بسم تعالی


# آزمایشگاه علوم اعصاب

## تمرین شماره ۳

امیرحسین زاهدی ۹۹۱۰۱۷۰۵

**بهار ۱۴۰۳**

## بخش اول:

الف) آزمایشی که داده از آن بدست آمده حاوی متغیر هایی است که توسط آزمون گیرنده به ازای هر آزمودنی می تواند تغییر کند. برای مثال در این آزمایش مدت زمان ترینینگ یا تعداد فرکتال های نشان داده شده در هر ترایال می توانند تغییر کنند و توسط آزمون گیرنده تعیین می شوند. به همین دلیل نوع دیتا از نوع Experimental است.

ب) داده Display Size و داده Training Duration هر دو از نوع Categorical هستند زیرا که حاوی مقادیر گسسته هستند. برای مثال تعداد فرکتال ها محدود و به میزان گسسته است. میزان مدت ترینینگ نیز به صورت گسسته تعیین شده است. البته درباره داده مدت زمان جست و جوی چشم این مسئله صدق نمی کند و به دلیل ماهیت زمانی بودن آن از نوع پیوسته است. البته به هر حال چون با فرکانسی، نمونه برداری شده است، این داده را نیز حتی می توان گفت که گسسته است.

## بخش دوم:

الف) همانطور که در بخش قبل نیز بیان شد، متغیر سرچ تایم، وابسته به متغیر های زمان ترینینگ و دیسپلی سایز است به همین دلیل با استفاده از داده های موجود مرتبط با سرچ تایم در دیتاست، مدلی خطی را با استفاده از دو متغیر گفته شده فیت می کنیم. مدل فیت شده با استفاده از دو ضریب و عرض از مبدا و خطا به صورت زیر است:

$$ST_i = \beta_1 DS_i + \beta_2 TD_i + \beta_0 + \epsilon_i$$

مدل را فیت می کنیم و استت ها را با استفاده از همان داده های بدست آمده از مدل داریم:

```
Regression Coefficients, significance, t-values and p-values :
```

| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| **(Intercept)** | 142.9 | 8.9474 | 15.971 | 3.3032e-56 |
| **x1** | 25.259 | 1.1222 | 22.509 | 1.4516e-107 |
| **x2** | 6.7084 | 1.6824 | 3.9874 | 6.7638e-05 |

به ترتیب از چپ به راست، مقادیر ضرایب، standard error، t-values و p-values را داریم. به ترتیب از بالا به پایین نیز مربوط به بتا۰ و بتا۱ و بتا۲ هستند. (ضرایب رگرسیون گفته شده.)

همانطور که میبینیم p-values در هر سه ضریب بسیار بسیار کم و نزدیک صفر است که به معنای معنا دار بودن و موثر بودن ضرایب به لحاظ آماری در مدل رگرسیون بدست آمده است.

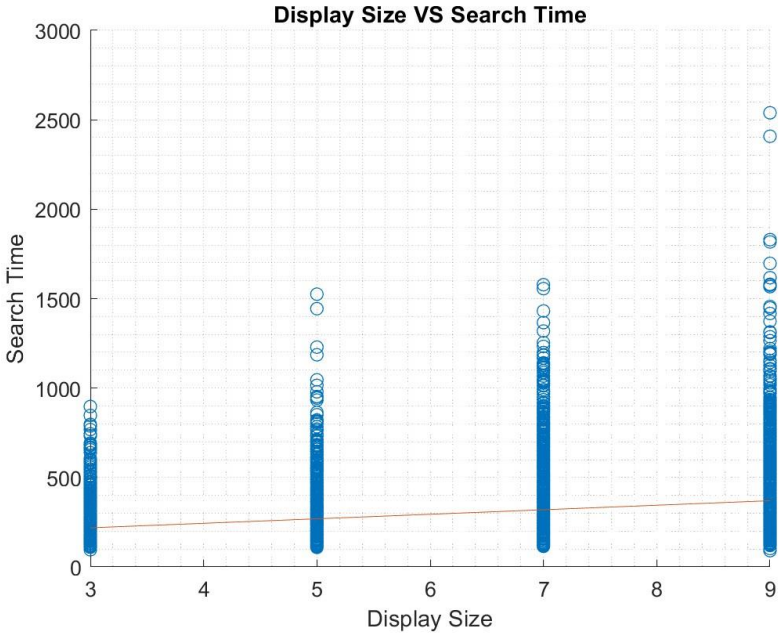
با استفاده از MSR و MSE، F-static را محاسبه می کنیم.

```
F-statistic :
  173.5660
```

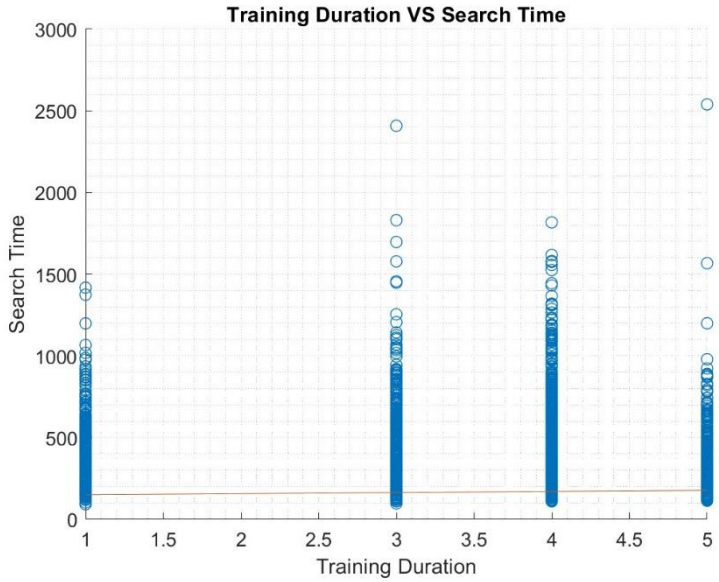
مولفه آماری F نشان دهنده میزان خوب بودن یا معنا دار بودن مدل فیت شده نسبت به حالت رندوم است که هر چه بزرگتر باشد، نشان می دهد مدل معنا دار تر و موثر تر است. در این رگرسیون نیز شاید هستیم که این مولفه بزرگ است و نشان از معنا دار بودن مدل سازی انجام شده بر روی دیتای داده شده است.

ب) در این بخش سرچ تایم را بر اساس هر کدام از متغیر های مستقل دسیپلی سایز و ترینینگ دیوریشن رسم می کنیم. همچنین در پلات سه بعدی بر اساس هر دو متغیر نیز رسم می کنیم. لازم به ذکر است که در هر پلات، خط یا صفحه رگرسیون بر حسب مدل فیت شده نیز رسم شده است.

ابتدا دیسپلی سایز:

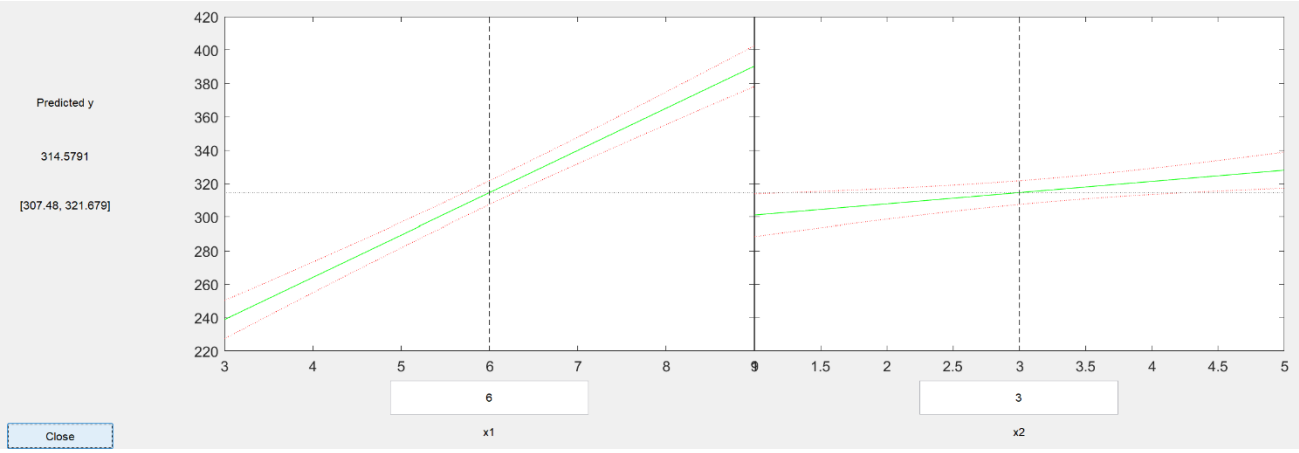



سپس ترینینگ دیوریشن:



حال بر اساس هر دو متغیر:

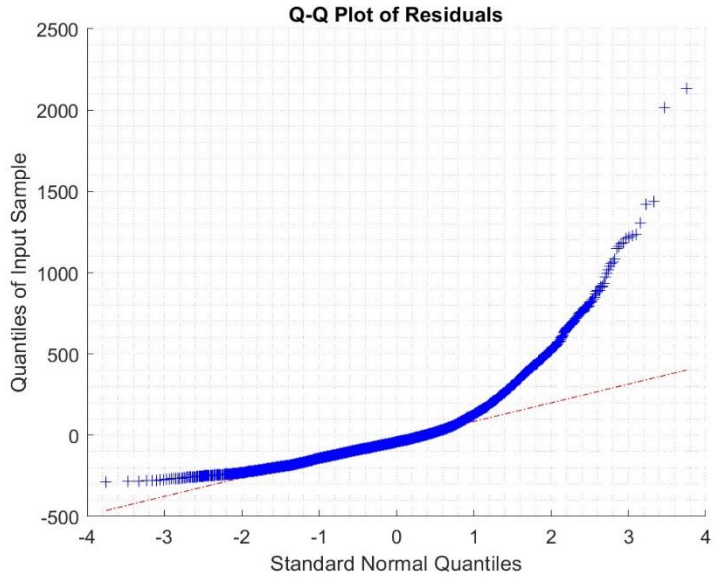

ج) در این بخش سرچ تایم تخمین زده شده بر اساس هر یک از متغیر ها به همراه باند ۹۵ درصد اطمینان آن، در پلات های ۲ بعدی رسم می کنیم.




پلات چپ بر اساس دیسپلی سایز و پلات راست بر اساس ترینینگ دیوریشن است.

## بخش سوم:

الف) نمودار Q-Q اختلافات را رسم می کنیم تا ببینیم از توزیع نورمال تولید می شوند یا خیر.




همانطور که مشاهده می شود، انطباق بر روی نیمساز ناحیه اول نمودار اتفاق نیفتاده است، به همین دلیل می توان گفت که باقی مانده ها یا همان تفاوت های مقادیر واقعی و تخمین زده شده از توزیع نورمال بدست نیامده اند.

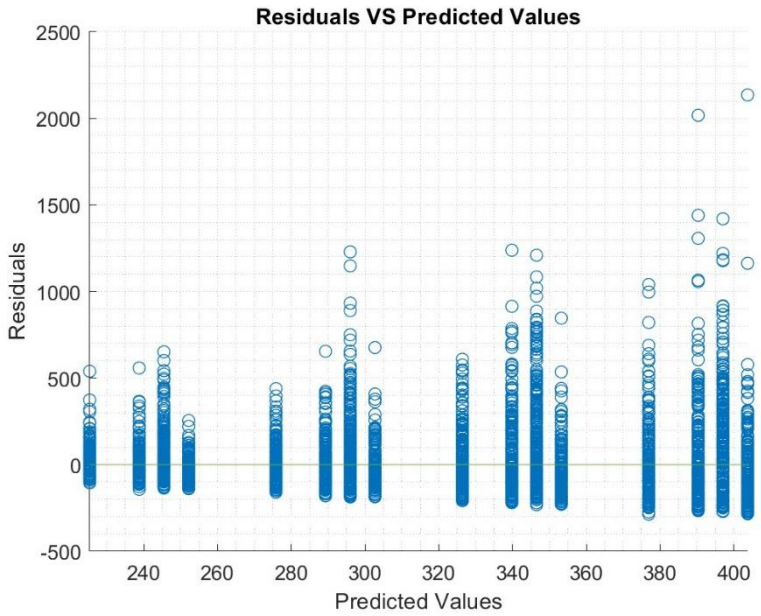
ب) در این بخش ابتدا میانگین اختلافات را محاسبه کرده و با استفاده از میانگین، سعی می کنیم از روی نمودار بدست آوریم که آیا واریانس ثابت است یا نه.

```
mean_residuals =

   -2.7720e-14
```

میانگین با تقریب عالی ۰ است.

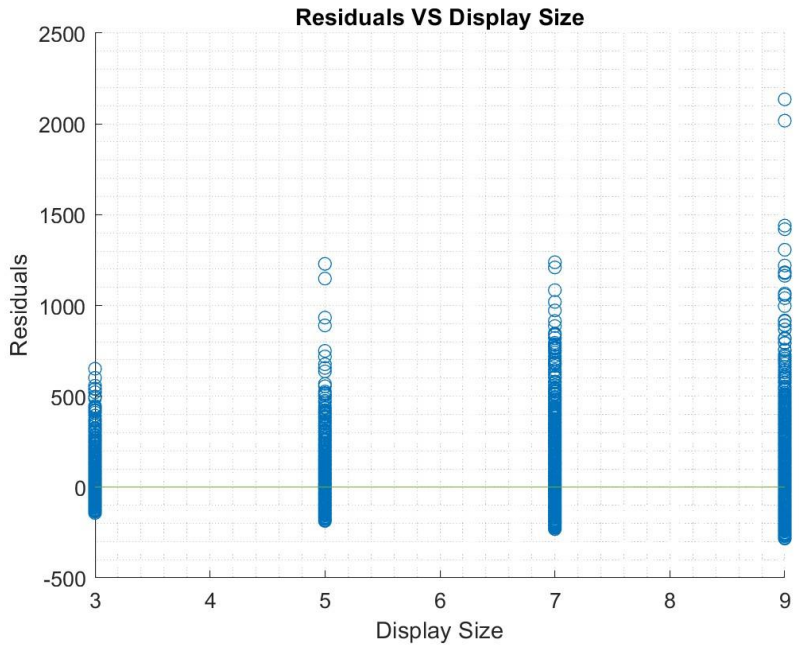
نمودار اختلافات را بر حسب مقادیر تخمین زده شده رسم می کنیم.




با توجه به میانگین ۰ می بینیم که مقادیر پراکندگی ثابت نیست و تغییر می کند.

به همین دلیل واریانس ثابت نیست.

ج) در این بخش می خواهیم ببینیم که اختلافات مستقل هستند یا نسبت به متغیر ها وابسته هستند. البته که از نمودار بخش قبل می توان گفت که وابسته هستند ولی برای مثال این اختلافات را بر خسب دیسپلی سایز نیز رسم می کنیم تا وابستگی را بهتر ببینیم.




با توجه به نمودار میبینیم که مقدار اختلاف نسبت به این متغیر وابسته است و تغییر می کند. واریانس آن نیز نسبت به این متغیر عوض می شود. احتمالا برای متغیر ترینینگ دیوریشن نیز به همین صورت است ولی در هر صورت وابستگی اختلافات ثابت شده است.

## بخش چهارم:

ابتدا سرچ تایم را با استفاده از دیسپلی سایز تخمین می زنیم و مدل فیت می کنیم. سپس با استفاده از باقی مانده ها، ترینینگ دیوریشن را تخمین می زنیم. به ترتیب می بینیم:

| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| (Intercept) | 164.98 | 7.0381 | 23.441 | 4.3287e-116 |
| x1 | 25.211 | 1.1236 | 22.438 | 6.2343e-107 |

| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| (Intercept) | 3.2491 | 0.019636 | 165.46 | 0 |
| x1 | 0.00041421 | 0.00010388 | 3.9875 | 6.7603e-05 |

در ابتدا می بینیم که ضریبی که برای متغیر دیسپلی سایز بدست آمده است شبیه ضریبی است که در مدلسازی با هر دو متغیر بدست آمده است.

مقدار ضریب باقی مانده ها در تخمین ترینینگ دیوریشن نیز نشان می دهد که باقی مانده ها ارتباطی با ترینینگ دیوریشن ندارند.

حال برعکس بالا را اجرا کرده و ابتدا سرچینگ تایم را به وسیله ترینینگ دیوریشن تخمین می زنیم و سپس با استفاده از اختلافات، دیسپلای سایز را فیت می کنیم.

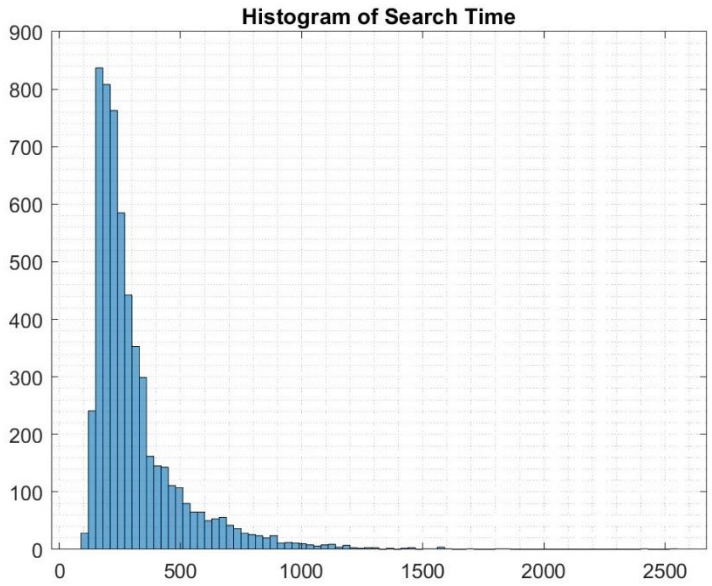
| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| (Intercept) | 292.1 | 6.2707 | 46.581 | 0 |
| x1 | 6.3044 | 1.7553 | 3.5917 | 0.00033128 |

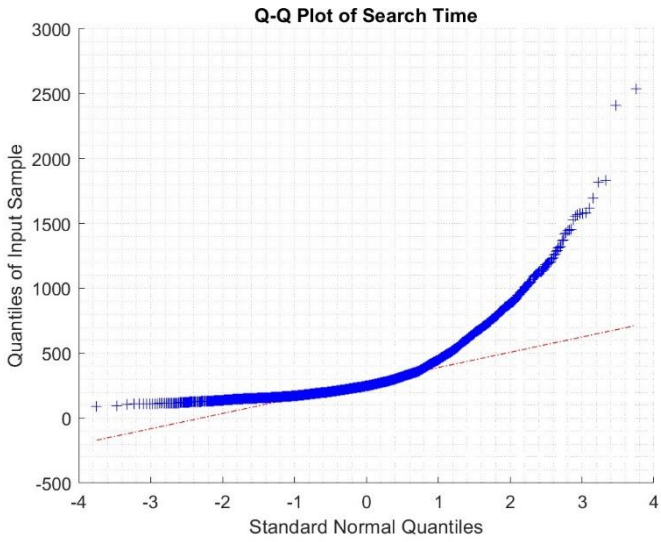
| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| (Intercept) | 5.8548 | 0.028253 | 207.23 | 0 |
| x1 | 0.0032287 | 0.00014344 | 22.51 | 1.4318e-107 |

اینبار نیز ضریب متغیر ترینینگ دیوریشن تا حدی شبیه مدل کامل است اما همچنان شبیه قبلی، اختلافات ربطی به دیسپلای سایز ندارند زیرا که ضریب نزدیک ۰ است. البته پی ولیو بسیار کم نشان دهنده معنا دار بودن تخمین است.

## بخش پنجم:

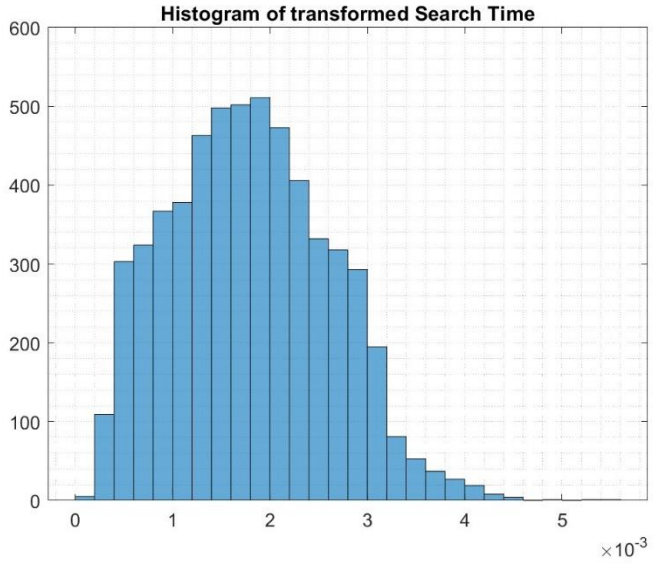
با هر دو روش هیستوگرام و پلات Q-Q بررسی می کنیم که آیا توزیع سرچ تایم نورمال است یا خیر.

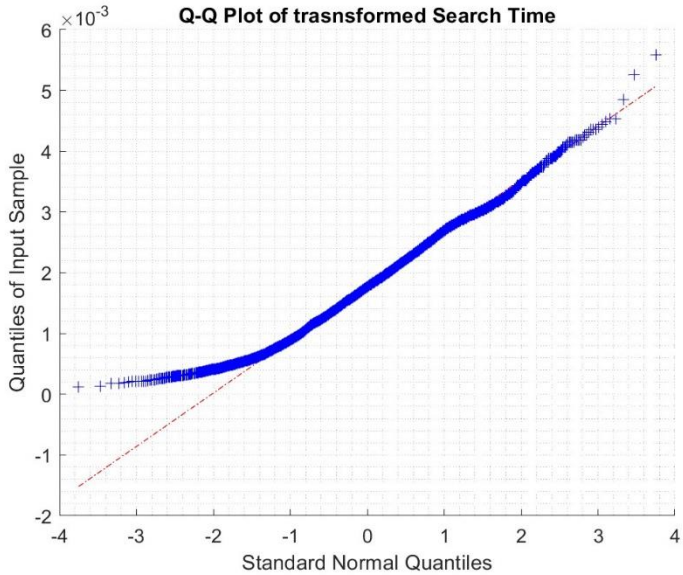





طبق هر دو نمودار می بینیم که توزیع به نورمال شباهتی ندارد به همین دلیل سعی می کنیم تا تبدیلی بر روی داده های سرچ تایم انجام دهیم تا توزیع آن به توزیع نرمال نزدیک تر شود.

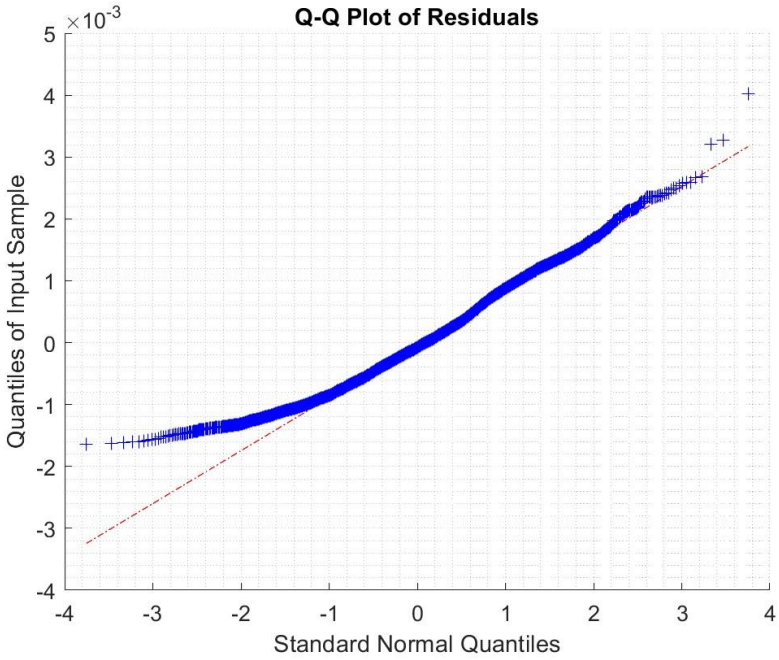
برای اینکار دیتا را به توان ۱.۱- می رسانیم و خروجی های هیستوگرام و پلات Q-Q را رسم می کنیم.



از نمودار ها این برداشت را می کنیم که تا حدی توانسته ایم توزیع نرمال را ایجاد کنیم ولی طبیعتا خیلی می توانست بهتر باشد، اما در بین تبدیل های معمول این تبدیل نتیجه بهتری داشته است.

حال با دیتای جدید مدل خطی فیت می کنیم و ضرایب رگرسیون و نمودار Q-Q باقی مانده ها را نیز رسم می کنیم.




Regression Coefficients, significance, t-values and p-values :

| | Estimate | SE | tStat | pValue |
|---|---|---|---|---|
| **(Intercept)** | 0.0023566 | 3.7866e-05 | 62.236 | 0 |
| **x1** | -8.5581e-05 | 4.7491e-06 | -18.021 | 1.1827e-70 |
| **x2** | -1.8578e-05 | 7.12e-06 | -2.6092 | 0.0090981 |

با توجه به ضرایب می بینیم که وابستگی نسبت به متغیر ها تا حد زیادی حذف شده است. مقادیر پی ولیو نیز معنا دار بودن ضرایب را تایید می کنند.

## بخش ششم:

الف) با توجه به آنکه متغیر های دیسپل سایز و ترینینگ دیوریشن، به صورت دستی و عمدی تعیین شده و مستقل هستند، می توان گفت که این مدل Fixed effect است و رندوم نیست.

به دلیل استقلال متغیر ها همان یکبار پیاده سازی **ANOVA** کافی است.

ب) آنووا را پیاده میکنیم. نتایج به شکل زیر است:

**Analysis of Variance**

| Source | Sum Sq. | d.f. | Mean Sq. | F | Prob>F |
|---|---|---|---|---|---|
| X1 | 1.91936e+07 | 3 | 6397874.5 | 192.9 | 3.74193e-119 |
| X2 | 1.4837e+07 | 3 | 4945657.4 | 149.12 | 4.92682e-93 |
| Error | 1.89113e+08 | 5702 | 33166.1 | | |
| Total | 2.22004e+08 | 5708 | | | |

Constrained (Type III) sums of squares.

متغیر اول دیسپلی سایز و متغیر دوم ترینینگ دیوریشن است. با توجه به مقادیر p-values، و بسیار کوچک بودن آن ها، معنا دار بودن و اهمیت دار بودن متغیر ها به لحاظ آماری دیده می شود.

ج) آنالیز را به هر سه روش انجام می دهیم:

Tukey HSD Post-Hoc Comparison:

| Group1 | Group2 | Difference | pValue |
|---|---|---|---|
| 1 | 2 | -64.268 | 6.0121e-12 |
| 1 | 3 | -125.64 | 0 |
| 1 | 4 | -170.47 | 0 |
| 2 | 3 | -78.646 | 3.868e-19 |
| 2 | 4 | -123.47 | 0 |
| 3 | 4 | -62.494 | 1.4314e-09 |

Scheffe Post-Hoc Comparison:

| Group1 | Group2 | Difference | pValue |
|---|---|---|---|
| 1 | 2 | -65.776 | 6.6159e-11 |
| 1 | 3 | -127.17 | 6.5821e-55 |
| 1 | 4 | -172.03 | 1.0567e-102 |
| 2 | 3 | -80.193 | 2.6999e-17 |
| 2 | 4 | -125.05 | 8.1858e-49 |
| 3 | 4 | -64.09 | 1.099e-08 |

Bonferroni Post-Hoc Comparison:

| Group1 | Group2 | Difference | pValue |
|---|---|---|---|
| 1 | 2 | -64.733 | 7.6171e-12 |
| 1 | 3 | -126.11 | 1.5738e-56 |
| 1 | 4 | -170.95 | 1.3845e-104 |
| 2 | 3 | -79.123 | 1.984e-18 |
| 2 | 4 | -123.96 | 2.1934e-50 |
| 3 | 4 | -62.987 | 1.5797e-09 |

در هر سه متود که میانگین گروه ها با یکدیگر مقایسه می شوند، در همه جفت ها گروه ها به صورت معنا داری می توانند تفیکیک شوند که این مسئله را مقادیر p-values نشان می دهند.

## بخش هفتم:

الف) چون که سابجکت ها محدود و غیر رندوم هستند، و به طور عمدی و انتخابی برای آزمایش استفاده شده اند، مدل Fixed effect است. همچنین همه سابجت ها همه ی آزمایش ها را به صورت یکسان انجام داده اند پس مدل Repeated measures است.

ب) تحلیل آنووا را با استفاده از دو متغیر مستقل قبلی و همچنین متغیر سابجکت جدید اجرا می کنیم. نتایج به شکل زیر هستند.

**Analysis of Variance**

| Source | Sum Sq. | d.f. | Mean Sq. | F | Prob>F |
|---|---|---|---|---|---|
| X1 | 1.86731e+07 | 3 | 6224362.8 | 190.36 | 1.20696e-117 |
| X2 | 1.47118e+07 | 3 | 4903920.4 | 149.97 | 1.51026e-93 |
| X3 | 2.76335e+06 | 3 | 921115.7 | 28.17 | 4.47732e-18 |
| Error | 1.8635e+08 | 5699 | 32698.6 | | |
| Total | 2.22004e+08 | 5708 | | | |

Constrained (Type III) sums of squares.

متغیر اول دیسپلی سایز، متغیر دوم ترینینگ دیوریشن و متغیر سوم سابجکت است. با توجه به مقادیر p-values، و بسیار کوچک بودن آن ها، معنا دار بودن و اهمیت دار بودن متغیر ها به لحاظ آماری دیده می شود.