

HAFTA 7

– Biçimsel yöntemler:

1. Park sınaması:

Önerilen ilişki $\sigma_i^2 = \sigma^2 X_i^\beta e^{v_i}$ olup, v_i = olasılıklı hata terimini göstermek üzere

$$\Rightarrow \ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$

doğrusal modeline dönüşür. Genellikle σ_i^2 bilinmediğinden Park bunun yerine $\hat{\varepsilon}_i^2$ 'yi kullanarak

$$\ln \hat{\varepsilon}_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i = \alpha + \beta X_i^* + v_i$$

α X_i^*

regresyon modelini bulmayı önerir. Eğer β istatistiksel anlamlı çıkarsa verilerde değişen varyans bulunduğunu gösterir. Anlamsız çıkarsa sabit varyans varsayımı kabul edilir.

I. Aşama: Değişen varyansa aldırılmaksızın EKK regresyonu bulunur ve artıklar hesaplanır ($\hat{\varepsilon}_i$)

II. Aşama: I. aşamada tanımlanan modelin kestirim denklemi bulunur.

Dezavantaj: $\hat{\varepsilon}_i$ 'ler değişen varyanslı olabilir.

Örnek: Dayanıksız mal üreten sanayilerde, 1958 de kuruluşta çalışan sayısına göre çalışan başına yapılan ödemeler (\$).

Y = ortalama ödeme (1000\$)

X = ortalama verimlilik (1000\$)

i = kuruluşun çalışan sayısına göre i . büyüklük sırası

<u>Y</u>	<u>X</u>	
3396	9355	Regresyon modeli: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$
3787	8584	Kestirim denklemi: $\hat{Y}_i = 1906.16 + 0.241 X_i$
4013	7962	
4014	8275	Std. error 920.657 0.098
4146	8389	t 2.070 2.457
4241	9418	p -değeri 0.077 0.044
4387	9795	$R^2 = 0.463$
4538	10281	$R = 0.68$
4843	11750	

Tahmin edilen $\hat{\beta}_1$ eğim katsayısı %5 düzeyinde anlamlı olduğu söylenir. İş gücü verimliliği bir dolar artarsa, işçiye ödeme ortalama 24 sent yükselir. Modelden elde edilen $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ artıklarının $\ln \hat{\varepsilon}_i^2 = \alpha + \beta \ln X_i + v_i$ modelinde önerildiği gibi X_i 'ye göre regresyon modeli

$$\ln \hat{\varepsilon}_i^2 = 32.802 - 2.477 \ln X_i$$

Std.hata 37.780 4.137

$$\begin{array}{r}
t \quad 0.868 \quad -0.599 \\
p\text{-değeri} \quad 0.414 \quad 0.568 \\
R^2 = 0.049
\end{array}$$

$p\text{-değeri}=0.568 > \alpha=0.05$ olduğundan H_0 hipotezi red edilemez. β anlamsız, Park sınamasına göre hata varyansında değişme olmadığı söylenir.

2. Spearman sıra korelasyon sınaması:

$$\text{Spearman sıra korelasyon katsayısı } r_s = 1 - 6 \left[\frac{\sum_{i=1}^n d_i^2}{n(n^2 - 1)} \right]$$

$d_i = i$. kişi ya da olgunun iki farklı özelliğine verilen sıra numaraları arasındaki fark
 $n =$ sıralana kişi ya da olgu sayısı

Bu sıra korelasyon katsayısı değişen varyansı bulmada kullanılan model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ olsun.

1.Adım: X ve Y verilerine regresyon analizi uygulanıp, kestirim modelinden $\hat{\varepsilon}_i$ artıkları bulunur.

2.Adım: $\hat{\varepsilon}_i$ artıklarının mutlak değerleri alınarak, hem $|\hat{\varepsilon}_i|$, hem de X_i 'i (ya da \hat{Y}_i 'i) artan ya da azalan bir sıraya dizilir. Daha sonra Spearman korelasyon katsayısı r_s bulunur.

3.Adım: $n > 8$ olduğu varsayımı altında

$$H_0 : \rho_s = 0 \text{ hipotezinin test istatistiği } t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} \sim t_{n-2} \text{ bulunur.}$$

4.Adım: Karar ve yorum: $t > t_{n-2}^*(\alpha)$ ise değişen varyans vardır.

$$t \leq t_{n-2}^*(\alpha) \text{ ise değişen varyans yoktur.}$$

Her bir $|\hat{\varepsilon}_i|$ ile X_i değerleri için r_s değişkenleri hesaplanır ve test istatistiği t bulunur.

Örnek: $E_i =$ portföyün beklenen getirisi (ortalama yıllık getiri %)

$\sigma_i =$ bu getirinin standart sapması

Portföy kuramının sermaye piyasası doğrusu

$$E_i = \beta_0 + \beta_1 \sigma_i \Rightarrow \hat{E}_i = 5.8194 + 0.4590 \sigma_i$$

E_i	σ_i	\hat{E}_i	$ \hat{\varepsilon}_i = E_i - \hat{E}_i $	$ \hat{\varepsilon}_i $ 'nin sırası	σ_i 'nin sırası	d_i fark	d_i^2
12.4	12.1	11.37	1.03	9	4	5	25
14.4	21.4	15.64	1.24	10	9	1	1

14.6	18.7	14.40	0.20	4	7	-3	9
16.0	21.7	15.78	0.22	5	10	-5	25
11.3	12.5	11.56	0.26	6	5	1	1
10.0	10.4	10.59	0.59	7	2	5	25
16.2	20.8	15.37	0.83	8	8	0	0
10.4	10.2	10.50	0.10	3	1	2	4
13.1	16.0	13.16	0.06	2	6	-4	16
11.3	12.0	11.33	0.03	1	3	-2	4
						0	110

Spearman sıra korelasyon katsayısı

$$r_s = 1 - 6 \left[\frac{\sum_{i=1}^n d_i^2}{n(n^2 - 1)} \right] = 1 - 6 \left[\frac{110}{10(100 - 1)} \right] = 0.3333$$

Test istatistiği

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} = \frac{(0.3333)\sqrt{10-2}}{\sqrt{1-(0.3333)^2}} = 0.9998$$

p -değeri $= 0.17 > \alpha = 0.05$ olduğu için H_0 hipotezi red edilemedi. Açıklayıcı değişken ile artıkların mutlak değerleri arasında düzenli bir ilişki yoktur. Yani değişen varyans yoktur.

3. Goldfeld-Quandt sınaması:

Değişen varyans σ_i^2 'nin regresyon modelindeki açıklayıcı değişkenlerden birine aynı yönlü bağlı olduğu varsayılırsa

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

modeli ele alınsın. σ^2 sabit olmak üzere σ_i^2 'nin X_i 'ye aynı yönlü bağlı olduğu varsayımı $\sigma_i^2 = \sigma^2 X_i^2$ olsun. Bu uygun bir varsayım ise X_i değerleri büyüdükçe σ_i^2 'nin de büyüyeceği anlamına gelir.

Goldfeld ile Quandt'ın önerdiği sınama:

1.Adım: Gözlemleri en küçükten başlayarak X_i değerlerine göre sıralanır.

2.Adım: Bir c sayısı belirlenip tam ortaya gelen c tane gözlem dışlanır. Kalan $n-c$ tane gözlem her yarıda $\frac{n-c}{2}$ tane gözlem kalacak şekilde ikiye ayrılır.

3.Adım: İlk $\frac{n-c}{2}$ tane gözlem ile son $\frac{n-c}{2}$ tane gözlem için ayrı ayrı regresyon en küçük kareler modelleri bulunup, SSE_1 ve SSE_2 değerleri bulunur. SSE_1 küçük X_i

değerlerine (küçük varyanslı), SSE_2 büyük X_i değerlerine (büyük varyanslı) göre serbestlik dereceleri parametre sayısı k olmak üzere $\frac{n-c}{2} - k$ ya da $\frac{n-c-2k}{2}$ dır.

4.Adım: Test istatistik değeri $\lambda = \frac{SSE_1/sd}{SSE_2/sd} \sim F_{sd, sd}$ dır. Eğer $\lambda > F^*$ tablo değerinden büyük ise sabit varyans varsayımı reddedilir. Değişen varyans olasıdır.

Not: Modelde birden fazla X değişkeni varsa bu sınama her bir X değişkeni için ayrı ayrı yapılır.

Monte-Carlo simülasyon denemelerine göre

$$\begin{array}{lcl} n = 30 & \Rightarrow & c = 8 \quad \Rightarrow \quad c = 4 \\ n = 60 & \Rightarrow & c = 16 \quad \Rightarrow \quad c = 10 \\ & & \text{Goldfelt-Quandt} \quad \quad \text{Judge} \end{array}$$

önermektedir.

Örnek: $n = 30$ ailenin tüketim harcamaları ile gelirleri arasındaki ilişkinin araştırılmasında;

$Y =$ tüketim harcamaları

$X =$ gelir

olmak üzere veri Temel İstatistik-Gujarati kitabının sayfa 376'dan alınmıştır.

$$n = 30 \Rightarrow c = 4 \text{ alınmıştır ve } \frac{n-c}{2} = \frac{30-4}{2} = 13 \text{ bulunmuştur.}$$

İlk 13 veriye dayalı kestirim modeli:

$$\hat{Y}_i = 3.4094 + 0.6968x_i$$

Std. hata 8.7049 0.0744

$$r^2 = 0.8887, \quad SSE_1 = 377.17, \quad sd=11$$

Son 13 veriye dayalı kestirim modeli:

$$\hat{Y}_i = -28.0272 + 0.7941x_i$$

Std. hata 30.6421 0.1319

$$r^2 = 0.7681, \quad SSE_2 = 1536.8, \quad sd=11$$

olarak bulunan sonuçlardan test istatistiği

$$\lambda = \frac{SSE_2/sd_2}{SSE_1/sd_1} = \frac{1536.8/11}{377.17/11} = 4.07$$

elde edilir. $\alpha = 0.05$ anlamlılık düzeyi için tablo değeri $F_{11,11}^*(0.05) = 2.82$ bulunur.

$\lambda = 4.07 > F_{11,11}^*(0.05) = 2.82$ olduğundan hata varyansının değıştiđi sonucuna varılır. Eğer $\alpha = 0.01$ sabit varyans varsayımı reddedilemeyecektir. $\lambda = 4.07$ test istatistiđi için p -deđeri 0.014 bulunur.