin tahminler verildiğinden diğer yöntemlerden daha üstündür. Ancak (2) denklemlerinin çözümünü elde etmek birçok hallerde kolay olmaz. Nonlineer denklem takımını ardışık yaklaşımlarla çözmek gerekir. Bundan dolayı maksimum olabilirlik yönteminin kullanılması nispeten azdır.

1.3) Güven Aralığı ve Güven Düzeyi:

Bir rastgele değişkenin toplumunun herhangi bir β parametresinin büyüklüğü N olan örneklerden hesaplanan b istatistiklerinin örnekleme dağılımının bilindiğini düşünelim. Eldeki bir örnekten belirlenen b istatistik değerinin iki yanına öyle bir (b_1, b_2) aralığı göz önüne alalım ki istatistiğin bu aralık içinde kalma olasılığı

Pc olsun. Buna göre toplumun bilinmeyen β parametre değerinin Pc olasılığı ile (b_1, b_2) aralığında kalacağı söylenebilir. Burada Pc'ye güven düzeyi, (b_1, b_2) aralığına bu düzeydeki güven aralığı denir. Burada dikkat edilecek bir nokta güven aralığının b_1 ve b_2 sınırlarının örnekten örneğe değişeceğidir. Zira her bir örnek için $E(\beta)=b$ kabulü yapılmaktadır.

1.4) Dağılım Fonksiyonunun Belirlenmesi

Kullanılabilecek olasık dağılım fonksiyonlarının sayısı pek çok olup kullanılacak fonksiyon tecrübelere dayanarak seçilir. Aşağıdaki şartları sağlayan herhangi bir F(x) fonksiyonu o.d.f olabilir.

- $0 \leq F(x) \leq 1$

- $x_1 < x_2$ için $F(x_1) \leq F(x_2)$

- $F(-\infty)=0$, $F(+\infty)=1$

Eldeki örneğin analiziyle belirlenen frekans dağılımına bir o.d.f'nin uydurulması şu adımlarla yapılır.

1. Denenecek o.d.f. seçilir. Bu seçimde tecrübelere dayanır.

2. Seçilen o.d.f.'nin parametreleri eldeki örnekten tahmin edilir. Her o.d.f.'nin belli sayıda parametresi vardır. Parametre sayısı optimum değerde olan o.d.f. kullanmak güvenilirlik sağlar.

3. bu o.d.f.'nin gözlenmiş frekans dağılımına uygunluk araştırılır.

4. uygunluğun istenilen düzeyde olmadığı görülürse başka bir o.d.f. seçilerek aynı adımlar tekrarlanır. Burada eldeki örneğin küçüklüğü dolayısıyla birçok hallerde gözlenen frekans dağılımına birden fazla o.d.f.'nin aynı derecede uyduğunun görülebileceğidir.

1.4.1.) Başlıca Olasılık Dağılım Fonksiyonları

a) Binom Dağılımı

Bir kesikli rastgele değişken için sadece 2 olay mevcut olduğunu düşünelim. Bu olayların olasılıkları p ve q=1-p ile gösterilsin. Bu değişkene ait birbirinden bağımsız n deneme yapılsın. Bu n elemanlı örnekte olasılığı p olan olayın x defa görülmesi olasılığının binon dağılıma uyduğu gösterilebilir. Bu dağılım kütle fonksiyonu:

$$p(x) = \binom{n}{x} p^x q^{n-x}$$

Bu ifadede $\binom{n}{x}$, n adet büyüklüğün x'li kombinezonlarının sayısı olup söyle hesaplanır.

$$\binom{n}{x} = \frac{n(n-1)........(n-x+1)}{1.2.........(n-1)x} = \frac{nI}{xI(n-x)I}$$

$$Ex = np \qquad , \quad Var_x = npq \quad , \quad Cs = \frac{q-p}{\sqrt{npq}} \qquad , \qquad k = 3 + \frac{1-6pq}{npq}$$

b) Poisson Dağılımı:

Rastgele değişken için 2 olay mevcut olsun. Ancak bunlardan birinin olasılığı çok küçük olsun. Buna karşın n deneme sayısı da çok büyük olsun. np çarpımının da sonlu olduğu kabul edilsin (np=∧) bu halde n denemede olasılığı p olan olayın x defa görülmesi olasılığı Poisson dağılımına uyar.

$$P(x) = \frac{\wedge^x e^{-x}}{xI}$$

Bu dağılımın özellikleri:

$$Ex = \wedge, \qquad Var_x = \wedge, \qquad Cs = \frac{1}{\sqrt{\wedge}}, \qquad k = 3 + \frac{1}{\wedge}$$

c) **Geometrik Dağılım**:

$$P(x) = q^{x-2}.p \qquad F(x) = 1 - q^x$$

Dağılımın Özellikleri;

$$Ex = \frac{1}{p}, \qquad Var_x = \frac{q}{p^2}$$

d) Normal Dağılım

Gauss dağılımı olarak da bilinen dağılımın olasılık yoğunluk fonksiyonu;

(i) $p(x) = \frac{1}{Q\sqrt{2\pi}} \exp\left[-(x-\mu)^2 / 2Q^2\right]$ şeklindedir.

Kısaca N(μθ) şekline gösterilen bu dağılımın iki parametresi vardır. Bunlardan μ rastgele değişkenin ortalaması, θ standart sapmasıdır. Normal dağılım simetrik olupçarpıklık katsayısı Cs=0 k=3'dür.

(*i*) denklemiyle verilen dağılımın a.g.f.'nu tablolaştırmak için standart normal değişken kullanılır.

$Z = \dfrac{x - \mu}{Q}$ şeklinde tanımlanan standart değişken boyutsuz olup ortalaması 0, standart sapması 1'dir. Z'nin dağılımı olan N(0,1) dağılımının a.g.f. ve o.d.f. tablolaştırılmıştır. Normal dağılım simetrik olduğundan bu tablolar sadece Z'nin pozitif değerleri için hazırlanmıştır.

Normal dağılımın parametreleri eldeki örnekten grafik yöntemle yada momentler yöntemiyle tahmin edilir. Normal dağılım simetrik olduğu için momentler yöntemi etkin tahminler verir. Normal dağılım istatistik uygulamalarda en çok karşılaşılan dağılımdır.

e) Lognormal Dağılım

Rastgele değişkene; y=*l*nx şeklinde logaritmik bir dönüşüm uygulandığında dönüştürülmüş y değişkeninin dağılımı normal ise x değişkeninin dağılımı lognormaldir.

$$p(x)P(y)\left|\frac{dy}{dx}\right| = \frac{1}{Qy\sqrt{2\pi}}.\exp\left[-(\ln x - \mu y)^2 / 2Qy^2\right]\frac{1}{x}$$

Yukarıdaki denklemde μy ve Qy, y değişkeninin ortalama ve standart sapması olup x'in parametreleri olan μx ve Qx'e şu şeklinde bağlıdır.

$$\mu x = \exp\left(\mu y + \frac{Qy^2}{2}\right) \qquad Qx = \mu x(e^{Qy2} - 1)^{1/2}$$

$$Cz = e^{Qy2} - 1)^{1/2}$$

$$Cs = Cv^3 + 3Cv$$

$$Ek = Cv^8 + 6Cz^6 + 15Cv^4 + 16Cv^2$$

Lognormal dağılımda rastgele değişken sadece pozitif değerler alabildiği ve dağılımın pozitif bir çarpıklığı bulunduğundan bu dağılım bir çok hidrolojik değişkenlere iyi uyar. Yıllık akışkanların dağılımı için lognormal dağılım çok kullanılır.

f) Gamma Dağılımları

Bir çok parametreli gamma dağılımının o.g.f.;

$$P(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}$$

$x \geq 0$ şeklindedir. Burada $\Gamma(\alpha)$ tablolaştırılmış gamma fonksiyonu olup $\alpha > 0$ için tanımlanır.

Dağılımın özellikleri;

$$Ex = \alpha \qquad Var_x = \alpha \qquad Cs = \frac{2}{\sqrt{\alpha}} \qquad Ek = \frac{6}{\alpha}$$

Yukarıdaki formülde x yerine $\beta > 0$ olmak üzere x/β konulursa 2 parametreli gamma dağılımının o.y.f. elde edilir:

$$P(x) = \frac{1}{\beta^{\alpha})} x^{\alpha-1} e^{-x/\beta} \qquad x \geq 0$$

Bu dağılımın özellikleri

$$Ex = \alpha.\beta \qquad Var_x = \alpha.\beta^2 \qquad Cs = \frac{2}{\sqrt{\alpha}} \qquad Ek = \frac{6}{\alpha}$$

Yine ilk denklemde x yerine $(x\text{-}x_0)/\beta$ konulursa 3 parametreli gamma dağılımına geçilir. (pearson Tip III dağılımı):

$$P(x) = \frac{1}{\beta^{\alpha})} (x - x_0)^{\alpha-1} e^{-(x-xo)/\beta} \qquad x \geq x_0$$

Bu dağılımın özellikleri

$$Ex = x_0 + \alpha.\beta \qquad Var_x = \alpha.\beta^2 \qquad Cs = \frac{2}{\sqrt{\alpha}} \qquad Ek = \frac{6}{\alpha}$$

Bir parametreli gamma dağılımının α parametresi şu şekilde bulunur.

$$L = \underset{i-1}{\overset{N}{\pi}} p(xi;\alpha) = \underset{i=1}{\overset{N}{\pi}} \frac{1}{\Gamma(\alpha)} xi^{\alpha-1} e^{-xi} = [\Gamma(\alpha)]^{-N} \underset{i=1}{\overset{N}{\pi}} xi^{\alpha-1} e^{-xi}$$

$$\frac{d(\ln L}{d\alpha} = -N\frac{d[\ln\Gamma(\alpha)]}{d\alpha} + \sum_{i=1}^{N}\ln xi = 0$$

g) Pearson Dağılımları:

Bir örnekten elde edilen frekans dağılımına en iyi uyan o.d.f.'yi seçmek için Pearson, örnekten hesaplanan Cs ve Ek katsayılarına bağlı olan Kp parametresinin kullanılmasını önermiştir.

$$Kp = \frac{Cs^2(Ek+6)^2}{4(2Ek-3Cs^2)(4Ek-3Cs^2+12)}$$

Kp<0 halinde Pearson Tip I, Kp=0 için normal, 0<Kp<1 için Pearson Tip IV, Kp>1 için Pearson Tip VI, Kp→∞ için Pearson Tip III dağılımları kullanılmaktadır.

h) Ektrem Değer Dağılımları:

i) İki Değişkenli Normal Dağılım:

Çok değişkenli dağılımlar arasında sadece normal dağılımın analitik ifadesi mevcuttur. 2 değişkenli normal dağılımın ortak o.y.f.:

$$P8x,y) = \frac{\exp[-Q/2(1-g^2)]}{2\pi QxQy\sqrt{1-g^2}}$$ şeklindedir. Burada;

$$Q = \left(\frac{x-\mu x}{Qx}\right)^2 + \left(\frac{y-\mu y}{Qy}\right)^2 - 2g\frac{(x-\mu x)(y-\mu y)}{Qx.Qy}$$

μx, μy, Qx, Qy sırasıyla x ve y değişkenlerinin marjinal dağılımlarının ortalama ve standart sapmaları, *g* ise x ile y arasındaki korelasyon katsayısıdır. x'in y'ye göre şartlı dağılımının o.y.f. normal olup:

$$Px(xIy) = \frac{\exp[-Q/2(1-g^2)+(y-\mu y)^2/2Qy^2]}{Qx\sqrt{2\pi(1-g^2)}}$$ şeklindedir.

Bu şartlı dağılımın parametreleri:

$$\mu xIy = \mu x + g\frac{Qx}{Qy}(y-\mu y)$$

$$QxIy = Qx\overline{)1-g^2}$$

Normal dağılmış iki değişkenin ortak dağılımları her zaman 2 değişkenli normal dağılım değildir, fakat uygulamalarda analitik ifadelerin basitliği nedeniyle çoğu zaman bu kabulün yapıldığı görülür.

Değişkenlere logaritmik dönüşüm uygulayarak 2 değişkenli lognormal dağılım da benzer şekilde tanımlanabilir. 2 değişkenli normal dağılım 2'den fazla değişkenli normal dağılımlar haline de genelleştirilebilir.

Örnek Problem 1.1)

Dicle nehrinin Cizre akım ölçüm istasyonunda 1956-1975 yıllarında ölçülen yıllık maksimum debiler aşağıda verilmiştir.

Bu debilerin frekans analizini yapınız.

YIL	1956	1957	1958	1959	1960
Qmax(m^3/s)	2324	6300	2340	2080	2262
YIL	1966	1967	1968	1969	1970
Qmax(m^3/s)	8820	4516	4866	6450	2250

Çözüm:

20 Elemanı olan bu küçük örneğin frekans analizinde önce elemanlar büyüklük sırasına göre dizilerek düzenlenirler. Düzenlenmiş örneğin m'inci elemanı x_m ise yıllık maksimum debi x_m'den küçük olması olasılığı şu denklemle hesaplanır. Ve hesaplar tabloda gösterilmiştir.

$$F(x_m) = \frac{M}{N+1}$$

m	1	2	3	4	5

x_m	963	1250	2080	2250	2262
$F(x_m)=\dfrac{M}{N+1}$	0,048	0,095	0,143	0,190	0,238
M	11	12	13	14	15
x_m	3450	4350	4516	4866	5300
$F(x_m)=\dfrac{M}{N+1}$	0,524	0,571	0,619	0,667	0,714

2) TAŞKIN DEBİSİ DAĞILIM MODELLERİ

Bir rastgele değişkene ait N=n-m elemanı olan bir örneğin her birinde m elemanı bulunan n adet alt örneklere ayrılsa, bu alt örneklerden her birindeki en büyük (yada en küçük) elemanı (ekstrem değeri) göz önüne alalım . Bu ekstrem değerlerinin dağılımının incelenmesi hidrolojide özel bir önem taşır. Zira alt örnekler birer yıllık olarak düşünülüp her birindeki günlük en büyük akış (taşkın debisi) yada en küçük günlük akış göz önüne alınırsa taşkınların yada kuraklıkların dağılımı ile ilgili istatistik bilgiler elde edilmiş olur.

Hidrolojide gözlenmiş akış serileri genellikle n=30-100 yıl uzunluğunda olduklarından bunlardan ekstrem değerlere ait elde edilecek örneklerdeki eleman sayısı da ancak 30-100 kadar olur. Bu sayı bazı hallerde daha azdır. Bu örneklerin frekans analiziyle elde edilen frekans dağılım çizgisini uzatarak gözlem süresinden daha uzun aralıklarla görülecek taşkın debilerini tahmin etmek büyük hatalara yol açabilir. Bu durumda ekstrem değerlerin örnekten elde edilen frekans dağılımına uyan o.d.f.'nin amprik olarak belirlenmesi yerine teorik ekstrem değer dağılımlarını kullanmak uygundur. Bu dağılımlar şöyledir;

2.1) Gumbel Dağılımı

Taşkın debileri için en çok kullanılan teorik dağılımdır. Fisher-Tippet Tip I dağılımı olarak da bilinir. Dağılımın o.y.f. ve o.d.f. şu ifadelerle verilir.

$$p(x) = \alpha.\exp[-y - \exp(-y)]$$

$$F(x) = \exp[-\exp(-y)] \qquad y = \alpha - (x - \beta)$$

α; ölçek parametresi β; yer parametresi

Gumbel dağılımının özellikleri;

$$\mu x = \beta + \frac{0,5772}{\alpha}, \qquad Qx = \frac{1,2825}{\alpha}, \qquad \text{Cs}=1,14, \qquad \text{k}=4,5$$

$$M = \beta + \frac{0,3665}{\alpha}, \qquad M = \beta\,, \qquad F(m) = e^{-1} = 0,368$$

$$\alpha = \frac{1,2825}{Qx} \qquad \beta = \mu_x - 0,450Qx$$

$$\alpha(x-\beta) =^{-} \ln\left[-\ln F(x)\right] \qquad \frac{1}{\alpha} = 0,636(x_{0,75} - x_{0,25})$$

2.2) Fisher Tippet Tip III Dağılımı (Weibull Dağılımı)

Dağılımın o.y.f. ve o.d.f. şu şekilde verilmiştir.

$$P(x) = \frac{\alpha}{x - x_0}\, y \exp(-y) \qquad x \geq x_0$$

$$F(x) = 2 - \exp(-y) \qquad x \geq x_0$$

$$y = \left(\frac{x - x_0}{\beta - x_0}\right)^{\alpha} \qquad \alpha > 0, \beta > x_0, x \geq x_0$$

α; ölçek parametresi, β; yer parametresi, x_0; alt sınır

Dağılım özellikleri;

$$\mu_x = x_0 + (\beta - x_0)\Gamma\left(1 + \frac{1}{\alpha}\right)$$

$$Qx^2 = (\beta - x_0)^2\left[\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right]$$

Bu dağılım hidrolojide en küçük debiler (kuraklıklar) için kullanılır. x_0 Çoğunlukla 0 olarak alınır.

2.3) Frekans Faktörü ve Dönüş Aralığı

Taşkın debisinin açılması olasılığı; $F_1(x)$

$$F_1(x) = 1 - F(x) \qquad\qquad T = \frac{1}{F_1(x)}$$

Taşkın debilerinin ortalaması; μ standart sapması Q ile gösterilirse T yıllık taşkın debisi için şöyle bir genel ifade yazılabilir.

$$x = \mu + Q.K$$

Bu ifadedeki K frekans faktörü taşkın debisinin olasılık dağılımının ve T dönüş aralığının fonksiyonudur. Çeşitli o.d.f.'ları için K'yı veren ifadeler elde edilebilir.

2.4) Kısmi Süreklilik Serilerinin Frekans Analizi

Çözülen debilerin arasından sadece her yılın en büyük değerinin değil, belli bir değerin üstünde kalanların hepsi göz önüne alınırsa bu taban değeri o şekilde seçilmelidir ki her yıl en az bir günün debisi taban değerinin üzerinde kalsın. Bu şekilde belirlenen seriye kısmi süreklilik serisi denir.

$$T_E = \frac{N+1}{m}$$

Burada Te, kısmi süreklilik serisinde m'inci büyük elemanın dönüş aralığı olup N, yıl sayısıdır.

$$1 - P = \left(1 - \frac{Pe}{m}\right)^m$$

Pelm; Bir yılda taşkınlardan birinin x değerini aşması olasılığı P; yıllık maksimum taşkın debileri için x'den büyük olma olasılığı

$$T_E \cong \frac{1}{\ln T - \ln(T-1)}$$

2.5) Proje Periyodu ve Risk

Proje hesaplarında göz önüne alınan T yıllık taşkın debisinin proje periyodu olan n yıllık bir süre içinde Pn ile gösterilen bir aşılma olasılığı vardır. Pn, kabul edilebilecek risk'i ifade etmekte olup ekonomik düşüncelerle belirlenecek olan bir proje kriteridir. Dönüş aralığı T yıl olan taşkın debisinin herhangi bir yılda aşılması olasılığı 1-1/T ise

$$Pn = 1 - \left(1 - \frac{1}{T}\right)^{n} \cong 1 - \exp\left(-\frac{n}{T}\right)$$

Örnek Problem 2:

Yıkıldığı takdirde can kaybına açmayacak bir barajın dolu savası 500 yıllık taşkın debisine göre projelendirilmiştir. Bu barajın 50 yıllık proje periyodu içinde daha büyük bir taşkın gelmesi riskini hesaplayınız.

Çözüm:

n=50 yıl, T=500 yıl için risk, $Pn = 1 - \left(1 - \frac{1}{T}\right)^{n}$ denklemine göre

$$Pn = 1 - \left(\frac{1}{500}\right)^{50} = 0{,}095 \text{ bulunur.}$$

3.) KORELASYON VE REGRESYON MODELLERİ:

İstatistikte rastgele değişkenler arasındaki bir bağıntıyı ifade eden matematik ifadeye regresyon denklemi denir. Bir rastgele değişkenin değerini bir veya daha fazla sayıda rastgele değişkenlerin değerlerine bağlı olarak en iyi şekilde tahmin etmeye yarayan regresyon denklemlerinin belirlenmesine regresyon analizi denir. Regresyon analizi şu adımlar halinde yapılır.

1. Önce incelenecek bağıntıda göz önüne alınacak değişkenler belirlenir. Buna göre ağıntı iki değişkenli yada çok değişkenli yada çok değişkenli olabilir.

2. Göz önüne alınan değişkenler arasındaki ilişkiyi gösteren regresyon denkleminin tipi seçilir. En basit olarak doğrusal regresyon yapılabilir.

3. Seçilen değişkenler arasındaki bağımlılığın derecesini belirlemek amacıyla korelasyon analizi yapılır. Bağımlılığı ölçmek için korelasyon katsayısı denen parametrelerin değerleri hesaplanır.

4. Bu parametrelerin hesaplanan değerlerinin anlamlı bir bağımlılık ifade edip etmediği istatistik testlerle kontrol edilir.

5. Kontrol sonucunda göz önüne alınan değişkenler arasında anlamlı bir bağımlılık bulunduğu sonucuna varılırsa önceden tipi seçilmiş olan regresyon denkleminin parametreleri hesaplanır.

6. Bu şekilde regresyon denklemi belirlendikten sonra bu denklemi kullanarak yapılacak tahminlerin güven aralıkları belirlenir.

3.1) İki Değişkenli Doğrusal Korelasyon ve Regresyon

3.1.1) Regresyon Çizgisi

X ve Y rastgele değişkenlerin birbirine karşı gelen x_i, y_i değerlerini x-y düzleminde noktalansın. X ile Y arasında fonksiyonel bir bağıntı söz konusu değilse $y=x_i$ değerine karşı Y değişkeni çeşitli değerler alabilir. Bu değerlerin ortalaması olan $y_{0i}=E(Y/X=X_i)$ değeri hesaplansın. Bu şekilde belirlenen y_{oi} noktalarıyla elde edilen çizgiye y'nin x'e göre regresyon çizgisi denir. Aynı şekilde x'in y'ye göre regresyon çizgisi de çizilebilir. Bunlar genelde birbirinden farklı çizgilerdir. Ancak aradaki bağıntı fonksiyonelse bu iki çizgi çakışır.

3.1.2) Korelasyon Katsayısı:

İki rastgele değişken arasındaki doğrusal bağımlılığın derecesini ölçen parametre korelasyon katsayısıdır.

$p = \dfrac{CoV_{xy}}{Q_X Q_Y}$ Bu katsayı boyutsuz olup 0 ile 1 arasında değişir.

g toplum parametresinin eldeki örnekten tahmini şu denklemle yapılır.

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{N.Sx.Sy} = \frac{\sum_{i=1}^{N} x_i y_i - N.\bar{x}.\bar{y}}{N.Sx.Sy} \qquad (1)$$

r; korelasyon katsayısının eldeki N elemanlı örnekten hesaplanan istatistik değeri, x_i, y_i ; örnekdeki gözlem çiftleri,

$\bar{x}\, ve.\, \bar{y}$; sırasıyla x ve y'nin ortalamaları,

Sx, Sy; x ve y'nin standart sapmalarıdır.

(1) denklemiyle hesaplanan korelasyon katsayısının değişkenler arasında anlamlı bir bağımlılık ifade edip etmediğini anlamak için korelasyon katsayısına ait

istatistik hipotezlerin kontrolü bazı yaklaşık testlerle yapılır. Değişkenlerin ortak dağılımının normal olması hali için geçerli olan bu testler şöyledir;

1. *g* değerinin 0'a yakın olması halinde r'nin örnekleme dağılımı, asimptotik olarak, standart sapması:

$Sr = \frac{1-r^2}{\sqrt{N}}$ olan normal dağılımdır. Sr değeri hesaplandıktan sonra 0 değeri (r-3Sr, r+3Sr) aralığının dışında kalıyorsa *g*=0 hipotezi reddedilir. Daha kuvvetli bir test olarak 0 değerinin (r-4Sr, r+4Sr) aralığının dışında kalıp kalmayışı kontrol edilebilir.

2. Aşağıdaki şekilde tanımlanan Z değişkeninin örnekleme dağılımının yaklaşık olarak normal olduğu bilinmektedir.

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}$$

Z'nin örnekleme dağılımının parametreleri:

$$\mu_z \frac{1}{2}\ln\frac{1+g}{1-g}, \qquad Q_z = \frac{1}{\sqrt{N-3}}$$

Buna göre toplumun korelasyon katsayısının herhangi bir *g* değerine eşit olduğu hipotezi yapılarak bu hipotez kontrol edilebilir.

3. Korelasyon katsayısı için en kuvvetli test aşağıdaki şekilde tanımlanan t değişkeni ile yapılır.

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

Toplumun korelasyon katsayısının *g*=0 olması halinde t'nin örnekleme dağılımı s.d.=N-2 olan standart t dağılımıdır. Buna göre *g*=0 hipotezi kolayca kontrol edilebilir. Çeşitli α anlamlılık düzeylerinde r'nin güven aralığının sınırları için Tablolar oluşturulmuştur. Buna göre seçilen α düzeyinde eldeki örnekte hesaplanan r'nin mutlak değeri tablodaki değerden küçükse hipotez kabul, aksi halde reddedilir.

3.1.3) Regresyon Doğrusu

Korelasyon katsayısı için hesaplanan r değeri rastgele değişkenler arasında anlamlı bir doğrusal bağımlılık bulunduğunu gösteriyorsa x'in verilen bir değeri için y'nin en iyi tahminini veren y'nin x'e göre regresyon doğrusunun denklemi olan:

$$y = a + bx$$

ifadesindeki a ve b regresyon katsayıları Ai gözlenmiş noktalarının regresyon doğrusuna düşey uzaklıkların karelerinin toplamını minimum yapacak şekilde belirlenir. Minimum yapılacak ifade:

$$\sum_{i=1}^{N} e^2 y_i = \sum_{i=1}^{N} (y_i - a - bx_i)^2$$ ve

$$\frac{d\sum_{i=1}^{N} e^2 y_i}{da} = 0, \frac{d\sum_{i=1}^{N} e^2 y_i}{db} = 0$$ denklemlerinin çözümüyle regresyon katsayıları için şu ifadeler bulunur.

$$b = \frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} = \frac{Sy}{Sx} r, a = \bar{y} - b\bar{x}$$

3.2.4) Regresyonda Güven Aralığı

Verilen denklemlerde hesaplanan a ve b regresyon katsayıları toplumun α ve β parametrelerinin tahminleridir. Bu tahminler, rastgele değişkenlerin ortak dağılımı normal ise tarafsız tahminlerdir. a ve b'nin örnekleme dağılımları doğrudan doğruya bilinmemekle beraber normal dağılmış değişkenler halinde + değişkenlerinin s.d = N-2 olan Studentt dağılımına uydukları bilinmektedir.

$$ta = \frac{r(a - \alpha)}{b} \sqrt{\frac{N-2}{(1 - r^2)\left(Sx^2 + \bar{x}^2\right)}}$$

$$tb = \frac{r - (b - \beta)}{b} \sqrt{\frac{N-2}{1 - r^2}}$$

Büyük örnekler halinde b'nin dağılımı N(β,Sey^2/N.Sx^2) normal dağılımına, α'nın dağılımı da n(α,Sey^2(2+x^2/Sx^2)/N) normal dağılımına yaklaşır. Bu dağılımları kullanarak regresyon katsayılarının güven aralıkları belirlenebilir, a'nın veya b'nin 0 olup olmadıklarına ait hipotezler kontrol edilebilir.

Öte yandan verilen bir x_i değeri regresyon denklemiyle tahmin edilen y değerinin dağılımı normal olup varyansı:

$$S^2 = \frac{S^2 ey}{N}\left(1 + \frac{\left(x_i - \bar{x}\right)^2}{Sx^2}\right)$$ şeklinde verilir.

S'e y'nin tahminindeki standart hata denir. Yukarıdaki denklemde görüleceği gibi bu hata Sey ile birlikte azalır, N azaldıkça artar, x_i-x'nın artışıyla da artar. Buna göre y'nin tahminindeki standart hata x_i-x değerindeki en küçük tür, x_i değeri x'den uzaklaştıkça tahminlerin güvenirliği azalır.

3.2) Dogrusal Olmayan Korelasyon ve Regresyon

X ve Y rastgele değişkenleri arasındaki regresyon çizgisi için herhangi bir: $y = f(x)$ ifadesi seçilebilir. Bu denklemdeki parametrelerin değerleri $\sum_{i=1}^{N} ey_i^2 = \sum_{i=1}^{N}\left[y_i - f(x_i)\right]^2$ 'yi minimum yapacak şekilde belirlenir. Elde edilen denklemler genellikle lineer olmadıklarından ardaşık yaklaşımlarla çözülmelidir.

Seçilen regresyon fonksiyonunu uygun bir dönüşümle doğrusal regresyona çevirmek mümkünse hesaplarda büyük bir kolaylık sağlanmış olur. Örneğin; $y = c_1.c^2$ ifadesi logaritmik dönüşümle doğrusal regresyona dönüştürülebilir;

$$\log y = \log c_1 + c_2 \log x, \qquad v = a + b.u$$

Doğrusal olmayan regresyondaki bağımlılığın derecesi korelasyon indisleriyle ölçülür. y'nin x'e göre doğrusal olmayan bağımlılığı için;

$$ry = \left(1 - \frac{S^2 ey}{Sy^2}\right)^{1/2}$$ korelasyon indisi kullanılır.

$$S^2ey = \frac{\sum_{i=1}^{N} e^2 yi}{N} = \frac{\sum_{i=1}^{N} [yi - f(xi)]^2}{V} ve..ry \neq rx$$

3.3) Çok Değişkenli Korelasyon ve Regresyon

Y rastgele değişkenin iki yada daha fazla sayıda x_1, x_2,.......... x_m, rastgele değişkenlerine göre çok değişkenli regresyonu genel olarak:

Y=f(x_1, x_2,............ x_m,) ifadesi ile verilir.

Çok değişkenli regresyonun en basit ve en çok kullanılan şekli doğrusal regresyonudur.

y= b_0,+ b_1 x_1 + b_2 x_2 +................+ b_m x_m

Bu ifadedeki m değişken sayısı örnekteki N eleman sayısına göre yeter derecede küçük seçilmelidir. Pratikte m değeri $\sum_{i=1}^{N} eyi^2$ 'i minimum yapacak şekilde grafik olarak veya deneme yoluyla belirlenir.

3 değişkenli doğrusal regresyon çizgisinin denklemi:

y= b_0,+ b_1 x_1 + b_2 x_2 şeklindedir. Bu denklemdeki regresyon katsayıları $\sum_{i=1}^{N} eyi^2 = \sum_{i=1}^{N} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i})^2$ 'yi minimum yapacak şekilde belirlendiğinde şu denklemlere varılır.

$$b_1 \sum_{i=1}^{N} (x_{1i} \bar{x}_1)^2 + b_1 \sum_{i=1}^{N} (x_{1i} \bar{x}_1)(x_{2i} \bar{x}_2) = \sum_{i=1}^{N} (y_{1i} \bar{y})(x_{1i} \bar{x}_1)$$

$$b_1 \sum_{i=1}^{N} (x_{1i} \bar{x}_1)(x_{2i} \bar{x}_2) + b_2 \sum_{i=1}^{N} (x_{2i} \bar{x}_2) = \sum_{i=1}^{N} (y_{1i} \bar{y})(x_{2i} \bar{x}_2)$$

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2$$

b_0, b_1, b_2 katsayıları yukarıdaki denklem takımı çözülerek bulunur. 3'den fazla değişkenli doğrusal regresyonlarda da benzer denklemlere varılır.

Çok değişkenli doğrusal regresyonlarda değişkenler arasındaki bağımlılığın derecesini ölçmek için çeşitli parametreler kullanılabilir.

1. İki değişkenli doğrusal regresyondaki korelasyon katsayısına benzer şekilde çok değişkenli korelasyon katsayısı tanımlanabilir.

$$R = \left(1 - \frac{S^2 ey}{Sy^2}\right)^{1/2}$$

Burada Sey^2 gözlenen y_i değeriyle regresyon denkleminden hesaplanan değerler arasındaki farkların varyansıdır:

$$S^2 ey = \frac{\sum_{i=1}^{N} e^2 y_i}{N - (m+1)}$$ denklemiyle tanımlanan R korelasyon katsayısı gözlenen y_i değerleriyle regresyon denkleminden hesaplanan değerler arasındaki bağımlılığı ifade eden boyutsuz bir sayıdır. R'nin değeri 1'e yaklaştıkça regresyon denkleminden hesaplanan değerlerin güven aralığı küçülür.

2. Y bağımlı değişkeninin X_i bağımsız değişkenlerinden biriyle bağımlılığı, kısmi korelasyon katsayıları ile ölçülür. r_i kısmi korelasyon katsayısını hesaplamak için önce Y'nin bütün X değişkenleriyle R_{-i} korelasyon katsayısı hesaplanır. r_i şu formülle bulunur.

$$(1)\ r_i = \left(1 - \frac{1 - R^2}{1 - R_{-i}^2}\right)^{1/2}$$

Ancak (1) denklemiyle verilen r_i'lerin hesabı vakit alıcıdır. Bunların yerine hesabı daha basit olan beta katsayıları kullanılabilir.

$$\beta_i = b_i \frac{Si}{Sy}$$

Bu katsayılar kısmi korelasyon katsayılarına yakın değerler verirler. Çok değişkenli doğrusal regresyonda b_i regresyon katsayılarının örnekleme dağılımlarının standart sapması:

$$Sbi = \frac{Sey}{Si} \cdot \sqrt{\frac{1}{(N - m - 1)(1 - R_i)^2}}$$ ifadesiyle verilir.

Burada R_i, X_i değişkeninin diğer bütün X değişkenleriyle çok değişkenli korelasyon katsayısıdır. Eğer X değişkenleri aralarında bağımsız iseler R_i=0 olur. R_i

büyüdükçe S_{bi}'nin büyüdüğüne, yani hesaplanan b_i katsayısının değerine daha az güvenilebileceğine dikkat edilmelidir.

Örnek Problem 3):

Dicle nehrindeki Rezuk akım ölçme istasyonu ile Sinan istasyonunda ölçülen yıllık akış hacimleri aşağıda verilmiştir.

Yıl	1956	1957	1958	1959	1960
R($10^6 m^3$)	16077	14867	11720	9352	10537
S($10^6 m^3$)	4629	4556	2507	1612	2125

Yıl	1964	1965	1966	1967	1968
R($10^6 m^3$)	16439	13019	17729	20368	26748
S($10^6 m^3$)	-	4041	5191	5328	6543

a) İki istasyondaki akışlar arasındaki korelasyon katsayısını hesaplayınız.

b) Hesaplanan korelasyon katsayısının sıfırdan anlamlı derecede farklı olup olmadığını çeşitli testlerle kontrol ediniz.

c) Hesaplanan korelasyon katsayısı için 0,05 anlamlık düzeyindeki güven aralığını belirleyiniz.

d) Sinan'daki akışların Rezuk'daki akışlara göre regresyon doğrusunun denklemini elde ediniz. Bu denklem Sinan'daki akışların varyansının yüzde kaçını izah eder?

e) 1964 yılında ölçülmemiş olan Sinan istasyonundaki yıllık akış hacmini tahmin ediniz. Bu değer için % 90 güven düzeyindeki güven aralığını belirleyiniz.

Çözüm:

a) Bu iki değişken arasındaki doğrusal korelasyon katsayısının tahmini:

$$r = \frac{\sum x_i y_i - N.\bar{x}.\bar{y}}{N.Sx.Sy}$$

Rezuk'daki akışlar için;

$\bar{x} = 16237x10^6 \qquad Sx = 7669x10^6$

Sinan'daki akışlar için;

$\bar{y} = 4397x10^6 \qquad Sy = 2768x10^6$

Rezuk'da ve Sinan'da aynı yılda ölçülen akışların çarpımlar toplamı;

$\sum x_i y_i = 1347x10^{18}$

Yukarıdaki denklemde yerlerine koyarak :

$$r = \frac{1347x10^{18} - 15x16233x10^6 x4397x10^6}{15x7669x10^6 x2768x10^6} = 0{,}868$$

b) Hesaplanan r değerinin anlamlı olup olmadığı çeşitli testlerle kontrol edilebilir. I-r'nin örnekleme dağılımının standart sapması,

$$Sr = \frac{1-r^2}{\sqrt{N}} = \frac{1-0{,}868^2}{\sqrt{15}} = 0{,}0637$$

$r - 4Sr = 0{,}868 - 4x0{,}0637 = 0{,}613 > 0$ ve

$r + 4Sr = 0{,}868 - 4x0{,}0637 = 1{,}123 > 0$ ise korelasyon katsayısı sıfırdan anlamlı derecede farklıdır.

II Hesaplanan r=0,868 değerine karşı gelen Z değeri ilgili tabloda Z=1,325 olarak okunur.

$$\mu_Z = 0, \qquad Q_Z = \frac{1}{\sqrt{15-3} = 0{,}289}, \qquad \frac{Z}{Q_Z} = \frac{1{,}325}{0{,}289} = 4{,}585$$

Tablodan normal dağılmış standart değişkenin 4,585 değerini aşması olasılığı ~ 0 olarak okunur. Buna göre toplumun korelasyon katsayısının 0 olduğu hipotezi büyük bir güvenirlikle reddedilir.

III. $t = \dfrac{r\sqrt{N-2}}{\sqrt{1-r^2}} = \dfrac{0{,}868\sqrt{15-2}}{\sqrt{1-0{,}868^2}} = 6{,}30$

Serbestlik derecesi N-2=13 için ilgili tablodan 0,01 anlamlılık düzeyinde r'nin kritik değeri 0,641 olarak okunur. Ya da t=6,3 için t dağılımı tablosunda aşılma olasılığının 0,005'den küçük olduğu görülür.

Böylece bütün testler toplumun korelasyon katsayısının sıfırdan anlamlı derecede farklı olduğu sonucunu verirler. Buna göre Rezuk akışları ile Sinan akışları arasında anlamlı bir ilişki olduğu kabul edilebilir.

c) g=0,868 için

μ_z=1,325 Q_z=0,289

0,05 düzeyindeki güven aralığının sınırları:

$$Z_1 = 1{,}325 - 1{,}96x0{,}289 = 0{,}76$$

$$Z_2 = 1{,}325 + 1{,}96x0{,}289 = 1{,}89$$

Tabloda bunlara karşı gelen r değerleri:

$$(0{,}64 - 0{,}95)$$

d) $b = \dfrac{Sy}{Sx}.r = \dfrac{2768x10^6}{7669x10^6}.0{,}868 = 0{,}313$

$$a = \bar{y} - b\bar{x} = 4397x10^6 - 0{,}313x16233x10^6 = -684x10^6$$

Buna göre Sinan akışlarının (y) Rezuk akışlarına (x) göre regresyon doğrusu: $y = -684x10^6 + 0{,}313x$

Bu denklem Sinan'daki akışların varyansının $0{,}868^2$= %75'ini izah eder.

e) 1964 yılında ölçülmemiş olan Sinan'daki akışın tahmini:

$$y = -648x10^6 + 0{,}313x16439x10^6 = 4461x10^6 m^3$$

Bu tahminlerin varyansı ise;

$$S^2 = \frac{(1-r^2)Sy}{N}\left(1 + \frac{(x-\bar{x})^2}{Sx^2}\right)$$

$$= \frac{(1-0{,}868^2)(2768x10^6)^2}{15}\left(1 + \frac{(16439x10^6 - 16233x10^6)^2}{(7669x10^6)^2}\right)$$

$$S = 357x10^6$$

Akış tahmini için %90 güven düzeyindeki güven aralığı:

$$4461x10^6 - 1{,}65x357x10^6 = 3872x10^6 m^3$$

$$4461x10^{6} + 1,65x357x10^{6} = 5050x10^{6}m^{3}$$

4) HİPOTEZ TEST MODELLERİ VE HİPOTEZLERİN KONTROLÜ

4.1) Parametrelerle İlgili Hipotezler

Bir rastgele değişkenin toplumunun tümü gözlenemediğinden bu toplumun olasılık dağılım fonksiyonu ve bu dağılımın parametrelerinin değerleri hiçbir zaman kesin olarak belirlenemez. Ancak pratikte parametrelerin değerleri ile ilgili bazı kararlar vermek gerekir. Bu durumda parametrelerin doğruluğu eldeki örnekten elde edilen bilgilerle karşılaştırılarak kontrol edilebilir. Söz konusu istatistiğin örnekleme dağılımı biliniyorsa değişme bölgesi kabul bölgesi ve red bölgesi olmak üzere ikiye ayrılır. Hesaplanan istatistik değeri kabul bölgesine düşüyorsa hipotez kabul edilir, kritik bölgeye düşüyorsa reddedilir.

Kabul ve red bölgelerinin belirlemek için iki büyüklüğün seçilmesi gerekir. Bunların birisi hipotezin kontrolünde kullanılacak anlamlılık düzeyidir. α anlamlılık düzeyi α=1-Pc şeklinde tanımlanır. Yapılan hipotezin (H_0=sıfır hipotezi) diğer bir hipoteze (H_1=karşıt hipotez) göre kontrolünde kullanılacak α anlamlılık düzeyi seçildikten sonra bu hipotez şu şekilde kontrol edilir:

a) H_0 hipotezinin karşıt hipotezi olan H_1 hipotezi parametrenin değerinin β_0 değerine eşit olmadığı şeklinde söz konusu istatistiğin örnekleme dağılımında aşılması olasılığı $\alpha/2$ ve $1-\alpha/2$ olan değerler belirlenir. β_0 örnekleme dağılımın ortalaması olarak alınır. Şeklen kritik bölgeler işaretlenir. Eldeki örnekten hesaplanan istatistik değeri kritik bölgeden birine düşüyorsa hipotez red, aksi halde kabul edilir. Buna iki uçlu test denir.

b) H_1 hipotezi parametrenin değerinin β_0'dan büyük (yada küçük) olduğu şeklinde ise söz konusu istatistiğin örnekleme dağılımında aşılması olasılığı α olan değer belirlenir. β_0 değeri yine örnekleme dağılımının ortalaması olarak alınarak şekil çizilerek bölge işaretlenir. Örnekteki hesaplanan istatisyik kritik bölgeye düşüyorsa hipotez red, aksi halde kabul edilir. Buna bir uçlu test denir.

İstatistik hipotezlerin kontrolü sonunda verilen kararların hatalı olabilmesinin kaçınılmaz olduğuna dikkat edilmelidir. Bir istatistik hipotezin kontrolünde verilen kararla gerçek durum arasındaki ilişki 4 farklı şekilde alınabilir.

KARAR	Gerçek Durum	
	H_0 doğru	H_0 yanlış
H_0 kabul et	Doğru karar	Yanlış karar (II. tip hata)
H_0'ı reddet H_1'i kabul et	Yanlış karar (I. tip hata)	Doğru karar

Görüldüğü gibi bir testte verilen kararın yanlışlığı 2 şekilde olabilmektedir.

1. I. Tip Hata: Yapılmış olan H_0 hipotezi gerçekte doğru olduğu halde testte reddedilmektedir.

2. II.Tip Hata: Yapılmış olan H_0 hipotezi gerçekte yanlış olduğu halde testte kabul edilmektedir.

Testte seçilen α anlamlılık düzeyi aynı zamanda I.Tip hata yapabilmesi olasılığını göstermektedir. Hidrolojik uygulamalarda anlamlılık düzeyi genellikle 0,05 yada 0,10 seçilir. II. Tip hata yapmak olasılığını azalttıkça (yani α'ı küçülmekte) II. Tip hata yapmak olasılığı arttırılmış olacağından α istenildiği kadar küçültülemez.

4.2) İki Örneğin Homojenliğinin Kontrolü

Eldeki iki ayrı örnekten hesaplanan aynı istatiğin değerleri arasında istatistik bakımından anlamlı bir fark bulunup bulunmadığı araştırılabilir. Böylece bu iki örneğin aynı toplumdan gelip gelmedikleri kontrol edilmiş olur. Bu yöntem hidrolojide özellikle eldeki verilerin homojen olup olmadıklarını ve verilerde sistematik hatalar bulunup bulunmadığını kontrolde kullanılır. herhangi bir nedenle bir ölçümler dizisindeki verilerin homojenliğinin bir noktadan başlayarak bozulduğundan şüphe ediliyorsa bu noktadan önce ve sonraki ölçümlerin oluşturduğu iki örnek bu şekilde kontrol edilir. Sistematik hataların varlığını araştırmak için de ölçümler dizisi herhangi bir şekilde iki örneğe ayrılarak kontrol edilir. İki komşu istasyonda ölçülen değerlerin aynı toplumdan geldiklerinin kabul edilip edilmeyeceği de aynı şekilde araştırılır.

4.3) Olasılık Dağılım Fonksiyonuyla İlgili Testler

Gözlenen bir örnekten elde edilen frekans dağılım fonksiyonunun seçilen bir teorik o.d.f.'ye uygunluğunu kontrol etmek için en çok kullanılan 2 test vardır:

a) x^2 testi: N elemanı örneğin m sınıf aralığına ayrılarak analizi sonunda elde edilen sınıf frekansları f_i (i=1,2,...,m) olsun. Seçilen o.d.f.'dan aynı sınıf aralıkları için hesaplanan olasılıklar p; (i=1,2,...m) ile gösterilsin.

* $x^2 = N \sum_{i=1}^{m} \frac{(f_i - p_i)^2}{p_i}$ istatistiğinin örnekleme dağılımı asimptotik olarak s.d.=m-1 olan x^2 dağılımdır. Buna göre gözlenen frekans dağılımının seçilen teorik dağılıma uygun olduğu hipotezini kontrol etmek için seçilen α anlamlılık düzeyine göre ilgili tablodan aşılması olasılığı α olan $x\alpha^2$ değeri okunur. * formulüyle hesaplanan x^2 değeri $x\alpha^2$ değeri okunur. * formulüyle hesaplanan x^2 değeri $x\alpha^2$'den küçükse hipotez kabul edilir.

x^2 testini uygularken sınıf aralığı sayısı m≥s ve herbir sınıf için N.pi≥5 olmalıdır. Sınıf aralıklarının aynı genişlikte seçilmesi şart değildir. Bunları pi olasılıkları eşit olacak şekilde (pi=1/m) seçmek uygun olur. X^2 testi ile bir frekans dağılımı çeşitli teorik dağılımlarla da karşılaştırılabilir, en küçük x^2 değerini veren dağılımın gözleme en uygun dağılım olduğu kabul edilir.

b) Smirnov-Kalmogorov testi:

Eldeki örneğin düzenlendiğini ve düzenlenmiş örnekten frekans dağılımının:

$Fa(x_i) = \frac{i}{N}$ şeklinde hesaplandığını düşüneli. Seçilen dağılım fonksiyonu F(x) ile gösterilirse:

1- $D = \max_i |F(x_i)-Fa(x_i)|$ istatistiğinin örnekleme dağılımı bilinmektedir, bu dağılım göz önüne alınan o.d.f.'den bağımsızdır. Bu dağılım bilindiğine göre seçilen α olan Δ_α değeri ilgili tablodan okunur. (Δ_α örnekteki, N, eleamn sayısına bağlıdır. Formül 1'den hesaplanan Δ değeri Δ_α'dan küçükse hipotez kabul, aksi halde reddedilir. Bu test küçük örnekler halinde x^2 testine göre daha uygundur.

4.4. İki Değişkenin Bağımlılığının Kontrolü

x^2 testini başka bir kullanılış yeri de iki rastgele değişkenin bağımsız olup olmadıklarının araştırılmasıdır. Bu iki değişkene ait N elemanlı bir örneği değişkenlerden birini m, diğerini n adet sınıf aralığına ayırarak incelendiğimizi ve fiğ ortak frekans değerlerini hesapladığımızı düşünelim. Değişkenlerin marjinal dağılımlarının frekansları da fi ve fj olarak bulunmuş olsun. Bu durumda:

$$* \; x^2 = N \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(fij - fi.fj)^2}{fi.fj} \quad \text{istatistiğinin dağılımı}$$

s.d.= (m-1) (n-1) olan x^2 dağılımıdır. Buna göre eldeki örnekten * denklemiyle hesaplanan x^2 istatistiğinin değeri, seçilen α olan $x\alpha^2$ değerinden küçükse bu iki değişkenin bağımsız olduğu hipotezi kabul, aksi halde reddedilir. (probleme geçiş...) $\rightarrow$ (26. sayfa)

5. HİDROLOJİK SÜREÇLER VE AKIŞ SERİLERİNİN MODELLERİ

Birçok hidrolojik değişkenlerde ardışık gözlemlerin birbirinden bağımsız olmadıkları görülür. Örneğin, bir akarsudaki ardışık günlerin akışları arasında havanın depolama özelliklerine dayanan kuvvetli bir bağımlılık bulunur. Akışın fazla olduğu bir günü yine akışı fazla olan günlerin takip ettiği, benzer şekilde akışın az olduğu günlerin de bir araya kümelendiği görülür. Böyle bir iç bağımlılığa stokastik bağımlılık denir. Aralarında stokastik bağımlılık bulunan gözlemlerden oluşan bir zaman serisine de stokastik süreç denir. Stokastik süreleri incelerken sadece rastgele değişkenin olasılık dağılımını bilmek yeterli olmaz, ayrıca değişkenin iç bağımlılığını ifade eden bir model de kurmak gerekir. Stokastik süreçlerin modellerini kurulmasındaki amaç bu modeller yardımıyla söz konusu değişken için sentetik seriler üretilmesidir. Bu serilerin yardımıyla su kaynaklarının planlanması ve işletilmesiyle ilgili çalışmalarda akışlar için sadece gözlenmiş olan örneği değil, aynı toplumdan geldiği kabul edilebilecek başka örnekleri de göz önüne almak mümkün olabilir.

Örnek Problem 4)

Göta nehrinde 150 yıl boyunca yapılan ölçmeler sonunda yıllık akışların ortalaması 530 m^3/s, standart sapması 97 m^3/s olmak üzere normal dağıldığı görülmüştür. Yıllık akışlarda serisel bağımlılık mevcut olup r_1 = 0,463 olarak hesaplanmıştır. Bu akışlara ait 30 yıllık örneklerden hesaplanacak ortalama ve varyans istatistikleri için %90 güven düzeyinde güven aralıklarını belirleyiniz.

Çözüm:

Serisel bağımlılığın Markov süreci şeklinde olduğunu kabul ederek;

Ortalama için:

$$Ne= N\frac{1-r}{1+r} = 30\ \frac{1-0{,}463}{1+0{,}463} = 11$$

Varyans için;

$$Ne = N\ \frac{1-r_1^{\ 2}}{1+r_1^{\ 2}} = 30\frac{1-0{,}463^2}{1+0{,}463^2} = 19$$

Ortalamanın örnekleme dağılımının varyansı:

$$\frac{Q^2}{Ne} = \frac{97^2}{11} = 855$$

Standart sapması:

$(855)^{1/2} = 29$

%90 güven düzeyindeki güven aralığı:

(530-1,65x29=482, 530+1,65x29=578)

Varyansın örnekleme dağılımının varyansı

$$\frac{2Q^4}{Ne} = \frac{2x97^4}{19} = 9318872$$

Bir stokastik sürecin Tam olarak belirlenebilmesi için şu özelliklerin bilinmesi gereklidir.

1. Sürecin rastgele değişkeninin olasılık: Bu dağılım genel olarak zaman içinde değişebilir, fakat ileride görüleceği gibi dağılımın zamanla değişmemesi yani sürecin bütün elemanlarının aynı toplumdan gelmesi halinde inceleme çok daha kolay olur.

2. Sürecin stokastik bağımlılığı: rastgele değişkenin ardışık gözlemlerde alacağı değerlerin arasındaki ilişkinin de bilinmesi gerekir.

Stokastik süreçler incelenirken ardışık gözlemler arasındaki zaman aralığı sonlu alınabileceği gibi sonsuz küçük de alınabilir, buna göre sırasıyla kesikli yada sürekli seriler söz konusu olur. Hidrolojide daha çok kesikli serilerle çalışır. Bunun nedeni ölçümlerin sürekli yapılanmadığı hallerde bile hesap kolaylığı bakımından çoğunlukla sonlu bir zaman aralığı boyunca alınan ortalamalarla çalışarak kesikli seriler kullanılır.

5.1) Başlıca Akış Serisi Modelleri:

Hidrolojide stokastik süreçleri incelemenin amacı bu süreçlerin yapısını ifade eden matematik modellerin kurulmasıdır. Kurulacak matematik modelin kullanılacağı amaca göre söz konusu sürecin istatistik özelliklerini yeterli bir şekilde ifade edebilmesi gerekir.modeli bu şartı sağlayacak en basit, yani parametre sayısı en az olan model seçmek uygun olur. Zira parametre sayısı arttıkça bunların değerlerinin eldeki örnekten hesabındaki güvenilirlik azalır. Akış serilerinin modellerinin kurulmasında sırasıyla şu çalışmalar yapılır.

1. Model tanımlaması: Eldeki verilere en iyi uyan model tipinin belirlenmesi

2. Parametrelerin tahmini: Seçilen modelin parametrelerinin değerlerinin eldeki verileri kullanarak hesaplanması (modelin kalibrasyonu)

3. Modelin kontrolü: Böylece tam olarak belirlenmiş olan modelin gözlenmiş seriye uygunluğunu istatistik testlerle kontrol ederek modelin kabul edilip edilmeyeceğine karar verilir.

Hidrolojik süreç modellerine genel olarak bakıldığında bunların iki bileşenin toplamı şeklinde olduğu görülür.

$x_i=d_i+\in_i$, Burada d_i modelin deterministik bileşenini göstermektedir. Bu bileşen sürecin daha önceki onlardan almış olduğu x_{i-1}, x_{i-2} değerlerinin bir fonksiyonudur. Bu fonksiyonun şekli ve modelin parametreleri modelden modele değişir. Modelin karmaşıklığına göre ortalama, standart sapma, oto korelasyon katsayıları gibi parametreler modelde görülürler. $\in_i$ bileşeni ise modelin rastgele bileşeni olup bağımsız bir süreç oluşturur.

5.2) Yıllık Akışların Modelleri

Yıllık akış serilerinin bazılarında iç bağımlılık çok küçük olduğundan bunların bağımsız süreç oldukları kabul edilebilir. Buna karşılık birçok yıllık akış serisinde ihmal edilmeyecek stokastik bağımlılıklar bulunduğu görülmüştür. Bunun nedeni yer altı biriktirme haznesinin akışlara katsayısıdır.

Bu gibi seriler için en çok kullanılan modeller aşağıda ele alınmaktadır.

5.2.1) Lineer otoregresif modeller:

$$x_i = \sum_{j=1}^{m} aj.x_{i-j} + \varepsilon_i$$

Burada x_i İ'inci yılda görülen akış, a; modelin parametreleri olan otoregresif katsayılar, $\in_i$ bağımsız bir süreç oluşturan normal dağılmış bir değişkendir, m modelin mertebesini göstermektedir. m'inci mertebe Markov modeline herhangi bir yılın akışının ondan önceki m yılın akışlarına bağlı olarak ifade edildiği görülür.

Morkov modellerinin en basit tipi 1. mertebe Markov modelidir. Bu modelde herhangi bir yılın akışı sadece ondan önceki yılın akışına bağlı olarak;

$x_i = a_i.x_{i-1} + \varepsilon_i$ ifadesine varılır. Bu model $y = (x - \mu_x)/Q_x$ standart değişkeni için kurulursa

$* y_i = g_1 y_{i-1} + \varepsilon_i$ ifadesine varılır. Bu modelde parametre akış serisinin 1. mertebe otokorelasyon katsayısı (g_1) olmaktadır. $\in_i$ normal dağılmış bağımsız değişken ortalaması =, varyansı $1\text{-}g^2$ dir. * denklemiyle verilen Markov sürecinin korelogramı belirlenirse;

$$g_k = g_1^{\ k}$$

Hidrolojide en çok 1. ve 2. mertebe Markov modellerin kullandığı görülür. 2. mertebe modelin katsayıları şu denklemlerle hesaplanır.

$$a_1 = \frac{g_1 - g_1.g_2}{1 - g_1^{\ 2}}, \qquad a_2 = \frac{g_2 - g_1^{\ 2}}{1 - g_1^{\ 2}}$$

5.2.2) Hareketli ortalama modelleri: Bu modellerde x_i, belli sayıda bağımsız $\in_i$ değişkenlerinin ağırlıklı bir ortalaması olarak ifade edilir.

$$x_i = \sum_{j=0}^{m} b_j.\varepsilon_{i-j}$$

Bu modelde m modelin mertebesini gösterir b_j katsayıları ile serinin otokorelasyon katsayıları arasındaki bağıntılar belirlenip b_j değeri burada bulunur.

Hareketli ortalama modelin korelogramı:

(1) $g_k = \dfrac{\sum_{j=0}^{m} b_j.b_j + k}{\sum_{j=0}^{m} b_j^{\,2}}$ ifadesi ile verilir, bu ifadede j+k>m için $b_{j+k}=0$ alınacaktır. (k>m için g_k=0) b_j=1/(m+1) olan basit hareketli ortalama modelinde (1) denklemine göre;

$g_k = 1 - \dfrac{k}{m}$ k≤m olduğu görülür.

5.2.3) ARMA modelleri: Lineer otoregresif süreçlerle hareketli ortalamaların karışımı olan bu modellerin denklemlerini kısa bir şekilde yazabilmek için önce bazı operatörleri tanımlamak gerekir. Geri kaydırma operatörü:

$B^m Z_i = Z_{i-m}$ şeklinde tanımlanır. Bu operatörü kullanarak $Z_i = x_i - \bar{x}$ değişkeni cinsinden modelin denklemi şu şekilde yazılabilir.

(1) $Z_i = \varepsilon_i - Q_1\varepsilon_{i-1}Q_2\varepsilon_{i-2} - \ldots\ldots\ldots\ldots - Q_q\varepsilon_{i-q}$ olduğuna göre Q(B) operatörü aşağıdaki şekilde tanımlanırsa:

$Q(B) = 1 - Q_1B - Q_2B^2 - \ldots\ldots\ldots\ldots - Q_qB^q$

(1) denklemi kısaca $Z_i = Q(B)\varepsilon_i$ denklemi şeklinde yazılabilir. Lineer otoregresif model için de benzer şekilde:

$Z_i = \phi_1 Z_{i-1} + \phi_2 Z_{i-2} + \ldots\ldots\ldots\ldots + \phi_q Z_{i-q}\varepsilon_i$ denklemi, ∅(B) operatörünü: $\phi(b) = 1 - \phi_1 B - \phi_2 B^2 - \ldots\ldots\ldots\ldots - \phi_p B^p$ şeklinde tanımlayarak kısaca $\phi(B)_{Zi} = \varepsilon_i$ ve $\phi(B)_{Zi} = Q(B)_{\varepsilon i}$ (2)

(2) denklemi açık şekilde yazılacak olursa;

$Z_i = \phi_1 Z_{i-1} + \ldots\ldots\ldots + \phi_{1p} Z_{i-p} + \varepsilon_i - Q_1\varepsilon_{i-1} \ldots\ldots\ldots\ldots Q_q\varepsilon_{i-q}$

(2)denklemiyle ifade edilen modele ARMA (p, q) modeli denir. Bu modelin parametrelerinin sayısı p+q+2'dir bunlardan p adedi, ∅ katsayıları, q adedi Q katsayıları, birisi $\bar{x}$ ve birisi $Q\varepsilon^2$'dir. p=0 için hareketli ortalama (µA), q=0 için Mardov (AR) modelleri elde edilir.

Yıllık akışlar için en çok kullanılan ARMA (1.1) modelinin denklemi

$Z_i = \phi_1 Z_{i-1} + \varepsilon_i - Q\varepsilon_{i-1}$ şeklindedir.

Bu modelin korelogramı şu şekildedir.

$$g_i = \frac{(1-\phi_1 Q_1)(\phi_1 - Q_1)}{1+Q_1^2 - 2\phi_1 Q_1}, g_k = \phi_1 g_{k-1} \qquad (k\geq 2)$$

Eldeki örnekten hesaplanan r_1 ve r_2 değerlerini yukarıdaki denklemde g_1 ve g_2 yerine koyarak $\varnothing_1$ ve Q_1 parametreleri hesaplanabilir.

Bir sürecin $\varnothing_{11}$ kısmi korelasyon katsayıları ise şu şekilde tanımlanırlar:

$$\phi_{11} = r_1, \qquad \phi_{11} = \frac{r_l - \sum_{j=1}^{l-1} \phi_{l-1}, j.r_{l-j}}{1 - \sum_{j=1}^{l-1} \phi_{l-1}, j.r_j} \qquad l \geq 2$$

Gerek otokorelasyon fonksiyonu, gerekse kısmi otokorelasyon fonksiyonu büyük k değerleri için de sıfıra yaklaşmıyorsa süreç hem hareketli ortalama, hem otoregresif bileşenleri olan bir ARMA (p, q) modeli ile temsil edilebilir.

5.3) Aylık Akışların Modelleri

5.3.1) Korelogramın periyodikliğini korumayan modeller:

Sürecin otokorelasyon katsayılarının periyodikliği ihmal edilebilirse, Yani bu katsayıların yıl boyunca aydan aya değişmeyip sabit bir değerde kaldıkları kabul edilirse, süreci stasyoner hale getirmek için (*) denklemiyle verilen dönüşümü uygulamak yeterli olur.

$$(*)\ y_{i,T} = \frac{x_{i,T} - \bar{x}_T}{S_T} \qquad T=1,2,........,12$$

Burada x_T ve s_T aylık akışların Fovrier açılımıyla belirlenen periyodik ortalama ve standart sapmalarını göstermektedir.

5.3.2) Korelasyon periyodikliğini koruyan modeller:

Otokorelasyon katsayılarının yıl boyunca değişiminin de modelde korunması isteniyorsa şu iki yoldan birine gidilebilir.

a) (*) denklemiyle tanımlanan y_i değişkeni için kurulan modeldeki katsayılar yıl boyunca periyodik olarak değiştirilir. 1. mertebe Markov modeli bu durumda:

$y_{i,T} = g_{1,T} \cdot y_{i-1,T-1} + \varepsilon_i$ şeklini alır. $G_{1,T}$ yılın T ve T-1'inci aylarındaki akışlar arasındaki 1. mertebeden otokorelasyon katsayısının Fovrier açılımıyla belirlenen periyodik bileşenidir.

b) Aylık akışlar için çok kullanılan bir model tipi de parametreleri yıl boyunca değişen 1. mertebe Markov modelidir.

$$x_{i,j} = \bar{x}_j + b_j(x_{i-1,j-1} - \bar{x}_{j-1}) + s_j(1 - j_j^{\,2})^{1/2}\varepsilon_i$$

Bu denklemde i indisi aydan aya sürekli olarak değişir. x_j ve s_j yılın j'inci akışlarının ortalaması ve standart sapmasıdır. b_j ve r_j yılın j ve j-1'inci aylarındaki akışlar arsındaki regresyon ve korelasyon katsayıları olup b_j şu şekilde hesaplanır.

$$b_j = \frac{s_j}{s_{j-1}} r_j$$

Burada r_j'nin j'inci mertebeden otokorelasyon katsayısını değil, yılın j ve j-1'inci aylarındaki akışlar arasındaki 1. mertebeden korelasyon katsayısını gösterdiğine dikkat edilmelidir.

Aylık akışlar türetilirken bir yandan yıllık akışların istatistik özelliklerinin de korunması istenilirse şu şekilde hareket edilebilir. Önce yıllık akışlar için bir model kurulup parametreleri belirlenerek yıllık akışlar serisi türetilir. Sonra aylık akışlar için kurulan modele göre her biri yılın aylık akışları türetilir, ancak türetilen bu aylık akışlar toplamları o yılın türetilen yıllık akışına eşit olacak şekilde belli bir oranda arttırılır veya azaltılır. Böylece gerek yıllık gerekse aylık akışların istatistik değerleri korunmuş olur. Takvim yılı yerine su yılıyla çalışmak uygundur.

5.3.3) ARIMA modelleri:

ARMA modelleri ve bunların stasyoner olmayan süreçler için genelleştirilmiş şekilleri olan ARIMA modelleri de aylık akışlar için kullanılabilir.

a) ARMA modellerini aylık akışlar için kullanırken önce;

$y_{i,T} = \dfrac{x_{i,T} - \bar{x}_T}{s_T}$ T:1,2,........,12 denklemiyle verilen dönüşüm yapılır.

Böylece elde edilen ve stasyoner olduğu kabul edilen y_i süreci için bir ARMA modeli kullanılır.

b) Önce ardışık yıllarda aynı ayların akışlarının farkı alınarak yeni bir değişken tanımlanır.

$$W_i = x_i - x_{i-s}$$

Aylık akışlar için s=12'dir W_i süreci ortalaması bakımından stasyoner hale getirilmiş olup şöyle bir denklemle ifade edilebilir.

$$\Phi(B^s)W_i = \Theta(B^s)\alpha_i$$

α_i için benzer denklem yazılırsa

$$\phi(B)\alpha_i = Q(B)\varepsilon_i$$

Yukarıdaki 2 denklem bir araya getirilerek;

$$\phi(B)\Phi(B^2)W_i = Q(B)\Theta(B^s)\varepsilon_i$$

ARIMA model olarak modeli olarak bilinen bu süreçlerin aylık akışlar için kullanılabilecek en basit şekli ∅ ve Q'yı 1. mertebeden alıp diğer operatörleri kullanmadan elde edilir.

$$(1-\phi_i B)W_i = (1-\Theta_i B^{12})\varepsilon_i$$

Bu denklem açık şekilde yazılırsa;

$$W_i - \phi W_{i-1} = \varepsilon_i - \Theta_{i-12}$$

Bu modelin $\varnothing_1$, Θ_1 ve Q_E^2 olmak üzere sadece 3 parametresi vardır. Modelin korelogramı:

$$g_k = \phi_1 g_{k-1} \qquad (k \neq 12n)$$

$$g_k = \phi_1 g_{k-1} - \Theta_1 \frac{Q\varepsilon^2}{Qw^2} \qquad (k = 12n) \text{ şeklindedir.}$$

Modelin parametreleri şu denklemlerden elde edilirler.

$$\phi_1 = g_1$$

$$g_{12} = \phi_1^{12} - \Theta_1 \frac{Q\varepsilon^2}{Qw^2} \qquad\qquad Q\varepsilon^3 = \frac{1 - Q_1^2}{1 - \Theta_1^2} Qw^2$$

Örnek Problem 5)

Bir akarsuyun 51 yıl boyunca ölçülen yıllık debileri aşağıda gösterilmiştir. (m^3/sn)

398	429	690	437	567	460	288
520	563	490	437	368	522	548
694	762	729	713	615	626	575
452	520	480	414	437	351	420
444	590	489	533	753	576	696
737						

Bu akışların matematik modelini kurunuz.

Çözüm:

Verilen değerlerden şu istatistikler hesaplanır.

$\bar{x} = 544 m^3 / sn$

$Sx = 130 m^3 / sn$

$r_1 = 0{,}535$

$r^2 = 0{,}463$

Değişkenin standart şekli;

$y = \dfrac{x - \bar{x}}{Sx} = \dfrac{x - 544}{130}$ şeklinde tanımlanarak 1. mertebeden Markov modeli;

$y_i = 0{,}535 y_{i-1} + \varepsilon_i$ şeklinde kurulur.

$\in_i$, varyansı $1-0{,}535^2=0{,}714$ olan normal değişkendir.

2. mertebe Markov modelinin katsayıları hesaplanır.

$$a_1 = \frac{0{,}535 - 0{,}535x0{,}463}{1 - 0{,}535^2} = 0{,}403$$

$$a_2 = \frac{0{,}463 - 0{,}535^2}{1 - 0{,}535^2} = 0{,}248$$

$$y_i = 0{,}403 y_{i-1} + 0{,}248 y_{i-2} + \varepsilon_i$$

$\in_i$'nin varyansı:

$$1 - \frac{0{,}535^2 + 0{,}463^2 - 2x0{,}535x0{,}463}{1-0{,}535^2} = 0{,}671$$

1. ve 2. mertebe Markov modellerinden birinin seçimi hakkında karar verirken ilk olarak kurulan modelin deterministik kısmının değişken varyansının ne kadarını açıkladığına bakılabilir. 1. mertebe modeli için bu $r_1^2=0{,}535^2=0{,}286$, 2. mertebe model için 1-0,671=0,329

Bu bakımdan 2. mertebe modelinin üstün olduğu görülür.

ARMA (1.1) modeli $Z_i = x_i - \bar{x} = x_i - 544$ değişkeni için

$Z_i = \phi Z_{i-1} + \varepsilon_i - Q_1 \varepsilon_{i-1}$ şeklinde yazılabilir.

$\varnothing_1$ ve Q_1 parametrelerinin tahmini

$$r_1 = \frac{(1-\phi_1 Q_1)(\phi_1 - Q_1)}{1 + Q_1^{\,2} - 2\phi_1 Q_1}$$

$$r_2 = \phi_1 r_1$$

denklemlerinden $r_1=0{,}535$ ve $r_2=0{,}463$ koyarak $\varnothing_1=0{,}865$ $Q_1=0{,}40$ olarak bulunur.

Sonuç olarak;

$$Z_i = 0{,}865 Z_{i-1} + \varepsilon_i - 0{,}40 \varepsilon_{i-1}$$