به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

هوش مصنوعی قابل اعتماد

تمرین سوم

نام و نام خانوادگی	زکیه امین
شماره دانشجویی	810102398

سؤال اول: شبکه‌های بیزی

بخش اول

1. **زیربخش اول:**

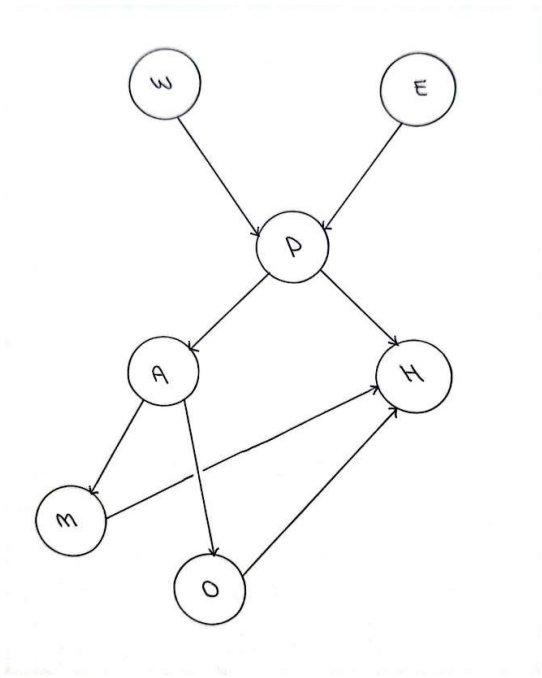

Figure 1_Bayesian Network

2. **زیربخش دوم:**

توزیع احتمال توام شبکه بیزی به صورت فاکتورشده:

$$P(X_1, \dots X_n) = \prod_{i=1}^{n} P(X_i \mid Parent(X_i))$$

$$P(W, E, P, A, M, O, H) =$$
$$P(W) \times P(E) \times P(P|E, W) \times P(A|P) \times P(M|A) \times P(O|A) \times P(H|O, M, P)$$

3. **زیربخش سوم:**

Markov Blanket مربوط به یه گره درواقع کوچک‌ترین مجموعه‌ای از گره‌ها که اگه اونا رو بدونیم آن گره از تمام گره‌های دیگه شبکه مستقل میشه.

$$Markov\ Blanket(M) = \{Parent(M), Coparent(M), Children(M)\}$$

مطابق شبکه بیزی خواهیم داشت:

$$Parent(M) = \{A\} \quad Coparent(M) = \{O, P\} \quad Children(M) = \{H\}$$
$$Markov\ Blanket(M) = \{A, O, P, H\}$$

4. زیربخش چهارم:

الف) A⊥W | P

مسیر بین A و W به صورت $W \rightarrow P \rightarrow A$ میباشد. با دادن P مسیر های بین A و W بلاک میشود. پس این دو گره از هم مستقل میشوند. (درست)

ب) M⊥O | A

مسیر بین M و O به صورت $M \leftarrow A \rightarrow O$ میباشد. با دادن A مسیر های بین O و M بلاک میشود. پس این دو گره از هم مستقل میشوند. (درست)

پ) A⊥H |{M,O}

مسیر بین A و H به صورت $A \leftarrow P \rightarrow H$ و $A \rightarrow O \rightarrow H$ و $A \rightarrow M \rightarrow H$ میباشد. چون P داده نشده مسیر $A \leftarrow P \rightarrow H$ بلاک نمیشه، پس تحت این شرایط هم این دو گره از هم مستقل نمیشوند. (نادرست)

ت) M⊥E | A

مسیر بین M و E به صورت $E \rightarrow P \rightarrow A \rightarrow M$ میباشد. با دادن A مسیر های بین E و M بلاک میشود. پس این دو گره از هم مستقل میشوند. (درست)

ث) O⊥W | {A,H}

مسیر بین O و W به صورت $W \rightarrow P \rightarrow A \rightarrow O$ و $W \rightarrow P \rightarrow H \leftarrow O$میباشد. بخاطر V-Structure بودن با دادن H استقلال این دو گره از بین میره. (نادرست)

بخش دوم

	مسیر های فعال	توضیح
الف) $P(w^1) > P(w^1\|f^1)$	$F \overset{-}{\rightarrow} E \overset{+}{\rightarrow} W$	f^1 باعث میشه احتمال ورزش کردن(e^1) کم بشه که این در ادامه منجر به کاهشw^1 میشه.
ب) $P(f^1\|e^1) < P(f^1\|e^1, h^1)$	$H \overset{+}{\rightarrow} E \overset{-}{\leftarrow} F$	e^1 به ما داده شده و مسیر ذکر شده فعاله. e^1 دو تا علت میتونه داشته باشه:(f^0 و h^1). h^1 ورزش کردن رو توضیح میده و دیگه نیاز به f^0 نداریم و احتمال f^1 افزایش پیدا میکنه.
پ) $P(h^1\|w^1) > P(h^1\|f^0, w^1)$	$H \overset{+}{\rightarrow} E \overset{-}{\leftarrow} F$ $F \overset{-}{\rightarrow} E \overset{+}{\rightarrow} W \overset{+}{\leftarrow} D \overset{+}{\leftarrow} H$	w^1 به ما داده شده، مسیر دوم فعاله. دو علت برای w^1 داریم $(d^1$ و$e^1)$. با f^0 احتمال e^1 زیاد میشه و w^1 توجیه میشه و نیاز به d^1 و والدش (H) نداریم و احتمال h^1 کم میشه.
ت) $P(e^1\|w^1),\ P(e^1\|c^1, w^1)$ غیرقابل مقایسه	$C \overset{-}{\leftarrow} D \overset{+}{\rightarrow} W \overset{+}{\leftarrow} E$ $C \overset{-}{\leftarrow} D \overset{+}{\leftarrow} H \overset{+}{\rightarrow} E$	w^1 به ما داده شده، مسیر v-structure فعاله. دو علت برای w^1 داریم: (d^1, e^1). c^1 افزایش پیدا کرده که یعنی d^1 کم شده، پس احتمال e^1 زیاد میشه. از طرفی تو مسیر دیگه کاهش d^1 باعث کاهش h^1 میشه و این هم رو e^1 رو کم میکنه. پس دو مسیر آثار متضادی دارن و نمیشه گفت کدوم اثر قوی تری داره، پس مقایسه ممکن نیست.
ث) $P(w^1\|d^1, e^1) = P(w^1\|d^1, e^1, h^1)$	$D \overset{+}{\leftarrow} H \overset{+}{\rightarrow} E \overset{+}{\rightarrow} W$ $D \overset{+}{\rightarrow} W$	در هر دو عبارت d^1 و e^1 داده شده. پس با داشتن parent های W، این گره از باقی گره های غیر فرزندش مستقل میشه. به بیان دیگه مسیر های ذکر شده بلاک میشن. از اونجایی هم که H فرزند W نیست پس از هم مستقل میشن.
د) $P(t^1\|w^0, e^1) > P(t^1\|w^0, e^1, f^1)$	$F \overset{-}{\rightarrow} E \overset{+}{\leftarrow} H \overset{+}{\rightarrow} D \overset{-}{\rightarrow} C \overset{+}{\rightarrow} T$	e^1 به ما داده شده، مسیر بین H و F فعاله. f^1 احتمال h^1 رو بالا میبره و این احتمال d^1 رو زیاد و موجب کم شدن احتمال c^1 میشه که یعنی t^1 کم میشه. در نتیجه دونستن f^1 احتمال t^1 رو کم میکنه.

Table 1_Answer of part2

بخش سوم

1. **زیربخش اول:**

n متغیر داریم که هرکدام می‌تواند L مقدار متفاوت بگیرد، یعنی باید برای هر کدام L حالت را بررسی کرد. لذا داریم L^n. در واقع احتمال حالت آخر رو میتوان از روی حالات دیگر به دست آورد، پس فقط محاسبه $L^n - 1$ پارامتر مستقل لازم است.

2. **زیربخش دوم:**

هر گره دارای L مقدار متفاوت است و حداکثر k تا والد داره. برای توزیع احتمال توام داریم:

اگر k تا والد داشته باشیم یعنی داریم L^k تا حالت مختلف برای بخش Parent. خود X_i هم L حالت متفاوت داره که یعنی L-1 پارامتر مستقل نیاز داریم. پس برای هر گره

$$(L-1)L^k$$

پارامتر لازم است که کران بالاس آن برای کل شبکه میشود:

$$n(L-1)L^k$$

3. **زیربخش سوم:**

با توجه به فرمول توزیع احتمال توام خواهیم داشت:

$$P(C, X_1, X_2, \dots, X_n) = P(C) \times P(X_1|C) \times \dots \times P(X_{i+1}|X_i, C) \times \dots \times P(X_n|X_{n-1}, C)$$

برای $P(C)$: m-1 پارامتر نیاز است.

برای $P(X_1|C)$: پدرش m حالت متفاوت دارد، خودش هم L حالت داره. در نتیجه مطابق زیربخش دو به $(L-1)m$ پارامتر نیاز داریم.

برای $P(X_{i+1}|X_i, C)$: والدین mL حالت دارند، خودش هم L حالت داره. در نتیجه مطابق زیربخش دو به $mL(L-1)$ پارامتر نیاز است. $(n-1)$ تا از این گره ها داریم

در مجموع

$$(m-1) + m(L-1) + (n-1)mL(L-1)$$

تا پارامتر مستقل لازم است.

بخش چهارم

1. **زیربخش اول:**

در شکل سمت راست با جا به جا کردن جهت یال های $A \to B$ یا $E \to D$ v-structure جدیدی تولید نمیشه و قبلی عوض نمیشه، روابط دیگه بهم نمیخوره و شبکه بیزی جدید I-equivalent میشه.

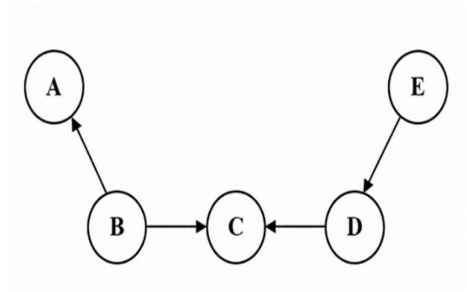

Figure 2_Example of I-equivalent Bayesian network

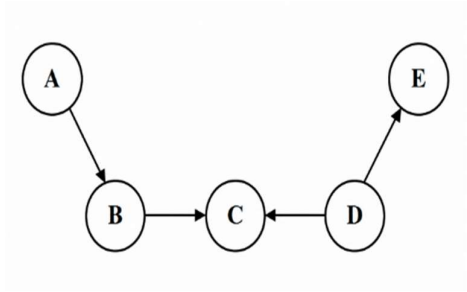

Figure 3_Example of I-equivalent Bayesian network

در شکل سمت چپ یه v-structure داریم به صورت $A \to B \leftarrow C$. با بررسی امکان تغییر جهت یال‌ها داریم:

نمیشه جهت یال های $A \to B$ و $C \to B$ رو عوض کرد چون v-structure اصلی‌مون از بین میره.

با عوض کردن جهت $B \to D$ گره D هم میشه والد B و یه v-structure ایجاد میشه و گره های A و B و D از هم مستقل میشن که چنین چیزی تو فرض سوال نیست. پس این تغییر شدنی نیست.

با عوض کردن جهت $D \to E$ هم یه v-structure جدید ساخته میشه، به طوری که $E \to D \leftarrow B$ و یعنی B و E از هم مستقل میشن. چنین چیزی تو شبکه بیزی صورت سوال نداریم، پس این تغییر هم شدنی نیست.

پس شبکه بیزی دیگری که با این شبکه I-equivalent باشه وجود ندارد.

2. **زیربخش دوم:**

طبق فرض سوال دو گره A و B از هم مستقل هستند، یعنی بین این دو در شبکه بیزی نباید مسیری باشد. در صورتی که در فرض بعدی داریم اگر C داده شود، A و B از هم مستقل میشوند

که یعنی مسیری بین این گره وجود دارد که با given شدن C بلاک میشود. این با فرض اول در تناقض است و نمیتوان برای این توزیع Perfect map رسم کرد.

سؤال دوم: علّیت

بخش اول: سوالات تئوری

1. زیربخش اول: (نامساوی)

مثال نقض:

از یک ساختار علّی با دو متغیر تصادفی باینری X و Y استفاده می‌کنیم که $Y \to X$ و توسط یه Cofounder هم به یکدیگر وابستگی دارند. به صورت $X \leftarrow Z \to Y$.

$$P(Z = 1) = P(Z = 0) = 0.5$$

روابط SCM به صورت:

$$X = Z$$

پس داریم:

$$P(X = 1) = P(X = 0) = 0.5$$

و برای متغیر Y:

$$Y = 2X + 2Z \xrightarrow{X=Z} Y = 4Z$$
$$P(Y = 0) = 0.5, \qquad P(Y = 4) = 0.5$$
$$P(Y = 1) = P(Y = 2) = P(Y = 3) = 0$$

با Force کردن $X = 0$ داریم:

$$Y = 2Z$$
$$P\big(Y = 2 \big| do(X = 0)\big) = 0.5$$

با Force کردن $X = 1$ داریم:

$$Y = 2 + 2Z$$
$$P(Y = 2 \,\big|\, do(X = 1)) = 0.5$$
$$\min_x P(Y = 2 \,\big|\, do\,(x)) = 0.5$$
$$\max_x P(Y = 2 \,\big|\, do\,(x)) = 0.5$$
$$\min_x P\big(Y = 2 \big| do\,(x)\big) \le P(Y = 2) \le \max_x P(Y = 2 \,\big|\, do\,(x))$$

در این حالت به نامساوی غیرمنطقی زیر میرسیم:

$$0.5 \leq 0 \leq 0.5$$

2. زیربخش دوم: (اثر علّی)

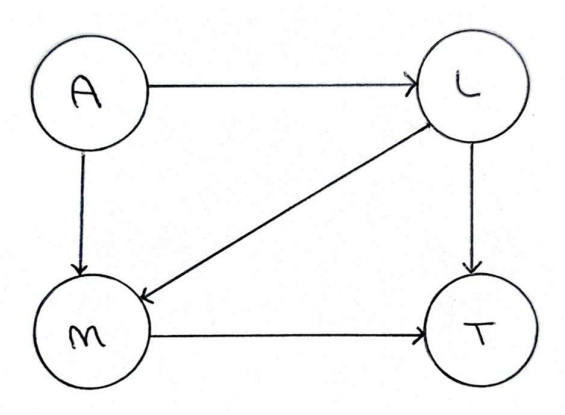

Figure 4_Causality graph

$E[M|L,A]$ اثر علّی بار CPU روی مصرف حافظه را نشان می‌دهد. درواقع A والد مشترک بین L و M میباشد که یعنی Confounder است. اگر این گره given شود، درواقع مسیر $M \leftarrow A \rightarrow L$ بلاک میشود و فقط مسیری باقی میماند که اثر خالص L و M را نشان میدهد.

بخش دوم: سوالات پیاده‌سازی

1. زیربخش اول: (پیاده‌سازی SCM)

add_variable: یه متغیر جدید رو به همراه Parents هاش و Equation مربوط بهش رو ذخیره میکنه.

Get_topological_order: در SCM متغیرها باید بر اساس ساختار والد-فرزندی مقداردهی شوند و هیچ متغیری پیش از مشخص شدن مقدار والدینش ارزیابی نشه. یک گراف جهت‌دار بدون دور از مدل میسازع و با الگوریتم مرتب‌سازی توپولوژیک ترتیب دقیق اجرا رو پیدا میکنیم.

sample: با حرکت بر اساس ترتیب توپولوژیک به دست آمده، نمونه‌ها را تولید می‌کند. داده‌ها کاملا منطبق بر توزیع توام شبکه بدست میان.

do_operator: در عملگر $do(X = x)$ ما تمام یال‌های ورودی به متغیر X را قطع می‌کنیم و مقدار x رو بهش force میکنیم.

plot_graph: گراف علّیت را رسم می‌کند.

2. زیربخش دوم: (تاثیر دوره آموزشی)

الف)

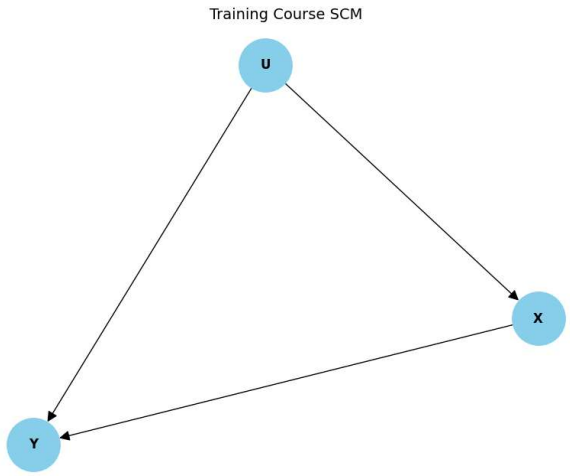

Figure 5_Training Course SCM

ب) تعداد ۱۰,۰۰۰ نمونه از جامعه آماری به صورت مشاهداتی(observational) استخراج شد.

	U	X	Y
0	-2.635612	False	-5.278651
1	0.160845	True	1.640465
2	1.927164	True	8.488907
3	0.188527	True	3.339134
4	-0.606901	False	-3.909879

Figure 6_Observational samples

$E[Y\|X=1]$	4.59112
$E[Y\|X=0]$	-1.60224
Observational difference	6.19336

Table 2_Observational samples' output

پ) برای محاسبه اثر خالص دوره آموزشی، از عملگر do استفاده شد. ابتدا تمام افراد به شرکت در دوره $do(X=1)$ و سپس به عدم شرکت $do(X=0)$ مجبور و نمونه‌برداری مجدد انجام شد.

$E[Y\|do(X=1)]$	2.99142
$E[Y\|do(X=0)]$	0.05267
ATE	2.93875

Table 3_Intervention samples' output

د) اختلاف میانگین مشاهده‌ای E[Y|X=1]−E[Y|X=0] بزرگ‌تر از اثر علّی واقعی است. علت این موضوع وجود متغیر U است که هم بر X و هم Y اثر دارد. در داده‌های مشاهده‌ای، اثر دوره و اثر انگیزه باهم ترکیب می‌شن و برآوردی بایاس‌دار ایجاد میکنند. با استفاده از عملگر do، وابستگی $U \to X$ حذف شده و اثر علّی واقعی دوره اندازه‌گیری می‌شود که مطابق معادله ساختاری برابر با ضریب X در معادله Y، یعنی تقریبا 3 است.

3. زیربخش سوم: (سیگار نکشیم!)

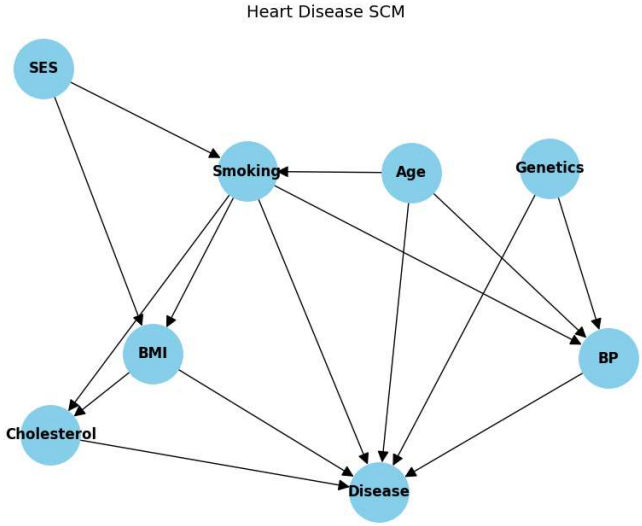

Figure 6_Heart Disease SCM

اثر علّی کل جامعه (ATE):

$E[Disease\|do\ Smoking=1]$	0.15634
$E[Disease\|do\ Smoking=0]$	0.06010
ATE	0.09624

Table 4_ATE for Smoking

مصرف سیگار به طور میانگین و خالص، ابتلا به بیماری قلبی را در کل جامعه 9.6 درصد افزایش می‌دهد.

اثر علّی روی افراد تحت آزمایش(ATT):

$E[Disease \| do\ Smoking = 1, Smoking = 1]$	0.16275
$E[Disease \| do\ Smoking = 1, Smoking = 0]$	0.06211
ATT	0.10065

Table 5_ATT for Smoking

احتمال ابتلا به بیماری قلبی در میان افرادی که در دنیای واقعی سیگاری هستند، حدود 16.3٪ است. (چون این افراد واقعا سیگاری بودند، مداخله علّی تاثیری روی رفتارشون نداره). اگه همین افراد سیگاری واقعی رو مجبور کنیم که سیگار رو ترک کنند احتمال بیماری قلبی در آن‌ها به 6.2% کاهش پیدا میکند. ATT نشان میده مصرف سیگار ابتلا به بیماری قلبی را در میان جامعه سیگاری‌ها به اندازه 10٪ درصد افزایش داده.

اثر علّی مشروط بر سن(CATE):

برای افراد بالای 55 سال:

$E[Disease \| do(Smoking = 1), Age > 55]$	0.20185
$E[Disease \| do(Smoking = 0), Age > 55]$	0.08083
$CATE(AGE > 55)$	0.12102

Table 6_CATE(Age>55) for Smoking

اگر تمام افراد بالای ۵۵ سال را مجبور به مصرف سیگار کنیم، احتمال ابتلای آن‌ها به بیماری قلبی به رقم 20.2٪ می‌رسد. اگر همین گروه سنی را از مصرف سیگار منع کنیم، احتمال ابتلا به بیماری قلبی به 8.1٪ کاهش پیدا میکند. مصرف سیگار ابتلا به بیماری قلبی را در افراد بالای ۵۵ سال به طور خالص و به میزان 12.1٪ افزایش می‌دهد.

برای افراد پایین 55 سال:

$E[Disease \| do(Smoking = 1), Age \leq 55]$	0.1366
$E[Disease \| do(Smoking = 0), Age \leq 55]$	0.05133
$CATE(AGE \leq 55)$	0.08527

Table 7_CATE(Age<=55) for Smoking

اگر تمام افراد ۵۵ سال و کمتر را مجبور به مصرف سیگار کنیم، احتمال ابتلای آن‌ها به بیماری قلبی به حدود 13.6٪ می‌رسد. اگر همین گروه سنی را از مصرف سیگار منع کنیم، احتمال ابتلا

به بیماری قلبی به 5.1٪ کاهش می‌یابد. مصرف سیگار ابتلا به بیماری قلبی را در افراد ۵۵ سال و کمتر به طور خالص و به میزان 8.5٪ افزایش می‌دهد.

مقایسه تحلیلی ATE، ATT و CATE:

1- مداخله عمومی (بر اساس ATE): حذف سیگار در کل جامعه ریسک بیماری قلبی را به طور متوسط 9.6٪ کاهش می‌دهد.

2- مداخله هدفمند روی رفتار (بر اساس ATT): تمرکز روی ترک سیگار برای افراد سیگاری فعلی، کارایی بالاتری داره و نرخ بیماری را در این گروه 10٪ کاهش می‌ده.

3- مداخله هدفمند روی مشخصات فردی (بر اساس CATE): متوقف کردن سیگار در گروه سنی بالای ۵۵ سال، بالاترین کاهش ریسک (12.1٪) رو داره.

4. زیربخش چهارم: (قالب جدید وب‌سایت)

بخش یک: حالت ساده‌لوحانه

میانگین زمان سپری‌شده برای چیدمان قدیمی در مقابل چیدمان جدید:

Old Layout	12.5017
New Layout	9.7305
Difference	-2.7712

Table 8_Old and New layout time

New Layout باعث شده کاربران به طور متوسط 2.77 دقیقه کمتر در سایت وقت بگذرونند.

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     12.5017      0.207     60.422      0.000      12.096      12.908
layout_new    -2.7712      0.275    -10.085      0.000      -3.310      -2.232
==============================================================================
```

Figure 7_Naive coefficient

ضریب منفی متغیر New layout نشان میدهد که کاربران در چیدمان جدید به طور متوسط حدود 2.77 دقیقه کمتر از چیدمان قدیمی در سایت وقت می‌گذرانند. مقدار p-value بسیار کوچک است که یعنی این نتیجه کاملا از نظر آماری معناداره و شانس و تصادف نقشی نداشته.

یک تحلیل‌گر داده بدون دیدگاه علّی با مشاهده این نتیجه‌گیری، میگه چیدمان جدید وب‌سایت شکست خورده و پیشنهاد میکنه که طرح جدید متوقف بشه و سیستم به همان چیدمان قدیمی

بازگردد. با این حال، این تحلیل دچار سوگیری شدیده، چون رفتارهای پنهان کاربران در این مدل نادیده گرفته شده.

بخش دو: تشخیص بصری

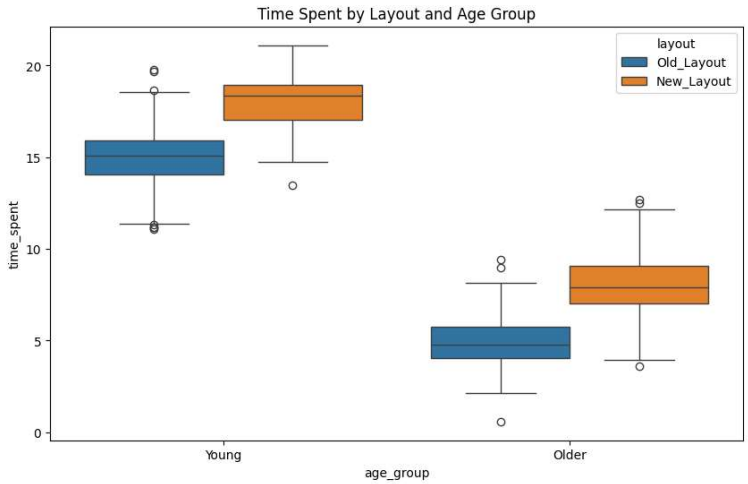

Figure 8_Box plot for time spent by layout and age group

این نمودار میانه و پراکندگی متغیر وابسته time spent را به تفکیک گروه سنی و نوع چیدمان نشون می‌ده. در گروه جوان، میانه زمان صرف‌شده در New Layout بالاتر از Old Layout است. در گروه بزرگسال میانه زمان صرف‌شده در چیدمان جدید بالاتر است. نمودار نشان می‌ده که چیدمان جدید یک اثر مثبت یکنواخت روی هر دو زیرگروه سنی داره. این نمودار مشخص می‌کنه که کاربران جوان به طور کلی و فارغ از نوع چیدمان، زمان بیشتری را نسبت به بزرگسالان در سایت میگذرونند.

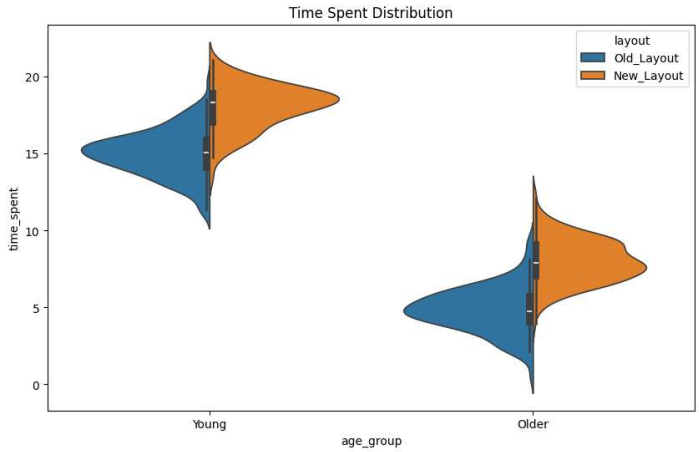

Figure 9_Violin plot for time spent distribution

این نمودار علاوه بر شاخص‌های مرکزی، چگالی احتمالی و شکل توزیع داده‌ها را نمایش می‌ده. شکل نمودار برای هر دو گروه سنی نشان می‌دهد که تمرکز و چگالی اصلی داده‌ها در حالت New Layout به سمت مقادیر بالاتر زمان متمایل شده است و پایداری اثر مثبت چیدمان جدید را در تمام سطوح توزیع جامعه تایید.

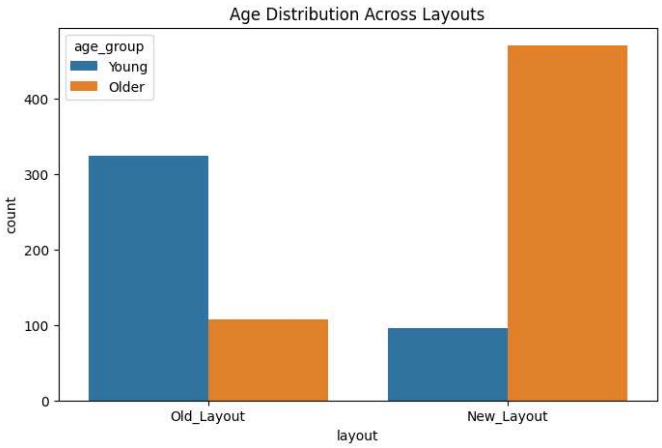

Figure 10_Count plot for age distribution across layouts

در چیدمان قدیمی بخش اعظم نمونه‌ها را کاربران جوان تشکیل می‌دهند که ذاتا زمان طولانی در سایت می‌گذرانند. در چیدمان جدید سهم عمده نمونه‌ها به کاربران بزرگسال اختصاص یافته که زمان بسیار کوتاهی در سایت سپری می‌کنند.

یک عدم تعادل شدید در تخصیص کاربران به دو چیدمان دیده می‌شه. سیستم به جای تخصیص تصادفی بخش عمده‌ای از کاربران جوان را به چیدمان قدیمی و بخش بزرگی از کاربران مسن را به چیدمان جدید مپ کرده. این عدم تعادل مهمه چون متغیر سن Confounder ایه که هم روی متغیر وابسته (زمان حضور در سایت) اثر می‌ذاره و هم روی متغیر مستقل(جوانان کلا زمان بیشتری در سایت می‌مانند). این عدم توازن باعث می‌شه اثر ذاتی سن با اثر واقعی چیدمان وب‌سایت ترکیب شده و نتایج آماری کلی منحرف بشه.

میانگین کل فریبنده است چون تفاوت‌های رفتاری درون گروه‌ها را نادیده می‌گیره. جوانان کلا زیاد در سایت می‌مانند (چه در طرح قدیم و چه جدید) و مسن‌ها کلا خیلی کم در سایت می‌مانند. از آنجا که سیستم به اشتباه تعداد خیلی زیادی از افراد مسن را در گروه چیدمان جدید قرار داده، زمان پایین این گروه میانگین کل چیدمان جدید را پایین کشید. وقتی در تحلیل ساده‌لوحانه داده‌ها را بدون در نظر گرفتن سن یک‌جا جمع می‌کنیم، به دلیل وزن بسیار بالای بزرگسالان در طرح جدید، میانگین کل طرح جدید به شدت سرکوب می‌شود. این ساختار علّی

معیوب، اثر واقعی و مثبت چیدمان جدید را پنهان کرده و به صورت یک اثر منفی کاذب نشون می‌ده.

در اینجا Simpson's Paradox رخ داده. یک پدیده آماری در تحلیل داده که در آن، یک رابطه یا جهت اثر که در گروه‌های مجزا و تفکیک‌شده از داده‌ها مشاهده می‌شود، پس از ترکیب کردن گروه‌ها ناپدید شده یا حتی معکوس می‌شود. این پارادوکس نتیجه مستقیم نادیده گرفتن ساختار علّی داده‌ها و وجود یک متغیر Confounder است. اینجا متغیر سن این نقش رو داره.

بخش سه: حل مساله علّی

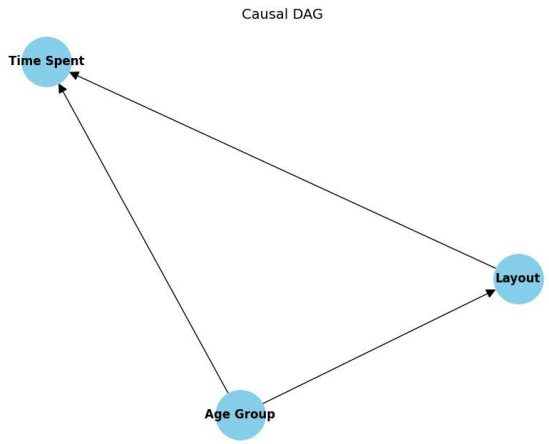

Figure 11_Causal DAG

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	15.0168	0.076	196.986	0.000	14.867	15.166
layout_new	3.0901	0.115	26.787	0.000	2.864	3.316
older	-10.0836	0.116	-87.099	0.000	-10.311	-9.856

Figure 12_Controlled coefficient

در مدل ساده‌لوحانه، ضریب منفی بود (-2.7712) که به اشتباه نشان میگفت طرح جدید استفاده کاربران را کاهش داده. اما در مدل جدید، ضریب به 3.0901 تغییر جهت داد. علت این تغییر علامت De-confounding است. مدل اثر ساختاری سن را تفکیک کرده و به ما می‌گه که تو یک گروه سنی تغییر Layout از قدیم به جدید چه اثری داره. این ضریب مثبت به طور کامل با میانگین‌های درون‌گروهی همخوانی دارد. ضریب منفی متغیر older (-10.08) تایید می‌کند که افراد مسن به طور ذاتی زمان بسیار کمتری در سایت می‌گذرانند و همین امر در مدل

ساده‌لوحانه باعث نتیجه‌گیری اشتباه راجب چیدمان جدید شده بود. مقدار p-value بسیار کوچک است که یعنی این نتیجه کاملا از نظر آماری معنادار است و شانس و تصادف نقشی نداشته.

توصیه نهایی این است که چیدمان جدید (New Layout) برای 100٪ کاربران وب‌سایت فعال شود. تحلیل علّی ثابت کرد که چیدمان جدید یک بهبود یکنواخت و سرتاسری ایجاد کرده و زمان تعامل هر دو گروه سنی را به شکل چشمگیری ارتقا داده. کاهش میانگین کل در دیتای اولیه، صرفا یک خطای آماری ناشی از تخصیص نامتوازن سیستم بود.

بخش چهار: اثبات پایداری و صحت مدل

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	15.6515	0.117	134.335	0.000	15.423	15.880
layout_new	0.1242	0.123	1.012	0.312	-0.117	0.365
older	-8.2754	0.123	-67.143	0.000	-8.517	-8.034

Figure 13_Placebo coefficient

ضریب layout new برابر با 0.1242 است و p-value برابر 0.312 به دست آمده است. $(pvalue > 0.05)$.

اگر مدل چندگانه و فرآیند کنترل در مرحله قبل معتبر باشد، انتظار داریم مقدار این ضریب در تست *Placebo* به صفر نزدیک شود و از نظر آماری غیرمعنادار گردد که خروجی ما این فرضیه را تایید کرد. ضریب به 0.1242 کاهش پیدا کرد. از طرفی $(pvalue > 0.05)$ که یعنی این ضریب کوچک از نظر آماری بی‌معنی است. پس اثر متغیر از نظر آماری غیرمعناداره و ضریب صرفا ناشی از نویز و شانس تصادفی در نمونه‌برداری است و هیچ رابطه واقعی رو نشان نمی‌ده.

این تست نشان داد اثری که ما در مرحله قبل پیدا کردیم ناشی از یک همبستگی کاذب یا شانس نبوده. وقتی ما ستون علت (*layout*) را به صورت تصادفی شافل میکنیم، مکانیزم و یال علّی بین علت و معلول قطع می‌شه. اگر مدل ما در مرحله قبل صرفا یک نویز یا رابطه ساختگی را به عنوان اثر نشان می‌داد، در این مرحله هم احتمالا عددی معنادار تولید می‌کرد. اما ناپدید شدن کامل اثر پس از تصادفی‌سازی (تبدیل شدن به اثری نزدیک به صفر و غیرمعنادار) ثابت می‌کنه که اثر مثبت دیده شده در مرحله قبلی*True Causal Effect* بوده است. این امر به

اطمینان صددرصدی می‌ده که بهبود حاصل شده ناشی از طراحی جدید است و نه تصادف آماری.