

جواب تمرین سری 8

بکتاش انصاری

99521082

سوال اول :

جواب تمرین سری 8

بکتاش انصاری

99521082

تمرین سری 8

1- $P(F=true) = ?$

جدول احتمالات:

F \ S	بله	خیر
True	0.12	0.05
False	0.13	0.45

جدول حاشیه‌ای:

F	P(S)
True	0.17
False	0.58

$P(F=True) = 0.17$

$P(F) = \sum_S P(F, S)$

جواب:

ب) $P(F=true | S = \text{بله})$

توجه: فرض شده است. $P(F=true | S = \text{بله}) = 0.12$

ج) $P(F=true | S = \text{بله}, H=true, D=true)$

$= \frac{P(F=true, S = \text{بله}, H=true, D=true)}{P(S = \text{بله}, H=true, D=true)}$

$= \frac{0.12 \times 0.05 \times 0.9 \times 0.1}{0.12 \times (0.9 + 0.18) \times 0.1} = \frac{0.12}{1.17} \approx 0.1026$

PAPCO

سوال 2 :

ابتدا کوئری ای که بخش اول به ما داده است را در کد پیاده سازی میکنیم تا تمامی حالت هایی که $PetalWidth = XS$ 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 :

```
[53]: df.loc[df.PetalWidthCm == "XS"][df.Species == "Iris-setosa"]
```

C:\Users\bakta\AppData\Local\Temp\ipykernel_12464\3252677934.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
df.loc[df.PetalWidthCm == "XS"][df.Species == "Iris-setosa"]

```
[53]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	S	L	XS	XS	Iris-setosa
1	2	XS	M	XS	XS	Iris-setosa
2	3	XS	M	XS	XS	Iris-setosa
3	4	XS	M	XS	XS	Iris-setosa
4	5	XS	L	XS	XS	Iris-setosa
7	8	XS	L	XS	XS	Iris-setosa
8	9	XS	S	XS	XS	Iris-setosa
9	10	XS	M	XS	XS	Iris-setosa
10	11	S	L	XS	XS	Iris-setosa
11	12	XS	L	S	XS	Iris-setosa
12	13	XS	M	XS	XS	Iris-setosa
13	14	XS	M	XS	XS	Iris-setosa

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-virginica میباشد را محاسبه کنیم که کوثری 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 :

$$P = 31/42$$

```
[57]: len(df.loc[df.SepalLengthCm == "L"])
[57]: 42

[58]: len(df.loc[df.SepalLengthCm == "L"][df.Species == "Iris-virginica"])
C:\Users\bakta\AppData\Local\Temp\ipykernel_12464\2182852149.py:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  len(df.loc[df.SepalLengthCm == "L"][df.Species == "Iris-virginica"])
[58]: 31
```

بخش آخر :

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

Sample sizes for conjoint studies generally range from about **150 to 1,200** respondents. If the purpose of your research is to compare groups of respondents and detect significant differences, you should use a large enough sample size to accommodate a minimum of about 200 per group.

[https://content.sawtoothsoftware.com > assets](https://content.sawtoothsoftware.com/assets/12464/2182852149.py) PDF

[Sample Size Issues for Conjoint Analysis - Sawtooth Software](https://content.sawtoothsoftware.com/assets/12464/2182852149.py)

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.