

הנחיות כלליות

1. זמן הבחינה: שעתיים אקדמיות
2. חומר פתוח
3. מותר לצטט כל תוצאה שראינו בכיתה, אלא אם השאלה מבקשת במפורש לפתח או להוכיח את התוצאה

הסעיפים א'-ד' עוסקים ברגרסיה ליניארית פשוטה של משתנה תוצאה Y_i על משתנה מסביר בודד X_i .

א. **(10 נק')** כתבו את המודל הליניארי (שעבדנו איתו בכיתה) עבור הקשר בין Y_i לבין x_i . הקפידו לציין את הנחות המודל, ולציין מהם הפרמטרים הלא ידועים.

ב. **(15 נק')** תחת המודל הליניארי, מהי הפרשנות של השיפוע β_1 ?

ג. **(15 נק')** תחת המודל הליניארי, חשבו את $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)$ כאשר $\hat{\beta}_0, \hat{\beta}_1$ הם אומדי הריבועים הפחותים של השיפוע והחותך. פשטו את התוצאה ככל הניתן.

הערה: אתם יכולים לחשב את זה בכל דרך שתמצאו. אחת הדרכים היא לכתוב את המודל בהצגה

וקטורית, ולהשתמש בתוצאות מוכרות לגבי מטריצת השונות של $\hat{\beta}$. אם אתם בוחרים בדרך הזו,

תוכלו להשתמש בנוסחה הכללית להופכי של מטריצה 2×2 :

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \quad \text{אז} \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{אם}$$

ד. **(15 נק')** תחת המודל הליניארי הנוורמלי, בנו רווח-סמך ברמת בטחון 95% עבור β_1 (השיפוע במודל).

הסעיפים ה'-ז' מתייחסים לקובץ נתונים ספציפי.

קובץ נתונים כולל $n = 533$ מדידות על שכרם של עובדים (Y) והמשתנים המסבירים:

X_1 = שנות ותק

X_2 = תחום עיסוק $\in \{\text{Clerical, Management, Professional, Sales, Service, Other}\}$

מצורפים פלט ניתוח רגרסיה של Y על X_1, X_2 , ותרשים פיזור של Y כנגד X_1 כשהנקודות מקודדות לפי

תחום העיסוק (X_2).

```
> fm <- lm(wage ~ experience + occupation)
> summary(fm)
```

Call:

```
lm(formula = wage ~ experience + occupation)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.4428	-2.9976	-0.8973	1.9859	19.4079

Coefficients:

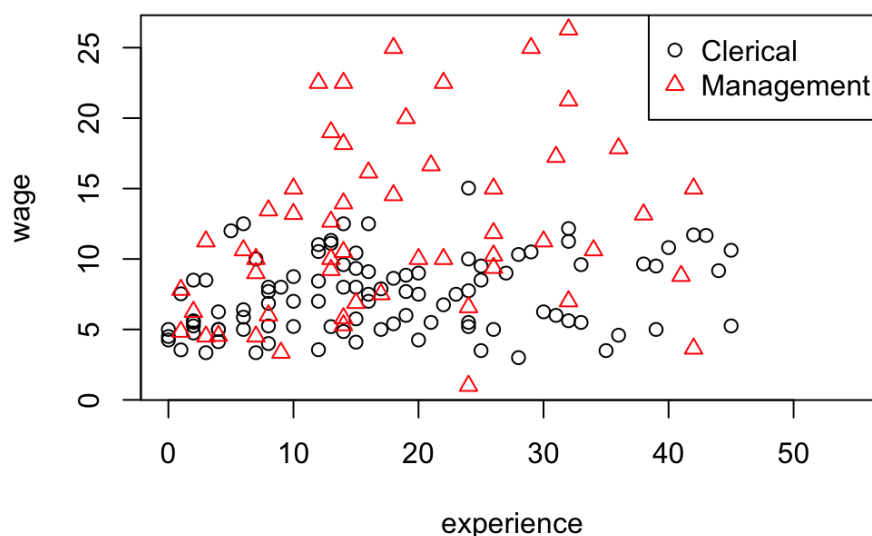
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.39640	0.52486	12.187	< 2e-16 ***
experience	0.05838	0.01557	3.750	0.000196 ***
occupationManagement	4.64523	0.74909	6.201	1.13e-09 ***
occupationOther	0.97548	0.57045	1.710	0.087850 .
occupationProfessional	4.64863	0.62212	7.472	3.31e-13 ***
occupationSales	0.11926	0.84432	0.141	0.887727
occupationService	-1.03780	0.66085	-1.570	0.116922

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.411 on 526 degrees of freedom

Multiple R-squared: 0.2013, Adjusted R-squared: 0.1922

F-statistic: 22.1 on 6 and 526 DF, p-value: < 2.2e-16



- ה. (15 נק') חשבו אומדן (ריבועים פחותים) לתוחלת השכר של עובד בתחום המכירות (sales) בעל 16 שנות ותק. באופן כללי, מה התכונות של האומד הזה מבחינת תוחלת השגיאה הריבועית (MSE)?
- ו. (15 נק') מהסתכלות בלתי-אמצעית בתרשים, האם נראה באופן ברור שיש אינטראקציה בין תחום העיסוק לבין שנות ותק? הסבירו. מה הייתם מציעים לשנות במודל כדי לבדוק פורמלית את ההשערה הזאת?
- ז. (15 נק') עכשיו נשווה את המודלים עם ובלי אינטראקציה בין X_1 ו- X_2 :

```
> fm <- lm(wage ~ experience + occupation)
> fm.int <- lm(wage ~ experience * occupation)
```

נסמן את האומד למקדם של X_1 תחת המודל ללא אינטראקציה, וב- $\hat{\gamma}_1$ את האומד למקדם של X_1 תחת המודל עם אינטראקציה. באופן כללי (בלי להתחשב בפלט הספציפי הנתון), האם זה נכון תמיד שמתקיים

$$\hat{\beta}_1 = \hat{\gamma}_1$$

בהצלחה!