

به نام خدا



دانشگاه صنعتی امیرکبیر

گزارش پروژه‌ی نهایی درس تحلیل آماری داده‌های پزشکی

نام استاد: دکتر فرناز قاسمی

نام مقاله‌ی مورد بررسی:

Let's talk over coffee: Exploring the effect of coffee flavor descriptions on consumer imagery and behavior.

از سپهر کلانکی

با شماره دانشجویی ۹۸۳۳۰۵۸

زمستان ۱۴۰۲

۱	مقدمه
۱	معرفی پژوهش و دادگان:
۴	بررسی دادگان پرت
۵	بررسی نرمال بودن توزیع دادگان
۶	تحلیل‌های آماری
۹	مروری بر ساختار دادگان
۹	انتخاب آزمون
۱۰	نتایج آزمون‌های ANOVA/Friedman
۱۱	نتایج آزمون‌های post-hoc
۱۲	توضیح مدل رگرسیون خطی با اثرهای مختلط (LMER)
۱۳	نتایج آزمون LMER برای متغیرهای جانبی
۱۳	برای همه‌ی Trial‌ها
۱۵	برای حالت میانگین گرفته‌شده
۱۷	نتایج آزمون LMER برای پیش‌بینی WTP
۱۷	توضیح متغیر Covariate
۱۷	برای همه‌ی Trial‌ها
۱۹	نتیجه‌گیری
۲۰	تحلیل‌های فراتر؟

فهرست تصاویر

- شکل ۱ : تصویر پرسش‌نامه.....۲
- شکل ۲ : قسمتی از دادگان میانگین گرفته شده.....۴
- شکل ۳ : نمودار جعبه‌ای دادگان ۳ متغیر برای ۴ دسته‌ی توصیفات قهوه.....۴
- شکل ۴ : مقادیر متغیر WTP برای کاربر شماره‌ی ۳۹.....۵
- شکل ۵ : هیستوگرام و نمودار Q-Q برای متغیر WTP.....۶
- شکل ۶ : نتایج آزمون‌های تست نرمال بودن توزیع متغیر WTP.....۶
- شکل ۷ : هیستوگرام و نمودار Q-Q برای متغیر DTT.....۷
- شکل ۸ : نتایج آزمون‌های تست نرمال بودن توزیع متغیر DTT.....۷
- شکل ۹ : هیستوگرام و نمودار Q-Q برای متغیر VOI.....۸
- شکل ۱۰ : نتایج آزمون‌های تست نرمال بودن توزیع متغیر VOI.....۸
- شکل ۱۱ : راهنمای انتخاب آزمون جهت آنالیز اختلاف بین ۲ یا چند متغیر.....۹
- شکل ۱۲ : نتایج آزمون‌های ANOVA و Friedman برای هر سه متغیر WTP, DTT, VOI.....۱۰
- شکل ۱۳ : نتایج آزمون‌های post-hoc پارامتریک و غیر پارامتریک برای متغیر VOI.....۱۱
- شکل ۱۴ : نتایج آزمون‌های post-hoc پارامتریک و غیر پارامتریک برای متغیر DTT.....۱۱
- شکل ۱۵ : نتایج آزمون‌های post-hoc پارامتریک و غیر پارامتریک برای متغیر WTP.....۱۲
- شکل ۱۶ : نتایج مدل LMER برای پیش‌بینی متغیر VOI برای همه‌ی trial‌ها به کمک متغیرهای self-reported.....۱۴
- شکل ۱۷ : نتایج مدل LMER برای پیش‌بینی متغیر DTT برای همه‌ی trial‌ها به کمک متغیرهای self-reported.....۱۴
- شکل ۱۸ : نتایج مدل LMER برای پیش‌بینی متغیر WTP برای همه‌ی trial‌ها به کمک متغیرهای self-reported.....۱۵
- شکل ۱۹ : نتایج مدل LMER برای پیش‌بینی متغیر VOI برای میانگین trial‌ها به کمک متغیرهای self-reported.....۱۵
- شکل ۲۰ : نتایج مدل LMER برای پیش‌بینی متغیر DTT برای میانگین trial‌ها به کمک متغیرهای self-reported.....۱۶
- شکل ۲۱ : نتایج مدل LMER برای پیش‌بینی متغیر WTP برای میانگین trial‌ها به کمک متغیرهای self-reported.....۱۶
- شکل ۲۲ : نتایج مدل LMER برای پیش‌بینی متغیر WTP برای همه‌ی trial‌ها به کمک متغیرهای Age, VOI, DTT.....۱۸
- شکل ۲۳ : نتایج مدل LMER برای پیش‌بینی متغیر WTP برای میانگین trial‌ها به کمک متغیرهای Age, VOI, DTT.....۱۸
- شکل ۲۴ : خروجی آزمون post-hoc برای متغیر DTT در حالت بدون حذف outlier (سوژه‌ی ۳۹).....۱۹

مقدمه

به دلیل این که خرید کردن کالاها به خصوص مواد خوراکی، بسیار به تصورات خریدار از طعم و مزه‌ی آن ماده‌ی غذایی وابسته است، می‌شود از توصیفات درست برای کالاها استفاده کرد تا فروختن آن کالا تهسیل یابد.

هدف از این پژوهش بررسی اثر انواع مختلف توصیفات که می‌شود از یک قهوه کرد بر رفتار مصرف کننده می‌باشد. به عبارت دیگر این مقاله قصد دارد که نشان دهد زبان (language) می‌تواند رفتار خریدن کردن یک کالای خوراکی مثل قهوه توسط مصرف کننده را تحت تاثیر قرار دهد.

معرفی پژوهش و دادگان:

به کاربران ۳ نوع مختلف توصیف قهوه که به صورت تصادفی از ۴ دسته توصیف‌های قهوه (صرفاً با کلمات مبتنی بر منبع (source)، کلمات انتزاعی، کلمات ارزشی و ترکیبی از این سه دسته) انتخاب شده‌اند نشان داده‌شد. سپس از کاربران خواسته شد تا به سه سوال زیر پاسخ دهند:

۱. **Vividness of imagery (VOI):** 'Can you imagine the flavor of this coffee?' (scale ۱-۵)
۲. **Desire to taste (DTT):** 'Would you like to try this coffee?' (scale ۱-۵)
۳. **Willingness to pay (WTP):** 'How much are you willing to pay for a bag of ۵۰۰ g of this coffee?' (slider £۰ - £۵۰)

انواع توصیفات در جدول ۱ قابل رویت است. همچنین نمونه‌ای از پرسش‌نامه‌ای که توسط آن دادگان جمع‌آوری شده‌اند، در شکل ۱ قابل مشاهده است.

Coffee Flavour Descriptions Experiment A.

	Coffee 1	Coffee 2	Coffee 3	Coffee 4	Coffee 5
Source-based	honey chocolate	blueberry caramel	apple lemon	chocolate peanut	vanilla orange
Abstract	pear sweet acidic mild	almond balanced woody sweet	caramel volatile acidic fruity	flowers floral balanced soft	berry creamy sweet smooth
Evaluative	exciting distinct appealing	pleasant exciting lovely	delicate refreshing pleasant	exquisite pleasant wonderful	delightful interesting extraordinary
Combination	honey sweet exciting	blueberry balanced pleasant	caramel fruity delicate	chocolate floral exquisite	orange creamy extraordinary

جدول ۱



Can you imagine the flavour of this coffee?

1. No, not at all

2. No, vague

3. Okay, moderately clear

4. Yes, pretty clear

5. Yes, perfectly clear and vivid

Would you like to try this coffee?

1. No, not at all

2. No, not necessarily

3. Perhaps

4. Yes, I can give it a try

5. Yes, would love to try it

How much are you willing to pay for a bag of 500 grams of this coffee? Answer in £ (\$1 = £0.72 / €1 = £0.86).

0 5 10 15 20 25 30 35 40 45 50

£

شکل ۱: تصویر پرسشنامه

درواقع سه متغیر وابسته داریم که در ۴ دسته‌ی مختلف مقادیر آن‌ها موجود است. مقادیر میانگین، میانه، و انحراف معیار متغیرها برای ۴ دسته‌ی توصیفات در جدول ۲ قابل رویت است.

Means, standard deviations (in brackets), and medians for each of the dependent variables (Vividness of Imagery, Desire to Taste, and Willingness to Pay) across the four description conditions (source-based, abstract, evaluative, or a combined description) for Experiment A.

	Number of participants	Vividness of Imagery	Desire to Taste	Willingness to Pay (in GBP)
Source-based terms	$n = 39$	$M = 3.49 (SD = 0.89), Mdn = 3.6$	$M = 2.99 (SD = 0.82), Mdn = 3.0$	$M = 6.04 (SD = 3.04), Mdn = 5.69$
Abstract terms	$n = 39$	$M = 2.83 (SD = 0.76), Mdn = 2.8$	$M = 2.63 (SD = 0.54), Mdn = 2.6$	$M = 4.92 (SD = 2.52), Mdn = 4.32$
Evaluative terms	$n = 39$	$M = 2.00 (SD = 0.98), Mdn = 1.8$	$M = 2.57 (SD = 0.89), Mdn = 2.2$	$M = 4.91 (SD = 2.85), Mdn = 3.42$
Combination	$n = 39$	$M = 3.24 (SD = 0.70), Mdn = 3.2$	$M = 3.09 (SD = 0.61), Mdn = 3.0$	$M = 5.91 (SD = 2.94), Mdn = 5.25$

جدول ۲

همچنین تعداد کاربران ۴۰ بوده است که همگی از کشور بریتانیا و انگلیسی زبان هستند (برای جلوگیری از بروز موانع زبانی (language barrier)). ۱۳ مرد و ۲۶ زن در این آزمایش شرکت نموده با بازه‌ی سنی بین ۱۹ و ۶۷ سال (میانگین سنی ۳۴ سال است).

تعداد کل مشاهدات برای هر متغیر، برابر است با :

۴۰ (تعداد کاربران) * ۴ (تعداد دسته‌های توصیف) * ۵ (دفعه تکرار (trial) برای ۵ قهوه‌ی مختلف که با توصیف‌های مختلف انجام شده بود) = ۷۸۰

در این پژوهش هم از داده‌های کلی (trial-wise) و هم از دادگان میانگین گرفته شده (شکل ۲) برای trialها استفاده شده‌است. تعداد کل دادگان میانگین گرفته شده برابر است با ۴ * ۴۰ = ۱۶۰

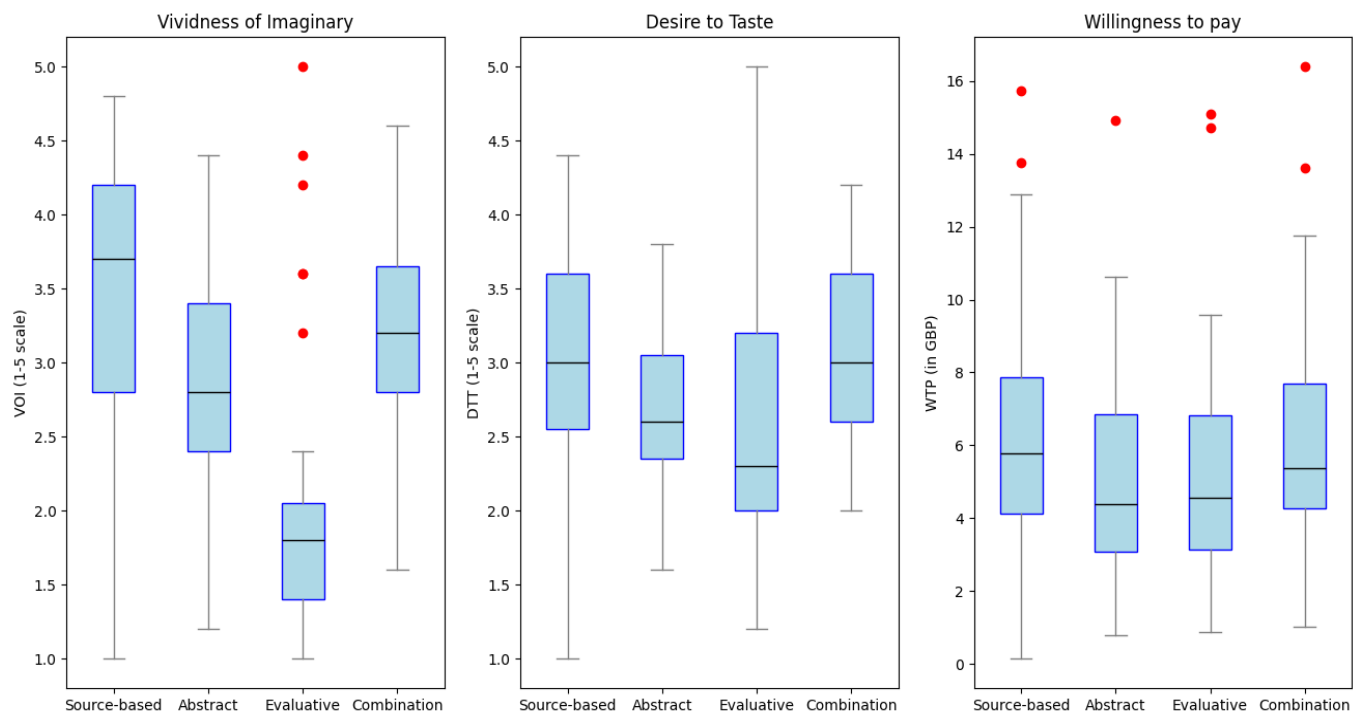
Participant	Description_type	VOI	DTT	WTP
1	Abstract	2.8	2.8	1.504
1	Combination	1.6	2.0	1.030
1	Evaluative	3.2	3.2	2.200
1	Source-based	1.0	1.0	0.140
2	Abstract	2.6	2.4	6.294
...	

160 rows × 31 columns

شکل ۲: قسمتی از دادگان میانگین گرفته شده

بررسی دادگان پرت

برای این منظور نمودارهای جعبه‌ای ۳ متغیر وابسته در ۴ دسته‌ی توصیفات مختلف رسم و بررسی شدند.
(شکل ۳)



شکل ۳: نمودار جعبه‌ای دادگان ۳ متغیر برای ۴ دسته‌ی توصیفات قهوه

با توجه به نمودار متغیر WTP و مقادیر بزرگ نزدیک ۱۵ بررسی شدند و معلوم شد که همگی متعلق به کاربر شماره‌ی ۳۹ بوده‌اند (شکل ۴)

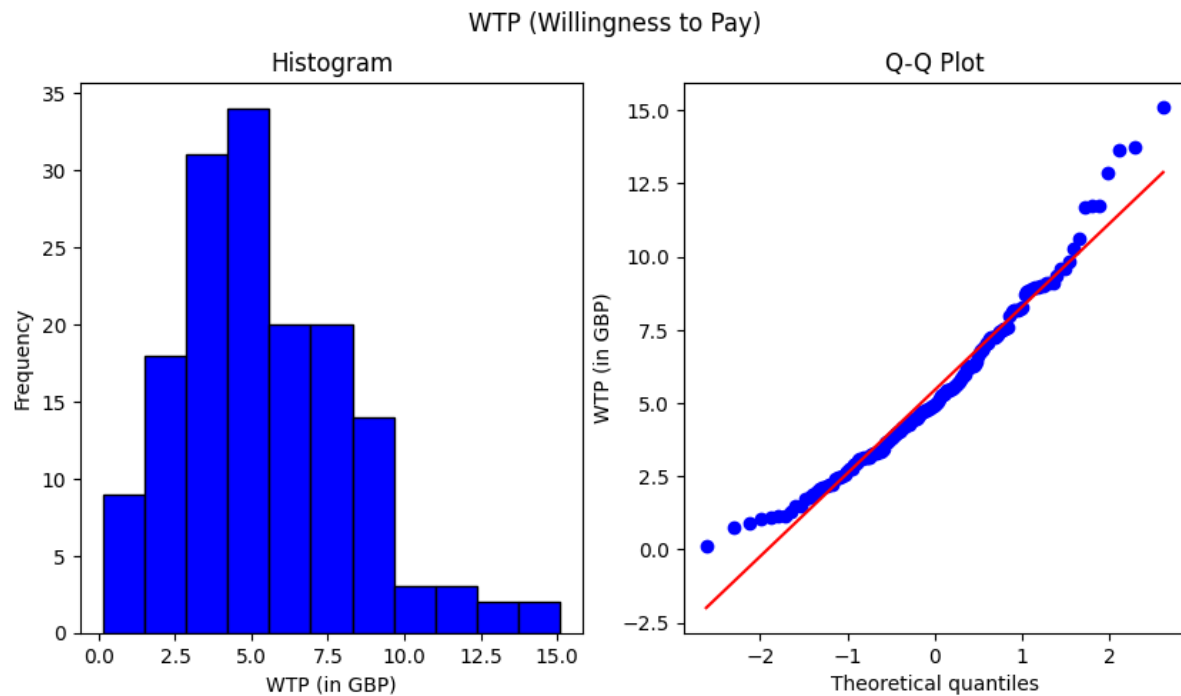
Participant	Description_type	WTP
33	Evaluative	15.098
39	Abstract	14.930
39	Combination	16.394
39	Evaluative	14.732
39	Source-based	15.746

شکل ۴ : مقادیر متغیر WTP برای کاربر شماره‌ی ۳۹

در این مقاله گفته‌شد که مقادیر WTP ی که فاصله‌ی بیشتر از سه انحراف معیار WTP و میانگین آن دارند، داده‌ی پرت در نظر گرفته شده‌اند و بایستی حذف شوند. البته این نکته قابل توجه است که نه تنها دادگان متغیر WTP برای کاربر ۳۹ حذف شدند، بلکه بقیه‌ی متغیرها یعنی VOI و DTT ی کاربر ۳۹ نیز حذف شدند (برای ثابت ماندن طول دادگان برای هر شرکت کننده)

بررسی نرمال بودن توزیع دادگان

برای بررسی توزیع دادگان، از هیستوگرام‌ها و نمودارهای Q-Q دادگان به همراه ۲ آزمون تست توزیع نرمال بودن کولموگروف اسمینورف و شپیرو-ویلک استفاده شد.



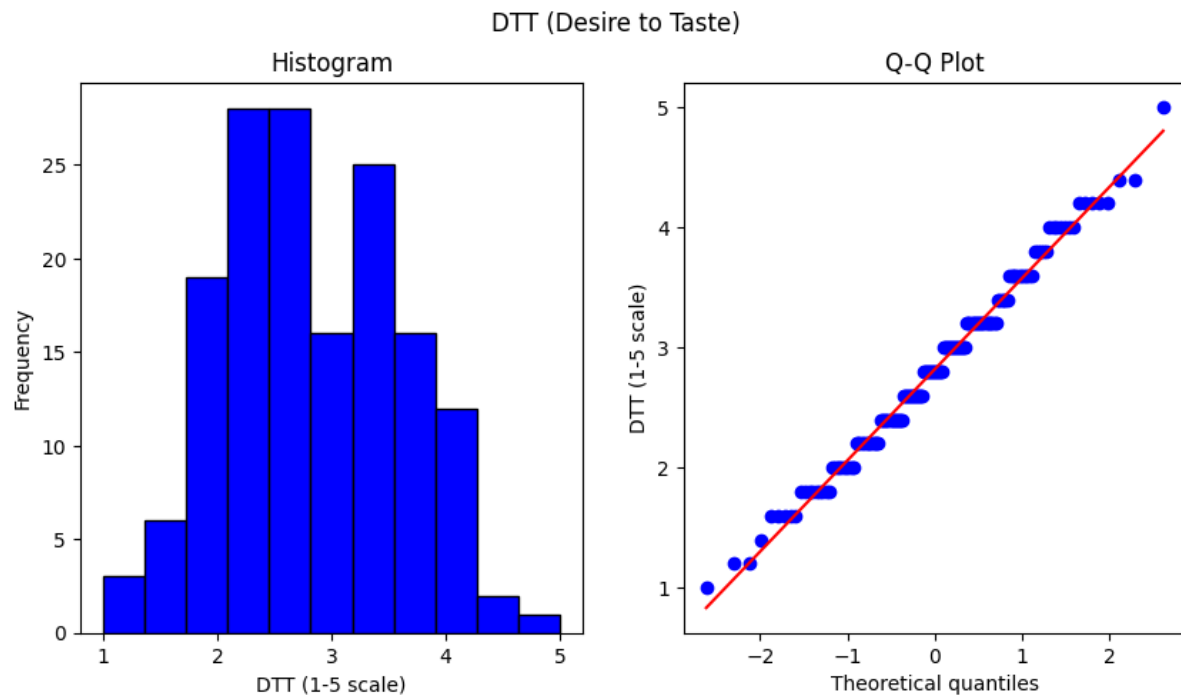
شکل ۵: هیستوگرام و نمودار Q-Q برای متغیر WTP

```

Shapiro-Wilk Test:
Statistic=0.962, p-value=0.0003
The data does not appear to be normally distributed.

Kolmogorov-Smirnov Test:
Statistic=0.893, p-value=0.0
The data does not appear to be normally distributed.
  
```

شکل ۶: نتایج آزمون‌های تست نرمال بودن توزیع متغیر WTP



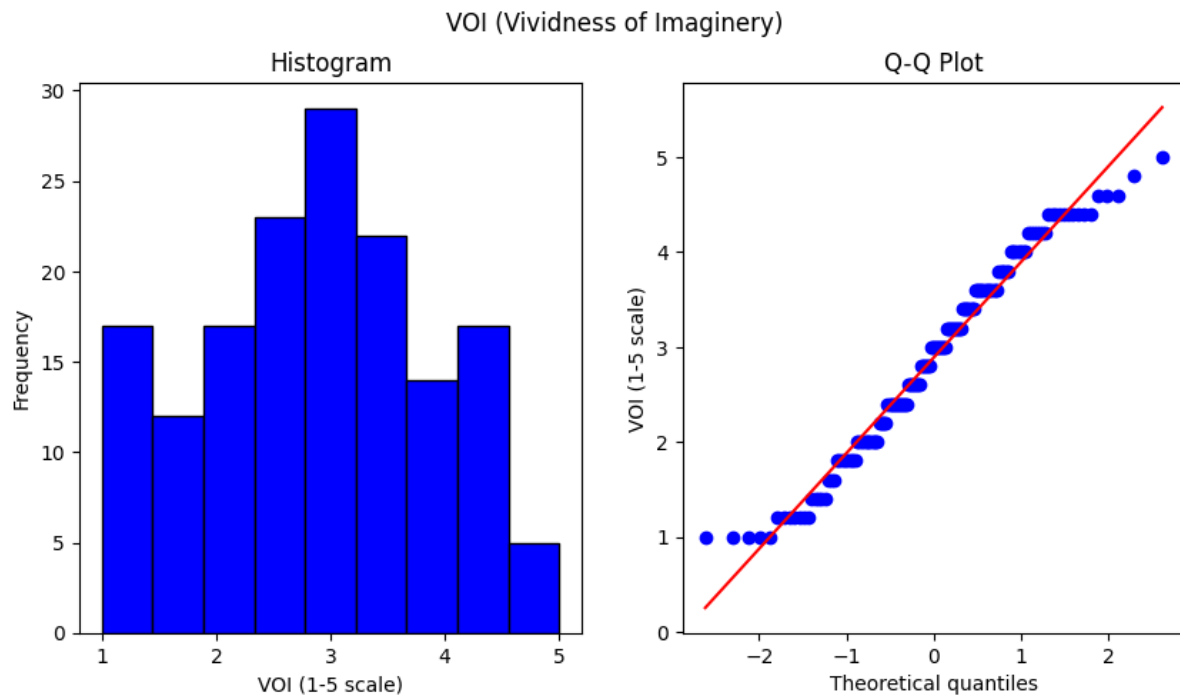
شکل ۷: هیستوگرام و نمودار Q-Q برای متغیر DTT

```

Shapiro-Wilk Test:
Statistic=0.989, p-value=0.2484
The data appears to be normally distributed.

Kolmogorov-Smirnov Test:
Statistic=0.92, p-value=0.0
The data does not appear to be normally distributed.
  
```

شکل ۸: نتایج آزمون‌های تست نرمال بودن توزیع متغیر DTT



شکل ۹: هیستوگرام و نمودار Q-Q برای متغیر VOI

```

Shapiro-Wilk Test:
Statistic=0.972, p-value=0.0026
The data does not appear to be normally distributed.

Kolmogorov-Smirnov Test:
Statistic=0.853, p-value=0.0
The data does not appear to be normally distributed.
  
```

شکل ۱۰: نتایج آزمون‌های تست نرمال بودن توزیع متغیر VOI

با توجه به نتایج بررسی نرمال بودن توزیع دادگان، متغیرهای WTP و VOI غیر نرمال و متغیر DTT صرفاً با آزمون شپیرو نرمال تلقی شد(با بقیه‌ی آزمون‌ها غیر نرمال تلقی گردید)

در نتیجه برای بیشتر شدن اطمینان تحلیل آماری و افزایش robustness، هم از آزمون‌های پارامتریک استفاده شده و هم از آزمون‌های غیر پارامتریک

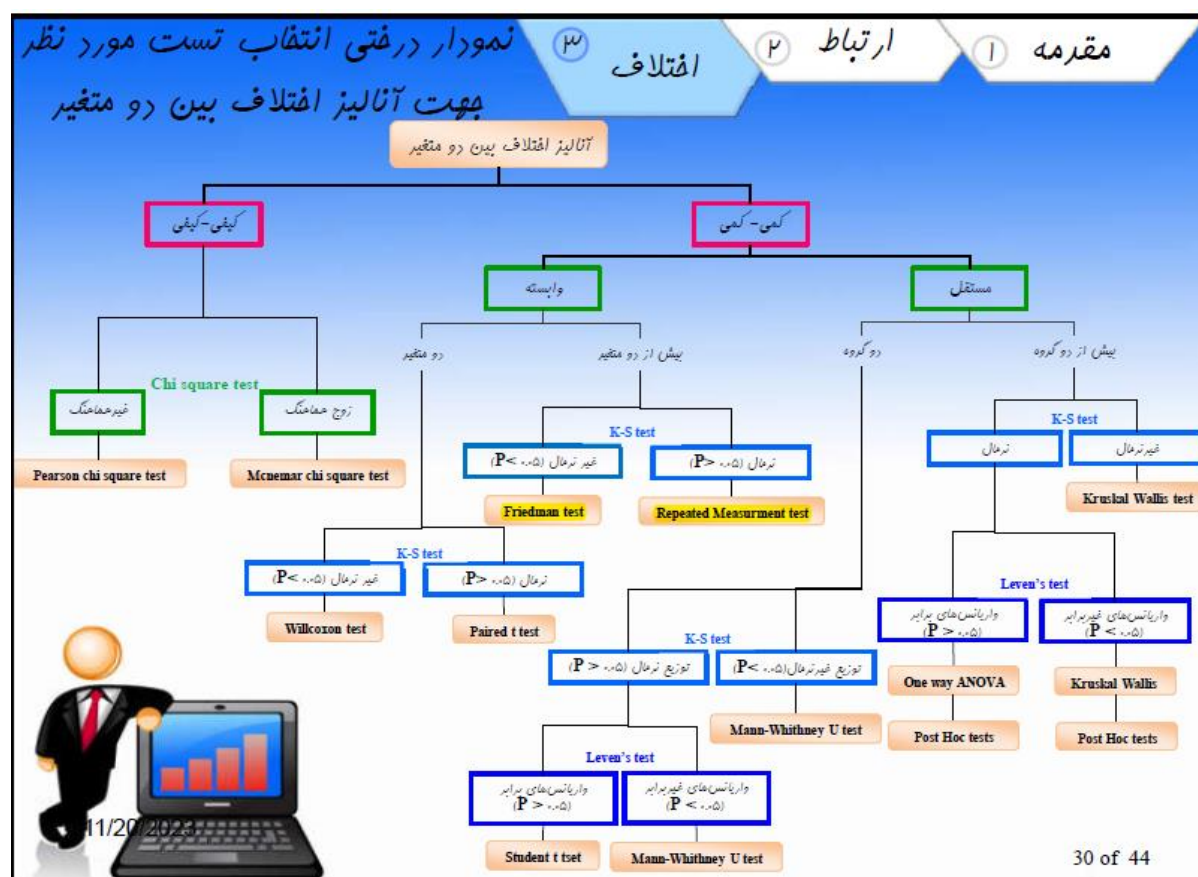
تحلیل‌های آماری

مروری بر ساختار دادگان

۳ متغیر وابسته با نام‌های VOI, DTT, WTP وجود دارند که برای ۴ دسته‌ی مختلف توصیف‌های قهوه، برای ۳۹ کاربر اندازه‌گیری شده‌اند. همچنین داده‌های هر شرکت‌کننده نسبت به هم وابستگی دارند.

انتخاب آزمون

با توجه به ساختار دادگان و چارت مربوط به انتخاب آزمون (از اسلایدهای خانم دکتر قاسمی) (شکل ۱۱) بایستی که از آزمون‌های repeated measure ANOVA و Friedman استفاده کرد (بیشتر از ۲ متغیر داریم و متغیرها به هم وابسته اند چون متغیرهای درون هر گرون مربوط به یک کاربر هستند و قطعاً وابسته اند). درست است که توزیع دادگان به نظر غیر نرمال می‌آیند ولی برای دقیق تر شدن نتایج و اطمینان بیشتر، از همتای پارامتریک آزمون Friedman یعنی repeated measure ANOVA هم استفاده شده است.



شکل ۱۱: راهنمای انتخاب آزمون جهت آنالیز اختلاف بین ۲ یا چند متغیر

بعد از انجام این ۲ آزمون بروی سه متغیر وابسته، در صورتی که اختلاف معنا داری بین حداقل ۲ گروه (از ۴ دسته‌ی توصیفات) وجود داشته باشد، از آزمون‌های pairwise برای مقایسه‌ی جفت دسته‌ها استفاده خواهیم کرد. برای آزمون‌های post-hok نیز هم از آزمون پارامتریک t و هم از همتای غیر پارامتریک آن wilcox paired test استفاده خواهیم کرد.

همچنین از معیار Bonferroni برای مقایسه‌ی p-value ها استفاده شده‌است. لذا مقدار آلفای اصلاح شده برابر است با آلفا تقسیم بر تعداد آزمایش‌ها.

$$\alpha_{adj} = \frac{\alpha}{m}, m = \binom{4}{2} = 6 \rightarrow \alpha_{adj} = 0.008$$

(آلفا برابر با 0.05 در نظر گرفته شده است)

نتایج آزمون‌های ANOVA/Friedman

نتایج این ۲ آزمون برای هر سه متغیر وابسته، در شکل ۱۲ قابل مشاهده‌است. از این نتایج استنباط می‌شود که حداقل بین ۲ تا از میانگین‌های جدول ۲، برای همه‌ی متغیرها اختلاف معناداری وجود دارد. برای بررسی اینکه بین کدام توصیف‌ها اختلاف وجود دارد، از آزمون‌های post-hok استفاده می‌کنیم.

Parametric test:

Repeated measure ANOVA for VOI:						
	Source	SS	DF	MS	F	Pr(>F)
0	Description_type	50.364	3	16.788	42.924	0.0
1	Error	44.586	114	0.391	NaN	NaN

Repeated measure ANOVA for DTT:						
	Source	SS	DF	MS	F	Pr(>F)
0	Description_type	7.798	3	2.599	7.916	0.0
1	Error	37.432	114	0.328	NaN	NaN

Repeated measure ANOVA for WTP:						
	Source	SS	DF	MS	F	Pr(>F)
0	Description_type	44.318	3	14.773	10.615	0.0
1	Error	158.648	114	1.392	NaN	NaN

Non-Parametric test:

Repeated measure Friedman Test Results for VOI:				
	n	Chi-squared statistic	df	P-value
0	39	54.64	3	0.0

Repeated measure Friedman Test Results for DTT:				
	n	Chi-squared statistic	df	P-value
0	39	24.59	3	0.0

Repeated measure Friedman Test Results for WTP:				
	n	Chi-squared statistic	df	P-value
0	39	28.88	3	0.0

شکل ۱۲ : نتایج آزمون‌های ANOVA و Friedman برای هر سه متغیر VOI, DTT, WTP

نتایج آزمون‌های post-hoc

نتایج برای متغیرهای VOI, DTT, WTP به ترتیب در شکل‌های ۱۳ و ۱۴ و ۱۵ قابل رویت می‌باشند.

Parametric test:

	Group 1	Group 2	P-value	Bonferroni Corrected
0	Abstract	Combination	0.0003	True
1	Abstract	Evaluative	0.0000	True
2	Abstract	Source-based	0.0001	True
3	Combination	Evaluative	0.0000	True
4	Combination	Source-based	0.0176	False
5	Evaluative	Source-based	0.0000	True

Non-Parametric test:

	Group 1	Group 2	P-value	Bonferroni Corrected
0	Abstract	Combination	0.0004	True
1	Abstract	Evaluative	0.0000	True
2	Abstract	Source-based	0.0002	True
3	Combination	Evaluative	0.0000	True
4	Combination	Source-based	0.0273	False
5	Evaluative	Source-based	0.0000	True

شکل ۱۳: نتایج آزمون‌های post-hok پارامتریک و غیر پارامتریک برای متغیر VOI

با توجه به شکل ۱۳، استنباط می‌شود که بین تمامی جفت دسته‌ها اختلاف معنی دار وجود دارد به غیر از جفت (Combination & Source-based). در ضمن، مقادیر p-value با مقدار اصلاح‌شده‌ی Bonferroni که ($\alpha_{adj} = 0.008$) بود مقایسه می‌شوند.

نکته: در مقاله، با وجود اینکه ذکر شده است که از آلفای اصلاح‌شده‌ی Bonferroni استفاده شده‌است، اما برای بررسی و گزارش نتایج آزمون‌های post-hok، p-value های کوچک تر از 0.05 را نیز معنادار در نظر گرفته‌است!

Parametric test:

	Group 1	Group 2	P-value	Bonferroni Corrected
0	Abstract	Combination	0.0000	True
1	Abstract	Evaluative	0.6209	False
2	Abstract	Source-based	0.0096	False
3	Combination	Evaluative	0.0005	True
4	Combination	Source-based	0.2061	False
5	Evaluative	Source-based	0.0289	False

Non-Parametric test:

	Group 1	Group 2	P-value	Bonferroni Corrected
0	Abstract	Combination	0.0001	True
1	Abstract	Evaluative	0.5153	False
2	Abstract	Source-based	0.0094	False
3	Combination	Evaluative	0.0011	True
4	Combination	Source-based	0.2843	False
5	Evaluative	Source-based	0.0255	False

شکل ۱۴: نتایج آزمون‌های post-hok پارامتریک و غیر پارامتریک برای متغیر DTT

با توجه به شکل ۱۴، استنباط می‌شود که بین تمامی جفت دسته‌ها اختلاف معنی دار وجود ندارد به غیر از جفت‌های (Combination & Abstract) و (Combination & Evaluative).

Parametric test:

	Group 1	Group 2	P-value	Bonferroni Corrected
0	Abstract	Combination	0.0000	True
1	Abstract	Evaluative	0.9802	False
2	Abstract	Source-based	0.0002	True
3	Combination	Evaluative	0.0017	True
4	Combination	Source-based	0.4514	False
5	Evaluative	Source-based	0.0043	True

Non-Parametric test:

	Group 1	Group 2	P-value	Bonferroni Corrected
0	Abstract	Combination	0.0000	True
1	Abstract	Evaluative	0.6637	False
2	Abstract	Source-based	0.0004	True
3	Combination	Evaluative	0.0001	True
4	Combination	Source-based	0.4512	False
5	Evaluative	Source-based	0.0013	True

شکل ۱۵: نتایج آزمون‌های post-hoc پارامتریک و غیر پارامتریک برای متغیر WTP

با توجه به شکل ۱۵، استنباط می‌شود که بین همه‌ی جفت دسته‌ها اختلاف معنی دار وجود دارد به غیر از جفت‌های (Evaluative & Abstract) و (Combination & Source-based).

نکته: در آزمون‌های ANOVA/Friedman و آزمون‌های پارامتریک و غیر پارامتریک Post-hok، نتایج مشابه بودند (با اختلافات جزئی). این نتیجه نشان می‌دهد که تحلیل‌های آماری مان درست عمل کرده و قدرت اطمینان بالایی دارند (به اصطلاح، نتایج robust هستند و به نوع توزیع دادگان وابسته نبودند).

توضیح مدل رگرسیون خطی با اثرهای مختلط (LMER)

این مدل مدل‌های رگرسیون خطی رایج را به منظور در نظر گرفتن همزمان تأثیرات ثابت و تأثیرات تصادفی گسترش می‌دهد. این مدل به ویژه در مواقعی که با داده‌های خوشه‌ای یا تکرار شده سر و کار داریم و مشاهدات مستقل نیستند، بسیار مفید است. به ترتیب، مفاهیم کلیدی رگرسیون خطی با اثرات مختلط عبارتند از:

اثرات ثابت (اثرات جمعیتی): این‌ها به مانند ضرایب در رگرسیون خطی سنتی هستند. آن‌ها نشان‌دهنده رابطه میان متغیرهای مستقل و وابسته در میانه همه سطوح متغیر گروهی هستند.

اثرات تصادفی (اثرات فردی): این‌ها پارامترهای اضافی هستند که به عنوان متغیرهای تصادفی در نظر گرفته می‌شوند و اجازه می‌دهند تغییرات میانگین افراد یا گروه‌ها را در نظر بگیرند. این اثرات تصادفی تنوع در سطوح فردی را که به طور صریح توسط اثرات ثابت نماینده نمی‌شود، ضبط می‌کنند.

ساختار سلسله مراتبی: داده‌ها اغلب به صورت سلسله مراتبی سازماندهی شده‌اند، با واحدهای سطح پایین‌تر (مثل اندازه‌گیری‌های تکراری یا افراد) درون واحدهای سطح بالاتر (مثل گروه‌ها یا خوشه‌ها). مدل همزمان با تنوع داخلی و بیرونی گروه‌ها سازگاری دارد.

اجزای تنوع: این مدل فرض می‌کند که تنوع کلیت در متغیر وابسته می‌تواند به اجزاء ثابت و تصادفی تقسیم شود. اجزاء تنوع نشان‌دهنده میزان تغییرات در سطوح مختلف سلسله مراتبی هستند.

ساختار همبستگی: این مدل امکان تعیین ساختارهای همبستگی را فراهم می‌کند، که احتمال همبستگی‌های ممکن بین مشاهدات در یک گروه را نشان می‌دهد.

فرمول‌بندی یک مدل رگرسیون خطی با اثرات مختلط ممکن است به صورت زیر باشد

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \text{random effects} + \epsilon_{ij}$$

نتایج آزمون LMER برای متغیرهای جانبی

برای همه‌ی *Trial*‌ها

برای این آزمون، از کل *Trial*‌ها استفاده شده است (۷۸۰ مشاهده)
دستور زیر نمونه‌ای از دستور اجرای مدل LMER می‌باشد.

```
formula = "VOI ~ Description_type + ImageryM + ExpertiseM + DrinkCoffeeNum"
# Fit the mixed-effects model
mixed_model = smf.mixedlm(formula, data_by_trial, groups=data_by_trial["Participant"], re_formula="~Description_type").fit()
```

در این جا قصد داریم اثرات متغیرهای **self-reported** که توسط کاربران پر شده است (پرسش‌نامه‌ی مربوط به این قسمت برای یک پژوهش درمورد شراب بود که ارجاع داده شده بود به مقاله‌ای دیگر. اما داده‌های برای هر دفعه آزمایش (*trial*) در دسترس می‌باشد)

- متغیر **ImageryM** میانگین دفعات‌یست که فرد قدرت تصور کردن خود را امتیاز دهی کرده (از ۱ تا ۵)

- متغیر **ExpertiseM** هم میانگین دفعات‌یست که فرد میزان تخصص خود را در مورد انواع قهوه و طعم‌های آن امتیاز دهی کرده (۱ تا ۵)

- متغیر DrinkCoffeeNum نیز تعداد دفعاتی که فرد در روز قهوه مصرف میکند (به صورت میانگین و معمول) را بیان می کند.

Mixed Linear Model Regression Results							
=====							
Model:	MixedLM	Dependent Variable:				VOI	
No. Observations:	780	Method:				REML	
No. Groups:	39	Scale:				0.6223	
Min. group size:	20	Log-Likelihood:				-1023.4003	
Max. group size:	20	Converged:				Yes	
Mean group size:	20.0						

		Coef.	Std.Err.	z	P> z	[0.025	0.975]

Intercept		2.315	0.671	3.450	0.001	1.000	3.631
Description_type[T.Combination]		0.415	0.104	3.982	0.000	0.211	0.620
Description_type[T.Concrete]		0.667	0.150	4.445	0.000	0.373	0.961
Description_type[T.Evaluative]		-0.831	0.120	-6.913	0.000	-1.066	-0.595
ImageryM		-0.031	0.177	-0.178	0.859	-0.378	0.315
ExpertiseM		-0.035	0.068	-0.506	0.613	-0.169	0.100
DrinkCoffeeNum		0.348	0.161	2.158	0.031	0.032	0.665

شکل ۱۶: نتایج مدل LMER برای پیش بینی متغیر VOI برای همه ی trialها به کمک متغیرهای self-reported

با توجه به شکل ۱۶، متغیر VOI صرفا توسط میزان نوشیدن قهوه به صورت رایج تحت تاثیر قرار گرفته. یعنی نوشیدن یک قهوه بیشتر در روز، باعث افزایش Vividness of Imagery به اندازه ی 0.348 می شود.

Mixed Linear Model Regression Results						
Model:	MixedLM	Dependent Variable:			DTT	
No. Observations:	780	Method:			REML	
No. Groups:	39	Scale:			0.8160	
Min. group size:	20	Log-Likelihood:			-1102.1943	
Max. group size:	20	Converged:			Yes	
Mean group size:	20.0					
		Coef.	Std.Err.	z	P> z	[0.025 0.975]
Intercept		2.061	0.539	3.823	0.000	1.004 3.117
Description_type[T.Combination]		0.462	0.104	4.428	0.000	0.257 0.666
Description_type[T.Concrete]		0.354	0.132	2.678	0.007	0.095 0.613
Description_type[T.Evaluative]		-0.062	0.124	-0.496	0.620	-0.305 0.182
ImageryM		0.127	0.147	0.866	0.387	-0.161 0.415
ExpertiseM		-0.012	0.055	-0.213	0.831	-0.119 0.095
DrinkCoffeeNum		0.047	0.130	0.358	0.720	-0.209 0.302

شکل ۱۷: نتایج مدل LMER برای پیش بینی متغیر DTT برای همه ی trialها به کمک متغیرهای self-reported

با توجه به شکل ۱۷، هیچ کدام از متغیرهای self-reported، تاثیری بر DTT ندارند.

Mixed Linear Model Regression Results							
=====							
Model:	MixedLM	Dependent Variable:				WTP	
No. Observations:	780	Method:				REML	
No. Groups:	39	Scale:				2.3051	
Min. group size:	20	Log-Likelihood:				-1553.7793	
Max. group size:	20	Converged:				Yes	
Mean group size:	20.0						

		Coef.	Std.Err.	z	P> z	[0.025	0.975]

Intercept		1.072	2.743	0.391	0.696	-4.304	6.448
Description_type[T.Combination]		0.993	0.218	4.561	0.000	0.566	1.420
Description_type[T.Concrete]		1.125	0.271	4.154	0.000	0.594	1.656
Description_type[T.Evaluative]		-0.006	0.230	-0.025	0.980	-0.456	0.444
ImageryM		0.965	0.725	1.330	0.184	-0.457	2.386
ExpertiseM		-0.343	0.267	-1.285	0.199	-0.866	0.180
DrinkCoffeeNum		0.737	0.635	1.162	0.245	-0.507	1.981

شکل ۱۸: نتایج مدل LMER برای پیش‌بینی متغیر WTP برای تمامی trialها به کمک متغیرهای self-reported با توجه به شکل ۱۸ هیچ کدام از متغیرهای self-reported، تاثیری بر WTP ندارند.

برای حالت میانگین گرفته‌شده

حال از داده‌های میانگین گرفته شده استفاده می‌کنیم (۱۵۶ مشاهده)

Mixed Linear Model Regression Results						
Model:	MixedLM	Dependent Variable:	VOI			
No. Observations:	156	Method:	REML			
No. Groups:	39	Scale:	0.1163			
Min. group size:	4	Log-Likelihood:	-159.3782			
Max. group size:	4	Converged:	Yes			
Mean group size:	4.0					
		Coef.	Std.Err.	z	P> z	[0.025 0.975]
Intercept		2.316	0.670	3.454	0.001	1.002 3.630
Description_type[T.Combination]		0.415	0.103	4.014	0.000	0.213 0.618
Description_type[T.Evaluative]		-0.831	0.120	-6.913	0.000	-1.066 -0.595
Description_type[T.Source-based]		0.667	0.150	4.447	0.000	0.373 0.960
ImageryM		-0.032	0.176	-0.179	0.858	-0.376 0.313
ExpertiseM		-0.035	0.068	-0.511	0.609	-0.168 0.098
DrinkCoffeeNum		0.348	0.161	2.157	0.031	0.032 0.664

شکل ۱۹: نتایج مدل LMER برای پیش‌بینی متغیر VOI برای میانگین trialها به کمک متغیرهای self-reported

Mixed Linear Model Regression Results						
=====						
Model:	MixedLM	Dependent Variable:			DTT	
No. Observations:	156	Method:			REML	
No. Groups:	39	Scale:			0.0829	
Min. group size:	4	Log-Likelihood:			-141.9618	
Max. group size:	4	Converged:			Yes	
Mean group size:	4.0					

	Coef.	Std.Err.	z	P> z	[0.025 0.975]	

Intercept	2.100	0.526	3.994	0.000	1.070	3.131
Description_type[T.Combination]	0.462	0.098	4.720	0.000	0.270	0.653
Description_type[T.Evaluative]	-0.062	0.123	-0.499	0.618	-0.303	0.180
Description_type[T.Source-based]	0.354	0.130	2.729	0.006	0.100	0.608
ImageryM	0.114	0.142	0.801	0.423	-0.165	0.392
ExpertiseM	-0.013	0.054	-0.244	0.808	-0.118	0.092
DrinkCoffeeNum	0.056	0.127	0.443	0.658	-0.193	0.305

شکل ۲۰: نتایج مدل LMER برای پیش‌بینی متغیر DTT برای میانگین trialها به کمک متغیرهای self-reported

Mixed Linear Model Regression Results						
=====						
Model:	MixedLM	Dependent Variable:			WTP	
No. Observations:	156	Method:			REML	
No. Groups:	39	Scale:			0.4323	
Min. group size:	4	Log-Likelihood:			-281.0453	
Max. group size:	4	Converged:			Yes	
Mean group size:	4.0					

	Coef.	Std.Err.	z	P> z	[0.025	0.975]

Intercept	1.067	2.734	0.390	0.696	-4.292	6.426
Description_type[T.Combination]	0.993	0.217	4.584	0.000	0.569	1.418
Description_type[T.Evaluative]	-0.006	0.230	-0.025	0.980	-0.456	0.444
Description_type[T.Source-based]	1.125	0.270	4.161	0.000	0.595	1.655
ImageryM	0.970	0.724	1.339	0.180	-0.450	2.390
ExpertiseM	-0.344	0.267	-1.290	0.197	-0.867	0.179
DrinkCoffeeNum	0.732	0.633	1.156	0.248	-0.509	1.974

شکل ۲۱: نتایج مدل LMER برای پیش‌بینی متغیر WTP برای میانگین trialها به کمک متغیرهای self-reported

با توجه به شکل‌های ۱۹، ۲۰، ۲۱، نتایج برای داده‌های میانگین گرفته شده، بسیار مشابه حالت همه‌ی trialهای می‌باشد.

نکته: در مقاله‌ی اصلی، صرفاً از داده‌های همه‌ی Trialها استفاده شده و مقایسه‌ای با حالت میانگین گرفته شده انجام نداده.

نتایج آزمون LMER برای پیش‌بینی WTP

در این آزمون قصد داریم که متغیر WTP را به کمک متغیرهای VOI و DTT و Age (به عنوان Covariate) پیش‌بینی کنیم.

توضیح متغیر Covariate

متغیر همراه یا متغیر هم‌بسته (به انگلیسی: Covariate)، یک متغیر است که با فرض یک ارتباط یا همبستگی با متغیر وابسته در یک مدل آماری در نظر گرفته می‌شود. به عبارت دیگر، انتظار می‌رود که این متغیر تأثیری بر روی نتیجه مورد نظر دارد.

در سیاق تحلیل آماری، متغیرهای هم‌بسته برای کنترل عوامل تداخلی و در نظر گرفتن اختلافات متغیر وابسته که موضوع اصلی نیستند، به کار می‌روند. اضافه کردن متغیرهای هم‌بسته به یک مدل به محققان کمک می‌کند تا تأثیر متغیرهای مستقل را بر متغیر وابسته با کنترل عوامل مرتبط دیگر، جدا کنند.

به عنوان مثال، اگر شما قصد دارید رابطه بین میزان تمرین (متغیر مستقل) و افت وزن (متغیر وابسته) را بررسی کنید، سن، جنسیت و وزن اولیه ممکن است به عنوان متغیرهای هم‌بسته در نظر گرفته شوند. با اضافه کردن این متغیرهای هم‌بسته به تحلیل، شما به هدف دارید تا تأثیر خاص تمرین بر افت وزن را در حالی که اثرات احتمالی سایر عوامل را ثابت نگه دارید، تمییز دهید.

خلاصه‌اش، متغیرهای هم‌بسته متغیرهای اضافی در یک مدل آماری هستند که برای کنترل اثرات تداخلی احتمالی و بهبود دقت تخمین‌ها از رابطه میان متغیرهای مستقل و وابسته استفاده می‌شوند.

برای همه‌ی *Trial*ها

دستور مورد نظر:

```
formula = "WTP ~ Description_type + VOI + DTT + Age"

# Fit the mixed-effects model
mixed_model = smf.mixedlm(formula, data_by_trial, groups=data_by_trial["Participant"], re_formula="~Description_type").fit()
```

Mixed Linear Model Regression Results						
Model:	MixedLM	Dependent Variable:		WTP		
No. Observations:	780	Method:		REML		
No. Groups:	39	Scale:		1.2832		
Min. group size:	20	Log-Likelihood:		-1332.1920		
Max. group size:	20	Converged:		Yes		
Mean group size:	20.0					
	Coef.	Std.Err.	z	P> z	[0.025	0.975]
Intercept	-0.008	1.204	-0.001	0.935	-2.457	2.261
Description_type[T.Combination]	0.386	0.162	2.374	0.018	0.067	0.704
Description_type[T.Concrete]	0.485	0.190	2.549	0.011	0.112	0.858
Description_type[T.Evaluative]	0.463	0.148	3.126	0.002	0.173	0.753
VOI	0.500	0.061	8.188	0.000	0.380	0.619
DTT	0.867	0.054	15.919	0.000	0.760	0.974
Age	0.038	0.033	1.164	0.244	-0.026	0.103

شکل ۲۲: نتایج مدل LMER برای پیش‌بینی متغیر WTP برای تمامی trial ها به کمک متغیرهای Age, VOI, DTT

با توجه به شکل ۲۲، نتیجه می‌گیریم که:

متغیرهای VOI و DTT به مقدار قابل توجهی، با WTP ارتباط خطی دارند. با افزایش ۱ واحد متغیر VOI یا DTT، به ترتیب متغیر WTP به اندازه‌ی 0.5 و 0.867 پوند افزایش پیدا می‌کند. همچنین متغیر Age یا سن، ارتباطی با افزایش یا کاهش WTP ندارد.

Mixed Linear Model Regression Results						
Model:	MixedLM	Dependent Variable:		WTP		
No. Observations:	156	Method:		REML		
No. Groups:	39	Scale:		0.2661		
Min. group size:	4	Log-Likelihood:		-240.6466		
Max. group size:	4	Converged:		Yes		
Mean group size:	4.0					
	Coef.	Std.Err.	z	P> z	[0.025	0.975]
Intercept	-0.712	1.226	-0.581	0.562	-3.115	1.691
Description_type[T.Combination]	0.280	0.171	1.636	0.102	-0.055	0.615
Description_type[T.Evaluative]	0.599	0.182	3.291	0.001	0.242	0.955
Description_type[T.Source-based]	0.349	0.202	1.725	0.085	-0.048	0.746
VOI	0.657	0.152	4.324	0.000	0.359	0.954
DTT	0.955	0.163	5.844	0.000	0.635	1.275
Age	0.037	0.033	1.126	0.260	-0.027	0.100

شکل ۲۳: نتایج مدل LMER برای پیش‌بینی متغیر WTP برای میانگین trial ها به کمک متغیرهای Age, VOI, DTT

با توجه به شکل ۲۳، نتیجه می‌گیریم که نتایج نسبتاً مشابه حالت تمامی Trial ها می‌باشد با این تفاوت

که این دفعه، با افزایش ۱ واحد متغیر VOI یا DTT، به ترتیب متغیر WTP به اندازه‌ی 0.675 و 0.955 پوند افزایش پیدا می‌کند.

نکته:

در فایل (code no outlier) ، دیگر داده‌های شرکت‌کننده‌ی ۳۹ را حذف نکردم و تمام تحلیل‌ها را تکرار کردم. نتایج کاملاً مشابه شد با این تفاوت جزئی که در آزمون post-hok برای متغیر DTT، بین جفت‌های (Abstract & Source-based) نیز تفاوت معنی دار وجود دارد (P-value = 0.007 < α_{adj} = 0.008) (شکل ۲۴)

	Group 1	Group 2	P-value	Bonferroni Corrected
0	Abstract	Combination	0.0000	True
1	Abstract	Evaluative	0.5649	False
2	Abstract	Source-based	0.0070	True
3	Combination	Evaluative	0.0003	True
4	Combination	Source-based	0.2060	False
5	Evaluative	Source-based	0.0210	False

شکل ۲۴: خروجی آزمون post-hok برای متغیر DTT در حالت بدون حذف outlier (سوژه‌ی ۳۹)

ولی در حالت قبل که سوژه‌ی ۳۹ را حذف کرده بودیم، بین جفت‌های (Abstract & Source-based) تفاوت معنی داری وجود نداشت . (P-value = 0.0096 > α_{adj} = 0.008) (شکل ۱۴)

نتیجه گیری

- متغیرهای VOI و DTT با WTP ارتباط خطی دارند، (نتیجه‌ی حاصل‌شده از مدل LMER پیش‌بینی WTP).
- همچنین این سه متغیر خودشان، تحت تاثیر نوع توصیف قهوه‌ی مورد نظر می‌باشند.
- بیشترین مقدار برای این سه متغیر زمانی اتفاق افتاد که از توصیفات combination استفاده می‌شد.
- لذا پاسخ سوال پژوهشی‌مان (آیا انواع مختلف توصیفات قهوه بر رفتار مصرف کننده و میزان تصور پذیری وی تاثیر دارد؟) بله است.

تحلیل‌های فراتر؟

در این پروژه‌ی صرفاً آزمایش A این مقاله بررسی شد که از کلمات تکی به عنوان توصیفات قهوه‌ها استفاده شده بود. در این مقاله یک آزمایش دیگر به نام آزمایش B انجام شده است که در آن به جای استفاده از کلمات، از جمله‌های کوتاه حاوی این صفت‌ها استفاده شده. همچنین نتایج این ۲ آزمایش باهم مقایسه‌گردیده است. اگر وقت کافی بود بسیار تمایل داشتم که آزمایش B را هم پیاده‌سازی نمایم.

با تشکر از توجه شما.

پایان