جزوه جلسه هفتم (نیمه دوم جلسه) دوره آمار با R دکتر مجید صابری

گردآورنده: شهرزاد احمدکرجی

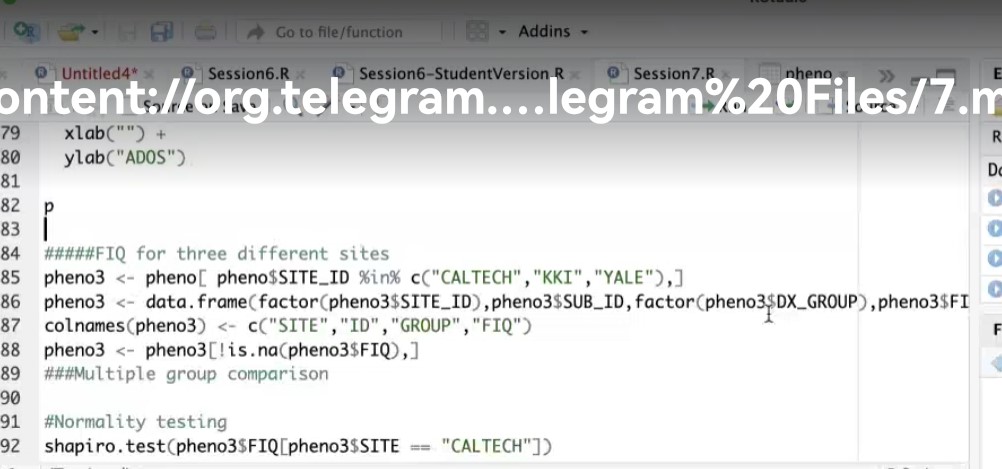
در ادامه نیمه اول این جلسه به سراغ مقایسه سه گروهی و چند متغیره می رویم. همانطور که قبلا توضیح داده شد، بر اساس اینکه متغیر وابسته ما چه تابع توزیعی دارد، که اگر تابع توزیع یکی از این­ها نرمال نیست، ما باید به اجبار به سراغ آمار ناپارامتریک برویم.

مثال: دو متغیر مستقل (گروه(سالم یا کنترل) و سایت(مکان جمع آوری دیتا)) و یک متغیر وابسته (FIQ) داریم. میخواهیم FIQ را براساس سایت مقایسه کنیم.

برای مثال در خط 85 دستور کد به این شکل است که ردیف هایی را برای من انتخاب کن که سایت ID داخل یکی از آن ها باشد:

اگر این دستور اجرا شود، فنوتیپیک ID هایی را می دهد که فقط مربوط به این سه دسته باشد.

در ادامه یک Dataframe جدید را می توانیم داشته باشیم که ستون هایی را انتخاب کرده و بعد از این ستون های انتخابی را با یک Dataframe با همjoin کرده و یکDataframe جدید را بسازیم ( خط 86).



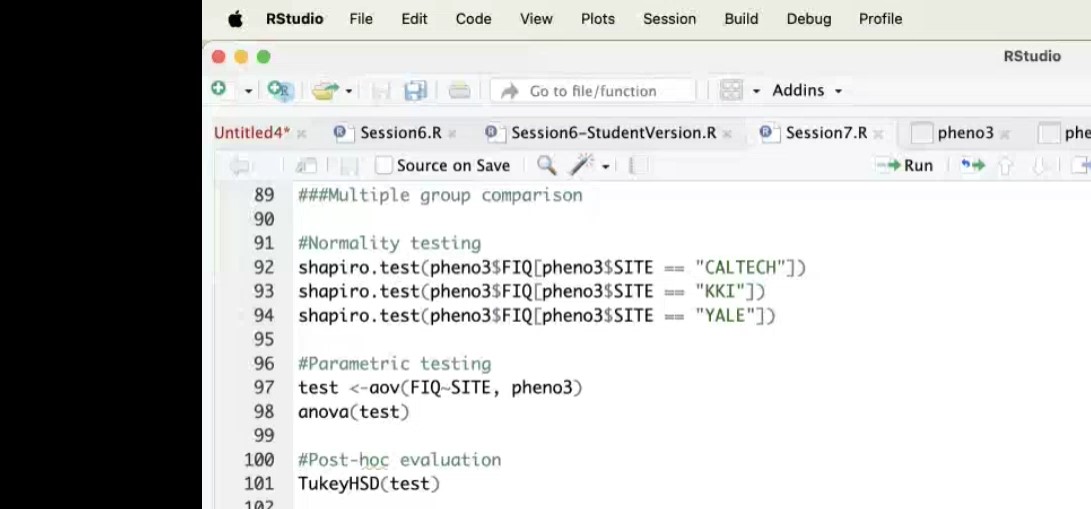
مثلا ستون های متفاوت SITE\_ID و SUBJECT\_ID را میتوانیم انتخاب کنیم.

نکته: با دستور factor بعضی از ستون ها که می خواهیم عدد در نظر گرفته نشوند را مشخص میکنیم. مثلا Group در اینجا مقدار 1 و 2 را دارد. این مقدارnumeric است، ما میدانیم که این مقدار عدد نیست و صرفا یکLable است. پس در انجام تست ها اگر عدد دیده شود ممکن است به اشتباه بجای آنکه 2 گروه را مقایسه کند، از linear regression استفاده کند. بنابراین جنس ورودی ها بسیار مهم است.

###Multiple group Comparison

#Normality testing

برای بررسی پارامتریک یا ناپارامتریک بودن، باید هر 3 Subject به طور جداگانه بررسی شده (یعنی ازShapiro-Wilk Normality Test) استفاده شود؛ حتی اگر یکی از گروه ها، تابع توزیع نرمال نداشته باشد، باید به سراغ آمار ناپارامتریک رفت.

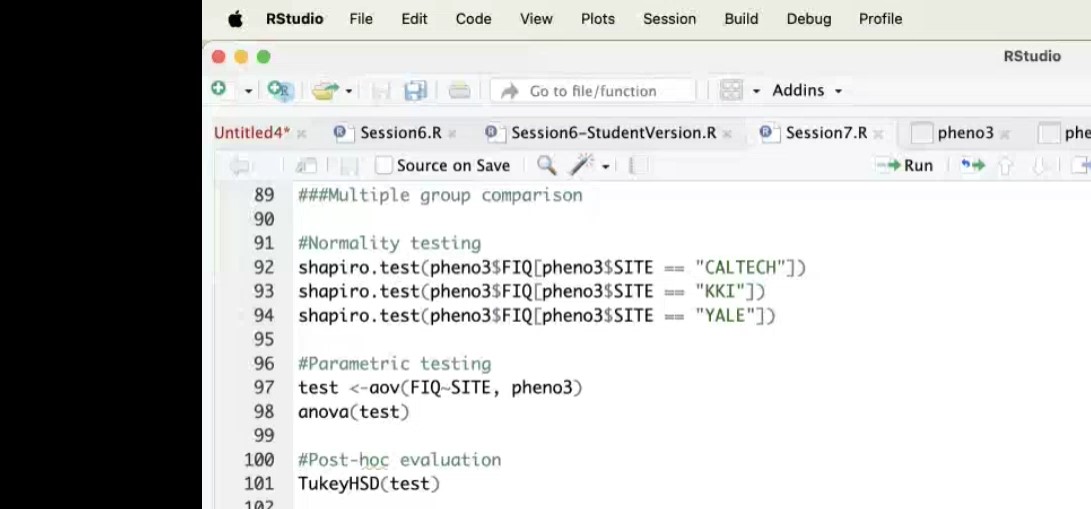


میتوانید از جدول راهنمای زیر استفاده کنید:

|  |  |  |  |
| --- | --- | --- | --- |
|  | تعداد گروه ها | مقایسه میانگین ها | Post-hoc evaluation |
| پارامتریک | 3 گروهی | One way Anova | Turkey HSD (Test) |
| ناپارامتریک | 3 گروهی | [Kruskal–Wallis](https://en.wikipedia.org/wiki/Kruskal%E2%80%93Wallis_test) | DunnTest |

#Prametric testing

برگردیم به مثال، بر فرض اگر از آمار پارامتریک استفاده گردد، از آزمون One way Anova به این شیوه استفاده می شود:

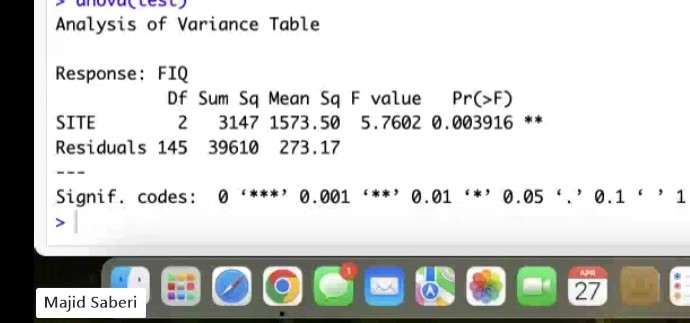


نکته: متغیر وابسته براساس متغیر مستقل بررسی می شود. به این معنی است که FIQ\_SITE، Full IQ را براساس سایت مقایسه میکند. در صورتی که P-Value معنی دار شود، این به این معنا است که یک تفاوت معنا دار وجود دارد. فرض صفر در اینجا، این بوده است که هیچ تفاوتی بین میانگین گروه 1 و 2 و 3 وجود ندارد. حال که فرض صفر رد شده است، P-Value معنا دار می شود.

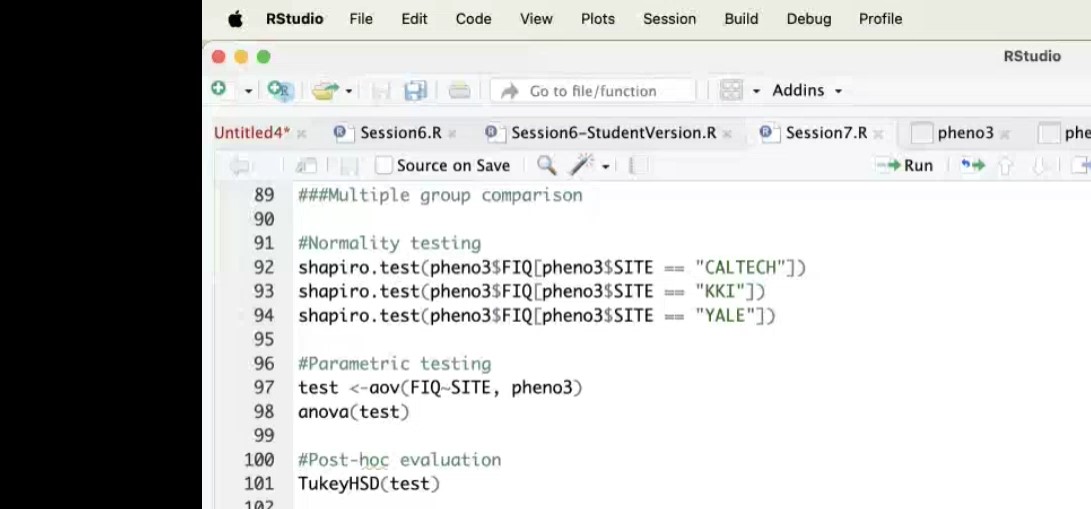
d= 3-1=2

آماره در T-Test : T-Value

در اینجا، آماره F-Value بوده و عدد آن از 0 به بالا است و هر چقدر بالاتر باشد، به معنای Effect Size بیشتر و تست معناداراست.

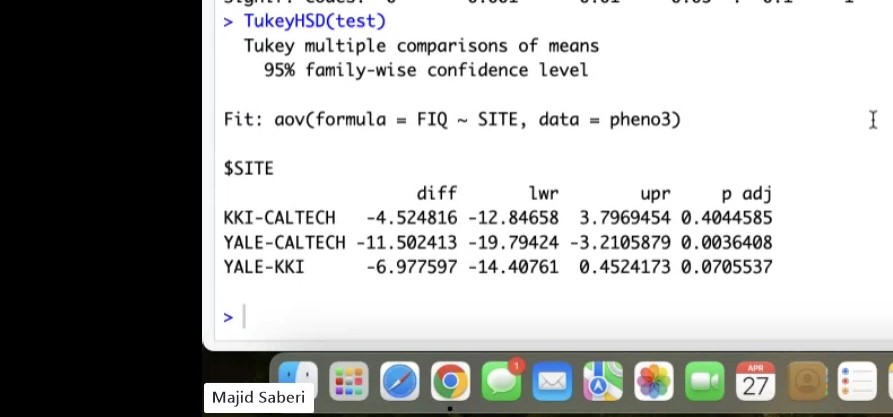


حال، در صورت معنا دار شدن تفاوت بین گروه ها ، برای بررسی دقیق تر از Post-hoc evaluation استفاده میکنیم که طبق جدول باید از TurkeyHSD Test استفاده شود:



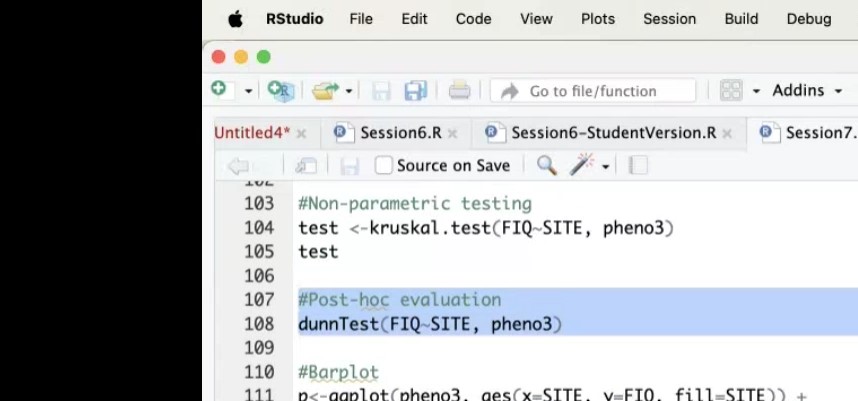
در واقع این دستور، گروه ها را 2 به 2 با یکدیگر از طریق T-Test مورد مقایسه قرار میدهد. در صورتی که P-Value <0.05 باشد، تست معنا دار است.

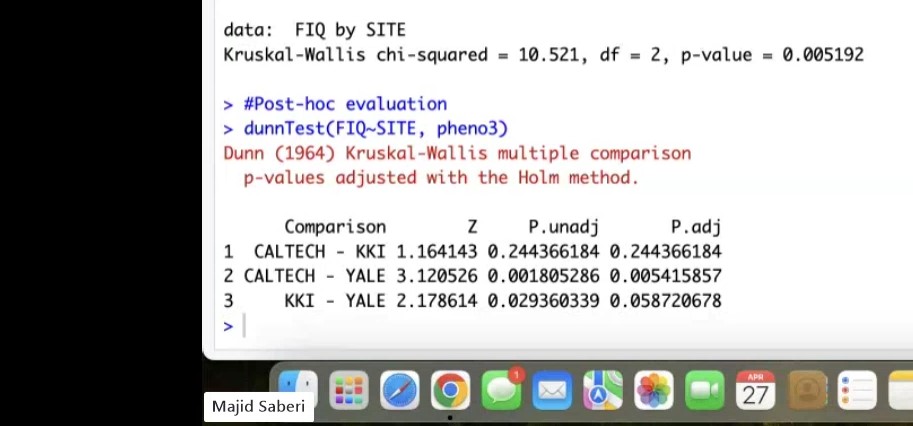
نکته: هر وقت چند مقایسه داشتیم، باید از P-Value ای که Adjust شده است استفاده کنیم تا اثر Multiple Comparison را از بین ببریم. تست Turkey ، P-Value adjust را به شما میدهد.



#Nonparametric testing

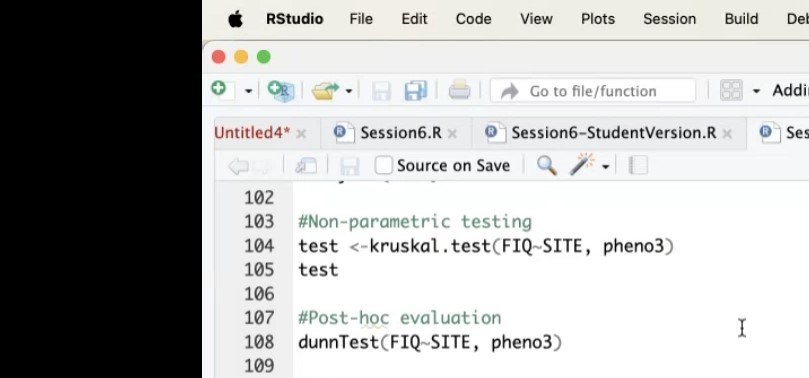
برگردیم به مثال، حال اگر بر فرض اگر از آمار ناپارامتریک استفاده گردد، از آزمون Kruskal–Wallis به این شیوه استفاده می شود:

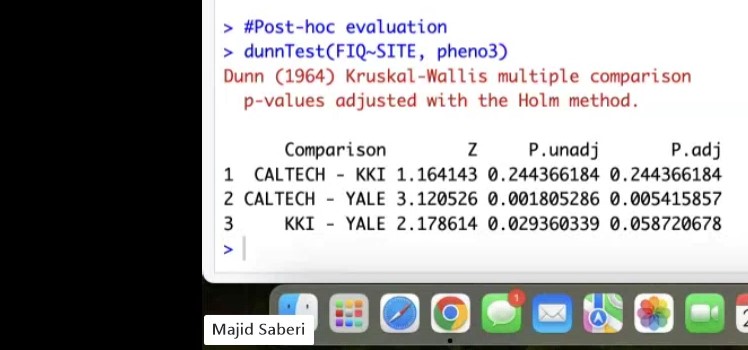




نکته: در اصل برای این مثال باید از آمار ناپارامتریک استفاده میشد بنابراین این p-value مورد استناد است.

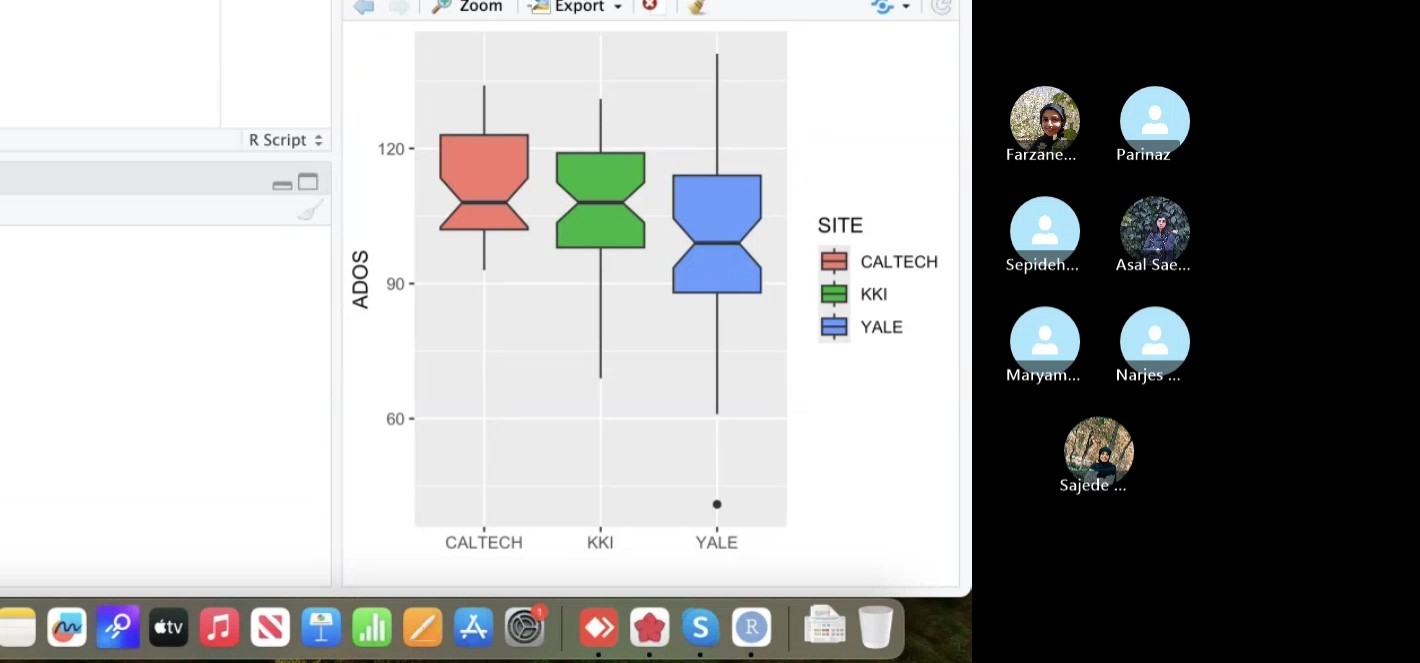
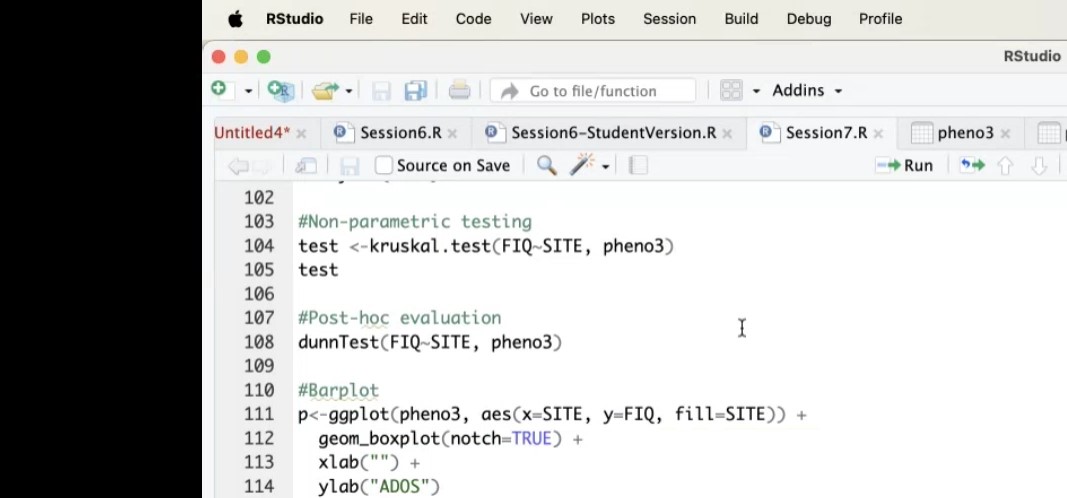
برای Post-hoc evaluation از DunnTest استفاده می شود. این تست به شما یک Z-Value و یک P-Value میدهد.





نکته» در آمار ناپارامتریک، همیشه از Boxplot استفاده میشود:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Index of central tendency | Index of dispersion | plot |
| پارامتریک | میانگین | Standard deviation | Bar plot |
| ناپارامتریک | میانه | Interquartile range | Box plot |



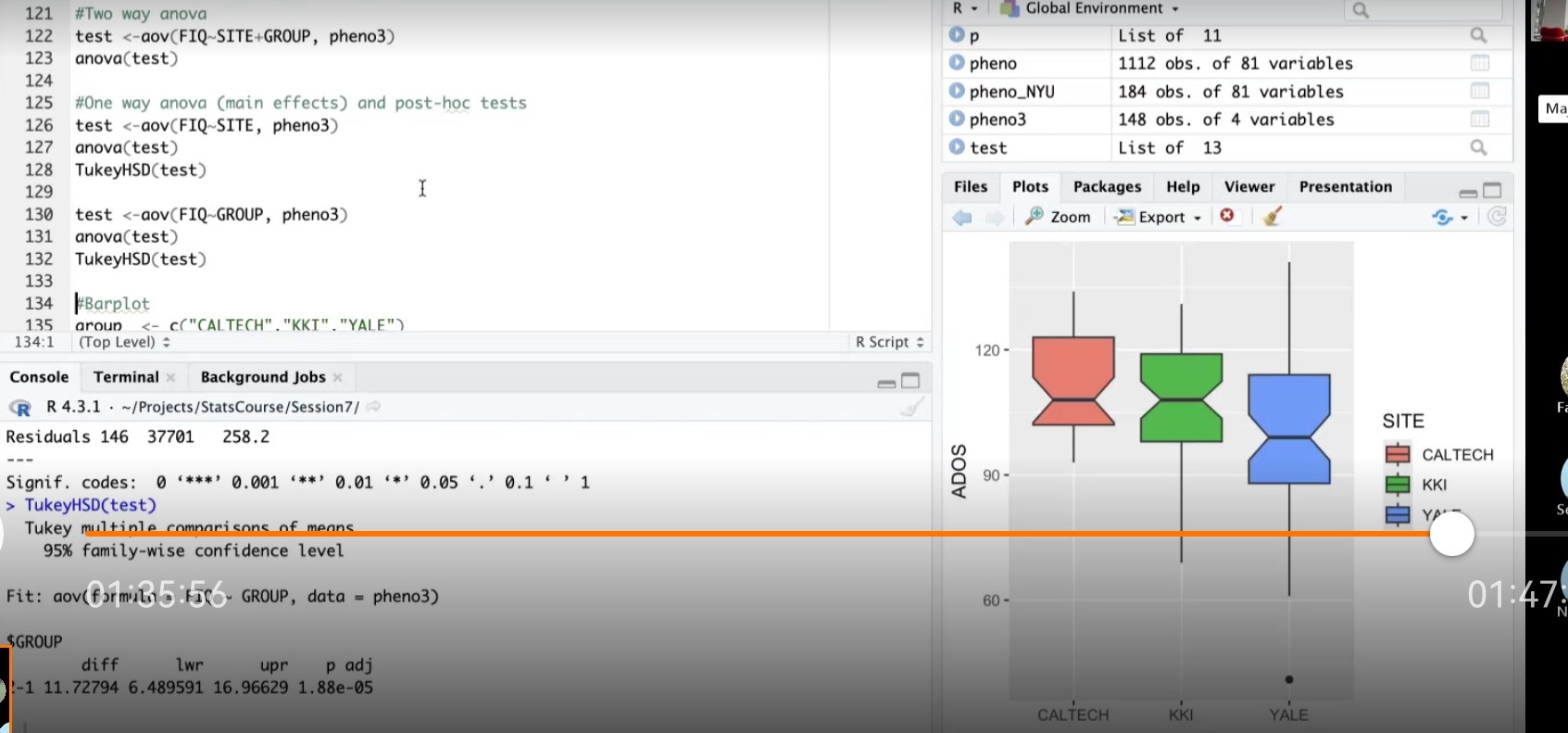
#Two Way Anova

حال تصور کنید که Two Way Anova داشته باشیم، یعنی متغیر مستقل 2 تا بیشتر باشد:

مثال: Full IQ بین سه سایت و در بین دو گروه سالم و ناسالم آیا متفاوت است؟

با فرض پارامتریک بودن، یک جدول 3 در 2 داریم که به این شکل میتوان خلاصه کرد :

Two Way -> One Way -> TurkeyHSD Test



با فرض ناپارامتریک بودن، عموما توصیه میشود که Two Way نگاه نکنید و با Kruskal–Wallis جلو بروید و در ادامه BoxPlot رسم کنید.

