



# University of Tehran

School of Electrical and Computer Engineering



## مقدمه‌ای بر علوم شناختی

تمرین دوم

ساسان کشاورز

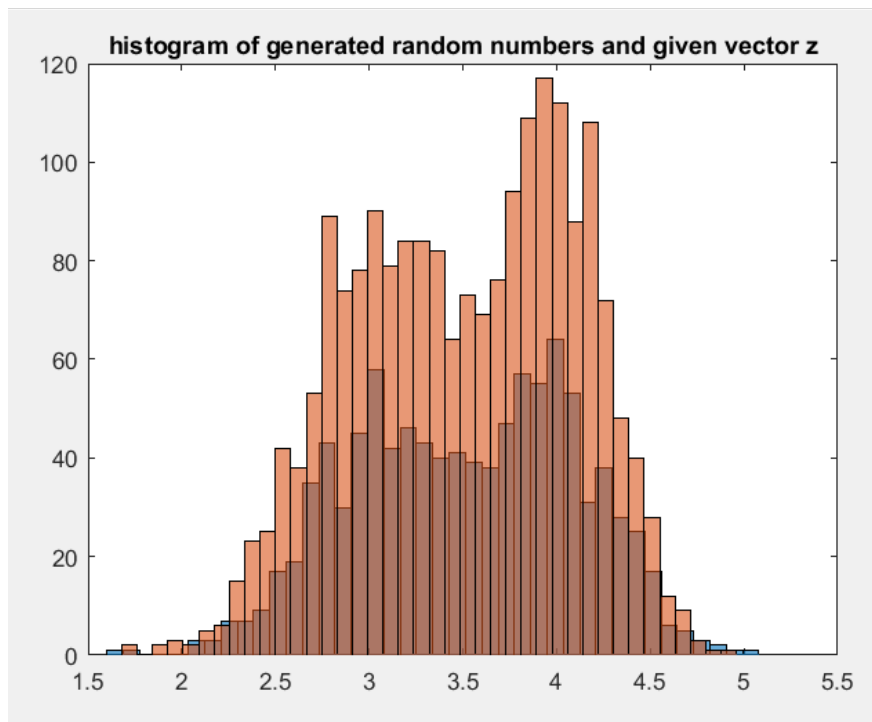
۸۱۰۱۹۹۲۵۳

تابستان ۱۴۰۰

## مسئله اول: تولید اعداد تصادفی

### ۱. تولید اعداد تصادفی با توزیع مشابه بردار داده Z

ابتدا توزیع بردار داده Z را تخمین میزنیم. این عمل با تابع `fitdist` در متلب انجام میشود. باید توجه کرد این تابع ورودی بردار ستونی میگیرد و باید ترانهاده بردار Z را به آن بدهیم. برای تعیین اینکه چه توزیعی به داده‌ها برازش شوند ابتدا در `distribution fitting tool` توزیع‌های مختلف را امتحان کردم و دیدم که بهترین انتخاب `non parametric` است در نتیجه توزیع `kernel` را برای برازش توزیع بردار Z انتخاب کردم. پس از آن با دستور `random` ۱۰۰۰ عدد رندوم مطابق توزیع استخراج شده تولید میکنیم. هیستوگرام توزیع Z و اعداد تولید شده در شکل ۱ مشاهده میشود.



شکل ۱ هیستوگرام توزیع Z و توزیع اعداد رندوم تولید شده

### ۲. تولید پاسخ های نرونی

از توزیع پواسون برای تولید توزیع اسپایک نرون در بازه زمانی میتوان استفاده کرد. ضریب لاندای توزیع پواسون همان نرخ اسپایک زدن هر نرون است. ابتدا برای این سه نرون پاسخ نرونی لازم را تولید میکنیم. سپس آن را طوری تغییر میدهیم که ماتریس همبستگی پاسخ‌ها به صورت جدول داده شده باشد.

در این مرحله ابتدا با دستور `sqrtm` ماتریس همبستگی داده شده را به جذر ماتریسش تبدیل میکنیم و این ماتریس را در ماتریس متشکل از پاسخهای نرونی تولید شده ضرب میکنیم. نتیجه یک ماتریس خروجی پاسخهای نرونی است که همبستگی بین پاسخهایش بسیار به ماتریس داده شده در سوال شبیه است.

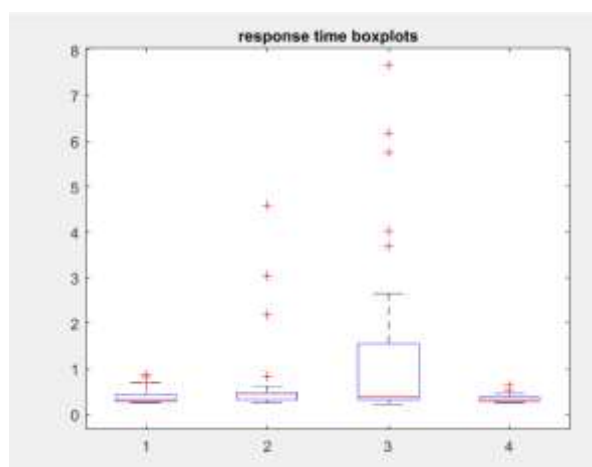
Variables - ccfinal				
	cc	ccfinal		
	3x3 double			
	1	2	3	4
1		0.0914	-0.0666	
2	0.0914	1	0.7210	
3	-0.0666	0.7210	1	
4				
5				
6				

شکل ۲ همبستگی پاسخهای نرونی تولید شده

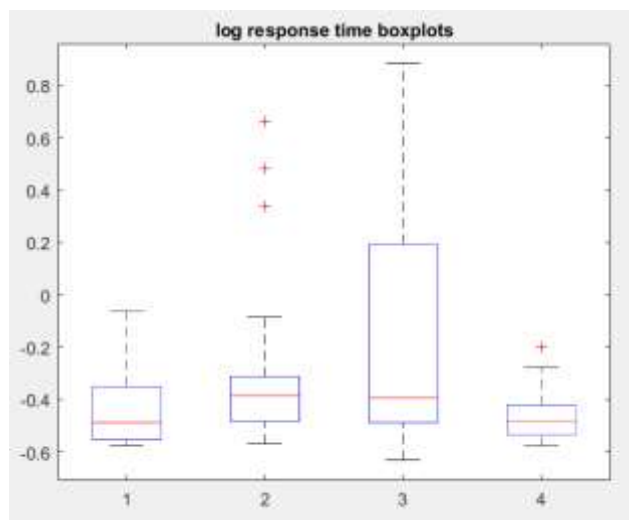
## مسئله دوم: آزمون آماری

### ۱. نمایش زمان پاسخ دهی و لگاریتم آن با نمودار جعبه‌ای

بسیار ساده چهار دسته کلمات مختلف را جدا میکنیم و با دستور `boxplot` نمودار جعبه ای داده ها را رسم میکنیم.



شکل ۳ نمودار جعبه ای مربوط به زمان پاسخ دهی فرد به ۴ دسته کلمات



شکل ۴ نمودار جعبه ای لگاریتم زمان پاسخ دهی فرد به ۴ دسته کلمات

## II. تست anova رو یزمان های پاسخ دهی

تست anova برای مقایسه و تحلیل میان میانگین ها گروه های مختلف داده ها به کار میرود. با انجام تحلیل anova در متلب نتایج به صورت زیر خواهند بود.

Source	SS	df	MS	F	Prob>F
Columns	23.212	3	7.73725	6.23	0.0006
Error	143.987	116	1.24127		
Total	167.199	119			

شکل ۵ جدول نتایج تحلیل anova بر زمانهای پاسخ

با توجه به اینکه مقدار  $0.0006$  خیلی کوچک است، پس حتما بین حداقل دو گروه از ۴ گروه کلمات اختلاف چشمگیری وجود داشته است. از تحلیل anova نمیتوان برای مشخص کردن این ۲ گروه خاص استفاده کرد.

### III. تست anova روی لگاریتم زمان های پاسخ دهی

با انجام تحلیل anova روی لگاریتم داده ها تایید به صورت زیر خواهند بود. عملیات لگاریتم باعث میشود که بهتر بتوانیم بین گروه های داده مقایسه انجام بدهیم.

Source	SS	df	MS	F	Prob>F
Columns	1.9066	3	0.63554	7.28	0.0002
Error	10.1307	116	0.08733		
Total	12.0373	119			

شکل ۶ جدول نتایج تحلیل anova بر لگاریتم زمانهای پاسخ

همانطور که از این جدول هم مشخص است. مقدار  $0.0002$  بسیار کوچک است و فرض  $H_0$  رد میشود. نشانگر اختلاف معنادار میان میانگینهای حداقل دو گروه از چهار گروه لگاریتم زمانهای پاسخ دهی است.

### IV. مقایسه میانگین زمان پاسخ دهی برای هر دو جفت گروه

برای آنکه بفهمیم بین کدام دو گروه از کلمات اختلاف معنادار وجود داشته است میتوانیم از آزمون  $t$ -test برای هر جفت از گروه ها بهره ببریم. آزمون  $t$  یک نوع از آمارهای استنباطی است که برای تعیین اینکه آیا بین میانگین دو گروه اختلاف معنی داری وجود دارد یا نه که ممکن است در ویژگیهای خاصی مرتبط باشد وجود دارد، بیشتر مورد استفاده قرار می گیرد. چون حدس میزنیم اختلاف معنی دار بین گروه کلمات مرتبط و نامرتبط است ابتدا آزمون را برای همین دو دسته انجام میدهیم. نتایج انجام  $t$ -test روی داده ها به صورت زیر است:

```
h_ttest_two_tail =  
0  
p_ttest_two_tail =  
0.0002  
h_ttest_one_tail =  
3  
p_ttest_one_tail =  
0.0001
```

شکل ۷ نتایج انجام  $t$ -test بین گروه کلمات مرتبط و غیر مرتبط

فرض  $H_0$  در حالت دوطرفه رد نمیشود. اما در حالت یک طرفه فرض  $H_0$  رد میشود و میتوان گفت priming رخ داده است.

همین کار را بر روی لگاریتم پاسخ های زمانی کلمات مرتبط و غیرمرتبط هم اعمال میکنیم. نتایج به شکل زیر خواهند بود.

```
Current Model  
  
b_two_tail_log_ttest =  
0  
  
p_two_tail_log_ttest =  
0.0595  
  
b_one_tail_log_ttest =  
1  
  
p_one_tail_log_ttest =  
0.0297
```

شکل ۸ نتایج انجام t-test بر روی لگاریتم پاسخ های زمانی کلمات مرتبط و غیر مرتبط

در این حالت مشابه حالت قبل در حالت دوطرفه  $H_0$  رد نمیشود و در حالت یک طرفه  $H_0$  رد میشود. پس priming رخ داده.

## ۷. شرایط اعمال تست anova

برای آنکه بتوان از تست anova استفاده کرد باید شروط مشخصی برقرار باشند. داده های هر گروه یا جامعه باید دارای توزیع نرمال باشند. واریانس در هر گروه ثابت باشد. این امر نشان می دهد که نباید داده ها شامل نقاط دورافتاده باشند. واریانس گروه ها با یکدیگر برابر باشند. میانگین در بین گروه ها متفاوت باشد. در حقیقت این همان عبارتی است که به عنوان فرض مقابل در تحلیل واریانس anova به دنبالش هستیم. واریانس های گروه ها برابر ولی نامشخص هستند. همچنین الزامی است که گروه ها از یکدیگر مستقل هستند.

با بررسی این شروط نتیجه میگیریم که شرایط انجام تست روی این دسته داده مهیا است و تست anova معتبر است.

## مسئله سوم: تست با رگرسیون

در این قسمت، با استفاده از یک مدل تعمیم یافته خطی، آزمون آماری بر روی داده‌ها کلمات مرتبط و غیرمرتبط انجام می‌دهیم.

```
Linear regression model:
```

```
y ~ 1 + x1
```

```
Estimated Coefficients:
```

	<b>Estimate</b>	<b>SE</b>	<b>tStat</b>	<b>pValue</b>
	<hr/>	<hr/>	<hr/>	<hr/>
<b>(Intercept)</b>	0.39576	0.12327	3.2105	0.0021615
<b>x1</b>	0.30096	0.17433	1.7264	0.089604

```
Number of observations: 60, Error degrees of freedom: 58
```

```
Root Mean Squared Error: 0.675
```

```
R-squared: 0.0489, Adjusted R-Squared: 0.0325
```

```
F-statistic vs. constant model: 2.98, p-value = 0.0896
```

فر \~

شکل ۱- مدل آماری و آزمون آماری

با توجه به مقدار p-value فرش  $H_0$  رد میشود و اختلاف معنی دار وجود دارد.