



به نام خدا



دانشگاه تهران
دانشکده مهندسی برق و کامپیوتر
آمار و احتمال مهندسی

گزارش پروژه‌ی نهائی

نام و نام خانوادگی	سید محمد امین اطمیابی
شماره دانشجویی	
تاریخ ارسال گزارش	۱۴۰۰/۴/۴

فهرست گزارش سوالات

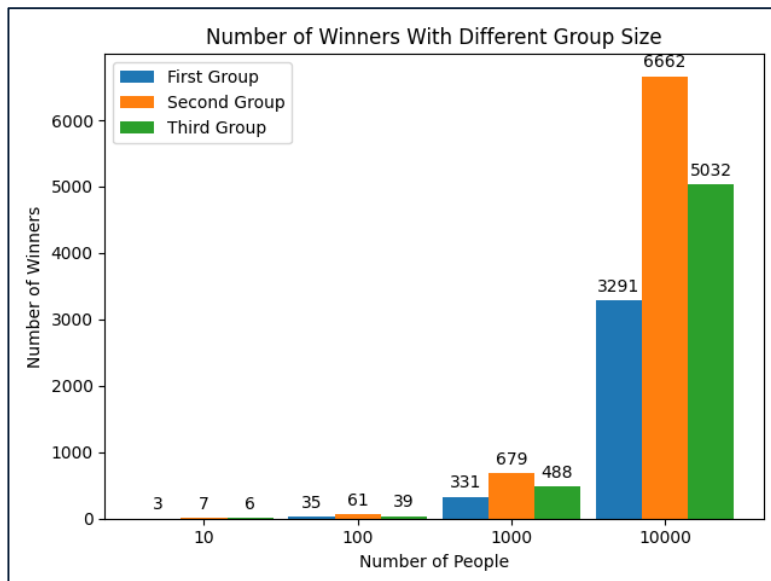
- سوال اول: مسابقه تلویزیونی ۳
- سوال دوم: تخمین عدد پی ۴
- سوال سوم: پارادوکس سنت پترزبورگ ۶
- سوال چهارم: داده بازی ۷

سوال اول: مسابقه تلویزیونی

در نگاه اول بعد از پوچ شدن یکی از گزینه‌ها توسط مجری مسابقه شانس برنده شدن ما $\frac{1}{3}$ است ، در نتیجه عوض کردن یا نکردن گزینه انتخابی تفاوتی در نتیجه نهایی نخواهد داشت . پس در ابتدا تصمیم به عوض کردن گزینه انتخابی نخواهم کرد . اما با شبیه سازی کردن این آزمایش در مقیاس‌های متفاوت نتایج جالبی بدست می‌آید . همانطور که نمودار ۱ مشخص است ، تعداد برندگان گروه دوم و سوم نسبت به برندگان گروه اول بسیار بیشتر است .

با بررسی دقیق تر در نتایج آزمایش در میابیم افراد گروه اول برای برنده شدن شانس معادل با $\frac{1}{3}$ دارند (انتخاب ۱ درب از ۳ درب موجود) . افراد گروه دوم با توجه به اینکه بعد از پوچ اعلام شدن یکی از گزینه ها انتخاب خود را عوض می‌کنند ، برای برنده شدن شانس معادل با $\frac{1}{3}$ دارند (انتخاب ۱ درب از ۲ درب موجود) . در نهایت افراد گروه سوم را داریم که با احتمال $\frac{1}{3}$ گزینه انتخابی خود را عوض می‌کنند که به بیانی دیگر ترکیبی از افراد دو گروه قبلی هستند .

طبق نتایج بدست آمده مشخص است شانس برنده شدن افراد گروه دوم و سوم که انتخاب خود را عوض می‌کنند از افرادی که انتخاب خود را عوض نمی‌کنند بیشتر است . حال می‌توان گفت انتخاب اولیه ما مبنی بر تغییر ندادن گزینه انتخابی می‌تواند بر ضرر ما باشد و شانس برنده شدن ما را تا نصف کاهش دهد.



نمودار ۱ (شبیه سازی سوال مسابقه تلویزیونی)

سوال دوم: تخمین عدد پی

۱.

$$\begin{aligned}\int_0^1 \sqrt{1-x^2} dx &= \int_0^1 \sqrt{1-\sin^2(t)} \cos(t) dt = \int_0^1 \sqrt{\cos^2(t)} \cos(t) dt = \int_0^1 \cos(t)^2 dt \\ &= \frac{1}{2} \int_0^1 1 + \cos(2t) dt = \left. \frac{\arcsin(x) + x\sqrt{1-x^2}}{2} \right|_0^1 = \frac{\pi}{4}\end{aligned}$$

۲.

$$f_{avg} = \frac{1}{1-0} \times \frac{\pi}{4} = \frac{\pi}{4}$$

۳.

$$\pi = 4 \times f_{avg}$$

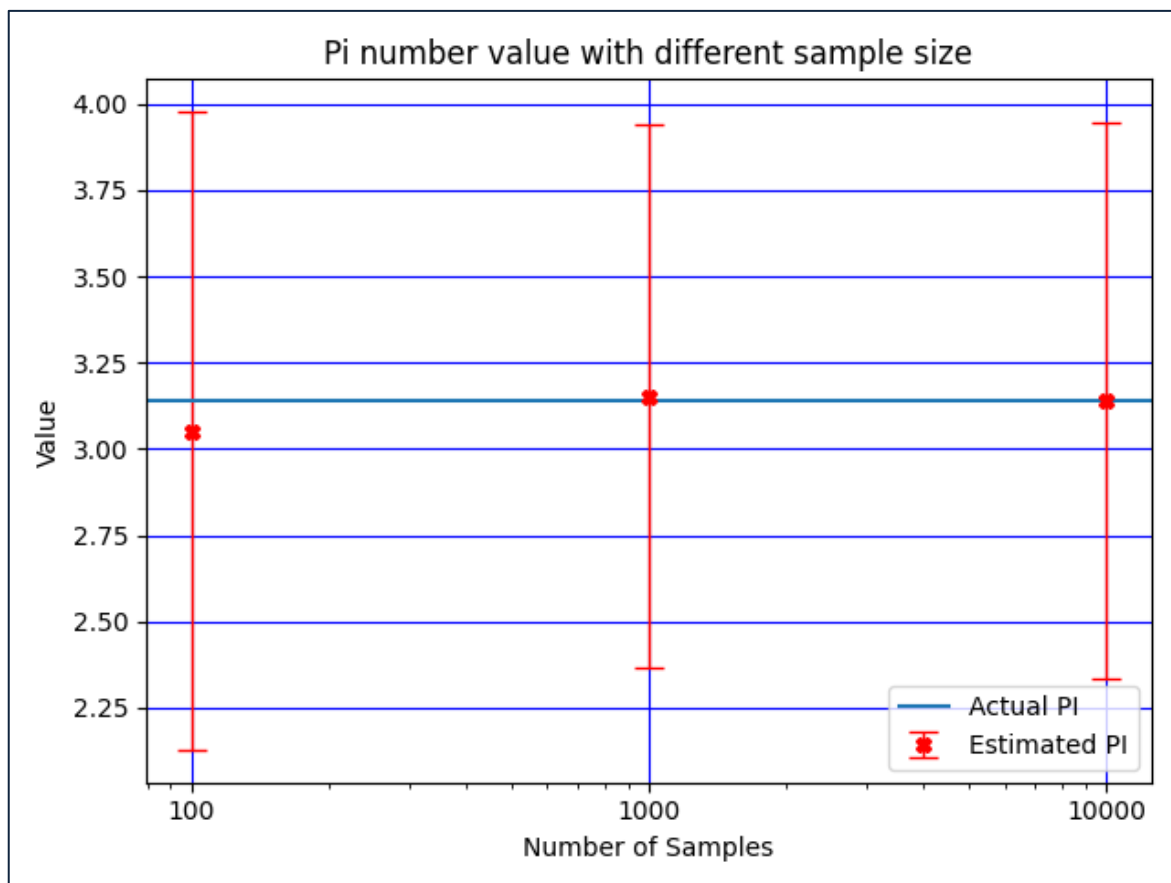
۴. در اجرای اول مقدار 3.072248324080613 و در اجرای دوم مقدار 3.2396749730657364 بدست آمد. این دو عدد اختلاف نسبت زیادی با هم دارند. بدلیل کم بودن تعداد نمونه‌هایی انتخابی در بازه مورد نظر، در هر بار اجرای شبیه سازی، مقدار متفاوتی بدست می‌آید و باعث می‌شود نتایج بدست آمده در هر بار اجرا متفاوت از دفعه قبلی باشد. البته این تفاوت در شبیه سازی با تعداد نمونه‌های بیشتر نیز مشاهده می‌شود، اما اختلاف مقادیر بدست آمده بسیار کمتر است.

۵. در این بخش شبیه سازی با نمونه‌های بیشتری انجام شده و به ازای تعداد نمونه‌ها ۱۰۰۰ و ۱۰۰۰۰ به ترتیب مقادیر 3.148781585275666 و 3.1453345044479337 بدست آمد.

۶. همانطور که در نمودار ۲ مشاهده می‌شود، هرچه تعداد نمونه‌گیری‌ها افزایش می‌یابد، میانگین داده‌ها به مقدار واقعی عدد پی نزدیکتر می‌شوند.

۷. از آنجایی که داده‌های ما به طور یکنواخت توزع شده‌اند، با زیاد شدن تعداد نمونه‌ها، پراکندگی آن‌ها حول میانگین فرقی نخواهد کرد، لذا توقع واریانس‌های مشابه هم داریم.

۸. به طور کلی واریانس بیان کننده میزان پراکندگی داده‌ها حول میانگین هستند و در این سوال نیز این ویژگی مشاهده شد.



نمودار ۲ (تخمین عدد پی)

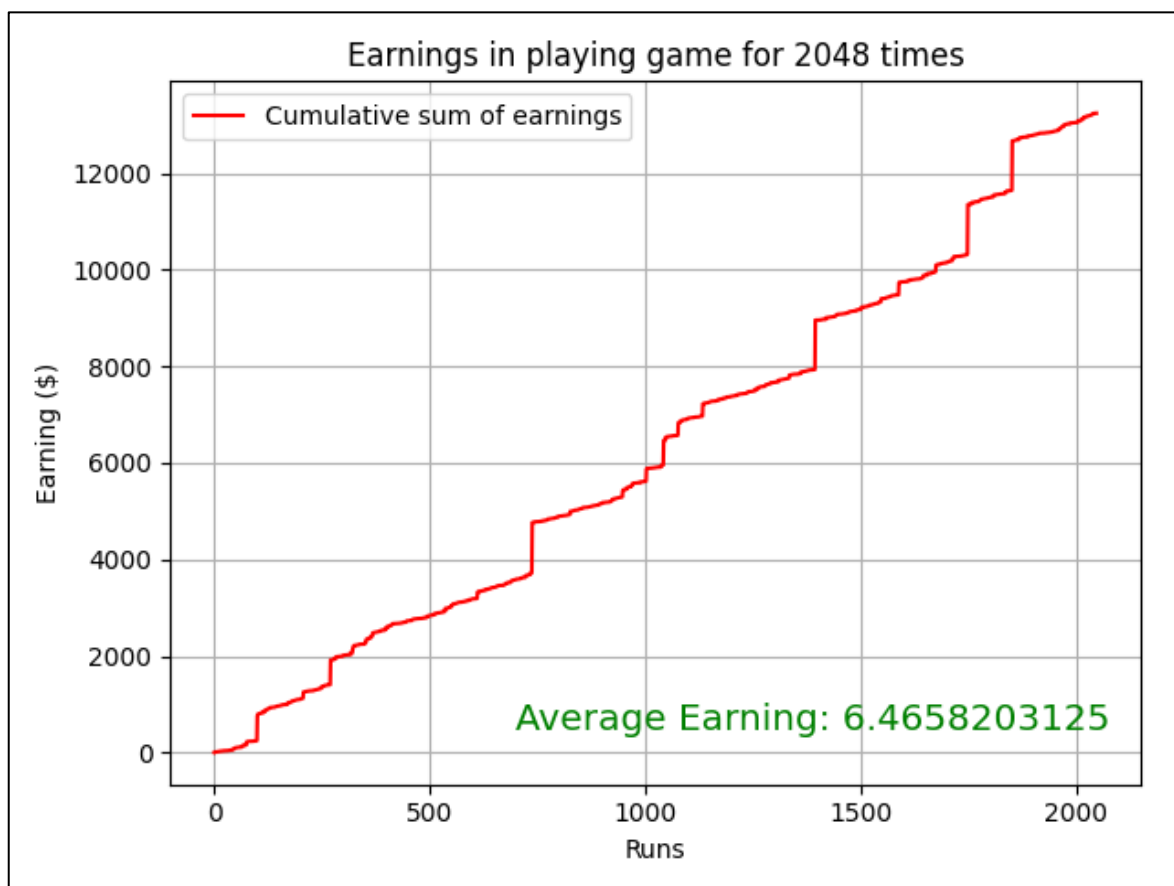
سوال سوم: پارادوکس سنت پترزبورگ

۱.

$$E = \frac{1}{2} \cdot 2 + \frac{1}{4} \cdot 4 + \frac{1}{8} \cdot 8 + \dots = 1 + 1 + 1 + \dots = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \cdot 2^n = \infty$$

همانطور که مشاهده می‌کنیم ، با میل کردن تعداد دفعات بازی به سمت بی‌نهایت ، مقدار پولی که ما برنده می‌شویم نیز به سمت بی‌نهایت میل می‌کند .

۲. با ۲۰۴۸ بار انجام این بازی و بدست آوردن مقداری پولی که در هر بار برنده می‌شویم می‌توان گفت در صورتی که مبلغ ورود به بازی حدود ۶.۵ دلار باشد ، بازی برای ما منصفانه خواهد بود. نتایج شبیه سازی این بخش را می‌توانید در [نمودار ۳](#) مشاهده کنید .



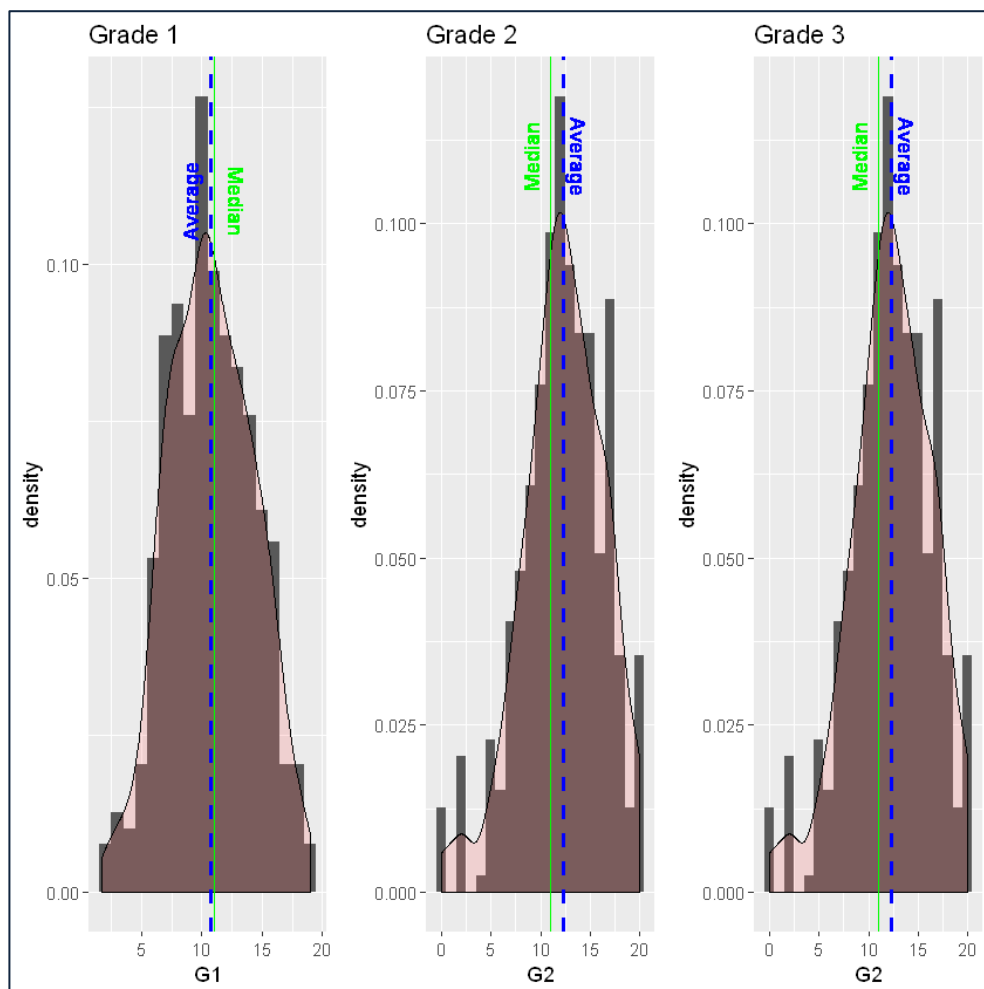
نمودار ۳ (شبیه سازی راه حل بوفون سوال پارادوکس سنت پترزبورگ)

سوال چهارم: داده بازی

۱. آ) این دیتاست دارای ۱۵ داده به ازای هر ورودی است. این داده ها که شامل اطلاعات کیفی و کمی در مورد دانش آموز و خانواده وی همچنین وضعیت او در مدرسه مانند تعداد غیبت ها، زمان درس خواندن، نمرات و ... اوست.

ب) به طور کلی داده های عددی دارای اهمیت بیشتری هستند چراکه نمی توان روی داده های کیفی تحلیل های دقیقی انجام داد. نگاشت ویژگی های کیفی به مقادیر عددی بسیار کمک کننده است. در دیتاست موجود پارامتر نمرات دانش آموز ($G1, G2, G3$) از اهمیت بالایی برخوردار هستند چراکه بهترین پارامتر توصیف کننده برای عملکرد وی از نظر درسی می باشند.

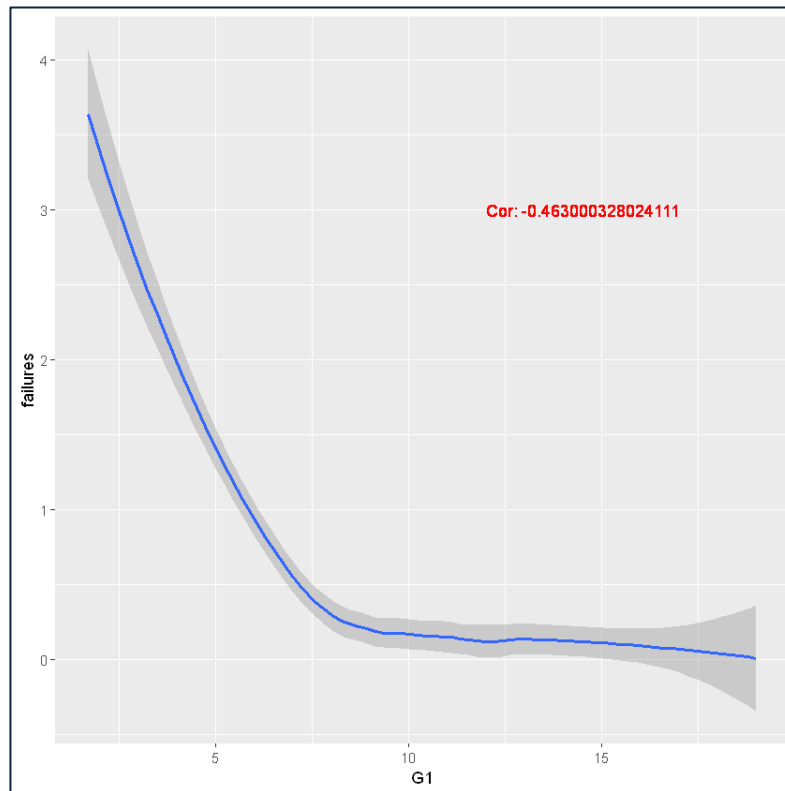
هیستوگرام نمرات $G1$ ، $G2$ و $G3$ دانش آموزان در نمودار ۴ قابل مشاهده است. در این نمودار میانگین و میانه روی شکل مشخص شده است.



نمودار ۴ (هیستوگرام نمرات دانش آموزان سوال ۴)

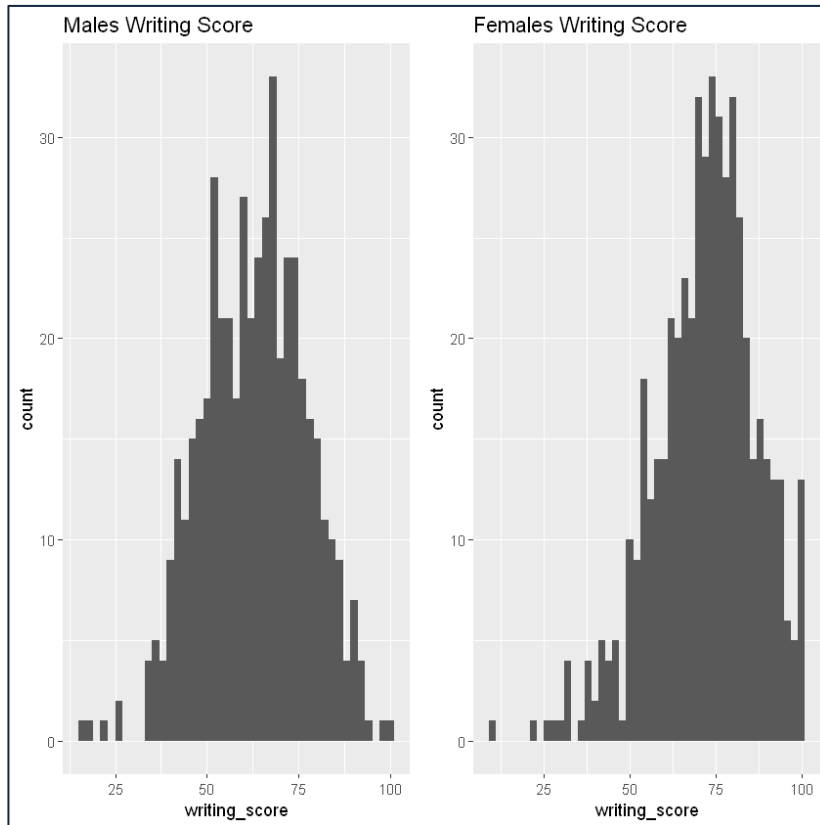
آ) پارامتر نمره G1 و دروس افتاده با هم در ارتباط هستند . ارتباط این دو کمیت بدین صورت است که هرچه تعداد دروس افتاده بیشتر باشد ، نمره دانش آموز کمتر است .

د) همانطور که در **نمودار ۵** مشاهده می کنید رابطه میان این دو داده به درستی بیان شده بود . همچنین منفی بودن مقدار ضریب همبستگی دلالت بر عکس بودن رابطه این دو داده دارد .

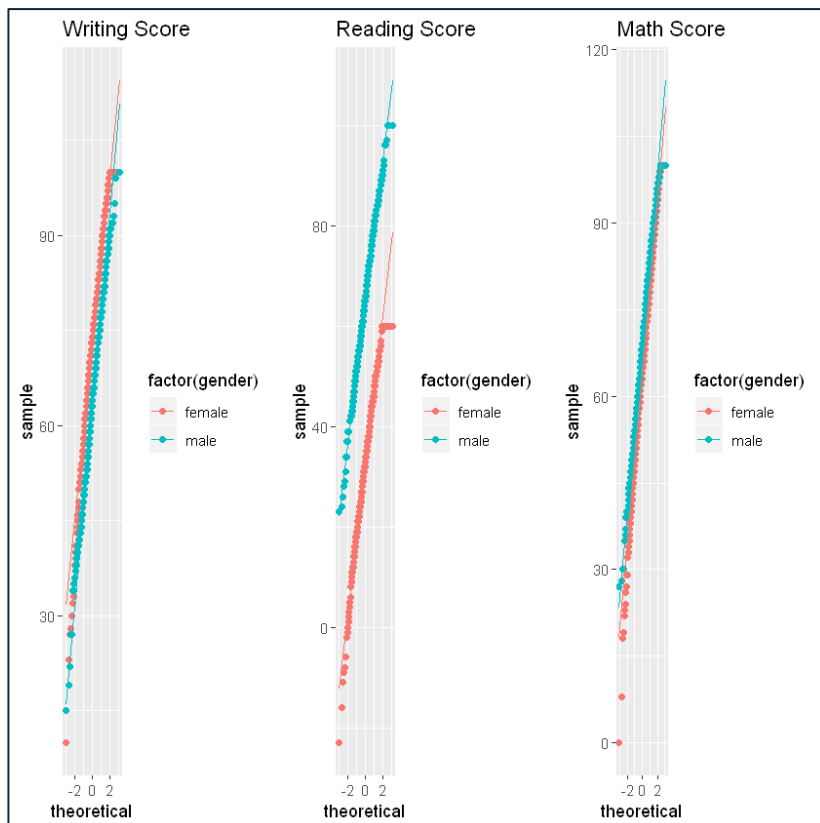


نمودار ۵ (پراکندگی دروس افتاده دانش آموز حول نمره G1)

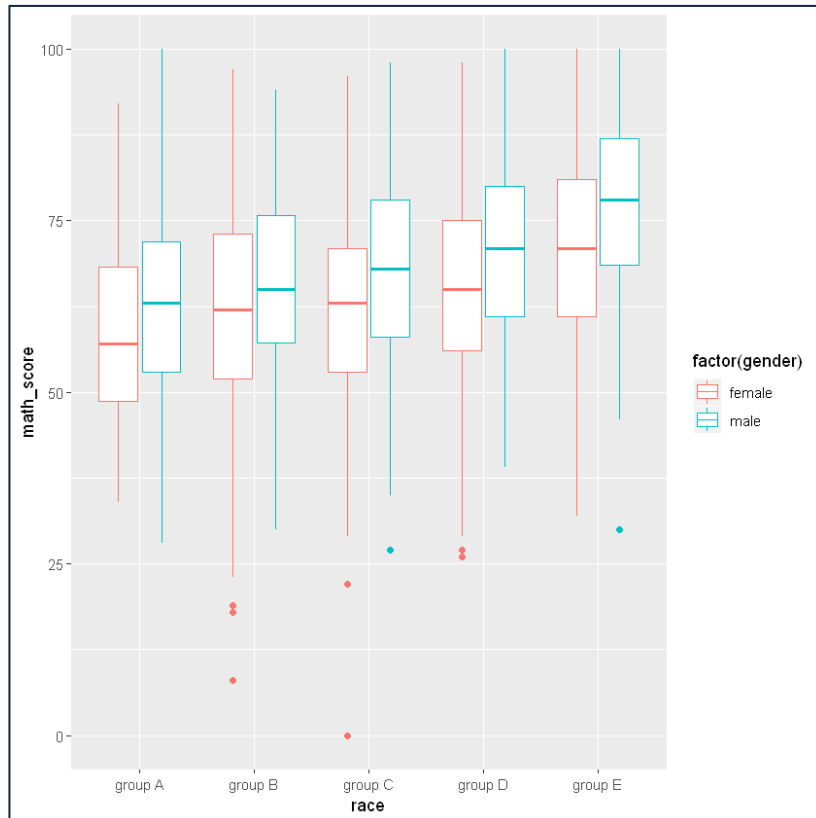
۲. آ) هیستوگرام نمرات دانش آموزان در **نمودار ۶** قابل مشاهده است .
- ب) بله . مطابق **نمودار ۷** می توان گفت داده های ما توزیعش مشابه توزی گوسی دارند . چراکه داده ها در دو طرف مرکز و به یک اندازه پخش شده اند .
- ج) نمودار جعبه ای این بخش در **نمودار ۸** قابل مشاهده است .
- د) همانطور که در **نمودار ۹** مشخص است ، دانش آموزانی که اولیا آنها دارای مدرک کارشناسی ارشد و کارشناسی هستند از میانگین نمرات بالاتری برخوردارند .



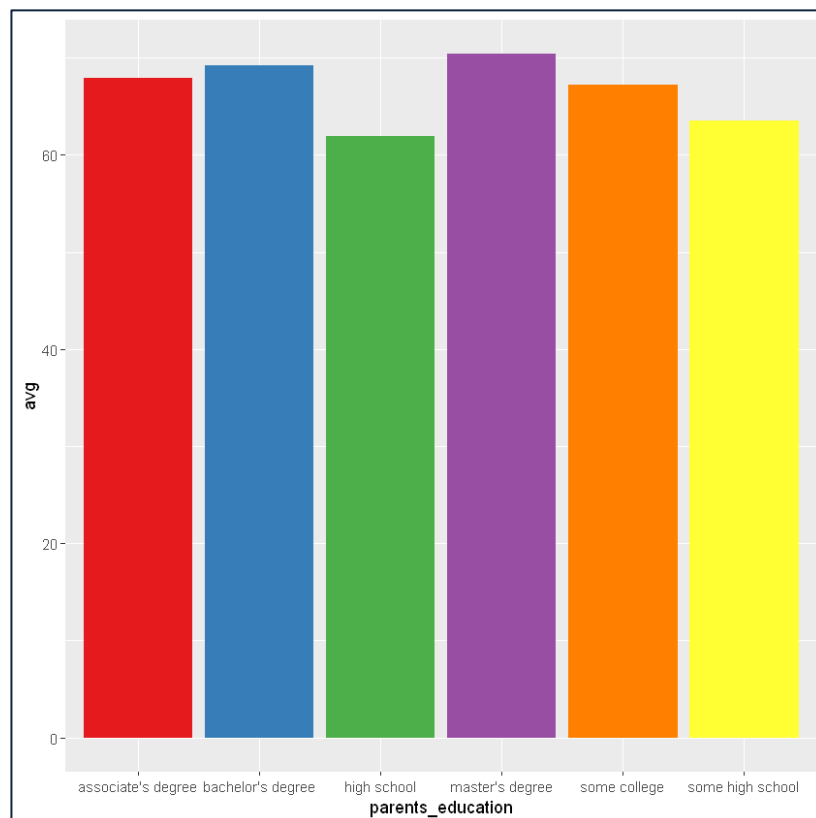
نمودار ۶ (هیستوگرام نمرات **Writing** دانش‌آموزان)



نمودار ۷ (نمودار چارک نمرات دانش‌آموزان)



نمودار ۸ (نمودار جعبه ای نمرات ریاضی دانش آموزان)



نمودار ۹ (میانگین نمرات دانش آموزان بر اساس مدرک تحصیلی اولیا آنها)