

به نام خدا



دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر



درس علوم شناختی

تمرین شماره ۲

نام و نام خانوادگی : مهیار ملکی

شماره دانشجویی : ۸۱۰۱۰۰۴۷۶

بهار ۱۴۰۱

۱. پاک سازی داده ها :

در این بخش مراحل زیر انجام شد :

۱. تصحیح موارد گزارش شده در گروه درس مانند زمان واکنش هایی که به جای ثانیه به میلی ثانیه ذخیره شده

بود یا لوکیش هایی که به جای بازه ۱-۱۳ از ۰ تا ۱۲ نام گذاری شده بود.

۲. حذف اوتلایرها (زمان واکنش های کمتر از ۱۰۰ میلی ثانیه و بیشتر از $1.5 * IQR + Q3$)

۳. بازسازی ستون دست استفاده شده (چپ یا راست) از روی ستون کلید

۴. پاکسازی ستون محرک با استفاده از کتابخانه `regex`

۵. حذف سطرهای دارای خانه های خالی که قابل بازیابی نیستند

۶. حذف ستون اندازه

در مرحله پاک سازی داده ها، ما تنها مجاز به حذف حداکثر ۲ درصد از داده ها هستیم، اما در اینجا پس از انجام مراحل گفته شده حدودا ۹ درصد داده ها حذف شدند. این امر در اینجا غیر قابل اجتناب بود زیرا اوتلایرها و خانه های خالی زیادی در داده های این مسئله وجود دارد.

۲. کاوش داده ها :

ابتدا تغییرات خواسته شده را اعمال و ستون هایی که خواسته شده را ایجاد می کنیم:

۱. تبدیل واحد زمان زمان واکنش از ثانیه به میلی ثانیه

۲. ایجاد ستون lrt : لگاریتم زمان واکنش

۳. ایجاد ستون ecc : گریز از مرکز یا موقعیت مکانی محرکها بر اساس شعاع (مرکز - دایره داخلی - دایره بیرونی)

۴. ایجاد ستون vf : میدان دیداری محرک ها (مرکز - راست - بالا - چپ - پایین)

۵. ایجاد ستون udh : زمانی که افراد از دست غالب خود استفاده میکنند مقدار ۱ و در غیر این صورت صفر

۶. ایجاد ستون acc : با توجه به جواب های درست و جواب های داده شده مقدار ۱ برای جواب درست و مقدار

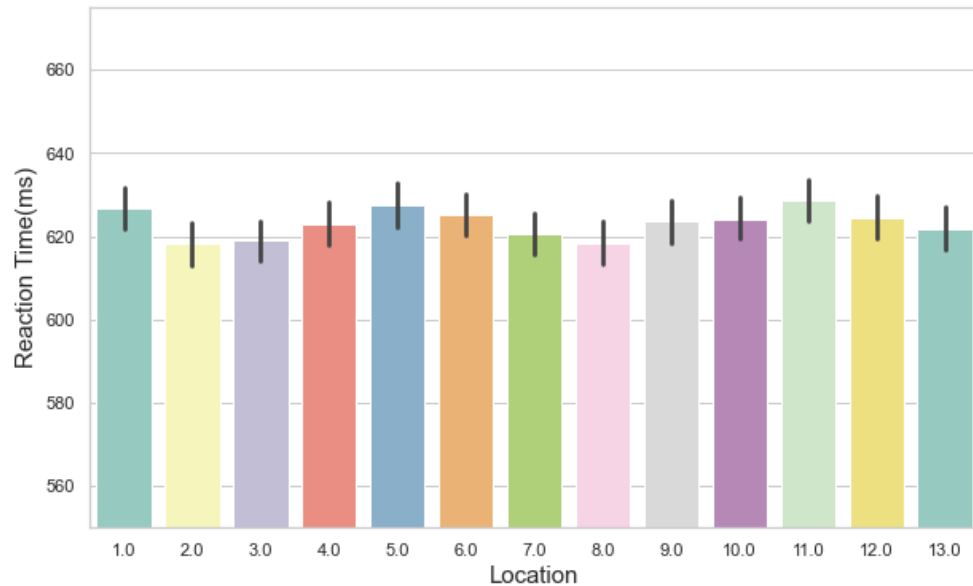
صفر برای جواب غلط قرار می گیرد

سپس در ادامه به رسم باریلات و انجام تست آماری ANOVA برای هر یک از متغیرهای مستقل به صورت جداگانه با هر یک از متحرک های وابسته (دقت و زمان واکنش) می پردازیم و ارتباط متغیرها با هم را بررسی می کنیم.

- **موقعیت مکانی محرک :**

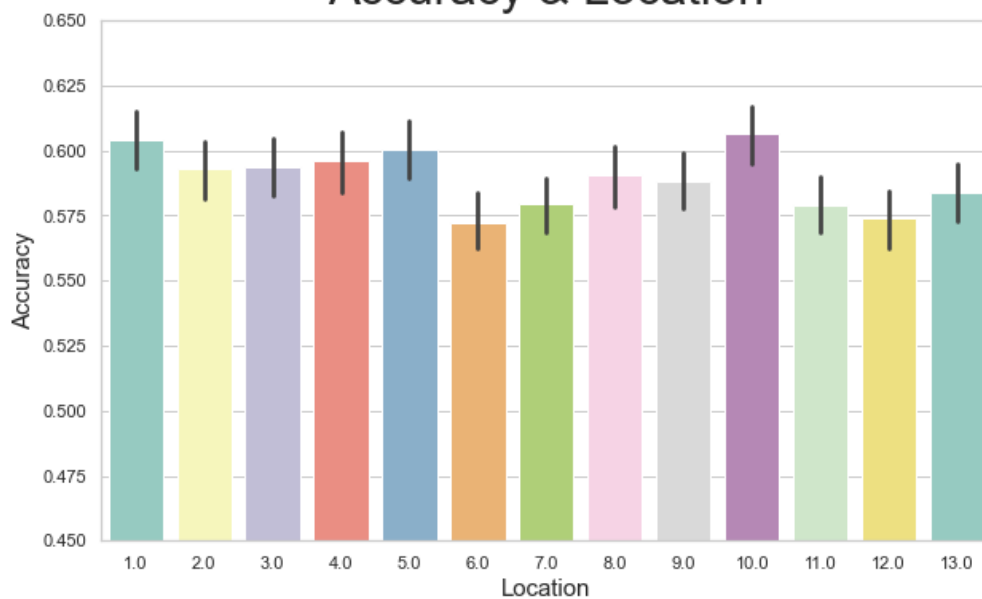
با بررسی نمودار شکل ۱ مشاهده می‌شود زمانی که محرک در مرکز یا سمت چپ قرار می‌گیرد کندترین زمان واکنش و وقتی در سمت راست قرار می‌گیرد سریعترین زمان واکنش اتفاق می‌افتد. در ادامه با بررسی نمودار شکل ۲ تقریباً می‌توان نتیجه گرفت که کندترین واکنش‌ها که در موقعیت‌های مکانی مرکز و چپ قرار داشتند، بیشترین دقت را حاصل شده‌اند.

Reaction Time & Location



شکل ۱- نمودار زمان واکنش بر اساس موقعیت محرک

Accuracy & Location



شکل ۲- نمودار دقت پاسخ بر اساس موقعیت محرک

حال با انجام تست آماری ANOVA می‌خواهیم بفهمیم که آیا برای جفت متغیرهای مکان محرک - زمان واکنش و مکان محرک - دقت پاسخ، فرض صفر رد می‌شود یا خیر.

جدول ۱

	P_value	R_squared	AIC	BIC
rt & Location	0.051137	0.000209	1.372040e+06	1.372163e+06

جدول ۲

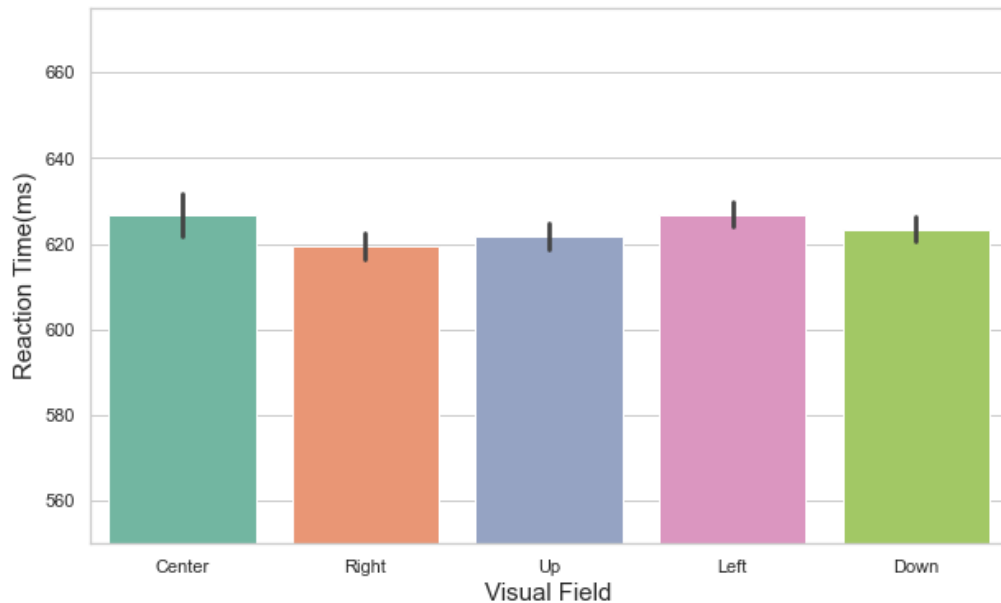
	P_value	R_squared	AIC	BIC
acc & Location	0.000005	0.000468	141972.152729	142095.828038

در جدول ۱ مشاهده می‌شود که مقدار p -value برابر با ۰.۰۵ شده است لذا نمی‌توان به طور قطع رد شدن یا تایید فرض صفر و تاثیرگذاری پارامتر مکان محرک بر زمان واکنش را نتیجه گرفت، اما در جدول ۲ مشاهده می‌شود که مقدار p -value بسیار کمتر از ۰.۰۵ بدست آمده لذا فرض صفر رد شده و می‌توان نتیجه گرفت که موقعیت مکانی محرک بر دقت پاسخ داده شده تاثیرگذار است.

- میدان دیداری :

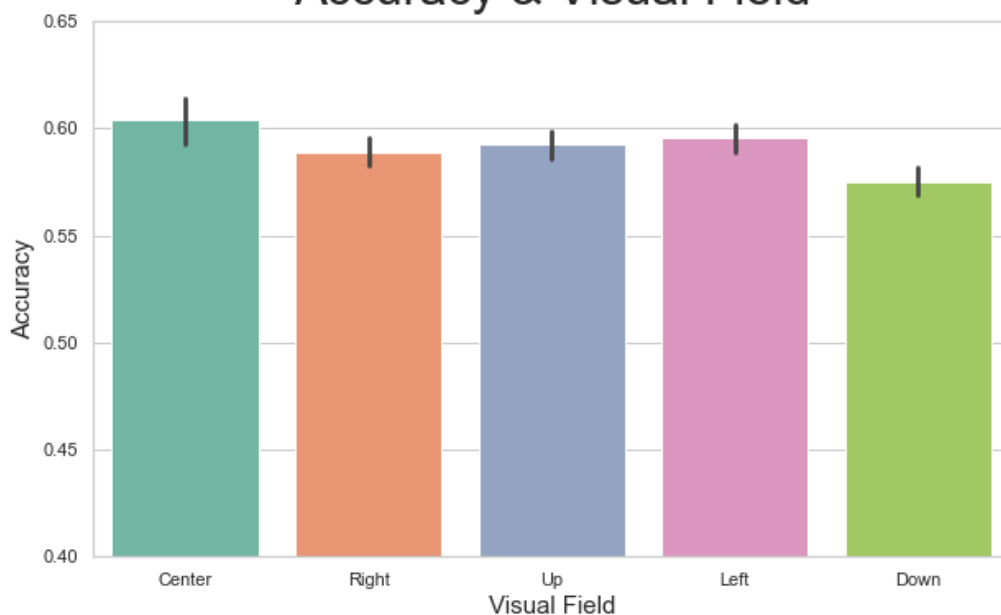
با بررسی نمودار شکل ۳ همان نتایج قبلی تایید می‌شوند زیرا زمانی که محرک در مرکز و سمت چپ قرار داشته کندترین پاسخ و زمانی که در سمت راست قرار داشته سریعترین پاسخ دریافت شده است. همچنین با توجه به نمودار شکل ۲ کندترین واکنش‌ها بیشترین دقت را نتیجه داده‌اند

Reaction Time & Visual Field



شکل ۳ - نمودار زمان واکنش بر اساس میدان دیداری

Accuracy & Visual Field



شکل ۴ - نمودار دقت پاسخ بر اساس میدان دیداری

جدول ۳

	P_value	R_squared	AIC	BIC
rt & Visual Field	0.005929	0.000145	1.372030e+06	1.372078e+06

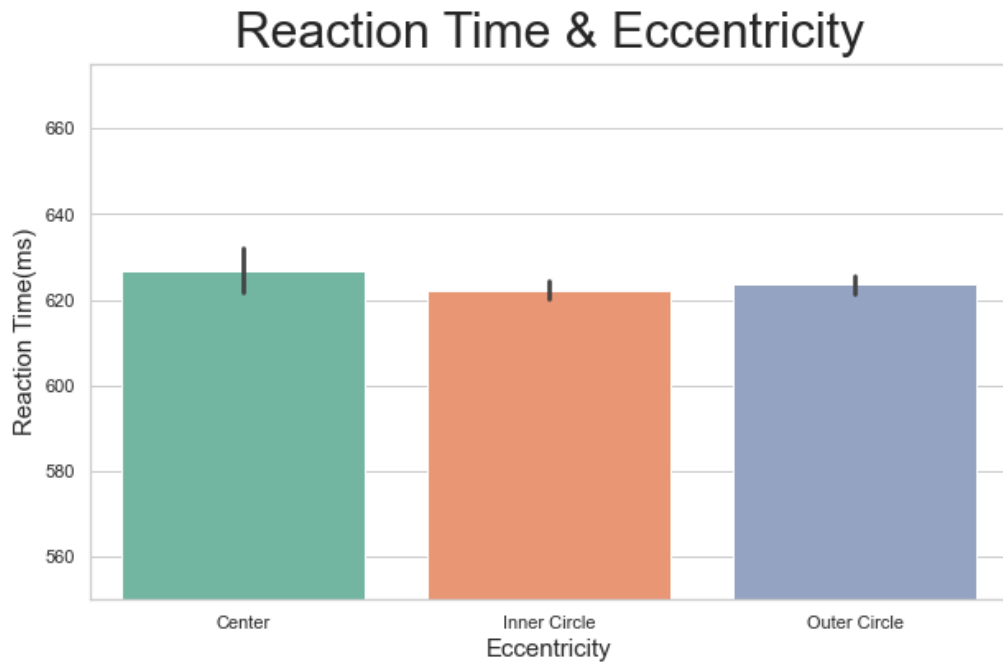
جدول ۴

	P_value	R_squared	AIC	BIC
acc & Visual Field	0.000004	0.000303	141972.740622	142020.308049

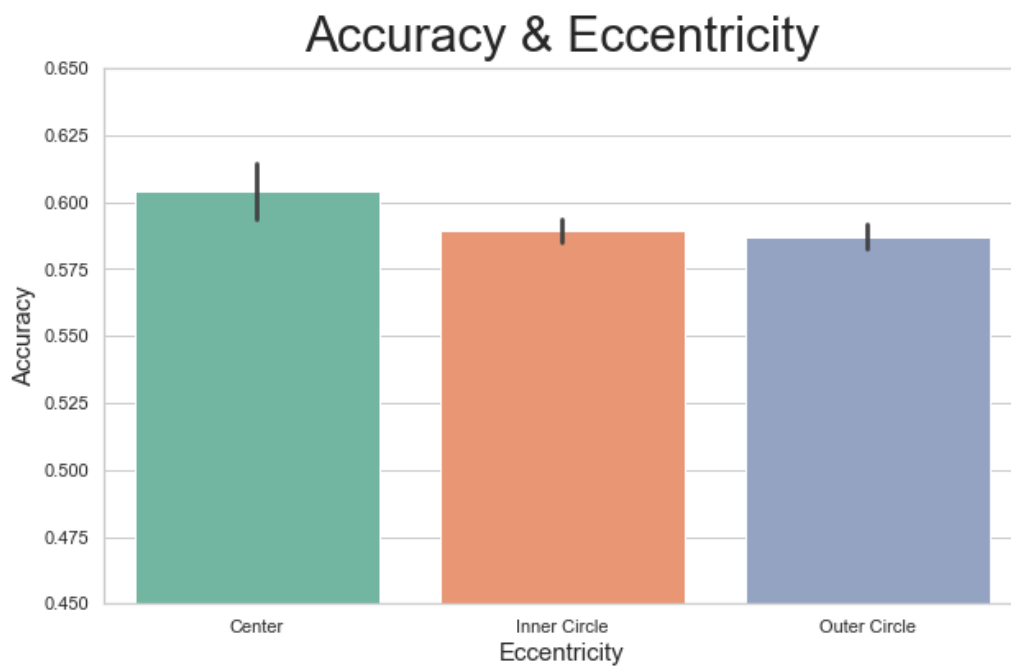
همچنین پس از انجام تست آماری ANOVA مشاهده می شود که در هر دو جدول ۳ و ۴ مقدار p-value کمتر ۰.۰۵ شده است لذا فرض صفر در هر دو مورد رد شده و می توان نتیجه گرفت که متغیر میدان دیداری هم بر زمان واکنش و هم بر دقت پاسخ به طور معناداری تاثیرگذار است.

- گریز از مرکز محرک :

با بررسی نمودار شکل ۵ و ۶ می توان نتیجه گرفت که وقتی محرک در مرکز تصویر قرار می گیر کندترین زمان پاسخ و بیشترین دقت را دارد. این موارد برای شعاع دایره داخلی و بیرونی تفاوت چندانی با هم ندارند.



شکل ۵- نمودار زمان واکنش بر اساس گریز از مرکز محرک



شکل ۶- نمودار دقت پاسخ بر اساس گریز از مرکز محرک

جدول ۵

	P_value	R_squared	AIC	BIC
rt & Eccentricity	0.227933	0.00003	1.372037e+06	1.372066e+06

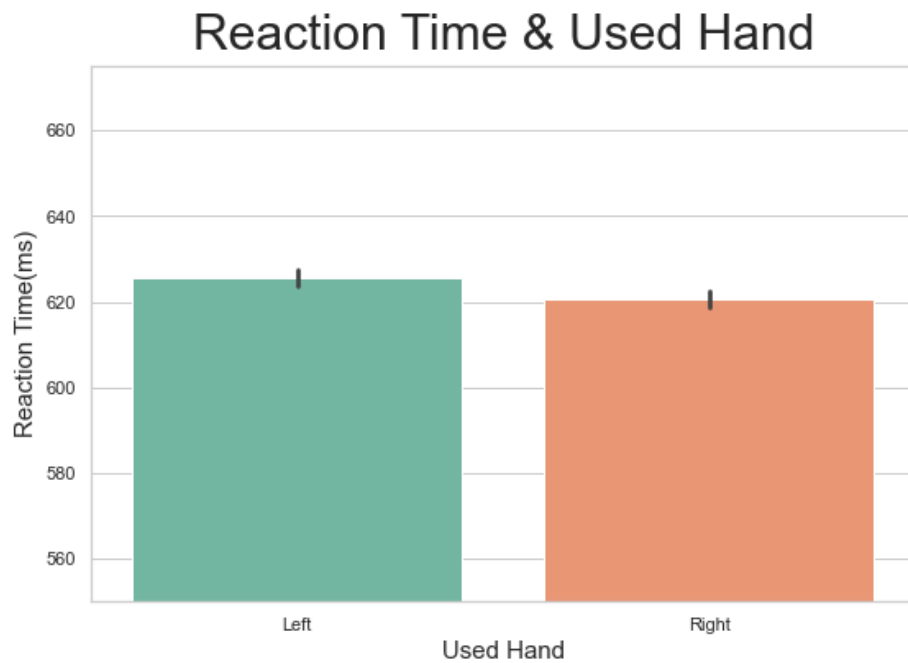
جدول ۶

	P_value	R_squared	AIC	BIC
acc & Eccentricity	0.020782	0.000077	141991.265891	142019.806347

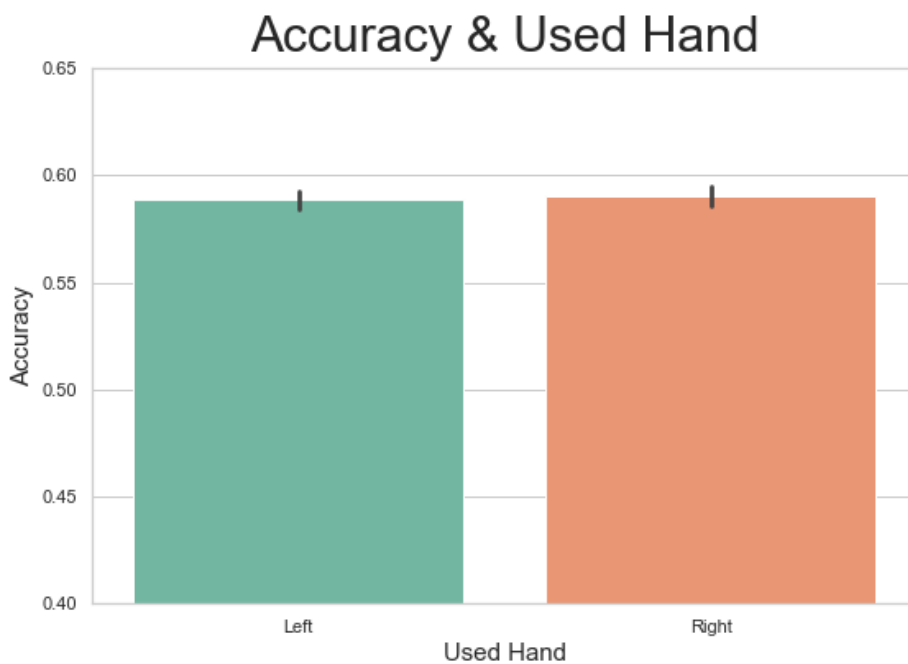
پس از انجام تست ANOVA و چاپ نتایج، در جدول ۵ مشاهده می‌شود که مقدار p-value بیشتر از ۰.۰۵ شده است لذا فرض صفر تایید می‌شود و نمی‌توان تاثیرگذاری گریز از مرکز محرک را بر زمان واکنش نتیجه گرفت. اما در جدول ۶ مشاهده می‌شود که مقدار p-value کمتر از ۰.۰۵ بدست آمده لذا فرض صفر رد شده و می‌توان نتیجه گرفت که گریز از مرکز محرک بر دقت پاسخ داده شده تاثیرگذار است.

- دست استفاده شده برای پاسخ :

در نمودار شکل ۷ مشاهده می‌شود پاسخ هایی که با دست راست داده شده اند اندکی سریعتر از دست چپ بوده است، اما با بررسی نمودار شکل ۸ می‌بینیم که دقت پاسخ این دو مورد تفاوت قابل توجهی با هم ندارد.



شکل ۷- نمودار زمان واکنش بر اساس دست استفاده شده



شکل ۸- نمودار دقت پاسخ بر اساس دست استفاده شده

جدول ۷

	P_value	R_squared	AIC	BIC
rt & Used Hand	0.000448	0.000123	1.372026e+06	1.372045e+06

جدول ۸

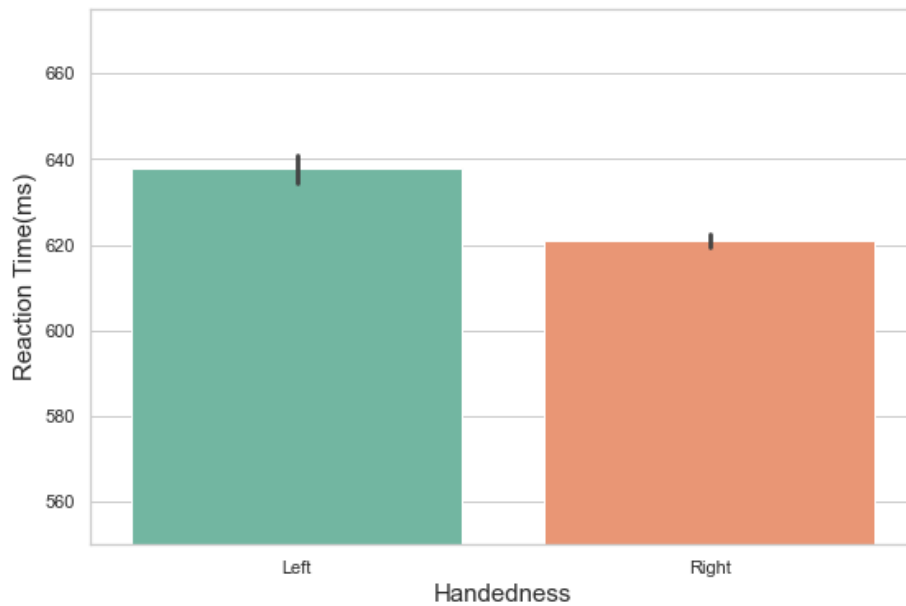
	P_value	R_squared	AIC	BIC
acc & Used Hand	0.549522	0.000004	141996.655309	142015.682279

پس از انجام تست ANOVA با توجه به جدول ۷ مشاهده می‌شود که مقدار p-value کمتر از ۰.۰۵ شده و فرض صفر رد می‌شود لذا دست استفاده شده برای پاسخ بر زمان پاسخ تاثیر گذار است. اما مقدار p-value در جدول ۸ بسیار بیشتر از ۰.۰۵ بدست آمده لذا فرض صفر تایید شده و نمی‌توان نتیجه گرفت که دست استفاده شده بر دقت پاسخ داده شده تاثیر گزار است.

- دست غالب :

در شکل ۹ مشاهده می‌شود که دست راست سریعتر از دس چپ ها بوده اند و علاوه بر آن با توجه به شکل ۱۰ درمی‌یابیم که دست راست ها دقت بیشتری نیز داشته اند. البته این موارد خیلی دقیق نیست زیرا در جامعه آماری موجود اکثر افراد راست دست بوده اند و تعداد افراد چپ دست فقط ۸ نفر در برابر ۵۷ نفر راست دست است.

Reaction Time & Handedness



شکل ۹ - نمودار زمان پاسخ بر اساس دست غالب

Accuracy & Handedness



شکل ۱۰ - نمودار دقت پاسخ بر اساس دست غالب

جدول ۹

	P_value	R_squared	AIC	BIC
rt & Handedness	1.248118e-14	0.000594	1.371979e+06	1.371998e+06

جدول ۱۰

	P_value	R_squared	AIC	BIC
acc & Handedness	0.000138	0.000145	141982.478862	142001.505832

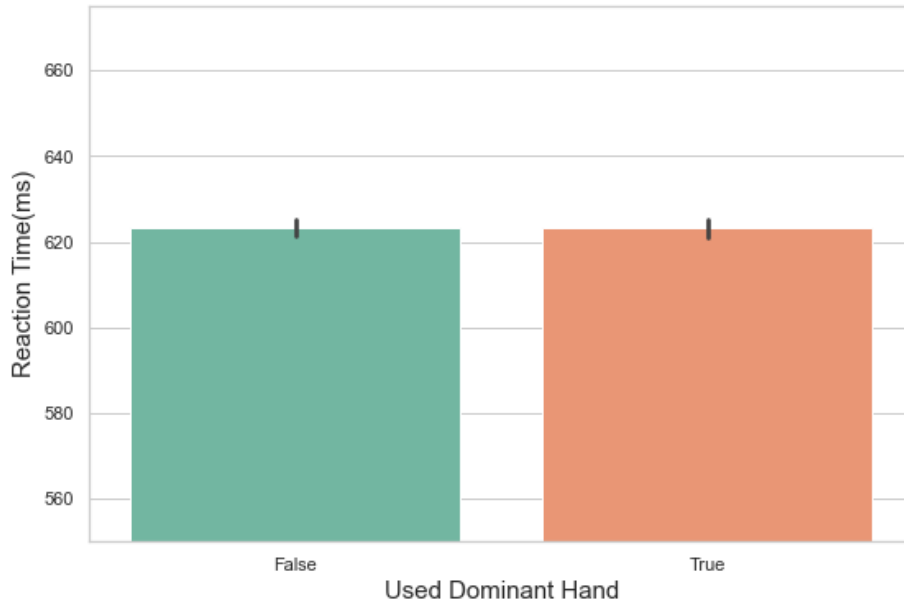
در هر دو جدول ۹ و ۱۰ مشاهده می شود که مقدار p -value کمتر از ۰.۰۵ بدست آمده است، لذا فرض صفر رد شده و می توان نتیجه گرفت که دست غالب هم بر زمان واکنش و هم بر دقت پاسخ تاثیرگذار است.

- استفاده از دست غالب :

در اینجا می‌خواهیم ببینیم زمانی که افراد با دست غالب خود پاسخ می‌دهند تفاوتی در زمان واکنش آنها و دقت پاسخشان ایجاد می‌شود یا خیر؟

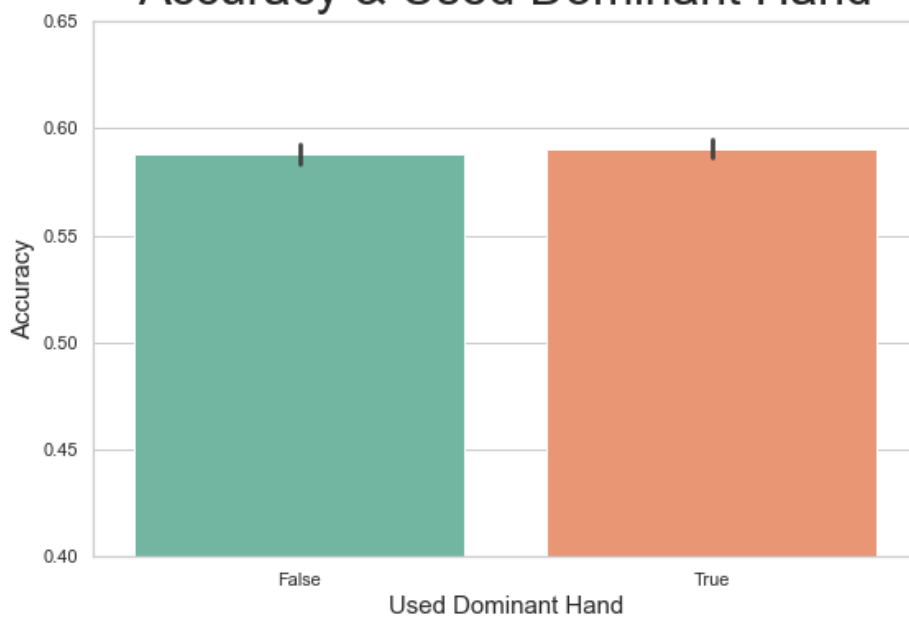
با مشاهده نمودارهای دو شکل ۱۱ و ۱۲ می‌بینیم که هیچ تفاوت معناداری در این دو مورد ایجاد نشده است.

Reaction Time & Used Dominant Hand



شکل ۱۱ - نمودار زمان واکنش بر اساس استفاده از دست غالب

Accuracy & Used Dominant Hand



شکل ۱۲ - نمودار دقت پاسخ بر اساس استفاده از دست غالب

جدول ۱۱

	P_value	R_squared	AIC	BIC
rt & Used Dominant Hand	0.917928	1.061241e-07	1.372038e+06	1.372057e+06

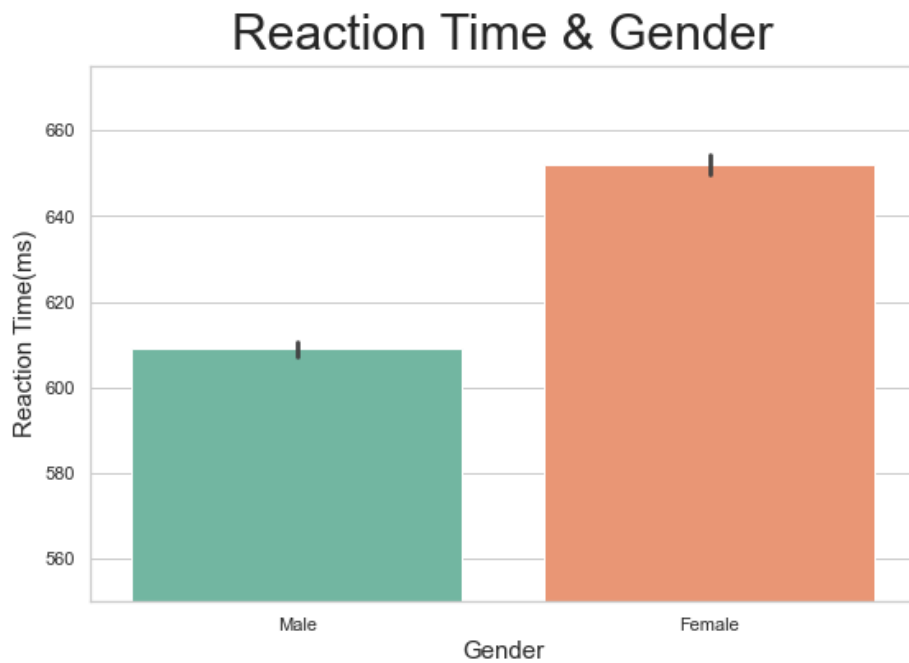
جدول ۱۲

	P_value	R_squared	AIC	BIC
acc & Used Dominant Hand	0.38733	0.000007	141996.266167	142015.293138

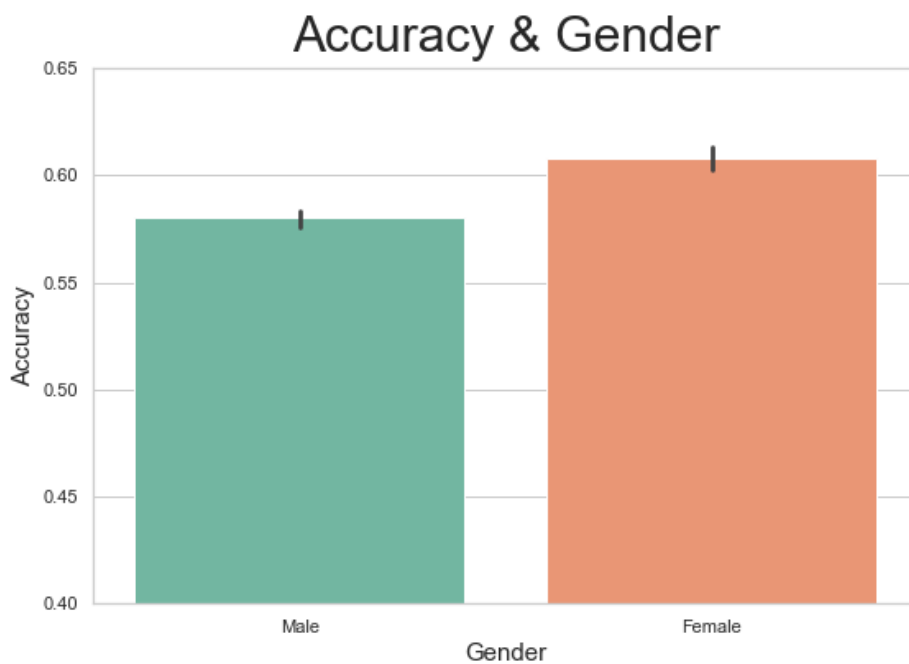
همچنین با انجام تست ANOVA می‌بینیم که نتایج قبلی تایید می‌شوند زیرا مقادیر p-value در دو جدول ۱۱ و ۱۲ بیشتر از ۰.۰۵ شده است و این امر فرض صفر را تایید می‌کند و لذا می‌توان نتیجه گرفت که استفاده از دست غالب تاثیری در زمان واکنش یا دقت پاسخ ندارد.

- جنسیت :

در اینجا تاثیرگذاری جنسیت افراد را بررسی می کنیم. در نمودار شکل ۱۳ مشاهده می شود که مردها سریعتر از خانمها واکنش نشان داده اند، اما در نمودار شکل ۱۴ مشاهده می شود که دقت خانمها بیشتر از مردها بوده است.



شکل ۱۳- نمودار زمان واکنش بر اساس جنسیت



شکل ۱۴- نمودار دقت پاسخ بر اساس جنسیت

جدول ۱۳

	P_value	R_squared	AIC	BIC
rt & Gender	4.107114e-172	0.007787	1.371256e+06	1.371275e+06

جدول ۱۴

	P_value	R_squared	AIC	BIC
acc & Gender	2.605562e-17	0.000716	141925.38911	141944.41608

با انجام تست آماری ANOVA مشاهده می‌شود که جنسیت افراد عاملی تاثیرگذار هم در زمان واکنش و هم در دقت پاسخ است، زیرا همان طور که در جدولهای ۱۳ و ۱۴ قابل مشاهده است مقادیر p-value بسیار بیشتر از ۰.۰۵ بدست آمده است، که این امر فرض صفر را رد خواهد کرد.

- چشم غالب :

در نمودار شکل ۱۵ مشاهده می شود که افراد چپ چشم و راست چشم تفاوت چندانی در زمان واکنش نداشته اند و این مسئله امری مستقل به نظر می رسد. اما در نمودار شکل ۱۶ مشاهده می شود که افراد راست چشم دقت بیشتری داشته اند.



شکل ۱۵ - نمودار زمان واکنش بر اساس چشم غالب



شکل ۱۶ - نمودار دقت پاسخ بر اساس چشم غالب

جدول ۱۵

	P_value	R_squared	AIC	BIC
rt & Eyedness	0.267876	0.000012	1.372037e+06	1.372056e+06

جدول ۱۶

	P_value	R_squared	AIC	BIC
acc & Eyedness	2.156439e-27	0.001174	141879.443135	141898.470105

با انجام تست آماری ANOVA مشاهده می‌شود که با توجه به جدول ۱۵ مقدار P-value بیشتر از ۰.۰۵ شده است لذا فرض صفر تایید می‌شود و نمی‌توان نتیجه گرفت که چشم غالب بر سرعت واکنش افراد تاثیرگذار است. اما همانطور که در جدول ۱۶ قابل مشاهده است، در اینجا مقدار p-value بسیار کمتر از ۰.۰۵ شده و فرض صفر رد می‌شود لذا می‌توان نتیجه گرفت که چشم غالب امری تاثیرگذار بر دقت پاسخ افراد می‌باشد.

۳. پیش‌بینی ویژگی‌های رفتاری

• زمان واکنش (reaction time)

ابتدا یک مدل رگرسیون خطی با ۳۵ متغیر زیر روی داده‌ها فیت می‌کنیم:

```
C(pos) + uhnd + hndns + eye + sex + C(ecc) + C(vf) + C(udh) +  
C(pos)*uhnd + C(pos)*hndns + C(pos)*eye + C(pos)*sex + C(pos)*C(ecc) +  
C(pos)*C(vf) + C(pos)*C(udh) + uhnd*hndns + uhnd*eye + uhnd*sex +  
uhnd*C(ecc) + uhnd*C(vf) + uhnd*C(udh) + hndns*eye + hndns*sex +  
hndns*C(ecc) + hndns*C(vf) + hndns*C(udh) + eye*sex + eye*C(ecc) +  
eye*C(vf) + eye*C(udh) + sex*C(ecc) + sex*C(vf) + sex*C(udh) +  
C(ecc)*C(vf) + C(ecc)*C(udh) + C(vf)*C(udh)
```

سپس ویژگی که بیشترین p-value را دارد حذف کرده و دوباره این حلقه را تکرار می‌کنیم.
در نهایت پس از ۳۳ بار تکرار به ۴ متغیر زیر می‌رسیم که p-value همه آنها کمتر از ۰.۰۵ است:

```
hndns + eye + sex + eye*sex
```

خروجی نهایی مدل رگرسیون در جدول ۱۷ قابل مشاهده است. (مراحل طی شده برای حذف ویژگی نیز در فایل ipynb قابل مشاهده است)

جدول ۱۷

OLS Regression Results

Dep. Variable:	rt	R-squared:	0.011			
Model:	OLS	Adj. R-squared:	0.011			
Method:	Least Squares	F-statistic:	307.7			
Date:	Wed, 30 Mar 2022	Prob (F-statistic):	8.85e-264			
Time:	21:41:11	Log-Likelihood:	-7.7031e+05			
No. Observations:	112419	AIC:	1.541e+06			
Df Residuals:	112414	BIC:	1.541e+06			
Df Model:	4					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

Intercept	681.3829	2.846	239.446	0.000	675.805	686.960
hndns[T.r]	-38.1270	2.189	-17.416	0.000	-42.418	-33.836
eye[T.r]	16.2530	2.404	6.761	0.000	11.541	20.965
sex[T.m]	-44.3948	2.376	-18.685	0.000	-49.052	-39.738
eye[T.r]:sex[T.m]	-10.9383	2.990	-3.658	0.000	-16.799	-5.078

• دقت پاسخ (accuracy)

ابتدا یک مدل رگرسیون لاجستیک با ۳۵ متغیر زیر روی داده ها فیت می کنیم:

$$\begin{aligned} &C(pos) + uhnd + hndns + eye + sex + C(ecc) + C(vf) + C(udh) + \\ &C(pos)*uhnd + C(pos)*hndns + C(pos)*eye + C(pos)*sex + C(pos)*C(ecc) + \\ &C(pos)*C(vf) + C(pos)*C(udh) + uhnd*hndns + uhnd*eye + uhnd*sex + \\ &uhnd*C(ecc) + uhnd*C(vf) + uhnd*C(udh) + hndns*eye + hndns*sex + \\ &hndns*C(ecc) + hndns*C(vf) + hndns*C(udh) + eye*sex + eye*C(ecc) + \\ &eye*C(vf) + eye*C(udh) + sex*C(ecc) + sex*C(vf) + sex*C(udh) + \\ &C(ecc)*C(vf) + C(ecc)*C(udh) + C(vf)*C(udh) \end{aligned}$$

سپس ویژگی که بیشترین p-value را دارد حذف کرده و دوباره این حلقه را تکرار می کنیم.
در نهایت پس از ۳۴ بار تکرار به ۳ متغیر زیر میرسیم که p-value همه آنها کمتر از ۰.۰۵ است:

$$eye + sex + eye*sex$$

خروجی نهایی مدل رگرسیون در جدول ۱۸ قابل مشاهده است. (مراحل طی شده برای حذف ویژگی در
نیز فایل ipynb قابل مشاهده است)

جدول ۱۸

Logit Regression Results						
=====						
Dep. Variable:	acc		No. Observations:	100056		
Model:	Logit		Df Residuals:	100052		
Method:	MLE		Df Model:	3		
Date:	Wed, 30 Mar 2022		Pseudo R-squ.:	0.002341		
Time:	22:48:17		Log-Likelihood:	-67589.		
converged:	True		LL-Null:	-67748.		
Covariance Type:	nonrobust		LLR p-value:	1.922e-68		
=====						
	coef	std err	z	P> z	[0.025	0.975]

Intercept	0.2424	0.017	14.315	0.000	0.209	0.276
eye[T.r]	0.3501	0.023	15.380	0.000	0.305	0.395
sex[T.m]	0.0511	0.021	2.407	0.016	0.009	0.093
eye[T.r]:sex[T.m]	-0.3026	0.028	-10.840	0.000	-0.357	-0.248