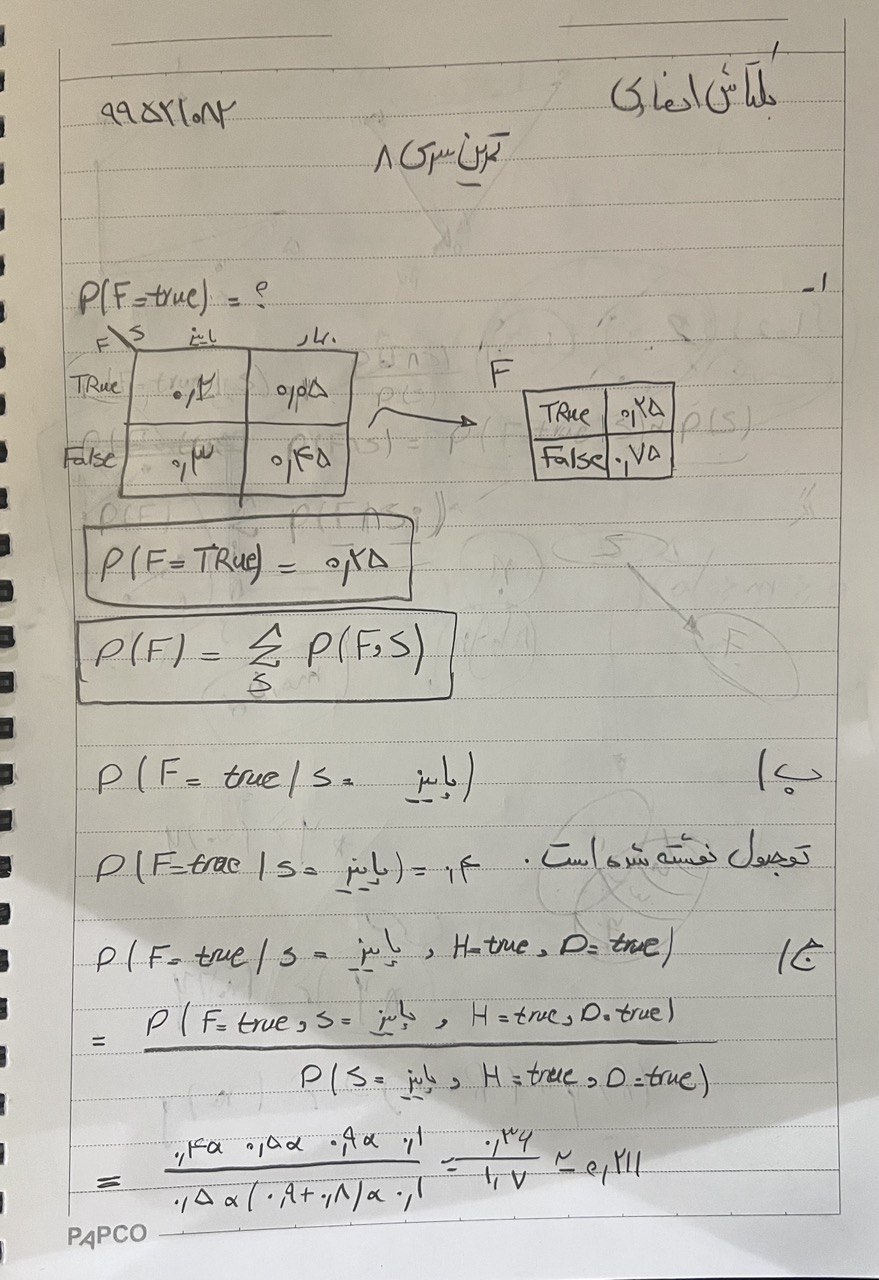
جواب تمرین سری 8

بکتاش انصاری

99521082

سوال اول :



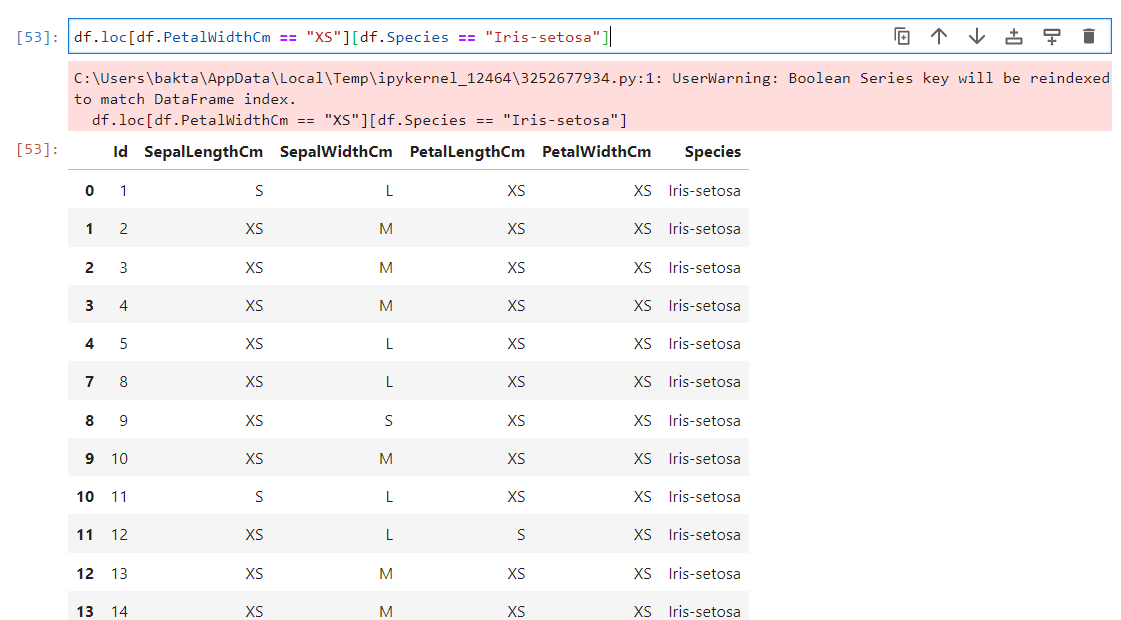
سوال 2 :

ابتدا کوئری ای که بخش اول به ما داده است را در کد پیاده سازی میکنیم تا تمامی حالت هایی که PedalWidth = XS و Species = Iris-setosa را از جدول استخراج کنیم.

برای این کار از دستور df.loc[df.PetalWidthCm == "XS"][df.Species == "Iris-setosa"]

استفاده میکنیم.

که نتیجه آن برابر است با :



حال برای محاسبه احتمال آن یک len() از عبارت بالا میگیریم تا تعداد instance های مورد نظر را بدست آوریم که این مقدار برابر 34 است و تعداد کل خانه هایی را که Species آن ها برابر با Iris-setosa هست نیز برابر 50 است (احتمال شرطی برابر است با اشتراک دو عبارت بر روی عبارت دوم) که احتمال آن برابر است با : p = 34/50 = 0.68.

بخش دوم :

مانند بخش اول ابتدا کوئری مورد نظر را ران میکنیم تا instance های مورد نظر بدست آیند.

که این کوئری برابر است با :

df.loc[df.SepalLengthCm == "L"][df.Species == "Iris-virginica"]

که مقدار len آن را نیز میگیریم که مقدار 31 به ما میدهد.

برای مخرج احتمال نیز باید حالت هایی که Species برابر با Iris-viriginica میباشد را محاسبه کنیم

که کوئری آن نیز برابر است با :

df.loc[df.SepalLengthCm == "L"]

که len آن را نیز محاسبه میکنیم و مقدار 42 را به ما میدهد.

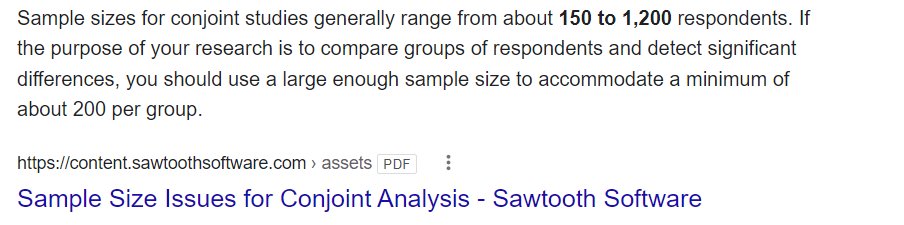
که احتمال ما برابر است با :

P = 31/42



بخش آخر :

طبق سرچ هایی که انجام دادم به این نتیجه رسیدم که :



که با توجه به توضیحات برای آنکه ما بتوانیم یک جدول توزیع احتمال توأم مناسب ایجاد کنیم باید از هر گروه از داده ها حداقل 200 رکورد داشته باشیم که در این سوال از هر گونه از گیاه حداقل 200 رکورد باید موجود باشد که در اینجا ما این مقدار را نداریم در نتیجه نمیتوان جدول توزیع احتمال توأم مناسبی ایجاد کرد.