به نام خدا

دانشگاه تهران

دانشکده مهندسی برق و کامپیوتر

آمار و احتمال مهندسی

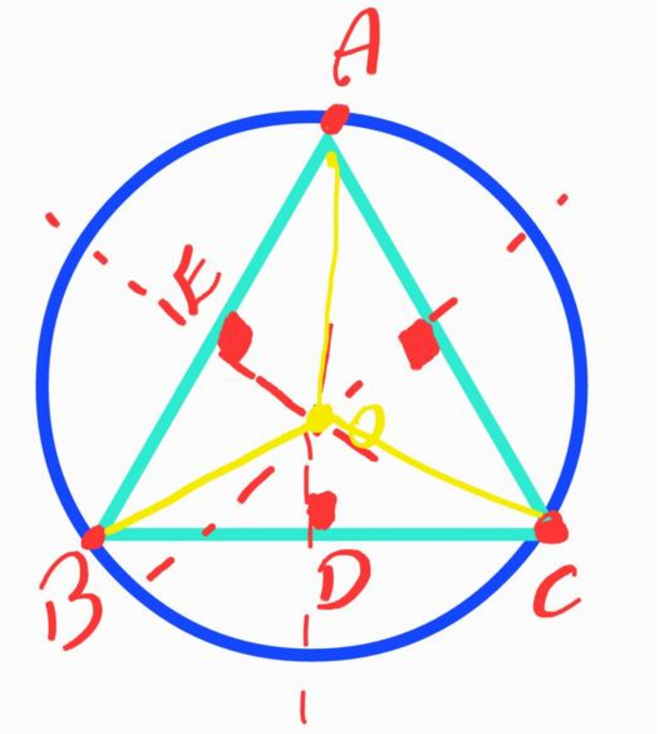
گزارش پروژه‌ی نهائی

نام و نام خانوادگی	فردین عباسی
شماره دانشجویی	810199456
تاریخ ارسال گزارش	1400/10/22

فهرست گزارش سوالات (لطفاً پس از تکمیل گزارش، این فهرست را به‌روز کنید.)

سوال اول: پارادوکس برتراند

ابتدا اندازه ضلع مثلث متساوی الاضلاع محاطی داخل دایره را محاسبه می کنیم:



رابطهٔ شعاع دایرهٔ محیطی و مثلث داخل آن به صورت زیر است :

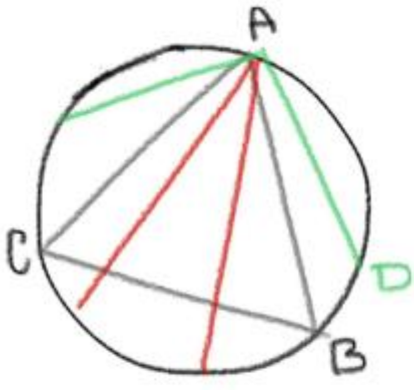
$$R = \frac{abc}{4S} \to a = b = c, S = \frac{\sqrt{3}}{4}a^2$$

$$\to 1 = \frac{a^3}{4 \times \frac{\sqrt{3}}{4}a^2} \to a$$

$$= \sqrt{3} = 1.732$$

درنتیجه برای آن که شرط مسأله برقرار باشد، طول وتر ما باید از $\sqrt{3} = 1.732$ بزرگتر باشد.

1.1.الف :



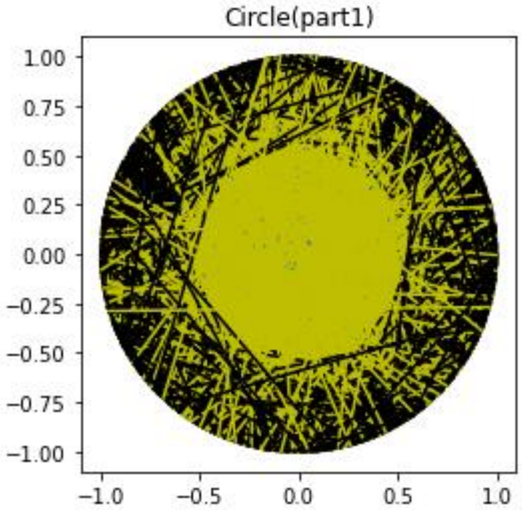
برای آن که وتر AD بزرگتر از طول ضلع مثلث باشد، نقطهٔ D باید در بازهٔ انتخاب انتخاب گردد. در نتیجه احتمال مطلوب ما، نسبت اندازهٔ BC به اندازهٔ محیط دایره است :

$$P = \frac{BC}{2\pi} = \frac{\frac{2\pi}{3}}{2\pi} = \frac{1}{3}$$

ب :

برای این بخش، مطابق گزارش کار، دو ماتریس 1000 تایی از زاویه ها درست کردیم، سپس در 4 ماتریس، مختصات سینوس و کسینوس آن ها ریخته شد. که چون مختصات ما دایره به شعاع واحد است، همان مختصات ابتدایی و انتهایی وتر به شمار می روند.

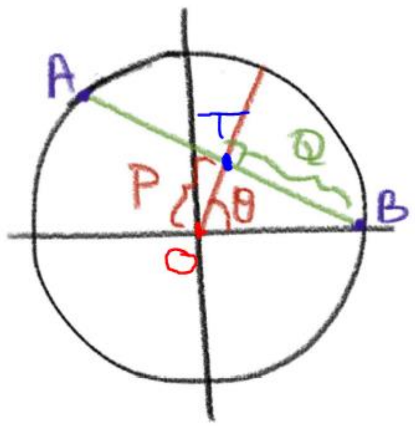
در نهایت هم وترها با توجه به این مختصات کشیده شده و نسبت آن هایی که طولشان بیشتر از 1.732 است به نسبت کل 0.338 است که با تقریب بسیار خوبی، نتایج تئوری ما را تأیید می کند.



Probability(part1)= 0.338

وترهای کوچکتر با رنگ مشکی، و وترهای بزرگتر با رنگ زرد نشان داده شده اند.

1،2،الف :

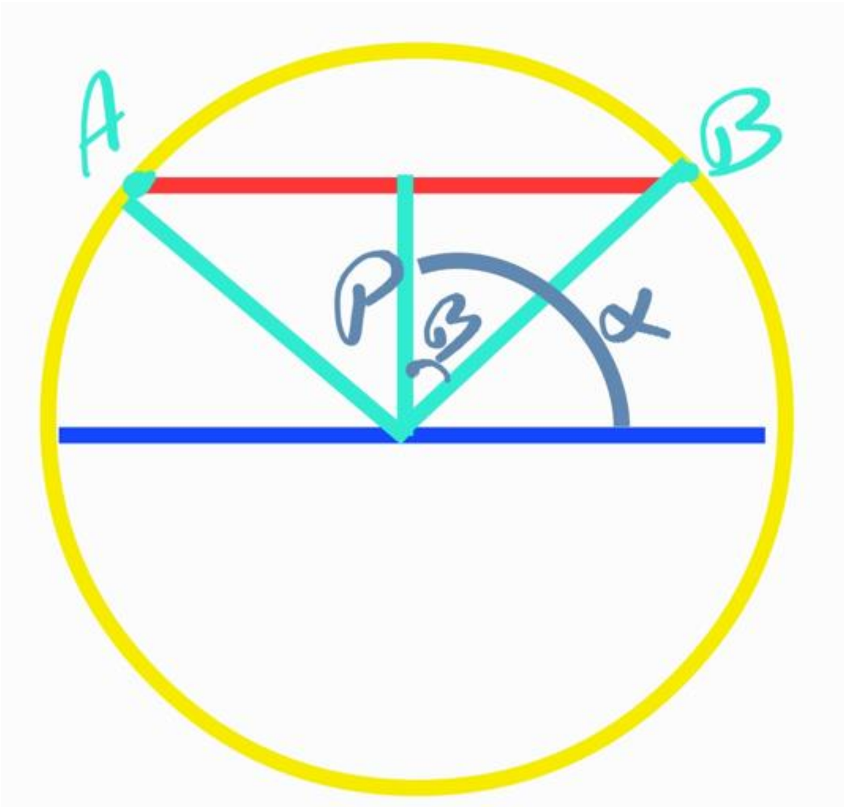


فرض کنید نقطهٔ تقاطع دو وتر عمود را T، و مرکز دایره را O بنامیم. طول OB، برابر طول شعاع دایره و برابر یک است. مثلث OBT نیز یک مثلث قائم الزاویه است. برای آن که طول وتر AB بزرگتر از $\sqrt{3}$ باشد، طول BT که نصف وتر است، باید بزرگتر از $\frac{\sqrt{3}}{2}$ باشد. طبق قضیه فیثاغورث :

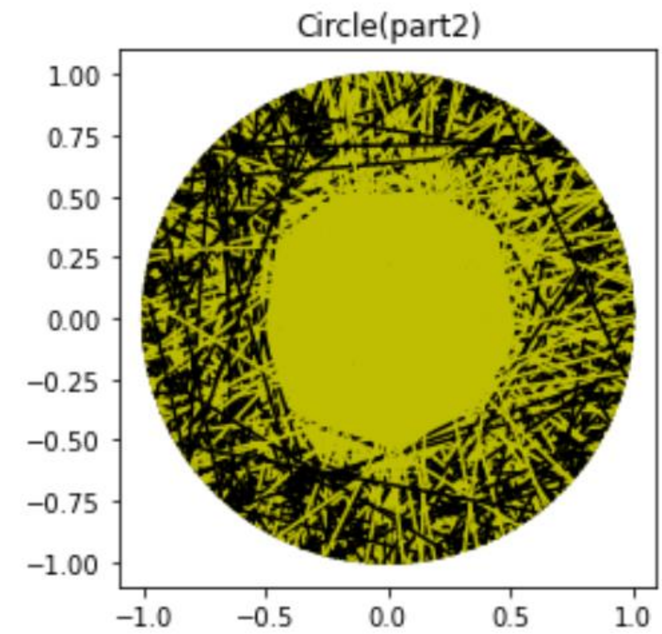
$$BT > \frac{\sqrt{3}}{2} \rightarrow 1 - OT^2 > \frac{3}{4} \rightarrow OT = P < \frac{1}{2}$$

چون که اندازهٔ P، یک توزیع یکنواخت بین 0 تا 1 است، درنتیجه احتمال اینکه اندازه آن کوچکتر از یک دوم باشد برابر یک دوم است.

1،2ب)



برای کشیدن وترها، باید مختصات A و B را بیابیم. با توجه به شکل، زاویهٔ A برابر آلفا منهای بتا و زاویهٔ B برابر آلفا بعلاوهٔ بتا است. در نتیجه ماتریس جمع و تفریق دو زاویه را بدست آورده و طبق بخش قبلی، با استفاده از سینوس و کسینوس مختصات آن ها را بدست آورده و رسمشان می کنیم.

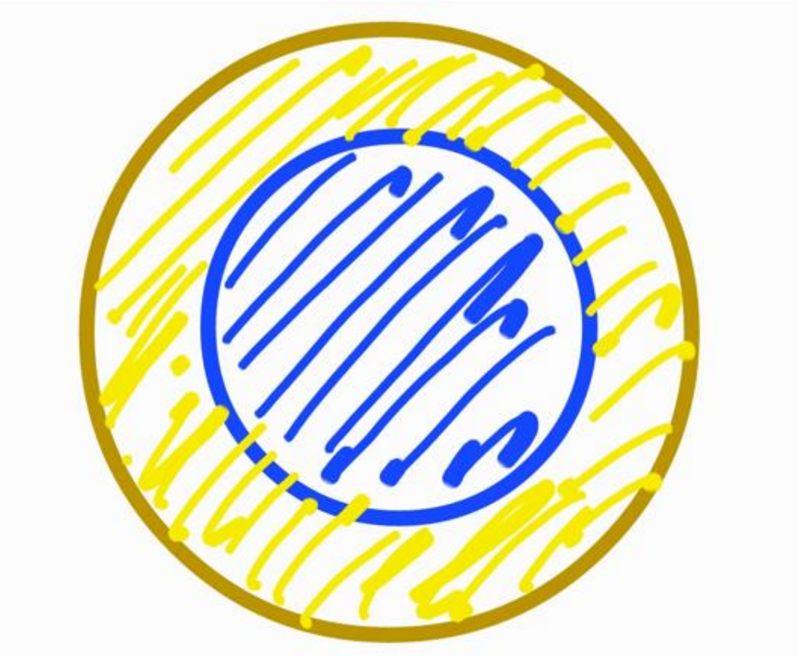


نتیجه شبیه سازی نیز، عدد 0.538را نشان می دهد که نتیج تئوری ما را تایید می کند :

```
Probability(part2)= 0.538
```

وترهای بزرگتر با رنگ زرد، و وترهای کوچکتر با رنگ مشکی نشان داده شده اند.

3،1الف) در قسمت 2، به این نتیجه رسیدیم که برای اینکه اندازهٔ وتر بزرگتر از ضلع مثلث باشد، فاصلهٔ نقطهٔ انتخابی تا مرکز باید بیشتر از یک دوم باشد. در قسمت سوم، چون ما با ناحیه سروکار داریم، برای اینکه این شرط محقق گردد، نقطهٔ انتخابی باید در ناحیهٔ آبی رنگ باشد که، یک دایره به شعاع یک دوم است. درنتیجه احتمال ما، نسبت مساحت دایرهٔ آبی رنگ به دایرهٔ زرد رنگ است :

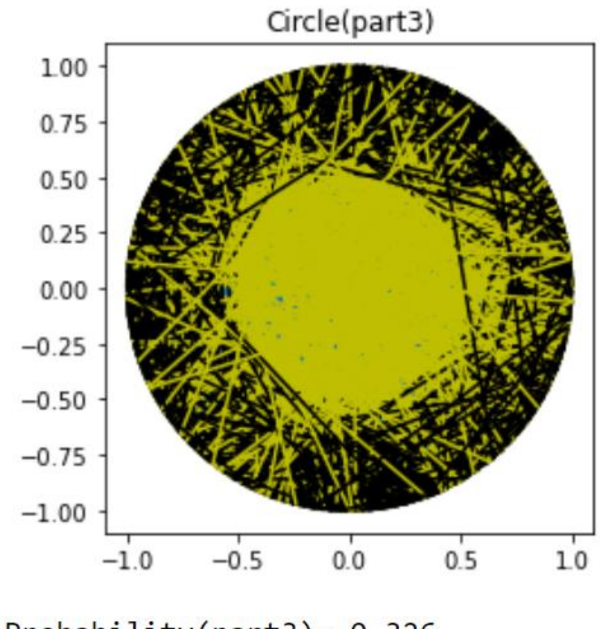


$$P = \frac{\pi(\frac{1}{2}^2)}{\pi 1^2} = \frac{1}{4}$$

3،1ب :

برای این بخش نیز دقیقا تمام مراحلی را که در بخش قبلی طی کردیم طی می کنیم، فقط این بار به جای P جذر آن را قرار می دهیم. دلیل این آن هم این است که در فضای احتمال را از فضای خطی بخش دوم، به فضای سطحی مدنظر انتقال دهیم. انگار که داریم نقطه موردنظرمان را در سطح دایره انتخاب می کنیم.

همانطور که می بینم، احتمال بدست آمده از شبیه سازی0.326است که داده های تئوری ما را تأیید می کند.



سوال دوم: تخمین عدد اویلر

دنباله $a_n = \left(1+\frac{1}{n}\right)^n$ را در نظر گرفته و چند جمله آنرا بدست می‌آوریم:

$$\left\{\left(1+\frac{1}{n}\right)^n\right\}_1^\infty = \left\{\left(1+\frac{1}{1}\right)^1, \left(1+\frac{1}{2}\right)^2, \left(1+\frac{1}{3}\right)^3, \cdots, \left(1+\frac{1}{n}\right)^n, \cdots\right\}$$
$$= \left\{\underbrace{2}_{n=1}, \underbrace{2.25}_{n=2}, \underbrace{2.370}_{n=3}, \cdots, \underbrace{2.692}_{n=50}, \cdots, \underbrace{2.705}_{n=100}, \cdots, \underbrace{2.717}_{n=1000}, \cdots\right\}$$

به نظر می‌رسد که این دنباله صعودی بوده و از بالا کراندار است. اگر چنین باشد می‌توان گفت دنباله همگرا بوده و دارای حد است.

حال این موضوع را به طریق ریاضی نشان میدهیم. ابتدا بررسی میکنیم که دنباله اکیدا صعودی است. برای این منظور:

$$\frac{a_{n+1}}{a_n} = \left(1+\frac{1}{n}\right)\left(1-\frac{1}{(n+1)^2}\right)^{n+1} > \left(1+\frac{1}{n}\right)\left(1-\frac{1}{n+1}\right) = 1$$

چرا که بر طبق نامساوی برنولی اگر $-1 < x$ و $r > 1$ آنگاه:

$$(1+x)^r > 1+rx \rightarrow \left(1+\frac{-1}{(n+1)^2}\right)^{n+1} > 1-\frac{(n+1)}{(n+1)^2} = 1-\frac{1}{n+1}$$

از آنجا که $a_1 = 2$ و دنباله a_n صعودی است، لذا $a_n > 2$. حال نشان می‌دهیم $a_n < 3$ نیز میباشد یعنی از بالا کراندار است.
برای این منظور از بسط دوجمله‌ای استفاده می‌کنیم. خواهیم داشت:

$$a_n = \left(1+\frac{1}{n}\right)^n = 1+n\frac{1}{n}+\frac{1}{2!}\frac{n(n-1)}{n^2}+\cdots+\frac{1}{n!}\frac{n(n-1)\cdots(n-n+1)}{n^n}$$
$$= 1+1+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\cdots+\frac{1}{n!}\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{n-1}{n}\right)$$
$$< 2+\frac{1}{2!}+\frac{1}{3!}+\cdots+\frac{1}{n!} < 2+\frac{1}{2}+\frac{1}{2^2}+\cdots+\frac{1}{2^{n-1}} = 2+\left(1-\left(\frac{1}{2}\right)^{n-1}\right) < 3$$

که در سطر آخر از نامساوی $\frac{1}{n!} < \frac{1}{2^{n-1}}$ استفاده شده است. بنابراین دنباله مورد نظر صعودی و از بالا کراندار است. لذا دارای یک حد خواهد بود که آنرا حد را بنام عدد نپر معرفی کرده و با e نمایش می‌دهیم. در نتیجه:

$$\lim_{n\to+\infty}\left(1+\frac{1}{n}\right)^n = e = 2.7182818\ldots\ldots$$

همچنین می‌توان نشان داد وقتی $n \to -\infty$ حد مورد نظر با زمانی که $n \to +\infty$ یکسان بوده و لذا آن هم برابر e خواهد شد.

توضیح ۱: نشان می‌دهیم این حد به ازای هر عدد حقیقی و مثبت x نیز برقرار است.

$$n=[x] \rightarrow n \le x < n+1 \rightarrow \frac{1}{n+1} < \frac{1}{x} \le \frac{1}{n} \rightarrow 1+\frac{1}{n+1} < 1+\frac{1}{x} \le 1+\frac{1}{n} \xrightarrow{n\le x<n+1}$$
$$\left(1+\frac{1}{n+1}\right)^n < \left(1+\frac{1}{x}\right)^x < \left(1+\frac{1}{n}\right)^{n+1} \rightarrow \frac{\left(1+\frac{1}{n+1}\right)^{n+1}}{\left(1+\frac{1}{n+1}\right)} < \left(1+\frac{1}{x}\right)^x < \left(1+\frac{1}{n}\right)^n\left(1+\frac{1}{n}\right)$$

حال اگر طرفین به $+\infty$ میل کند، حد دو طرف برابر e می‌باشد، لذا حد وسط نیز تحت فشار طرفین e خواهد شد. به همین ترتیب می‌توان برای هر عدد حقیقی و منفی x نیز نشان داد که این حد تغییر نمی‌کند. لذا در نهایت:

$$\boxed{\lim_{x\to\pm\infty}\left(1+\frac{1}{x}\right)^x = e}$$

خانم شاکری اجازه قرار دادن تصویر به جای تایپ را دادند!

روش دوم:

Solution. We will find E[N] by obtaining a more general result. For $x \in [0, 1]$, let

$$N(x) = \min\left\{n: \sum_{i=1}^{n} U_i > x\right\}$$

and set

$$m(x) = E[N(x)]$$

That is, $N(x)$ is the number of uniform (0, 1) random variables we must add until their sum exceeds x, and $m(x)$ is its expected value. We will now derive an equation for $m(x)$ by conditioning on U_1. This gives, from Equation (5.1b),

$$m(x) = \int_0^1 E[N(x)|U_1 = y]\,dy \tag{5.6}$$

Now,

$$E[N(x)|U_1 = y] = \begin{cases} 1 & \text{if } y > x \\ 1 + m(x - y) & \text{if } y \le x \end{cases} \tag{5.7}$$

The preceding formula is obviously true when $y > x$. It is also true when $y \le x$, since, if the first uniform value is y, then, at that point, the remaining number of uniform random variables needed is the same as if we were just starting and were going to add uniform random variables until their sum exceeded $x - y$. Substituting Equation (5.7) into Equation (5.6) gives

$$\begin{aligned} m(x) &= 1 + \int_0^x m(x - y)\,dy \\ &= 1 + \int_0^x m(u)du \qquad \text{by letting } u = x - y \end{aligned}$$

Differentiating the preceding equation yields

$$m'(x) = m(x)$$

or, equivalently,

$$\frac{m'(x)}{m(x)} = 1$$

Integrating this equation gives

$$\log[m(x)] = x + c$$

or

$$m(x) = ke^x$$

Since $m(0) = 1$, it follows that $k = 1$, so we obtain

$$m(x) = e^x$$

Therefore, $m(1)$, the expected number of uniform (0, 1) random variables that need to be added until their sum exceeds 1, is equal to e. ■

سوال سوم: کبریت باناخ

میدانیم برای آن که کبریت های یکی جعبه به اتمام برسد، باید N بار آن جعبه انتخاب شود. به همین شیوه، برای آن که در جعبه دیگر نیز K کبریت باقی بماند، باید N-K بار از آن کبریت بیرون آورده شود. در نتیجه تعداد 2N-K بار آزمایش میکنیم، تا کبریت یکی از جعبه ها تمام شده باید و در دیگری، K عدد کبریت باقی مانده باشد.

همچنین چون که انتخاب جعبه ها نسبت به یکدیگر برتری ندارند، در نتیجه احتمال انتخاب هریک آن ها مساوی و برابر یک دوم است.

$$P_X(x=k)=\binom{2N-K}{N}\times\left(\frac{1}{2}\right)^{\mathrm{N-K}}\times\left(\frac{1}{2}\right)^{N}\times\left(\frac{1}{2}\right)=\binom{2N-K}{N}\times\left(\frac{1}{2}\right)^{\mathrm{2N-K}}$$

تشریح عبارت بالا نیز به این صورت است :

2N-K بار باید جعبه ها بیرون آورده شوند. در N بار از این حالات، باید یکی از جعبه ها (مثلا راست) بیرون آورده شود که احتمال آن برابر یک دوم به توان N است. برای N-K بار بعدی نیز جعبهٔ دیگر باید بیرون آورده شود که احتمال آن برابر دو به توان N-K می باشد.

.2

دو نکته:

۱- فرمول استرلینگ:

$$if\ n\to\infty\ :\ \Gamma(n+1)\sim\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \quad \text{(Stirling's formula)}$$

که اگر $n\in\mathbb{N}$ بهتر است $\Gamma(n+1)$ را بصورت ساده‌تر $n!$ بیان کرد. اثبات رابطه استرلینگ در مثال ۱۱-۴۱ ارائه خواهد شد.

※ **مثال ۱۱-۴۱** با استفاده از تابع گاما، فرمول استرلینگ را ثابت کنید. یعنی اگر $n\to\infty$ آنگاه: $n!\approx\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

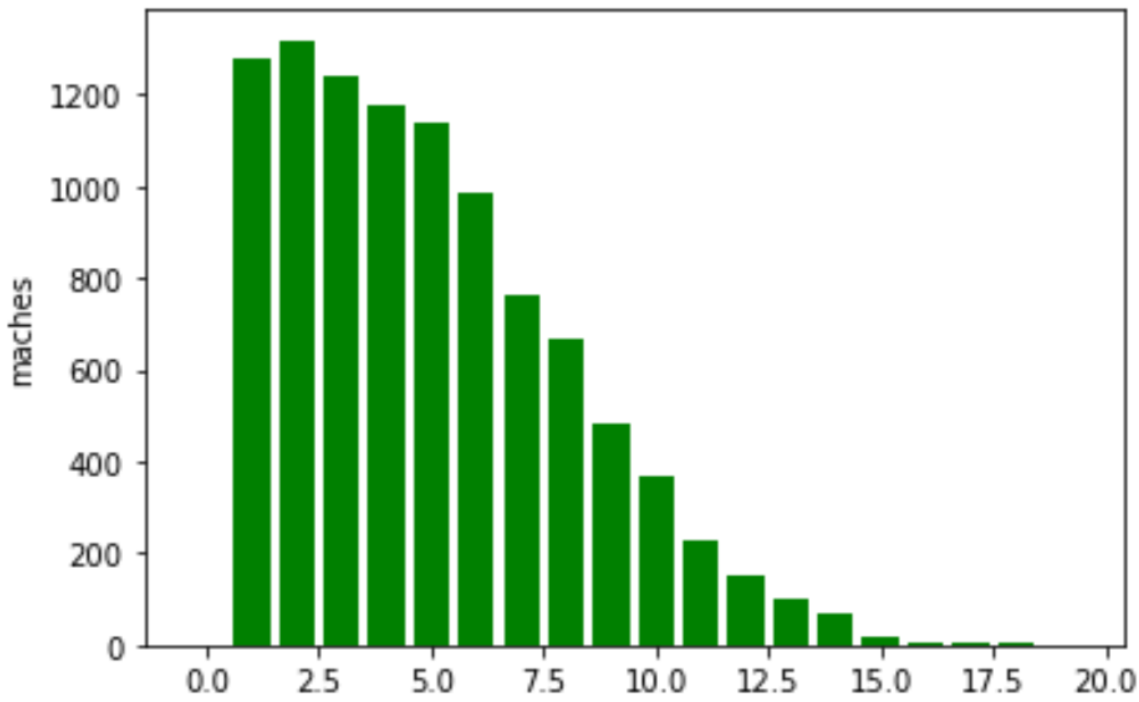
$$n!=\Gamma(n+1)=\int_0^{+\infty}e^{-t}x^n dx=\int_0^{+\infty}e^{n\ln x-x}dx\xrightarrow{x=n+y}n!=e^{-n}\int_{-n}^{+\infty}e^{n\ln(n+y)-y}dy$$

$$=e^{-n}\int_{-n}^{+\infty}e^{n\ln n+n\ln\left(1+\frac{y}{n}\right)-y}dy=e^{-n}n^n\int_{-n}^{+\infty}e^{n\left(\frac{y}{n}-\frac{y^2}{2n^2}+\frac{y^3}{3n^3}-\cdots\right)-y}dy$$

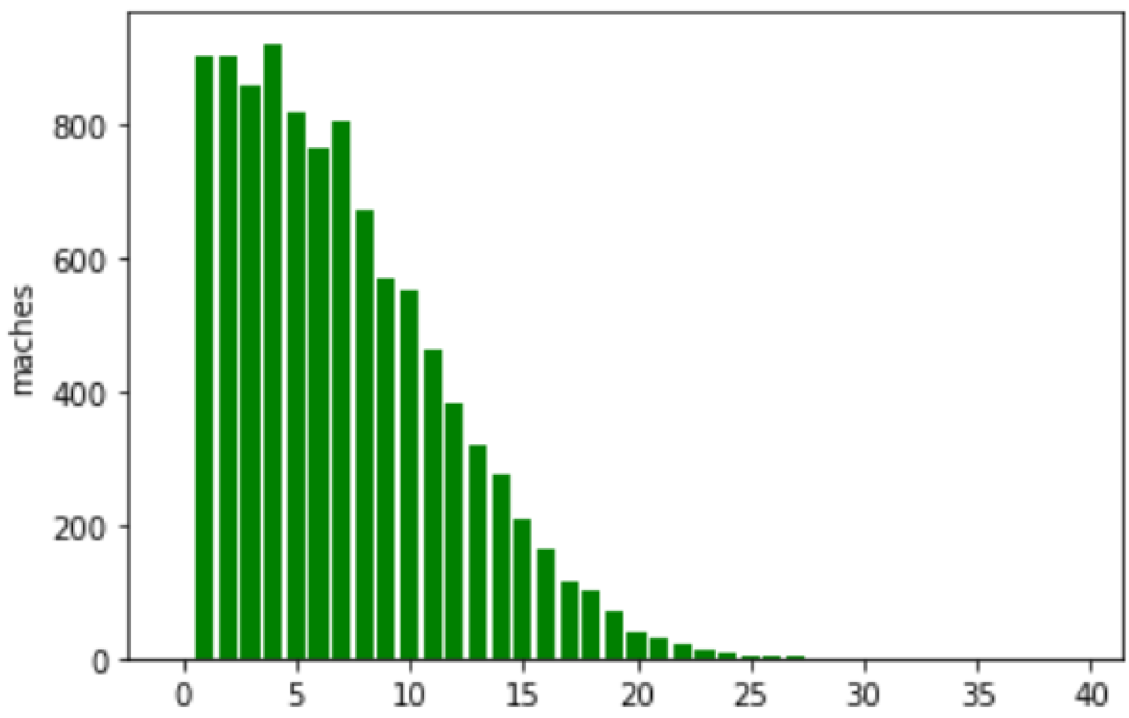
$$\xrightarrow{y=\sqrt{n}t}n!=e^{-n}n^n\sqrt{n}\int_{-n}^{+\infty}e^{-\frac{t^2}{2}+\frac{t^3}{3\sqrt{n}}-\cdots}dt\approx e^{-n}n^n\sqrt{n}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt\to n!\approx\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \quad \blacksquare$$

(3.3

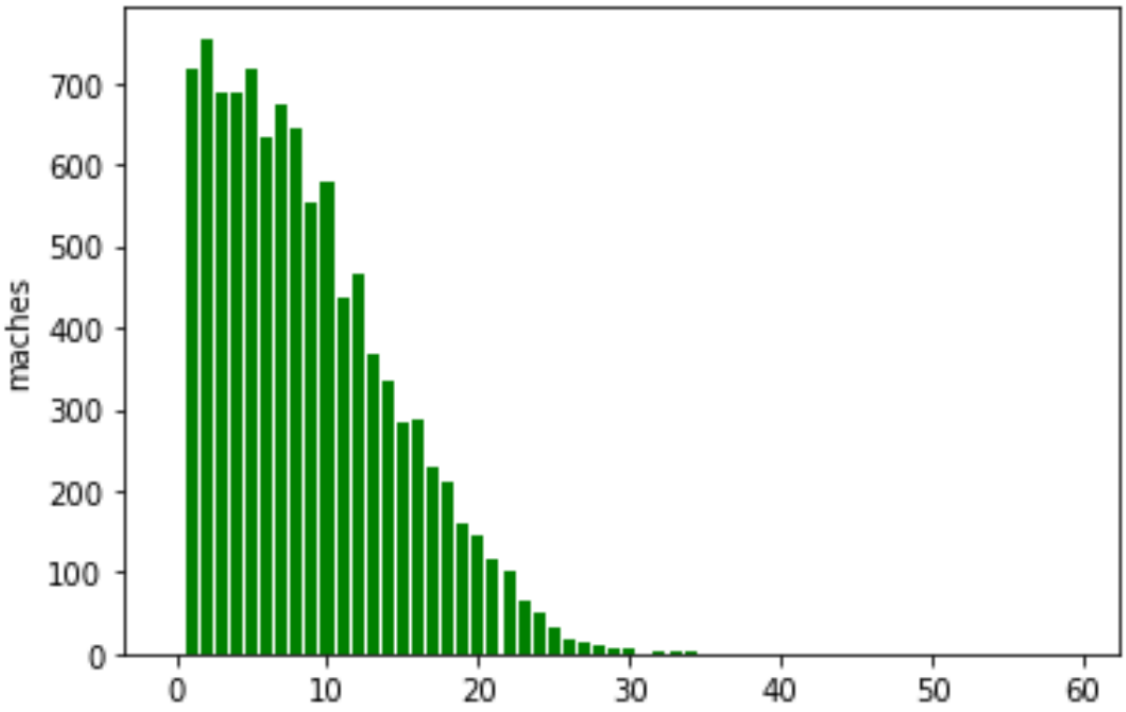
N=20:


maches
0
200
400
600
800
1000
1200
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

N=40:


maches
0
200
400
600
800
0
5
10
15
20
25
30
35
40

N=40:



برای تایید جواب شبیه سازی، احتمال را به ازای K = 2 و K = N در هر 3 حالت محاسبه می کنیم :

$$\binom{38}{20} \times \left(\frac{1}{2}\right)^{38} =$$

$$\binom{78}{40} \times \left(\frac{1}{2}\right)^{78} =$$

$$\binom{118}{60} \times \left(\frac{1}{2}\right)^{118} =$$

$$\binom{20}{20} \times \left(\frac{1}{2}\right)^{20} =$$

$$\binom{40}{40} \times \left(\frac{1}{2}\right)^{40} =$$

$$\binom{60}{60} \times \left(\frac{1}{2}\right)^{60} =$$

سوال چهارم: محاسبه انتگرال

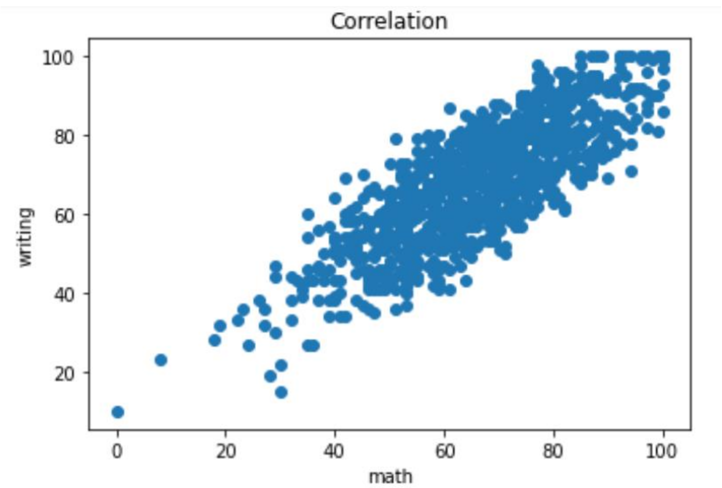
برای x تعدادی نقطه یکنواخت در بازه انتگرال میسازیم که هرچه بیشتر باشد دقت بالاتر است سپس به ازای هر یک مقدار تابع را براساس ضابطه میابیم و طبق فرمول محاسبه میکنیم:

```
integral 0:1 x^3=0.2501048904850936,eror=0.00010489048509360055
integral 0:pi sin(x)=2.0002340222379553,eror=0.0002340222379553225
integral 0:4 1/(1+x^2)=1.3255933170932053,eror=0.0002066829067948195
integral 0:1 sqrt(1-x^4)dx is: 0.8739088413866322
integral 0:1 e^(x^2)dx is: 1.4625781382586627
```

سوال پنجم: کار با داده

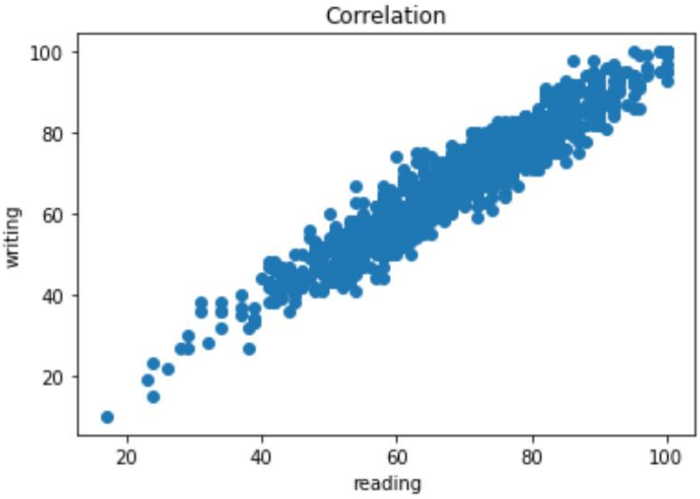
1.آ)

نمره ریاضی با نوشتار نسبت مستقیم دارد و فرض مطرح شده صحیح است



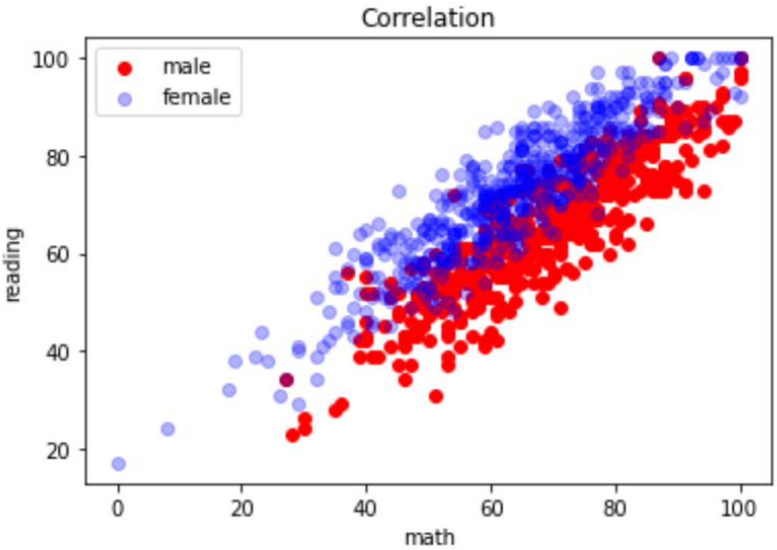
1.ب)

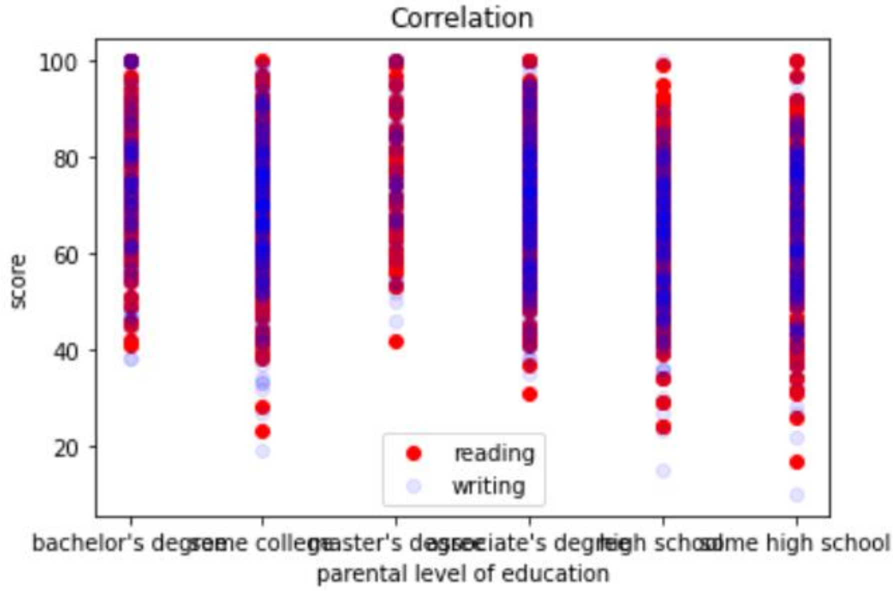
نمره نوشتار با خوانش نسبت مستقیم دارد و
فرض مطرح شده غلط است



1.ج)

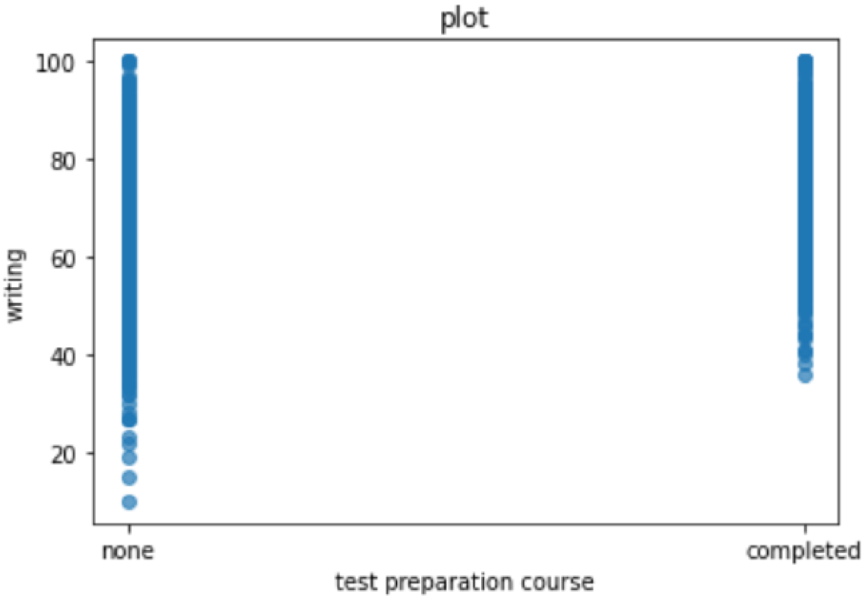
درهر دو گروه نمره خوانش و ریاضی
نسبت مستقیم دارد ولی دختران به طور
میانگین عملکرد بالاتری داشتند.





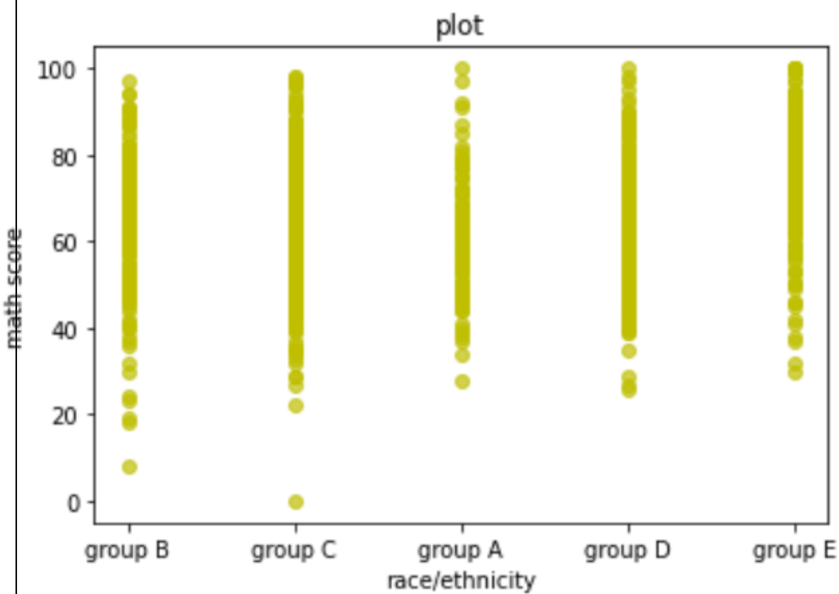
1.د)

به طور شهودی هرچه تحصیلات بیشتر باشد حداقل نمره دانش آموزان بالاتر است،همچنین به طور کلی عملکرد نوشتار بهتر از خوانش است.



.2

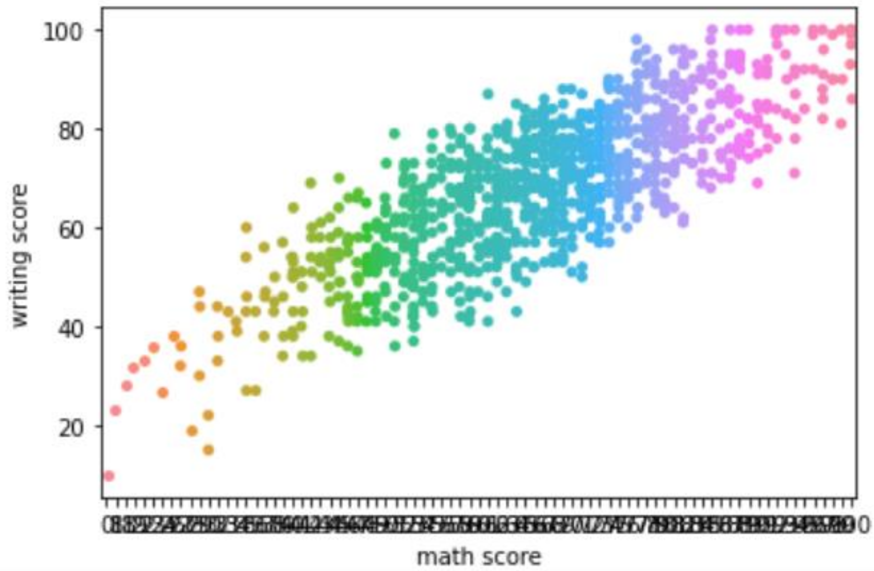
اگر دانش آموزان دوره آمادگی قبل از آزمون را گذرانده باشند عملکرد بهتر و حداقل نمره بالاتری دارند.



فرض:چینش دانش آموزان درگروه ها با توجه به سطح درسی آنها بوده است.

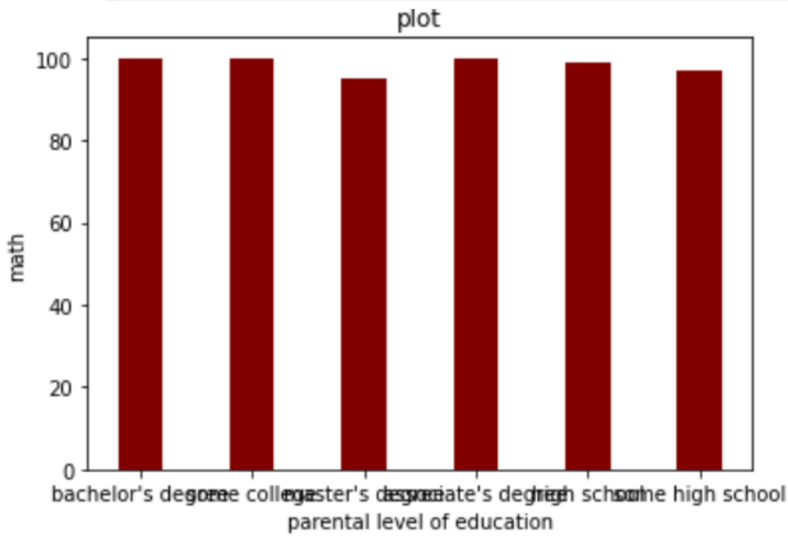
این فرض تقریبا غلط است و توزیع نمرات گروه ها تقریبا یکسان است.

3.در رسم نمودار پراکندگی گاهی نقاط بر روی هم می افتد و دیده نمیشوند برای رفع این مشکل از jitter استفاده میشود که نقاط را کمی جابه جا میکند تا بتوانیم تمامی نقاط را مشاهده کنیم،به این منظور از strip plot کتابخانه seaborn استفاده میکنیم.

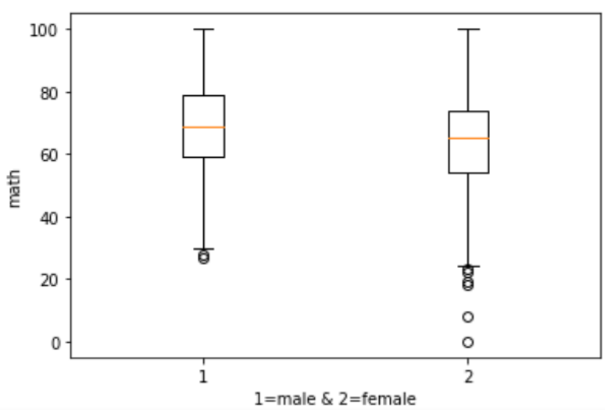


.4

نمودار ستونی ریاضی بر حسب سطح تحصیلات اولیا



نمودار جعبه ای نمره ریاضی دختران و پسران که برای یافتن چارک ها مناسب است.



.5

الف)

ابتدا در نظر بگیریم که یک نمونه 1000 تایی از شرکت کنندگان انتخاب کرده و مشاهده میکینم 40 درصد آن ها در تست موردنظر شرکت کرده اند. آیا می توانیم بگوییم اکثر شرکت کنندگان در تست شرکت کرده اند ؟بله یا خیر؟

$$\hat{P} = 0.4 \rightarrow H_0 = P = 0.5, H_1 = P > 0.5$$

$$\hat{P} \sim N\left(0.5, \sqrt{\frac{\frac{1}{4}}{1000}}\right) = N(0.5,0.015)$$

$$P\{\hat{P} > 0.7 \middle| H_0\} = P\left(Z > \frac{0.4 - 0.5}{0.015}\right) = P\{Z > -2.93\}$$

احتمال بالا زیاد است، درنتیجه فرض صفر رد باقی می ماند.

ب)

حال بیایید در یک نمونه 45 تایی، متوسط نمرۀ خوانش، 65 و واریانس آن 30 است. اگر فرض صفر ما، 68 بودن نمرۀ میانگین باشد :

$$H_0 = Mean = 68, H_1 = Mean > 68$$

$$\bar{X} = Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$$

$$P\{\bar{X} > 65 | H_0\} = P\left(Z > \frac{68 - 65}{\frac{\sqrt{30}}{\sqrt{45}}}\right) = P\{Z > 1.161\} = 0.13$$

با توجه به زیاد بودن احتمال بالا، نمی توان فرض صفر را رد کرد.