دانشکده‌ی علوم ریاضی

آمار و کاربرد ها مدرس: دکتر محمد هادی مستفید

تمرین سری یک

نام و نام‌خانوادگی : کژال باغستانی *شماره دانشجویی:* ۴۰۱۱۰۰۰۷۱

پرسش ۱

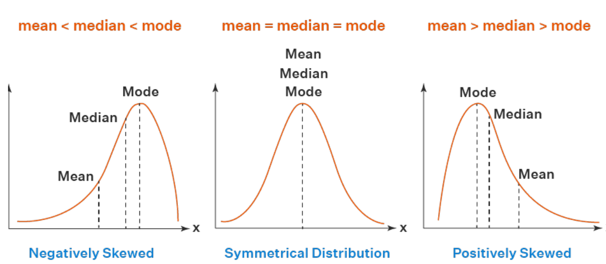
آ) به سادگی مشخص است که این عبارت غلط است زیرا شخصی که از این رستوران سفارش داده است یا از قبل سفارش داده بوده که به وضوح راضی بوده و مشتری رستوران است و نظر او مثبت و در غیر این صورت این شخص از طرف شخصی دیگر که سلیقه های مشابهی دارند معرفی شده یا خودش در نگاه اول از رستوران راضی بوده که این رستوران را برای سفارش انتخاب کرده و در کل تعداد حالات کمی وجود دارد که شخص نظر منفی داشته باشد و این نمونه گیری به وضوح اریب است.

ب) این نمونه گیری نیز به دلایل مختلفی که چند نمونه را ذکر خواهم کرد اریب است. یکی اینکه ریاضی دو درسی پایه و اجباری است. یکی دیگر اینکه درس فلسفه ریاضی فقط مربوط به دانشکده ریاضی و همچین درس اختیاری است در صورتی که ریاضی دو مربوط به کل دانشگاه و اجباری است. یکی دیگر از دلایل این است که برای خیلی از دانشجویان نمره خوب را به سادگی گرفتن ملاک است ولی دکتر شهشهانی استادی سخت گیر است.

پرسش ۲

آ) از لحاظ مفهومی چولگی به معنای میزان دوری داده های پرت است یا به عبارتی علامت آن نشان دهنده راست یا چپ بودن داده های دور تر از میانگین یا مد است.

ب)



شکل بالا به وضوح به ما نشان میدهد که ترتیب میانگین و میانه و مد در چولگی به چپ و راست و توزیع متقارن به چه صورت است.

پ) برابر بودن ضریب اول و دوم پیرسن به ما رابطه زیر را می دهند:

$$\frac{\bar{x} - M}{s} = \frac{3(\bar{x} - m)}{s} = 0.32$$

که با قرار دادن میانگین و انحراف معیار به ترتیب داریم:

$$\frac{29.6 - M}{6.5} = 0.32$$

که به ما می‌دهد:

$$M = 29.6 - (0.32 \times 6.5) = 27.52(\text{مد})$$

و برای معادله دوم:

$$\frac{3(29.6 - m)}{6.5} = 0.32$$

که به ما می‌دهد:

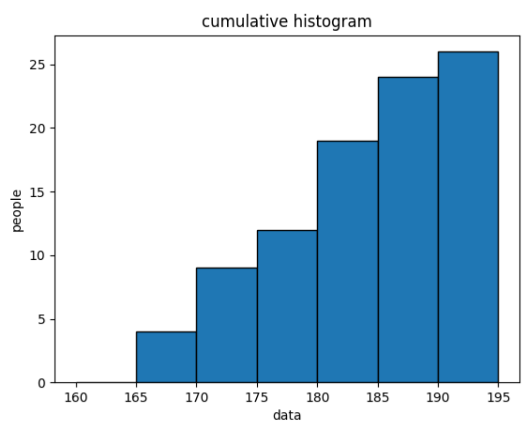
$$29.6 - m = 0.693$$

بنابراین:

$$m \approx 29.6 - 0.693 = 28.907 \approx 28.9$$

پرسش ۳

آ)

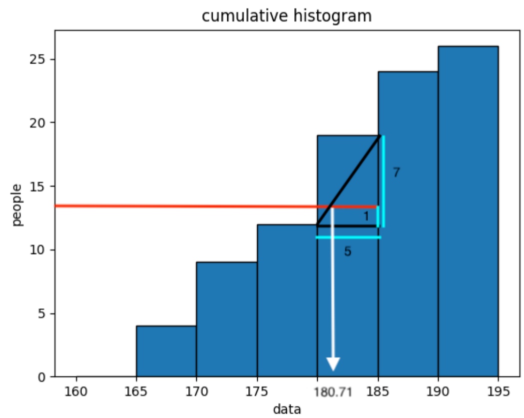


شکل ۱: نمودار توزیع تجمعی رسم شده توسط python

ب) برای به دست اوردن میانگین کافیست مقدار مرکز دسته ها را محاسبه کنیم و برای هر دسته مقدار مرکز دسته را در فراوانی نسبی دسته ضرب کنیم و در نهایت با هم جمع کنیم.

$$\bar{X} = \frac{4 * 167.5 + 5 * 172.5 + 3 * 177.5 + 7 * 182.5 + 5 * 187.5 + 2 * 192.5}{26} = 179.423$$

حالا برای پیدا کردن میانه کافیست با ببینیم در نمودار کدام دسته است که دسته قبل از آن کمتر از 13 و دسته بعد از آن بیشتر از 13 فراوانی تجمعی دارد. که با توجه به شکل دسته 180 تا 185 دسته مورد نظر ماست که ابتدای دسته 12 و انتهای آن 19 است و با قضیه تالس به سادگی مشخص میشود که اگر یک خط از نقطه $(180, 12)$ و $(185, 19)$ رد کنیم در نقطه $180 + \frac{5}{7}$ یعنی تقریبا 180.71میانه ما به دست می آید. که نسبت تالس به صورت زیر است:



شکل ۲: پیدا کردن میانه در نمودار هیستوگرام به کمک روش تالس

پ) اعداد به دست آمده توسط پایتون به صورت زیر هستند که دقیقا با رابطه گفته شده در کلاس یکی هستند.

$$MeanRootfor-1 : 178.201273382513$$

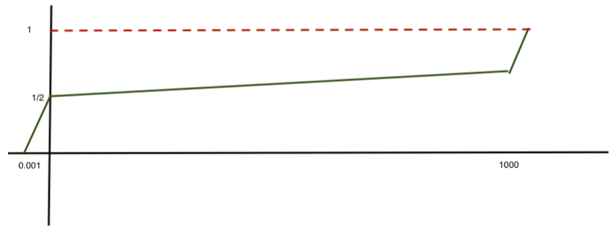
$$GeometricMean : 178.37000921518623$$

$$MeanRootfor1 : 178.53846153846152$$

$$MeanRootfor2 : 178.7064632295094$$

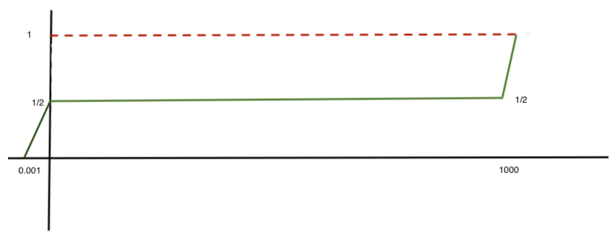
پرسش ۴

آ) کافیست یک مثال نقض از توزیعی بیاورم که میانگین آن (که برای توزیع پیوسته به صورت امید ریاضی تعریف میشود) جایی باشد که تابع توزیع آن برابر با یک دوم نشود. برای مثال در شکل زیر تنها نقطه ای که F(x) برای آن برابر با یک دوم میشود صفر است اما به وضوح امید ریاضی یا همان میانگین این نمودار بزرگ تر از صفر است.



شکل ۳: مثال نقض برای قسمت آ سوال چهار

ب) اگر میانه توزیع را برای حالت پیوسته تعریف کنیم نقطه ای در تابع چگالی که در آن انتگرال قبل و بعد از أن برابر با یک دوم میشود این دقیقا تعریفی هست که در صورت سوال گفته شده و میانه در آن می افتد.
پ) برای این قسمت کافیست یک توزیع نامتقارن ارایه دهیم که هر دوی میانه و میانگین عوض این مجموعه باشند.



در شکل بالا به صورت شهودی مشخص است که میانگین و میانه هر دو در قسمت وسط نمودار می افتند و توزیع دقیقا متقارن نیست.

پرسش ۵

$$e = P(E)$$
$$f = P(F)$$
$$s = P((F \cup E)^c) = 1 - e - f$$

با تعریف احتمالات بالا و فرض مستقل بودن آزمایش‌ها، احتمال رخ دادن E قبل از F برابر است با مجموع تمام حالت‌هایی که مجموعه‌ای ازمایش رخ داده و نه F و نه E رخ داده است و در آخرین آزمایش E رخ داده است. از آنجایی که این پیشامدها اشتراکی ندارند، احتمال اجتماع آنها جمع احتمالاتشان است. بنابراین:

$$P(E \text{ قبل } F) = \sum_{i=0}^{n}(s^i \times e) = e \times \sum_{i=0}^{n}(s^i)$$

که یک سری هندسی است و جمع آن به صورت زیر حساب می‌شود:

$$\lim_{n\to\infty}\left(\frac{(1-(s^n))}{1-s}\right)=\frac{1}{1-1+e+f}$$

در نتیجه داریم:

$$P(E \text{ قبل } F)=\frac{e}{f+e}=\frac{P(E)}{P(E)+P(F)}$$

پرسش ۶

آ) در این حالت هر استراتژی ما در نهایت به یک دنباله از شماره و خال کارت ها تقسیم میشود که وقتی هیچ اطلاعاتی ا ز کارت ها نداریم، دنباله ما صرفا یک جایگشت رندوم از کل کارت ها است. پس برای محاسبه $E[N]$ می توانیم متغیر تصادفی X_i را تعریف کنیم که بر نولی درست بودن کارت i ام را نشان می‌دهد.

$$E[N]=\sum_{i=1}^{52}E[X_i]=52\times\frac{1}{52}=1$$

ب) در این قسمت، در هر استراتژی، نهایت اطلاعات ما در مرحله i ام این است که کارت های ۱ تا $i-1$ چه بوده اند که می‌توانیم از گفتن آنها برای کار ت i ام صرف نظر کنیم. پس دوباره مثل قسمت قبل می‌توانیم X_i ها را تعریف کنیم و این بار داریم:

$$E[N]=\sum_{i=1}^{n}E[X_i]=\sum_{i=1}^{n}\frac{1}{n-i+1}\approx\ln n$$

بله، درست است. می‌توانید از تعریف توابع لاگرانژ برای نقطه x_i و استفاده از خاصیت $L_i(x_i)=1$ برای نشان دادن جمع ضرایب c_i برابر با ۱ استفاده کنید.
بنابراین، با استفاده از تعریف توابع لاگرانژ به صورت زیر:

$$L_i(x)=\prod_{j\neq i}\frac{x-x_j}{x_i-x_j}$$

و خاصیت $L_i(x_i)=1$، می‌توان نشان داد که جمع ضرایب c_i برابر با ۱ است.