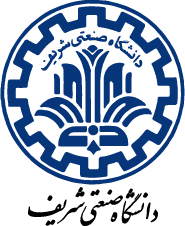
**به نام خدا**



**تمرین شماره­ی3**

**زبان­های شبیه سازی**

**احمد امامی**

**99207521**

\*\* تمام بررسی­ها و نتایج ارائه شده در این تمرین به کمک پایتون انجام شده است و کد آن به پیوست ضمیمه شده است.

# فرضیات

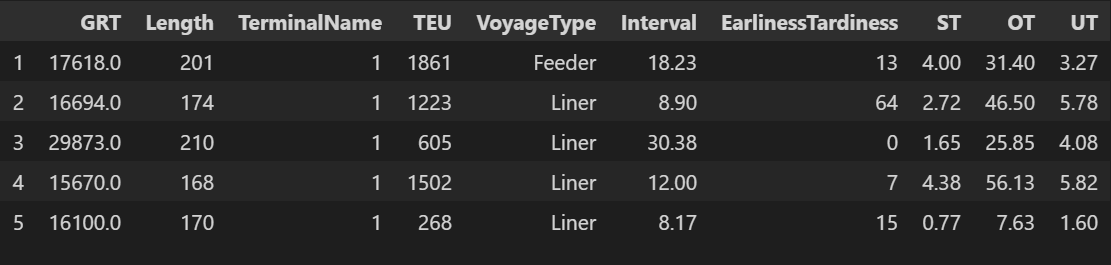
در این تمرین داده­های توناژ کشتی GRT، طول کشتی Length، تعداد کانتینر­ها TEU و زمان بین ورود Interarrival time را به عنوان ورودی سیستم شبیه­سازی در نظر گرفتیم. در ارتباط با ویژگی setup time که در کلاس نیز مورد بحث واقع شد، تصمیم گرفتیم که به عنوان خروجی سیستم در نظر گرفته شود. زیرا عوامل مختلفی در سیستم می­تواند در زمان آماده­سازی کشتی دخیل باشد و این موضوع سبب می­شود که نیاز به تحلیل نتایج خروجی داشته باشیم. در نتیجه در این تمرین 4 ویژگی عنوان شده در بالا مورد بررسی قرار می­گیرند.

هم­چنین در این تمرین مقادیر مشخصات فوق را به صورت جداگانه برای هر کشتی مورد بررسی قرار داده­ایم.

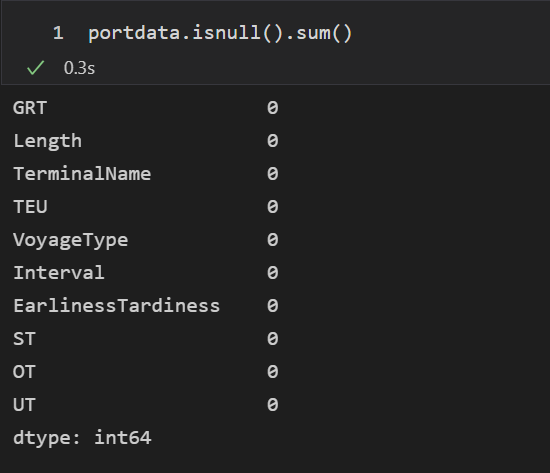
# پیش­پردازش داده­ها Data Cleaning

پیش از پاسخ به پرسش ­های تمرین بهتر است دیتاست مربوطه را بررسی کرده و در صورت وجود داده­ی null و یا داده­های پرت آن­ها را حذف کنیم. زیرا در فرایند فیت کردن توزیع به هرکدام از مشخصه­ها، در صورت وجود داده­های پرت با مشکل مواجه می­شویم.

ابتدا ستون No از دیتاست را حذف می­کنیم زیرا نیازی به آن نخواهیم داشت. 5 سطر ابتدایی دیتاست جدید را مشاهده می­کنید.

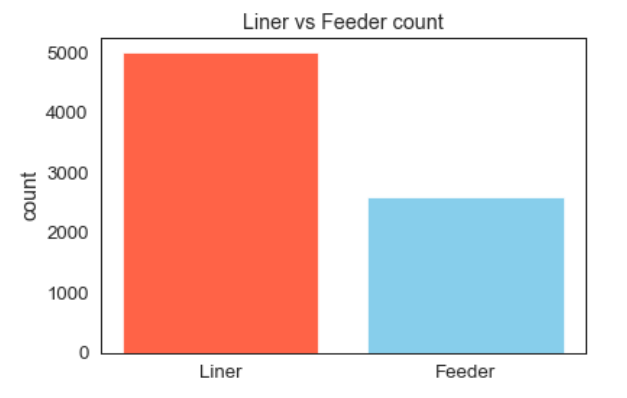


داده­های null را بررسی می­کنیم:



همان­طور که می­بینیم در هیچ یک از ستون­های دیتاست داده­ی null مشاهده نمی­شود و از این نظر مشکلی در دیتاست مشاهده نمی­شود.

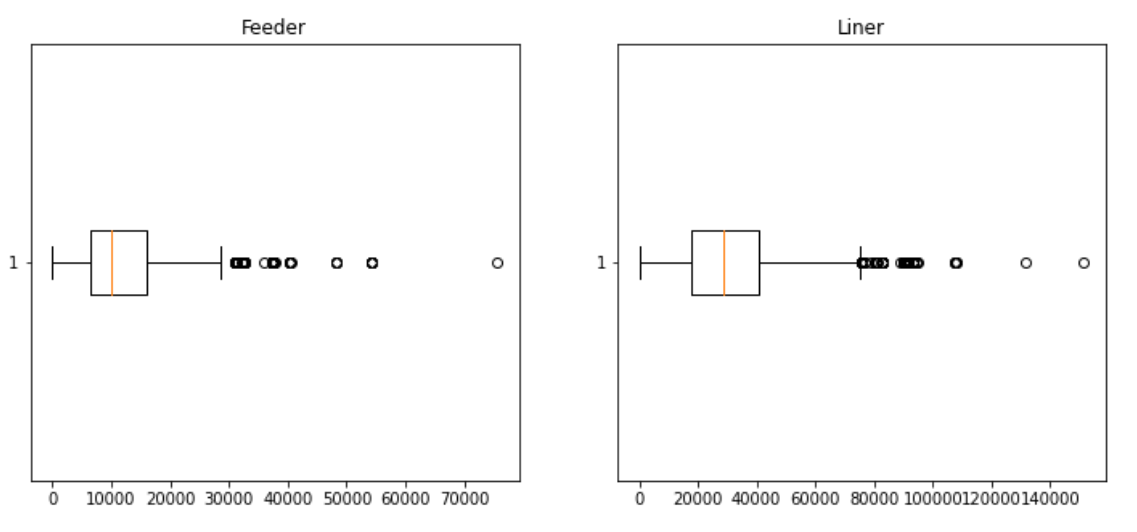
در مرحله­ی بعدی بهتر است که داده­هایمان را به دو دسته تقسیم کنیم. یک قسمت مربوط به کشتی­های لاینر و دیگری مربوط به کشتی­های فیدر. از آنجایی که مشخصات مرتبط با هر کشتی مختص خودش است، بهتر است آن­ها را به صورت جداگانه و در دو دیتاست مختلف مورد بررسی قرار دهیم. همان طور که در تصویر زیر می­بینیم حدودا 5 هزار کشتی لاینر و 2500 عدد نیز فیدر هستند که به صورت مجزا مورد بررسی قرار می­دهیم.



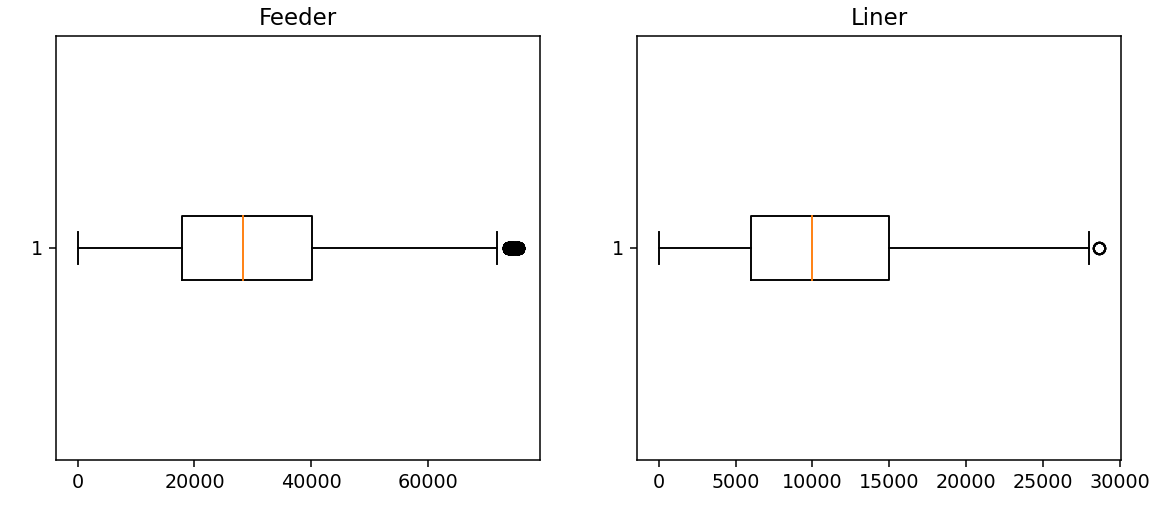
**حذف داده­های پرت**

برای حذف داده­های پرت از نمودار جعبه­ای (boxplot) کمک می­گیریم. برای هر مشخصه این نمودار را رسم می­کنیم و داده­های بسیار بزرگ و غیر معمول را حذف می­نماییم.

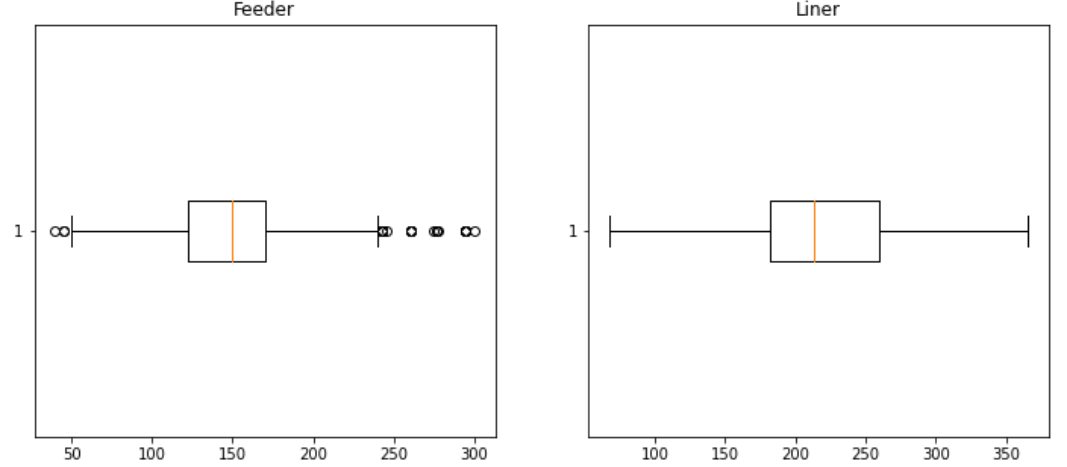
**GRT**



پس از حذف داده­های پرت نمودار باکس­پلات شاخصGRT به شکل زیر خواهد بود:

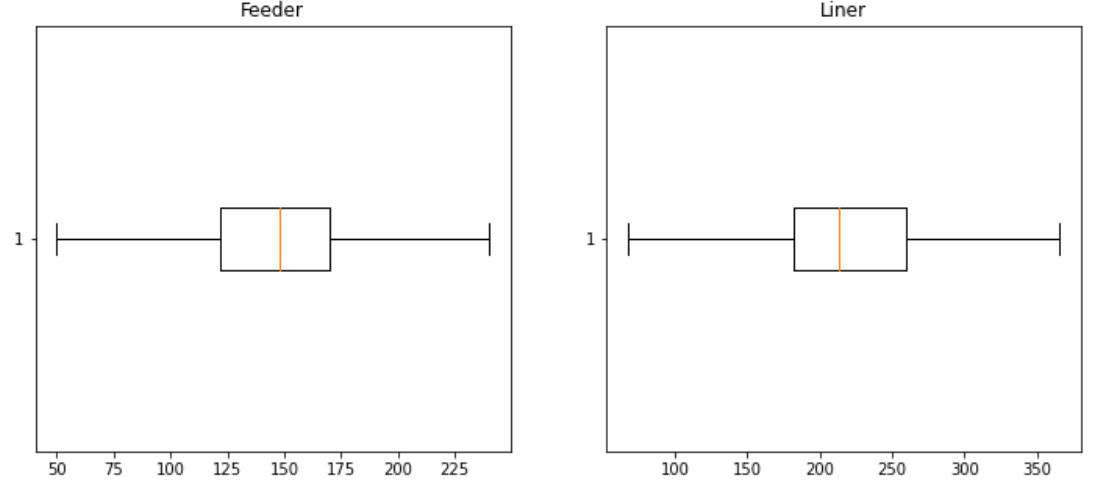


**Length**

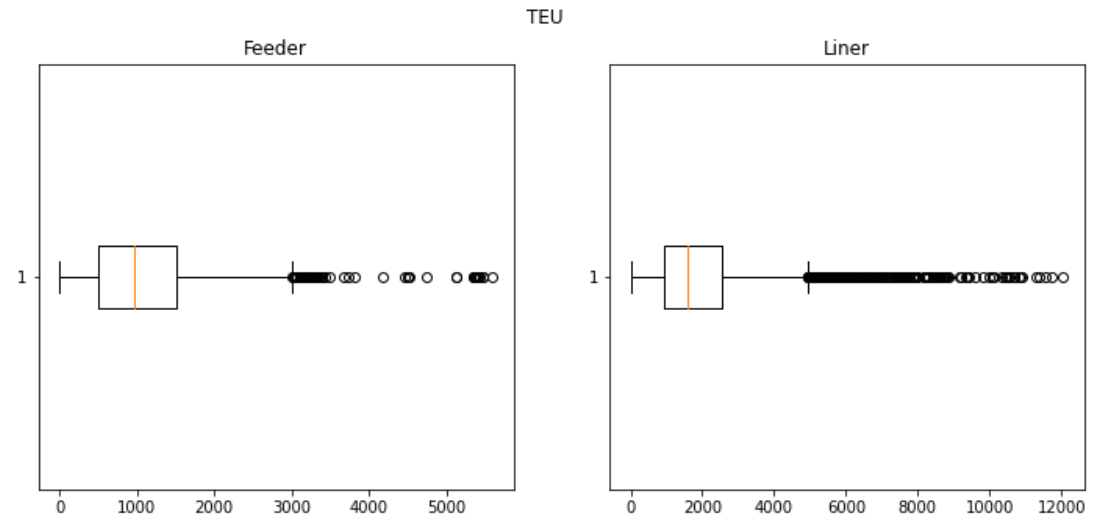


برای طول کشتی­ها وضعیت بهتری را شاهد هستیم و برای کشتی­های Liner داده­ی پرتی مشاهده نمی­شود. برای کشتی­های Feeder تعدادی مشاهده outlier داریم که تصمیم گرفتیم که آن­ها را حذف بنماییم.

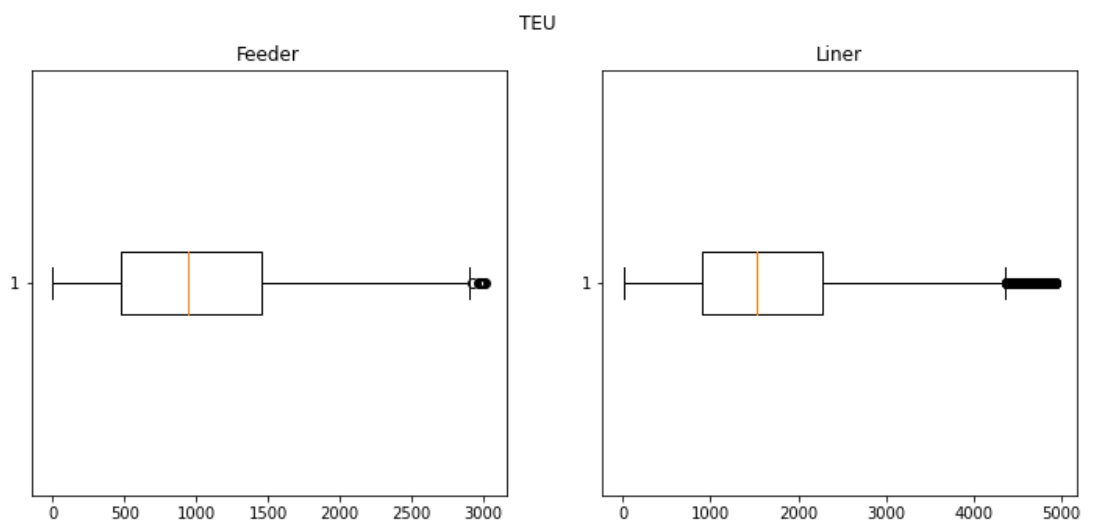
باکس­پلات داده­های جدید طول کشتی به شکل زیر است:



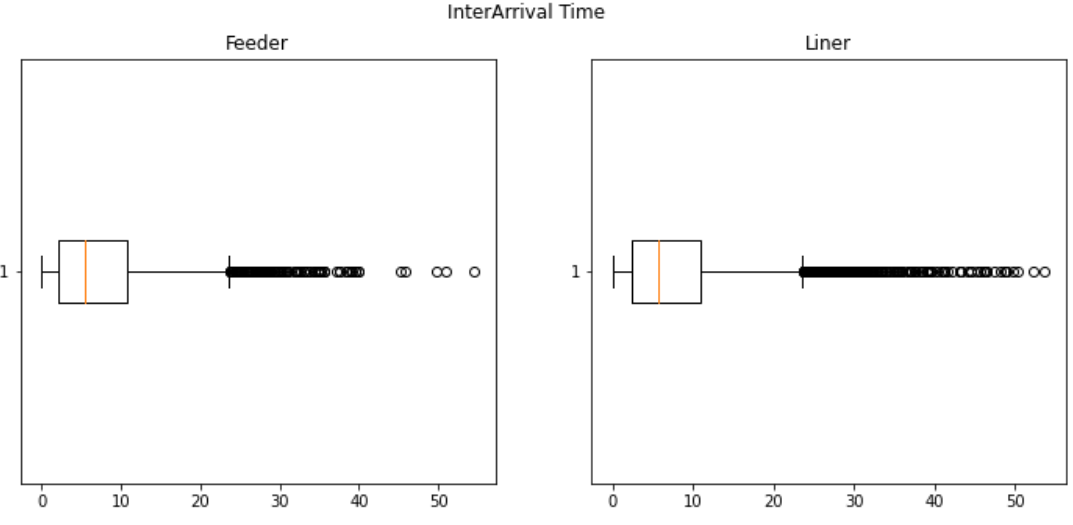
**TEU**

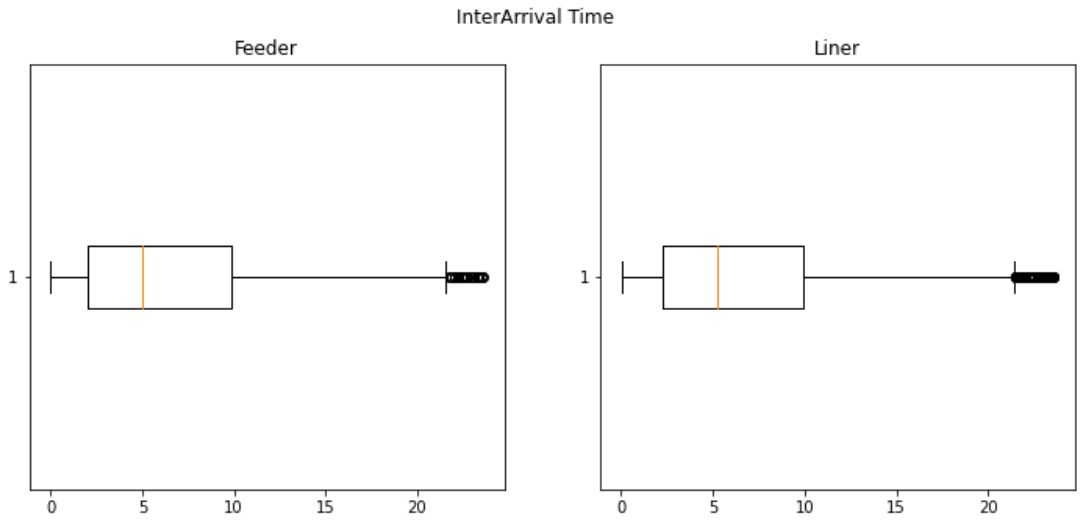


همان­طور که مشاهده می­شود واضح است که داده­ها توزیع چندان مناسبی ندارند و بهتر است داده­های پرت را حذف کنیم. نمودار باکس­پلات بروز شده در شکل زیر قابل مشاهده است.



**Interarrival time**





# سوال 1

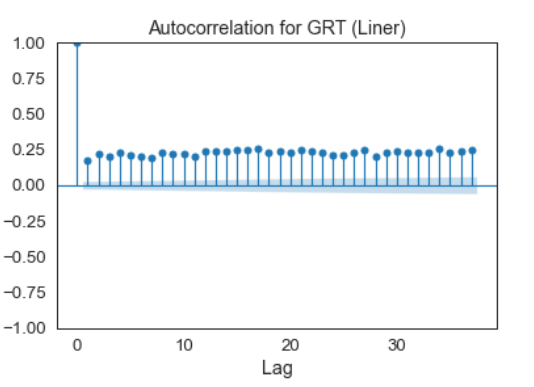
# بررسی فرض iid بودن

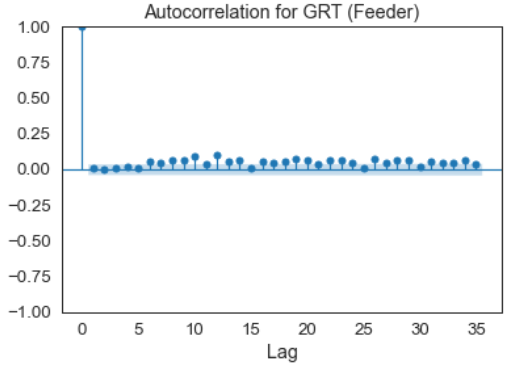
--------------------------------------------------------------------------------------------------------

برای بررسی iid بودن باید چک شود که آیا داده­های مربوطه در هر ستون به یکدیگر مرتبط هستند یا خیر. به بیان دیگر آیا خوهمبستگی autocorrelation میان داده­ها دیده می­شود یا نه. اگر فرض iid بودن برقرار باشد آنگاه داده­های ما مانند یک نویز رفتار می­کنند و بین هیچ لگی همبستگی­ای وجود نخواهد داشت. یکی از بهترین راه­ها برای بررسی این موضوع استفاده از نمودار خودهمبستگی (ACF) می­باشد. در صورت مشاهده­ی یک لگ قابل توجه در این نمودار میتوان نتیجه گرفت که داده­ها دارای خودهمبستگی هستند. نتایج زیر را برای مشخصه­های مختلف دیتاست مربوطه مشاهده می­کنید:

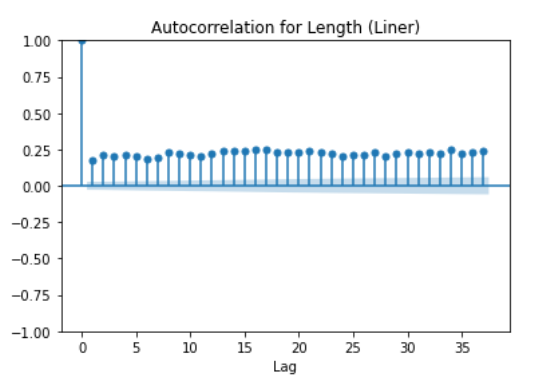
## توناژ ناخالص کشتی GRT

همان­طور که در تصاویر زیر مشاهده می­شود مقدار همبستگی میان توناژ ناخالص در کشتی­های لاینر و فیدر دارای خودهمبستگی است و در لگ­های متعددی مقادیر قابل توجه مشاهده می­شود. در نتیجه فرض i.i.d بودن در ارتباط با مشخصه­ی توناژ ناخالص کشتی قابل قبول نمی­باشد.

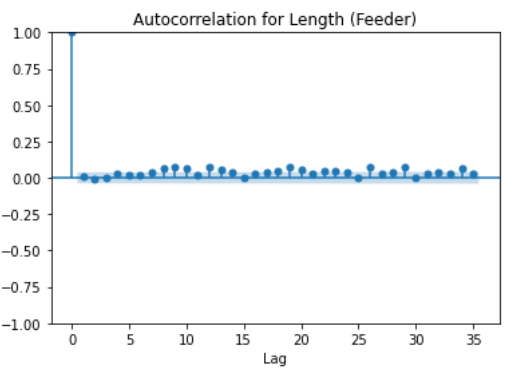




## طول کشتی­ها Length

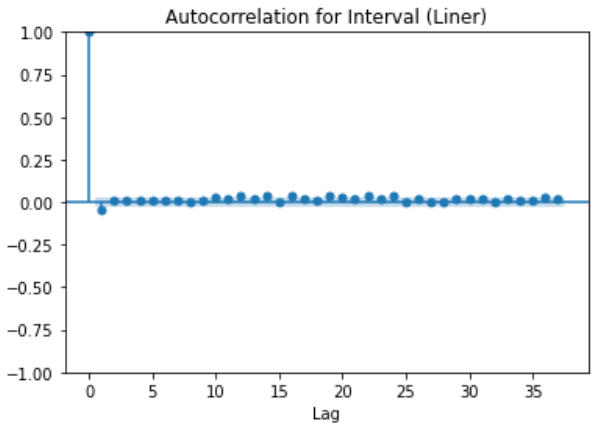


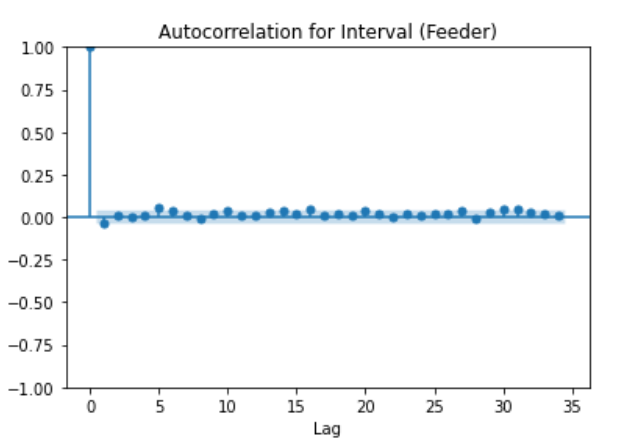
برای کشتی­های لاینر واضح است که طول کشتی دارای خودهمبستگی است و فرض i.i.d بودن برقرار نیست.



برای کشتی­های Feeder مشاهده می­کنیم که طول کشتی­ها نزدیک به iid است و میتوان فرض کرد استقلال خطی میان آن­ها برقرار است.

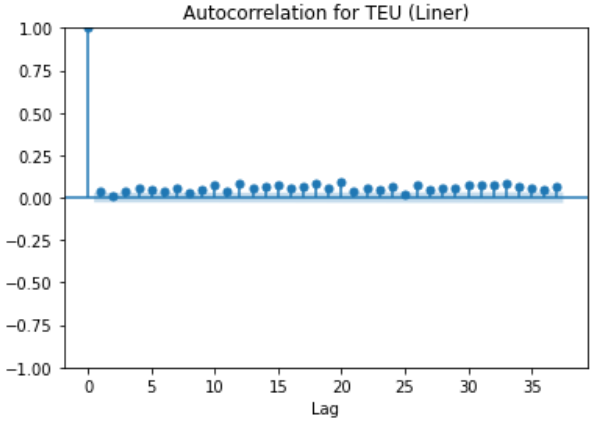
## زمان بین ورود (Interarrival time)

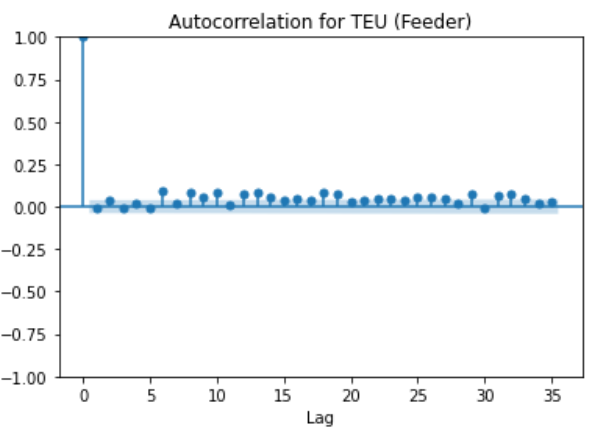




برای زمان بین ورود هر دو کشتی مشاهده می­کنیم که شرط iid بودن برقرار است.

## تعداد کانتینر­های کشتی TEU



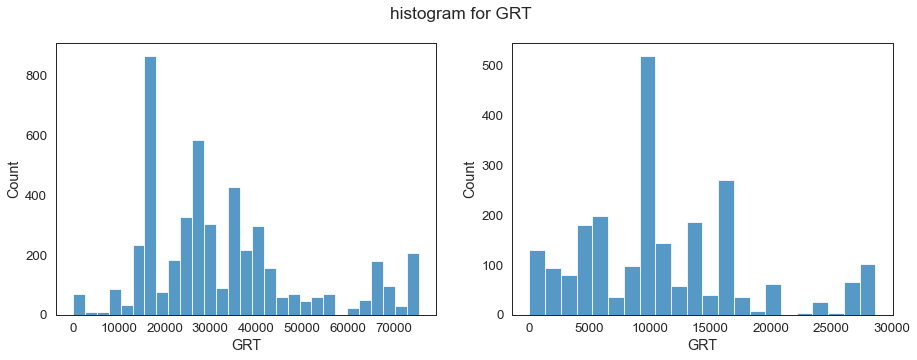


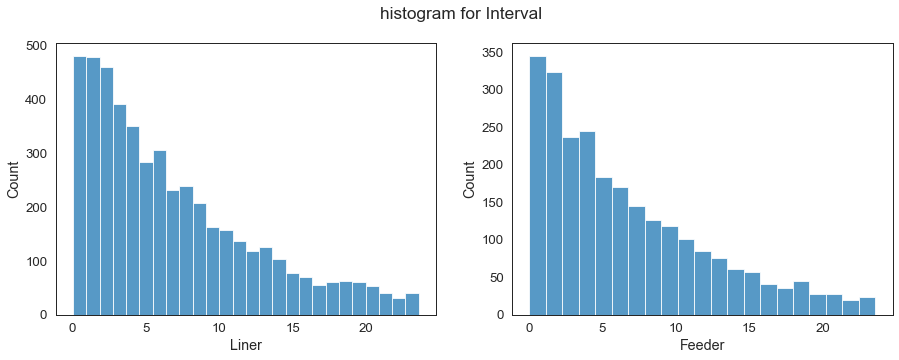
بار کشتی­های لاینر و فیدر نیز تا حد خوبی به iid بودن نزدیک است و از این رو میتوان فرض کرد از توزیع یکسانی برخوردار هستند.

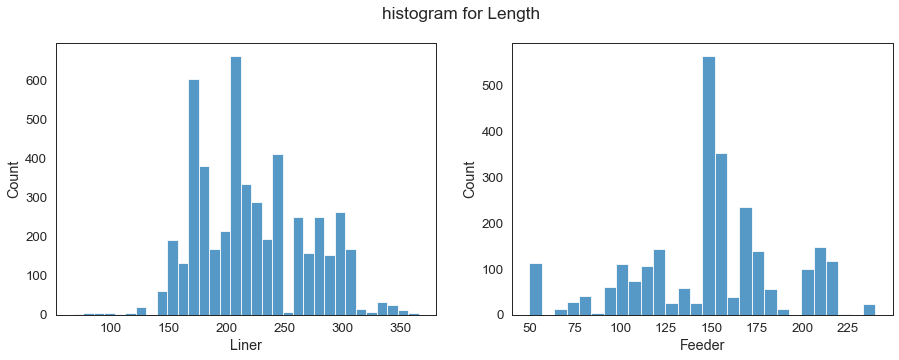
در ادامه نمودار هیستوگرام مشخصه­های فوق را رسم می­کنیم و توزیع­های احتمالی را برای هر یک بیان می­کنیم.

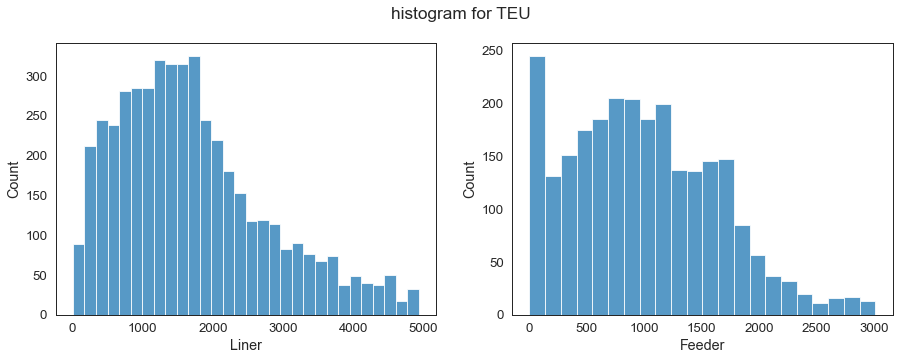
# بررسی هیستوگرام داده­های ورودی

برای رسم هیستوگرام­ها نیز مجددا از پایتون استفاده کردیم و نتایج را در شکل­های زیر مشاهده می­نمایید.









# همبستگی میان ویژگی­ها

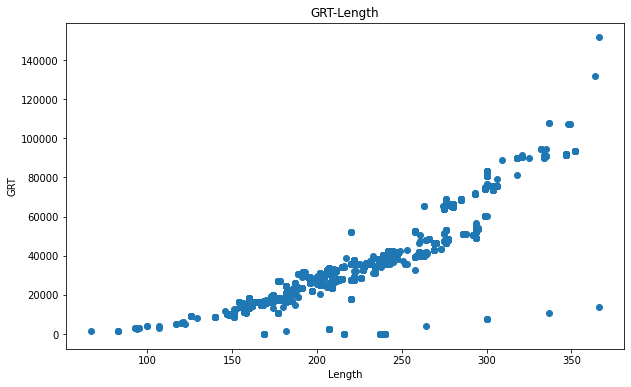
برای بهتر نشان دادن ارتباط و همبستگی میان مشخصه­های مختلف دیتاست از ماتریس همبستگی استفاده کردیم. نتایج حاصله برای کشتی­های Liner و Feeder به شکل زیر است.

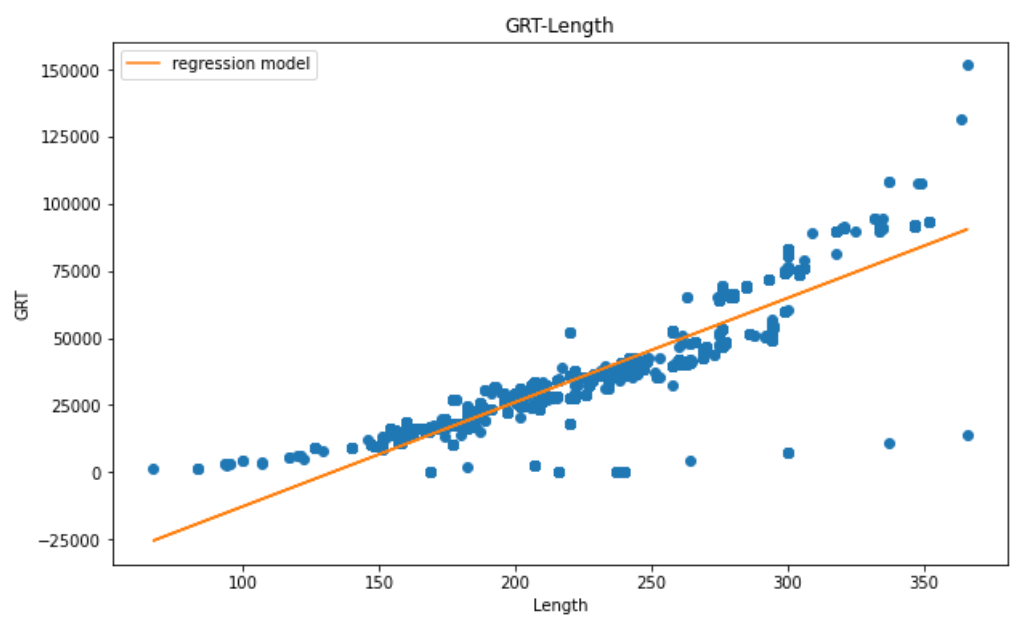




همان طور که مشاهده می­شود دو ویژگی TEU و GRT همبستگی بسیار زیادی با یکدیگر دارند. میتوان برای درک بیشتر رگرسیون خطی بین این دو مقدار را نیز به دست آورد. پیش از آن نمودار توناژ کشتی بر حسب طول آن را در رسم می­کنیم. همان طور که در نمودار مشاهده می­شود و انتظار داشتیم یک همبستگی خطی میان این دو ویژگی دیده می­شود.

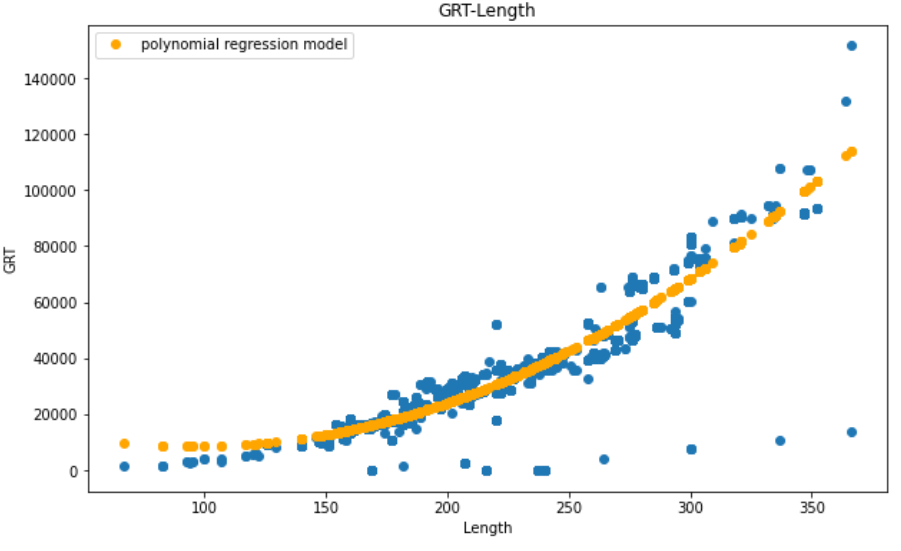
زمانی که دو یا چند ویژگی در دیتاست مورد بررسی­مان دارای همبستگی خطی هستند بهتر است از رگرسیون کمک بگیریم و داده­ی دیگر را بر حسب دیگری بیان کنیم. در این صورت میتوان یکی از آن­ها را از مدل شبیه­سازی حذف نمود. در ادامه مدل رگرسیونی میان این دو ویژگی را بیان می­کنیم.





با توجه به خروجی نرم افزار معادله­ی خط به صورت می­باشد

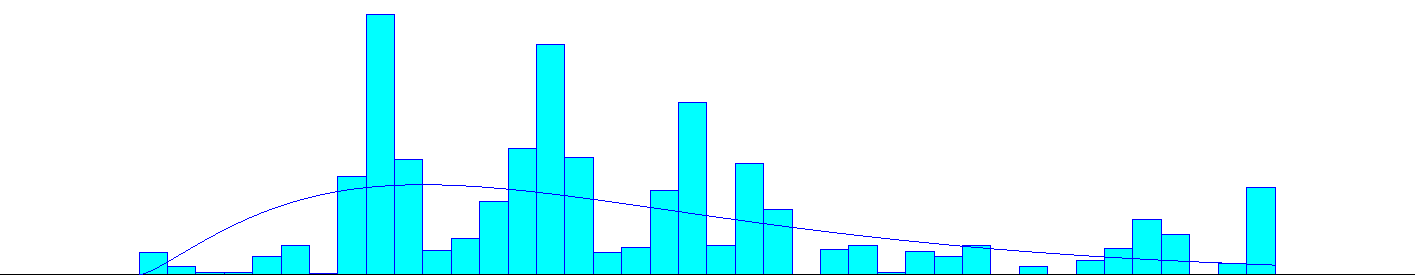
*هم­چنین میتوان رگرسیون­های مرتبه­ی بالاتر را نیز پیاده­سازی نمود. به عنوان مثال در شکل زیر یک رگرسیون درجه دوم را فیت نموده­ایم که نتیجه را مشاهده می­کنید:*

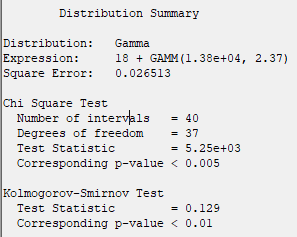
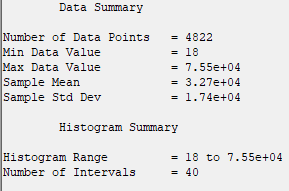


# فیت کردن داده­ها و آزمون­های برازش نیکویی

برای این قسمت از تمرین ابتدا داده­های پیش­پردازش شده را در پایتون فراخوانی می­کنیم و به تفکیک نوع کشتی در فایل txt مجزا قرار می­دهیم. سپس به کمک ابزار input analyzer در arena بهترین توزیع را به آن فیت می­کنیم و نتایج آزمون­های برازش نیکویی را مشاهده می­کنیم. نتایج حاصله به تفکیک در ادامه آورده شده اند:

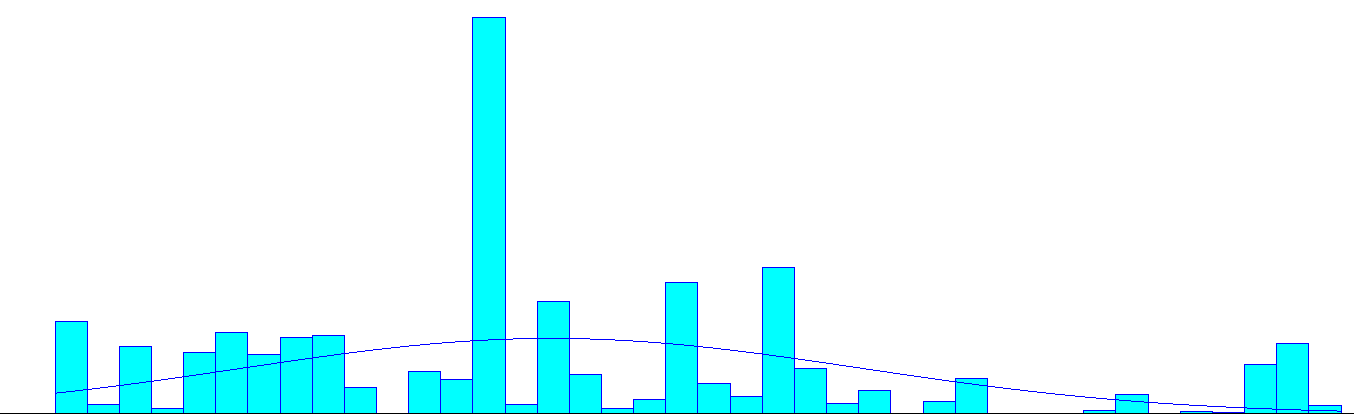
## GRT-Liner

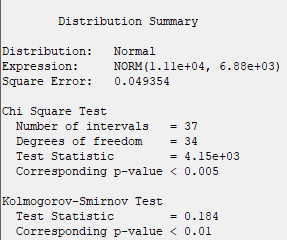
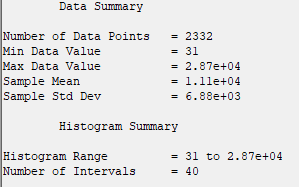


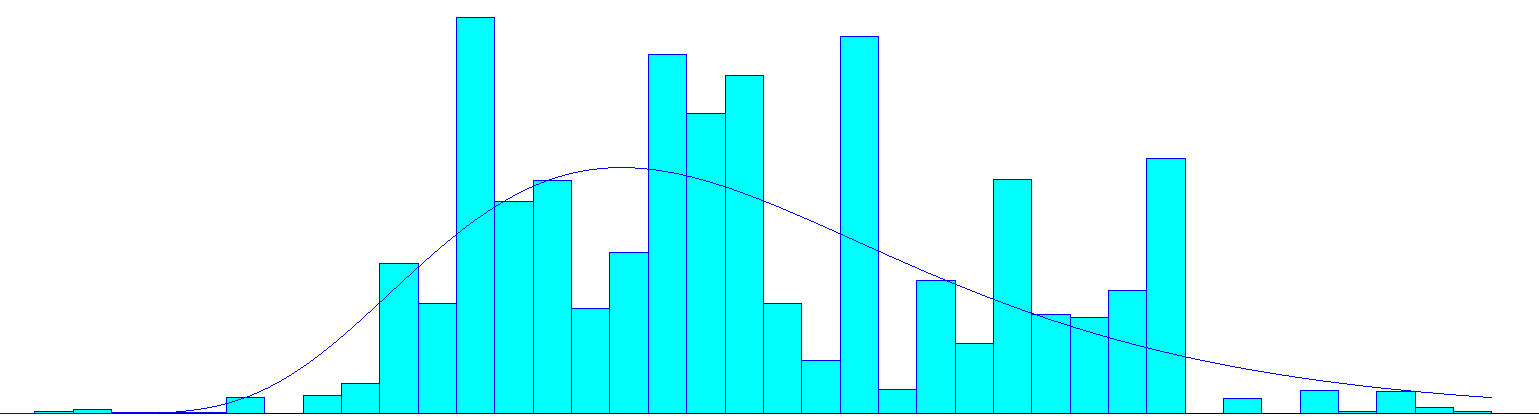
همان­طور که در نتایج مشاهده می­شود فرض 0 در هر دو تست برازش رد می­شوند و بدین معنی است که توزیع مورد نظر فیت مناسبی برای داده­های ما نیست. دلیل اصلی این اتفاق iid نبودن داده­ها است.

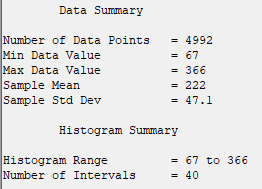
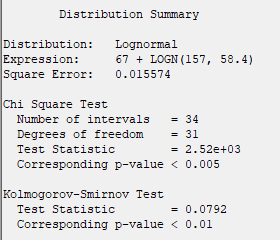
## GRT-Feeder



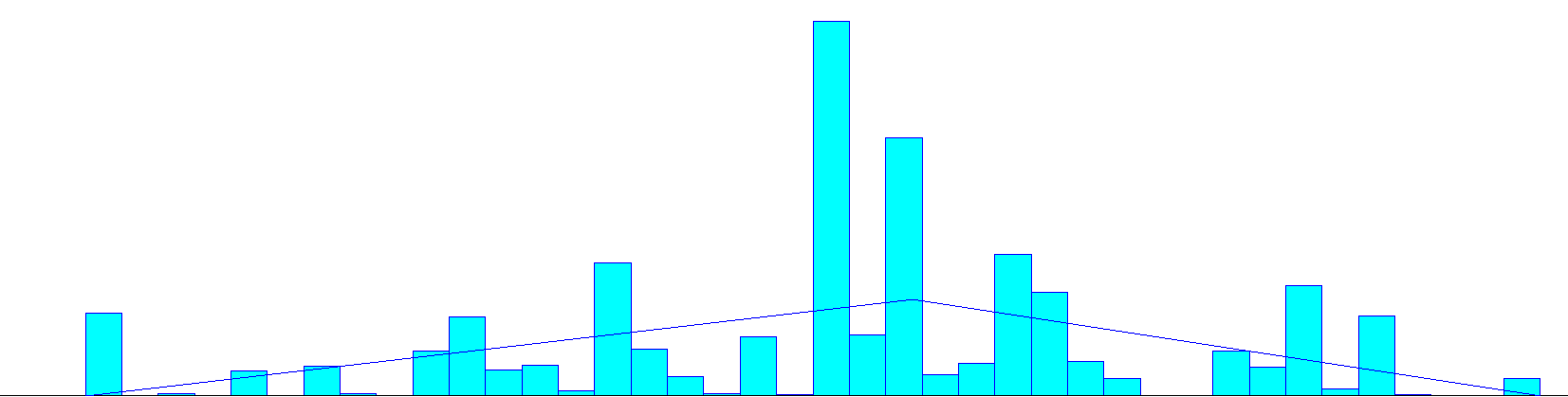
 

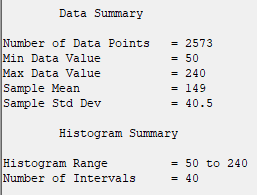
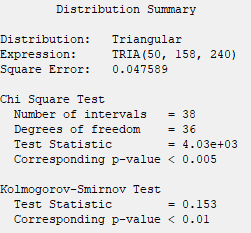
## Length-Liner



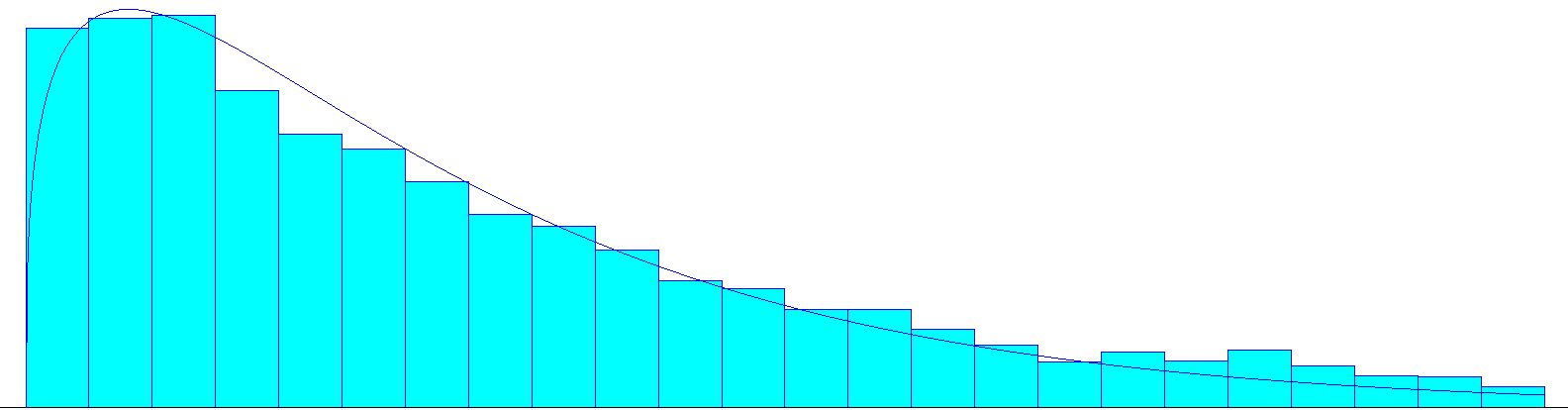


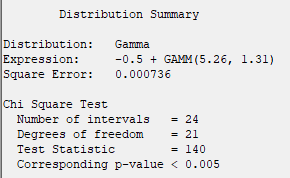
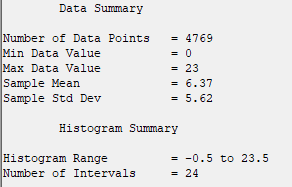
## Length-Feeder



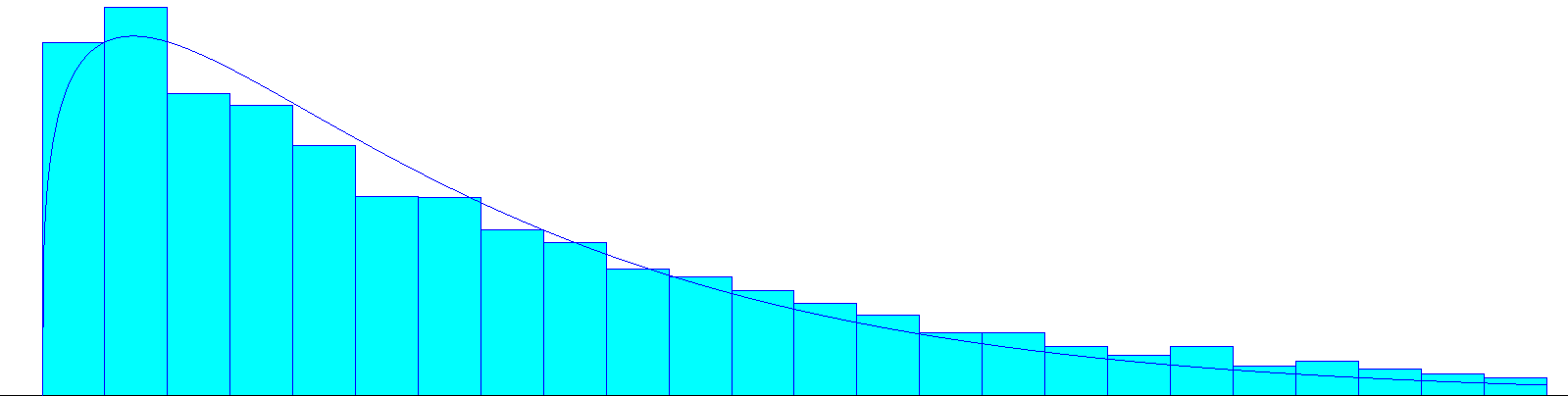


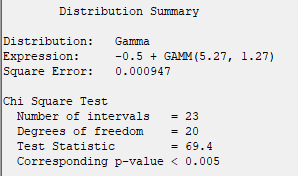
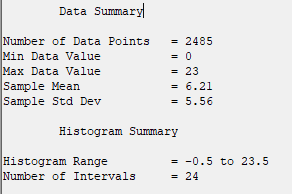
## Arrival-Liner



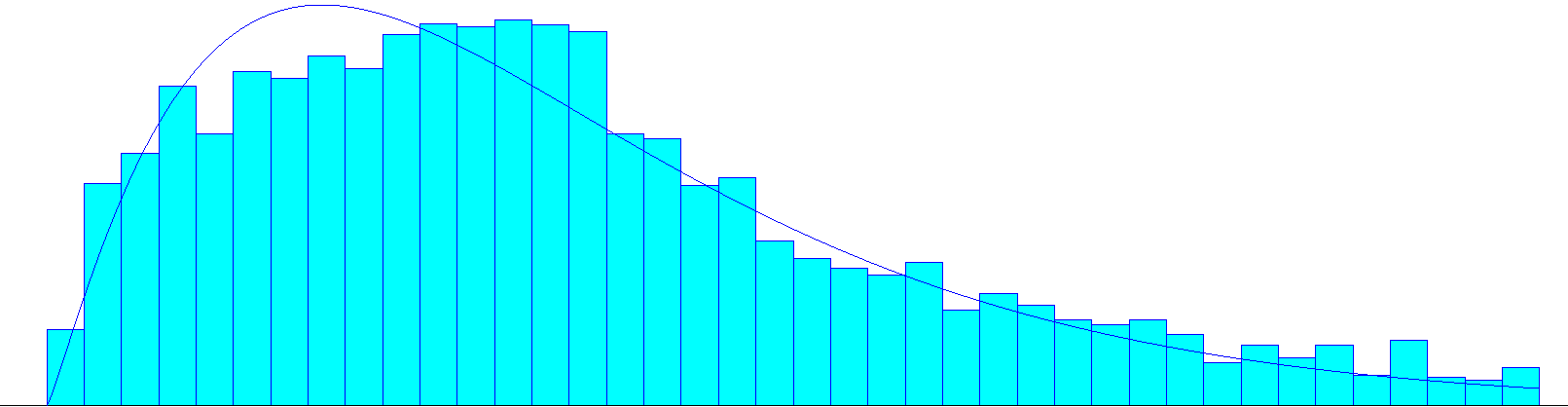
 

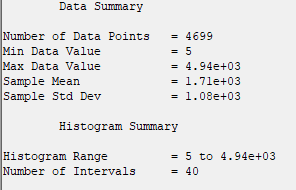
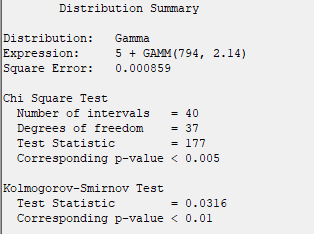
## Arrival-Feeder



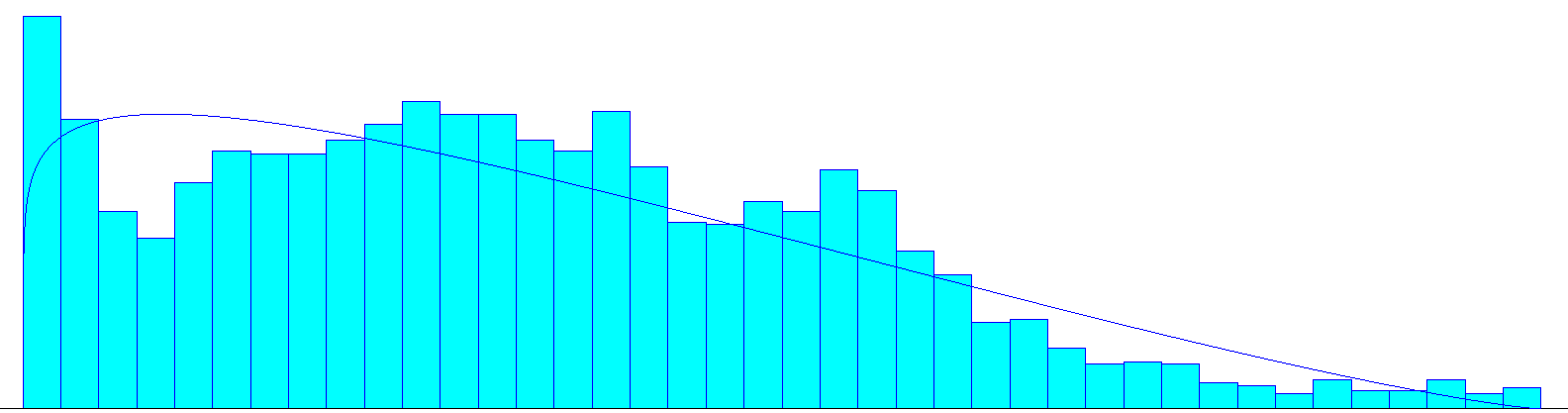
 

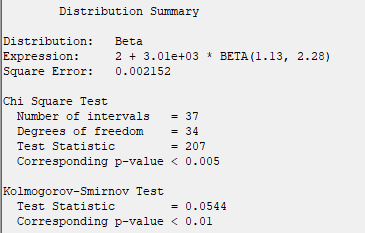
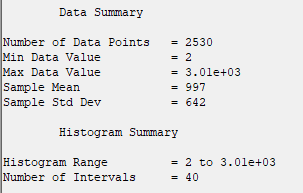
## TEU-Liner





## TEU-Feeder



از آنجایی که داده­ها iid نبودند، نمیتوان با اطمینان کامل توزیع مناسبی را برای داده­ها در نظر گرفت. این امر باعث شده که آماره ها نیز نتوانند پذیرش فرض صفر مبنی بر فیت بودن داده­ها را به خوبی تایید کنند. تفکیک داده­ها بر حسب نوع کشتی نتایج بهتری را حاصل کرد ولی لزوما باعث بهتر شدن آماره آزمون­ها نشد. در صورتی که نتوان به نتایج خوبی از توزیع داده­ها رسید بهتر است از توزیع­ تجربی برای توصیف داده­های ورودی بهره بگیریم. این کار نیز به کمک نرم­افزار arena قابل انجام است.