2024. április 2-i gyakorlat [1]

Előjel, sorozat és rangösszeg próba

1. Egy mozitulajdonos állítása szerint az egy-egy rajzfilmre hetente eladott gyermekjegyek mediánja 300. Állításának alátámasztására kiválasztott 8, a moziban vetített rajzfilmet, és feljegyezte, hogy egy-egy filmre egy adott héten mennyi gyermekjegyet váltottak. A következő eredményeket kapta:

   $$412,\ 232,\ 197,\ 454,\ 251,\ 114,\ 256,\ 318.$$

   Hipotéziseit pontosan megfogalmazva, az előjel próba segítségével döntsön 90%-os megbízhatósági szinten, igaz-e a mozitulajdonos állítása!

2. Egy olajtársaság állítása szerint az általuk az üzemanyagokhoz használt adalék lényegesen csökkenti a gépjárművek fogyasztását. Ennek ellenőrzésére 12 különböző középkategóriás benzinüzemű személygépkocsi fogyasztását tesztelték azonos körülmények között. A kapott eredmények az alábbiak (l/100 km):

   $$6.8,\ 6.5,\ 6.9,\ 7.3,\ 6.4,\ 6.7,\ 7.0,\ 6.8,\ 6.9,\ 6.5,\ 7.2,\ 6.4.$$

   Ismeretes, hogy a vizsgálatban résztvevő személygépkocsik fogyasztásának mediánja 7 l/100 km. Hipotéziseit pontosan megfogalmazva vizsgálja meg az olajtársaság állítását! Döntsön 95%-os megbízhatósági szinten!

3. Péter szeretné elvinni a feleségét a házassági évfordulójukon egy olyan étterembe vacsorázni, ahol még nem jártak. Hosszas keresés után sikerült kettőre, a Táncoló rákra és a Vidám cápára, szűkíteni a lehetőségek listáját. Mivel Péter jól ismeri a feleségét, ezért azokat a paramétereket, amelyeket a felesége fontosnak tart egy étteremben, egy 18 pontból álló listába gyűjtötte. Mindkét éttermet meglátogatva és összefoglalva az eredményeket, a következőt kapta:

   11 darab „+”, 3 darab „–” és 4 darab „0”,

   ahol „+” jelöli, hogy az adott szempontnál a Táncoló rák bizonyult jobbnak, „–” hogy a Vidám cápa, és „0” hogy nincs különbség a két étterem között. 5%-os szignifikancia szinten vizsgálja meg Péter azon állítását, hogy a Táncoló rák a jobb választás számukra!

4. Egy tanteremben a férfiak és nők elhelyezkedése a következőképpen alakult:

   $$NNNNNFFNNNNNFNNNNNNFNNF$$

   Döntsön 95%-os megbízhatósági szinten, hogy a nők és férfiak elhelyezkedése véletlen-e!

5. Annának 2 macskája van, Leó és Micóka, akik gondoskodnak gazdájuk reggeli ébresztéséről. Anna egy éven keresztül minden nap feljegyezte, hogy aznap melyik macskája keltette fel, s az adatokat az ebresztes.csv fájlban rödzítette. (Leo 178-szor, Micóka 187-szer ébresztette gazdáját, s a futamok száma 183.) Döntsön 5%-os szignifkancia szinten arról, hogy véletlennek tekinthető-e, hogy melyik reggel melyik macska ébresztette fel Annát!

6. Egy óvoda dolgozói kíváncsiak voltak arra, hogy jobban teljesítenek-e a kreatív foglalkozásokon azok a gyerekek, akik óvoda előtt bölcsődébe is jártak. A dolgozók a gyerekek teljesítményét különböző kritériumok alapján egy 0-tól 30-ig terjedő skálán értékelték. Az eredményeket az alábbi táblázat tartalmazza.

Járt bölcsődébe	25	24	20	23	19		
Nem járt bölcsődébe	23	22	24	20	21	28	23

Hipotéziseit pontosan megfogalmazva egy alkalmas nemparaméteres próba segítségével döntsön 5%-os szignifikancia szinten, van-e különbség a bölcsödei ellátásban részesülő és nem részesülő gyerekek teljesítménye között!

7. 2004 januárjában a budapesti Nagycsarnokban és a Lehel téri csarnokban 10-10 egymástól függetlenül kiválasztott gyümölcsárusnál a mandarin kilogrammonkénti árának alakulása a következő volt:

Nagycsarnok:	190;	170;	260;	275;	320;	340;	168;	280;	250;	250;
Lehel téri csarnok:	250;	360;	252;	190;	180;	290;	340;	210;	220;	240;

Állítható-e 5%-os szignifikancia szinten, hogy a két piacon azonos a mandarin árának eloszlása?

8. A közgazdasági alapképzés két évfolyamáról egymástól függetlenül, véletlenszerűen kiválasztott 5-5 hallgató év végi tanulmányi átlaga a következő:

A hallgató sorszáma	I. évfolyam	II. évfolyam
	tanulmányi átlaga	
1.	2,7	4,1
2.	4,0	3,3
3.	3,5	4,7
4.	4,2	3,6
5.	3,8	3,4

Vizsgálja meg 1%-os szignifikancia szinten, hogy a hallgatók tanulmányi átlagának eloszlása azonos-e a két évfolyamon!

2024. április 8-i gyakorlat

Kétváltozós lineáris regresszió

Kétváltozós lineáris regressziós modell

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad\qquad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

β_0, β_1 becslése legkisebb négyzetek módszerével:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} d_{x_i} d_{y_i}}{\sum_{i=1}^{n} d_{x_i}^2} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sum_{i=1}^{n} x_i^2 - n(\overline{x})^2}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Varianciafelbontás a kétváltozós lineáris modellre:

$$SST = SSR + SSE$$

$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum d_y^2$ teljes négyzetösszeg

$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} e_i^2$ belső négyzetösszeg, a hiba okozta (reziduális) négyzetösszeg

$SSR = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$ külső négyzetösszeg, regressziós vagy magyarázott négyzetösszeg

A *determinációs együttható* azt mutatja, hogy a regressziós modellel az y_i adatokban meglévő variancia hány százaléka szüntethető/magyarázható meg:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST},$$

$$R^2_{\text{adjusted}} = 1 - \frac{n-1}{n-k-1}(1 - R^2)$$

$R^2 \approx 1$ - jó illeszkedés, nagy magyarázó erő
$R^2 \approx 0$ - gyenge modellteljesítmény.

A mintából számolt becsült *lineáris korrelációs együttható* a magyarázó- és eredményváltozó között:

$$r = \frac{\sum d_x d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\,\overline{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right) \cdot \left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right)}}$$

$|r| \approx 1$ - szoros, közel lineáris függvényszerű kapcsolat
$r \approx 0$ - lineáris kapcsolat hiánya (korrelálatlanság)
$r > 0$ - a változók egy irányba mozognak
$r < 0$ - a változók ellentétes irányba mozognak

Kétváltozós lineáris esetben: $R^2 = r^2$.

Az *elaszticitás* (rugalmasság) azt méri, hogy az X változó 1%-os növekedése hány százalékos növekedést/csökkenést eredményez az Y változónál. Az elaszticitás kiszámítása a becsült eredményváltozóra:

$$El(\hat{y}, x) = \frac{\partial \hat{y}}{\partial x} \cdot \frac{x}{\hat{y}} = \hat{\beta}_1 \cdot \frac{x}{\hat{y}} = \frac{\hat{\beta}_1 x}{\hat{\beta}_0 + \hat{\beta}_1 x}$$

Intervallumbecslés a függvényértékekre:

- az átlagos értékre

$$Int_{1-\alpha}(E(Y_*)) = \hat{y}_* \pm t_{1-\frac{\alpha}{2}}(n-2)s_e\sqrt{\frac{1}{n}+\frac{(x_*-\overline{x})^2}{\sum d_x^2}},$$

- az egyedi értékre

$$Int_{1-\alpha}(Y_*) = \hat{y}_* \pm t_{1-\frac{\alpha}{2}}(n-2)s_e\sqrt{1+\frac{1}{n}+\frac{(x_*-\overline{x})^2}{\sum d_x^2}},$$

ahol $s_e = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}$ a *korrigált reziduális szórás.*

1. A fogászati világnapon az életkor és a fogak száma közötti összefüggést vizsgálták tíz önként vállalkozó segítségével:

Életkor (év)	Fogak száma (db)
18	30
40	21
20	29
71	8
36	23
55	14
34	25
62	12
17	31
30	28
$\sum$	$\sum$

(a) Határozza meg és értelmezze a lineáris regresszió paramétereit!

	$d_x = x_i - \overline{x}$	$d_y = y_i - \overline{y}$	x_i^2	$x_i y_i$	$\hat{y}$	$e_i = y_i - \hat{y}$
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
$\sum$						

(b) Számítsa ki a két változó lineáris korrelációs együtthatóját! r = -0,9923

(c) Számítsa ki és értelmezze a lineáris regresszió determinációs együtthatóját! R^2 = 0,9846

(d) Adjon becslést a 30 évesek fogainak átlagos számára, majd szerkesszen konfidenciaintervallumot ugyanerre 95%-os megbízhatósági szinten! $\hat{y}_{30}$ = 25,67; $Int_{0,95}$(25,67) = (25,6686; 25,6714)

(e) Adjon 95%-os konfidenciaintervallumot egy 30 éves ügyfél fogainak a számára!

(f) Számolja ki és értelmezze az elaszticitást az átlagos életkorra! El(y^, x) = -0,7445

SPSS: Graphs ⇒ Legacy Dialogs ⇒ Scatter/Dot

Transform ⇒ Compute variable: becslés, hiba

Analyze ⇒ Regression ⇒ Curve Estimation

Analyze ⇒ Regression ⇒ Linear

	x	y	xy	x^2	d_x	d_y	d_x^2	d_y^2	$d_x d_y$	y^	e_i
	18	30	540	324	-20,3	7,9	412,09	62,41	-160,37	30,8211	-0,8211
	40	21	840	1600	1,7	-1,1	2,89	1,21	-1,87	21,3702	-0,3702
	20	29	580	400	-18,3	6,9	334,89	47,61	-126,27	29,9619	-0,9619
	71	8	568	5041	32,7	-14,1	1069,29	198,81	-461,07	8,05298	-0,053
	36	23	828	1296	-2,3	0,9	5,29	0,81	-2,07	23,0885	-0,0885
	55	14	770	3025	16,7	-8,1	278,89	65,61	-135,27	14,9264	-0,9264
	34	25	850	1156	-4,3	2,9	18,49	8,41	-12,47	23,9477	1,05229
	62	12	744	3844	23,7	-10,1	561,69	102,01	-239,37	11,9193	0,08074
	17	31	527	289	-21,3	8,9	453,69	79,21	-189,57	31,2507	-0,2507
	30	28	840	900	-8,3	5,9	68,89	34,81	-48,97	25,6661	2,33394
Σ	383	221	7087	17875	2,8422E-14	-1,4211E-14	3206,1	600,9	-1377,3	221,0048	-0,0048

x átlag	38,3
y átlag	22,1
n	10

b)
c)
d)
e)
f)

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\left(\bar{x}^2\right)} = \frac{7087 - 10 * 38,3 * 22,1}{17875 - 10 * (38,3\text{^}2)} = \frac{-1377,3}{3206,1} = -0,42959$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 22,1 - (-0,4296) * 38,3 = 38,5537$$

$$\sum y_i = \hat{\beta}_0 n + \hat{\beta}_1 \sum x_i \,,$$

$$\sum x_i y_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2$$

$$El(\hat{y}, \bar{x}) = \frac{\hat{\beta}_1 \bar{x}}{\hat{\beta}_0 + \hat{\beta}_1 \bar{x}} = \frac{-16,454}{22,1} = -0,7445$$

$$r = \frac{-1377,3}{\text{sqrt}(3206,1 * 600,9)} = -0,9923$$

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum d_y^2} = \frac{\sum(\hat{\beta}_0 + \hat{\beta}_1 x_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{x})^2}{\sum d_y^2} = \hat{\beta}_1^2 \cdot \frac{\sum d_x^2}{\sum d_y^2} = r^2 . \quad = 0,9846$$

y30^ = 25,67

$$s_e = \sqrt{\frac{1}{n-2} \sum e_i^2} = 0,0017$$

$$s_{\hat{y}_*} = s_e \sqrt{\frac{1}{n} + \frac{(x_* - \bar{x})^2}{\sum d_x^2}} = 0,0006 \qquad 0,0018$$

$$Int_{1-\alpha}\left(\mathbb{E}\left(Y_*\right)\right) = \hat{y_*} \pm t_{1-\frac{\alpha}{2}}\left(n-k-1\right) s_{\hat{y}_*} = 25,67 \text{ +- } 2,31 * 0,0006 = (25,6686;\ 25,6714)$$

$$= 25,67 \text{ +- } 2,31 * 0,0017 * 0,0018 = (25,66999;\ 25,67001)$$

2024. április 15-i gyakorlat [1]

Kétváltozós lineáris és lineárisra visszavezethető regresszió

Az *elaszticitás* (rugalmasság) azt méri, hogy az X változó 1%-os növekedése hány százalékos növekedést/csökkenést eredményez az Y változónál. Az elaszticitás kiszámítása a becsült eredményváltozóra:

$$El(\hat{y}, x) = \frac{\partial \hat{y}}{\partial x} \cdot \frac{x}{\hat{y}}$$

Kétváltozós lineáris esetben:

$$El(\hat{y}, x) = \hat{\beta}_1 \cdot \frac{x}{\hat{y}} = \frac{\hat{\beta}_1 x}{\hat{\beta}_0 + \hat{\beta}_1 x}$$

1. Egy bank 10 ügyfelét vizsgálva az életkor (X, év) és a havi jövedelem (Y, eFt-ban) kapcsolatát elemzi. Az alábbi részeredményeket kapták:

$$\sum_{i=1}^{n} x_i = 260, \qquad \sum_{i=1}^{n} y_i = 2\,040, \qquad \sum_{i=1}^{n} x_i y_i = 53\,754,$$

$$\sum_{i=1}^{n} x_i^2 = 6\,862, \qquad \sum_{i=1}^{n} y_i^2 = 421\,866, \qquad \sum_{i=1}^{n} e_i^2 = 708,$$

$$\sum_{i=1}^{n} \ln x_i = 50, \qquad \sum_{i=1}^{n} \ln y_i = 90, \qquad \sum_{i=1}^{n} x_i \ln y_i = 675.$$

   (a) Határozza meg és értelmezze a lineáris regresszió paramétereit!

   (b) Számítsa ki a két változó lineáris korrelációs együtthatóját!

   (c) Számítsa ki és értelmezze a lineáris regresszió determinációs együtthatóját!

   (d) Adjon becslést a 30 évesek átlagos jövedelmére, majd szerkesszen konfidenciaintervallumot ugyanerre 98%-os megbízhatósági szinten!

   (e) Adjon 98%-os konfidenciaintervallumot egy 30 éves ügyfél egyedi jövedelmére!

   (f) Számolja ki és értelmezze az elaszticitást az átlagos életkorra!

2. Egy borkereskedő 10 óbor életkora (X) és ára (Y) közötti összefüggést vizsgálta. Az alábbi részeredményeket kapta:

$$\sum_{i=1}^{n} x_i = 200, \qquad \sum_{i=1}^{n} y_i = 280\,520, \qquad \sum_{i=1}^{n} x_i y_i = 4\,237\,000,$$

$$\sum_{i=1}^{n} x_i^2 = 2\,000, \qquad \sum_{i=1}^{n} y_i^2 = 7\,869\,152\,746, \qquad \sum_{i=1}^{n} e_i^2 = 708,$$

$$\sum_{i=1}^{n} \ln x_i = 20, \qquad \sum_{i=1}^{n} \ln y_i = 120, \qquad \sum_{i=1}^{n} x_i \ln y_i = 1\,900, \qquad \sum_{i=1}^{n} d_{y_i}^2 = 5\,706.$$

   Illesszen exponenciális ($Y = \beta_0 \beta_1^X \nu$) regressziófüggvényt az adatokra! (SPSS-ben: Compound)

   (a) Becsülje és értelmezze a β_0 és β_1 paramétereket!

   (b) Mennyibe kerül átlagosan egy 15 éves óbor? Mennyi idő alatt duplázódik átlagosan a bor ára?

   (c) Jellemezze a kapcsolat szorosságát egy alkalmas mutatóval!

[1] A feladatok Keresztély-Sugár-Szarvas: Statisztika közgazdászoknak Példatár és feladatgyűjteményből, továbbá korábbi ZH feladatokból származnak.

3. A Milliárdos Veteránok Klubja aukcióján árverésre került 19 veterán autó licitálási adatai alapján vizsgáljuk az autók kikiáltási ára (X, eFt) és leütési ára (Y, eFt) közötti sztochasztikus kapcsolatot. Ismertek az alábbi részeredmények:

$$\overline{x} = 837.368, \qquad \overline{\ln x} = 6.536, \qquad \overline{x \ln y} = 8\,645.798, \qquad \overline{\ln y} = 9.859,$$

$$\overline{\ln x \log y} = 65.127, \qquad \overline{(\ln x)^2} = 43.22, \qquad \overline{x^2} = 881\,952.632.$$

Illesszen hatványkitevős ($Y = \beta_0 X^{\beta_1} \nu$) regressziófüggvényt az adatokra! (SPSS-ben: Power)

(a) Becsülje és értelmezze a β_0 és β_1 paramétereket!

(b) Mekkora a várható leütési ára egy 500 eFt kikiáltású árú VW Typ3-asnak?

(c) Számolja ki és értelmezze a regressziófüggvény elaszticitását!

4. Az eszpresszó kávék hőmérséklet-változását vizsgálták az idő függvényében. Ennek érdekében 14 eszpresszót főztek, s különböző időpontokban megmérték az egyes kávék hőmérsékletét. A kapott eredményeket az alábbi táblázat tartalmazza:

Idő (perc) X	Hőmérséklet (°C) Y
1	82
5	76
8	70
11	65
15	61
18	57
22	52
25	51
30	47
34	45
38	43
42	41
45	39
50	38

Néhány számítási eredmény:

$$\sum_{i=1}^{14} x_i = 344, \qquad \sum_{i=1}^{14} x_i^2 = 11\,658 \qquad \sum_{i=1}^{14} \ln x_i = 40 \qquad \sum_{i=1}^{14} (\ln x_i)^2 = 129$$

$$\sum_{i=1}^{14} y_i = 767, \qquad \sum_{i=1}^{14} y_i^2 = 44\,629 \qquad \sum_{i=1}^{14} \ln y_i = 55.63 \qquad \sum_{i=1}^{14} (\ln y_i)^2 = 222$$

$$\sum_{i=1}^{14} x_i y_i = 16\,048, \qquad \sum_{i=1}^{14} x_i(\ln y_i) = 1\,316.19 \qquad \sum_{i=1}^{14} (\ln x_i) y_i = 2\,003.9 \qquad \sum_{i=1}^{14} \ln x_i \ln y_i = 155.76$$

(a) Írja fel az adatokra illeszkedő exponenciális regressziós modellt!

(b) Hány fokos a frissen főzött eszpresszó?

(c) 1 óra múlva várhatóan mennyi lesz egy csésze kávé hőmérséklete?

(d) 1992-ben egy nő beperelte a McDonald's gyorséttermet, ugyanis a kapott 82°C-os kávét magára borította, ami égési sérüléseket okozott. A szakértői vizsgálat szerint a 82°C-os kávé 2-7 másodperc alatt harmadfokú égési sérülést okoz az emberi bőrön. Megállapították, hogy 68°C -ra kell lehűteni a kávét, hogy elkerülhető legyen a súlyos égési sérülés. A nő 2,7 millió dollár kártérítést kapott. Ezen nevezetes eset kapcsán sok étterem 68°C-os kávét szolgál fel. Mennyi ideig kell várnia az étteremnek mielőtt felszolgálja a kávét (a kávéscsészébe töltés után), hogy az ne legyen melegebb 68°C -nál?

5. 1912-ben Tokió kormányzója 3 000 cseresznyefát ajándékozott Washington D.C.-nek a barátság jelképeként. A Nyugat-Potomac nemzeti parkban elültetett fák virágzása minden tavasszal rendkívüli látványosság és sok turistát csal a fővárosba. Ezen fák növekedési ütemét vizsgálták, s néhány véletlenszerűen kiválasztott fa magassága alapján 11 különböző időpontban megnézték a cseresznyefák átlagmagasságát. Elültetéskor mindegyik fa 1 éves volt és 6 láb magas (1 láb = 0.3048 m).

A fa kora (év) X	Átlagos magasság (láb) Y
1	6
2	9.5
2.5	10.5
4	15
5.5	17
6	17.5
7	18.5
8	19
8.5	19.5
10	19.7
11	19.8

Néhány számítási eredmény:

$$\sum_{i=1}^{11} x_i = 66, \qquad \sum_{i=1}^{11} x_i^2 = 500 \qquad \sum_{i=1}^{11} \ln x_i = 17 \qquad \sum_{i=1}^{11} (\ln x_i)^2 = 33$$

$$\sum_{i=1}^{11} y_i = 172, \qquad \sum_{i=1}^{11} y_i^2 = 2\,920 \qquad \sum_{i=1}^{11} \ln y_i = 29.6 \qquad \sum_{i=1}^{11} (\ln y_i)^2 = 81$$

$$\sum_{i=1}^{11} x_i y_i = 1\,171.8, \qquad \sum_{i=1}^{11} x_i(\ln y_i) = 187.64 \qquad \sum_{i=1}^{11} (\ln x_i) y_i = 307.42 \qquad \sum_{i=1}^{11} \ln x_i \ln y_i = 49.61$$

(a) Írja fel az adatokra illeszkedő hatvány regressziós modellt!

(b) Várhatóan milyen magas lesz egy 20 éves cseresznyefa? Reális-e az előrejelzés? Indokolja meg az állítását!

(c) Találtak egy régi fényképet, melyen a cseresznyefák átlagosan 10 láb magasak voltak. Hány évesek lehettek akkor a fák?

2024. április 22-i gyakorlat

Többváltozós lineáris regresszió

Egy orvosi vizsgálatnál valamely gyógyszer hatástartamát mérték percekben. Feljegyezték még a páciens életkorát, testsúlyát, lakóhelyét és vércukorszintjét. A `hatástartam.sav` adatállomány alapján illesszen lineáris regressziós modellt a Hatás változóra Backward eljárással!

1. Adja meg a legjobb modell illeszkedését leíró mutató értékét.
   Model Summary -> R Square: 0,846
2. Adja meg a legjobb modell illeszkedését leíró módosított mutató értékét.
   Model Summary -> Adjusted R Square: 0,838
3. Adja meg a legjobb modell többszörös determinációs együtthatójának az értékét.
   Model Summary -> R Square: 0,846
4. Adja meg a végső modellhez tartozó globális F-próba próbastatisztikájának értékét.
   ANOVA -> F: 101,687
5. Adja meg a függő változó és a vele legkevésbé korreláló magyarázó változó korrelációs együtthatóját.
   Correlations -> Pearson Correlation: 0,075 (lakóhely)
6. Adja meg a függő változó és a vele legjobban korreláló magyarázó változó korrelációs együtthatóját.
   Correlations -> Pearson Correlation: 0,889 (vércukor)
7. Melyik magyarázó változó hat a legerősebben a hatástartam értékére?
   vércukor
8. Melyik magyarázó változóval korrelál legerősebben az Életkor változó?
   Correlations -> Pearson Correlation: testsúly (0,338)
9. Mely magyarázó változó(k) marad(nak) ki a hatástartamot legjobban leíró lineáris modellből?
   Variables Entered/Removed -> Variables Removed: lakóhely, életkor
10. Mely magyarázó változóktól függ a hatástartamot legjobban leíró lineáris modell?
    Variables Entered/Removed -> vércukor, testsúly
11. Mely magyarázó változó(k)tól nem függ a hatástartamot legjobban leíró lineáris modell?
    Variables Entered/Removed -> Variables Removed: lakóhely, életkor
12. A végső modellből kihagyott magyarázó változók közül adja meg a legkevésbé szignifikáns parciális korrelációját.
    Excluded Variables -> Partial Correlation: 0,028 (lakóhely)
13. A végső modell esetén adja meg a konstans becsült standard hibáját.
    Coefficients -> Std. Error: 9,393
14. Adja meg a legjobban illeszkedő modellben a Vércukor együtthatójának standard hibáját.
    Coefficients -> Std. Error: 9,792
15. A végső modell esetén adja meg a konstans szignifikanciájára vonatkozó teszt próbastatisztikájának értékét.
    Coefficients -> t: 13,212
16. Adja meg a legjobban illeszkedő modellben a Vércukor együtthatójának szignifikanciájára vonatkozó teszt próbastatisztikájának értékét.
    Coefficients -> t: -10,051
17. Adjon 95%-os konfidenciaintervallumot a legjobb modell konstans tagjára.
    Coefficients -> 95,0% Confidence Interval for B: (105,071; 143,136)

18. Adjon 95%-os konfidenciaintervallumot a legjobb modellben a Vércukor együtthatójára.

Coefficients -> 95,0% Confidence Interval for B: (-15,339; -10,184)

19. A végső modellben szereplő magyarázó változók közül melyiknek a legkisebb a parciális korrelációs együtthatója?

Coefficients -> Correlations-partial: 0,514

20. Adja meg a végső modell reziduális és teljes négyzetösszegét.

ANOVA -> Sum of Squares-Residual/Total: 574,793 és 3734,188

21. Adja meg a legjobban illeszkedő modell maradékai normalitásának ellenőrzésére szolgáló teszt aszimptotikus p-értékét. Amennyiben az SPSS alsó határt ad meg, úgy azt írja be.

One Sample Kolmogorov-Szmirnov Test: 0,2

22. Adja meg a legjobban illeszkedő modell maradékai normalitásának ellenőrzésére szolgáló teszt próbastatisztikájának értékét.

One Sample Kolmogorov-Szmirnov Test: 0,067

23. 10%-os döntési szintet használva normálisnak tekinthetőek-e a végső modell maradékai?

Igen

24. Várhatóan hány óráig tart a vizsgált gyógyszer hatása egy olyan személynél, aki 70 kg, 40 éves, falun lakik és a vércukra 7 mmol/l?

43,688 perc

25. A illesztett modell alapján mennyivel csökken a vizsgált gyógyszer hatása, ha a páciens vércukra 1 mmol/l-rel nő, a többi értéke viszont változatlan marad?

12,838 perccel csökken

2024. április 29-i gyakorlat

Többváltozós lineáris regresszió

1. Nyissa meg a *Ketvaltozos_regresszio_lerablos.pdf* fájlt. Kiegészítő kérdések a feladathoz:

   (a) Adja meg a legjobb modell illeszkedését leíró mutató értékét.

   Árbevétel = 297,267 + 1,589 * Kisegítők száma Linear -> R Square: 0,966

   (b) Adja meg a legjobb modell illeszkedését leíró módosított mutató értékét.

   Linear -> Adjusted R Square: 0,961

   (c) A legjobb modell esetén adja meg a konstans becsült standard hibáját.

   Linear -> Coefficients -> Std. Error: 2,636

   (d) A legjobb modell esetén adja meg a konstans szignifikanciájára vonatkozó tesztpróbastatisztikájának értékét.

   Linear -> Coefficients -> t: 112,764

   (e) Adja meg a legjobb modell reziduális és teljes négyzetösszegét.

   ANOVA -> residual/total: 48,289 és 1411,556

   (f) Adja meg a legjobban illeszkedő modell maradékai normalitásának ellenőrzésére szolgáló teszt próbastatisztikájának értékét.

   0,576

2. Nyissa meg a *tobbvaltozos_regresszio_forward_bonsai.pdf* fájlt. Kiegészítő kérdések a feladat-hoz:

   (a) Adja meg a legjobb modell többszörös determinációs együtthatójának az értékét.

   0,923

   (b) Adja meg a legjobb modell többszörös korrelációs együtthatójának az értékét.

   0,961

   (c) Mely magyarázó változó(k) marad(nak) ki az árat legjobban leíró lineáris modellből?

   típus, alapterület

   (d) Mely magyarázó változóktól függ az árat legjobban leíró lineáris modell?

   távolság, átmérő, kor

   (e) Adjon 95%-os konfidenciaintervallumot a legjobb modellben a Távolság együtthatójára.

   (156,463; 194,638)

   (f) A végső modellből kizárt magyarázó változók közül melyiknek a legkisebb a parciális korrelációs együtthatója?

   alapterület: 0,042

   (g) Adja meg a végső modellhez tartozó globális F-próba próbastatisztikájának értékét.

   448,178

   (h) Adja meg a legjobban illeszkedő modell maradékai normalitásának ellenőrzésére szolgáló teszt próbastatisztikájának értékét.

   0,758

   (i) 10%-os döntési szintet használva normálisnak tekinthetőek-e a végső modell maradékai?

   normálisnak tekinthető

   (j) A illesztett modell alapján mennyi a különbség két bonsai ára között, ha az egyik lombkoronájának átmérője 1 cm-rel nagyobb, de azonos korúak és ugyanabból a kertészetből származnak?

   476,323

3. Nyissa meg a tobbvaltozos_regresszio_backward_porcelanvazak.pdf fájlt. Kiegészítő kérdések a feladathoz:

(a) Melyik magyarázó változó hat a legerősebben az ár értékére?

(b) Adja meg a függő változó és a vele legkevésbé korreláló magyarázó változó korrelációs együtthatóját.

(c) Adja meg a végső modell illeszkedését leíró módosított mutató értékét.

(d) Mely magyarázó változó(k)tól nem függ az árat legjobban leíró lineáris modell?

(e) A végső modellből kihagyott magyarázó változók közül adja meg a legkevésbé szignifikáns parciális korrelációját.

(f) Adjon 95%-os konfidenciaintervallumot a legjobb modell konstans tagjára.

(g) Adja meg a legjobban illeszkedő modellben az Idő együtthatójának szignifikanciájára vonatkozó teszt próbastatisztikájának értékét.

(a) Adja meg a legjobban illeszkedő modell maradékai normalitásának ellenőr zésére szolgáló teszt aszimptotikus p-értékét. Amennyiben az SPSS alsó határt ad meg, úgy azt írja be.

(b) A illesztett modell alapján mennyivel nő egy váza ára, ha a festésére 1 órával többet szánunk, a többi érték viszont változatlan marad?

2024. május 6-i gyakorlat

Idősorok

1. Az alábbi adatok egy vállalkozás bevételei (millió forint) 2005-ben és 2006-ban negyedéves bontásban:

$$20, \quad 24, \quad 24, \quad 25, \quad 28, \quad 31, \quad 32, \quad 33.$$

   (a) Illesszen lineáris trendet a bevételek idősorára! Értelmezze a kapott együtthatókat!

   (b) Adja meg és értelmezze a második negyedéves additív szezonális eltérést!

   (c) Mekkora az éves átlagos növekedés?

   (d) Mekkora bevétel jelezhető előre 2007 második negyedévére?

2. Az alábbi adatok ifj. Hörömpő Ödön szesz- és cigarettacsempész, valamint műbútorasztalos három heti bevételét mutatja (euróban) a hét munkanapjaira lebontva.

I.Hét	Bevétel	II. Hét	Bevétel	III. Hét	Bevétel
Hétfő	154	Hétfő	297	Hétfő	448
Kedd	182	Kedd	361	Kedd	477
Szerda	431	Szerda	626	Szerda	832
Csütörtök	252	Csütörtök	594	Csütörtök	895
Péntek	386	Péntek	653	Péntek	684

Ismertek továbbá a következőek:

$$\sum_{t=1}^{15} t = 120, \qquad \sum_{t=1}^{15} t^2 = 1240, \qquad \sum_{t=1}^{15} \log(t) = 27.8993, \qquad \sum_{t=1}^{15} \big(\log(t)\big)^2 = 60.4520,$$

$$\sum_{t=1}^{15} y_t = 7272, \qquad \sum_{t=1}^{15} \log(y_t) = 91.0037, \qquad \sum_{t=1}^{15} t y_t = 70200, \qquad \sum_{t=1}^{15} t \log(y_t) = 756.3849,$$

$$\sum_{t=1}^{15} \log(t)\log(y_t) = 174.3574, \qquad \sum_{t=1}^{15} y_t \log(t) = 15523.1980.$$

   (a) Illesszen exponenciális trendet a bevételek idősorára! Értelmezze a kapott együtthatókat!

   (b) Adja meg és értelmezze a csütörtöki és pénteki multiplikatív szezonális indexet!

   (c) Az adatfelvétel utolsó napján Ödön meglátott egy 1500 eurós ujjnyi vastag aranyláncot és úgy döntött, a következő hét péntekén az aznapi bevételből megveszi. Elegendő lesz-e az adatok alapján arra a napra előrejelzett bevétel? Ha igen, mennyi pénze marad még, ha nem, mennyivel kell kipótolnia a lánc megvásárlásához?

3. Az alábbi adatok egy vállalkozás bevételei (millió forint) 2005-ben és 2006-ban negyedéves bontásban:

$$22, \quad 22, \quad 22.5, \quad 26.6, \quad 32, \quad 32.4, \quad 32.5, \quad 39.$$

   (a) Illesszen exponenciális trendet a bevételek idősorára! Értelmezze a kapott együtthatókat!

   (b) Adja meg és értelmezze az első negyedéves multiplikatív szezonális indexet!

   (c) Mekkora az éves átlagos növekedés üteme?

   (d) Mekkora bevétel jelezhető előre 2008 első negyedévére?

4. Egy fagylaltárus forgalma (ezer gombócban) az alábbi volt:

	I.	II.	III.	IV.
Év	negyedév			
2005	208	574	456	184
2006	231	703	560	239
2007	284	874	710	216

Néhány számolási eredmény:

$$\sum_{t=1}^{12} y_t = 5\,239, \quad \sum_{t=1}^{12} t y_t = 36\,363, \quad \sum_{t=1}^{12} t = 78, \quad \sum_{t=1}^{12} t^2 = 650,$$

$$\sum_{t=1}^{12} \ln y_t = 71.19, \quad \sum_{t=1}^{12} \ln t = 20, \quad \sum_{t=1}^{12} t \ln y_t = 467.21, \quad \sum_{t=1}^{12} \ln^2 t = 39.57.$$

(a) Határozza meg a lineáris trend egyenletét és értelmezze a paramétereket!

(b) Határozza meg a nyers szezonális indexeket, majd a II. negyedévhez tartozó korrigált (tisztított) szezonindexet és értelmezze az utóbbit!

(c) Becsülje meg a forgalmat 2009 II. negyedévére!

5. Az alábbi adatok Pálpusztai-Trappista Elvira humán szolgáltatóipari kisvállalkozó három heti bevételét mutatja (euróban) a hét munkanapjaira lebontva.

I.Hét	Bevétel	II. Hét	Bevétel	III. Hét	Bevétel
Hétfő	70	Hétfő	244	Hétfő	471
Kedd	171	Kedd	382	Kedd	539
Szerda	462	Szerda	599	Szerda	738
Csütörtök	243	Csütörtök	536	Csütörtök	772
Péntek	400	Péntek	646	Péntek	656

Ismertek továbbá a következőek:

$$\sum_{t=1}^{15} t = 120, \quad \sum_{t=1}^{15} t^2 = 1240, \quad \sum_{t=1}^{15} \log(t) = 27.8993, \quad \sum_{t=1}^{15} \big(\log(t)\big)^2 = 60.4520,$$

$$\sum_{t=1}^{15} y_t = 6929, \quad \sum_{t=1}^{15} \log(y_t) = 89.7867, \quad \sum_{t=1}^{15} t y_t = 66875, \quad \sum_{t=1}^{15} t \log(y_t) = 751.3058,$$

$$\sum_{t=1}^{15} \log(t)\log(y_t) = 173.4258, \quad \sum_{t=1}^{15} y_t \log(t) = 14873.5657.$$

(a) Illesszen exponenciális trendet a bevételek idősorára! Értelmezze a kapott együtthatókat!

(b) Adja meg és értelmezze a csütörtöki és pénteki additív szezonális ingadozást!

(c) Az adatfelvétel utolsó napján Elvira meglátott egy 1465 eurós gyémánt nyakéket és úgy döntött, a következő hét csütörtökén az aznapi bevételből megveszi. Elegendő lesz-e az adatok alapján arra a napra előrejelzett bevétel? Ha igen, mennyi pénze marad még, ha nem, mennyivel kell kipótolnia a bevételt a nyakék megvásárlásához?