**실전 연습문제 세트 1번**

글로벌 유통 회사에서 매출액을 증대시킬 수 있는 마케팅 방법을 찾기 위하여, 채널별 반응과 매출금액과의 관계를 분석하고자 한다.

**01\_ADS\_sample\_1.csv** (구분자: comma(“,”), 169 Rows, 5 Columns, UTF-8 인코딩)

|  |  |  |
| --- | --- | --- |
| **컬럼** | **정의** | **Type** |
| YEAR\_MONTH | 년월, YYYY-MMM 형태 (2018-Jan) | String |
| SOCIAL | 일평균 SNS 조회수 | Double |
| TV | 일평균 TV광고 시청자수 | Double |
| NEWSPAPER | 일평균 신문기사 구독자수 | Double |
| SALES\_AMT | 매출금액 | Double |

**1. (ADS\_sample\_1.csv를 활용하여)** 년월(YEAR\_MONTH) 변수를 제외한 변수들에 대해 상관분석을 수행하고자 한다. 매출금액(SALES\_AMT)과 상관 정도가 가장 강한(상관계수 절대값이 가장 큰) 변수를 찾고 해당 변수와 매출금액(SALES\_AMT)과의 상관계수 값을 기술하시오.

* 피어슨(Pearson) 상관계수 값을 구한다.
* 상관계수는 절대값을 취하지 않은 상관계수 값 그대로, 소수점 넷째 자리에서 반올림하여 셋째 자리까지 기술하시오.

**(답안예시) 0.123**

**2. (ADS\_sample\_1.csv를 활용하여)** 2009년 매출금액(SALES\_AMT) 대비 2019년 매출금액(SALES\_AMT)의 변화 정도를 확인하고자 한다. 다음 수식을 이용하여 증감률(%)을 기술하시오.

증감률(%) =

* 연도는 년월(YEAR\_MONTH) 변수로부터 추출하며, 연도별 매출금액합계는 1월부터 12월까지의 매출 총액을 의미한다.
* 증감률(%)은 소수점 넷째 자리에서 반올림하여 셋째 자리까지 기술하시오.

**(답안예시) 1.234**

**3. (ADS\_sample\_1.csv를 활용하여)** 2009년과 2019년의 매출금액(SALES\_AMT) 평균이 통계적으로 유의한 차이가 있는지 검정하고자 한다. 적절한 검정 수행 후, 유의 확률(P-value)의 값을 기술하시오.

단, 검정 시 세운 대립 가설(H1)은 다음과 같다.

대립 가설(H1): 2009년과 2019년의 매출금액의 평균이 같지 않다

* 등분산 가정 하에서 검정을 수행한다.
* 2009년 매출금액(SALES\_AMT) 12건과 2019년 매출금액(SALES\_AMT) 12건에 대한 평균 검정으로, 해당 검정의 검정통계량은 자유도가 22인 t 분포를 따른다
* 유의 확률(P-value)값은 소수점 넷째 자리에서 반올림하여 셋째 자리까지 기술하시오.
* 툴별 가이드

|  |  |
| --- | --- |
| Brightics | 문제 지시 외 Default 값 사용 |
| R | t.test() 함수의 var.equal=T 옵션 사용  문제 지시 외 Default 값 사용 |
| Python | from scipy import stats |

**(답안예시) 0.123**

**실전 연습문제 세트 2번**

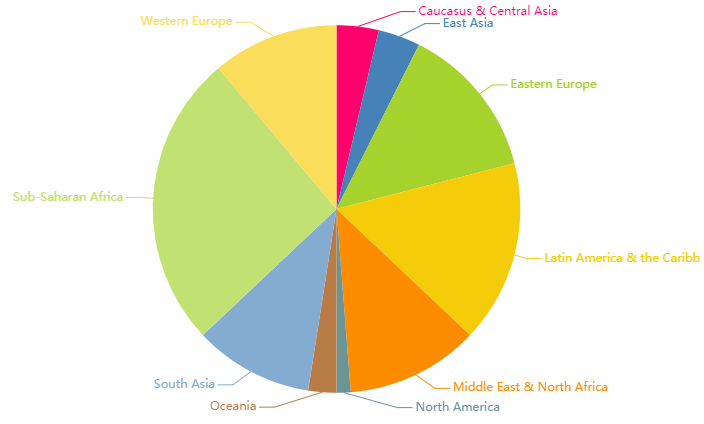
경제적 자유(economic freedom) 지수는 국가별로 정부, 법률, 돈, 무역, 규제와 같이 세분화된 항목으로 측정해 활용되고 있다. 각 지수들에 대한 분석을 시작해보자.

**01\_ADS\_Sample\_2.csv** (구분자: comma(“,”), 486 Rows, 9 Columns, UTF-8 인코딩)

|  |  |  |
| --- | --- | --- |
| **컬럼** | **정의** | **Type** |
| year | 연도 | Integer |
| ISO\_code | 국가의 ISO 코드 | String |
| region | 지역 | String |
| ef\_score | 경제적 자유(economic freedom)지수 | Double |
| ef\_government | 경제적 자유 지수\_정부 | Double |
| ef\_legal | 경제적 자유 지수\_법률 | Double |
| ef\_money | 경제적 자유 지수\_돈 | Double |
| ef\_trade | 경제적 자유 지수\_무역 | Double |
| ef\_regulation | 경제적 자유 지수\_규제 | Double |

**4. (ADS\_Sample\_2.csv를 활용하여)** 2015년 데이터에서 각 국가들이 속한 지역(region)의 비율을 살펴보던 중, 가장 많은 국가(ISO\_code)가 속한 지역(region)이 ‘Sub-Saharan Africa’라는 것을 확인하였다. 이 지역에 속한 국가(ISO\_code)의 수는 전체 국가 수(162개) 중 몇 %인가?

**[2015년 국가들의 지역(region) 비율]**



비율(%)은 0~100사이의 숫자이며, 소수점 아래는 버리고 정수만을 기술하시오

**(답안예시) 12**

**5. (ADS\_Sample\_2.csv를 활용하여)** 경제적 자유지수(ef\_score)가 결정되는 요인을 설명하기 위해 다음 단계로 회귀모형을 생성하고 질문에 답하시오.

1단계)

2015년 데이터를 학습데이터로 한다. 학습데이터의 6개 변수(ef\_score, ef\_government, ef\_legal, ef\_money, ef\_trade, ef\_regulation)에 대해, 결측치가 있는 행을 삭제한다.

2단계)

1단계를 거친 학습데이터로 선형회귀모형(Linear Regression)을 학습한다.

* 종속 변수: ef\_score
* 설명 변수(5개): ef\_government, ef\_legal, ef\_money, ef\_trade, ef\_regulation
* 툴별 가이드

|  |  |
| --- | --- |
| Brightics Studio | 문제 지시 외 Default 값 사용  (Fit Intercept=True, VIF=True) |
| R | lm() 함수 사용  문제 지시 외 Default 값 사용 |
| Python | from sklearn.linear\_model import LinearRegression  LinearRegression에서 fit\_intercept=True  문제 지시 외 Default 값 사용 |

각 설명 변수의 증감에 따라 종속 변수의 변동폭이 가장 큰 설명 변수(즉, 설명 변수가 1 커질 때 종속 변수 변화량이 가장 큰 설명 변수)가 무엇인지 확인하고, 그때의 회귀계수를 구하시오.

* 각 설명 변수가 증가되거나 감소할 때, 해당 설명 변수 외에 다른 설명 변수는 고정된   
  상태라고 간주한다.
* 회귀계수는 소수점 넷째 자리에서 반올림하여 셋째 자리까지 기술하시오.

**(답안예시) 1.234**

6.문제 5에서 생성한 회귀모형을 2016년데이터에 대입하여 경제적 자유지수(ef\_score)를 예측한 후 질문에 답하시오.

1단계) 예측 값과 실제 값을 활용하여 오차(error)의 절대값을 계산한다.

* 오차(error)는 경제적 자유지수(ef\_score)의 예측된 값()과 실제 값()의 차이를 의미한다.

계산된 오차의 절댓값의 총합()을 기술하시오.

* 오차의 절댓값의 총합은 소수점 넷째 자리에서 반올림하여 셋째 자리까지 기술하시오.

**(답안예시) 0.123**

**실기툴별 정답**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **문제** | **배점** | **Brightics** | **Python** | **R** |
| **1** | 5 | **-0.183** | **-0.183** | **-0.183** |
| **2** | 10 | **0.194** | **0.194** | **0.194** |
| **3** | 10 | **0.991** | **0.991** | **0.991** |
| **4** | 10 | **25** | **25** | **25** |
| **5** | 5 | **0.202** | **0.202** | **0.202** |
| **6** | 10 | **1.702** | **1.702** | **1.702** |

\* 시험환경에서 실행한 정답 기준으로 작성하였으며, 정답은 문항에 따라 사용 라이브러리/패키지의 버전 차이로 달라질 수 있습니다.