به نام خدا

دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

**درس استنتاج آماری**

**آزمایشگاه (بخش )**

**محمد بزم**

**810196126**

اردیبهشت ماه 1397

**فهرست**

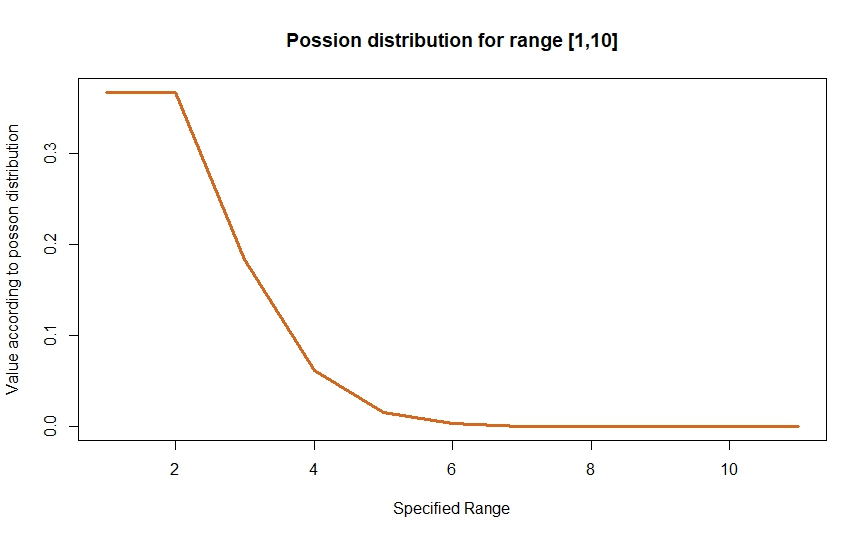
|  |  |
| --- | --- |
| عنوان | شماره صفحه |
| چکیده | ۳ |
| تمرین 1 | 4 |
| تمرین 3 | 6 |
| تمرین 4 (تمرین عملی) | 8 |
| پیوست 1 : روند اجرای برنامه | 11 |

# سوال 1 :

نمونه‌هایی از پدیده‌های طبیعی که از توزیع پوآسون پیروی می‌کنند:

1. نسبت سلول‌هایی که با یک ویروس آلوده می‌شوند.
2. تعداد دفعات وقوع جهش روی یک رشته ی ثابت از مولکول DNA.
3. هدف‌گیری بمب‌ها قبل از رها شدن از بمب افکن.

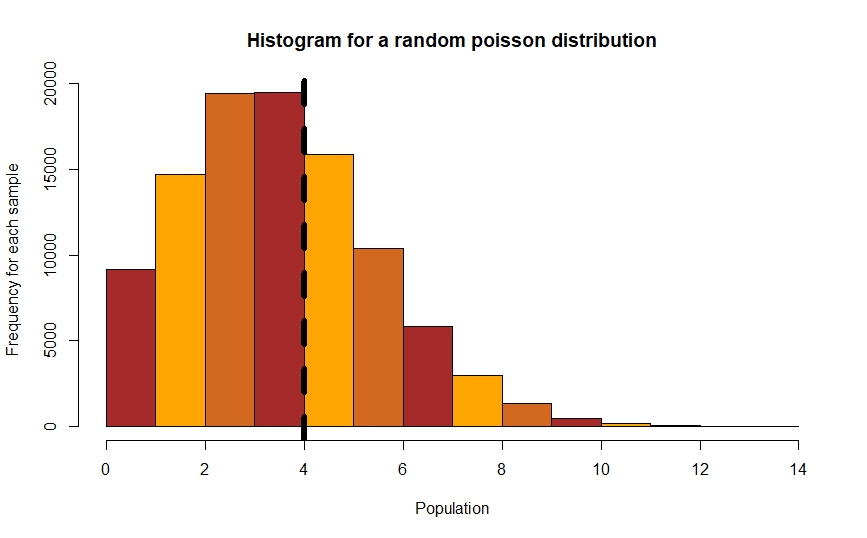
# سوال 2:



نمودار 1 : نمودار توزیع پوآسون برای بازه‌ی 1 تا 10

با افزایش عدد عدد لامبدا، قله‌ی نمودار به سمت مرکز حرکت می‌کند و درست وقتی در وسط (عدد 5) قرار می‌گیرد، نمودار مانند نمودار توزیع نرمال بدون چولگی می‌شود.

# سوال 3:



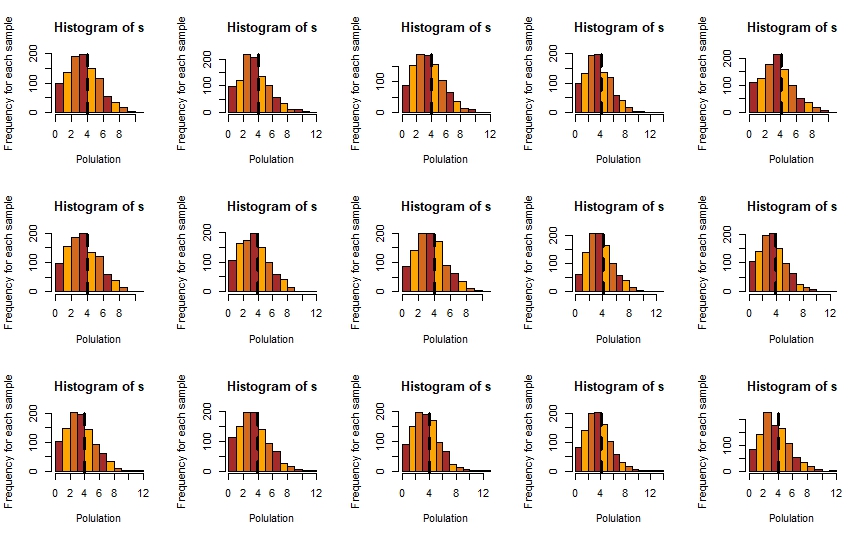
شکل 2: نمودار هیستوگرام مربوط به توزیع تصادفی خواسته شده.

این نمودار با مجموعه دستورات زیر رسم شده است.



دستور اول نمودار هیستوگرام را رسم می‌کند، دستور دوم پارامتر‌های جدید برای رسم خط مطابق با مقدار میانگین را مشخص می‌کند و دستور سوم خط معادل با این نقطه (x=3.99) را رسم می‌کند.

# سوال 4:

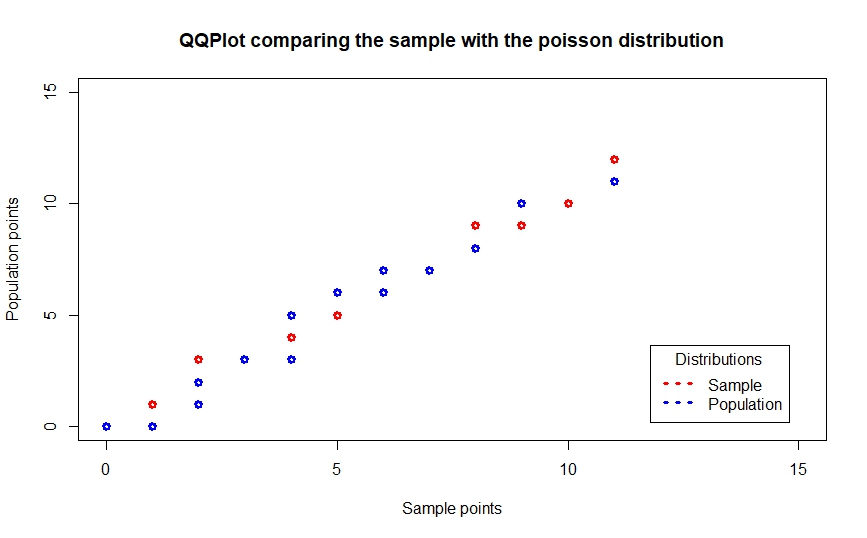


شکل 3: نمودار مربوط به 15 نوبت تکرار آزمایش یاد شده



کد مربوط به این سوال

تابع اول شامل مراحل نمونه‌برداری، رسم نمودار هیستوگرام و در نهایت رسم خط عمودی در محل میانگین است که به کمک دستور abine انجام می‌شود. تابع دوم تابع اول را 15 با تکرار می‌کند و در پایان تعداد قسمت‌های صفحه‌ی نمایش را از 15 قسمت به 1 قسمت (حالت پیش‌فرض) برمی‌گرداند.

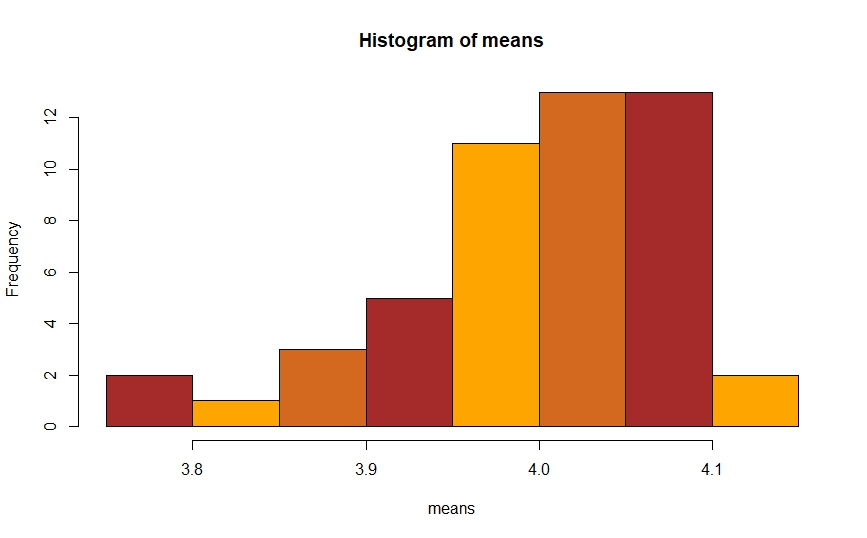


شکل 4: نمودار qqplot مقایسه‌ی نمونه با توزیع اصلی.

مقایسه‌ی نمونه‌ی برداشته شده از جامعه با توزیع پوآسون جامعه در شکل بالا که توسط دستورات زیر انجام شده است نشان می‌دهد دو توزیع بسیار به یکدیگر نزدیک است.

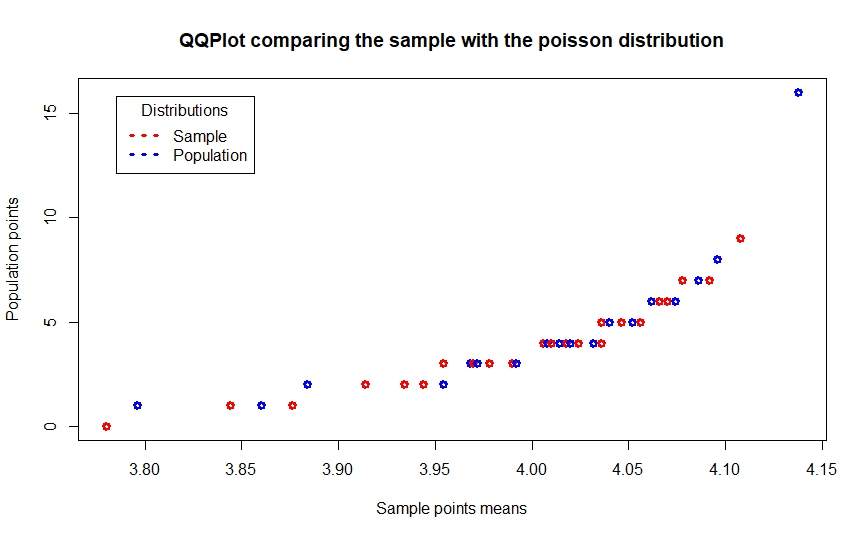


# سوال 5:

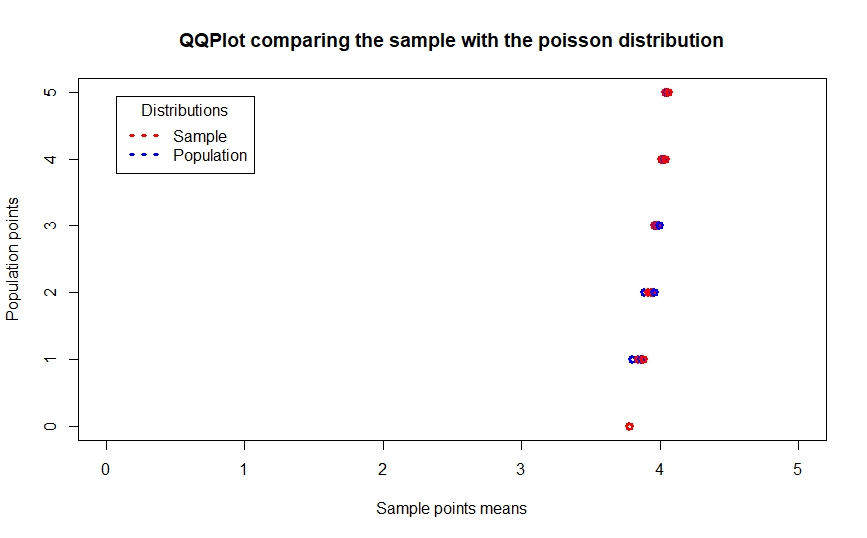


شکل 5 : نمودار هیستوگرام میانگین‌های نمونه‌های ساخته شده از جمعیت اولیه.

نمودار هیستوگرام میانگین نمونه مانند شکل بالا می‌باشد.

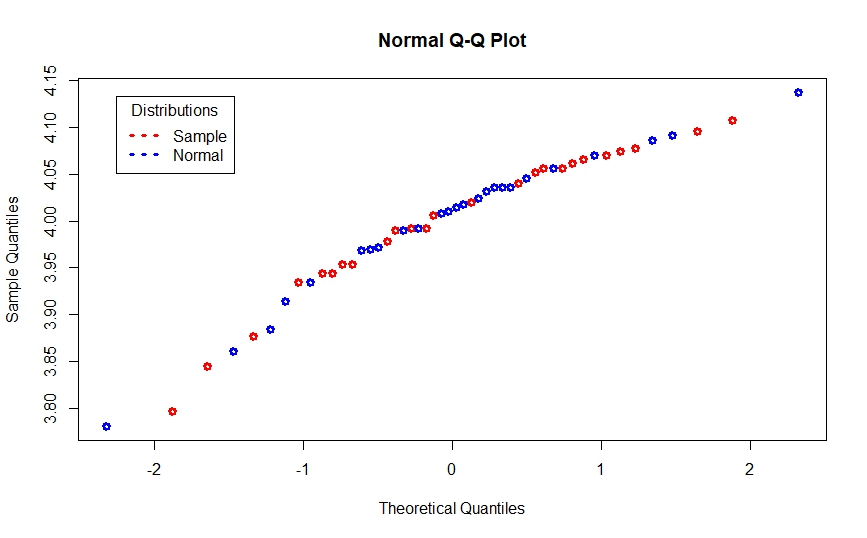


شکل 6 : مقایسه‌ی میانگین‌های نمونه‌ها با توزیع جامعه با محورهای ناهمسان.

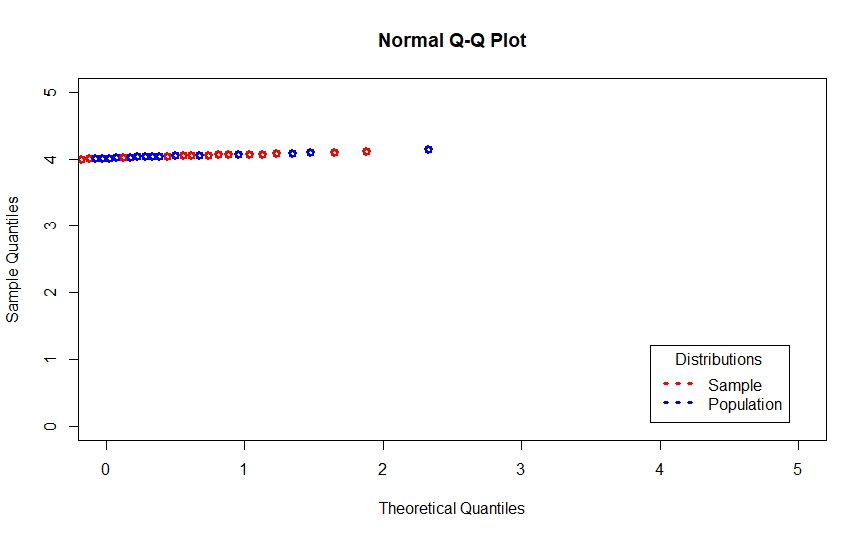


شکل 7 : مقایسه‌ی میانگین‌های نمونه‌ها با توزیع جامعه با محورهای همسان.

مقایسه‌ی توزیع میانگین‌ها و توزیع جامعه نشان می‌دهد در نمایشی که محورها مقیاس همسان ندارند، اگر چه تطابق توزیع‌ها منطبق بر یکدیگر نیست اما تا حدود زیادی نزدیک می‌باشند. اما این مسئله در مورد نموداری که مقیاس محورهای آن همسان است برقرار نیست.



شکل 8 : مقایسه‌ی توزیع میانگین نمونه‌ها با توزیع نرمال با محورهای ناهمسان.



شکل 9 : مقایسه‌ی توزیع میانگین نمونه‌ها با توزیع نرمال با محورهای همسان.

گزاره‌های عنوان شده در مورد مقایسه‌ی توزیع میانگین‌ها با توزیع نرمال نتایجی مانند مقایسه‌ی آنها با توزیع جامعه داشت. از این مقایسه نتیجه ‌می‌شود توزیع میانگین‌ها از توزیع نرمال پیروی نمی‌کند که این خلاف گزاره‌های ثابت شده است. به همین دلیل برای ارائه‌ی گزاره‌ی دقیق لازم است از آزمون فرض استفاده شود. همچنین کد استفاده شده برای این بخش در ادامه آمده است.

بر اساس قضیه‌ی حد مرکزی میانگین‌های تعدادی نمونه از جامعه، خود از توزیع همان جامعه پیروی می‌کنند. برخلاف آنچه انتظار می‌رفت، همانطور که نمودار qqplot نشان می‌دهد، توزیع میانگین‌ها از توزیع جامعه فاصله دارد و منطبق بر آن نمی‌باشد.

# سوال 6:

اگر اندازه‌های نمونه‌های برداشته شده از جامعه به اندازه‌ی کافی بزرگ باشد (مشاهدات تجربی نشان می‌دهند اندازه‌های نمونه‌ها باید بزرگتر از 30 باشد تا ویژگی‌ای که در ادامه بیان می‌شود برقرار باشد )، توزیع میانگین‌های این نمونه‌ها از توزیع نرمال پیروی خواهد کرد. لازم به ذکر است این گزاره مستقل از توزیع جامعه می‌باشد، به این معنی که توزیع جامعه هر چه باشد، نمونه‌های به اندازه‌ی کافی بزرگ میانگین‌هایی بدست می‌دهند که از توزیع نرمال پیروی می‌کند.

# سوال 7 :

کد مربوط به این بخش بصورت زیر است:



# سوال 8 :

این بخش با دستور زیر قابل انجام است.



# سوال 9 :

در 9662/. (97%) موارد میانگین جمعیت درون بازه‌ی اطمینان بدست آمده قرار گرفته است. نتیجه‌ی بدست آمده به طور کامل منطبق بر تعریف بازه‌ی اطمینان می‌باشد. طبق اگر n بار از جامعه نمونه برداری کنیم و برای این نمونه‌ها بازه‌ی اطمینان x% بسازیم، در x درصد از مشاهدات انجام شده (x%) از کل نمونه‌های اخذ شده، میانگین جامعه درون بازه‌ی اطمینان قرار می‌گیرد. به منظور تاکید یادآور می‌شود همین نتیجه در آزمایش انجام شده مشاهده شد.

این آزمایش توسط کد زیر انجام شده است.



# سوال 10:

در آزمایش انجام شده اندازه‌ی نمونه‌ها کوچکتر از 30 می‌باشد. به این دلیل توزیع آماره‌ی مورد بررسی برای نمونه‌های بدست آمده از توزیع نرمال پیروی نمی‌کنند و نمی‌توان از نتایج قضیه‌ی حد مرکزی برای آنها استفاده کرد. در نهایت میانگین تعداد آزمایشاتی که در آنها عدد میانگین جامعه درون بازه‌ی اطمینان مورد نظر قرار گرفته برابر با 5169/. (52%) می‌باشد. در صورت برقرار بودن شرایط قضیه انتظار می‌رفت میانگین برابر با پهنای بازه‌ی اطمینان (90%) باشد. مشاهده‌ی بالا نشان می‌دهد در صورت عدم برقراری شرایط قضیه، نمی‌توان انتظار نتیجه‌ی درست داشت.

این آزمایش با کد زیر پیاده‌سازی شده است.

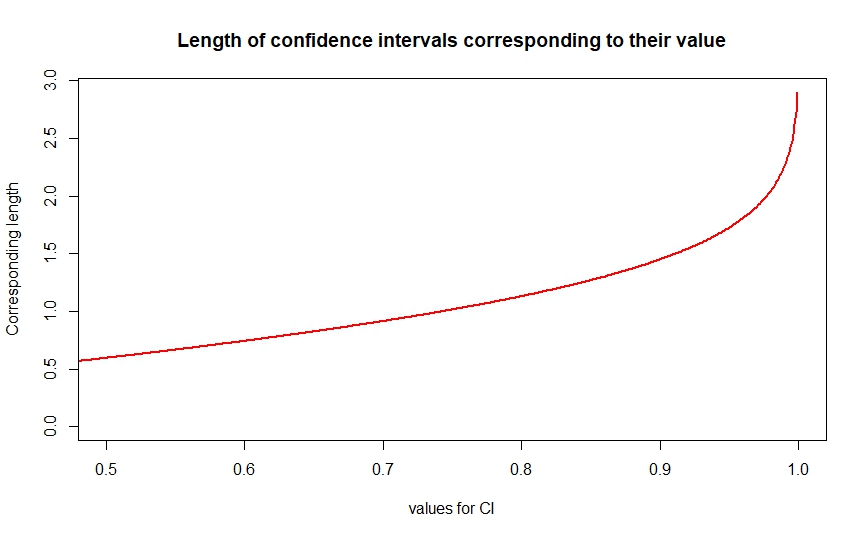


# سوال 11:

پیاده سازی مورد نظر توسط کد زیر انجام شده است.



# سوال 12:



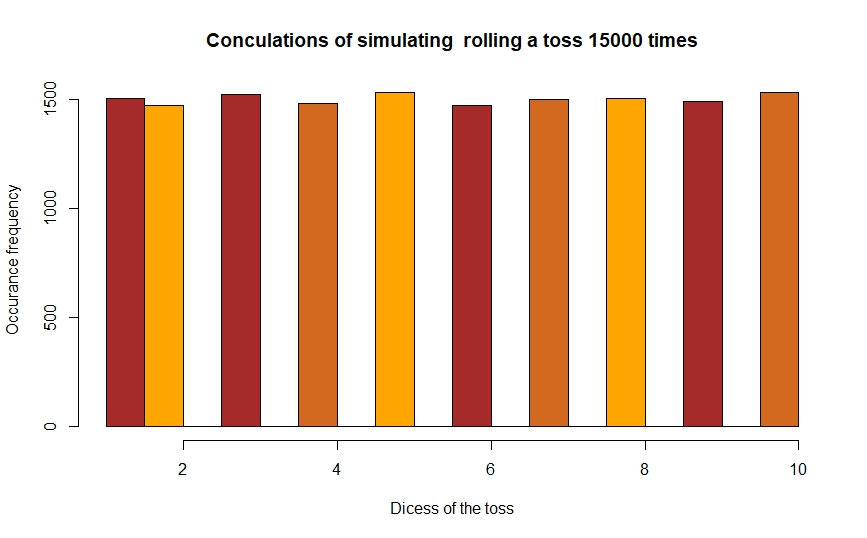
شکل 10: نمودار پهنای بازه‌ی اطمینان بازای مقادیر مختلف بازه

همانطور که مشاهده می‌شود با افزایش عدد بازه‌ی اطمینان، پهنای آن نیز افزایش میابد. کد مربوط به این بخش به صورت زیر است:



# سوال 13:

تعداد دفعات پیشامد هر وجه از تاس ده وجهی در 15000 پرتاب توزیعی یکنواخت مانند شکل زیر داشته است.

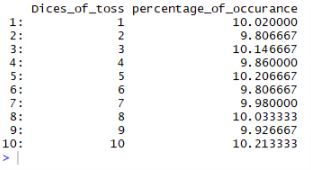


شکل 11: توزیع پیشامد‌های مختلف در 15000 با شبیه‌سازی پرتاب تاس.

کد مربوط به این سوال به صورت زیر است:



همچنین خروجی table بصورت زیر است:

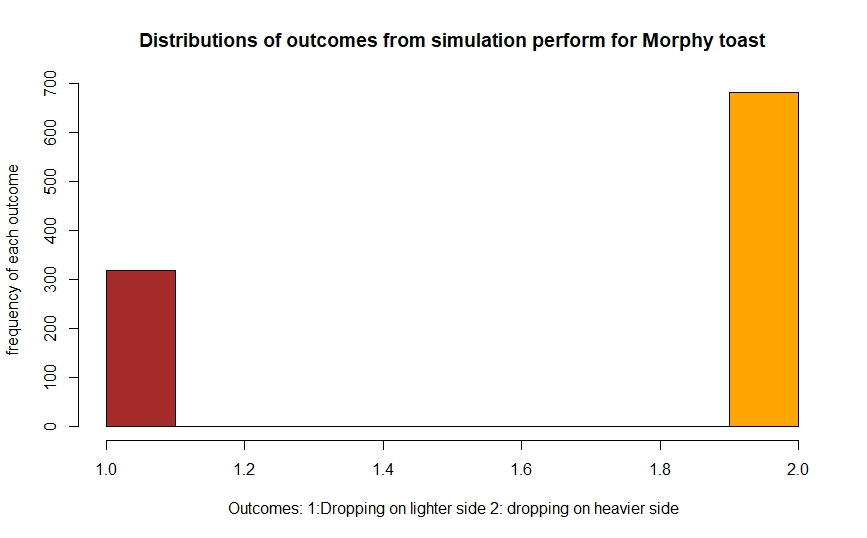


# سوال 14:

اگر نان توست را از پهنا برش دهیم، به دو قسمت تقسیم خواهد شد. هر نیمه 10 گرم وزن دارد. در یکی از این برش‌ها 12 گرم هم مربا و کره وجود دارد که وزن آن را در مجوع به 22 گرم می‌رساند. در صورتیکه احتمال فرودآمدن نان روی هر قسمت متناسب با وزن آن قسمت باشد، احتمال آمدن یک سمت  و روی دیگر برابر با  می‌باشد.

شبیه‌سازی مورد نظر بر اساس احتمالات یاده شده و با استفاده از کد زیر انجام شده است و نتایج آن در ادامه آمده است.





شکل 12 : توزیع پیشامد‌های مختلف در شبیه‌سازی انجام شده. لازم به ذکر است در اینجا عدد 1 معادل با فرود آمدن نان روی قسمت سبکتر و عدد 2 معادل با فرود آمدن روی سطح سنگین‌تر می‌باشد.

# سوال 15:

از بین 36 حالت ممکن برای پرتاب همزمان دو تاس، 15 حالت است که در آنها مجموع اعداد بر روی تاس‌ها بزرگتر از 8 می‌باشد. در نتیجه احتمال چنین پیشامده برابر با  می‌باشد. شبیه‌سازی انجام شده که کد آن در ادامه آمده است، احتمالی برابر 0.25 برای این پیشامد را نشان می‌دهد که با محاسبه‌ی انجام شده سازگار نیست.



# سوال 16:

بررسی شرایط:

1. استقلال: چون نمونه‌ها مربوط به روز‌های مختلف هستند، مستقل از یکدیگرند. همچنین لازم است در زمان انتخاب نمونه تصادفی بودن انتخاب روز‌ها رعایت شده باشد.
2. نرمال بودن و چولگی: تعداد نمونه‌ها بیشتر از 30 است چولگی قابل ملاحظه‌ای وجود ندارد.





چون عدد بدست آمده از سطح اطمینان 5% کوچکتر است، پس فرض اولیه رد می‌شود. به این ترتیب نتیجه می‌شود این داده‌ها، موید گزاره‌ی اعلام شده از جانب دانشمندان نمی‌باشد.

# سوال 17:

با توجه به نزدیک بودن میانگین واقعی و میانگین گزارش شده، رد کردن مقدار گزارش شده بدون در نظر گرفتن نتایج آزمون انجام شده دشوار است. علت این است که p-value تنها می‌گوید یک گزاره رد می‌شود یا نمی‌شود اما در مورد اینکه با چه اختلافی از مقدار واقعی این اتفاق می‌افتد چیزی نمی‌گوید.

# سوال 18:

بررسی شرایط:

1. استقلال: نمونه‌ها باید به تصادف انتخاب شده باشند و نیز به تصادف در گروه‌های مختلف قرار گرفته باشند (در صورتی که پژوهش از نوع آزمایش (Experiment) باشد نه مشاهده (Observation)).
2. نمونه‌های انتخاب شده نباید چولگی شدید داشته باشند. این مسئله از روی نمودار هیستوگرام وعلامت عبارت  قابل تحقیق است که اینجا علامت منفی داشته و در نتیجه چولگی کمی به چپ دارد. اما به دلیل کمی بودن چولگی قضیه استفاده است.



با توجه به بزرگتر بودن p-value از 05/0 فرضیه‌ی اولیه رد نمی‌شود. عدد بدست آمده به این معناست که احتمال مشاهده‌ی میانگین برابر با 318/. با فرض درست بودن میانگین 3/0 برای جامعه برابر با 5/. است. این عدد بزرگتر از آن است که بتوان ادعا کرد به تصادف بدست آمده است. بنابراین نمی‌توان فرض اولیه را در کرد.

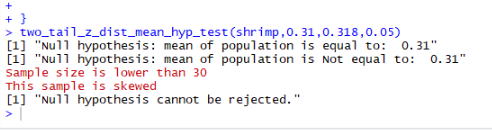
# سوال 19:

پیاده‌سازی تابع خواسته شده توسط کد زیر انجام می‌شود:



# سوال 20:

نتیجه‌ی اجرای تابع بر روی مسئله‌ی 18 بصورت زیر است:



همچنین نتیجه‌ی اجرا بر روی دادگان مسئله‌ی 16 مانند ذیل است.

