

אקונומטריקה א



תוכן העניינים

1.	רווח סמך לשונות וסטיית תקן	1
2.	רווח סמך ליחס שונות	6
3.	בדיקת השערות על שונות	10
4.	ניתוח שונות חד גורמי (ללא ספר)	21
5.	מבוא לקורס	27
6.	אומדי הריבועים הפחותים	35
7.	מודלים לא ליניאריים	39
8.	מבחני המובהקות וקריאת פלטים - תוכנת SAS	46
9.	שינוי יחידות מדידה	48
10.	הרגרסיה המרובה	57
11.	משתנה דמי	74
12.	מבחן 1	78
13.	מבחן 2	83
14.	מבחן 3	89
15.	מבחן 4	95
16.	מבחן 5	101
17.	שאלות חזרה למבחן מבוססות על תוכנת ATATS	

אקונומטריקה א

פרק 1 - רווח סמך לשונות וסטיית תקן

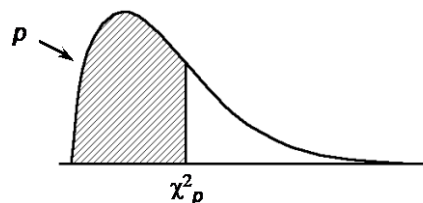
תוכן העניינים

1. רווח סמך לשונות וסטיית תקן.....1

רווח סמך לשונות וסטיית תקן:

רקע:

בפרק זה נדון על בניית רווח סמך לשונות האוכלוסייה. התנאי לבניית רווח הסמך: המשתנה הנחקר מתפלג נורמלית, למרות שנהוג לא לדרוש את התנאי הזה אם המדגם מספיק גדול. רווח הסמך יתבסס על התפלגות הנקראת חי בריבוע. התפלגות זו היא התפלגות אסימטרית חיובית המתחילה מהערך אפס ותלויה בדרגות חופש. דרגות החופש במקרה זה יהיו: $n-1$.



$$\text{רווח הסמך לשונות: } \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}}$$

$$\text{כאשר: } S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2}{n-1}$$

אם נרצה לבנות רווח סמך לסטיית תקן אז נוציא שורש לרווח סמך לשונות.

דוגמה:

זמן התגובה מתפלג נורמלית. במטרה לאמוד את שונות זמן התגובה נדגמו 4 תצפיות. להלן התוצאות בשניות: 4.7, 5.2, 4.6, 5.3. בנו רווח סמך, ברמת סמך של 95%, לשונות זמן התגובה באוכלוסייה.

פתרון:

פרמטר: σ^2 .

$$X \sim N(\mu, \sigma^2) = \text{זמן תגובה (בשניות)}.$$

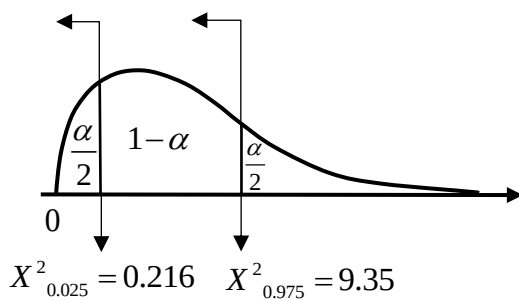
תוצאות מדגם: $n = 4$.

$$\bar{X} = \frac{4.7+5.2+4.6+5.3}{4} = 4.95$$

$$d.f = n-1 = 4-1 = 3$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - n \cdot \bar{X}^2}{n-1} \quad \text{נציב:}$$

$$S^2 = \frac{4.7^2 + 5.2^2 + \dots - 4 \cdot 4.95^2}{4-1} = 0.123$$



$$1-\alpha = 0.95$$

$$\alpha = 0.05$$

$$\frac{\alpha}{2} = 0.025$$

(טבלת התפלגות חי-בריבוע מופיעה בעמוד האחרון).

$$\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}, n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}, n-1}} \quad \text{נציב:}$$

$$\frac{(4-1) \cdot 0.123}{9.35} < \sigma^2 < \frac{(4-1) \cdot 0.123}{0.216}$$

$$0.039 < \sigma^2 < 1.708$$

שאלות:

(1) חמישה מטופלים קבלו תרופה מסוימת. בדקו לכל מטופל את זמני התגובה שלו. להלן הזמנים שהתקבלו בדקות: 18, 17, 21, 26, 28. בהנחה וזמני התגובה מתפלגים נורמאלית, בנו רווח סמך ברמת סמך של 95% לשונות זמן התגובה.

(2) נדגמו 20 ימים אקראיים מחודשי יולי-אוגוסט ונמדדה בהם הטמפ' במעלות צלזיוס בת"א. במדגם התקבל טמפ' ממוצעת 30.8 וסטיית תקן מדגמית 1.1. בהנחה והטמפ' מתפלגת נורמאלית:

א. בנו רווח סמך לתוחלת הטמפ' בחודשים אלה בת"א ברמת סמך של 95%.

ב. בנו רווח סמך לסטיית התקן של הטמפ' בחודשים אלה בת"א ברמת סמך של 95%.

(3) ציוני IQ בארה"ב מתפלגים נורמאלית עם ממוצע 100 וסטיית תקן 5. נבחנו 20 נבחנים ישראלים במבחן ה-IQ.

$$\sum_{i=1}^{20} X_i = 2080, \quad \sum_{i=1}^{20} X_i^2 = 218,220$$

להלן התוצאות שהתקבלו:

נניח שגם בישראל הציונים מתפלגים נורמאלית.

- א. מצאו אומדנים לממוצע הציונים בישראל ולשונות הציונים בישראל באמצעות אומדנים חסרי הטיה.
- ב. אמדו ברמת ביטחון של 95% את תוחלת הציונים של נבחנים בישראל.
- ג. אמדו ברמת סמך של 90% את סטיית התקן של הציונים של נבחנים ישראלים.
- ד. על סמך הסעיפים הקודמים, האם בישראל ממוצע הציונים וסטיית התקן של הציונים שונה מבארה"ב? הסבירו.

(4) באוכלוסייה מסוימת נדגמו 10 תצפיות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{10} X_i = 750, \quad \sum_{i=1}^{10} (X_i - \bar{X})^2 = 900$$

$$X_i \sim N(\mu, \sigma^2)$$

נתון ש:

- א. בנו רווח סמך ל- μ ברמת סמך של 95%.
- ב. בנו רווח סמך ל- σ^2 ברמת סמך של 95%.

תשובות סופיות:

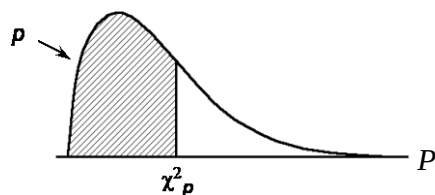
(1) $8.4 < \sigma^2 < 194.2$

(2) א. $30.285 < \mu < 31.315$ ב. $0.836 < \sigma < 1.606$

(3) א. ממוצע: 104, שונות: 100. ב. $99.32 \leq \mu \leq 108.68$ ג. $7.94 < \sigma < 13.7$

ד. בביטחון של 95% ממוצע הציונים איננו שונה, ובביטחון של 90% סטיית התקן שונה.

(4) א. $68.75 < \mu < 82.15$ ב. $47.4 < \sigma^2 < 333.3$

נספח - טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p :


df	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.00157	0.00982	0.02393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

אקונומטריקה א

פרק 2 - רווח סמך ליחס שונויות

תוכן העניינים

1. רווח סמך ליחס שונויות 6

רווח סמך ליחס שונויות:

רקע:

נרצה לאמוד את ההבדל בין שתי שונויות משתי אוכלוסיות שונות.

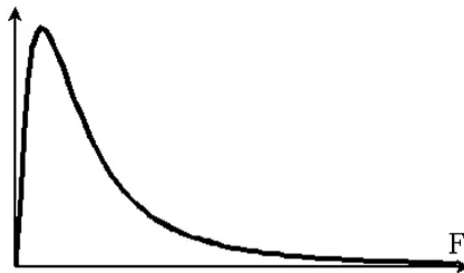
הפרמטר יהיה: $\frac{\sigma_1^2}{\sigma_2^2}$: כלומר היחס בין השונויות.

התנאים:

• $X_1, X_2 \sim N$ או מדגמים גדולים.

• מדגמים בלתי תלויים.

רווח הסמך יבנה על סמך התפלגות הנקראת התפלגות F. התפלגות זו היא אסימטרית חיובית ומושפעת משתי דרגות החופש, זו של המונה וזו של המכנה.



$$df_1 = n_1 - 1$$

$$df_2 = n_2 - 1$$

$$\frac{s_1^2}{s_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} F_{1-\frac{\alpha}{2}}(n_2-1, n_1-1) \quad \text{רווח הסמך יהיה:}$$

דוגמה (פתרון בהקלטה):

מחקר סוציולוגי מעוניין לחקור את הרגלי הבילויים בקבוצות גיל שונות: במדגם שנעשה על סטודנטים בגילאי 21-26 התקבל אומדן חוסר הטיה לשונויות ההוצאה החודשית על בילויים 10,000. כמות הסטודנטים שנדגמו 16. במדגם שנעשה על 11 מבוגרים בשנות השלושים התקבל אומדן חוסר הטיה לשונויות ההוצאה החודשית על בילויים 490,000. נניח שההוצאה החודשית לבילוי בכל קבוצת גיל מתפלג נורמאלית. בנו רווח סמך ברמת סמך של 95% ליחס בין השונויות.

שאלות:

- (1) בתחום הבינוי משתמשים בשני סוגי מתכות: מתכת A ומתכת B. מחקר מעוניין לבדוק האם קיים הבדל בין שני סוגי המתכות מבחינת שונויות החוזק שלהן. דגמו מספר יחידות מתכת מכל סוג והתקבלו התוצאות הבאות:

סוג המתכת	A	B
N	8	10
$\sum X_i$	16	30
$\sum X_i^2$	60	198

- יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.
- א. בנו רווח סמך ליחס השונויות של רמות החוזק בין שני סוגי המתכות ברמת סמך של 90%.
- ב. בנו רווח סמך ליחס סטיות התקן של רמות החוזק בין שני סוגי המתכות ברמת סמך של 90%.

- (2) מעוניינים להשוות בין נשים וגברים מבחינת השונויות בזמנים שלהם לבצע משימה מסוימת. במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמני ביצוע המשימה: $\sum (y_i - \bar{y})^2 = 204$. במדגם של 13 נשים התקבלו התוצאות הבאות: $\sum (x_i - \bar{x})^2 = 200$.
- אמוד ברמת ביטחון של 95% פי כמה גדולה השונויות של הגברים באוכלוסייה מהשונויות של הנשים.
- מה יש להניח לצורך פתרון?

תשובות סופיות:

$$(1) \quad \text{א. } 0.1013 < \frac{\sigma_A^2}{\sigma_B^2} < 1.2267 \quad \text{ב. } 0.318 < \frac{\sigma_A}{\sigma_B} < 1.108$$

$$(2) \quad 0.39 \leq \frac{\sigma_m^2}{\sigma_F^2} \leq 5.26$$

טבלת התפלגות F לפי זנב ימני של α

α	$df.2$	$df.1$											
		1	2	3	4	5	6	7	8	9	10	12	15
.05	1	161	200	216	225	230	234	237	239	241	242	244	246
.025		648	800	864	900	922	937	948	957	963	969	977	985
.01		4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157
.005		16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630
.05	2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43
.025		38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43
.01		98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.42	99.43
.005		198.50	199.01	199.16	199.24	199.30	199.33	199.36	199.38	199.39	199.39	199.42	199.43
.05	3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
.025		17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25
.01		34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05	26.87
.005		55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.68	43.39	43.08
.05	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
.025		12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66
.01		21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20
.005		31.33	26.28	24.26	23.15	22.46	21.98	21.62	21.35	21.14	20.97	20.70	20.44
.05	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
.025		10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43
.01		16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72
.005		22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15
.05	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
.025		8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27
.01		13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56
.005		18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81
.05	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
.025		8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57
.01		12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31
.005		16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97
.05	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
.025		7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10
.01		11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52
.005		14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81

$d.f.1.$													
α	$d.f.2.$	1	2	3	4	5	6	7	8	9	10	12	15
.05	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
.025		7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77
.01		10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96
.005		13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.23	6.03
.05	10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
.025		6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52
.01		10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56
.005		12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.66	5.47
.05	12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
.025		6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18
.01		9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01
.005		11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.91	4.72
.05	15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40
.025		6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86
.01		8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52
.005		10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.25	4.07
.05	20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20
.025		5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57
.01		8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09
.005		9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.68	3.50
.05	30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01
.025		5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31
.01		7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70
.005		9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.18	3.01
.05	60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84
.025		5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.17	2.06
.01		7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35
.005		8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.74	2.57
.05	120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75
.025		5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.05	1.94
.01		6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19
.005		8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.54	2.37
.05	∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67
.025		5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83
.01		6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04
.005		7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.36	2.19

אקונומטריקה א

פרק 3 - בדיקת השערות על שונויות

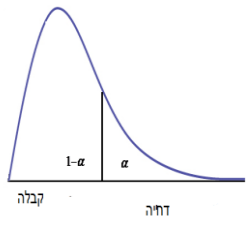
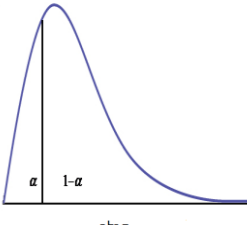
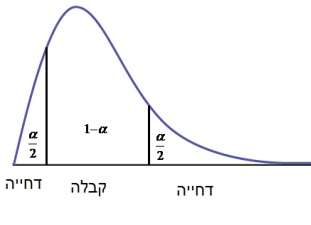
תוכן העניינים

- 1. בדיקת השערות על שונות וסטיית תקן 10
- 2. בדיקת השערות על שתי שונויות 15

בדיקת השערות על שונות וסטיית תקן:

רקע:

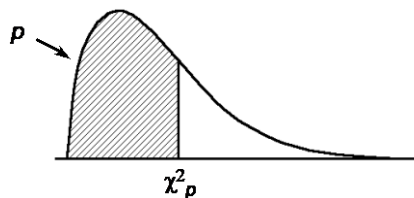
בדיקת השערות על שונות האוכלוסייה כאשר התוחלת לא ידועה:

השערת האפס: השערת אלטרנטיבית:	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$
תנאים:	$X \sim N$		
נדחה את השערת האפס אם:	 $\chi^2 > \chi_{1-\alpha}^{2(n-1)}$	 $\chi^2 < \chi_{\alpha}^{2(n-1)}$	 או $\chi^2 < \chi_{\frac{\alpha}{2}}^{2(n-1)} \text{ או } \chi^2 > \chi_{1-\frac{\alpha}{2}}^{2(n-1)}$

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \quad \text{סטטיסטי המבחן:}$$

התפלגות חי בריבוע:

$$\frac{(n-1)S^2}{\sigma_0^2} \sim \chi^{2(n-1)} \quad \text{אם } X_i \sim N(\mu, \sigma^2), \text{ והפרמטר } \mu \text{ אינו ידוע, מתקיים ש:}$$



התפלגות זו היא התפלגות אסימטרית חיובית המתחילה מהערך אפס וערכיה שואפים לאינסוף.

התפלגות זו תלויה בדרגות החופש.

אם μ אינו ידוע, אז: $d.f = n - 1$.

דוגמה:

ציוני IQ לפי סטנדרטים אמריקאים מתפלגים נורמאלית עם $\sigma = 15$. מעוניינים לבדוק האם שונות הציונים של נבחנים ישראלים שונה מאמריקה.

$$\sum_{i=1}^{20} (x_i - \bar{x})^2 = 3420 \quad \text{במדגם של 20 ישראלים התקבל:}$$

מה המסקנה ברמת מובהקות של 5%?

פתרון:

האוכלוסייה: נבחנים ישראלים במבחן I.Q.

המשתנה: $X =$ ציון I.Q.

פרמטר: σ^2

$$H_0: \sigma^2 = 15^2 = 225$$

השערות: $H_1: \sigma^2 \neq 225$

הנחה: $X \sim N$

כלל הכרעה: $d.f = n - 1 = 20 - 1 = 19$

נדחה את H_0 אם $X^2 > 32.9$ או $X^2 < 8.91$

תוצאות המדגם: $n = 20$

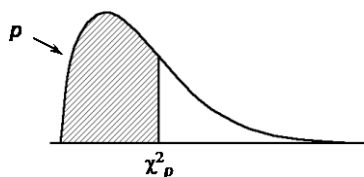
$$\sum (X_i - \bar{X})^2 = 3420$$

$$S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{3420}{20-1} = 180$$

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{19 \cdot 180}{225} = 15.2$$

מסקנה: לא נדחה את H_0 , לא נסיק ששונות הציונים של נבחנים ישראלים במבחן I.Q.

שונה מזו של אמריקאים. ($\alpha = 5\%$)

טבלת התפלגות חי-בריבוע – ערכי החלוקה χ^2_p


df	p												
	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.004393	0.00157	0.00982	0.02393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

שאלות:

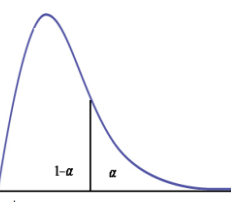
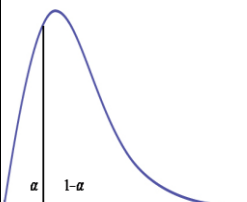
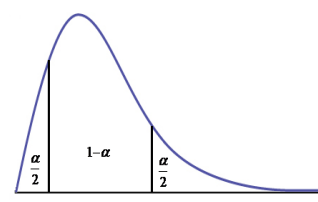
- (1) חברה אורזת סוכר במשקל עם סטיית תקן 20 גרם. משקל הסוכר באריזה מתפלג נורמאלית. החברה החליפה את מכוונת האריזה במטרה לדייק יותר במשקל הנארוז. (רוצים שסטיית התקן תהיה קטנה יותר).
לצורך בדיקה דגמו 5 אריזות סוכר ולהלן משקלן (בגרם):
1008, 1024, 1005, 996, 997.
מה המסקנה ברמת מובהקות של 5%?
- (2) זמן ההחלמה ממחלה מסוימת כאשר משתמשים בטיפול מסוים מתפלג נורמלית עם סטיית תקן של 80 שעות. תרופה חדשה נוסתה על 5 חולים. זמני ההחלמה שלהם בשעות היו: 110, 90, 72, 50, 38.
א. ברמת מובהקות של 5% בדקו האם סטיית התקן של זמן החלמה של התרופה החדשה נמוכה מהתרופה המקורית?
ב. האם ניתן לדעת מה תהיה התשובה לסעיף א', אם נגדיל את רמת המובהקות?
ג. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נקטין את רמת המובהקות?
ד. האם ניתן לדעת מה תהיה התשובה לסעיף א אם נוסיף תצפית שערכה 70?
- (3) הגובה של אוכלוסייה מסוימת נחשב כמתפלג נורמלית עם ממוצע של 174 ס"מ וסטיית תקן 12. במדגם של 20 אנשים מהאוכלוסייה התקבל ממוצע 171 וסטיית תקן מדגמית 23.
א. בדקו ברמת מובהקות של 5% האם חל שינוי בשונות הגבהים באוכלוסייה.
ב. בדקו ברמת מובהקות של 5%, האם חל שינוי בתוחלת הגבהים באוכלוסייה, בבחירת המבחן המתאים הסתמך על המסקנה מסעיף א'.
- (4) השערות המחקר הן: $H_0: \sigma^2 = 100$ $H_1: \sigma^2 > 100$
מתכננים לבצע מדגם בגודל 10 תצפיות. רמת המובהקות היא 5%.
א. מה תהיה עוצמת המבחן אם $\sigma_1^2 = 150$?
ב. איזו השערה אלטרנטיבית תיתן עוצמה של 90%?
- (5) השערות המחקר הן: $H_0: \sigma = 2$, $H_1: \sigma < 2$.
במדגם של 21 תצפיות התקבל סטיית תקן 1.143.
תנו הערכה למובהקות התוצאה.

תשובות סופיות:

- (1) לא נדחה H_0 .
- (2) א. נדחה את H_0 .
ב. לא תשתנה.
ג. לא ניתן לדעת.
ד. לא תשתנה.
- (3) א. נדחה את H_0 .
ב. לא נדחה את H_0 .
- (4) א. בין 25% ל-50%.
ב. 405.3.
- (5) $0 < P_V < 0.005$.

בדיקת השערות על שתי שוניות:

רקע:

השערת האפס : השערה אלטרנטיבית :	$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ $H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1$	$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ $H_1 : \frac{\sigma_1^2}{\sigma_2^2} < 1$	$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$ $H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$
תנאים :	1. מדגמים בלתי תלויים 2. $X_1, X_2 \sim N$		
נדחה את השערת האפס אם :	 קבלה דחייה $F \geq f_{1-\alpha}^{(n_1-1, n_2-1)}$	 דחייה קבלה $F \leq \frac{1}{f_{1-\alpha}^{(n_2-1, n_1-1)}}$	 דחייה קבלה דחייה $F \geq f_{1-\alpha/2}^{(n_1-1, n_2-1)}$ $F \leq \frac{1}{f_{1-\alpha/2}^{(n_2-1, n_1-1)}}$

$$F = \frac{S_1^2}{S_2^2} : \text{סטטיסטי המבחן}$$

התפלגות F:

$$\frac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1) \quad \text{אם } X_1 \sim N(\mu_1, \sigma^2) \text{ ו- } X_2 \sim N(\mu_2, \sigma^2) \text{ אזי}$$

התפלגות F הינה התפלגות אסימטרית חיובית התלויה בדרגות חופש של המונה ושל המכנה.

$$\text{כמו כן בהתפלגות F מתקיימת התכונה הבאה : } F_{1-\alpha}(n_1-1, n_2-1) = \frac{1}{F_{\alpha}(n_2-1, n_1-1)}$$



$$df_1 = n_1 - 1$$

$$df_2 = n_2 - 1$$

דוגמה:

מעוניינים להשוות בין נשים וגברים מבחינת השונות בזמנים שלהם לבצע משימה מסוימת. במדגם של 10 גברים התקבלו התוצאות הבאות לגבי זמן ביצוע המשימה: $\sum (y_i - \bar{y})^2 = 204$.

במדגם של 13 נשים התקבלו התוצאות הבאות: $\sum (x_i - \bar{x})^2 = 200$. בדקו ברמת מובהקות של 2% האם קיים הבדל בין השונות? מה יש להניח?

פתרון:

האוכלוסיות: נשים מול גברים.

משתנה: $y =$ זמן ביצוע משימה של גבר, $x =$ זמן ביצוע משימה של אישה

$$\frac{\sigma_x^2}{\sigma_y^2} : \text{פרמטר}$$

$$H_0 : \frac{\sigma_x^2}{\sigma_y^2} = 1$$

השערות:

$$H_1 : \frac{\sigma_x^2}{\sigma_y^2} \neq 1$$

הנחות: 1. מדגימים ב"ת 2. $x, y \sim N$

כלל הכרעה:

$$\alpha = 2\%$$

$$n_1 = 13, d.f_1 = n_1 - 1 = 12$$

$$n_2 = 10, d.f_2 = n_2 - 1 = 9$$

נדחה את H_0 אם $F > 5.11$ או $F < 0.23$

$$S_y^2 = \frac{204}{10-1} = 22\frac{2}{3} : \text{תוצאות המדגם}$$

$$S_x^2 = \frac{200}{13-1} = 16\frac{2}{3}$$

$$F = \frac{S_1^2}{S_2^2} = \frac{16\frac{2}{3}}{22\frac{2}{3}} : \text{סטטיסטי המבחן}$$

מסקנה: ברמת מובהקות של 2% נקבל את H_0 .

לא קיים הבדל מובהק בין גברים לנשים מבחינת השונות שלהם.

טבלת ערכים קריטיים לפי התפלגות F																	
α = 0.05																	
ד"ח מנהל"ח מנבה	1	2	3	4	5	6	7	8	9	10	12	16	20	24	60	120	∞
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	246.46	248.01	249.05	252.20	253.25	254.31
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.69	8.66	8.64	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.84	5.80	5.77	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.60	4.56	4.53	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.92	3.87	3.84	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.49	3.44	3.41	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.20	3.15	3.12	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	2.99	2.94	2.90	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.83	2.77	2.74	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.70	2.65	2.61	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.60	2.54	2.51	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.51	2.46	2.42	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.44	2.39	2.35	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.38	2.33	2.29	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.33	2.28	2.24	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.29	2.23	2.19	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.25	2.19	2.15	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.21	2.16	2.11	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.18	2.12	2.08	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.16	2.10	2.05	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.13	2.07	2.03	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.11	2.05	2.01	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.09	2.03	1.98	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.07	2.01	1.96	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.05	1.99	1.95	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.04	1.97	1.93	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.02	1.96	1.91	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.01	1.94	1.90	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	1.99	1.93	1.89	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.90	1.84	1.79	1.64	1.58	1.51
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.95	1.85	1.78	1.74	1.58	1.51	1.44
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.82	1.75	1.70	1.53	1.47	1.39
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.86	1.76	1.69	1.64	1.46	1.39	1.30
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.73	1.66	1.61	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.64	1.57	1.52	1.32	1.22	1.00

טבלת ערכים קריטיים לפי התפלגות F $\alpha = 0.01$ ראה איור מטה.																	
ד"ח מונה/ד"ח מכנה	1	2	3	4	5	6	7	8	9	10	12	16	20	24	60	120	∞
1	4052.18	4999.50	5403.35	5624.58	5763.65	5858.99	5928.36	5981.07	6022.47	6055.85	6106.32	6170.10	6208.73	6234.63	6313.03	6339.39	6365.86
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.44	99.45	99.46	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.83	26.69	26.60	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.15	14.02	13.93	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.68	9.55	9.47	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.52	7.40	7.31	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.28	6.16	6.07	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.48	5.36	5.28	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.92	4.81	4.73	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.52	4.41	4.33	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.21	4.10	4.02	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	3.97	3.86	3.78	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.78	3.66	3.59	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.62	3.51	3.43	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.49	3.37	3.29	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.37	3.26	3.18	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.27	3.16	3.08	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.19	3.08	3.00	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.12	3.00	2.92	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.05	2.94	2.86	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	2.99	2.88	2.80	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.94	2.83	2.75	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.89	2.78	2.70	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.85	2.74	2.66	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.81	2.70	2.62	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.78	2.66	2.58	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.75	2.63	2.55	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.72	2.60	2.52	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.69	2.57	2.49	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.66	2.55	2.47	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.48	2.37	2.29	2.02	1.92	1.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.56	2.38	2.27	2.18	1.91	1.80	1.68
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.31	2.20	2.12	1.84	1.73	1.60
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.39	2.21	2.09	2.00	1.72	1.60	1.46
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.15	2.03	1.95	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.00	1.88	1.79	1.47	1.32	1.00

שאלות:

- (1) להלן נתונים על שטחי דירות במ"ר עבור דירות חדשות שנבנו בשנת 2012 ובשנת 2013:

120	94	90	130	95	112	120	2012
	69	74	105	91	82	100	2013

- א. בדקו ברמת מובהקות של 10% את ההשערה ששונות שטחי הדירות החדשות בשנת 2012 ובשנת 2013 שוות. מה הן ההנחות הדרושות לביצוע הבדיקה?
 ב. האם וכיצד הייתה משתנה המסקנה מהסעיף הקודם אם מסתבר שחלה טעות ברישום ויש להפחית 10 מ"ר מכל הדירות שמופיעות במדגם?

- (2) בתחום הבינוי משתמשים בשני סוגי מתכות: מתכת A ומתכת B. מחקר מעוניין לבדוק האם קיים הבדל בין שני סוגי המתכות מבחינת החוזק שלהן. דגמו מספר

B	A	סוג המתכת
10	8	n
30	16	$\sum X_i$
198	60	$\sum X_i^2$

- יחידות מתכת מכל סוג והתקבלו התוצאות הבאות:
 יש להניח שרמת החוזק של המתכות מתפלגת נורמאלית.
 א. האם קיים הבדל בין שונות החוזק של מתכות?
 ב. האם קיים הבדל בין תוחלות החוזק של מתכות?
 בכל סעיף רמת מובהקות של 10%.

- (3) מחקר סוציולוגי מעוניין לחקור את הרגלי הבילויים בקבוצות גיל שונות. ידוע כי בקרב האוכלוסייה הבוגרת (מעל 18) ההוצאה החודשית על בילויים מתפלגת נורמאלית עם תוחלת של 500 ₪ וסטיית תקן של 300 ₪.
 במדגם שנעשה על סטודנטים בגילאי 21-26 התקבל אומדן חוסר הטיה לשונות ההוצאה החודשית על בילויים של 10,000 ₪. כמות הסטודנטים שנדגמה 16.
 במדגם שנעשה על 11 מבוגרים בשנות השלושים התקבל אומדן חסר הטיה לשונות ההוצאה החודשית על בילויים של 490,000 ₪.
 א. בדקו ברמת מובהקות של 5% האם שונות ההוצאה על בילויים בקרב סטודנטים בקבוצת גילאי 21-26 נמוכה מהשונות אצל כלל המבוגרים.
 ב. בדקו ברמת מובהקות של 1% האם הפיזור של ההוצאה החודשית לבילויים גדולה יותר בקבוצת גיל ה-30 מאשר בקבוצת גיל 21-26.

- (4) נתון $X_i \sim N(\mu_x, \sigma^2)$, וכמו כן $Y_i \sim N(\mu_y, \sigma^2)$.
 מאוכלוסייה X נדגמו 7 תצפיות ומאוכלוסייה Y נדגמו 13 תצפיות.

א. כיצד $\frac{S_x^2}{S_y^2}$ מתפלג?

- ב. מה ההסתברות ש- S_x^2 גדולה ביותר מפי 3 מאשר S_y^2 ?

תשובות סופיות:

- (1) א. לא נדחה את H_0 . ב. מסקנה לא תשתנה.
- (2) א. לא נדחה את H_0 . ב. לא נדחה את H_0 .
- (3) א. נדחה את H_0 . ב. נדחה את H_0 .
- (4) א. $F(6,12)$. ב. 5%.

אקונומטריקה א

פרק 4 - ניתוח שונות חד גורמי

תוכן העניינים

1. כללי (ללא ספר)

אקונומטריקה א

פרק 5 - מבוא לקורס

תוכן העניינים

1. כללי 21

מבוא לקורס:

רקע:

הגדרות וסימונים:

משתנה אמפירי – תוצאותיו ידועות מראש (למשל: רמת הכנסה, גיל, מס' שנות לימוד במדגם מסוים).

משתנה מקרי – תוצאותיו לא ידועות מראש (כגון תוצאה בהטלת קובייה או בהטלת מטבע). באקונומטריקה נעסוק בעיקר במשתנים מקריים.

שני סוגי המשתנים יסומנו באות לועזית עם אינדקס (למשל: X_t או Y_t).

קבוע – מקבל ערך אחד בלבד (מסומן באות לועזית ללא אינדקס – למשל a או b).

לכל משתנה מקרי X_t יש תוחלת המייצגת את מרכז ההתפלגות (μ_x או $E(X)$).

השונות – מייצגת את מידת הפיזור של ההתפלגות (σ_x^2 או $V(X)$).

סטית התקן – היא השורש של השונות (σ_x).

שונות משותפת (covariance) – מדד להתפלגות המשותפת של שני משתנים מקריים ומייצגת את הכיוון של הקשר ביניהם ($\text{Cov}(X, Y)$):

$$X, Y \Leftrightarrow \text{Cov}(X, Y) = 0 \text{ בלתי מתואמים.}$$

$$\text{Cov}(X, Y) > 0 \Leftrightarrow \text{מתאם חיובי בין המשתנים.}$$

$$\text{Cov}(X, Y) < 0 \Leftrightarrow \text{מתאם שלילי בין המשתנים.}$$

$$X, Y \Leftarrow \text{בלתי תלויים } X, Y \Leftarrow \text{בלתי מתואמים.}$$

מקדם המתאם של פירסון – מדד לכיוון ולעוצמת הקשר הליניארי בין שני

$$\text{משתנים: } \eta_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

$$-1 \leq \eta \leq 1$$

$$\eta = 1 \text{ מתאם ליניארי חיובי מלא בין שני המשתנים.}$$

$$\eta = -1 \text{ מתאם ליניארי שלילי מלא בין שני המשתנים.}$$

$$\eta = 0 \text{ לא קיים מתאם ליניארי בין שני המשתנים.}$$

אמידה:

פרמטר – ערך המשתנה הנחקר המתאר את כל האוכלוסייה.
סטטיסטי/אומד – ערך המשתנה הנחקר המתאר את המדגם.

מדגם	אוכלוסייה
$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$	$E(X) = \mu$
$S_X^2 = \frac{S_{XX}}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$	$V(X) = \sigma^2 = E(X - E(X))^2$
$\frac{S_{XY}}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$	$\text{cov}(X, Y) = E(X - E(X))(Y - E(Y))$
$r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}} \sqrt{S_{YY}}}$	$\eta_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{V(X)} \sqrt{V(Y)}}$

נוסחאות וחוקים בסטטיסטיקה:

יהיו X ו- Y משתנים מקריים, ו- a, b קבועים:

חוקי הסיגמה:

$$1. \sum_{t=1}^T X_t = X_1 + X_2 + X_3 + \dots + X_T$$

$$2. \sum_{t=1}^T a = Ta \quad \text{סכום של קבוע:}$$

$$3. \sum_{t=1}^T aX_t = a \sum_{t=1}^T X_t \quad \text{סכום של קבוע כפול משתנה = לקבוע כפול הסכום:}$$

$$4. \sum_{t=1}^T (X_t \pm Y_t) = \sum_{t=1}^T X_t \pm \sum_{t=1}^T Y_t \quad \text{סכום של סכום/הפרש = לסכום/הפרש הסכומים:}$$

$$5. \sum_{t=1}^T X_t^2 \neq \left(\sum_{t=1}^T X_t \right)^2 \quad \text{יש לשים לב כי:}$$

$$\sum_{t=1}^T X_t Y_t \neq \sum_{t=1}^T X_t \sum_{t=1}^T Y_t$$

הגדרות ופיתוחים:

1. סכום הסטיות מהממוצע $= 0$: $\sum_{t=1}^T (X_t - \bar{X}) = 0$
2. סכום הסטיות הריבועיות מהממוצע (מונה השונות):

$$S_{XX} = \sum_{t=1}^T (X_t - \bar{X})^2 = \sum_{t=1}^T X_t^2 - T\bar{X}^2 = \sum_{t=1}^T (X_t - \bar{X})X_t$$
3. מונה של השונות המשותפת:

$$S_{XY} = \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^T X_t Y_t - T\bar{X}\bar{Y} = \sum_{t=1}^T (X_t - \bar{X})Y_t = \sum_{t=1}^T (Y_t - \bar{Y})X_t$$

חוקי התוחלת:

1. תוחלת של קבוע = קבוע: $E(a) = a$
2. תוחלת של סכום/הפרש = לסכום/הפרש התוחלות:

$$E(X \pm Y) = E(X) \pm E(Y)$$

$$E\left(\sum (X_i)\right) = \sum E(X_i)$$
3. תוחלת של כפל/חילוק $\neq$ לכפל/חילוק התוחלות:

$$E(X \cdot Y) \neq E(X) \cdot E(Y)$$

$$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$$

$$E(X^2) \neq [E(X)]^2$$
4. השפעת טרנספורמציה ליניארית על התוחלת:

$$E\left(a + \frac{1}{a}X \pm b\right) = a + \frac{1}{a} \cdot E(X) \pm b$$

חוקי השונות:

1. עבור X ו- Y בלתי תלויים/בלתי מתואמים מתקיים:
שונות של סכום/הפרש = סכום השונות:

$$V(X \pm Y) = V(X) + V(Y)$$

$$V\sum(X_i) = \sum V(X_i)$$
2. עבור X ו- Y תלויים/מתואמים מתקיים:
שונות של סכום/הפרש $\neq$ סכום השונות:

$$V(X \pm Y) = V(X) + V(Y) \pm 2 \cdot Cov(X, Y)$$

$$V(a) = 0$$

3. שונות של קבוע = 0 : $V(a \pm x) = V(X)$

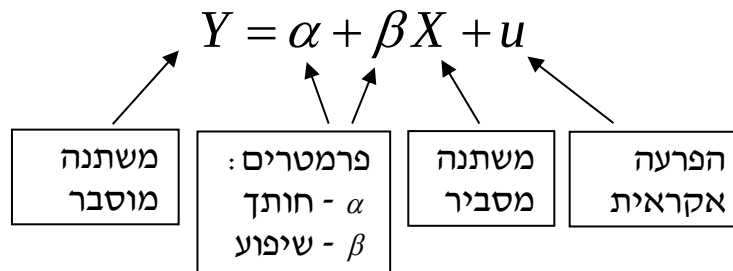
4. השפעת טרנספורמציה ליניארית על השונות : $V(aX + b) = a^2 V(X)$

- חוקי התוחלת והשונות מתייחסים למשתנים אמפיריים כאל קבועים (יוצאים מחוץ לתוחלת או לשונות).
חוקי הסכימה מתייחסים למשתנים אמפיריים כמשתנים הנשארים בתוך הסיגמא (רק הקבועים ייצאו מחוץ לסיגמא).

חוקי השונות המשותפת :

1. שונות משותפת בין משתנה לקבוע = 0 : $\text{cov}(X, a) = 0$
2. שונות משותפת של משתנים המוכפלים בקבוע : $\text{cov}(aX, bY) = ab \cdot \text{cov}(X, Y)$
3. שונות משותפת של משתנה עם עצמו = שונות המשתנה : $\text{cov}(X, X) = V(X)$
 $\text{cov}(Y, Y) = V(Y)$

המודל האקונומטרי :



1. במודל : $Y_t = \alpha + \beta X_t + u_t$, α ו- β הם מספרים קבועים אך לא ידועים. אנו יכולים להעריך אותם ולקבל אומדים (תהליך קבלת האומדנים נקרא אמידה).
2. $\hat{\alpha}$ הוא האומד ל- α ו- $\hat{\beta}$ הוא האומד ל- β .
3. אומדי ריבועים פחותים (אר"פ) הם אומדים שחושבו בשיטת הריבועים הפחותים. מסומנים בד"כ ע"י 'כובעי' - $\hat{\beta}$.
אומדים אחרים מסומנים בד"כ ע"י 'תלתלי' - $\tilde{\beta}$.

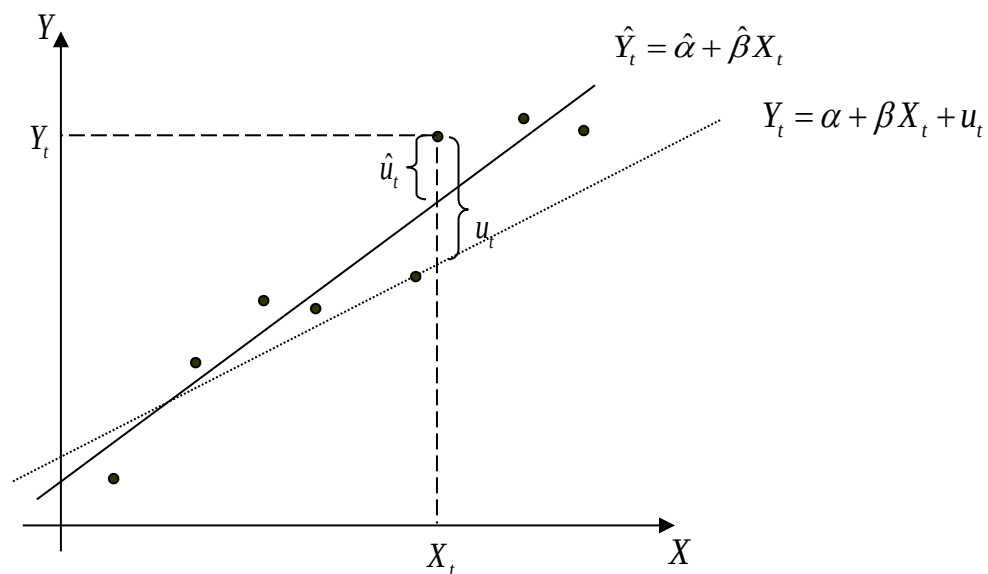
4. בעוד α ו- β הם קבועים, $\hat{\alpha}$ ו- $\hat{\beta}$ הם משתנים מקריים כיוון שבכל מדגם מתקבלים $\hat{\alpha}$ ו- $\hat{\beta}$ אחרים.

5. את α ו- β ו- u_t לא ניתן לדעת (אלא רק לאמוד מנתוני המדגם) – הקו האמיתי באוכי לא ידוע.

6. אפשר לדעת את $\hat{u}_t$, שהיא הסטיה מקו הרגרסיה במדגם:

- עבור X_t , הערך הצפוי של המשתנה המוסבר $(\hat{Y}_t)$ המתקבל לפי הרגרסיה הוא: $\hat{Y}_t = \hat{\alpha} + \hat{\beta}X_t$.

- הסטיה של התצפית (Y_t) מהערך הצפוי לפי הרגרסיה $(\hat{Y}_t)$ היא: $\hat{u}_t = Y_t - \hat{Y}_t$.



— קו הרגרסיה הנאמד (במדגם)
 קו הרגרסיה האמיתי באוכלוסייה
 • תצפית בודדת

שאלות:

(1) הבא נוכיח את הזהויות הבאות:

$$א. \sum_{t=1}^T (X_t - \bar{X})^2 = \sum_{t=1}^T X_t^2 - T\bar{X}^2$$

$$ב. \sum_{t=1}^T (X_t - \bar{X})^2 = \sum_{t=1}^T (X_t - \bar{X})X_t$$

$$ג. \sum_{t=1}^T (X_t - \bar{X}) = 0$$

$$ד. \sum \frac{(X_t - \bar{X})^2}{\sum (X_t - \bar{X})X_t} = 1$$

$$ה. \sum (\hat{\alpha} + \hat{\beta}x_i)(x_i - \bar{x}) = \hat{\beta}(x_i - \bar{x})^2$$

$$ו. \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^T X_t Y_t - T\bar{X}\bar{Y}$$

$$ז. \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^T (X_t - \bar{X})Y_t$$

(2) בטא באמצעות: $\text{cov}(x, y), \text{var}(x), \text{var}(y)$ והקבועים a ו- b את הביטויים

הבאים:

$$א. \text{Var}(ax)$$

$$ב. \text{Var}(x + y)$$

$$ג. \text{Var}(ax + b)$$

$$ד. \text{Cov}(x, ay)$$

$$ה. \text{Cov}(x + a, y + b)$$

$$ו. \text{מקדם המתאם בין } x \text{ ל-} y$$

תשובות סופיות:

(1) הוכחה.

(2) א. $a^2 \text{var}(x)$ ב. $\text{var}(x) + \text{var}(y) + 2\text{cov}(x, y)$ ג. $a^2 \text{var}(x)$ ד. $a \text{cov}(x, y)$

$$ה. \text{cov}(x, y) \quad ו. r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)} \cdot \sqrt{\text{var}(y)}}$$

אקונומטריקה א

פרק 6 - אומדי הריבועים הפחותים

תוכן העניינים

1. כללי 27

אומדי הריבועים הפחותים:

רקע:

שיטת האמידה של α ושל β לקבלת אומדים $\hat{\alpha}$ ו- $\hat{\beta}$ – Ordinary Least Squares (OLS) שיביאו למינימום את סכום ריבועי טעויות האמידה:

$$\min_{\hat{\alpha}\hat{\beta}} \sum \hat{u}_t^2 = \min_{\hat{\alpha}\hat{\beta}} \sum (y_t - \hat{y}_t)^2 = \min_{\hat{\alpha}\hat{\beta}} \sum \left[y_t - (\hat{\alpha} + \hat{\beta}x_t) \right]^2 = ?$$

מתוך גזירת הפונקציה הזו מתקבלים האומדים $\hat{\alpha}$ ו- $\hat{\beta}$.

מודל רק עם חותך $Y_t = \alpha + u_t$	מודל ללא חותך $Y_t = \beta X_t + u_t$	מודל עם חותך ושיפוע $Y_t = \alpha + \beta X_t + u_t$	
$\hat{\alpha} = \bar{Y}$	$\hat{\beta} = \frac{\sum_{t=1}^T X_t Y_t}{\sum_{t=1}^T X_t^2}$	$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^T (X_t - \bar{X})^2}$ $= \frac{\sum_{t=1}^T (X_t - \bar{X})Y_t}{\sum_{t=1}^T (X_t - \bar{X})^2}$ $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$	חישוב האומדים
$E(\hat{\alpha}) = \alpha$	$E(\hat{\beta}) = \beta$	$E(\hat{\beta}) = \beta$ $E(\hat{\alpha}) = \alpha$	תוחלת האומדים
$V(\hat{\alpha}) = \frac{\sigma_u^2}{T}$	$V(\hat{\beta}) = \frac{\sigma_u^2}{\sum_{t=1}^T X_t^2}$	$V(\hat{\beta}) = \frac{\sigma_u^2}{S_{XX}}$ $V(\hat{\alpha}) = \sigma_u^2 \left(\frac{1}{T} + \frac{\bar{X}^2}{S_{XX}} \right)$	שונות האומדים

"המשוואות הנורמליות" מתקבלות בתהליך הגזירה של פונקציית הריבועים הפחותים וחייבות להתקיים על מנת שהפונקציה תתקיים $(\sum \hat{u}_t^2 = \min)$:

עבור המודל הקלאסי (עם חותך):

$$\sum \hat{u}_t = 0 \quad \alpha$$

$$\sum \hat{u}_t \cdot x_t = 0 \quad \beta$$

עבור מודל ללא חותך:

$$\sum \hat{u}_t \cdot x_t = 0 \quad \beta \text{ בלבד}$$

מן המשוואות הנורמליות נובעות:

1. התכונות הגיאומטריות:

$$\sum \hat{u}_i = 0 \quad \text{א.}$$

$$\sum x_i \hat{u}_i = 0 \quad \text{ב.}$$

- ברגרסיה ללא שיפוע מתקיימת רק התכונה הגיאומטרית הראשונה. ברגרסיה ללא חותך מתקיימת רק התכונה הגיאומטרית השנייה.

2. התכונות האלגבריות:

$$\text{cov}(x_i, \hat{u}_i) = 0 \quad \text{א.}$$

$$\text{cov}(\hat{y}_i, \hat{u}_i) = 0 \quad \text{ב.}$$

$$\bar{y} = \hat{\alpha} + \hat{\beta} \bar{x} = \bar{\hat{y}} \quad \text{ג.}$$

- התכונות האלגבריות תקפות עבור קו הרגרסיה הקלאסי (עם חותך ושיפוע) במדגם בלבד.

ההנחות הקלאסיות של מודל הרגרסיה:

1. קיים קשר ליניארי בין המשתנה המוסבר למשתנה המסביר.

$$2. X \text{ איננו קבוע: } S_{XX} = \sum_{t=1}^T (X_t - \bar{X})^2 \neq 0$$

3. תוחלת ההפרעה האקראית היא אפס לכל תצפית: $E(u_t) = 0$ לכל t .

4. X_t אינם משתנים מקריים $\Leftrightarrow$ ניתן להוציא אותם מחוץ לתוחלת ולשונוות $\Leftrightarrow$
 $\text{cov}(X_t, u_t) = 0$

5. הומוסקדסטיות: שונות ההפרעה האקראית קבועה לכל תצפית:

$$V(u_t) = \sigma_u^2 \text{ לכל } t.$$

6. u_t ב"ת: $\text{cov}(u_t, u_s) = 0$ לכל $t \neq s$.

7. ההפרעות האקראיות מתפלגות נורמלית: $u_t \approx N$.

תכונות האומדים:

אומדי הריבועים הפחותים הם לינאריים, חסרי הטיות, יעילים ועקיבים.

1. לינאריות:

אר"פ ניתנים להצגה כטרנספורמציה לינארית של Y_t .

כדי ש- $\hat{\beta}$ למשל, יהיה אומד ליניארי צריך להתקיים: $\hat{\beta} = \sum W_t \cdot Y_t$.

כאשר W_t היא קומבינציה של ערכי X בדרך כלל. למשל: $\hat{\beta} = \frac{\sum X_t \cdot Y_t}{\sum X_t^2}$.

כדי להביא את האומד לצורה: $\tilde{\beta} = \sum w_t \cdot y_i$ נעזר בשוויון: $\frac{\sum 0}{\sum 0} = \sum \frac{0}{\sum 0}$.

אומד זה ניתן להצגה בצורה הבאה:

$$\hat{\beta} = \sum \frac{X_t}{\sum X_t^2} Y_t = \sum W_t \cdot Y_t$$

$$W_t = \frac{X_t}{\sum X_t^2}$$

לפיכך מדובר באומד ליניארי.

• שימו לב כי:

W_t אסור שיכלול את Y_t .

Y_t אסור שיהיה במכנה או בשורש/חזקה (אלא אם כן במודל הנתון הוא מצוי בשורש/חזקה).

2. חוסר הטיה :

אומד $\hat{\theta}$ מסוים יהווה אח"ה לפרמטר θ אותו הוא אומד באוכלוסייה אם מתקיים: $E(\hat{\theta}) = \theta$.

כיצד יודעים אם אומד הוא חסר הטיה?

1. בשלב הראשון יש לבצע עבודת הכנה – מבטאים את האומד באמצעות הפרמטר האמיתי – מציבים במקום ה- Y_t את המודל ומפתחים אלגברית.

• יש לזכור כי:

$$\begin{aligned} & u_t \\ & y_t \\ & \sum \text{וה-} \\ & x_t \text{ איננו משתנה מקרי (על פי הנחה מס' 4)} \Leftarrow \text{יוצא מחוץ לתוחלת} \\ & \text{ולשונות אך נשאר בתוך ה-} \sum \text{ו-} \frac{\alpha}{\beta} \text{ קבועים} \Leftarrow \text{יוצאים מחוץ} \\ & \text{לתוחלת, לשונות ול-} \sum. \end{aligned}$$

2. בשלב השני מפעילים תוחלת על האומד המפותח ואם התוחלת שווה לפרמטר האמיתי אז האומד חסר הטיה.

• חוסר הטיה מחייב את התקיימותן של הנחות (3) $E(u_t) = 0$ לכל t
ו- (4) $\text{cov}(X_t, u_t) = 0$.

3. יעילות :

יעילות פירושה השונות הקטנה ביותר. ככל שהשונות של האומד קטנה יותר, כך יש הסתברות גבוהה יותר שהוא יהיה קרוב לפרמטר האמיתי באוכלוסייה אותו הוא אומד.
 $\hat{\theta}_1$ יקרא אומד יעיל יותר מ- $\hat{\theta}_2$ אם מתקיים שהשונות שלו קטנה יותר:
 $V(\hat{\theta}_1) < V(\hat{\theta}_2)$

משפט גאוס מרקוב – אר"פ הם בעלי השונות הנמוכה ביותר בקבוצה שלהם (קבוצת האומדים הלינאריים חסרי ההטיה), והם נקראים: B.L.U.E.
(Best Linear Unbiased Estimation).

כיצד מחשבים שונות של אומד?

חייבות להתקיים הנחות (4) $\text{cov}(X_t, u_t) = 0$, (5) $V(u_t) = \sigma_u^2$ לכל t

ו-(6) $\text{cov}(u_t, u_s) = 0$ לכל $t \neq s$. אם הן מתקיימות, מחשבים את השונות של האיברים המכילים את u_t מהפיתוח הקודם (לפי כללי הסיגמא והשונות).

4. עקיבות:

ככל שהמדגם יגדל כן יתקרב האומד לערך האמיתי של הפרמטר. אם נגדיל את המדגם לאינסוף תצפיות ונחשב את האומד, הוא יהיה שווה

$$\left(\hat{\theta} \rightarrow \theta \right) \\ \left(T \rightarrow \infty \right)$$

תנאי הכרחי לעקיבות:

האומד חייב להיות פונקציה של גודל המדגם. במילים אחרות, האומד צריך להיות מושפע מגודל המדגם. ברגע שהאומד עונה על תנאי זה הוא יהיה עקיב. אומד המחושב במדגם סופי בהגדרה לא יוכל להיות עקיב לפרמטר באוכלוסייה.

סיכום: השלבים להוכחת התכונות:

1. הוכחת ליניאריות.

2. הכנת האומד $\Leftarrow$ להציב במקום Y_t את המודל האמיתי.

במודל עם חותך: $Y_t = \alpha + \beta X_t + u_t$

במודל ללא חותך: $Y_t = \beta X_t + u_t$

3. פיתוח האלגברה.

4. חישוב תוחלת, שונות, עקיבות.

- ליניאריות מהווה תנאי הכרחי לחוסר הטיה.
- ליניאריות וחוסר הטיה מהוות תנאי הכרחי לבחינת היעילות של האומד לפי משפט גאוס-מרקוב.
- עקיבות איננה תלויה בתכונות האחרות, אלא רק בהיותו של האומד פונקציה של גודל המדגם (לא מחושב על מדגם סופי). כך שאומד לא חייב להיות ליניארי או חסר הטיה כדי להיות עקיב.
- העקיבות משפיעה על היעילות של האומד. עבור אומדים התלויים בגודל המדגם: ככל שגודל המדגם גדול יותר כך שונות האומד קטנה והאומד יהיה יעיל יותר לפרמטר באוכלוסייה.

שאלות:

תרגול ממבחנים:

(1) נתון המודל: $Y_t = \alpha + \beta X_t + u_t$, $T = 100$, כאשר מתקיימות כל ההנחות הקלאסיות.

$$\tilde{\beta} = \frac{\sum_{t=51}^{100} Y_t - \sum_{t=1}^{50} Y_t}{\sum_{t=51}^{100} X_t - \sum_{t=1}^{50} X_t} : \text{נתון האומד}$$

- האומד $\tilde{\beta}$ הינו אומד חסר הטיה ל- β . נכון / לא נכון
- האומד $\tilde{\beta}$ הינו אומד עקיב ל- β . נכון / לא נכון
- האומד $\tilde{\beta}$ הינו אומד לינארי ל- β . נכון / לא נכון
- האומד $\tilde{\beta}$ הינו אומד יעיל ל- β . נכון / לא נכון
- השונות האמיתית של $\tilde{\beta}$ היא?

(2) נתון המודל: $Y_t = \beta X_t + u_t$, כאשר כל ההנחות הקלאסיות מתקיימות. (יש לשים לב המודל ללא חותך).

$$\tilde{\beta} = \frac{\sum Y_t}{\sum X_t} : \text{נתון האומד}$$

- האומד $\tilde{\beta}$ הינו אומד מוטה ל- β : נכון / לא נכון / אי אפשר לדעת
- על סמך משפט גאוס-מרקוב ניתן להסיק כי $\tilde{\beta}$ איננו אומד יעיל יותר מאומד הריבועים הפחותים: נכון / לא נכון / לא ניתן לדעת
- מהי השונות האמיתית של $\tilde{\beta}$?

(3) נתון המודל: $Y_t = \beta X_t + u_t$, כאשר כל ההנחות הקלאסיות מתקיימות. (יש לשים לב המודל ללא חותך).

$$\tilde{\beta} = \frac{\sum X_t Y_t}{\sum (X_t - \bar{X})^2} : \text{נתון האומד}$$

- מהי התוחלת של $\tilde{\beta}$?
- $E(\tilde{\beta}) < \beta$. נכון / לא נכון / אי אפשר לדעת
- על סמך משפט גאוס-מרקוב ניתן להסיק כי אומד הריבועים הפחותים הינו אומד יעיל יותר מ- $\tilde{\beta}$. נכון / לא נכון / לא ניתן לדעת
- מהי השונות האמיתית של האומד $\frac{\sum X_t Y_t}{\sum X_t^2}$?

4) בכל השאלות ההנחות הקלאסיות מתקיימות.

האומדים הם אר"פ, והמודל הוא: $Y_t = \alpha + \beta X_t + u_t$.

א. $E(Y_t) = E(\hat{Y}_t)$ נכון / לא נכון / אי אפשר לדעת

ב. $\sum_{t=1}^T (X_t - \bar{X}) \bar{Y} \neq 0$ נכון / לא נכון / אי אפשר לדעת

ג. אמידת המודל בשיטת הריבועים הפחותים תיתן את

התוצאה: $\sum_{t=1}^T u_t = 0$ נכון / לא נכון / אי אפשר לדעת

ד. אם נתון ש- $r_{XY} = 0.57$, אזי $\hat{\beta}$:

i. הוא בהכרח שלילי.

ii. הוא בהכרח חיובי.

iii. הוא בהכרח שווה לאפס.

iv. לא ניתן לקבוע את סימנו על סמך הנתונים הקיימים.

ה. סמן את הטענה הנכונה בהכרח:

i. $\sum_{t=1}^T (X_t - \bar{Y}) \hat{u}_t = 0$

ii. $S_{XX} = \sum_{t=1}^T X_t^2 - (T\bar{X})^2$

iii. $\sum_{t=1}^T X_t u_t = 0$

iv. אף אחת מהטענות הנ"ל אינה נכונה בהכרח.

ו. אומדי הריבועים הפחותים אינם חסרי

הטיה, אם נתון שהשונות של u_t

אינה קבועה. נכון / לא נכון / אי אפשר לדעת

ז. אומד חסר הטיה הוא אינו בהכרח

גם אומד עקיב. נכון / לא נכון / אי אפשר לדעת

תשובות סופיות:

(1) א. נכון. ב. לא נכון. ג. נכון. ד. לא נכון.

$$V(\tilde{\beta}) = \frac{100\sigma_u^2}{\left(\sum_{t=51}^{100} X_t - \sum_{t=1}^{50} X_t\right)^2} \quad \text{ה.}$$

(2) א. לא נכון. ב. נכון. ג. $V(\tilde{\beta}) = \frac{T\sigma_u^2}{(\sum X_t)^2}$ ד. לא נכון.

(3) א. $E(\tilde{\beta}) = \frac{\beta \sum X_t^2}{\sum (X_t - \bar{X})^2}$ ב. לא נכון. ג. לא נכון. ד. לא נכון.

$$\cdot \frac{\sigma_u^2}{\sum X_t^2} \quad \text{ד.}$$

(4) א. נכון. ב. לא נכון. ג. לא נכון. ד. ii. ה. i. ו. לא נכון. ז. נכון.

אקונומטריקה א

פרק 7 - מודלים לא ליניאריים

תוכן העניינים

1. כללי 35

מודלים לא ליניאריים:

רקע:

הגמישות $\left(\frac{\partial Y}{\partial X} \cdot \frac{X}{Y}\right)$ בכמה % ישתנה Y אם נגדיל את X ב-1%?	השינוי השולי $\left(\frac{\partial Y}{\partial X}\right)$ בכמה ישתנה Y אם נגדיל את X ביחידה?	משמעות ה- β	המודל
$\frac{\beta X}{Y}$	β	השינוי השולי אם נגדיל את X ביחידה Y ישתנה ב- β יחידות	ליניארי: $Y = \alpha + \beta X + u$
βX	βY	שיעור השינוי השולי אם נגדיל את X ביחידה Y ישתנה ב- $100 \cdot \beta\%$	חצי לוגריתמי: $\ln Y = \alpha + \beta X + u$ $(Y = e^{\alpha + \beta X + u})$
β	$\frac{\beta Y}{X}$	הגמישות אם נגדיל את X ב-1% Y ישתנה ב- $\beta\%$	לוגריתמי כפול: $\ln Y = \alpha + \beta \ln X + u$ $(Y = e^\alpha \cdot X^\beta \cdot e^u)$
$\frac{\beta}{Y}$	$\frac{\beta}{X}$	אין משמעות כלכלית אם נגדיל את X ב-1% Y ישתנה ב- β	לוג ליניארי: $Y = \alpha + \beta \ln X + u$ $(e^Y = e^\alpha \cdot X^\beta \cdot e^u)$

- המשתנה שיש בו LN השינוי בו יהיה באחוזים.

תזכורת של חוקי לוגים:

$$\begin{aligned}
 LN(e^x) &= x \\
 LN(X^Y) &= Y \cdot LN(X) \\
 LN(X \cdot Y) &= LN(X) + LN(Y) \\
 LN\left(\frac{X}{Y}\right) &= LN(X) - LN(Y)
 \end{aligned}$$

שאלות:

(1) על מנת לאמד את התשואה להשכלה בישראל בשנים 1948-1990 נאמדו המודלים הבאים:

$$1. \quad MWAGE_t = 139.547 + 118.628 \cdot SCL_t$$

$$2. \quad MWAGE_t = -1445.08 + 1239.60 \cdot LN(SCL)_t$$

$$3. \quad LN(MWAGE)_t = 5.244 + 0.778 \cdot LN(SCL)_t$$

$$4. \quad LN(MWAGE)_t = 6.292 + 0.070 \cdot SCL_t$$

א. הסבירו את המשמעות של β בכל אחד מהמודלים.

ב. חשבו את הגמישות בנקודת הממוצעים: (12.311, 1600.01) עבור כל אחד מהמודלים.

(2) נתונים תוצאות האמידה של המודלים הבאים:

$$1. \quad \hat{Y} = e^{4.5} \cdot X^{0.05}$$

$$2. \quad \hat{Y} = e^{4.5+0.05X}$$

$$3. \quad \hat{Y} = 4.5 + \frac{0.05}{X}$$

$$4. \quad \hat{Y} = \frac{1}{1 + e^{4.5+0.05X}}$$

א. כתבו את המודלים בצורה ליניארית בעזרת טרנספורמציה מתאימה.

ב. עבור כל אחד מהמודלים ערכו תחזית נקודתית עבור $X = 6$.

(3) נתונים המודלים הבאים עבור התוצר במשק:

$$1. \quad Q_i = AK_i^{\beta_1} e^{u_i}$$

$$2. \quad Q_i = Ae^{\beta_1 L_i + u_i}$$

$$3. \quad Q_i = A + K_i^{\beta_1} + e^{u_i}$$

$$4. \quad Q_i = A + \frac{\beta_1}{L_i} + u_i$$

$$5. \quad Q_i = A + \beta_1 \sqrt{K_i} + u_i$$

$$6. \quad Q_i = e^{A + \beta_1 K_i + u_i}$$

$$7. \quad Q_i = A \left(\frac{K_i}{2} + 7 \right)^{\beta_1} e^{u_i}$$

$$8. \quad Q_i = A + \beta_1 L_i + u_i$$

$$9. \quad Q_i = A + \beta_1 \left(\frac{K_i}{L_i} \right) + u_i$$

כאשר :

Q - הוצאות צריכה על מוצר מסוים על ידי פרט מסוים.

A - הוצאות צריכה על המוצר בהינתן רמת הכנסה אפסית.

K - הכנסת הפרט.

L - שנות לימוד.

א. מי מהמודלים הבאים ניתן לאמידה בשיטת OLS?

ב. מי מבין המודלים שלא ניתנים לאמידה בשיטת OLS ניתן להביא למודל ליניארי בפרמטרים ועל כן לאמוד את הפרמטרים שלו?

ג. עבור כל אחד מהמודלים קבעו מיהו המשתנה המוסבר ומיהו המסביר במשוואת הרגרסיה הליניארית.

ד. עקומת אנג'ל מתארת את גמישות הצריכה של הפרט מוצר מסוים ביחס להכנסתו. איזה מהמודלים מתאים כדי לתאר את עקומת אנג'ל?

$$(4) \quad \text{נתון המודל הבא: } Q_i = \frac{A}{K_i^{\beta_1}} e^{u_i}$$

א. האם ניתן לאמוד את המודל בשיטת OLS?

ב. מה המשוואה שצריך לאמוד על מנת לקבל את הפרמטרים למודל זה (כלומר כיצד הופכים את המודל לליניארי בפרמטרים)?

ג. נאמד המודל הבא: $\ln(Q_i) = \alpha_0 + \alpha_1 \ln(K_i) + u_i$, והתקבלו התוצאות

הבאות: $\hat{\alpha}_0 = 3$, $\hat{\alpha}_1 = 0.8$.

מהם האומדנים עבור A , β_1 ?

(5) נתון כי הקשר באוכלוסייה בין X ל- Y נתון על ידי המודל

הבא: $\ln Y = \alpha + \beta \ln X + u$. נתון גם כי עבור המודל הנ"ל כל ההנחות הקלאסיות מתקיימות.

$$\tilde{\beta} = \frac{\sum_{t=1}^T (\ln X_t - \ln \bar{X}) \ln Y_t}{\sum_{t=1}^T (\ln X_t - \ln \bar{X})^2} : \beta \quad \text{כלכלן הציע את האומדן הבא עבור } \beta$$

א. האם האומדן ליניארי?

ב. האם האומדן חסר הטיה?

ג. האם האומדן $blue$?

ד. מהי שונותו?

תשובות סופיות:

- (1) א.1. השינוי השולי. 2. אין משמעות כלכלית. 3. גמישות. 4. שיעור השינוי השולי.
 ב.1. 0.912 2. 0.77 3. 0.778 4. 0.861
- (2) א.1. $\ln(Y) = 4.5 + 0.05 \cdot \ln(X)$ 2. $\ln(Y) = 4.5 + 0.05X$
 3. אין צורך. 4. $\ln\left(\frac{1-\hat{Y}}{\hat{Y}}\right) = 4.5 + 0.05X$
 ב.1. 98.45 2. 121.51 3. 4.50833 4. 0.00816
- (3) א. מודלים: 4, 5, 8 ו-9. ב. מודלים: 1, 2, 6 ו-7.
 ג.1. מסביר: $\ln(K_i)$, מוסבר: $\ln(Q_i)$ 2. מסביר: L_i , מוסבר: $\ln(Q_i)$
 3. אינו ליניארי. 4. מסביר: $\frac{1}{L_i}$, מוסבר: Q_i
 5. מסביר: $\sqrt{K_i}$, מוסבר: Q_i 6. מסביר: K_i , מוסבר: $\ln(Q_i)$
 7. מסביר: $K_i = \frac{K_i}{2} + 7$, מוסבר: $\ln(Q_i)$ 8. מסביר: L_i , מוסבר: Q_i
 9. מסביר: $\frac{K_i}{L_i}$, מוסבר: Q_i
 ד. מודלים: 1 ו-7.
- (4) א. לא. ב. $\ln(Q_i) = \ln(A) - \beta_1 \ln(K_i) + u_i$ ג. $\beta_1 = -0.8$, $A = 20$
- (5) א. כן. ב. כן. ג. כן. ד. $V(\hat{\beta}) = \frac{\sigma_u^2}{SS \ln x}$

אקונומטריקה א

פרק 8 - מבחני המובהקות וקריאת פלטים - תוכנת SAS

תוכן העניינים

1. כללי..... 39

מבחני המובהקות וקריאת פלטים – תוכנת SAS:

רקע:

פלט ניתוח שונות (Analysis of Variance):

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	k	RSS	$RSS/k = MSR$	$F = \frac{MSR}{MSE}$	PF
Error	$T - k - 1$	ESS	$ESS/T - k - 1 = MSE$		
C Total	$T - 1$	TSS			

Root MSE		$\sqrt{MSE} = s_u$	R-square	$R^2 = \frac{RSS}{TSS}$	
Dep Mean		$\bar{Y}$	Adj R-sq	$\bar{R}^2 = 1 - \frac{ESS}{TSS} \cdot \frac{T - 1}{T - k - 1}$	
C.V.		$\frac{s_u}{\bar{Y}} \cdot 100$			

פלט מקדמי הרגרסיה (Parameter Estimates):

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	$\hat{\alpha}$	$s_{\hat{\alpha}}$	$\frac{\hat{\alpha}}{s_{\hat{\alpha}}} = t_{(\hat{\alpha}=0)}$	$Pt_{\hat{\alpha}}$
X	1	$\hat{\beta}$	$s_{\hat{\beta}}$	$\frac{\hat{\beta}}{s_{\hat{\beta}}} = t_{(\hat{\beta}=0)}$	$Pt_{\hat{\beta}}$

פלט ה – Covariance of Estimates :

פלט שמתאר את השונות המשותפת (covariance) של האומדנים $\hat{\alpha}$ ו- $\hat{\beta}$:

Covariance of Estimates		
COVB	INTERCEP	X
INTERCEP	$s_{\hat{\alpha}}^2$	$\text{cov}(\hat{\alpha}, \hat{\beta})$
X	$\text{cov}(\hat{\alpha}, \hat{\beta})$	$s_{\hat{\beta}}^2$

עריכת תחזית וקריאת פלטים (תוכנת SPSS) :

אמידה נקודתית :

אמידה נקודתית עבור X_0 מסוים (תחזית).

מחושבת על פי קו הרגרסיה במדגם : $\hat{Y}_0 = \hat{\alpha} + \hat{\beta} \cdot X_0$.

אמידת מרווח ל- $E(Y)$:

אמידת התחזית באוכלוסייה עבור X_0 מסוים. נחשב רווח בר סמך לערך ממוצע של Y

באוכ' עבור X_0 מסוים ($E(Y)$) ברמת סמך $1-\alpha$.

נוסחת הרב"ס : $\hat{Y} \pm t_{n-2; 1-\frac{\alpha}{2}} \hat{\sigma}_u \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$

$\hat{\sigma}_u = MSE = \frac{SSE}{n-2}$, $\sum (X_i - \bar{X})^2 = S_{xx} = (n-1)S_X^2$

רישום הרב"ס : $p(\text{---} \leq E(Y) \leq \text{---}) = 1-\alpha$

אמידת מרווח ל- Y :

אמידת ערך בודד של Y באוכלוסייה עבור X_0 מסוים. נחשב רווח בר סמך לערך בודד

של Y באוכ' עבור X_0 מסוים (Y_0) ברמת סמך $1-\alpha$.

נוסחת הרב"ס : $\hat{Y} \pm t_{n-2; 1-\frac{\alpha}{2}} \hat{\sigma}_u \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$

רישום הרב"ס : $p(\text{---} \leq Y \leq \text{---}) = 1-\alpha$

- רב"ס לערך בודד יהיה רחב יותר מאשר רב"ס לערך ממוצע משום שטעות התקן בראשון גדולה מאשר באחרון.

שאלות:

פלט ניתוח שונות:

- (1) חוקר רצה לבחון את השפעת ההכנסה ($INCOME$) על גובה המס (TAX) (במיליארדי \$) שגובה מדינה במערב לפי המודל: $TAX_t = \alpha + \beta \cdot INCOME_t + u_t$. לשם כך אסף נתונים מ-51 מדינות. להלן התוצאות:

Model: MODEL1

Dependent Variable: TAX

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2046.89694	2046.89694	8798.672	0.0001
Error	49	11.39922	0.23264		
C Total	50	2058.29615			
Root MSE	0.48232		R-square	0.9945	
Dep Mean	5.4242		Adj R-sq	0.9943	
C.V.	8.88711				

בדקו את ההשערה כי המודל מובהק ברמת מובהקות של 0.05.

פלט מקדמי הרגרסיה:

- (2) בהמשך לדוגמא הקודמת – בדיקת השפעת ההכנסה על גודל המס, התקבלו גם התוצאות הבאות:

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	-0.086912	0.08953904	-0.971	0.3365
INCOME	1	0.152232	0.0016229	93.801	0.0001

- א. אמדו את המודל: $TAX = \alpha + \beta \cdot INCOME + U$. מהי המשמעות הכלכלית של β ?
- ב. האם המודל מובהק? בדקו על סמך הפלט הנ"ל ברמת מובהקות של 0.05.
- ג. מהי רמת המובהקות הקטנה ביותר, עבורה עדיין תידחה השערת האפס מסעיף ב'?

ד. בדקו את ההשערה כי ככל שההכנסה עולה כך עולה גם המס (שיפוע β חיובי) ברמת מובהקות של 0.01.

ה. בנו רווח-סמך ברמת סמך של 95% עבור β .

ו. בדקו את ההשערה שתוספת של מיליארד \$ להכנסה תגדיל את המס ב-0.2 מיליארד \$, ברמת מובהקות של 0.05.

• שימו לב כי:

במודל עם משתנה מסביר אחד בלבד קיימת זהות בין מבחן F למובהקות המודל לבין מבחן t למובהקות ה- β :

$$F_{(1, T-2; 1-\alpha)} = t_{\left(T-2; 1-\frac{\alpha}{2}\right)}^2$$

$$F = t_{\beta}^2$$

כלומר: כל החלטה המתקבלת במבחן אחד חייבת להיות זהה להחלטה המתקבלת במבחן השני.

פלט שוניות משותפות:

(3) נתון פלט האמידה של המודל: $Y_t = \alpha + \beta X_t + u_t$, שלצורך אמידתו נאספו 240 תצפיות:

Parameter Estimates

Variable	DF	Parameter	Standard	T for H0:	
		Estimate	Error	Parameter=0	Prob> T
INTERCEP	1	5.25	0.25	21	0.0000
X	1	0.96	0.12	8	0.0000

Covariance of Estimates

	INTERCEP	X
INTERCEP	0.0625	-0.003
X	-0.003	0.0144

יש לבדוק את ההשערה: $H_0: \alpha = 5\beta$.

שאלה מסכמת:

(4) חוקר רצה לבדוק את השפעת הותק בעבודה (EXP) על השכר ($SALARY$) לפי המודל: $\ln(SALARY_t) = \alpha + \beta \cdot EXP_t + u_t$. הוא אסף 403 תצפיות, ואמד את הפרמטרים בתוכנת SAS. להלן חלקים מהפלט ויש להשלימו:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	---	---	5.68015	---	---
Error	---	205.22539	---		
C Total	---	---			

Root MSE	---	R-square	---
Dep Mean	7.14247	Adj R-sq	0.0245
C.V.	10.01602		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	---	---	---	---
EXP	1	-0.008740	---	---	0.0009

Covariance of Estimates

COVB	INTERCEP	EXP
INTERCEP	0.0047463101	---
EXP	-0.000154685	6.882844 E-6

- נתון נוסף: $EXP = 22$.

- קיים קשר חיובי מובהק בין ותק ללוג השכר. נכון / לא נכון
- שיעור התשואה בשכר לשנת ותק הוא?
- תחזית לוג השכר עבור אדם בעל 10 שנות ותק היא?

ביצוע תחזיות:

- 5) במדגם של 30 דירות מושכרות לסטודנטים ברדיוס של עד 2 ק"מ מסביב למכללה נחקר הקשר בין שכר דירה למספר הסטודנטים הגרים בדירה. להלן התוצאות:

Descriptive Statistics			
	Mean	Std. Deviation	N
שכר הדירה	1386.7667	509.46027	30
מספר הסטודנטים	3.0000	1.31306	30

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.602 ^a	.362	.339	414.05503

a. Predictors: (Constant), number of students

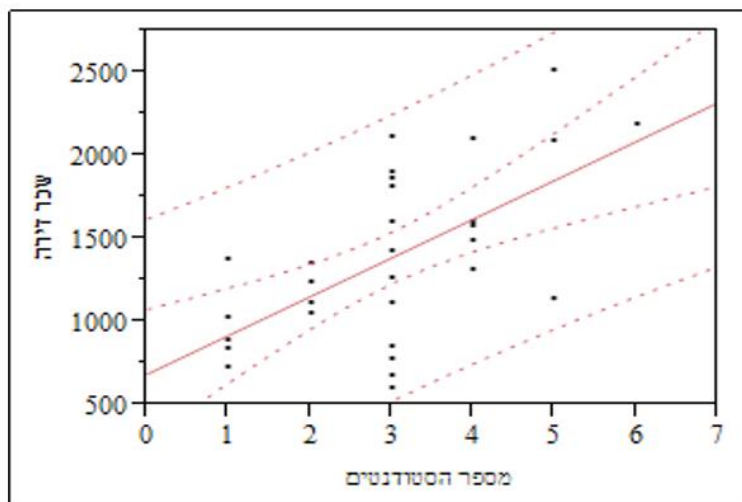
b. Dependent Variable: rent

ANOVA ^b					
Model	Sum of Squares	Df	Mean Square	F	Sig.
1 Regression	2726579.520	1	2726579.520	15.904	.000 ^a
Residual	4800363.847	28	171441.566		
Total	7526943.367	29			

a. Predictors: (Constant), number of students

b. Dependent Variable: rent

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	686.207	191.244		.001
	מספר הסטודנטים	233.520	58.556	.602	.000



- חשב אומדן נקודתי לשכר הדירה אותו ישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד.
- אמוד את שכר הדירה הממוצע שישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד, ברמת בטחון של 95%.
- אמוד את שכר הדירה שישלם סטודנט יחיד החולק את הדירה עם שותף אחד בלבד, ברמת ביטחון של 95%.

תשובות סופיות:

- (1) יש עדות לכך.
- (2) א. ראה סרטון. ב. יש עדות לכך. ג. $Pt_{\hat{\beta}} = 0.0001$.
- ד. יש עדות לכך. ה. $P(0.1488 \leq \beta \leq 0.1554) = 0.95$.
- ו. יש עדות לכך.
- (3) אין עדות לכך.
- (4) א. לא נכון. ב. -0.87% . ג. 7.24735 .
- (5) א. 1153.247 . ב. $p(957.4 \leq \mu_{Y_{X=2}} \leq 1349.08) = 0.95$. ג. $p(282.94 \leq Y_{X=2} \leq 2023.55) = 0.95$.

אקונומטריקה א

פרק 9 - שינוי יחידות מדידה

תוכן העניינים

1. כללי.....46

שינוי יחידות מדידה:

רקע:

טרנספורמציה ליניארית: הוספה/החסרה של קבוע ו/או הכפלה/חילוק של קבוע של אחד או שני המשתנים (התלוי והב"ת).

- טרנספורמציה ליניארית של המשתנים לא תשפיע על: R^2 , F , $t_{\hat{\beta}}$ ו- PF .

השינויים מסוכמים בטבלה הבאה:

$S_{\hat{\alpha}}$	$S_{\hat{\beta}}$	$\hat{\alpha}'$	$\hat{\beta}'$	
$s_{\hat{\alpha}'} \neq s_{\hat{\alpha}}$	$s_{\hat{\beta}'} = s_{\hat{\beta}}$	$\hat{\alpha}' = \hat{\alpha} - \hat{\beta}d$	$\hat{\beta}' = \hat{\beta}$	הוספת קבוע ל- X: $Y = \alpha' + \beta'(X + d) + v$
$s_{\hat{\alpha}'} = s_{\hat{\alpha}}$	$s_{\hat{\beta}'} = s_{\hat{\beta}}$	$\hat{\alpha}' = \hat{\alpha} + d$	$\hat{\beta}' = \hat{\beta}$	הוספת קבוע ל- Y: $Y + d = \alpha' + \beta'X + v$
$s_{\hat{\alpha}'} = s_{\hat{\alpha}}$	$s_{\hat{\beta}'} = \frac{s_{\hat{\beta}}}{d}$	$\hat{\alpha}' = \hat{\alpha}$	$\hat{\beta}' = \frac{\hat{\beta}}{d}$	הכפלת X פי קבוע: $Y = \alpha' + \beta'(dX) + v$
$s_{\hat{\alpha}'} = ds_{\hat{\alpha}}$	$s_{\hat{\beta}'} = ds_{\hat{\beta}}$	$\hat{\alpha}' = d\hat{\alpha}$	$\hat{\beta}' = d\hat{\beta}$	הכפלת Y פי קבוע: $dY = \alpha' + \beta'X + v$

- $t_{(\hat{\beta}'=0)} = t_{(\hat{\beta}=0)}$ תמיד.
- $t_{(\hat{\alpha}'=0)} = t_{(\hat{\alpha}=0)}$ רק בהכפלות.

שאלות:

- (1) חוקר ביקש לאמוד את הקשר בין שכר ב- $WAGE$ לבין שנות לימוד (SCL) באמצעות 2 מודלים שונים.
להלן תוצאות האמידה:

$$א. MWAGE_t = 139.54 + 118.62 \cdot SCL_t$$

$$ב. MWAGE_t = -1445.08 + 1239.60 \cdot \ln(SCL)_t$$

חשבו מחדש את מקדמי הרגרסיה וסטטיסטי המבחן F בכל אחד מהמודלים כתוצאה:

1. התברר כי נעשה טעות בחישוב מספר שנות הלימוד, ויש צורך להוסיף 20% למשתנה המקורי.
2. התברר כי הקשר בין שכר לשנות לימוד הוא ריבועי ולכן יש צורך להעלות את המשתנה המקורי של מספר שנות הלימוד בריבוע.

- (2) בהמשך לנתוני השאלה לדוגמא מהפרק החמישי:
- החוקר טען כי יש לבדוק את הקשר בין שכר לוותק ע"י שימוש בשכר נטו (NET) ולא בשכר ברוטו ($SALARY$). (קיים שיעור מס קבוע של 20%).
- המודל הוא: $\ln(NET_t) = \alpha' + \beta' \cdot EXP_t + v_t$
- מה יהיו ערכי האומדים, סטיות התקן שלהם וטיב ההתאמה באמידת מודל זה?

תשובות סופיות:

1. א. $\hat{\alpha}' = \alpha = 139.59$, $\hat{\beta}' = 98.85$, סטטיסטי F לא משתנה.
 - ב. $\hat{\alpha}' = -1671$, $\hat{\beta}' = 1239.6$, סטטיסטי F לא משתנה.
 2. לא ניתן לדעת.
- (2) $\hat{\beta}' = \hat{\beta} = -0.00874$, $\hat{\alpha}' = 7.11161$, $S_{\hat{\beta}'} = S_{\hat{\beta}} = 0.0026235$, $S_{\hat{\alpha}'} = S_{\hat{\alpha}} = 0.0688935$
- $R^2 = 0.0269$

אקונומטריקה א

פרק 10 - הרגרסיה המרובה

תוכן העניינים

1. רגרסיה מרובה..... 48

רגרסה מרובה:

רקע:

מבחן F - T :

כאשר יש יותר ממשתנה מסביר אחד, מדובר ברגרסה מרובה.
המודל הקלאסי: $Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + u_t$

- קבוע α יש אחד.
- מספר ה- β טות כמספר המשתנים ה"ת במודל.

מבחן F למובהקות המודל:

$$\begin{aligned} H_0 &= \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1 &: \text{OTHERWISE} \end{aligned}$$

השערות:

סטטיסטי המבחן F וכלל ההכרעה:

$$F = \frac{\frac{RSS}{k}}{\frac{ESS}{T-k-1}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{T-k-1}} > F(k, T-k-1; 1-\alpha)$$

מבחן t למובהקות ה- β טות:

מבחן לבדיקת מובהקות β ספציפית:

$$\begin{aligned} H_0 &= \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}$$

השערות:

סטטיסטי המבחן t וכלל ההכרעה:

$$|t_{\hat{\beta}_i}| = \left| \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}} \right| > t_{\left(T-k-1; 1-\frac{\alpha}{2}\right)}$$

השוואה בין מודלים – $\bar{R}^2$ וחוק חיטובסקי:

בכדי להחליט האם כדאי לנו להוסיף למודל משתנה ב"ת מסוים:
נשווה את פרופורציית השונות המוסברת המתוקנת $\bar{R}^2$ בין המודל ללא המשתנה
המסביר לבין המודל עם המשתנה המסביר שהוספנו.

- ניתן להשתמש גם באומד המוטטה - R^2 להשוואה בין מודלים אם מתקיימים
שני התנאים הבאים:

1. מספר המשתנים זהה.
2. המשתנה המוסבר זהה.

לפי חוק חיטובסקי – בהוספת משתנה מסביר אחד בלבד למודל ה- $\bar{R}^2$ יעלה אך

ורק אם: $|t_{\hat{\beta}}| > 1$.

כאשר: $|t_{\hat{\beta}}| < 1$ אז $\bar{R}^2$ ירד בהוספת המשתנה והוא גם לא יהיה רלוונטי למודל (מובהק).

כאשר: $|t_{\hat{\beta}}| > 2$ אז $\bar{R}^2$ יעלה והמשתנה שהוסף יהיה גם מובהק.

כאשר: $1 < |t_{\hat{\beta}}| < 2$ אז ה- $\bar{R}^2$ יעלה אך יש לבדוק את רלוונטיות המשתנה שהוסף למודל
על פי מבחן t .

שאלות:

מבחן T ו-F:

(1) נאמד המודל: $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבלו התוצאות הבאות:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	-----	646169.84	-----	-----	0.0000
Error	-----	-----	-----		
C Total	203	646790.01			

Root MSE	-----	R-square	-----
Dep Mean	178.6645	Adj R-sq	0.999022
C.V.	0.988075		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	5.067731	0.456604	11.09874	0.0000
X	1	-----	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-----	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

א. השלם את הנתונים החסרים בפלט.

ב. האם המודל מובהק? בדקו ברמת מובהקות של 0.05.

ג. האם משתנה W רלוונטי למודל? בדקו ברמת מובהקות של 0.01.

השוואה בין מודלים:

- (2) במודל לניבוי ההכנסה על פי שנות לימוד וותק במקום העבודה, התקבל: $\bar{R}^2 = 0.266$. הוסף המשתנה היקף המשרה. במבחן למובהקות המשתנה הנוסף התקבל: $t_{\beta} = 0.456$. האם ערך $\bar{R}^2$ יעלה/ירד/לא ישתנה בהוספת המשתנה הנוסף למודל?

מבחן Wald ו-T מורכב:

- (3) נאמד המודל: $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבלו התוצאות הבאות:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	646169.84	161542.46	51835.84	0.0000
Error	199	620.1683	3.1164236		
C Total	203	646790.01			

Root MSE	1.7653395	R-square	0.999041
Dep Mean	178.6645	Adj R-sq	0.999022
C.V.	0.988075		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	5.067731	0.456604	11.09874	0.0000
X	1	0.975736	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-68.75141	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

הועלתה ההשערה כי ההשפעה על Y של משתנה S היא פי 3 מזו של משתנה Z, וכן כי החותך הוא 5.

א. מהי השערת האפס?

ב. מהו המודל המוגבל שאותו צריך לאמוד?

להלן אמידת המודל המוגבל:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	646166.01	323083.01		
Error	201	623.9983	3.104469		
C Total	203	646790.01			

Root MSE	1.7619504	R-square	0.999035
Dep Mean	173.6645	Adj R-sq	0.999026
C.V.	1.0145714		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
X	1	0.978491	0.036399	26.88240	0.0000
Z+3S	1	2.999995	0.003669	817.6080	0.0000
W	1	-5.043109	0.071218	-70.81249	0.0000

ג. חשב את הסטטיסטי של W.L.D.

ד. כמה דרגות חופש יש במונה וכמה במכנה?

ה. האם דוחים או מקבלים את השערת האפס?

(4) על מנת לאמוד את פונקציית התצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה: $C_t = \alpha + \beta_1 \cdot W_t + \beta_2 \cdot P_t + u_t$.
להלן תוצאות האמידה של המשוואה הנ"ל:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	---	-----	-----	-----	-----
Error	---	-----	52968		
C Total	---	-----			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	-107.226	-----	-----	-----
W	1	0.743	-----		
P	1	0.561	-----		

Covariance of Estimates

COV	INTERCEP	W	P
INTERCEP	-----	-----	-----
W	-----	0.0046	-0.0090
P	-----	-0.0090	0.016

על מנת לבדוק את ההשערה שהנטייה השולית לצרוך מתוך ההכנסה זהה לנטייה השולית לצרוך מתוך ההון, נאמדה גם המשוואה הבאה:
 $C_t = \alpha + \beta_1 \cdot Y_t + u_t$, כאשר: Y_t = סה"כ ההכנסה של משק בית t.
התקבל: $ESS = 0.4566$.
בדקו את ההשערה בשתי דרכים.

תרגיל מסכם:

- 5) חוקר אמד את התצורות של 500 משקי בית כפונקציה של הכנסה שלהן לפי המשוואה: $EXPENSE_t = \alpha + \beta \cdot INCOME_t + u_t$.
- $EXPENSE_t$ - התצורות של משק הבית ה-t-י באלפי שקלים.
- $INCOME_t$ - ההכנסה של משק הבית ה-t-י באלפי שקלים.
- ההפרעות האקראיות מקיימות את כל ההנחות הקלאסיות התקבל הפלט הבא:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2013.105	2013.105	6495.745	0.0000
Error	498	154.3358	0.3099112		
C Total	499	2167.441			

Root MSE	0.556697	R-square	0.928794
Dep Mean	3.990208	Adj R-sq	0.928651
C.V.	13.95157		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	0.041995	0.054951	0.764236	0.4451
INCOME	1	0.713503	0.008853	80.59618	0.0000

- מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?
- מהו אחוז השונות בתצורות המוסבר ע"י ההכנסה?
- מהו אומדן לתצורות ההתחלתית של משק בית?
- האם אומדן זה מובהק?
- על עוזר מחקר הטיל החוקר לבדוק את ההשערה כי על כל 1000 ₪ נוספים בהכנסה צורך הפרט 700 ₪, כנגד ההשערה כי הוא צורך יותר מ-700 ₪. נסח את השערת האפס ואת ההשערה האלטרנטיבית.
- מהו הסטטיסטי t לבדיקת ההשערה?
- מהו הסטטיסטי WALD לבדיקת ההשערה?
- התברר כי הייתה טעות בנתונים, וכי יש להוסיף 1000 ₪ לתצורות של כל משק בית:
- ההוספה תגדיל את האומד ל- α : נכון / לא נכון / לא ניתן לדעת.

- ii. בעקבות ההוספה האומד ל- α יהיה מובהק : נכון / לא נכון / לא ניתן לדעת.
- iii. ההוספה תשנה את האומד ל- β : נכון / לא נכון / לא ניתן לדעת.
- iv. ההוספה תשנה את R^2 : נכון / לא נכון / לא ניתן לדעת.

החוקר טען כי יש להוסיף לפונקציית התצרוכת גם את השפעת העושר. העושר של משק בית מורכב מתוכניות החסכון שלו (SAVINGS) ומניירות הערך שיש לו (NE). שתי סדרות הנתונים הן באלפי שקלים. החוקר אמד את המשוואה :

$$EXPENSE_t = \alpha + \beta_1 \cdot INCOME_t + \beta_2 \cdot SAVINGS_t + \beta_3 \cdot NE_t + u_t$$

וקיבל כי סכום ריבועי הסטיות של הטעויות הוא 121.

ט. מהי השערת האפס לבדיקת הטענה של החוקר (שהמודל החדש נכון ולא המקורי)?

י. מהו הסטטיסטי WALD לבדיקת ההשערה?

החוקר רצה לבדוק את ההשערה כי הנש"צ מתוך ההכנסה שווה ל-0.6 וכי השפעת ניירות הערך על התצרוכת היא פי 2 מהשפעת תוכניות החסכון.

יא. מהי השערת האפס לבדיקה זו?

יב. המודל המוגבל לבדיקת ההשערה יהיה מהצורה : $Z_t = \gamma_0 + \gamma_1 W_t + v_t$.

בטא את Z_t ו- W_t באמצעות המשתנים המקוריים.

תשובות סופיות:

(1) א.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	646169.84	161542.46	51835.84	0.0000
Error	199	620.1683	3.1164236		
C Total	203	646790.01			
Root MSE		1.7653395	R-square	0.999041	
Dep Mean		178.6645	Adj R-sq	0.999022	
C.V.		0.988075			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	5.067731	0.456604	11.09874	0.0000
X	1	0.975736	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-68.75141	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

ב. יש עדות לכך. ג. יש עדות לכך.

(2) ירד.

(3) א. $H_0: \alpha = 5, \beta_s = 3\beta_z$ ב. $Y_t - 5 = \beta_x X_t + \beta_z (Z_t + 3S_t) + \beta_w W_t + u_t$ ג. $WALD_{stat} = 0.6145$ ד. מונה: 2, מכנה: 199.

ה. מקבלים.

(4) בדיקה ע"י מבחן WALD ו-t: אין עדות לכך.

(5) א. $PF = 0.000$ ב. 92% ג. $\hat{\alpha} = 0.04195$ ד. לא.ה. $H_0: \beta = 0.70, H_1: \beta > 0.70$ ו. $t_{\hat{\beta}} = 1.583$ ז. $WALD_{stat} = 2.505$

ח. i. נכון. ii. נכון. iii. לא נכון. iv. לא נכון.

ט. $H_0: \beta_2 = \beta_3 = 0, H_1: OTHERWISE$ י. $WALD_{stat} = 68.32$ יא. $H_0: \beta_3 = 2 \cdot \beta_2, \beta_1 = 0.6$ יב. $W_t = SAVINGS_t + 2 \cdot NE_t, Z_t = EXPENCE_t - 0.6 \cdot INCOME_t$

אקונומטריקה א

פרק 11 - משתנה דמי

תוכן העניינים

1. כללי 57

משתנה דמי:

רקע:

הכנסת משתנים ב"ת איכותיים למודל הרגרסיה.

למשל, נתונה משוואת הרגרסיה: $W_t = \alpha + \beta \cdot S_t$.

W_t = השכר (התלוי).

S_t = שנות לימוד (הב"ת) שניהם כמותיים.

נניח שאנו סבורים שגם משתנה המגדר (משתנה איכותי) משפיע על השכר.

כדי להכניסו למשוואת הרגרסיה יש להגדיר משתני דמי (dummy variable):

נגדיר משתנה D שיקבל את הערך 0 אם מדובר ב"אישה" ואת הערך 1 אם מדובר ב"גבר". ניתן להכניס את משתנה הדמי למודל בשלושה אופנים שונים:

1. משתנה דמי לחותך – המגדר משפיע על השכר ההתחלתי בלבד.
2. משתנה דמי לשיפוע – המגדר משפיע על התוספת לשכר בגין שנות הלימוד.
3. משתנה דמי לכל הפונקציה – המגדר משפיע גם על החותך וגם על השיפוע.

משתנה דמי לחותך:

המין משפיע על השכר ההתחלתי בלבד.

המודל: $W_t = \alpha_0 + \alpha_1 D + \beta \cdot S_t + u_t$ החותך מייצג כאן את השכר ההתחלתי.

שכר ההתחלתי של אישה: α_0 .

שכר התחלתי של גבר: $\alpha_0 + \alpha_1$.

הבדל בשכר בין נשים וגברים: α_1 (הפרש בין החותכים).

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש החותכים: $H_0: \alpha_1 = 0$.

- השיפוע מייצג את התוספת בשכר כפונקציה של מס' שנות הלימוד והוא זהה עבור נשים וגברים.

פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

המגדר הוא המשתנה היחיד במשוואה: $W_t = \alpha_0 + \alpha_1 D + u_t$.

החותך מייצג כאן את השכר הממוצע עבור כל קטגוריה:

שכר הממוצע של אישה: α_0 .

שכר הממוצע של גבר: $\alpha_0 + \alpha_1$.

הבדל בשכר הממוצע בין נשים וגברים: α_1 (הפרש בין החותכים).

בדיקת השערות על משתנה הדמי: מבחן t : $H_0: \alpha_1 = 0$ (מבחן זהה למבחן t להבדל בין ממוצעים).

משתנה דמי לשיפוע:

- המגדר משפיע על התוספת לשכר בגין שנות הלימוד: $W_t = \alpha + \beta_0 S_t + \beta_1 DS_t + u_t$.
 השיפוע מייצג כאן את התוספת לשכר בגין שנות לימוד.
 אצל אישה: התוספת לשכר בגין שנות לימוד: β_0 .
 אצל גבר: התוספת לשכר בגין שנות לימוד: $\beta_0 + \beta_1$.
 הבדל בין גברים לנשים בתוספת לשכר בגין שנות הלימוד: β_1 (הפרש השיפועים).
 בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש השיפועים: $H_0: \beta_1 = 0$.
 • החותך, המייצג את השכר ההתחלתי, יהיה זהה עבור גברים ונשים.

משתנה דמי לכל הפונקציה:

- המין משפיע גם על החותך וגם על השיפוע – גם על השכר ההתחלתי וגם על התוספת לשכר ההתחלתי בגין שנות הלימוד.
 המודל: $W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 DS_t + u_t$.
 השכר ההתחלתי של אישה: α_0 .
 השכר ההתחלתי של גבר: $\alpha_0 + \alpha_1$.
 הבדל בשכר ההתחלתי בין המינים: α_1 (הבדל בחותכים).
 אצל אישה: התוספת לשכר בגין שנות הלימוד: β_0 .
 אצל גבר: התוספת לשכר בגין שנות הלימוד: $\beta_0 + \beta_1$.
 הבדל בין המינים בתוספת לשכר בגין שנות הלימוד: β_1 (הבדל בשיפועים).

2 דרכים לבדיקה האם יש השפעה למשתנה האיכותי:

1. בדיקת השערות למשתני הדמי:
 באמצעות מבחן WALT יש לבדוק: $H_0: \alpha_1 = \beta_1 = 0$.
 לפחות אחד הפרמטרים שונה מ-0: H_1 .
 אם דוחים את השערת האפס, יש לבצע מבחני t עבור כל אחד מהפרמטרים
 בנפרד: $H_0: \alpha_1 = 0$ ו- $H_0: \beta_1 = 0$.
2. מבחן CHOW:
 דרך נוספת לבדיקת ההבדל בין הקטגוריות בלא יצירת משתני דמי:
 חלוקת המדגם לפי הקטגוריות של המשתנה האיכותי.
 מדגם של גברים (T_m) ושל נשים (T_f).
 עבור כל קבוצה לאמוד משוואות רגרסיה לניבוי שכר על ידי שנות לימוד:
 נשים: $W_t = \alpha_f + \beta_f X_t + u_t$.
 גברים: $W_t = \alpha_m + \beta_m X_t + u_t$.
 השערות: $H_0: \alpha_f = \alpha_m; \beta_f = \beta_m$.

לבדיקת ההשערה נשתמש במבחן CHOW (הזהה למבחן WALT) :
המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן יכול להיות
המדגם המאוחד : $W_t = \alpha + \beta X_t + u_t$

המודל הלא מוגבל (U) כולל את שני חלקי המדגם :
 $ESS_U = ESS_f + ESS_m$
 $DF_U = DF_f + DF_m$

$$CHOW_{stat} = \frac{\frac{ESS_R - (ESS_f + ESS_m)}{DF_R - (DF_f + DF_m)}}{\frac{ESS_f + ESS_m}{DF_f + DF_m}} = WALT_{stat}$$

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה :

1. אם דחינו את H_0 במבחן CHOW נתקשה לברר את מקור ההבדל שנמצא.
2. בהרצת שתי גרסיות נפרדות אנו בודקים הבדל בכל הפונקציה ואילו שיטת משתני הדמי מאפשרת לבדוק הבדל רק בחותך או רק בשיפוע.

סיכום ביניים :

משתנה דמי לכל הפונקציה	משתנה דמי לשיפוע	משתנה דמי לחותך	
$Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$	המודל
קיים הבדל בין הקטגוריות במשוואת הרגרסיה כולה (בחיתוך ושיפוע).	קיים הבדל בין הקטגוריות בתוספת ל-Y בגין X (בשיפוע).	קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחיתוך).	ההשערה במילים
מבחן WALT להפרש בין הפונקציות (החותכים והשיפועים) : $H_0 : \alpha_1 = \beta_1 = 0$ **ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW. אם דוחים את H_0 יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב- $H_0 : \alpha_1 = 0$ WALT) : $H_0 : \beta_1 = 0$	מבחן t להפרש השיפועים : $H_0 : \beta_1 = 0$	מבחן t להפרש החותכים : $H_0 : \alpha_1 = 0$	בדיקת ההשערה

משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים:

כאשר המשתנה האיכותי כולל יותר משני ערכים/קטגוריות נגדיר מס' משתני דמי כמספר הקטגוריות פחות אחד.

למשל, את המשתנה האיכותי של עונות השנה הכולל 4 ערכים: אביב, קיץ, סתיו, חורף נייצג באמצעות 3 משתני דמי:

D_1 יקבל את הערך 1 אם מדובר באביב ו-0 אחרת.

D_2 יקבל את הערך 1 אם מדובר בקיץ ו-0 אחרת.

D_3 יקבל את הערך 1 אם מדובר בסתיו ו-0 אחרת.

אם מדובר בחורף אז כל משתני הדמי יקבלו את הערך 0 ולכן החורף היא קבוצת הייחוס. נניח שאנו רוצים לבדוק עונתיות במחירי הירקות:

$$V_t = \text{מדד מחירי הירקות.}$$

$$p_t = \text{מדד המחירים לצרכן.}$$

1. משתני דמי לחותך:

הטענה: יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות.

$$\text{המודל: } V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta \cdot P_t + u_t$$

כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות

ב- β . למחיר זה יתווסף α_0 בחורף, $\alpha_0 + \alpha_1$ באביב, $\alpha_0 + \alpha_2$ בקיץ ו- $\alpha_0 + \alpha_3$ בסתיו.

ניתן לראות כי: α_0 - החותך בקטגוריה שהושמטה, $\alpha_0 + \alpha_1$ - החותך בקטגוריה i.

בדיקת השערות:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$H_1: \text{OTHERWISE}$$

המבחן הסטטיסטי – מבחן WALD:

$$(U) \quad V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta \cdot P_t + u_t$$

$$(R) \quad V_t = \alpha + \beta \cdot P_t + u_t$$

- שימו לב שהחותך במשוואה המוגבלת איננו α_0 שכן המשתנה המסביר של עונות השנה ירד.

אם נדחה את H_0 במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין החותכים על ידי מבחני t :

1. האם יש הבדל במחיר ההתחלתי של הירקות בין האביב לחורף:
 $H_0: \alpha_1 = 0$

2. האם יש הבדל במחיר ההתחלתי של הירקות בין הקיץ לחורף:
 $H_0: \alpha_2 = 0$

3. האם יש הבדל במחיר ההתחלתי של הירקות בין הסתיו לחורף:
 $H_0: \alpha_3 = 0$

2. משתני דמי לשיפוע:
 הטענה: יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן. המודל: $V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t$. המחיר ההתחלתי של הירקות שווה בין עונות השנה (α) אולם כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב: β_0 בחורף, $\beta_0 + \beta_1$ באביב, $\beta_0 + \beta_2$ בקיץ ו- $\beta_0 + \beta_3$ בסתיו. ניתן לראות כי- β_0 : השיפוע בקטגוריה שהושמטה $\beta_0 + \beta_i$: השיפוע בקטגוריה i.

בדיקת השערות:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

השערות:

$$H_1: \text{OTHERWISE}$$

המבחן הסטטיסטי – מבחן WALD:

$$V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t \quad (U)$$

$$V_t = \alpha + \beta \cdot P_t + u_t \quad (R)$$

- שימו לב שהשיפוע במשוואה המוגבלת איננו β_0 שכן המשתנה המסביר של עונות השנה ירד.

אם נדחה את H_0 במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין השיפועים על ידי מבחני t .

3. משתני דמי לכל הפונקציה :

הטענה : יש הבדל בין עונות השנה בפונקציית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן.
המודל :

$$V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t$$

בדיקת השערות :

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$$

המבחן הסטטיסטי - מבחן Wald :

(U) :

$$V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t$$

$$V_t = \alpha + \beta \cdot P_t + u_t \quad (R)$$

אם דוחים את H_0 , יש לבדוק במבחן Wald האם ההבדל הוא בין החותכים

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0, H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$$

אם דוחים את H_0 יש להמשיך לבדוק באמצעות מבחן t :

$$H_0 : \beta_j = 0, H_0 : \alpha_j = 0$$

משתני דמי עבור שני משתנים איכותיים :

לדוגמא – שני משתנים איכותיים המשפיעים על פונקציית השכר :
מגדר (אישה, גבר) וגזע (לבן, שחור).

נגדיר משתנה דמי G שיקבל 1 אם מדובר בגבר ו-0 אחרת (אישה).

נגדיר משתנה דמי R שיקבל 1 אם מדובר בלבן ו-0 אחרת (שחור).

נבדוק כיצד מגדר וגזע משפיעים על השכר ההתחלתי (החותך), כאשר השכר תלוי

גם בשנות לימוד (S_t) .

1. הבדל בחותך ללא אינטראקציה :

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t$$

המודל זה – אין השפעה משולבת של מגדר וגזע על השכר ההתחלתי.

ניתן לבדוק השערות על כל אחד מהמשתנים הב"ת האיכותיים בנפרד :

$$1. H_0 : \alpha_1 = 0 \quad \text{הבדל בשכר ההתחלתי בין גברים לנשים :}$$

$$2. H_0 : \alpha_2 = 0 \quad \text{הבדל בשכר ההתחלתי בין שחורים ללבנים :}$$

2. הבדל בחותך עם אינטראקציה :

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$$

המודל : $W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$.
 במודל זה הטענה היא כי קיימת, בנוסף להשפעה של מגדר וגזע בנפרד על השכר, גם השפעה משולבת (אינטראקציה) של מגדר וגזע על השכר ההתחלתי.
 במודל זה, לעומת הקודם, נוספת ההשערה לבדיקת השפעת האינטראקציה בין מגדר לגזע על השכר ההתחלתי :

$$H_0 : \alpha_3 = 0$$

3. דרך נוספת ליצירת מודל עם אינטראקציה :

הגדרת משתני דמי המייצגים שילוב בין המשתנים האיכותיים גזע ומגדר באופן הבא :

D_1 יקבל 1 אם מדובר בגבר לבן ו-0 אחרת.

D_2 יקבל 1 אם מדובר בגבר שחור ו-0 אחרת.

D_3 יקבל 1 אם מדובר באשה לבנה ו-0 אחרת.

הנשים השחורות מהוות כאן את קבוצת הייחוס.

$$W_t = \gamma_0 + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \delta \cdot S_t + u_t$$

נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה :

הפרש	אישה	גבר	
$\gamma_1 - \gamma_3$	$\gamma_0 + \gamma_3$	$\gamma_0 + \gamma_1$	לבן
γ_2	γ_0	$\gamma_0 + \gamma_2$	שחור
	γ_3	$\gamma_1 - \gamma_2$	הפרש

ההשערות לבדיקת קיום האינטראקציה : $H_0 : \gamma_1 - \gamma_3 = \gamma_2$ או $H_0 : \gamma_1 - \gamma_3 = \gamma_2$ התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך

$$WALD = t^2$$

$$PF = Pt$$

שאלות:

משתנה דמי לחותך:

(1) על בסיס מדגם של 50 איש העובדים בחברה מסוימת התקבלו התוצאות הבאות:

$$W_t = 5500 + 1043 \cdot D + 119 \cdot S_t$$

$$(S.E) \quad (134) \quad (56) \quad (24)$$

המספרים בסוגריים הם טעויות התקן של מבחני המובהקות לפרמטרים.

א. מהו השכר ההתחלתי של גבר בעל 12 שנות לימוד?

ב. מה ההבדל בשכר ההתחלתי בין גברים לנשים?

ג. האם הבדל זה מובהק באוכלוסייה?

ד. בדקו את הטענה כי השכר ההתחלתי של גברים גבוה ביותר מ-500 ₪ מזה של נשים.

ה. בדקו את הטענה שהשכר ההתחלתי של נשים נמוך ב-600 ₪ מזה של גברים.

פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

(2) על אותו המדגם של 50 איש העובדים בחברה מסוימת ביקש החוקר לבדוק

האם יש הבדל בשכר הממוצע בין גברים לנשים.

$$W_t = 5200 + 1120 \cdot D$$

$$S_{\hat{\alpha}_1} = 63$$

בדקו האם קיים הבדל מובהק בשכר הממוצע בין נשים וגברים?

משתנה דמי לשיפוע:

(3) על בסיס אותו מדגם, ביקש החוקר לדעת האם קיים הבדל מובהק בין גברים

לנשים בתוספת לשכר בגין שנות הלימוד.

תוצאות האמידה נתונות להלן:

$$W_t = 5000 + 110 \cdot S_t + 120 \cdot D \cdot S_t + u_t$$

$$(68) \quad (23) \quad (25)$$

בדוק את ההשערה.

משתנה דמי לכל פונקציה:

(4) חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה. החוקר בדק האם הפונקציה של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את ארבע המשוואות הבאות:

$$1. \quad NUM_t = \gamma_1 + \delta_1 \cdot AVGD_t + \varepsilon_{1t} \quad \text{כבישים מהירים בלבד.}$$

$$2. \quad NUM_t = \gamma_2 + \delta_2 \cdot AVGD_t + \varepsilon_{2t} \quad \text{כבישים לא מהירים בלבד.}$$

$$3. \quad NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t} \quad \text{שני סוגי הכביש (כל המדגם).}$$

$$4. \quad NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_2 \cdot AVGD_t + \beta_3 \cdot (AVGD \cdot TYPE)_t + U_t$$

כאשר:

NUM_t - מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה.

$AVGD_t$ - נפח התנועה בקטע כביש t ליום באלפים.

$TYPE_t$ - משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר, ו-0 כאשר הכביש לא מהיר.

תוצאות אמידת המשוואות מופיעות בהמשך השאלה.

א. בדקו את טענת החוקר בשתי דרכים שונות. ציינו איזה מן המשוואות רלוונטיות עבור כל דרך.

ב. חשבו את הערכים המספריים עבור אומדני משוואה (4).

ג. מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על ארבעת מכוניות ליום בקטע הכביש האמור?

הועלתה הטענה כי המקדם להשפעה של נפח התנועה בדרכים מהירות הינו כפול מזה שבדרכים לא-מהירות.

ד. מהי השערת האפס לבדיקת הטענה (במונחי משוואה (4))?

ה. מהי הרגרסיה "תחת" H_0 למבחן WALT?

משוואה (1) - כבישים מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 344

Number of Observations Used 344

Analysis of Variance

Source	DF	Sum of	Mean	F Value	Pr > F
		Squares	Square		
Model	1	4700.81174	4700.81174	89.12	<.0001
Error	342	18039	52.74684		
Corrected Total	343	22740			

Root MSE	7.26270	R-Square	0.2067
Dependent Mean	5.10465	Adj R-Sq	0.2044
Coeff Var	142.27617		

Parameter Estimates

Variable	DF	Parameter	Standard	t Value	Pr > t
		Estimate	Error		
Intercept	1	1.55289	0.54303	2.86	0.0045
avgd	1	0.02098	0.00222	9.44	<.0001

משוואה (2) - כבישים לא מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 410

Number of Observations Used 410

Analysis of Variance

Source	DF	Sum of	Mean	F Value	Pr > F
		Squares	Square		
Model	1	971.99073	971.99073	145.83	<.0001
Error	408	2719.34830	6.66507		
Corrected Total	409	3691.33902			
Root MSE					
			2.58168	R-Square	0.2633
Dependent Mean					
			1.38780	Adj R-Sq	0.2615
Coeff Var					
			186.02612		

Parameter Estimates

Variable	DF	Parameter	Standard	t Value	Pr > t
		Estimate	Error		
Intercept	1	0.14978	0.16360	0.92	0.3605
avgd	1	0.02877	0.00238	12.08	<.0001

משוואה (3) - שני סוגי הכביש (כל המדגם):

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 754

Number of Observations Used 754

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8052.00804	8052.00804	288.84	<.0001
Error	752	20964	27.87730		
Corrected Total	753	29016			

Root MSE 5.27990 R-Square 0.2775

Dependent Mean 3.08355 Adj R-Sq 0.2765

Coeff Var 171.22758

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.73903	0.23665	3.12	0.0019
avgd	1	0.02330	0.00137	17.00	<.0001

משוואה (4):

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 754

Number of Observations Used 754

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8256.966	2752.322	99.44	<.0001
Error	750	20759	27.678		
Corrected Total	753	29016			

Root MSE	5.26102	R-Square	0.2846
Dependent Mean	3.08355	Adj R-Sq	0.2817
Coeff Var	170.61553		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.14978	0.33340	0.45	0.6534
type	1				0.0067
avgd	1				<.0001
avgdtype	1				0.1283

משתנה איכותי עם יותר משתי קטגוריות:

(5) ענה על הסעיפים הבאים:

- א. הועלתה הטענה כי יש הבדל במחיר ההתחלתי בין האביב לקיץ.
 i. מהי השערת האפס לבדיקת הטענה?
 ii. פרטו שני מבחנים סטטיסטיים בעזרתם ניתן לבדוק את הטענה.
- ב. הועלתה הטענה כי יש רק שתי עונות המשפיעות על מחיר הירקות ההתחלתי: קיץ + אביב, חורף + סתיו.
 i. מהי השערת האפס לבדיקת הטענה?
 ii. מהו המבחן הסטטיסטי המתאים? פרטו.

משתנה דמי עבור שני משתנים איכותיים:

(6) חוקר בדק השפעות של השכלה, גזע (שחור, לבן) וניסיון (EXP) על לוג השכר ($\ln(Y)$) במדגם בן 306 תצפיות:

$$\ln(Y)_t = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta_1 EXP_t + \beta_2 EXP_t^2 + u_t$$

$\ln(Y)$ - לוג השכר.

EXP - שנות ניסיון.

D_1 - מקבל את הערך 1 עבור שחורים בעלי השכלה גבוהה (ו-0 אחרת).

D_2 - מקבל את הערך 1 עבור שחורים בעלי השכלה נמוכה (ו-0 אחרת).

D_3 - מקבל את הערך 1 עבור לבנים בעלי השכלה גבוהה (ו-0 אחרת).

תוצאות אמידת משוואת הרגרסיה מוצגות בפרט להלן:

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	-----	-----	-----	-----
Error	300	140	-----		
Corrected Total	305	210			
Root MSE		-----	R-Square	-----	
Dependent Mean		-----	Adj R-Sq	-----	
Coeff Var		-----			

Parameter Estimates

Variable	DF	Parameter		t Value	Pr > t
		Estimate	Standard Error		
Intercept	1	-----	-----	60.84	0.00
D1	1	-----	-----	-3.20	0.00
D2	1	-----	-----	-5.56	0.00
D3	1	-----	-----	7.23	0.00
EXP	1	-----	-----	8.11	0.00
EXP ²	1	-----	-----	-7.45	0.00

- א. לפי המשוואה הניסיון זהה עבור שחורים ולבנים :
- ב. בדוק את הטענה כי בקרב אנשים בעלי השכלה נמוכה אין השפעה לגזע.
- ג. בדוק את הטענה כי אין השפעות השכלה בקרב לבנים.
- ד. מהי השערת האפס לבדיקת הטענה כי אין אינטראקציה בין גזע להשכלה?
- ה. לבדיקת ההשערה של הסעיף הקודם בוצע מבחן WALT. הרגרסיה המוגבלת תחת השערת האפס הינה :
- $$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + v$$
- מהם ה-Zים?
- ו. בדוק את ההשערה אם ידוע שבמודל המוגבל $R^2 = 0.33$.
- ז. החוקר החליט לאמוד במקום את המשוואה המקורית את המשוואה :
- $$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \lambda_3 (S \cdot E) + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$
- כאשר :
- S מקבל את הערך 1 עבור שחורים ו-0 אחרת (לבנים).
- E מקבל את הערך 1 עבור השכלה גבוהה ו-0 אחרת (השכלה נמוכה).
- מה הקשר בין המקדמים של שני המודלים?
- ח. אם יאמוד החוקר את המשוואה :
- $$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$
- ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד', ו' ו-ז').

(7) חוקרת בדקה השפעות השכלה, מגדר וניסיון על הכנסה מעבודה לפי המשוואה הבאה :

$$\ln(MWAGE) = \alpha_0 + \alpha_1 \cdot S + \alpha_2 \cdot E + \alpha_3 \cdot (S \cdot E) + \beta_0 \cdot EXP + \beta_1 \cdot (EXP \cdot S) + \beta_2 \cdot (EXP \cdot S) + \beta_3 \cdot (EXP \cdot S \cdot E) + U$$

כאשר :

S משתנה דמי : 1 = עבור נשים, 0 = גברים.

E משתנה דמי : 1 = עבור השכלה גבוהה ($scl > 12$), 0 = השכלה נמוכה.

א. רשמו את הפונקציה לחישוב :

- i. תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו- 10 שנות ניסיון.
- ii. תחזית לוג השכר ההתחלתי עבור נשים משכילות.
- iii. לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות :

- i. אין השפעה של מגדר והשכלה על השכר.
- ii. השפעת ההשכלה אינה תלויה במגדר.
- iii. אין השפעות השכלה אצל גברים.
- iv. אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.

תשובות סופיות:

(1) א. $W_t = 7971$. ב. 1,043 נח. ג. כן. ד. יש עדות לכך.

ה. יש עדות לכך.

(2) יש עדות לכך.

(3) יש עדות לכך.

(4) א. יש עדות לכך, מבחן CHOW : 1, 2 ו-3, משתנה דמי : 3 ו-4.

ב. $\hat{\alpha} = 0.14978$, $\hat{\beta}_1 = 1.40311$, $\hat{\beta}_2 = 0.002877$, $\hat{\beta}_3 = -0.008$.

ג. $NUM_t = 1.532398$. ד. $H_0 : \beta_2 + \beta_3 = 2 \cdot \beta_2$.

ה. $H_0 : \beta_3 = \beta_2$.

ה. $NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_3 \cdot (AVGD_t + AVGD \cdot TYPE_t) + U_t$.

(5) א. i. $H_0 : \alpha_1 = \alpha_2$. ii. WALT ו-t. ב. i. $H_0 : \alpha_1 = \alpha_2, \alpha_3 = 0$.

ii. WALT .

(6) א. נכון. ב. יש עדות לכך. ג. יש עדות לכך.

ד. $H_0 : \alpha_3 = \alpha_1 - \alpha_2$ או $H_0 : \alpha_2 = \alpha_1 - \alpha_3$.

ה. $Z_0 = \ln(Y)_t$, $Z_1 = D_1 + D_3$, $Z_2 = D_2 - D_3$, $Z_3 = EXP_t$, $Z_4 = EXP_t^2$.

ו. אין עדות לכך. ז. $\lambda_0 = \alpha_0$, $\lambda_1 = \alpha_2$, $\lambda_2 = \alpha_3$, $\lambda_3 = \alpha_1 - \alpha_2 - \alpha_3$.

ח. לא.

(7) א. i. $\ln(MWAGE) = \hat{\alpha}_0 + \hat{\beta}_0 \cdot 10$.

ii. $\ln(MWAGE) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3$. iii. $EXP_t = \frac{-(\alpha_1 + \alpha_3)}{\beta_1 + \beta_3}$.

ב. i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$. ii. $H_0 : \alpha_3 = \beta_3 = 0$.

iii. $H_0 : \alpha_2 = \beta_2 = 0$. iv. $H_0 : \beta_2 + \beta_3 = 0$.

אקונומטריקה א

פרק 12 - מבחן 1

תוכן העניינים

1. כללי 74

מבחן 1:

שאלות:

- (1) חוקר רצה לבדוק את השפעת התל"ג על ההשקעה במשק לפי המודל הבא: $\ln I_t = \alpha + \beta \ln Y_t + u_t$, כאשר: I_t היא ההשקעה באלפי שקלים, Y_t הוא התוצר באלפי שקלים, וההרעה האקראית, u_t , מקיימת את כל ההנחות הקלאסיות. באמידה התקבל הפלט הבא:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.38523	0.38523	72.14	<.0001
Error	199	1.06266	0.00534		
C Total	200	1.44789			

Root MSE	0.073075	R-square	0.733936
Dep Mean	10.01722	Adj R-sq	0.732104
C.V.	0.729494		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T	95% conf. lim.
INTERCEPT	1	3.472013	0.85463	4.06259	0.0002	1.79 – 5.15
lnY	1	0.570042	0.06452	8.493526	0.0000	---- - ----

- מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?
 - אם נגדיל את התוצר ב-1% בכמה תגדל ההשקעה?
 - מהו רווח הסמך ל- α ? מהו רווח הסמך ל- β ?
 - הועלתה הטענה כי הגמישות שווה ל-0.4. מהן ההשערות לבדיקת הטענה?
 - מהי הרגרסיה המוגבלת למבחן WALT תחת H_0 ?
 - מהו הסטטיסטי של WALT למבחן זה (אם ניתן לחישוב)?
 - אם ההשקעה נמדדת בשקלים במקום באלפי שקלים:
- המקדם של $\ln Y$ לא ישתנה. נכון / לא נכון / לא ניתן לדעת
 - החותך לא ישתנה. נכון / לא נכון / לא ניתן לדעת

- iii. הסטטיסטי t לבדיקת המובהקות של β
לא ישתנה.
נכון / לא נכון / לא ניתן לדעת
- iv. הסטטיסטי F לבדיקת מובהקות המודל
לא ישתנה.
נכון / לא נכון / לא ניתן לדעת
- v. R^2 לא ישתנה.
נכון / לא נכון / לא ניתן לדעת

החוקר טען כי גם גודל האוכלוסייה, P , משפיע על ההשקעה לפי המודל
הבא: $\ln I_t = \alpha + \beta_1 \ln Y_t + \beta_2 \ln P_t + u_t$.
ח. מהי השערת האפס לבדיקת הטענה?

התקבל הפלט הבא:

Parameter Estimates

Variable	DF	Parameter	Standard	T for H0:	
		Estimate	Error	Parameter=0	Prob> T
INTERCEPT	1	1.131853	1.43547	0.788489	0.4435
lnY	1	1.035467	0.25756	4.020294	0.0004
lnP	1	-1.77456	0.94657	-1.874727	0.0736

- ט. באיזו רמת מובהקות נקבל את טענת החוקר?
י. R^2 של המשוואה החדשה קטן מזה של
המשוואה המקורית.
נכון / לא נכון / לא ניתן לדעת

- במשוואה החדשה הועלתה הטענה כי סכום הגמישויות שווה ל-0.
יא. מהי השערת האפס לבדיקת הטענה?
יב. מהו הסטטיסטי t לבדיקת ההשערה? (נתון כי: $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -0.25$).
יג. האם ניתן לדחות את השערת האפס?

(2) ענה על הסעיפים הבאים:

- א. ברגרסיה מרובה, כמו ברגרסיה חד משתנית,
מבחן F למובהקות המודל שווה לריבוע של
מבחן t למובהקות של β .
נכון / לא נכון / לא ניתן לדעת
- ב. אם הערך 0 נמצא בתוך רווח הסמך ל- β ,
אזי β מובהקת.
נכון / לא נכון / לא ניתן לדעת
- ג. בהוספת משתנה לא רלוונטי למודל האומד
המתוקן לפרופורציית השונות המוסברת
ירד בהכרח.
נכון / לא נכון / לא ניתן לדעת

- ד. אומדי הריבועים הפחותים אינם חסרי הטיה אם ידוע שהשונות של u_t אינה קבועה (הפרה של הנחה קלאסית).
 נכון / לא נכון / לא ניתן לדעת
- ה. אם דוחים H_0 ברמת מובהקות מסוימת, אזי דוחים H_0 בכל רמות המובהקות הקטנות יותר.
 נכון / לא נכון / לא ניתן לדעת
- ו. אומד חסר הטיה הוא אינו בהכרח אומד עקיב.
 נכון / לא נכון / לא ניתן לדעת

$$(3) \quad \tilde{\beta} = \frac{\sum_{t=1}^T X_t Y_t}{S_{XX}} \quad \text{נתון מודל ללא חותך: } Y_t = \beta X_t + u_t, \text{ ונתון האומד:}$$

- א. האומד $\tilde{\beta}$ הוא אר"פ.
 נכון / לא נכון / לא ניתן לדעת
- ב. האומד $\tilde{\beta}$ הוא אומד חסר הטיה.
 נכון / לא נכון / לא ניתן לדעת
- ג. האומד $\tilde{\beta}$ הוא אומד לינארי.
 נכון / לא נכון / לא ניתן לדעת
- ד. אר"פ יעיל יותר מ- $\tilde{\beta}$.
 נכון / לא נכון / לא ניתן לדעת
- ה. מהי השונות של $\tilde{\beta}$?

$$(4) \quad \hat{\beta} = \frac{\sum_{t=1}^T X_t Y_t}{\sum_{t=1}^T X_t^2} \quad \text{נתון מודל ללא חותך: } Y_t = \beta X_t + u_t, \text{ ונתון האומד:}$$

- א. האומד $\hat{\beta}$ הוא אר"פ.
 נכון / לא נכון / לא ניתן לדעת
- ב. האומד $\hat{\beta}$ הוא אומד חסר הטיה.
 נכון / לא נכון / לא ניתן לדעת
- ג. האומד $\hat{\beta}$ הוא אומד לינארי.
 נכון / לא נכון / לא ניתן לדעת
- ד. מהי השונות של $\hat{\beta}$?
- ה. האומד $\hat{\beta}$ הוא אומד עקיב.
 נכון / לא נכון / לא ניתן לדעת

תשובות סופיות:

$$(1) \quad \text{א. } PF = 0.0001 \quad \text{ב. } 0.57\% \quad \text{ג. } p(1.79 \leq \alpha \leq 5.15) = 0.95$$

$$\begin{aligned} & \text{ד. } \begin{aligned} H_0: \beta &= 0.4 \\ H_1: \beta &\neq 0.4 \end{aligned} \quad \text{ה. } p(0.026 \leq \beta \leq 1.11) = 0.95 \end{aligned}$$

$$\text{ו. } WALD_{stat} = 7.054 \quad \text{ז. } \ln I_t - 0.4 \ln Y_t = \alpha + u_t$$

$$\text{ח. } H_0: \beta_2 = 0 \quad \text{ט. } Pt_{\hat{\beta}} = 0.0736 \quad \text{י. לא נכון.} \quad \text{יא. } H_0: \beta_1 + \beta_2 = 0$$

$$(2) \quad \begin{aligned} & \text{יב. } t = -1.089 \quad \text{יג. אין סיבה מספקת.} \\ & \text{יד. לא נכון.} \quad \text{טו. לא נכון.} \quad \text{טז. לא נכון.} \quad \text{יז. לא נכון.} \quad \text{יח. לא נכון.} \end{aligned}$$

$$(3) \quad \text{א. לא נכון.} \quad \text{ב. לא נכון.} \quad \text{ג. נכון.} \quad \text{ד. לא ניתן לדעת.}$$

$$\text{ה. } V(\tilde{\beta}) = \frac{\sum X_t^2 \sigma^2}{S_{xx}^2}$$

$$(4) \quad \begin{aligned} & \text{א. נכון.} \quad \text{ב. נכון.} \quad \text{ג. נכון.} \quad \text{ד. } V(\tilde{\beta}) = \frac{\sigma_u^2}{\sum X_t^2} \end{aligned}$$

$$\text{ה. נכון.}$$

אקונומטריקה א

פרק 13 - מבחן 2

תוכן העניינים

1. כללי 78

מבחן 2:

שאלות:

- (1) חוקר בדק את השפעת שעות העבודה בשבוע (HOURS) על השכר החודשי ברוטו בשקלים (SALARY) לפי המודל: $SALARY_t = \alpha + \beta \cdot HOURS_t + u_t$. הסטייה המקרית מקיימת את כל ההנחות הקלאסיות. השלם את הפלט הבא, אם ידוע כי: $\bar{X} = 46.040873$, $S_{xx} = 35079$:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	---	---	---	---
Error	401	402271435	---		
C Total	---	449757359			
Root MSE	---		R-square	---	
Dep Mean	1580		Adj R-sq	---	
C.V.	---				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEPT	1	---	---	---	0.7476
HOURS	1	36.06745	---	---	0.0001

א. מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?

ב. מהו האומדן לשכר התחלתי?

החוקר רצה לבדוק את הטענה כי אם יעבוד שעה אחת נוספת בשבוע, שכרו יגדל ב-40 ₪.

ג. מהן ההשערות לבדיקת הטענה?

ד. מהו הסטטיסטי t למבחן?

ה. מהו הסטטיסטי WALT למבחן?

ו. מהי התחזית לשכר של עובד העובד 55 שעות בשבוע?

ז. החוקר טען כי יש לבדוק את הקשר בין השכר לשעות העבודה ע"י שימוש בנתונים שנתיים, כלומר, שכר שנתי (בהנחה שהשכר החודשי קבוע כל השנה) ושעות עבודה שנתיות (בהנחה ששעות העבודה קבועות בכל 52 השבועות בשנה). שימוש בנתונים שנתיים:

- i. ישנה את הסטטיסטי t לבדיקת המובהקות של α .
נכון / לא נכון / לא ניתן לדעת
- ii. יכפיל את האומד של β ב-0.23.
נכון / לא נכון / לא ניתן לדעת
- iii. יכפיל את סטית התקן של $\hat{\beta}$ ב-0.23.
נכון / לא נכון / לא ניתן לדעת
- iv. ישנה את Pvalue לבדיקת מובהקות המודל.
נכון / לא נכון / לא ניתן לדעת

החוקר טען כי יש להוסיף למשוואה גם את השפעת הגיל (AGE) ומספר שנות הלימוד (SCL). לשם כך הוא אמד את המשוואה הבאה:

$$SALARY_t = \alpha + \beta_1 \cdot HOURS_t + \beta_2 \cdot AGE_t + \beta_3 \cdot SCL_t + u_t$$

- ח. מהי השערת האפס לבדיקת הטענה?
- ט. מהו הנתון הנדרש כדי לחשב את הסטטיסטי של WALD לבדיקת טענת החוקר?
- י. בפלט האמידה של המשוואה החדשה לא היה ברור אם ערכו של נתון זה הוא 315968434 או 515968434 (בשל בעיה במדפסת). מהו הסטטיסטי של WALD לבדיקת טענת החוקר?
- יא. מהם הנתונים הנדרשים לחישוב הסטטיסטי t ?

החוקר רוצה לבדוק את הטענה כי השפעת ההשכלה על השכר גדולה פי 8 מהשפעת הגיל על השכר.

Parameter Estimates

Variable	DF	Parameter	Standard	T for H0:	
		Estimate	Error	Parameter=0	Prob> T
INTERCEPT	1	-1995.0275	331.7857	-6.013	0.0001
HOURS	1	36.408461	4.710021	7.730	0.0001
AGE	1	13.674254	3.816426	3.583	0.0004
SCL	1	109.93799	10.63745	10.335	0.0001

- יב. הנתונים בפלט אינם מספיקים לבדיקת ההשערה לפי מבחן t .
מהו הנתון החסר? באיזה פלט של SAS ניתן למצוא אותו?
- יג. בהנחה שנתון זה הוא 8.3969, חשב את הסטטיסטי t לבדיקת הטענה.
מהי מסקנתך לגבי נכונות הטענה?

יד. אם תרצה לבדוק את הטענה לפי מבחן WALD, יהיה המודל המוגבל:

$$Z_0 = \gamma_0 + \gamma_1 \cdot Z_1 + \gamma_2 \cdot Z_2 + v$$

$$Z_0 = \underline{\hspace{2cm}}$$

$$Z_1 = \underline{\hspace{2cm}}$$

$$Z_2 = \underline{\hspace{2cm}}$$

טו. אם יש מספיק נתונים, חשב את הסטטיסטי של WALD לבדיקת הטענה?

(2) נתון המודל: $Y_t = \alpha + \beta \cdot X_t + u_t$. ידוע כי כל ההנחות הקלאסיות מתקיימות.

$$\text{נתון האומד: } \tilde{\beta} = \frac{S_{XY}}{\sum_{t=1}^T X_t^2}$$

א. אומד זה הוא הפתרון של המשוואות

$$\text{הנורמליות: } \sum_{t=1}^T \hat{u}_t = 0, \sum_{t=1}^T \hat{u}_t X_t = 0$$

נכון / לא נכון / לא ניתן לדעת

ב. התוחלת של $\tilde{\beta}$ היא:

$$\text{i. } \beta$$

$$\text{ii. } \frac{\beta \cdot \sum_{t=1}^T X_t}{S_{XX}}$$

$$\text{iii. } \frac{\beta \cdot \sum_{t=1}^T X_t^2}{S_{XX}}$$

$$\text{iv. } \frac{\beta \cdot S_{XX}}{\sum_{t=1}^T X_t^2}$$

v. כל התשובות אינן נכונות.

ג. הטענה כי: $E(\tilde{\beta}) < \beta$:

i. תמיד נכונה.

ii. אינה נכונה.

iii. נכונה אם ורק אם: $\bar{X} > 0$.

iv. נכונה אם ורק אם: $\bar{X} \neq 0$.

v. כל התשובות אינן נכונות.

ד. אם $\bar{X} = 0$ אז השונות של $\tilde{\beta}$ היא:

$$\text{i. } \frac{\sigma^2}{\left(\sum_{t=1}^T X_t\right)^2}$$

$$\text{ii. } \frac{\sigma^2}{S_{XX}^2}$$

$$\text{iii. } \frac{\sigma^2 \sum_{t=1}^T X_t^2}{S_{XX}}$$

$$\text{iv. } \frac{\sigma^2}{S_{XX}}$$

v. כל התשובות אינן נכונות.

ה. אם $\bar{X} = 0$, אז $\tilde{\beta}$ הינו האומד הלינארי

חסר ההטיה בעל השונות הקטנה ביותר. נכון / לא נכון / לא ניתן לדעת

(3) נתון המודל: $Y_t = \alpha + \beta \cdot X_t + u_t$. ידוע כי כל ההנחות הקלאסיות מתקיימות.

נתון כי $\tilde{\beta}$ הוא אומד לינארי וחסר הטיה ל- β , אך איננו אומד עקיב ל- β .

מאחר ש- $\tilde{\beta}$ אינו אומד עקיב, לא נוכל להשתמש במשפט גאוס מרקוב ולקבוע

כי: $\hat{\beta} = \frac{S_{XY}}{S_{XX}}$ (ארי"פ) הינו אומד יעיל יותר.

נכון / לא נכון / לא ניתן לדעת

תשובות סופיות:

- (1) א. $PF = 0.00$ ב. $\hat{\alpha} = -80.5246$ ג. $H_0 : \beta = 40$
 $H_1 : \beta \neq 40$ ד. $t_{\hat{\beta}} = -0.75$ ה. $WALD_{stat} = 0.5625$ ו. $SALARY_t = 1903.16$
 ז. i. לא נכון. ii. נכון. iii. נכון. iv. לא נכון.
 ח. $H_0 : \beta_2 = \beta_3 = 0$ ט. SSE י. $WALD_{stat} = 54.49$ יב. Covariance of Estimates, $S^2_{\hat{\beta}_2, \hat{\beta}_3}$ יא. לא ניתן לחשב.
 $Z_0 = SALARY_t$ יג. נכונה. יד. $Z_1 = HOURS_t$ טו. $WALD_{stat} = 0.000324$
 $Z_2 = AGE_t + 8 \cdot SCL_t$ (2) א. לא נכון. ב. ה. ג. ה. ד. ה. ה. לא נכון.
 (3) לא נכון.

אקונומטריקה א

פרק 14 - מבחן 3

תוכן העניינים

1. רשימת שאלות.....83

מבחן 3:

שאלות:

1) על מנת לאמוד את פונקציית הייצור נאספו נתונים על 150 פירמות בשנת 2007 ונאמדה המשוואה הבאה:

$$1. \ln(Y_t) = \alpha + \beta_1 \cdot \ln(L)_t + U_t$$

כאשר:

$\ln(Y)_t$ - תפוקה שנתית באלפי ₪ בלוגים.

$\ln(L)_t$ - מספר העובדים בלוגים.

U_t - הטעות המקרית המקיימת את כל ההנחות הקלאסיות.

משוואה מס' 1 נאמדה בפלט מס' 1.

Dependent Variable: $\ln Y$

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	1	8.54211			0.0001
Error	35969	40.42584			
<u>C. Total</u>	35970	48.96795			
Root MSE	0.52264		R-square	0.1744	
Dep Mean	5.54003		Adj R-sq	0.1689	
C. V.	9.43380				

Parameter Estimates

Variable	D F	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	4.389949	0.21003743	20.901	0.0001
<u>lnL</u>	1	0.257487	0.04767276		0.0001

א. סטטיסטי F לבדיקת מובהקות המודל:

i. לא ניתן לחשב את סטטיסטי F בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: _____

ב. סטטיסטי t לבדיקת מובהקות המודל:

i. לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה

ii. לא ניתן לחשבו בעזרת הנתונים הקיימים.

iii. ניתן לחשבו וערכו הוא: _____

הועלתה הטענה כי עליה ב-1% במס' העובדים תגדיל את התפוקה בפחות מ-1%.

ג. ההשערות לבדיקת הטענה הן: H_0 : _____
 H_1 : _____

ד. הסטטיסטי לבדיקת הטענה הינו:

i. לא ניתן לחשבו בנתונים הקיימים.

ii. 5.5

iii. -5.5

iv. -15.5

v. 15.5

ה. הסטטיסטי של WALT לבדיקת הטענה:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: _____

ו. לאור התשובות לסעיפים הקודמים,

אחוז התפוקה קטן ככל שאחוז מס'

העובדים גדל: נכון/לא נכון/אי אפשר לדעת

החוקרת טענה כי יש משתנים נוספים המסבירים את תפוקת הפירמה ואמדה את המשוואה הבאה:

$$\ln(Y_t) = \alpha + \beta_1 \cdot \ln(L)_t + \beta_2 \cdot \ln(K)_t + \beta_3 \cdot \ln(PY)_t + U_t \quad 2.$$

כאשר:

$\ln(K)_t$ - מלאי ההון של הפירמה באלפי ש' בלוגים.

$\ln(PY)_t$ - הוצאות למחקר ופיתוח באלפי ש' בלוגים.

משוואה מס' (2) נאמדה בפלט מס' 2.

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	3	15.63370	5.21123	22.825	0.0001
Error	146	33.33425	0.22832		
C Total	149	48.96795			
Root MSE		0.47783	R-square	0.3193	
Dep Mean		5.54003	Adj R-sq	0.3053	
C. V.		8.62496			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	0.542062	1.66317350	0.326	0.7450
lnL	1	0.267771	0.08146608	3.287	0.0013
lnK	1	0.405694	0.09700769	4.182	0.0001
lnPY	1	0.406149	0.30781185	1.319	0.1891

ז. ההשערות לבדיקת הטענה הינן: H_0 : _____
 H_1 : _____

ח. הסטטיסטי של WALD לבדיקת הטענה הינו:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

ii. ניתן לחישוב וערכו: _____

ט. הסטטיסטי של t לבדיקת הטענה הינו:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

ii. לא ניתן לחשב סטטיסטי t לטענה מסוג זה

iii. ניתן לחישוב וערכו: _____

החוקרת טענה כי השפעת הוצאות למחקר ופיתוח אינה מובהקת ולכן יש
 לאמוד את המשוואה הבאה:

$$\ln(Y_t) = \alpha + \beta_1 \cdot \ln(L)_t + \beta_2 \cdot \ln(K)_t + U_t \quad 3.$$

כאשר:

משוואה מס' (3) נאמדה בפלט מס' 3.

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	2	15.23620	7.61810	33.199	0.0001
Error	147	33.73175	0.22947		
C Total	149	48.96795			

Root MSE	0.47903	R-square	0.3111
Dep Mean	5.54003	Adj R-sq	0.3018
C. V.	8.64667		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	2.681787	0.37024512	7.243	0.0001
lnL	1	0.177813	0.04470595	3.977	0.0001
lnK	1	0.465154	0.08612163	5.401	0.0001

Covariance of Estimates

COVB	INTERCEP	lnL	lnK
INTERCEP	0.1370814505	-0.003289697	-0.02723683
lnL	-0.003289697	0.0019986217	-0.001270417
lnK	-0.02723683	-0.001270417	0.0074169359

י. ההשערות לבדיקת הטענה הינן : H_0 : _____
 H_1 : _____

יא. הסטטיסטי של WALD לבדיקת הטענה הינו :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

ii. ניתן לחישוב וערכו : _____

הועלתה הטענה כי גמישות התפוקה ביחס להון גדולה פי 2 מגמישות התפוקה ביחס לעבודה.

בדקו את הטענה במשוואה (3).

יב. השערת האפס לבדיקת הטענה היא : H_0 : _____

יג. הסטטיסטי t לבדיקת הטענה הינו :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחישוב וערכו : _____

יד. הרגרסיה המוגבלת כאשר H_0 נכונה ("תחת H_0 ") למבחן WALD

$$Z_0 = \gamma_0 + \gamma_1 \cdot Z_1 + V$$

כאשר :

$$Z_1 = \underline{\hspace{2cm}}$$

$$Z_0 = \underline{\hspace{2cm}}$$

טו. הסטטיסטי של WALD לבדיקת הטענה (חשבי ישירות) :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחישוב וערכו : _____

טז. נטען כי אם נמדוד את המשתנים הב"ת

במודל בדולרים במקום בשקלים, האומדים

ל- β ול- α יישארו ללא שינוי

(הנח כי שער הדולר הוא 3.5 ₪) : נכון/לא נכון/אי אפשר לדעת

יז. נטען שאם נוריד את משתנה PY מהמודל

ה- $\bar{R}^2$ יעלה : נכון/לא נכון/אי אפשר לדעת

(2) ענו על כל השאלות הבאות. כל שאלה בפני עצמה. בכל השאלות מונח

המודל : $Y = \alpha + \beta X + U$ (ומתקיימות כל ההנחות הקלאסיות).

א. במודל לוגריתמי כפול β מייצגת את

שיעור השינוי השולי : נכון/לא נכון/אי אפשר לדעת

ב. במודל ללא חותך מתקיימת המשוואה

הנורמאלית : $\sum \hat{u}_i x_i = 0$ בלבד : נכון/לא נכון/אי אפשר לדעת

- ג. כאשר מוסיפים משתנה ב"ת למודל, עליה
ב- $\bar{R}^2$ מעידה על כך שהמשתנה שהוסף
מובהק באוכלוסייה:
נכון/לא נכון/אי אפשר לדעת
- ד. אם הנחה מס' 3 ($E(\hat{u}) = 0$ לכל t) איננה מתקיימת,
האומדים של המודל לא יהיו יעילים:
נכון/לא נכון/אי אפשר לדעת
- ה. ככל ש- S_{xx} גדול יותר, קל יותר לדחות
את H_0 למובהקות ה- β :
נכון/לא נכון/אי אפשר לדעת
- ו. $R^2 > \bar{R}^2$ מתקיים תמיד:
נכון/לא נכון/אי אפשר לדעת
- ז. מבחן F למובהקות המודל מהווה מקרה
פרטי של מבחן t למובהקות ה- β :
נכון/לא נכון/אי אפשר לדעת
- ח. ככל שגודל המדגם גדל כך האומד יהיה
יעיל יותר לפרמטר באוכלוסייה:
נכון/לא נכון/אי אפשר לדעת
- ט. ה- PVALUE גדל ביחס הפוך לרמת
המובהקות של המבחן (ה- α):
נכון/לא נכון/אי אפשר לדעת
- י. אם דחינו את H_0 במבחן t למובהקות ה- β כאשר
האומד חיובי, נדחה אותה בהכרח גם ביחס להשערה
כי מקדם השיפוע חיובי באוכלוסייה:
נכון/לא נכון/אי אפשר לדעת
- יא. אם ידוע כי הקשר בין X ל- Y מובהק
באוכלוסייה, הדבר מעיד בהכרח על
מובהקות המודל:
נכון/לא נכון/אי אפשר לדעת

$$(3) \quad Y_t = \beta X_t + U_t \quad \text{נתון המודל:}$$

$$\tilde{\beta} = \frac{\sum (X_t - \bar{X}) Y_t}{\sum (X_t - \bar{X})^2} \quad \text{נתון האומד:}$$

- א. $\tilde{\beta}$ הינו אומד חסר הטייה ל- β :
נכון/לא נכון/אי אפשר לדעת
- ב. שונותו של האומד: _____.

- ג. על סמך משפט גאוס מרקוב ניתן להסיק
כי אר"פ הינו אומד יעיל יותר מ- $\tilde{\beta}$:
נכון/לא נכון/אי אפשר לדעת
- ד. המשוואות הנורמאליות: $\sum \hat{u}_t = 0$
ו- $\sum \hat{u}_t X_t = 0$ הינן המשוואות לאמידת הפרמטרים
של המודל בשיטת הריבועים הפחותים:
נכון/לא נכון/אי אפשר לדעת
- ה. אם נתון ש: $\bar{X} = 0$ אזי $\tilde{\beta}$ הינו אומד
הריבועים הפחותים:
נכון/לא נכון/אי אפשר לדעת

תשובות סופיות:

- (1) א. ii , $F = 31.273$. ב. iii , $t = 5.5$. ג. $H_0 : \beta = 1$. ד. iv .
ה. i . ו. לא נכון . ז. ראו סרטון .
ח. ראו סרטון . ט. ראו סרטון . י. $H_0 : \beta_3 = 0$, $H_1 : \beta_3 \neq 0$.
יא. ii , $WALD_{stat} = 1.74$. יב. $H_0 : \beta_2 = 2 \cdot \beta_1$.
יג. ii , $t = 0.1417$. יד. $Z_0 = \ln(Y)_t$, $Z_1 = \ln(L)_t + 2 \ln(K)_t$.
טו. ii , $WALD_{stat} = 0.585$. טז. לא נכונה . יז. לא נכון .
(2) א. לא נכון . ב. נכון . ג. לא נכון . ד. לא נכון .
ה. נכון . ו. נכון . ז. נכון . ח. נכון . ט. לא נכון .
י. נכון . יא. נכון .
(3) א. נכון . ב. $V(\tilde{\beta}) = \frac{\sigma^2}{S_{xx}}$. ג. נכון . ד. לא נכון .
ה. נכון .

אקונומטריקה א

פרק 15 - מבחן 4

תוכן העניינים

1. רשימת שאלות.....89

מבחן 4:

שאלות:

- (1) בנק מעוניין לאמוד את סך הפעילות בכרטיסי אשראי של לקוחותיו. לשם כך אסף נתונים על 35,971 מלקוחותיו ואמד את המשוואה הבאה:

$$CREDIT_t = \alpha + \beta \cdot SAVINGS_t + U_t \quad 1.$$

כאשר:

$CREDIT_t$ - סך הפעילות בכרטיסי אשראי ב- t .

$SAVINGS_t$ - סך הפעילות בחשבונות חיסכון ב- t .

U_t - סטיה מקרית המקיימת את כל ההנחות הקלאסיות.

משוואה (1) נתונה בפלט מס' 1.

Dependent Variable: credit

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	---	---	---	---	<0.0001
Error	---	---	---		
C Total	---	---			
Root MSE	43859	R-square	0.0106		
Dep Mean	7433.60809	Adj R-sq	0.0106		
C. V.	589.99662				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0:	Prob> T	95% Confidence	
INTERCE							
P	1	11151.91516	394.35144	2.92	0.0035	378.97	1924.8
savings	1	0.56719	0.02884	19.67		0.51	0.623

א. סטטיסטי F לבדיקת מובהקות המודל הינו:

- לא ניתן לחשב את סטטיסטי F בעזרת הנתונים הקיימים.
- ניתן לחשבו וערכו הוא: _____.

ב. PVALUE של סטטיסטי t לבדיקת מובהקות ה- β :

- לא ניתן לחשבו בעזרת הנתונים הקיימים.
- לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה.
- ניתן לחשבו וערכו: _____.

הבנק טען שאם יגדילו לקוחותיו את הפעילות בחשבונות חיסכון שלהם אפילו בשקל אחד, הפעילות בכרטיסי אשראי תגדל ביותר מ 40 אגורות.

ג. ההשערות לבדיקת הטענה הינן :
 H_0 : _____
 H_1 : _____

ד. הסטטיסטי לבדיקת טענת הבנק הינו :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. הסטטיסטי לבדיקת הטענה צריך להיות שלילי.

iii. 19.67

iv. 5.797

ה. הסטטיסטי של WALD לבדיקת טענת הבנק :

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו : _____

ו. ברמת ביטחון של 95% מהו טווח הגידול בפעילות בכרטיסי אשראי, על כל שקל נוסף בפעילות בחשבונות חיסכון?

ז. ברמת ביטחון 95% מהו האומד לתוחלת פעילות בכרטיסי אשראי עבור סך פעילות בחשבונות חיסכון של 50,000 ₪?

ח. אם פעילות כרטיסי האשראי של כל לקוח תגדל ב- 1000 ₪ :

i. האומד של α ישתנה : נכון/לא נכון/ אי אפשר לדעת

ii. האומד של β ירד : נכון/לא נכון/ אי אפשר לדעת

iii. סטטיסטי F לבדיקת מובהקות

המודל לא ישתנה : נכון/לא נכון/ אי אפשר לדעת

נטען שסה"כ פעילות הלקוח בחשבונות חיסכון איננו המשתנה המשפיע על הפעילות בכרטיסי האשראי, אלא הרכב החסכונות.
 לשם כך נאמדה המשוואה הבאה :

$$CREDIT_t = \alpha + \beta_1 \cdot PIKADON1_t + \beta_2 \cdot PIKADON2_t + U_t \quad 2.$$

כאשר :

$PIKADON1_t$ - סה"כ הפקדה לפקדונות יומיים ב-₪.

$PIKADON2_t$ - סה"כ הפקדה לפקדונות חודשיים ב-₪.

משוואה (2) נאמדה בפלט מס' 2.

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	2	1.00791E12	5.003955E11	261.10	0.0001
Error	35968	6.893195E13	1916479937		
C Total	35970	6.993274E13			
Root MSE	43778		R-square	0.0143	
Dep Mean	7433.68809		Adj R-sq	0.0143	
C. V.	588.90847				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	1259.36230	379.00751	3.32	0.0009
Pikadon1	1	0.07552	0.05539	1.36	0.1728
Pikadon2	1	0.72350	0.03199	22.62	0.0001

Covariance of Estimates

COVB	INTERCEP	Pikadon1	Pikadon2
INTERCEP	143646.69097	-8.178835194	-9.154578973
Pikadon1	-8.176835154	0.0030678685	0.0003564263
Pikadon2	-9.15457897	0.0003564263	0.0010231462

ט. השערת האפס לבדיקת הטענה הינה: H_0 : _____.

י. הסטטיסטי של WALD לבדיקת הטענה:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו: _____.

יא. הסטטיסטי של t לבדיקת הטענה:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה.

iii. ניתן לחישוב וערכו: _____.

נטען שהגדלת הפעילות בחשבונות חיסכון של הלקוח על ידי העברה לפקדונות חודשיים משפיעה על הפעילות בכרטיסי אשראי פי 10 מאשר הגדלת הפעילות בחשבונות חיסכון על ידי העברה לפקדונות יומיים.

יב. השערת האפס לבדיקת הטענה הינה: H_0 : _____.

יג. הסטטיסטי t לבדיקת הטענה הינו:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: _____.

יד. PVALUE של סטטיסטי t מהסעיף הקודם:

i. לא ניתן לחשב את הסטטיסטי t בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: _____.

טו. הרגרסיה המוגבלת כאשר H_0 נכונה למבחן WALT

$$D_0 = \gamma_0 + \gamma_1 \cdot D_1 + \gamma_2 \cdot D_2 + v$$

$$D_0 : \underline{\hspace{2cm}}$$

$$D_1 : \underline{\hspace{2cm}} \text{ כאשר :}$$

$$D_2 : \underline{\hspace{2cm}}$$

טז. על פי משוואה מס' 2, כל שקל שיועבר

לפיקדון הראשון יוסיף כ-0.07552 ₪

לסה"כ הפעילות בכרטיסי אשראי :

נכון/לא נכון/אי אפשר לדעת

(2) ענו על השאלות הבאות (כל שאלה בפני עצמה, בכל שאלה מונח המודל: $Y = \alpha + \beta \cdot X + U$ ומתקיימות כל ההנחות הקלאסיות).

- א. אם המודל מובהק אזי שיפוע הרגרסיה מובהק בהכרח :
נכון/לא נכון/אי אפשר לדעת
- ב. הגמישות במודל חצי לוגריתמי היא קבועה :
נכון/לא נכון/אי אפשר לדעת
- ג. אם X_2 מהווה קומבינציה ליניארית של X_1 לא ניתן לאמוד את הרגרסיה המרובה : $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + U$:
נכון/לא נכון/אי אפשר לדעת
- ד. $\bar{R}^2 > R^2$ רק בתנאי שהמודל מובהק :
נכון/לא נכון/אי אפשר לדעת
- ה. ליניאריות וחוסר הטיה של האומדים מהווים תנאי הכרחי לעקיבותם :
נכון/לא נכון/אי אפשר לדעת
- ו. נתון כי רווח הסמך לאמידת β ברמת סמך של 95% הוא : [-2, -5].
מכך ניתן להסיק כי שיפוע הרגרסיה מובהק ברמת מובהקות של 5% :
נכון/לא נכון/אי אפשר לדעת
- ז. ככל שפיזור U_i גדול יותר כך קשה יותר לדחות את H_0 למובהקות המודל :
נכון/לא נכון/אי אפשר לדעת
- ח. מודלים לא ליניאריים מתארים קשרים שאינם ליניאריים בין המשתנה המסביר למוסבר :
נכון/לא נכון/אי אפשר לדעת.
- ט. אם הנחה 5 (שונויות קבועה) לא מתקיימת, אומדי הריבועים הפחותים אינם חסרי הטיה :
נכון/לא נכון/אי אפשר לדעת
- י. אם דחינו את H_0 לבדיקת הטענה כי שיפוע הרגרסיה הוא שלילי בוודאי שמודל הרגרסיה הוא מובהק :
נכון/לא נכון/אי אפשר לדעת

(3) נתון המודל: $Y_t = \beta \cdot X_t + U_t$, כאשר כל ההנחות הקלאסיות מתקיימות.

נתון האומד: $\tilde{\beta} = \frac{\sum Y_t}{S_{xx}}$.

א. $E(\tilde{\beta}) = \underline{\hspace{2cm}}$.

- ב. על סמך משפט גאוס מרקוב אומד זה יעיל פחות מאומד הריבועים הפחותים: נכון/ לא נכון/ אי אפשר לדעת
- ג. אומד $\tilde{\beta}$ מוגדר רק כאשר $S_x^2 \neq 0$: נכון/ לא נכון/ אי אפשר לדעת
- ד. חשבו את השונות של $\tilde{\beta}$ עבור מודל שבו $\alpha \neq 0$.
- ה. שונות האומד (שחושבה בסעיף הקודם) הינה גדולה משונות המודל הנתון: נכון/ לא נכון/ אי אפשר לדעת

תשובות סופיות:

- (1) א. ii, $F = 386.9089$, ב. iii, $PF < 0.0001 = Pt$, ג. $H_0: \beta = 0.4$, $H_1: \beta > 0.4$.
- ד. iv, ה. i, ו. $p(0.51 \leq \beta \leq 0.623) = 0.95$, ז. $p(-32,387,174.83 \leq E(Y) \leq 32,458,197.67) = 0.95$.
- ח. i. נכון, ii. לא נכון, iii. נכון, ט. $H_0: \beta_1 = \beta_2 = 0$, י. i, ii, $H_0: \beta_2 = 10 \cdot \beta_1$, יג. ii, $t = -0.0574$.
- יד. ii, $PVALUE > 0.1$, טו. $D_0: CREDIT_t$, $D_1: SAVINGS_t$, $D_2: PIKADON1_t + 10 \cdot PIKADON2_t$.
- טז. נכון.
- (2) א. נכון, ב. לא נכון, ג. נכון, ד. לא נכון, ה. לא נכון, ו. נכון, ז. נכון, ח. לא נכון, ט. לא נכון, י. לא נכון.
- (3) א. $E(\tilde{\beta}) = \frac{\beta \sum X_t}{S_{xx}}$, ב. לא ניתן לדעת, ג. נכון, ד. $V(\tilde{\beta}) = \frac{T\sigma^2}{S_{xx}^2}$, ה. לא נכון.

אקונומטריקה א

פרק 16 - מבחן 5

תוכן העניינים

1. רשימת שאלות.....95

מבחן 5:

שאלות:

1) על מנת לאמוד את הקשר בין רמת המחירים במשק (P) לכמות הכסף (M), נאספו נתונים חודשיים בשנים 86-94 (סה"כ 105 תצפיות) ונאמדה המשוואה הבאה:

$$1. M_t = e^\alpha + p^\beta + e^u$$

כאשר:

m - כמות הכסף במשק לחודש (מזומנים + עו"ש).

p - מדד המחירים לצרכן במשק.

U_t - סטיה מקרית המקיימת את כל ההנחות הקלאסיות.

משוואה מס' (1) נאמדה בפלט מס' 1.

Dependent Variable: lnm

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	1				<0.0001
Error	103				
C Total	104	44.91976			
Root MSE	0.09251		R-square	0.9804	
Dep Mean	8.53854		Adj R-sq	0.9802	
C. V.	1.08344				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCE					
P	1	1.49372	0.09862	15.15	<.0001
lnp	1	1.69267	0.02360		<.0001

א. כתבו את המשוואה בצורה ליניארית בעזרת הטרנספורמציה המתאימה.

ב. האומדן למשוואה (1) הינו: _____.

ג. המשמעות הכלכלית של β היא: _____.

ד. גבולות רווח-סמך ברמת סמך של 95% עבור β הינם:

גבול תחתון: _____.

גבול עליון: _____.

ה. ערך t לחישוב מובהקות ה- β הינו:

i. לא ניתן לחשב ערך זה בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: _____.

ו. אם נגדיל את מדד המחירים לצרכן ביחידה אחת, כמות הכסף במשק תגדל ב:

i. 71.7233

ii. 1.69267

iii. 169.267

iv. 1.69267%

v. אף תשובה איננה נכונה.

הועלתה הטענה שתוספת של אחוז אחד במדד המחירים לצרכן תגדיל את כמות הכסף במשק ביותר מאחוז אחד.

ז. ההשערות לבדיקת הטענה: _____.

ח. סטטיסטי t לבדיקת הטענה הינו:

i. לא ניתן לחשבו באמצעות הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: _____.

ט. על פי התשובות לסעיפים הקודמים ניתן להסיק כי ערכו של סטטיסטי F לבדיקת מובהקות המודל הינו:

i. לא ניתן לחשב את ערכו של סטטיסטי F על סמך סטטיסטי t .

ii. 861.4225

iii. 5144.23

iv. 71.7233 4

י. אם נוציא שורש ריבועי למדד המחירים לצרכן במשק:

i. האומד של α ישתנה: נכון/לא נכון/ אי אפשר לדעת

ii. האומד של β יעלה: נכון/לא נכון/ אי אפשר לדעת

iii. סטטיסטי F לבדיקת מובהקות המודל לא ישתנה: נכון/לא נכון/ אי אפשר לדעת

הועלתה הטענה כי יש צורך להוסיף למשוואה גם את הפעילות הכלכלית במשק (Y) כמשתנה מסביר, ולכן יש לאמוד את המשוואה הבאה:

$$LN(M)_t = \alpha + \beta_1 \cdot LN(P)_t + \beta_2 \cdot LN(Y)_t + U_t \quad 2.$$

משוואה (2) נתונה בפלט מס' 2.

Dependent Variable: lnm

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	2	44.05069	22.02535	2585.05	<0.0001
Error	102	0.86907	0.00852		
C Total	104	44.91976			
Root MSE	0.09231			R-square	0.9807
Dep Mean	8.53854			Adj R-sq	0.9803
C. V.	1.08104				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	0.78242	0.59739	1.31	0.1932
lnp	1	1.63491	0.05332	30.66	<.0001
lny	1	0.20001	0.16568	-----	0.2302

Covariance of Estimates

COVB	INTERCEP	lnp	lny
INTERCEP	0.35687	0.025884	-0.09762
lnp	0.02588	0.002843	-0.00792
lny	-0.09762	-0.00792	0.02745

יא. סטטיסטי t לבדיקת הטענה הינו :

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא : _____.

יב. על פי התשובה לסעיף הקודם, ניתן להסיק

את ערכו של סטטיסטי F למובהקות המודל. נכון/ לא נכון/ אי אפשר לדעת

יג. על פי התשובה לסעיף יא' ניתן להסיק את

ערכו של סטטיסטי WALD לבדיקת הטענה. נכון/ לא נכון/ אי אפשר לדעת

הועלתה הטענה כי הגמישות ביחס למחיר גבוהה פי 10 מהגמישות ביחס לפעילות הכלכלית במשק.

יד. סטטיסטי WALD לבדיקת הטענה הינו :

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא : _____.

טו. הרגרסיה המוגבלת כאשר H_0 נכונה למבחן WALD הינה : _____

כאשר : D_0 : _____
 D_1 : _____

טז.

i. איזה מבין המודלים המוצעים
במשוואות 1 ו-2 עדיף?

משוואה 1/משוואה 2/אין הבדל בין המודלים

ii. אם משתנה רמת המחירים במשק היה
מובהק במשוואה מס' 1, הוא יהיה מובהק
בהכרח גם במשוואה מס' 2 :
נכון/לא נכון/אי אפשר לדעת

(2) ענו על השאלות הבאות (כל שאלה בפני עצמה, בכל שאלה מונח
המודל: $Y = \alpha + \beta \cdot X + U$ ומתקיימות כל ההנחות הקלאסיות).

- א. $\bar{R}^2 < R^2$ מתקיים תמיד :
נכון/לא נכון/אי אפשר לדעת.
- ב. אם דוחים H_0 במבחן חד צדדי ברמת
מובהקות α , אזי בהכרח גם נדחה H_0
במבחן הדו צדדי באותה רמת מובהקות :
נכון/לא נכון/אי אפשר לדעת.
- ג. אם ערך האומד ל- β גבוה, השערת האפס
למובהקות השיפוע תידחה בוודאות :
נכון/לא נכון/אי אפשר לדעת.
- ד. הוספת משתנה מסביר למשוואת הרגרסיה
עשויה להקטין את R^2 :
נכון/לא נכון/אי אפשר לדעת.
- ה. אם דוחים H_0 במבחן דו צדדי ברמת
מובהקות α , אזי בהכרח גם נדחה H_0
במבחן החד צדדי באותה רמת מובהקות :
נכון/לא נכון/אי אפשר לדעת.
- ו. אם רווח בר סמך לשיפוע כולל את הערך
אפס, ניתן לומר כי השערת האפס למובהקות
השיפוע מתקבלת בהכרח :
נכון/לא נכון/אי אפשר לדעת.
- ז. האומדים היעילים ביותר לפרמטרים באוכלוסייה
יהיו בהכרח אומדי הריבועים הפחותים :
נכון/לא נכון/אי אפשר לדעת.
- ח. בהוספת משתנה מסביר מובהק למודל,
ערך $\bar{R}^2$ יעלה בהכרח.
נכון/לא נכון/אי אפשר לדעת.
- ט. מבחן WALD הוא מקרה פרטי של מבחן F
למובהקות המודל :
נכון/לא נכון/אי אפשר לדעת.
- י. שיטת הריבועים הפחותים מביאה
למקסימום את $\bar{R}^2$:
נכון/לא נכון/אי אפשר לדעת.

(3) נתון המודל: $Y_t = \alpha + \beta X_t + U_t$.

נתון כי אר"פ למודל זה הינו: $\hat{\beta} = \frac{S_{XY}}{S_{XX}}$.

א. הוכיחו כי $\hat{\beta}$ אומד ליניארי וחסר הטיה של β .

ב. חשבו את $VAR(\hat{\beta})$.

ג. נתון האומד: $\tilde{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2}$.

הוכיחו כי $\tilde{\beta}$ אומד ליניארי אך איננו חסר הטיה ל- β .

ד. מהם התנאים בהם מתקיים: $E(\tilde{\beta}) = \beta$?

תשובות סופיות:

- (1) א. $LN(M)_t = \alpha + \beta \cdot LN(P)_t + U_t$. ב. $LN(M)_t = 1.49372 + 1.69267 \cdot LN(P)_t$. ג. גמישות.
ד. גבול תחתון: 1.64527 , גבול עליון: 1.73987 .

ה. ii, $t_{\beta=0} = 71.7233$. ו. v. ז. $H_0: \beta = 1$, $H_1: \beta > 1$.

- ח. ii, $t = 29.35$. ט. i. י. לא ניתן לדעת.
ii. אי אפשר לדעת. iii. אי אפשר לדעת. יא. ii, $t = 1.2$. יב. לא נכון. יג. נכון. יד. ii, $WALD = 0.048$.

טו. $D_0 = LN(M)_t$, $D_1 = 10 \cdot LN(P)_t + LN(Y)_t$. טז. i. משוואה 1 .

- ii. לא נכון. (2) א. נכון. ב. לא נכון. ג. לא נכון. ד. לא נכון.
ה. לא נכון. ו. לא נכון. ז. לא נכון. ח. נכון.
ט. לא נכון. י. נכון.

(3) א. הוכחה. ב. $V(\hat{\beta}) = \frac{\sigma^2}{S_{xx}}$. ג. הוכחה.

ד. ראו סרטון.

אקונומטריקה א

פרק 17 - שאלות חזרה למבחן מבוססות על תוכנת STATA

תוכן העניינים

1. כללי 101

שאלות חזרה למבחן מבוססות על תוכנת STATA:

שאלות:

- 1) כדי לבדוק האם יש קשר בין שכר המורים וההוצאה לתלמיד בבי"ס ציבוריים נאמד המודל הבא: $Pay_i = \beta_1 + \beta_2 Spend + u_i$. כאשר pay הוא שכר המורים ו-Spend מייצג את ההוצאה לתלמיד שנמדדו בדולרים לשנה:

```
. reg pay spending
```

Source	SS	df	MS	Number of obs = 51		
Model	608555015	1	608555015	F(1, 49) = 112.60		
Residual	264825250	49	5404596.94	Prob > F = 0.0000		
				R-squared = 0.6968		
				Adj R-squared = 0.6906		
Total	873380265	50	17467605.3	Root MSE = 2324.8		

	pay	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
spending		3.307585	.3117043	10.61	0.000	2.681192	3.933978
_cons		12129.37	1197.351	10.13	0.000	9723.204	14535.54

- א. פרשו את הרגרסיה (רמז: משמעות האומדים לפרמטרים, האם התוצאות מובהקות?). האם יש הגיון כלכלי לתוצאות שהתקבלו?
- ב. מצאו תחזית נקודתית לשכר במוצע אם ההוצאה לתלמיד היא \$5000.
- ג. נבדקו מדדים תיאוריים לגבי המשתנה הב"ת:

```
. su spending
```

Variable	Obs	Mean	Std. Dev.	Min	Max
spending	51	3696.608	1054.761	2297	8349

- ד. הסבירו את משמעות ההנחות הקלאסיות במונחי המשתנים.
- ה. האם ההוצאה הממוצעת לתלמיד מסבירה הרבה משונות שכר המורים?
- ו. מהו הפער החזוי בשכר המורים בין שני בתי ספר שהפער בהוצאה לתלמיד ביניהם הוא \$20?
- ז. האם ניתן לומר על סמך התוצאות כי המודל מובהק? אם כן, באיזה רמת מובהקות?
- ח. חוקר ביצע מבחן שמטרתו לבדוק האם המודל יוצא מראשית הצירים:
- השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן.
 - ערך t סטטיסטי לבדיקת המבחן.
 - מסקנת המבחן: דוחים/לא דוחים את H_0 ברמת סמך של 95%.

- ט. חוקר ביצע מבחן שמטרתו לבדוק כי עבור עליה בדולר אחד בהוצאה השנתית הממוצעת לתלמיד שכר המורים הממוצע עולה בלפחות 3 דולר:
- השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן.
 - ערך t סטטיסטי לבדיקת המבחן.
 - ערך t קריטי לבדיקת המבחן.
 - מסקנת המבחן: דוחים/לא דוחים את H_0 ברמת ביטחון של 95%.
- י. חשבו רווח בר סמך לאמידת השיפוע ברמת סמך של 90%.
- יא. חוקר רצה לאמוד את המשוואה בשקלים במקום בדולרים ולכן אמד את המשוואה הבאה: $\hat{pay}_1 = \alpha_0 + \alpha_1 spend_1$.
- הניחו כי שער הדולר עומד על 3.5 ש.
- מהי הפקודה ב-stata ליצירת המשתנים החדשים?
 - מה יהיה האומדן ל- α_1 ? מה יהיה האומדן לסטיית התקן של α_1 ?
- על אותם הנתונים נאמדה גם המשוואה הבאה:
- $$Pay_i = \beta_0 + \beta_1 Spend + \beta_2 Spend^2 + u_i$$
- כאשר: $spend_2 = spend^2$.
- יב. כתוב את הפקודה ב-STATA ליצירת המשתנה החדש.
- יג. מתוצאות האמידה התקבל ש:
- $$\hat{pay}_i = 23128.44 + 1.223spend_i - 0.0000635spend_2$$
- (9.55) (10.55) (8.78)
- הערכים בסוגריים הם ערכי t סטטיסטי.
- השערת האפס לבדיקה האם יש השפעה ליניארית של ההוצאה על התשלום למורים? T סטטיסטי?
- מסקנת הבדיקה: דוחים/לא דוחים את H_0 .
- יד. מהי הפקודה ב-STATA לבדיקת המתאם בין שני המשתנים הב"ת?
- טו. בהינתן תוצאות המודל כיצד נראה פרופיל השכר כפונקציה של ההוצאה על התלמיד (אין צורך בציור מדויק רק במגמה).
- טז. מהי ההוצאה לתלמיד שאחריה שכר המורים מתחיל לרדת?
- יז. התקבל מתאם של 0.97 בין שני המשתנים הב"ת. לנוכח המתאם הגבוה החוקר טען כי האומדים במשוואה (3) הם מוטים.
- הטענה נכונה/הטענה אינה נכונה.
- יח. לאור תוצאות האמידה של משוואה (3) ניתן להסיק כי האומדים של משוואה (1) הינם מוטים.
- הטענה נכונה/הטענה אינה נכונה.
- יט. הניחו כי יש הטרוסקדסטיות במשוואה (3).
- מה יהיו ההשלכות לגבי האומדים של המשוואה:
- האומדים יהיו חסרי הטיה ויעילים.
 - האומדים יהיו חסרי הטיה אך לא יעילים.
 - האומדים יהיו מוטים אך יעילים.
 - האומדים יהיו מוטים ולא יעילים.

(2) נתונים 2 המודלים הבאים :

$$1. Y_t = \alpha + \beta X_t + u_t$$

$$2. Y_t = \beta X_t + u_t$$

- א. בהנחה כי R^2 של המודל הראשון גבוה מזה של המודל השני, האם ניתן לומר שהמודל הראשון טוב יותר?
הניחו כי מודל (1) הוא המודל האמיתי.
- ב. מהן תכונות האומדן לשיפוע של מודל (2)? מוטת/חסר הטיות?
- ג. האם ניתן לומר כי האומדן הינו עקיב?
- ד. חוו דעתכם על הטענה: "על סמך משפט גאוס מרקוב, ניתן להסיק כי אומדן הריבועים הפחותים של משוואה (1) הינו יעיל יותר מאומדן הריבועים הפחותים של משוואה (2)". הסבירו את תשובתכם.
- ה. חשבו את שונות האומדן.
- ו. לאיזה משני המודלים יהיה אומדן לשיפוע גבוה יותר? לראשון/לשני/לא ניתן לדעת.

(3) על מנת לאמוד את הקשר שבין השכלה (בשנים) להכנסה (באלפי שקלים) נאמדו שני המודלים הבאים :

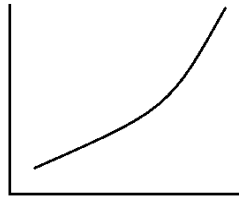
. reg wage educ

Source	SS	df	MS	Number of obs = 1000		
Model	7888.51144	1	7888.51144	F(1, 998) = 253.20		
Residual	31092.9858	998	31.1552964	Prob > F = 0.0000		
				R-squared = 0.2024		
				Adj R-squared = 0.2016		
Total	38981.4972	999	39.0205177	Root MSE = 5.5817		
wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.138517	.0715497	15.91	0.000	.998112	1.278922
_cons	-4.912181	.9667875	-5.08	0.000	-6.80935	-3.015011

. reg lwage educ

Source	SS	df	MS	Number of obs = 1000		
Model	65.5213155	1	65.5213155	F(1, 998) = 272.72		
Residual	239.767622	998	.240248118	Prob > F = 0.0000		
				R-squared = 0.2146		
				Adj R-squared = 0.2138		
Total	305.288937	999	.305594532	Root MSE = .49015		
lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.1037608	.0062831	16.51	0.000	.0914313	.1160904
_cons	.7883743	.0848975	9.29	0.000	.6217761	.9549724

א. ידוע כי ניתן לתאר את הקשר בין שכר להשכלה על ידי הגרף הבא :



- באיזה מודל כדאי לבחור? תנו 3 נימוקים לבחירתכם.
 ב. מהי התשואה להשכלה על סמך המודל הנבחר?
 ג. על פי מודל (1), מהי גמישות השכר ביחס להשכלה בנקודת הממוצעים : (7,12)?
 ד. על בסיס נתוני המדגם חושב השכר הממוצע עבור התצפיות שלהן 16 שנות השכלה.

```
. su wage if educ==16
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	186	13.30328	7.575015	2.54	60.19

- הממוצע הוא 13.303.
 שימו לב : סימן ה- "=" משמש להשוואה בין משתנים
 (למשל : $price_new = price * 1.2$), לעומת זאת, סימן ה- "==" משמש להגדרת ערך משתנה ($educ == 8$ if $educ$)).
 i. מהי התחזית לשכר לעובד עם 16 שנות השכלה עבור כל אחד מהמודלים?
 ii. איזה מהמודלים נותן תחזית נקודתית מדויקת יותר : המודל הראשון/המודל השני/ לא ניתן להשוות.
 ה. איזה מודל מסביר חלק גדול יותר של השונות של המשתנה התלוי : המודל הראשון/המודל השני/ לא ניתן להשוות.
 ו. לפי האומדים בפלט (לא לפי מבחן סטטיסטי) התשואה להשכלה חיובית בהכרח : נכון/לא נכון.
 ז. רצו לבחון כיצד השכר מושפע גם מוותק העובד ולכן הכניסו את המשתנה : $\log(vetek)$. מה משמעות מקדם השיפוע של המשתנה החדש במודל הראשון ובמודל השני.
 ח. בהתייחס למודל השני תוצאות האמידה מראות כי הן התשואה להשכלה והן גמישות הוותק אינם מובהקים. יחד עם זאת הרגרסיה עם שני המשתנים הב"ת יצאה מובהקת. כיצד ניתן להסביר זאת? האם מומלץ להשמיט את שני המשתנים ביחד מהמודל (לאמוד את השכר באמצעות משתנים אחרים)? האם מומלץ לאמוד את המודל ללא משתנה הוותק?

- (4) על מנת לבחון את פונקציית הייצור של אורז נאמד המודל הבא :
- $$\ln(PROD) = \beta_1 + \beta_2 \ln(AREA) + \beta_3 \ln(LABOR) + \beta_4 \ln(FERT) + \varepsilon_i$$
- כאשר :
 $PROD$ - כמויות אורז מדושן (נמדד בטונות).
 $AREA$ - גודל החלקות בהם האורז נשתל (נמדד בעשרות דונמים).
 $LABOR$ - סה"כ ימי עבודה של עובדים ובני משפחה (של החקלאי).
 $FERT$ - כמויות דשן בשימוש (נמדד בק"ג).
 להלן תוצאות האמידה :

```
. reg lprod larea llabor lfert
```

Source	SS	df	MS	Number of obs = 352		
Model	226.084875	3	75.361625	F(3, 348) = 646.51		
Residual	40.5653554	348	.116567113	Prob > F = 0.0000		
				R-squared = 0.8479		
				Adj R-squared = 0.8466		
Total	266.65023	351	.759687266	Root MSE = .34142		

	lprod	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
larea		.3617359	.0639678	5.65	0.000	.2359237	.4875481
llabor		.4328479	.0668825	6.47	0.000	.301303	.5643928
lfert		.2095023	.0382654	5.47	0.000	.1342417	.2847628
_cons		-1.546786	.2556536	-6.05	0.000	-2.049607	-1.043966

- א. בחנו את ההשערה כי גמישות הייצור ביחס לגודל החלקות (AREA) שווה לגמישות ביחס לימי העבודה (LABOR). השתמשו ברמת מובהקות של 5%, נסחו את ההשערה בצורה פורמאלית ודווחו את התוצאות תוך שימוש בפלט הבא :

```
. test larea= llabor
```

```
( 1) larea - llabor = 0
```

```
F( 1, 348) = 0.34
Prob > F = 0.5592
```

- ב. כתבו את הפקודות ב-STATA לאמידת הרגרסיה המוגבלת מהסעיף הקודם.
 ג. כעת, בחנו את ההשערה המורכבת מההשערה של סעיף א' + ההשערה כי פונקציית הייצור מקיימת תק"ל ($\beta_2 + \beta_3 + \beta_4 = 1$), תוך שימוש בפלט הבא :

```
. gen l_pr_fe=log( prod/ fert)
. gen l_ar_la_fe=log( area* labor/fert^2)
. reg l_pr_fe l_ar_la_fe
```

Source	SS	df	MS	Number of obs = 352		
Model	51.0075377	1	51.0075377	F(1, 350) = 439.63		
Residual	40.6079092	350	.116022598	Prob > F = 0.0000		
				R-squared = 0.5568		
				Adj R-squared = 0.5555		
Total	91.6154469	351	.261012669	Root MSE = .34062		

	l_pr_fe	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
l_ar_la_fe		.3940824	.018795	20.97	0.000	.3571172	.4310477
_cons		-1.402958	.0913195	-15.36	0.000	-1.582562	-1.223354

באיזה מבחן אתם משתמשים?
נסחו את ההשערה, את הסטטיסטי והקריטי.
מהי מסקנתכם?

- 5) חוקרים ביקשו לאמוד את פונקציית החיסכון המצרפי במשק הישראלי.
מפתח שמות משתנים:
GDS87 - חסכון מקומי גולמי.
GDP87 - תוצר מקומי גולמי.
GC87 - הוצאות הממשלה.
כל הנתונים הינם במיליוני ₪ במחירים קבועים של שנת 1987.
נאמדו 2 המודלים הבאים:

. reg gds87 gdp87

Source	SS	df	MS	Number of obs =
Model	268713647	1	268713647	26
Residual	60820138.1	24	2534172.42	F(1, 24) = 106.04
Total	329533785	25	13181351.4	Prob > F = 0.0000
				R-squared = 0.8154
				Adj R-squared = 0.8077
				Root MSE = 1591.9

gds87	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
gdp87	.1852456	.0179896	10.30	0.000	.1481169 .2223742
_cons	-2624.027	1018.32	-2.58	0.017	-4725.737 -522.3182

. reg gds87 gdp87 gc87

Source	SS	df	MS	Number of obs =
Model	307085221	2	153542611	26
Residual	22448563.7	23	976024.509	F(2, 23) = 157.31
Total	329533785	25	13181351.4	Prob > F = 0.0000
				R-squared = 0.9319
				Adj R-squared = 0.9260
				Root MSE = 987.94

gds87	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
gdp87	.2955591	.0208369	14.18	0.000	.2524547 .3386635
gc87	-.7784411	.1241513	-6.27	0.000	-1.035268 -.5216146
_cons	4217.674	1260.961	3.34	0.003	1609.177 6826.171

- א. כתבו את המודל האקונומטרי שנאמד בכל אחת מהאמידות.
ב. איזה מבין שני המודלים הקודמים הייתם מעדיפים? למה?
ג. מהו ההבדל בין המשמעות של האומדן למקדם המשתנה GDP87 בשני המודלים?
ד. ביחס לאומד במשוואה 2 האומד $\hat{\alpha}_2$ במשוואה 1 יהיה:
מוטה כלפי מעלה/מוטה כלפי מטה/ חסר הטיה/לא ניתן לדעת
ה. ביחס לאומד במשוואה 3 האומד $\hat{\alpha}_2$ במשוואה 1 יהיה:
יעיל/לא יעיל/ לא ניתן לדעת
ו. בהינתן תוצאות משוואה 1 הקורלציה בין שני המשתנים ה"ת" היא:
חיובית/שלילית/אפס/לא ניתן לדעת
ז. החוקר החליט להוסיף משתנה המודד את הצריכה הממשלתית במיליוני דולרים
במקום בשקלים: GC87\$ ואמד את המשוואה:
 $GDS87_t = \beta_1 + \beta_2 GDP87_t + \beta_3 GC87_t + \beta_4 GC87 + u_t$ 3.

- האומדים של משוואה 3 יהיו :
- חסרי הטיה ויעילים/מוטים ולא יעילים/חסרי הטיה אך לא יעילים/לא מוגדרים.
- ח. החוקר החליט להוסיף מדד נוסף לתוצר המקומי. כתוצאה מהוספת המדד הנוסף משתנה התוצר המקומי הפך להיות לא מובהק. כיצד ניתן להסביר זאת :
- מולטיקוליניאריות מלאה.
 - מולטיקוליניאריות חלקית.
 - הוספת משתנה לא רלוונטי.
 - השמטת משתנה רלוונטי.
- ט. החוקר טען כי : "אם נוסיף משתנה נוסף לרגרסיה כלשהי אז האומדן ל- σ^2 לעולם לא יעלה". נכון/לא נכון?
- י. החוקר טוען כי "אם נוסיף משתנה נוסף לרגרסיה, אז האומדן ל- $\bar{R}^2$ יעלה בהכרח". נכון/לא נכון.

6) מידע שנאסף לאחרונה על מכירות של 880 בתים בסטוקטון, קליפורניה, נמצא

בקובץ Stockton2.

המשתנים הם :

PRICE - מחיר בית בדולרים.

SQFT - גודל הבית (square feet).

BEDS - מספר חדרי שינה.

BATHS - מספר חדרי שירותים.

AGE - גיל הבית.

STORIES - הקומה של הבית.

VACANT - משתנה דמי המקבל 1 אם הבית היה פנוי בזמן מכירתו ו-0 אחרת.

להלן תוצאות אמידה המתבססת על המשתנים הנ"ל :

```
. gen price1000= price/1000
. gen lprice1000=ln( price1000)
. gen sqft100= sqft/100
. reg lprice1000 sqft100 age beds baths stories vacant
```

Source	SS	df	MS	Number of obs =	880
Model	92.5168833	6	15.4194805	F(6, 873) =	420.24
Residual	32.0322735	873	.03669218	Prob > F =	0.0000
Total	124.549157	879	.141694149	R-squared =	0.7428
				Adj R-squared =	0.7410
				Root MSE =	.19155

lprice1000	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sqft100	.0637441	.0020353	31.32	0.000	.0597495 .0677387
age	-.0024514	.000366	-6.70	0.000	-.0031698 -.0017331
beds	-.0848595	.0133702	-6.35	0.000	-.1111011 -.058618
baths	.0089069	.0181165	0.49	0.623	-.02665 .0444638
stories	-.0182728	.0219074	-0.83	0.404	-.06127 .0247245
vacant	-.0803092	.0132448	-6.06	0.000	-.1063045 -.0543138
_cons	3.994605	.037782	105.73	0.000	3.920451 4.068759

א. מהי משוואת הרגרסיה שנאמדה?

ב. פרשו את תוצאות האמידה. דונו בסימנים ובמשמעויות של כל אחד מהמשתנים.

ג. מה ההבדל במחיר הממוצע בין בית פנוי בזמן מכירתו לבין בית המאוכלס בזמן מכירתו?

ד. חוקר אחר הגדיר את משתנה VACANT = משתנה דמי המקבל 1 אם הבית היה מאוכלס בזמן מכירתו ו-0 אחרת.

האם יש צורך לחשב מחדש את משוואת הרגרסיה? כן/לא.

ה. נניח כי רצו לתרגם את תוצאות המודל מ-sqrft למטרים מרובעים.

יחידה אחת של sqrft שווה ל-0.093 מטר רבוע. מה יהיה הערך של β_1

במודל החדש? מה יהיה ערך t סטטיסטי לדחיית H_0 של המשתנה החדש?

ו. מה הפער החזוי בין מחיר של בגודל 130 מטר בת 15 שנה לבין דירה

בגודל 115 מטר בת 20 שנה, בהינתן שכל שאר המשתנים נותרים קבועים?

ז. מהי רמת הסמך הנמוכה ביותר עבורה ניתן לדחות את הטענה כי $\beta_5 = 0$?

ח. החוקר החליט לאמוד את המשוואה ללא משתנה VACANT כמשתנה מסביר. להלן תוצאות האמידה:

```
. reg lprice1000 sqft100 age beds baths stories
```

Source	SS	df	MS	Number of obs =	880
Model	91.1678763	5	18.2335753	F(5, 874) =	477.40
Residual	33.3812805	874	.038193685	Prob > F =	0.0000
				R-squared =	0.7320
				Adj R-squared =	0.7304
Total	124.549157	879	.141694149	Root MSE =	.19543

lprice1000	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sqft100	.0654365	.0020569	31.81	0.000	.0613995 .0694734
age	-.0021886	.0003708	-5.90	0.000	-.0029163 -.0014608
beds	-.0823237	.0136344	-6.04	0.000	-.1090837 -.0555638
baths	.0062791	.0184781	0.34	0.734	-.0299876 .0425459
stories	-.0260982	.0223123	-1.17	0.242	-.0698901 .0176938
_cons	3.925237	.0367376	106.85	0.000	3.853132 3.997341

כיצד השמטת המשתנה השפיעה על משוואת הרגרסיה ועל מקדמיה?

ט. בנוסף אמדו על בסיס המדגם הנוכחי את שתי הרגרסיות הבאות:

```
. reg lprice1000 sqft100 age beds baths stories if vacant==0
```

Source	SS	df	MS	Number of obs =	415
Model	50.040644	5	10.0081288	F(5, 409) =	233.63
Residual	17.5202976	409	.042836913	Prob > F =	0.0000
				R-squared =	0.7407
Total	67.5609416	414	.16319068	Adj R-squared =	0.7375
				Root MSE =	.20697

lprice1000	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sqft100	.0684748	.0030057	22.78	0.000	.0625662 .0743833
age	-.001996	.0005387	-3.70	0.000	-.003055 -.000937
beds	-.0977677	.019958	-4.90	0.000	-.1370009 -.0585346
baths	.0193179	.0251737	0.77	0.443	-.030168 .0688038
stories	-.0654738	.033737	-1.94	0.053	-.1317933 .0008457
_cons	3.979725	.0554426	71.78	0.000	3.870737 4.088713


```
. reg lprice1000 sqft100 age beds baths stories if vacant==1
```

Source	SS	df	MS	Number of obs =	465
Model	37.2518359	5	7.45036717	F(5, 459) =	242.82
Residual	14.0830841	459	.030682101	Prob > F =	0.0000
				R-squared =	0.7257
Total	51.33492	464	.110635603	Adj R-squared =	0.7227
				Root MSE =	.17516

lprice1000	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sqft100	.0593134	.0027696	21.42	0.000	.0538708 .0647561
age	-.0028851	.0004996	-5.78	0.000	-.0038668 -.0019034
beds	-.0678451	.0178018	-3.81	0.000	-.1028283 -.0328619
baths	-.0103401	.026461	-0.39	0.696	-.0623399 .0416596
stories	.0265265	.0284661	0.93	0.352	-.0294135 .0824665
_cons	3.924588	.0472086	83.13	0.000	3.831817 4.01736

השוו את תוצאות האמידה של שתי הרגרסיות הנ"ל.

י. ערכו מבחן Chow לבדיקת שקילות (יציבות) המקדמים בשתי הרגרסיות מהסעיף הקודם.

(7) חוקר מעוניין ללמוד על הקשר בין הכנסה של משפחה לבין מספר שנות הלימוד של הבעל, מספר שנות הלימוד של האישה והימצאות ילדים קטנים בבית. להלן מפתח שמות המשתנים:

FAMINC - הכנסת המשפחה (דולרים בשנה).

HEDU - מספר שנות הלימוד של הבעל.

WEDU - מספר שנות הלימוד של האישה

$$KID6 = \begin{cases} 1 & \text{אם יש ילדים מתחת לגיל 6} \\ 0 & \text{אחרת} \end{cases}$$

הוא אמד שלוש רגרסיות וקיבל את התוצאות הבאות:

./ Table 1

. reg faminc Heduc Weduc k16

Source	SS	df	MS	Number of obs =	428
Model	1.4725e+11	3	4.9082e+10	F(3, 424) =	30.43
Residual	6.8384e+11	424	1.6128e+09	Prob > F =	0.0000
Total	8.3109e+11	427	1.9463e+09	R-squared =	0.1772
				Adj R-squared =	0.1714
				Root MSE =	40160

faminc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Heduc	3211.526	796.7026	4.03	0.000	1645.547 4777.504
Weduc	4776.907	1061.164	4.50	0.000	2691.111 6862.704
k16	-14310.92	xxxxxxx	xxxxx	xxxx	xxxxxxxxxxx xxxxxxxxxxx
_cons	-7755.331	11162.93	-0.69	0.488	-29696.91 14186.25

./ Table 2

. reg faminc Heduc Weduc

Source	SS	df	MS	Number of obs =	428
Model	1.3405e+11	2	6.7027e+10	F(2, 425) =	40.87
Residual	6.9703e+11	425	1.6401e+09	Prob > F =	0.0000
Total	8.3109e+11	427	1.9463e+09	R-squared =	0.1613
				Adj R-squared =	0.1574
				Root MSE =	40498

faminc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Heduc	3131.509	802.908	3.90	0.000	1553.344 4709.674
Weduc	4522.641	1066.327	4.24	0.000	2426.711 6618.572
_cons	-5533.631	11229.53	-0.49	0.622	-27605.97 16538.71

./ Table 3

. reg faminc Heduc

Source	SS	df	MS	Number of obs =	428
Model	1.0455e+11	1	1.0455e+11	F(1, 426) =	61.30
Residual	7.2654e+11	426	1.7055e+09	Prob > F =	0.0000
Total	8.3109e+11	427	1.9463e+09	R-squared =	0.1258
				Adj R-squared =	0.1237
				Root MSE =	41297

faminc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
Heduc	5155.484	658.4573	7.83	0.000	3861.254 6449.713
_cons	26191.27	8541.108	3.07	0.002	9403.308 42979.23

- א. פרשו את משמעותות המקדמים ברגרסיה בטבלה הראשונה, התייחסו למובהקות המקדמים (אין צורך בבדיקה פורמאלית של מובהקות).
- ב. התקבל כי ממוצע משתנה $KL6$ במדגם הוא 0.432. מה משמעות נתון זה?
- ג. חשבו את סטית התקן הנאמדת של אומד הריבועים הפחותים ל- β_{KL6} מהמודל הנאמד בטבלה הראשונה. הראו את החישובים שלכם ואת הנוסחאות עליהן אתם מתבססים.
- רמז: השתמשו באינפורמציה שניתן לבדוק מובהקות המקדם הנ"ל ביותר מדרך אחת.
- ד. הסבירו את הסיבה להבדל המשמעותי בין אומדני β_{HEDU} בשלוש הטבלאות? איך אפשר להסביר את העדר השוני (כמעט העדר שוני) באומדני β_{HEDU} ו- β_{WEDU} בטבלה הראשונה והשנייה?
- (8) הורצו 2 רגרסיות על מדגם בן 400 תצפיות והתקבלו התוצאות הבאות:

$$1. \hat{CM}_i = 263.6416 - 0.0056 \cdot PGNP_i - 2.2316 \cdot FLR_i$$

$$(s.e) \quad (11.5932) \quad (0.0019) \quad (0.2099) \quad R^2 = 0.7077$$

$$2. \hat{CM}_i = 168.3067 - 0.0055 \cdot PGNP_i - 1.768 \cdot FLR_i + 12.8686 \cdot TFR_i$$

$$(s.e) \quad (11.5932) \quad (0.0019) \quad (0.2099) \quad (?) \quad R^2 = 0.7474$$

כאשר:

$CM = Child Mortality$ - מס' מקרי המוות של ילדים מתחת לגיל 5 לכל 1000 לידות חיים.

$PGNP = Per Capita GNP$ - תוצר לנפש במחירים קבועים בדולרים.

$FLR = Female Literacy Rate$ - אחוז נשים שיודעות לקרוא ולכתוב.

$TFR = Total Fertility Rate$ - מס' הלידות הממוצע לאישה במדינה. הנתונים הם עבור 64 מדינות.

- א. כיצד ניתן לפרש את המקדם למשתנה TFR ? אפרוורית, האם תצפו לקשר חיובי/שלילי בין CM ל- TFR ? הסבירו.
- ב. האם ערכי המקדמים של המשתנים $PGNP, FLR$ מרגרסיה 1 שונים מאלו ברגרסיה 2? אם כן, מה יכולה להיות הסיבה/ות לשינוי זה?
- ג. חוו דעתכם על הטענה הבאה: "כיוון ש FLR ו- TFR כל כך מתואמים אין לשים אותם באותה הרגרסיה".
- ד. באיזה מודל תבחרו מבין השניים? באיזה מבחן סטטיסטי יש להשתמש כדי לענות על שאלה זו? הראו חישובכם. (רמז: הביעו את הסטטיסטי של מבחן F במונחי R^2).
- ה. האם תוכלו לחשב את סטיית התקן הנאמדת של המקדם למשתנה TFR ? (רמז: היזכרו בקשר בין התפלגות T להתפלגות F).

- ו. האם ניתן להשוות את מקדם ההסבר של שתי הרגרסיות? האם ניתן להשוות את מקדם ההסבר המתוקן? אם כן, השוו ודווחו על התוצאות.
- ז. ענו על השאלות התיאורטיות הבאות:
 - i. איזה מן הגורמים הבאים יכול לגרום לכך שאומדי OLS יהיו מוטים:
 1. הטרוסקדסטיות.
 2. השמטת משתנה מסביר רלוונטי.
 3. מקדם מתאם גבוה מאוד בין שני משתנים מסבירים במודל.
 - ii. איזה מהגורמים הנ"ל יכול לגרום לכך שסטטיסטי t של OLS לא יהיה תקף?
 - iii. התייחסו לטענה הבאה: "אם האומדים הינם עקיבים הם יהיו בהכרח גם חסרי הטיה". נכון/לא נכון.
 - iv. אם נתון ש-u לא מתפלג נורמאלית אז אמידת המשוואה בשיטת OLS תניב אומדים שאינם עקיבים. נכון/לא נכון.

9) חוקרת רצתה לבדוק עונתיות במחירי הירקות. לשם כך הגדירה את משתני הדמי הבאים:

- D_1 יקבל את הערך 1 אם מדובר באביב ו-0 אחרת.
 - D_2 יקבל את הערך 1 אם מדובר בקיץ ו-0 אחרת.
 - D_3 יקבל את הערך 1 אם מדובר בסתיו ו-0 אחרת.
 - D_4 יקבל את הערך 1 אם מדובר בחורף ו-0 אחרת.
- כאשר:

V_t - מדד מחירי הירקות.

p_t - מדד המחירים לצרכן.

לשם כך אמדה את הרגרסיה הבאה על פני 30 שנה:

$$V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 \cdot P_t + u_t$$

א. מדוע לא הכניסה החוקרת למשוואת הרגרסיה את משתנה D_4 ?

תוצאות האמידה שהתקבלו הן:

$$V_t = 1379.11 + 99.18 D_{1t} + 2209.47 D_{2t} - 476.56 D_{3t} + 489.92 \cdot P_t, \quad R^2 = 0.0844$$

נאמדה בנוסף גם המשוואה הבאה: $V_t = \beta_0 + \beta_1 \cdot P_t + \varepsilon_t$.

$$V_t = 11114.14 + 536.36 \cdot P_t, \quad R^2 = 0.0670$$

תוצאות האמידה שהתקבלו:

ב. על סמך התוצאות שהתקבלו, דרגו את העונות לפי רמת המחיר הבסיסית שלהן. הציגו לכל עונה את המיקום שלה ואת רמת המחיר הבסיסית כפי שבא לידי ביטוי במודל.

ג. בדקו את ההשערה כי עונתיות לא משפיע על מחיר הירקות.

ד. כתבו את ההשערות הבאות:

- i. מדד מחירי הירקות זהה בחורף ובאביב.
- ii. מדד מחירי הירקות גבוה בקיץ מאשר בחורף ביותר מ-600.

ה. האם יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן (בהנחה שהמחיר ההתחלתי של הירקות זהה בין עונות השנה)?

10 חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה. החוקר בדק האם הפונקציה של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את המשוואות הבאות:

$$1. \quad NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t}$$

$$2. \quad NUM_t = \alpha + \beta_1 \times TYPE_t + \beta_2 \times AVGD_t + \beta_3 \times (AVGD \times TYPE)_t + U_t$$

כאשר:

NUM_t - מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה.

$AVGD_t$ - נפח התנועה בקטע כביש t ליום באלפים.

$TYPE_t$ - משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר, ו-0 כאשר הכביש לא מהיר.

תוצאות אמידת המשוואות מוצגות להלן:

$$1. \quad NUM_t = 0.739 + 0.0233 \cdot AVGD_t$$

$$2. \quad NUM_t = 0.14978 + 1.40311 \cdot TYPE_t + 0.002877 \cdot AVGD_t - 0.008 \cdot (AVGD \cdot TYPE)_t$$

$$1. \quad ESS = 20963, \quad Pt_{\hat{\alpha}} = 0.0019; \quad Pt_{\hat{\beta}} = 0.0001$$

$$2. \quad ESS = 20759, \quad Pt_{\hat{\alpha}} = 0.6534; \quad Pt_{\hat{\beta}_1} = 0.0067; \quad Pt_{\hat{\beta}_2} = 0.0001; \quad Pt_{\hat{\beta}_3} = 0.1283$$

א. בדקו את טענת החוקר.

ב. מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על 4 אלפי מכוניות ליום בקטע הכביש האמור?

ג. מהי השערת האפס לבדיקת הטענה?

ד. הרגרסיה המוגבלת "תחת" H_0 למבחן F (WALD) הינה:

$$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + v$$

כאשר:

$$Z_0 = \underline{\hspace{2cm}}$$

$$Z_1 = \underline{\hspace{2cm}}$$

$$Z_2 = \underline{\hspace{2cm}}$$

$$Z_4 = \underline{\hspace{2cm}}$$

11 ברשותכם נתונים על מחירי ארוחת ביג מק במסעדות מקדונלדס ברחבי הארץ ב-1/1/2008 וב-1/1/2009. חלק מהחנויות מנוהלות ע"י הרשת והשאר מנוהלות ע"י זכיינים. אתם מעוניינים לבדוק את השפעתה של פסיקת בית משפט מ-7/2008 אשר מאפשרת לרשת לקבוע מחיר מקסימום עבור מוצרים הנמכרים בחנויות המנוהלות ע"י זכיינים.

לפניך הנתונים הבאים:

Price - מחיר ארוחת ביג-מק.

$$D_{2009} = \begin{cases} 0 & \text{year} = 2008 \\ 1 & \text{year} = 2009 \end{cases}$$

$$D_{Franchise} = \begin{cases} 0 & \text{חנות רשת} \\ 1 & \text{חנות זכיון} \end{cases}$$

נתון המודל הבא המתאר את המחיר:

$$Price_i = \beta_1 + \beta_2 D_{2009} + \beta_3 D_{Franchise} + \beta_4 D_{2009} * D_{Franchise} + \varepsilon_i$$

א. מה המקדם שמגלם את הפרש השכר הצפוי בין חנות רשת לחנות זכיון כשכל יתר המשתנים מוחזקים קבוע?

ב. נסחו את H_0 ו- H_1 לבדיקת ההשערות הבאות:

i. מחיר ארוחת ביג מק ב-2008 היה זהה בחנויות המנוהלות ע"י הרשת ובחנויות המנוהלות ע"י זכיינים.

ii. פסיקת בית המשפט שינתה את פער המחירים בין חנויות המנוהלות ע"י הרשת לבין חנויות המנוהלות ע"י זכיינים בין השנים 2008 ו-2009.

ג. הגדירו את משתנה האינטראקציה על ידי הגדרת 4 הקבוצות הבאות:

D_1 - רשת לפני 2009.

D_2 - רשת אחרי 2009.

D_3 - זכיון לפני 2009.

D_4 - זכיון אחרי 2009.

i. מהי הפקודה הרלוונטיות ב-STATA ליצירת D_1 ?

ii. כמה ממשתני הדמי יש להכניס לתוך הרגרסיה?

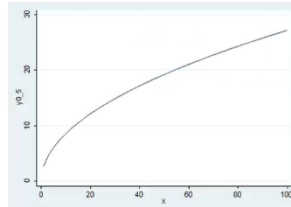
iii. כתבו את משוואת הרגרסיה.

iv. נסחו שוב את ההשערות של סעיף א'.

v. חוקר טוען כי התוצאות שהתקבלו בבדיקת ההשערות על ידי הגדרת משתני הדמי בשני האופנים יהיו זהות. נכון/לא נכון.

תשובות סופיות:

- (1) א. משמעות: ראו סרטון, תוצאות: מובהקות, היגיון כלכלי: יש.
 ב. $p\hat{y} = 28,666.87$. ג. $p\hat{y} = 24,355.9$. ד. ראו סרטון.
 ה. כן. ו. 66.15 . ז. כן, $p_v = 0.00$.
 ח. i. $H_0: \beta_1 = 0$. ii. $t = 10.13$. iii. דוחים.
 ט. i. $H_0: \beta_2 = 3$. ii. $t = 0.98$. iii. $t_{(0.05,48)} = 2$. iv. לא דוחים.
 י. $(2.90 < \beta_2 < 3.71)$. יא. i. $gen\ pay_1 = pay * 3.5$. ii. אין שינוי.
 יב. $gen\ spend_2 = spend * spend$. יג. $H_0: \beta_2 = 0$, $t = -8.78$, דוחים.
 יד. $Corr\ spend\ spend_2$. טו. $X^* = 9,629.92$. טז.



- (2) יז. אינה נכונה. יח. נכונה. יט. ii. ד. ראו סרטון.
 א. לא ניתן להשוות. ב. מוטה. ג. לא. ד. ראו סרטון.
 ה. $V(\hat{\beta}) = \frac{\sigma_u^2}{\sum X_t^2}$. ו. לשני.
 (3) א. במודל השני. ב. ראו סרטון. ג. $\eta_{y,x} = 1.95$.
 ד. i. $WAGE = 13.304$, $WAGE_F = 11.571$. ii. המודל הראשון.
 ה. לא ניתן להשוות. ו. לא נכון. ז. ראו סרטון.
 ח. לא מומלץ.
 (4) א. מבחן WALD: $prob(F - ST) = 0.5592 > 0.05$, לא דוחים.
 ב. $Gen\ Z_1 = \ln(AREA) + \ln(LABOR)$.
 ג. $Reg\ \ln(PROD) Z_1 \ln(FERT)$.
 ג. מבחן F.
 השערה: $H_0: \beta_2 + \beta_3 + \beta_4 = 1, \beta_2 = \beta_3 \Leftrightarrow \beta_4 = 1 - 2\beta_2, \beta_2 = \beta_3$.
 $H_1: else$.
 סטטיסטי: $F - st = 0.1825$.
 (5) א. המודל הראשון: $GDS87_t = \alpha_1 + \alpha_2 GDP87_t + \varepsilon_t$.
 המודל השני: $GDS87_t = \beta_1 + \beta_2 GDP87_t + \beta_3 GC87_t + u_t$.
 ב. המודל השני. ג. ראו סרטון. ד. מוטה כלפי מטה.

- ה. לא יעיל. ו. שלילית. ז. לא מוגדרים.
ח. ii. ט. לא נכון. י. לא ניתן לדעת.
א. (6)

$$\ln(PRICE1000) = \beta_0 + \beta_1 sqft100 + \beta_2 age + \beta_3 beds + \beta_4 baths + \beta_5 stories + \beta_6 vacant + \varepsilon_i$$

- ב. ראו סרטון. ג. $e^{-0.080309} = 0.9228$. ד. לא.
- ה. $sqrft1 = \frac{sqrft}{0.093}$, t לא ישתנה. ו. 0.957. ז. 0.404.
- ח. ראו סרטון. ט. ראו סרטון. י. ראו סרטון.
- (7) א. ראו סרטון. ב. פרופורציה. ג. $se(\hat{\beta}_{KL6}) = 5003.818$.
- ד. ראו סרטון.

- 8) א. ראו סרטון, חיובי. ב. שונים, מולטיקולנאריות. ג. ראו סרטון.
ד. השני, מבחן F. ה. כן. ו. מקדם ההסבר: אין טעם להשוות, מקדם ההסבר המתוקן: כן.
i. 1 ii. 2

9. א. הייתה נוצרת מולטיקולינאריות מלאה.
 ב. סתיו (902.58), חורף (1379.11), אביב (1478.29) וקיץ (3588.58).
 ג. אין עדות.
 ד. i. $H_0: \alpha_1 = 0$
 ii. $H_0: \alpha_2 = 600$
 iii. לא נכון.
 iv. ראו סרטון.
 ה. לא ניתן לבדוק.

- 10) א. ראו סרטון. ב. $NUM_t = 1.532398$ ג. $H_0: \beta_2 + \beta_3 = 2 \cdot \beta_2$ ד. $H_0: \beta_3 = \beta_2$

$$Z_0 = NUM_t$$

$$Z_1 = TYPE_t$$

$$Z_2 = (AVGD_t + AVGD \cdot TYPE)_t$$

$$Z_4 = 0$$

- $$\begin{array}{ll} H_0: \beta_4 = 0 & \text{.ii} \\ H_1: \beta_4 \neq 0 & \end{array} \quad \begin{array}{ll} H_0: \beta_3 = 0 & \text{.i.}\square \\ H_1: \text{else} & \end{array} \quad \begin{array}{ll} \beta_3: D_{franchise} & \text{.N (11)} \end{array}$$

$$genD_1 = 1 \text{ if } D_{franchise} = 0 \text{ and } D_{2009} = 0 \quad \text{i.i.d.}$$

replace $D_1 = 0$ if $D_1 =$

$$\text{Price} = \alpha_1 + \alpha_2 D_1 + \alpha_3 D_2 + \alpha_4 D_3 + u_i \quad .\text{iii} \quad .3.\text{ii}$$

- $$\text{v. נכון.} \quad \begin{array}{l} H_0: \alpha_3 - \alpha_2 - \alpha_4 = 0 \\ H_1: \alpha_3 - \alpha_2 - \alpha_4 \neq 0 \end{array} \quad \text{.ii} \quad \begin{array}{l} H_0: \alpha_2 - \alpha_4 = 0 \\ H_1: \alpha_2 - \alpha_4 \neq 0 \end{array} \quad \text{.i.iv}$$