

## ▼ HW06 (제출기한-6월7일자정)

Datascience 모듈의 Table 연습을 위한 문제입니다.

- 반드시 수업시간 배운 내용만을 사용하여 코드를 작성합니다.
- 한 셀에 여러줄의 코드를 사용해도 무방합니다.

제출시 다음 사항에 유의하기 바랍니다.

- 텍스트 셀에 설명되어 있는 각 문항을 잘 읽은 뒤, '답안을 작성하시오'라고 적힌 코드 셀에 적절한 코드를 작성합니다.
- '지우지 마시오'라고 적힌 코드 셀은 절대로 지우면 안 됩니다.
- 작성 후 파일명에 학번을 자신의 학번으로 고친 후 ULMS 해당과제 제출란에 업로드합니다.

```
1 from datascience import *
2 import matplotlib
3 %matplotlib inline
4 import matplotlib.pyplot as plots
5 plots.style.use('fivethirtyeight')
6 import numpy as np
7 np.random.seed(0)
8
9 fpath='https://raw.githubusercontent.com/data-8/textbook/878363106db4313de925a709b1617d6f328c56c'
```

```
/usr/local/lib/python3.7/dist-packages/datascience/tables.py:17: MatplotlibDeprecationWarning:
  matplotlib.use('agg', warn=False)
/usr/local/lib/python3.7/dist-packages/datascience/util.py:10: MatplotlibDeprecationWarning:
  matplotlib.use('agg', warn=False)
```

## ▼ 다음 데이터를 이용하자.

```
1 baby = Table.read_table(fpath+'baby.csv')
2 baby
```

## ▼ 문제1

산모체중(*Maternal Pregnancy Weight*) 평균을 구하여 `maternal_weight_mean`에 저장하시오.

```
1 ### 답안을 작성하시오.  
2  
3 maternal_weight_mean=np.mean(baby.column('Maternal Pregnancy Weight'))  
4
```

```
1 ### 지우지 마시오.  
2 maternal_weight_mean
```

128.4787052810903

## ▼ 문제2

위 `baby` Table에서 100명의 산모들을 `sampling`하여 `baby_sample` 변수에 저장하시오.

(단, `replacement=False`)

```
1 ### 답안을 작성하시오.  
2  
3 baby_sample=baby.sample(100)  
4
```

```
1 ### 지우지 마시오.  
2 baby_sample
```

### ▼ 문제3

위 `baby_sample` 을 10000번 Bootstrapping 하여 계산한 평균값 결과 `array`를 `bstrap_mean` 이라는 변수에 저장하시오.

```
1 ### 답안을 작성하시오.
2 bstrap_mean=np.empty(10000)
3 for i in np.arange(10000):
4     bstrap_mean[i]=np.mean(baby_sample.sample().column('Maternal Pregnancy Weight'))
```

```
1 ### 지우지 마시오.
2 bstrap_mean
```

```
array([131.86, 130.21, 130.56, ..., 129.08, 126.63, 129.38])
```

### ▼ 문제4

위 `bstrap_mean` 배열로부터 95% 신뢰도를 갖는 평균값 구간을 추정하여 `(left, right)` `array`에 저장하시오.

```
1 ### 답안을 작성하시오.
2 left = percentile(2.5, bstrap_mean)
3 right = percentile(97.5, bstrap_mean)
```

```
1 ### 지우지 마시오.
2 (left, right)
```

```
(124.66, 133.66)
```

### ▼ 문제5

위에서 구한 `bstrap_mean`에 대하여 histogram을 그리고, `(left, right)`에 관한 노란색 선을 그리시오. (수업시간 내용 그대로 적용)

(단, `bins`의 개수는 20개로 설정한다.)

```
1 ### 답안을 작성하시오.
2 Table().with_column('Bootstrap Sample Mean',bstrap_mean).hist(bins=20)
3 plots.plot(make_array(left, right), make_array(0, 0), color='yellow', lw=10);
```

다음 문제6부터 문제10을 위해 동전의 앞뒤를 나타내는 Table 변수 coin를 이용하자.

```
1 coin = Table().with_column('Coin', ['Head', 'Tail'])
2 coin
```

## ▼ 문제6

동전을 n번 던진 후 동전의 앞면이 나오는 비율을 리턴하는 함수 ratio\_head를 작성하시오.  
(위 coin Table을 이용하여 sampling한다)

```
1 def ratio_head(n):
2     ### 답안을 작성하시오.
3     return coin.sample(n, ("Head")).column('Coin')
```

```
1 ### 지우지 마시오.
2 ratio_head(20)
```

## ▼ 문제7

위 `ratio_head` 함수를 이용하여 동전을 20번 던진 후 앞면(Head)이 나오는 비율을 구하는 시행을 10000번 시행하여 그 값들을 `ratio_20` 배열 변수에 저장하시오.

```
1 ### 답안을 작성하시오.  
2 ratio_20=np.empty(10000)  
3 for i in np.arange(10000):  
4     ratio_20[i]=ratio_head(20)
```

```
1 ### 지우지 마시오.  
2 ratio_20
```

```
array([0.4 , 0.65, 0.45, ..., 0.45, 0.65, 0.25])
```

## ▼ 문제8

위 문제와 마찬가지로 위 `ratio_head` 함수를 이용하여 동전을 **100**번 던져 앞면(Head)가 나온 비율을 구하는 시행을 10000번 시행하여 그 값들을 `ratio_100` 배열 변수에 저장하시오.

```
1 ### 답안을 작성하시오.  
2  
3  
4  
5
```

```
1 ### 지우지 마시오.  
2 ratio_100
```

```
array([0.46, 0.53, 0.44, ..., 0.51, 0.49, 0.51])
```

## ▼ 문제9

위 `ratio_20`, `ratio_100` 배열을 `column`으로 지정하여 `results` 라는 `Table`을 생성하시오.

(단, 각 `column`의 이름은 `ratio_20` 과 `ratio_100` 으로 정하시오.)

```
1 ### 답안을 작성하시오.  
2  
3  
4  
5
```

```
1 ### 지우지 마시오.  
2 results
```

## ▼ 문제10

위 `results` `Table`의 두 `column`을 중첩하여 `histogram`을 그리시오.

(단, `bins`는 `np.arange(0.1,0.91,0.05)`로 설정한다.)

```
1 ### 답안을 작성하시오.  
2  
3  
4
```

