**DFMBA 경영통계 – 4th Assignmnet (기말고사 대체과제)**

학번 : 20224071

이름 : 오태건

※ 가설검정이 필요한 문제는 *가설, 검정통계량의 관찰값, 유의확률(p-value) 산출값, 결론*을 명확히 작성해주세요.

※ R을 이용하여 도출된 결과는 답안에 R 코드와 출력화면이 포함되도록 작성해주세요.

1. 어느 투자자문회사에서 운영하는 포트폴리오의 1-기간 단순수익률의 평균이 3 %를 넘는지를 확인하고자 한다. 이 회사에서 운영하는 포트폴리오 중 15개를 랜덤하게 뽑아 조사해본 결과 평균이 3.2 %, 표준편차는 % 였다고 한다. 이를 토대로 유의수준 1%에서 검정하여라. 단, 이 회사의 포트폴리오의 1-기간 단순수익률은 정규분포를 따른다고 가정할 것. (15점)

|  |  |
| --- | --- |
|  | > 1 - pt(0.5477, 14)  [1] 0.2962617 |

2. 포트폴리오에 채워지는 금융상품 수익률의 표준편차가 3%보다 작아야만 금융상품 수익률이 안정적이라고 말할 수 있다고 하자. 이를 기준으로 어느 포트폴리오의 안정성을 살펴보고자, 이 포트폴리오에 포함된 금융상품 중 10개를 임의로 추출하여 수익률을 측정한 결과 표본 표준편차는 1.5%였다. 이 포트폴리오에 포함된 금융상품의 수익률은 정규분포를 따른다고 할 때, 이 포트폴리오는 안정적이라고 말할 수 있는지를 유의수준 5%에서 검정하여라. (15점)

|  |  |
| --- | --- |
|  | > pchisq(2.25, 9)  [1] 0.01313062 |

3. 모평균이 인 어느 모집단에서 추출한 30개 확률 표본의 평균을 이용하여 모평균이 보다 작은가에 대한 가설검정을 수행하였을 때, 귀무가설 의 유의수준 2%에서 기각되지 않았다고 하자. 다음 물음에 '예', '아니오', 혹은 '알수 없음'으로 답하시오. (12점)

1. 귀무가설 는 유의수준 1%에서 검정하였다고 해도 기각되지 않았을 것인가? \_\_\_\_\_\_\_\_\_
2. 이 검정의 p-value는 0.02보다 큰 값인가? \_\_\_\_\_\_\_\_\_
3. 실제 귀무가설 가 사실인가? \_\_\_\_\_\_\_\_\_
4. 동일한 조건으로 30개의 표본을 다시 추출하여 유의수준 2%로 검정한다면 귀무가설 이 기각되지 않는다는 결론은 달라질 수도 있는가? \_\_\_\_\_\_\_\_\_

4. 다음은 단순 선형 회귀 모형(Simple Linear Regression)에 관한 내용이다. 아래와 같이 절편 항을 포함한 모형을 가정하고, 의 자료를 이용하여 최소자승법(Least square Method)에 의해 모형을 적합한다고 하자. 추정된 회귀선을 , 잔차 는 라고 할 때, 다음 설명 중 맞는 것은 T, 틀린 것은 F로 답하시오. (12점)



(1) 위 모형에서 회귀 계수 (또는 독립 변수 )의 유의성에 대한 t-검정과 모형 전체에 대한 유의성 검정인 F-검정을 하는 경우, 두 검정의 유의확률은 동일하다. \_\_\_\_\_\_

(2) 의 변동성이 클 수록, 최소자승 추정량인 의 변동성이 커져 추정의 정확도가 떨어진다. \_\_\_\_\_\_

(3) 위 모형의 추정결과를 이용하여 반응값에 대한 구간추정을 한다고 하자. 로 주어졌을 때, 에 대한 신뢰구간은, 에 대한 예측구간에 비해, 동일한 신뢰수준에서 구간의 크기가 항상 더 작다. \_\_\_\_\_\_

(4) 적합된 회귀선으로 구한 잔차의 평균은 0이다. \_\_\_\_\_\_

5. 어느 보험회사의 관리수행정도를 평가하기 위하여 10개의 지점을 무작위로 추출한 뒤 보험 판매건에 대한 마케팅 경비를 조사하였다. 이 자료를 이용하여 선형회귀모형을 적합하고자 할 때 다음 물음에 답하여라. (12점)

|  |  |  |
| --- | --- | --- |
| 지점번호 | 마케팅경비 | 보험판매수 |
| 1 | 28 | 220 |
| 2 | 12 | 170 |
| 3 | 45 | 260 |
| 4 | 30 | 180 |
| 5 | 33 | 200 |
| 6 | 65 | 290 |
| 7 | 75 | 300 |
| 8 | 80 | 450 |
| 9 | 24 | 120 |
| 10 | 50 | 200 |

1. 선형회귀모형의 절편과 기울기 계수 추정치를 구하여라.

|  |
| --- |
| > col1 <- c(1,2,3,4,5,6,7,8,9,10)  > col2 <- c(28,12,45,30,33,65,75,80,24,50)  > col3 <- c(220,170,260,180,200,290,300,450,120,200)  > df <- data.frame(col1, col2, col3)  > m <- lm(col3 ~ col2, df)  > coef( m )  (Intercept) col2  84.940823 3.485502 |

1. 종속변수의 변동 중 적합된 회귀식으로 설명되는 변동의 비율을 구하여라.

|  |
| --- |
| > summary(m)  Multiple R-squared: 0.7457 |

1. 마케팅 경비가 70이라고 할 때, 보험판매수는 얼마로 예상되는지 그 예측값을 구하여라.

|  |
| --- |
| > predict(m, newdata=data.frame(col2=70))  1  328.9259 |

6. 한 소형 대출업체는 고객들을 대상으로 그들의 현재 부채, 월소득, 그리고 나이를 고려하여 추가 대출 가능성이 있는 사람들을 연구해 오고 있다. 추가대출에 영향을 미치는 요인들을 추출하기 위하여 19개의 관찰치를 추출하여 아래와 같은 다중회귀모형을 설정한 뒤 추정하였다. (모형식의 는 추가대출금, 는 현재부채, 은 월 소득, 는 현재나이를 각각 의미하며, 괄호 안은 각 회귀계수 추정량에 대한 표준오차 추정값을 나타낸 것임) (10점)

1. 다른 변수 값이 고정되어 있을 때, 월 소득이 추가대출금에 미치는 영향이 유의한가를 5%의 유의수준에서 검정하여라.

|  |  |  |
| --- | --- | --- |
| |  | | --- | | 독립변수 월소득은 종속변수 추가대출금을 설명하기에 유용한 변수라고 해석할 수 있다 | | > 0.3883 / 0.136  [1] 2.855147  > (1-pt(2.855147, 15))\*2  [1] 0.01204045 |

1. 적합된 모형에 따르면, 현재부채가 2배로 늘어날 때, 추가대출은 어떤 영향을 받는가?

|  |  |  |
| --- | --- | --- |
| |  | | --- | | 독립변수 현재부채는 종속변수 추가대출금을 설명하기에 적합한 변수이다. 따라서, 현재부채가 2배 늘어날 때, 추가대출금은 -7.4787X1bar 만큼 줄어든다 | | > -7.4787/1.193  [1] -6.268818  > (1-pt(6.268818, 15))\*2  [1] 1.504383e-05 |

7. 30개의 관찰치를 이용하여, 9개의 설명변수가 포함된 다중 선형 회귀 모형을 적합하였더니, 오차항의 분산 의 추정값이 이고, 모형을 적합도를 나타내는 결정계수가 로 도출되었다고 한다. (12점)

1. 모형의 전반적인 유의성을 검정하기 위한 F 검정통계량의 관찰값을 구하여라.
2. 원 모형에서 사용한 9개의 설명변수 중 4개의 설명변수만 골라서 새로 다중 선형 회귀 모형을 적합하였더니 잔차제곱합이 2200으로 도출되었다고 한다. 새로 적합한 모형의 오차항의 분산 에 대한 추정값을 구하여라.
3. 원 모형과 (2)에서 정의한 새로운 모형 중 어느 것이 더 우월하며, 그 이유는 무엇인지를 설명하라.

|  |
| --- |
| 수정결정계수 정의에 따라 SSE/(n-k-1)는 추가된 독립변수가 종속변수를 설명하는데 기여하는 바가 큰 경우에만 감소하고, 그렇지 않은 경우는 증가하게 된다.  (2)에서 정의한 새로운 모형의 SSE가 더 작아 좀 더 우월한 모형으로 볼 수 있다. |

8. 어느 건식벽체를 제조업체의 사장은 자신의 제품 수요에 영향을 미치는 변수를 분석하려고 합니다. 이를 위하여 건식벽체의 월간 판매량(단위: 100개)인 Drywall을 종속변수로 하고 이를 아래와 같이 정의되는 독립변수들로 설명하는 다중회귀분석을 수행해보고자 한다. (12점)

- Permits : 카운티에서 발급된 건축 허가 수

- Mortgage : 5년 모기지 이자율 (단위: %)

- A Vacancy : 아파트 공실률 (단위: %)

- O Vacancy : 오피스 건물의 공실률(단위: %)

지난 2년의 기간 동안 월별로 수집된 자료 ‘drywall.csv’를 이용하여 분석을 수행한 뒤 다음 물음에 답하여라.

1. 추정된 회귀방정식을 구하여라.

|  |
| --- |
| > Drydata <- read.csv('drywall.csv')  > lmfit <- lm( Drywall ~ Permits + Mortgage + A.Vacancy + O.Vacancy, data=Drydata )  > lmfit$coefficients  (Intercept) Permits Mortgage A.Vacancy O.Vacancy  -111.827562 4.763052 16.988538 -10.527775 1.307940 |

1. 건축 허가 건수가 50건, 5년 모기지 이자율이 9.0%, 아파트 공실률이 3.6%, 오피스 건물의 공실률이 14.3%일 때 해당 월의 건식벽체의 판매량을 예측하여라.

|  |
| --- |
| > predict (lmfit, +  + data.frame( Permits=50, Mortgage=9.0, A.Vacancy=3.6, O.Vacancy=14.3))  1  260.0254 |

1. 다중공선성 문제가 있는지 여부를 진단하여라.

- VIF가 5또는 10보다 큰 경우 다중공선성이 심각한 것으로 판단하는데, 문제의 독립변수 VIF 값은 전부 5보다 작아 다중공선성 문제가 없는걸로 볼 수 있음

|  |
| --- |
| > vif(lmfit)  Permits Mortgage A.Vacancy O.Vacancy  1.032332 1.002554 1.027269 1.014223 |