به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

آمار و احتمال مهندسی

گزارش پروژه‌ی نهائی

نام و نام خانوادگی	آرین باستانی
شماره دانشجویی	۸۱۰۱۰۰۰۸۸
تاریخ ارسال گزارش	۱۴۰۲/۴/۱۶

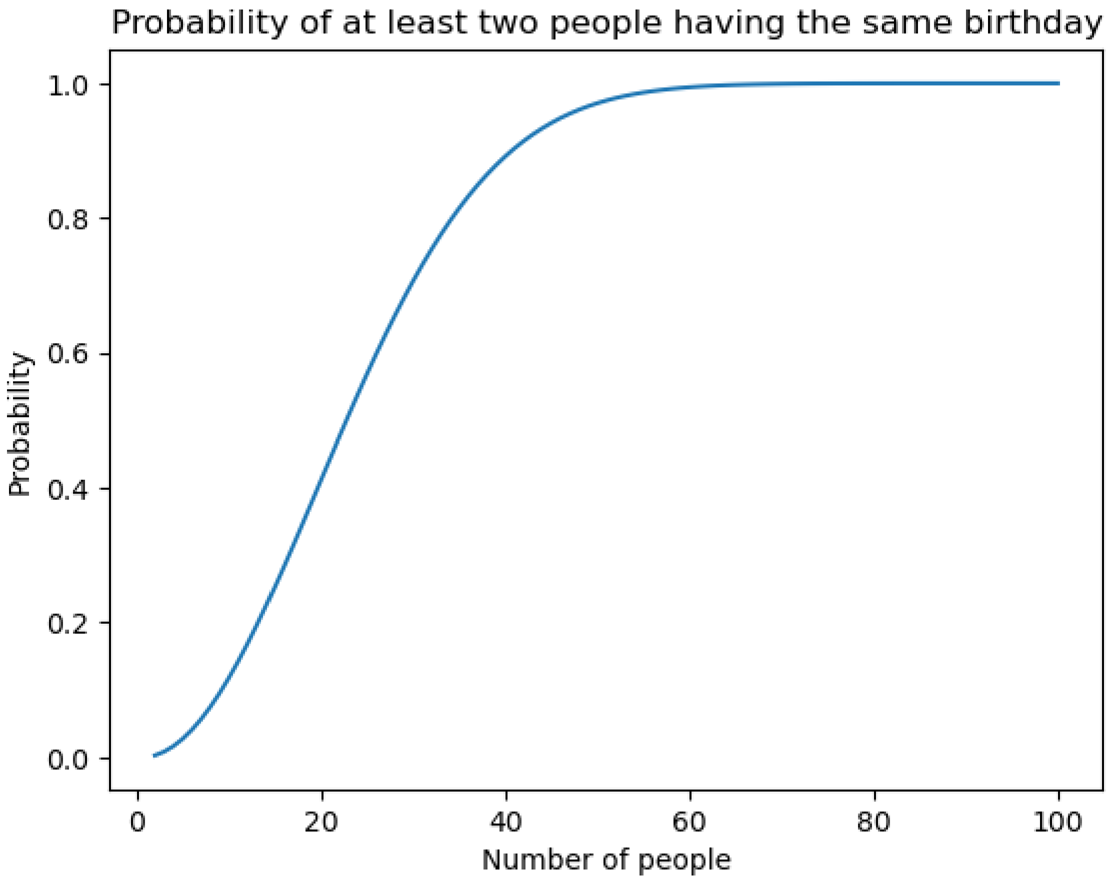
فهرست گزارش سوالات

سوال اول: تناقض تاریخ تولد

۱.۱_

$$P_n = 1 - \frac{(365)!}{(365)^n.(365-n)!}$$

نمودار احتمال برای n های کوچکتر از ۱۰۱:



در ابتدا مطابق با گفته ی سوال به طور شهودی همچین انتظاری را نداشتیم چراکه تصور میشد تا قبل از n برابر با ۱۸۳ احتمال انقدر زیاد رشد نمیکرد! ولی رابطه ی ریاضی متمم این احتمال با توجه به کسر آن به سرعت با زیاد شدن n کاهش میابد و بنابراین P_n افزایش خواهد یافت.

حداقل مقدار n برای اینکه احتمال برابر ۰.۵ شود:

$$1-\frac{(365)!}{(365)^n.(365-n)!}\geq 0.5$$

$$\Longrightarrow\ (365)!\leq 0.5(365)^n.(365-n)!$$
$$\Longrightarrow\ n\geq 23$$

پس باید n برابر با ۲۳ باشد.

به طور مشابه برای اینکه احتمال برابر ۰.۹۹ شود:

$$1-\frac{(365)!}{(365)^n.(365-n)!}\geq 0.99$$

$$\Longrightarrow\ (365)!\leq 0.01(365)^n.(365-n)!$$
$$\Longrightarrow\ n\geq 57$$

۱.۲_

$$P_n=1-\frac{(365)!}{(365)^n.(365-n)!}=\ 1-\left(\frac{365}{365}\times\frac{364}{365}\times ....\ \times\frac{365-n}{365}\right)$$

$$\Longrightarrow P_n=1-1\ \times\left(1-\frac{1}{365}\right)\times\left(1-\frac{2}{365}\right)\times ....\ \times\left(1-\ \frac{n}{365}\right)$$

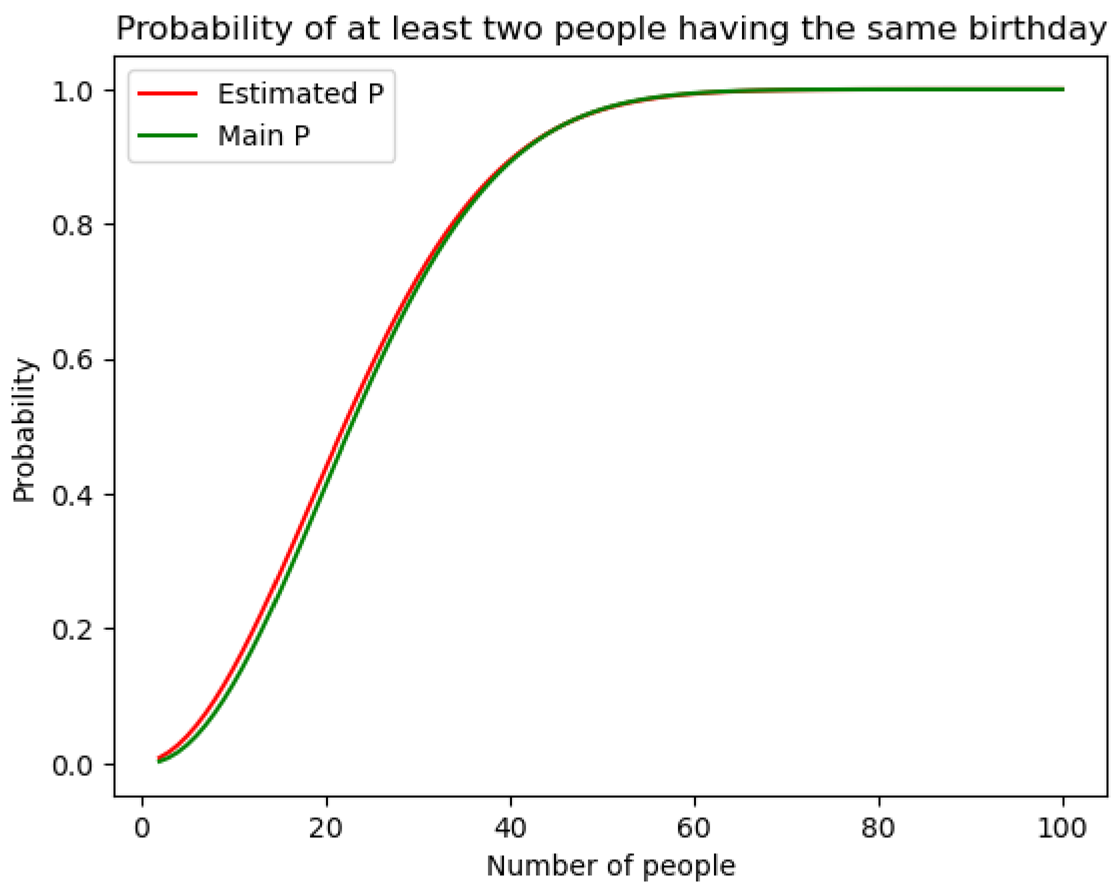
$$e^x\simeq 1+\frac{x}{1!}\qquad\Longrightarrow\quad 1-\frac{i}{365}\ \simeq\ e^{-\frac{i}{365}}$$

$$\Longrightarrow P_n\simeq 1-1\ \times\ e^{-\frac{1}{365}}\ \times\ e^{-\frac{2}{365}}\ \times ....\ \times\ e^{-\frac{n}{365}}$$

$$= 1 - e^{\frac{-(1+2+\cdots+n)}{365}} = 1 - e^{\frac{-n(n+1)}{730}}$$

$$\Longrightarrow P_n \simeq 1 - e^{\frac{-n(n+1)}{730}}$$

نمودار احتمال اصلی و احتمال تخمین زده شده:



بنابراین این تخمین بسیار خوبی برای احتمال است چراکه احتمال تخمین زده شده بسیار به احتمال اصلی نزدیک است و با افزایش n این تقریب بهتر میشود.

سوال دوم: سکه‌بازی

_۲.۱

آ) ابتدا $P_1(n_1, n_2)$ را احتمال بردن فرد اول با n_1 دلار در دست فرد اول و n_2 دلار در دست فرد دوم تعریف میکنیم و به همین صورت $P_2(n_1, n_2)$ را برای بردن فرد دوم در نظر میگیریم.

حالا برای پرتاب اول بازی دو حالت داریم:

a) شیر بیاید و نفر اول یک دلار سود کند. بنابراین باید با احتمال جدید $P_1(n_1 + 1, n_2 - 1)$ ببرد. و باتوجه به سالم بودن سکه در کل میشود : $0.5 \times P_1(n_1 + 1, n_2 - 1)$

b) خط بیاید و نفر دوم یک دلار سود کند. بنابراین همانند حالت قبل اینبار احتمال میشود:$0.5 \times P_1(n_1 - 1, n_2 + 1)$

پس در کل احتمال برد نفر اول میشود:

$$P_1(n_1,n_2) = 0.5 \times P_1(n_1+1,n_2-1) + 0.5 \times P_1(n_1-1,n_2+1)$$

و با توجه به اینکه با تمام شدن پول هر فرد ، فرد دیگر برنده میشود:

$$P_1(any\ natural\ number,0) = 1\ ,\ P_1(0,any\ natural\ number) = 0$$

و به همین شکل برای احتمال برد نفر دوم داریم:

$$P_2(n_1,n_2) = 0.5 \times P_2(n_1-1,n_2+1) + 0.5 \times P_2(n_1+1,n_2-1)$$

$$P_2(0,any\ natural\ number) = 1\ ,\ P_2(any\ natural\ number,0) = 0$$

ب) در این حالت فقط کافیست به جای ۰.۵ احتمال p و $1 - p$ را جایگذاری کنیم:

$$P_1(n_1,n_2) = p \times P_1(n_1+1,n_2-1) + (1-p) \times P_1(n_1-1,n_2+1)$$

$$P_1(any\ natural\ number,0) = 1\ ,\ P_1(0,any\ natural\ number) = 0$$

$$P_2(n_1,n_2) = (1-p) \times P_2(n_1-1,n_2+1) + p \times P_2(n_1+1,n_2-1)$$

$$P_2(0,any\ natural\ number) = 1\ ,\ P_2(any\ natural\ number,0) = 0$$

_۲.۲

با توجه به روابط بازگشتی قسمت قبل ، هرچه پول در دست بازیکنان (یعنی n_1 یا n_2) کمتر باشد، زودتر این روابط بازگشتی به شرایط اولیه میرسند. یعنی اگر n_1 کمتر از n_2 باشد ، آنگاه احتمال آنکه پول نفر اول زودتر تموم شود بیشتر میشود و یعنی در احتمال بردن نفر اول به شرایط اولیه ی $P_1(0, any\ natural\ number) = 0$ میرسیم و برای احتمال برد نفر دوم به شرایط اولیه ی $P_2(0, any\ natural\ number) = 1$ میرسیم.

بنابراین با کوچکتر بودن n_1 احتمال برد نفر دوم افزایش میابد و به شکل مشابه ، با کوچکتر بودن n_2 احتمال بردن نفر اول افزایش میابد.

و با توجه به اینکه p برابر با احتمال شیر آمدن است ، پس با افزایش مقدار p احتمال برد نفر اول افزایش میابد و با کاهش آن احتمال برد نفر دوم افزایش میابد.

پس به صورت کلی تاثیر این پارامتر ها بدین شکل است:

افزایش n_1 ، کاهش n_2 ، افزایش p (کاهش (۱ - p)) ⟵⟵ افزایش P_1 و کاهش P_2

کاهش n_1 ، افزایش n_2 ، کاهش p (افزایش (۱ - p)) ⟵⟵ کاهش P_1 و کاهش P_2

_۲.۳

برای ۱۰۰ دور اجرا کردن بازی میانگین را چند بار محاسبه کردم:

9.26

10.53

11.45

9.76

سوال سوم: محاسبه مساحت

_۳.۱

$$E[S_n] = \frac{1}{n}E[X_1 + X_2 + \cdots + X_n] = \frac{1}{n}(E[X_1] + E[X_2] + \cdots + E[X_n])$$

متغیرهای تصادفی X_i هم توزیع و مستقل و با توزیع برنولی هستند.

و باتوجه به اینکه مساحت مربع ۱ میباشد :

$$P_X(x) = \frac{A_s}{1} = A_s$$

که A_s مساحت زیر مجموعه میباشد.

با توجه به رابطه های توزیع برنولی :

$$E[X_i] = P_X(x) = A_s$$

$$\Rightarrow E[S_n] = \frac{1}{n} \times n \times A_s = A_s$$

حالا اگر واریانس هر X_i را σ^2 در نظر بگیریم:

$$Var(X_1 + X_2 + \cdots + X_n) = n \times \sigma^2 \Rightarrow Var(S_n) = \frac{1}{n^2} \times n \times \sigma^2 = \frac{\sigma^2}{n}$$

$$\Rightarrow \lim_{n\to\infty} \frac{\sigma^2}{n} = \frac{\sigma^2}{\infty} = 0$$

_۳.۲

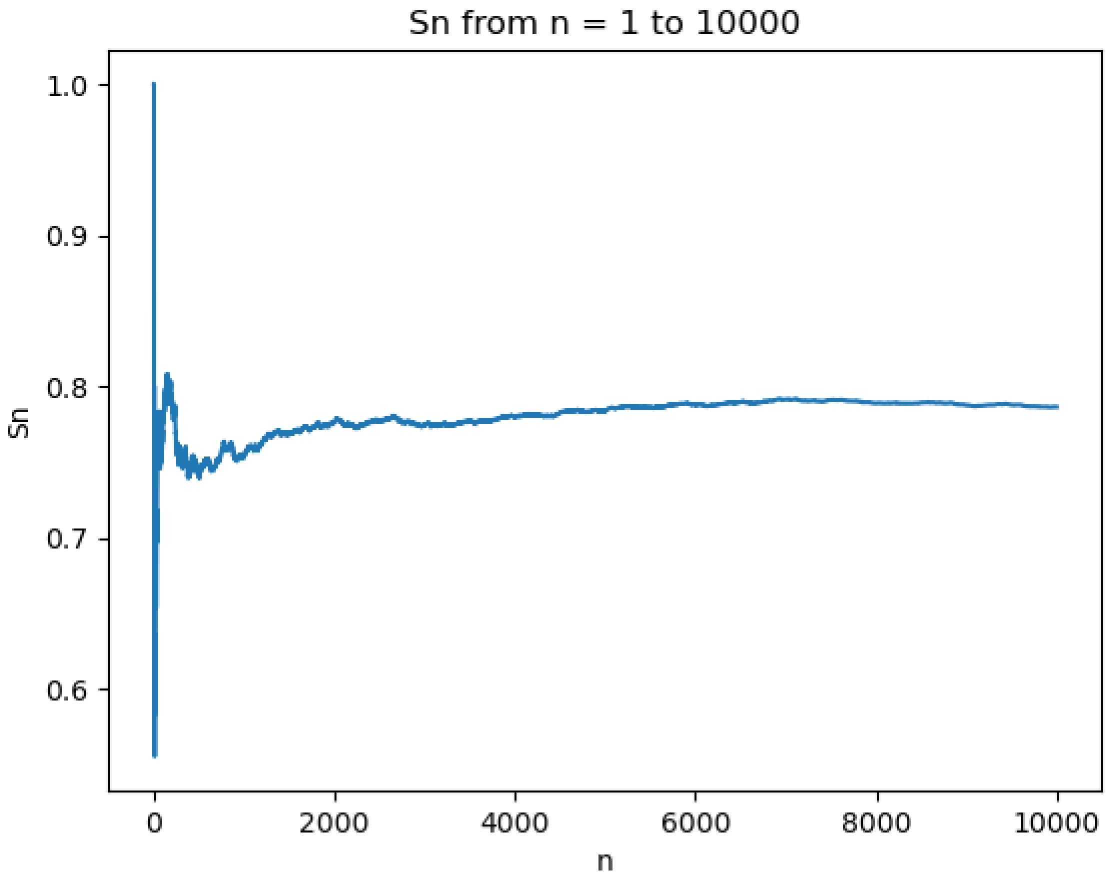
$$S_{n-1} = \frac{X_1 + X_2 + \cdots + X_{n-1}}{n-1} \Rightarrow (n-1)S_{n-1} = X_1 + X_2 + \cdots + X_{n-1}$$

$$and\ \ S_n = \frac{(X_1 + X_2 + \cdots + X_{n-1}) + X_n}{n} \Rightarrow S_n = \frac{(n-1)S_{n-1} + X_n}{n}$$

_۳.۳

از رابطه ی زیر برای دایره استفاده میکنیم:

$$(x-0.5)^2+(y-0.5)^2 \leq 0.25$$



از آنجایی که $A=\pi r^2$ بنابراین با محاسبه ی مساحت این دایره و تقسیم کردن آن به مجذورش

میتوان عدد پی را محاسبه کرد: $\pi = \frac{A}{r^2}$

و با توجه به بخش های قبل برای بدست آوردن مساحت دایره کافیست امید S_n را حساب کنیم.

$$\Rightarrow \pi = \frac{r^2}{E[S_n]} = \frac{1}{4 \times E[S_n]}$$

_۳.۴

برای محاسبه مساحت ، کافیست از S_n امید ریاضی بگیریم.

این نتایج مختلف برای چند بار میانگین گرفتن است(که باید طبق انتظار ما به هم نزدیک باشند):

0.37942214761490545

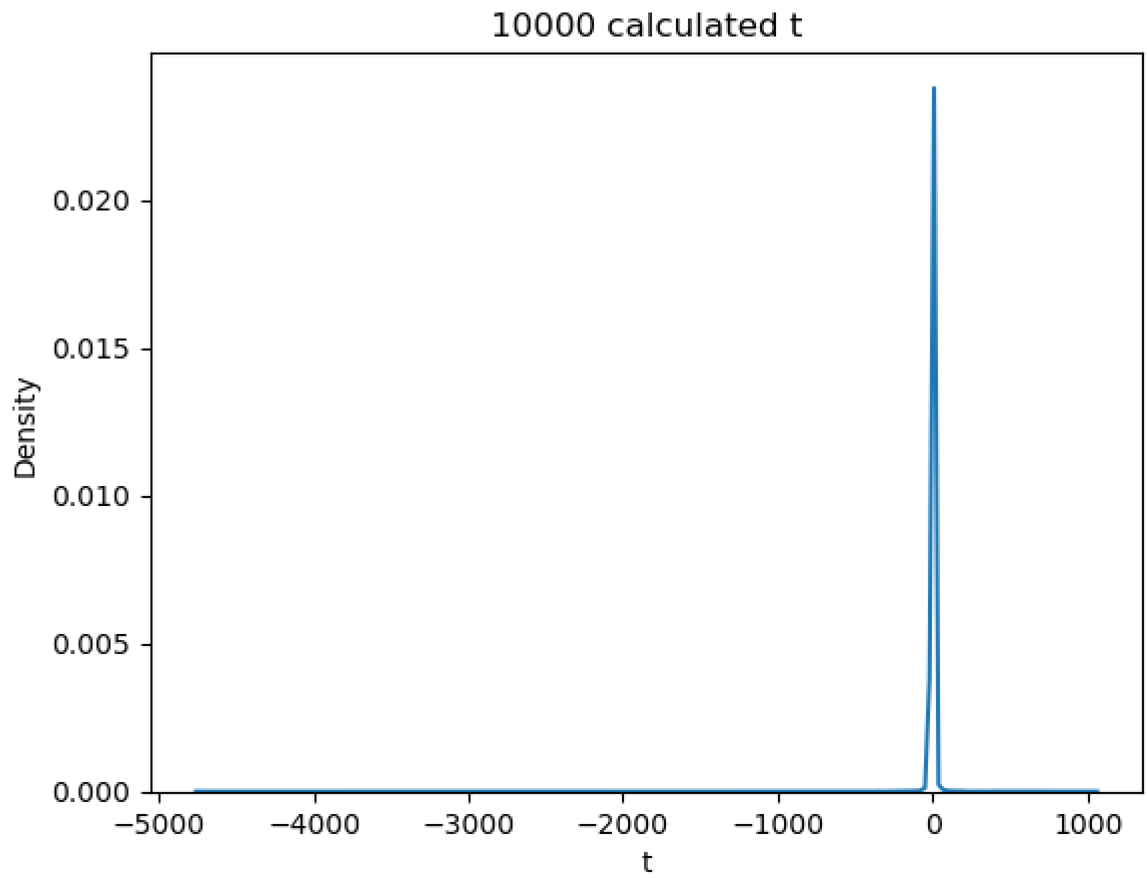
0.38243326646710807

0.37995772297800272

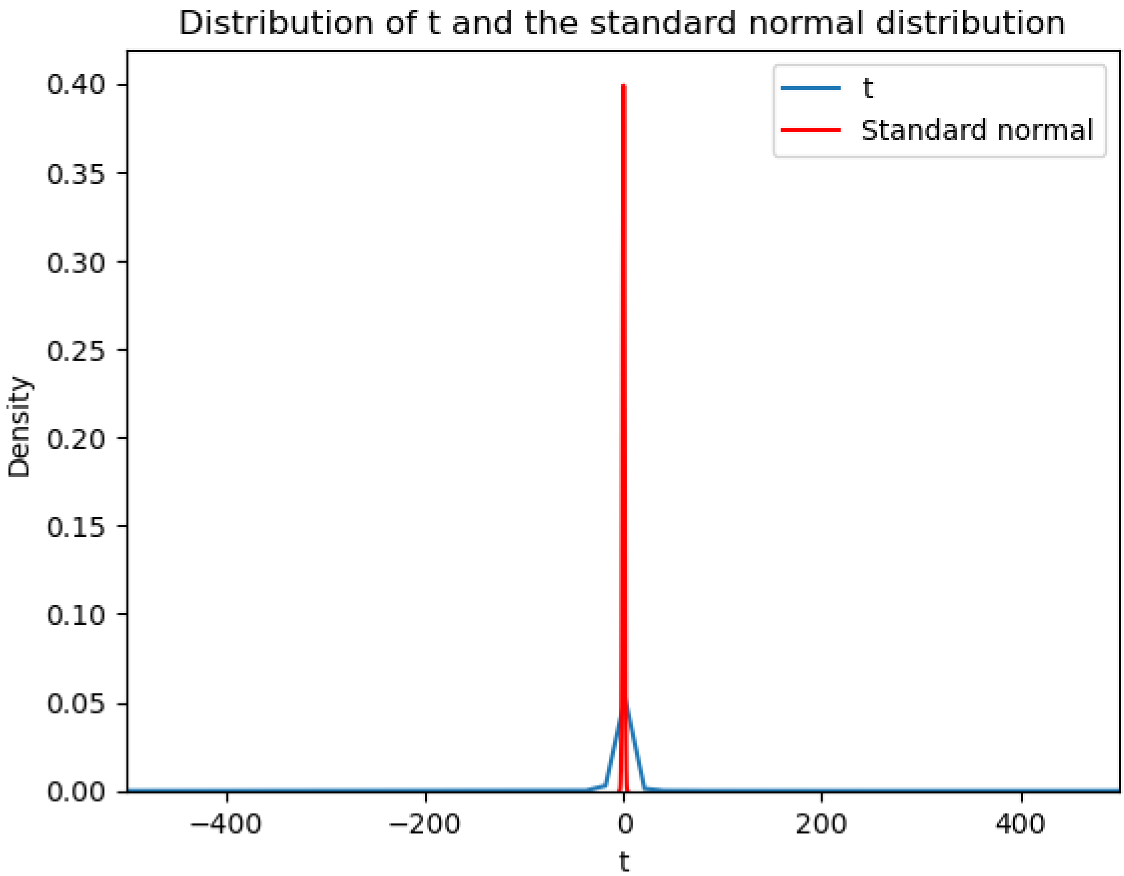
0.37957919573838766

سوال چهارم: کارخانه گلاب‌گیری

_۱



_۲

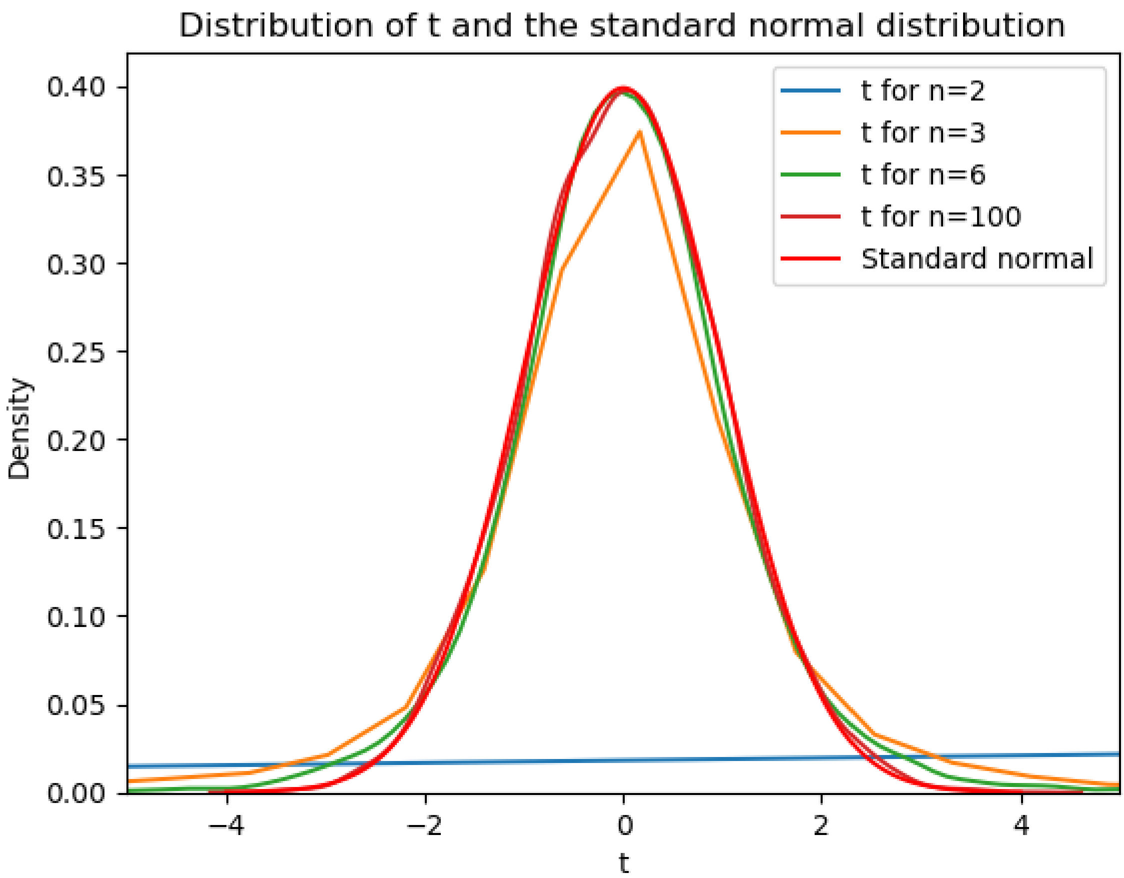


همانطور که در نمودار میبینیم ، نمودار نرمال استاندارد در دمها زودتر صفر میشود ولی نمودار

t های بدست آمده در دم ها ضخیم تر است و دیرتر به صفر میل میکند.

و در قسمت میانگین ، احتمال استاندارد نرمال بالاتر است ولی برای t این مقدار بسیار پایین تر است.

_۳



همانطور که انتظار میرفت با افزایش تعداد نمونه ها ، این نمودار به توزیع نرمال استاندارد نزدیک تر میشود.

مثلا برای تعداد نمونه های برابر با ۲ ، نمودار بسیار با نمودار استاندارد فرق دارد ولی با تبدیل این عدد به ۳ تا ، تغییر قابل توجهی در نمودار ایجاد میشود چرا که با تبدیل ۲ به عدد ۳ این عدد ۱.۵ برابر میشود و برای ۱۰۰۰۰ تا آزمایش نتایج بسیار بهتر خواهد بود.

با تغییر این عدد به ۶ ، درمقادیر مربوط به اطراف میانگین ، تقریبا میتوان گفت این نمودار بر نمودار استاندارد منطبق شده است اما هنوز در دم های نمودار تفاوت وجود دارد.

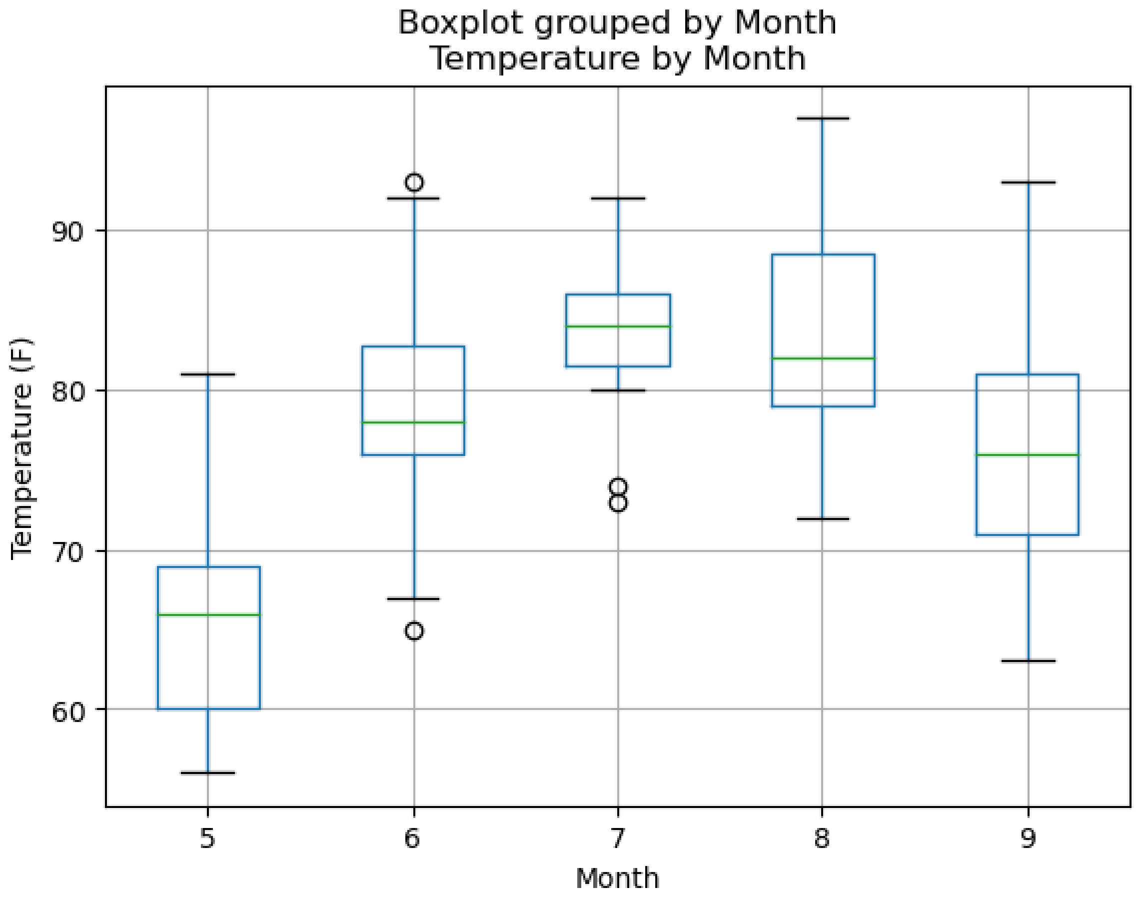
با تبدیل این عدد به ۱۰۰ دیگر تقریبا در همه جای این نمودار، میتوان شباهت و یا حتی انطباق بر نمودار استاندارد را دید.

در نتیجه با افزایش تعداد نمونه ها تقریب بهتری خواهیم داشت.

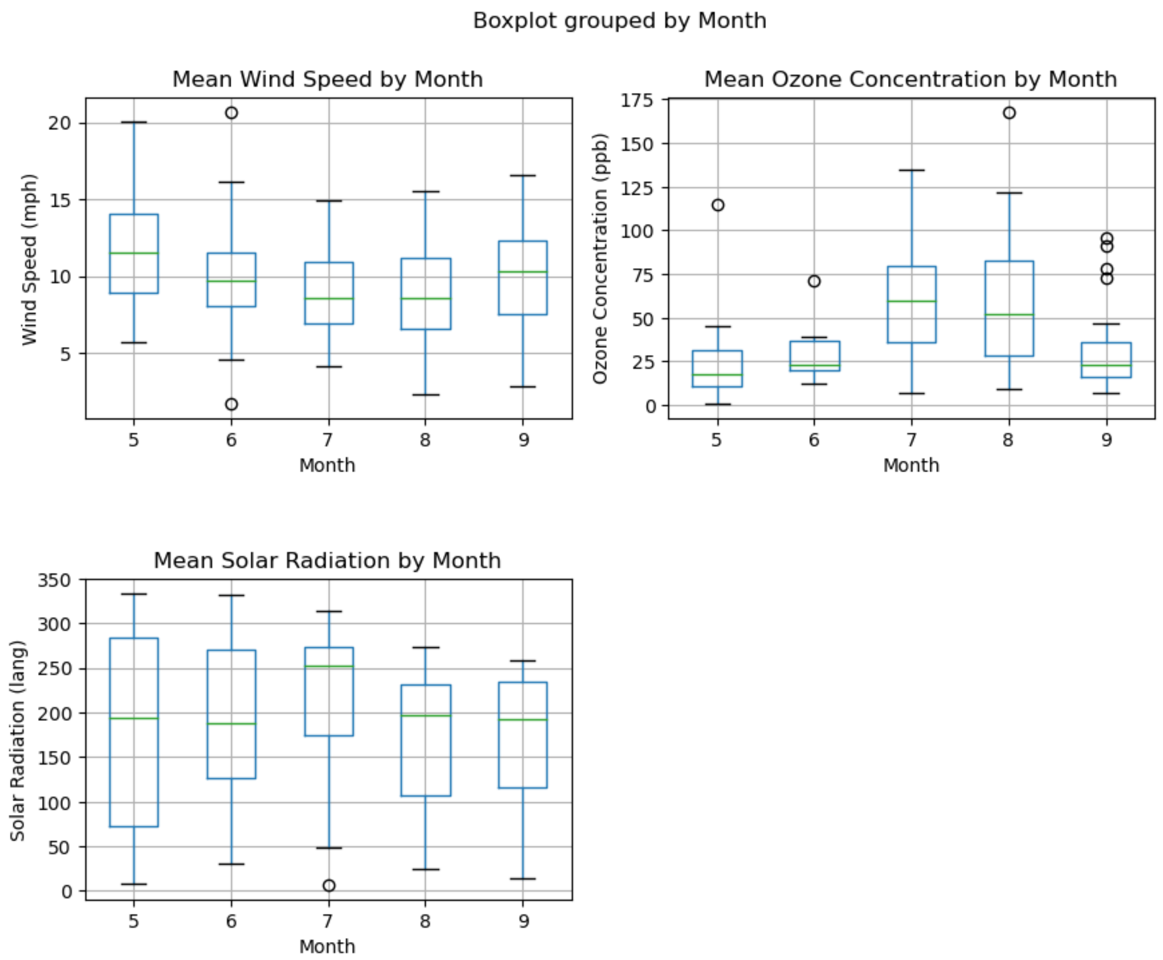
سوال پنجم: داده‌بازی

بخش اول

۲_



_۳



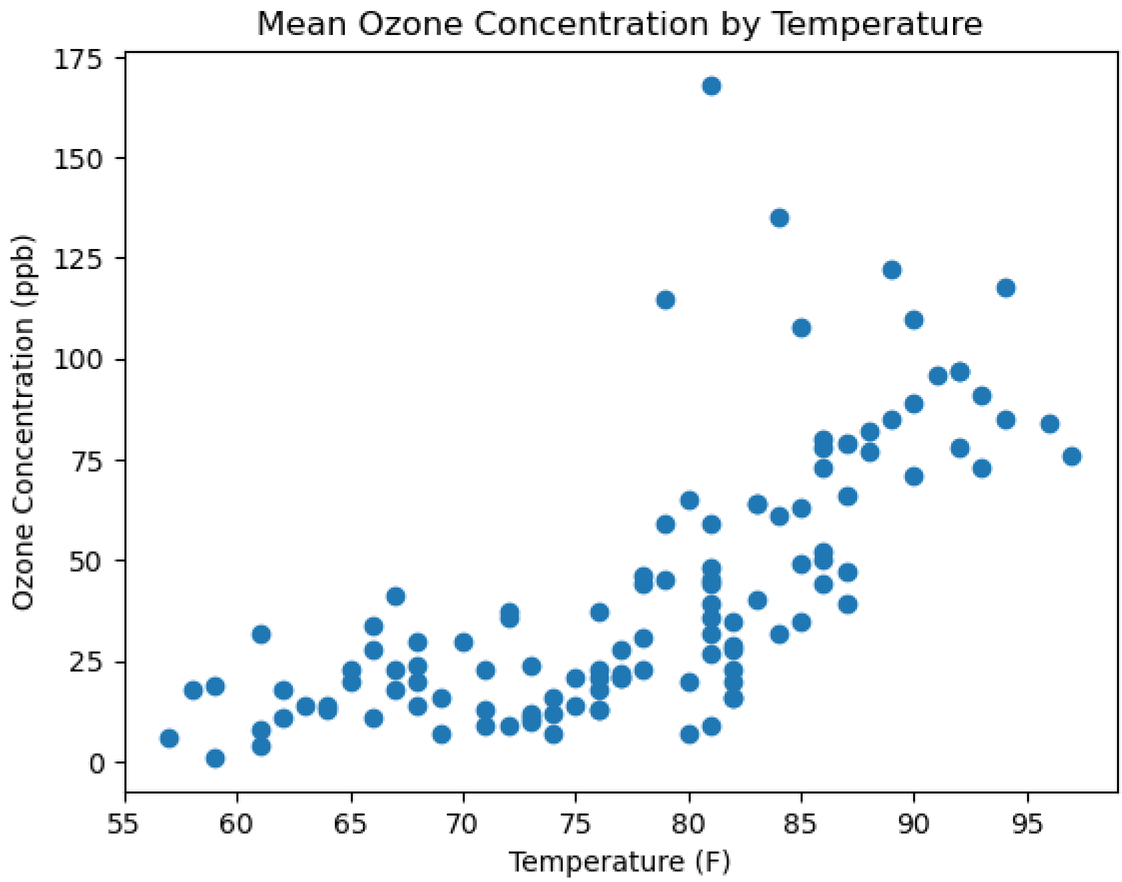
در ماه هایی که در بخش قبل دما بالا میرفت ، در این سه نمودار همانطور که انتظار میرود تابش خورشید و غلظت اوزون نیز بالا میرود و وزش باد کم میشود.

یعنی افزایش دما را میتوان ناشی افزایش تابش خورشید و کاهش وزش باد را ناشی از افزایش دما دانست و با این اوصاف غلظت اوزون نیز بالا خواهد بود.

و به طور برعکس در ماه هایی که دما پایین آمده ، طبق انتظارمان با توجه به دلایل علمی جوی ، وزش باد افزایش و تابش خورشید و غلظت اوزون افزایش یافته.

در واقع اگر از قبل اثرات دما و باد و تابش خورشید و غلظت اوزون را بر همدیگر نمیدانستیم ، میتوانستیم با این نمونه های داده ای به همچین نتیجه هایی برسیم(البته این داده ها برای این نتیجه گیری کم هستند).

_۴



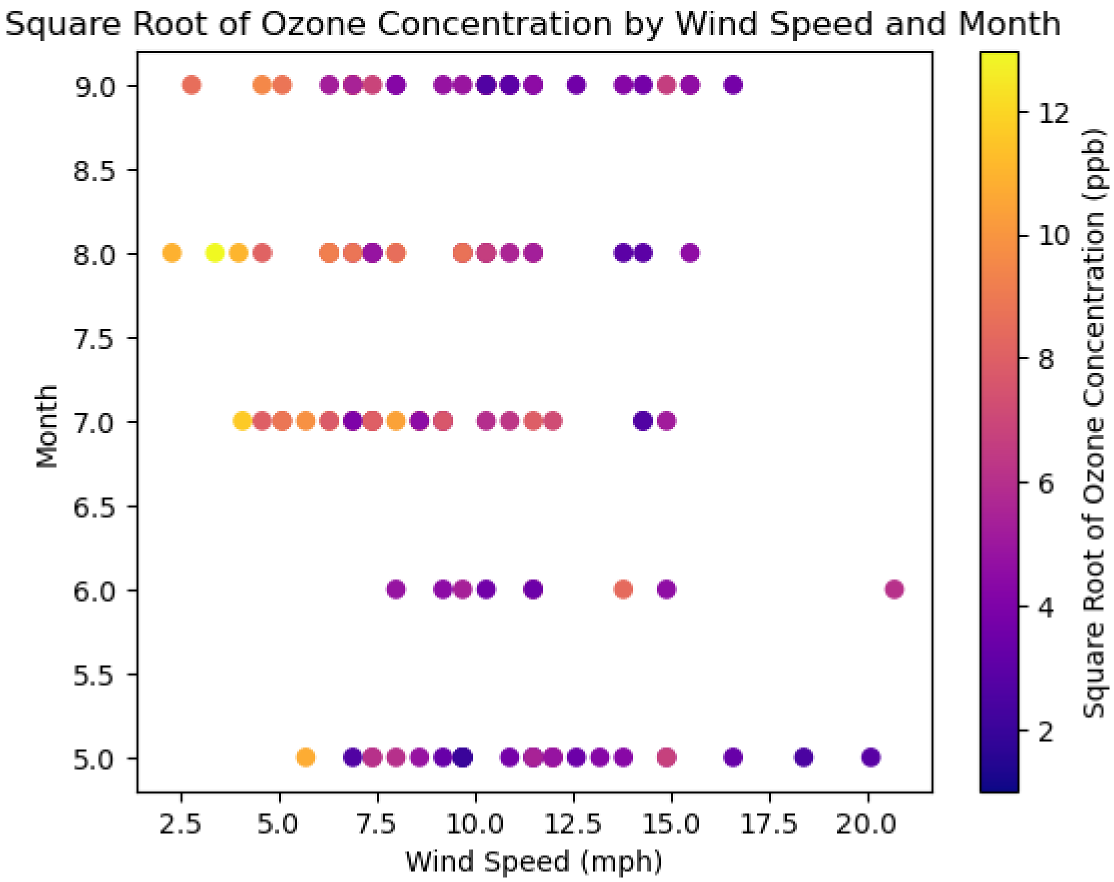
در این نمودار نیز با افزایش دما ، غلظت اوزون نیز به طور کلی افزایش میابد.

ولی این افزایش به صورت خطی و یکنواخت نیست و مثلا در دماهای بین ۶۰ تا ۶۵ غلظت اوزون کمی کاهش یافته ولی به صورت کلی افزایش میابد.

از طرفی در دماهای پایین تجمع غلظت اوزون بالاست و پراکندگی کمی دارد اما برای دماهای بالاتر پراکندگی افزایش میابد و میتوان گفت بیشترین پراکندگی برای دماهای بین ۸۰ تا ۹۰ اتفاق می افتد.

و پس از دمای ۸۰ میبینیم که تغییرات غلظت اوزون زیاد میشود و نسبت به قبل خیلی بیشتر افزایش میابد. ولی باتوجه به نمودار ، میانگین های غلظت اوزون پس از دمای حدودا ۹۳ دیگر افزایش نمیابند بلکه شروع به کاهش یافتن کرده است!

۵_

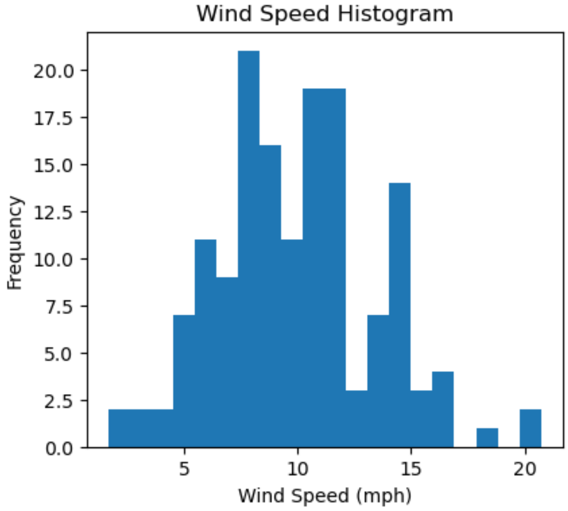


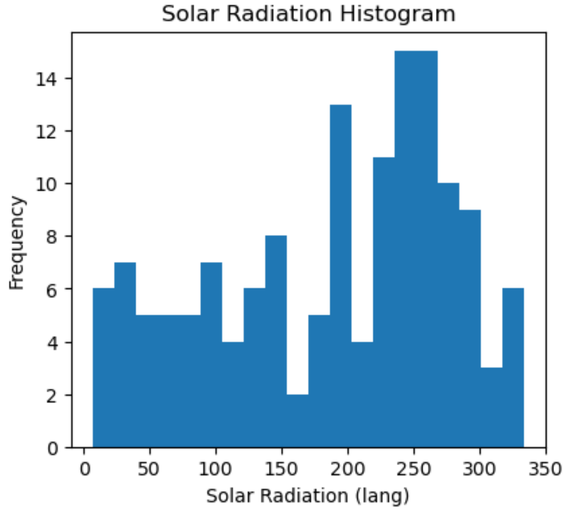
پیرو نتیجه گیری هایی که در قسمت های قبل کردیم ، در اینجا نیز بر آن نتیجه گیری ها تاکید میشود.

با بررسی این نمودار ، میبینیم که در جاهایی که سرعت باد کمتر است غلظت اوزون بیشتر میشود یعنی رنگ آن در نمودار بیشتر به زرد مایل میشود.

و همچنین در ماه هایی از سال که گرمتر است به نسبت غلظت اوزون بیشتر خواهد شد و به همان نسبت در ماه های گرم ، سرعت باد هم منطقا کاهش میابد.

ـ۶





با دقت در این دو نمودار درمیابیم که در زمانهایی که تابش خورشید بیشتر میشود سرعت باد کاهش میابد و با کاهش تابش خورشید نیز سرعت باد افزایش میابد و برعکس.

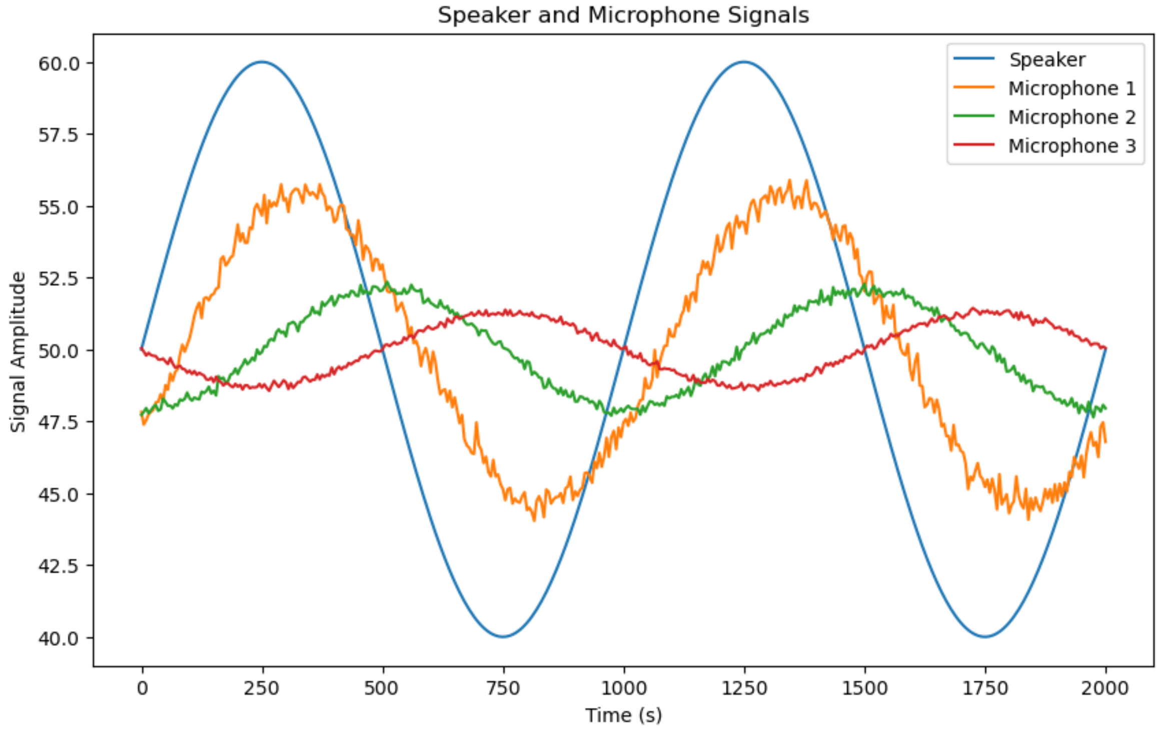
یعنی وقتی درنمودار مقدار سرعت باد بالا میرود ، تقریبا به همان نسبت تابش خورشید نیز کاهش یافته است.

و برای مقایسه ی این دو نمودار با نمودار نرمال، در رابطه با نمودار تابش خورشید میتوان گفت که شباهت چندانی وجود ندارد چراکه به جای اینکه وسط این نمودار ماکسیمم شود ، سمت راست نمودار ماکسیمم شده و ساختار کلی آن نیز متقارن نیست.

در رابطه با سرعت باد میتوانیم بگوییم نسبت به تابش خورشید شباهت بیشتری با نمودار وجود دارد چراکه ماسکیمم این نمودار تقریبا در وسط صورت گرفته اما خب بازهم این شکل متقارن نیست و سمت چپ آن کمی به نرمال شبیه است اما سمت راست آن بسیار تفاوت دارد چرا که ناگهان به صفر میل میکند و گاهی مقدار آن کمی بالا میرود.

بخش دوم

۱_



۲_

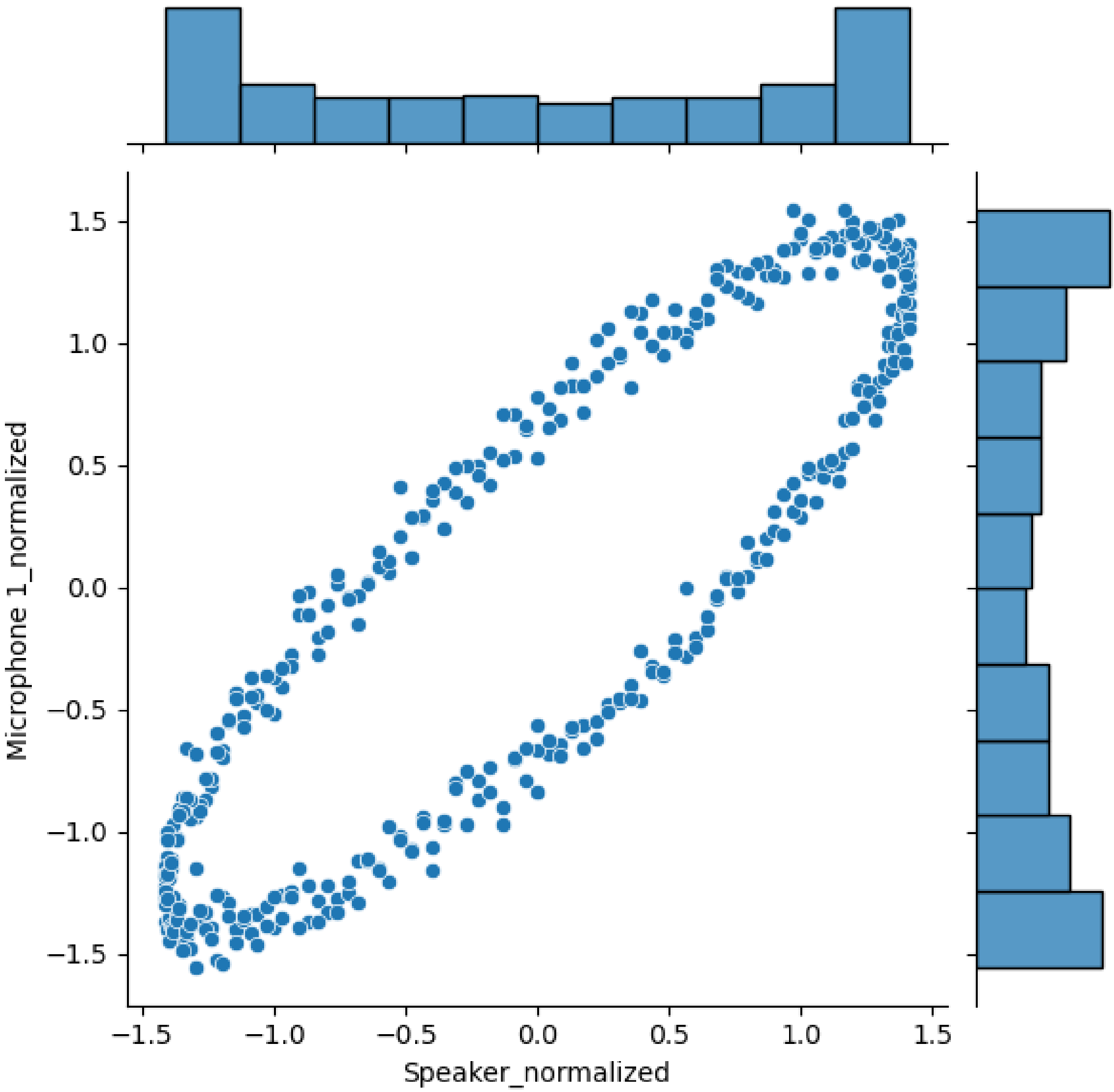
با توجه به اینکه باید توزیع داده ها نرمال استاندارد شده باشد ، بنابراین انتظار میرود میانگین آنها برابر ۰ و انحراف معیار آنها برابر ۱ باشد.

تست داده های جدید:

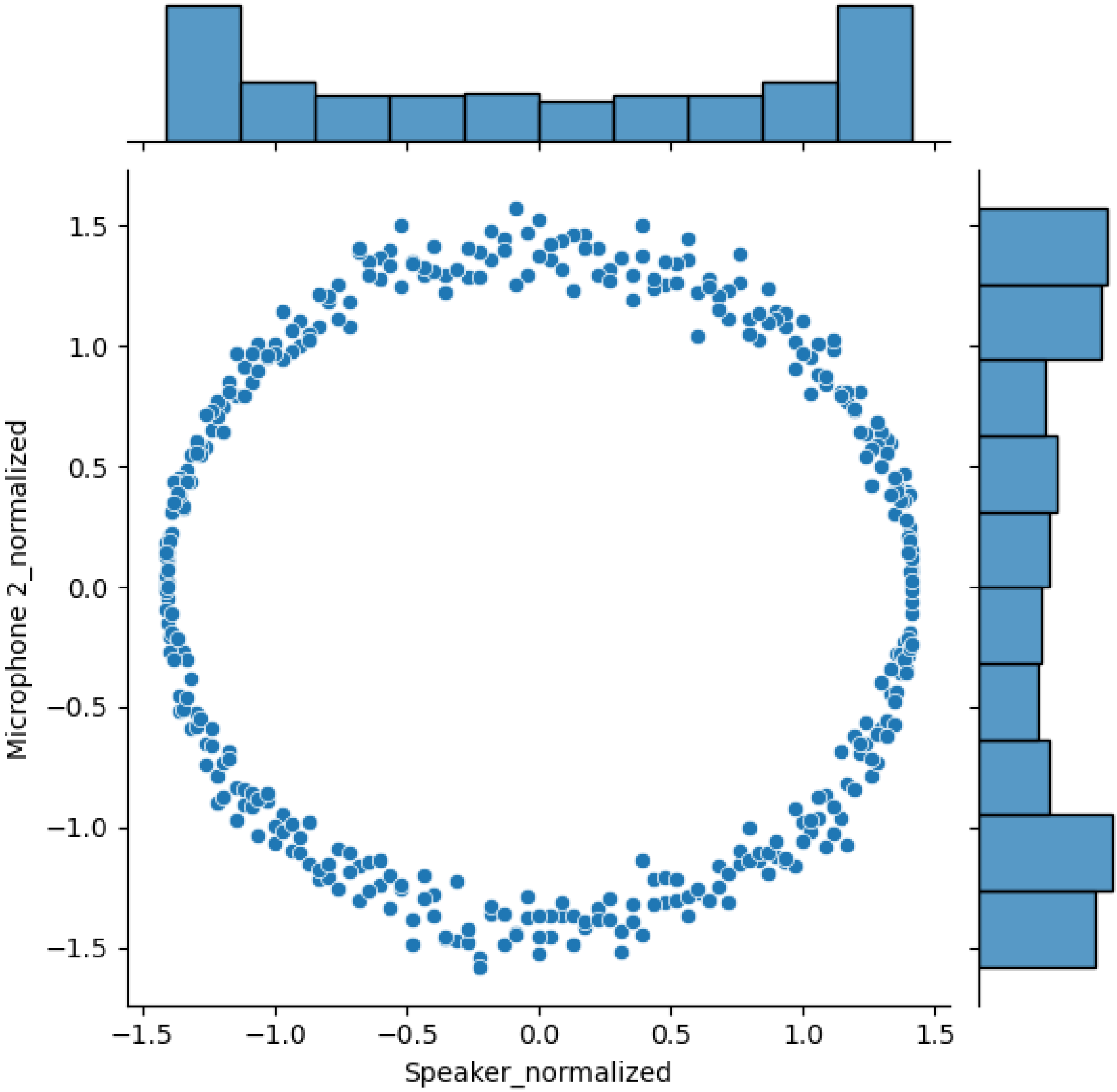
```
Speaker signal normalized mean: 0
Speaker signal normalized standard deviation: 1
Microphone 1 signal normalized mean: 0
Microphone 1 signal normalized standard deviation: 1
Microphone 2 signal normalized mean: 0
Microphone 2 signal normalized standard deviation: 1
Microphone 3 signal normalized mean: 0
Microphone 3 signal normalized standard deviation: 1
```

بنابراین داده ها به طرز درستی استاندارد شده اند.

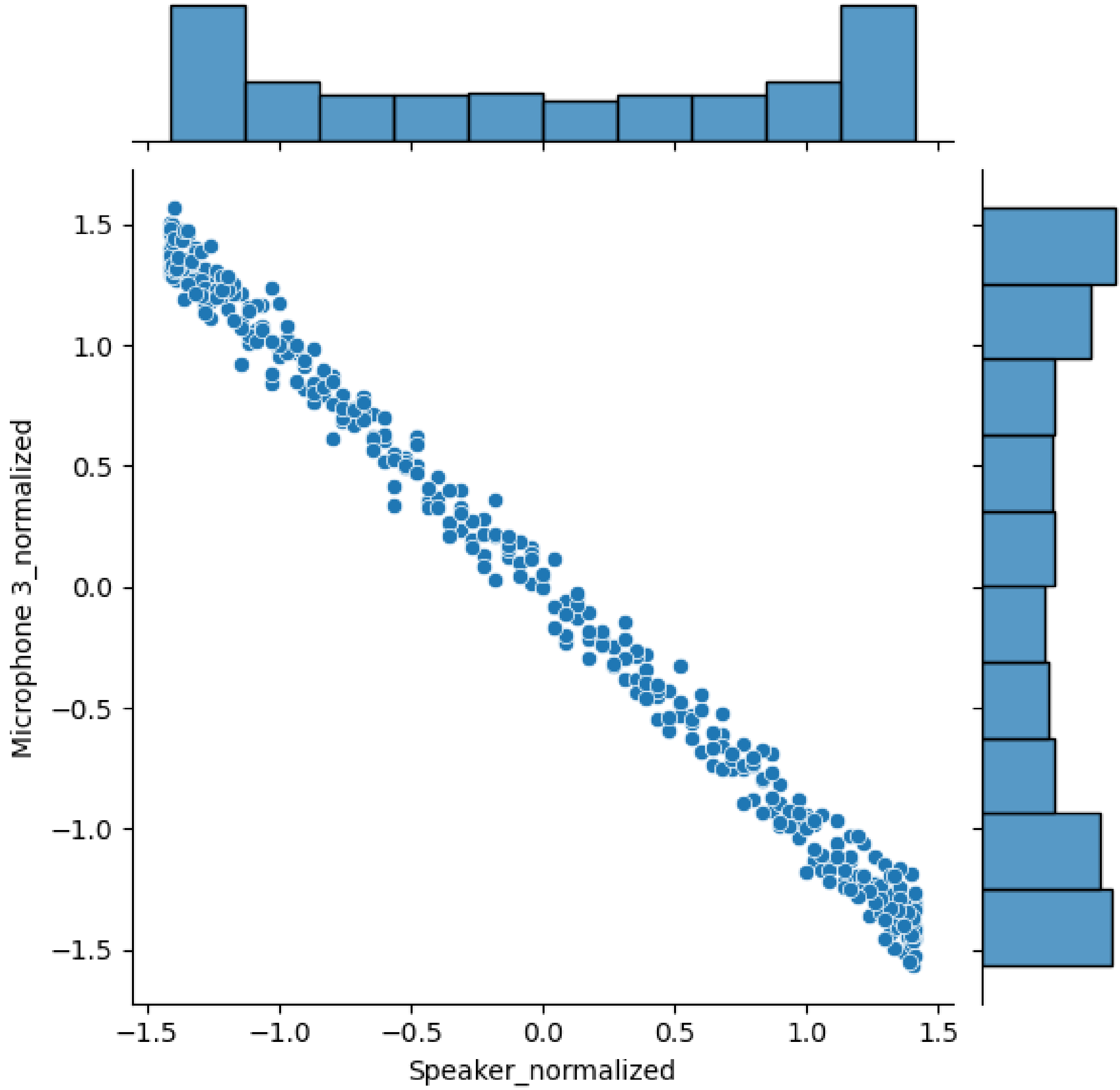
_٣



Microphone 1_normalized
Speaker_normalized
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5



1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Microphone 2_normalized
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Speaker_normalized



۴_

```
Covariance between Speaker and Microphone 1: 0.8628302937797431
Covariance between Speaker and Microphone 2: -0.007245529889605529
Covariance between Speaker and Microphone 3: -0.9967574051703081
```

با توجه به مفهوم کوواریانس ، هرچه اندازه ی کواریانس بزرگتر باشد نشانگر وابستگی بیشتر بین متغیرهاست و هر چه اندازه ی آن به صفر نزدیکتر باشد بیانگر وابستگی کمتر و برای مقدار صفر نشانگر عدم رابطه بین متغیر ها است.

مثبت بودن کوواریانس بیانگر رابطه ی مستقیم بین متغیرهاست و منفی بودن آن بیانگر رابطه عکس است.

پس با این تفاسیر با توجه به نمودار توزیع توام برای میکروفون اول و بلندگو ، رابطه ی بین این دو متغیر مستقیم است و میتوان گفت به همدیگر وابستگی قابل توجهی دارند(بیزی بسته است)

بنابراین انتظار میرود کوواریانس این دو ، عددی مثبت و بزرگتر از ۰.۵ باشد که باتوجه به نتایج بدست آمده این پیش بینی محقق شده(۰.۸۶۳)

و با توجه به نمودار توزیع توام میکروفون دوم و بلندگو ، رابطه ی چندانی بین این دو متغیر وجود ندارد و به هم وابستگی ندارد بنابراین انتظار میرود که کوواریانس عددی نزدیک به صفر باشد که باتوجه به نتیجه بدست آمده این پیش بینی درست بود(۰.۰۰۷-)

حالا با توجه به نمودار توزیع توام میکروفون سوم و بلندگو ، باتوجه به اینکه این نمودار خیلی شبیه به نمودار خطی است بنابراین وابستگی بسیار زیادی بین این دو متغیر وجود دارد و با توجه به جهت این خط میتوان گفت رابطه ی این دو متغیر عکس است . پس کوواریانس عددی منفی و اندازه ی آن بسیار نزدیک به یک پیش بینی میشود که با توجه به نتیجه این ادعا صحیح است(۰.۹۹۹-)