

R/데이터분석자격증(ADP)

데이터분석전문가(ADP) 실기시험 기출문제 조사 - 1. 통계 분석

👤 분석하는다니엘 2020. 11. 12.



이번 포스트는 데이터 분석 전문가(ADP) 실기시험을 준비하는 분들에게 도움이 되고자 확인된 내용을 공유하는 글입니다.

아래 기출 문제들의 분류는 제가 임의로 정리하여 오분류가 있을 수 있습니다. ^^

- 최종 업데이트 일자 : 2022-02-21

통계 분석 (Statistical Analysis)

[1회, 2014-06-24]

- 예측 : 세분화 집단별 예측 모형 정의, 예측 모형 개발, 인사이트 제시

[2회, 2014-10-18]

**[5회, 2015-01-24]**

- 당뇨병 환자 나이/성별 별 사망 여부 분석 : 비율/그래프, 상호작용 판단, 범주형/연속형 분석 결과 비교, 교호작용 판단 후 최종 파라미터 해석

[6회, 2016-04-30]

- risk분석(성별/인종/나이/활동성/risk) : 변수간 차이 분석, 교호작용 분석

[7회, 2016-10-29]

- 집단간 의료비 지출 차이 분석 : t-test, ANOVA, 회귀분석

[8회, 2017-04-29]

- 폐활량(Fev) 예측 문제 (나이, 키, 성별, 흡연 유무)
1. EDA 및 상관관계 분석
 2. 적절한 회귀모형 선택
 3. 회귀모형 해석 (변수 별 증가에 따른)
 4. 평균 키, 나이 (여자, 흡연자) 일경우 폐활량 예측



멍 때릴 시간 없는 데이터분석

프로젝트 300+30개, 복붙 바로
10000줄.이거 다 하려면 멍때릴

패스트캠퍼스

[11회, 2018-10-27]

- 출산률 분석 : 독립/반응변수 관계를 회귀분석으로 정의 및 결과 해석

[12회, 2019-03-30]



2. 회귀분석 결과의 해석 방법

3. 회귀 계수에 대한 standard error가 가지는 의미
4. 회귀분석에서 잔차 분석 및 시각화
5. 회귀분석에서 영향력 관측치와 그 영향 분석

[13회, 2019-06-29]

- 거래내역 데이터를 통해 부정사용여부 탐지 모델 개발
1. 타겟변수 불균형 문제 처리 : resampling, undersampling, oversampling(SMOTE) 특징(장단점) 서술 및 적용(패키지 활용가능)
 2. 간단한 시각화
 3. 불균형 문제가 해결된 resampled data로 binary classification model 생성
 4. Confusion matrix와 AUC 등 다양한 성능 지표를 적용하여 결과 해석
 - 분류문제에서 어떤 것을 주로 봐야하는지 판단 필요

[14회, 2019-09-07]

- 로지스틱 회귀 분석 문제
1. 주성분 분석(PCA)
 2. 주성분별 의미 해석
 3. 주성분을 독립변수로 하여 다항 로지스틱 회귀분석 수행
 4. Confusion Matrix 해석

[15회, 2019-12-14]

- 데이터 처리 및 통계 분석
- timestamp 처리 / date 기준 데이터 병합
 - hh:mm, A/B/C/D/E, 전력 사용량 데이터
 - yyyymmdd, 평균 기온
1. 아래 형태의 데이터 생성 : 3개 데이터를 date기준으로 병합 필요
-> yyyy-mm / A / B / C / D / E / 사용량





멍 때릴 시간 없는 데이터분석

프로젝트 300+30개, 복불 바로
10000줄.이거 다 하려면 멍때릴

패스트캠퍼스

[17회, 2020-06-21]

- 통계 분석 (설문데이터 분석) -> (사전에 역문항들에 대한 처리 필요)
 1. 그룹별 평균, 표준편차, 왜도, 첨도 산출
 2. (문제가 기억이 나질 않네요;)
 3. (문제가 기억이 나질 않네요;)
 4. (문제가 기억이 나질 않네요;)
- 시각화 및 시계열 분석 (코로나 데이터) (20점)
 1. 전체 인구대비 누적 사망률이 가장 높은 5개 국가 추출 후, 국가별 일일확진자, 누적확진자, 일일사망자, 누적사망자 시계열 그래프 출력
 2. 위험지수 생성 및 해석
 3. 시계열 분석 및 예측 모델 생성

[18회, 2020-09-19]

- 시계열 분석
 - 데이터 구성 : Year / month / amount
- 1. 정상성 확인 : 평균과 분산 일정 + 근거 & 해석 (10점)
- 2. ARIMA 모델 3가지 제시 (10점)
- 3. 최적 모델 선택 및 근거 서술 (15점)
- 4. 최종 예측 후 실제 결과와 비교 평가, 평가 방법 사용 이유 제시 (15점)





- 날짜 및 주가 수익률 데이터 (Time-Series 변환 전 데이터 제공)

1. 데이터 로드, 정상성/이분산성 검증
2. 정상성 파악 근거에 따른 고정시계열 여부 파악
3. SARIMA 분석 및 최적 모형 파라미터 선택
4. 잔차 그래프 출력

[21회, 2021-06-05]

- 선형 회귀(Linear Regression / 릿지(Ridge) / 라쏘(Lasso))
 - 데이터 : 연속형 독립변수가 여러개 있는 소규모 데이터 제공 ($x_1 \sim x_{10}$ 형식)
 - 1. 데이터를 8:2로 분할하고 선형 회귀 적용하시오. 결정계수와 RMSE를 구하시오.
 - 2. 데이터를 8:2로 분할하고 릿지 회귀 적용하시오. alpha 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 RMSE를 계산하시오.
 - 3. 데이터를 8:2로 분할하고 라쏘 회귀 적용하시오. alpha 값을 0부터 1까지 0.1단위로 모두 탐색해서 결정계수가 가장 높을때의 알파를 찾고, 해당 알파로 다시 모델을 학습해서 결정계수와 RMSE를 계산하시오.
- 다항회귀 시각화 문제
 - 1. 다항 회귀를 3차까지 적용하고 각 차수별 스캐터 플롯과 계수 및 기울기 선을 출력하시오. (총 3개 차트 출력)
- 이원분산분석, 통계표 작성
 - 데이터 구성 : 연속형 변수 2개, 범주형 변수 1개(alpha값을 가짐)
 - 1. 이원 분산 분석을 수행하고 통계표를 작성하시오.

[22회, 2021-09-12]

- 제품에 금속 재질 함유량의 분산이 1.3을 넘으면 불량이라고 보는데 제조사별로 차이가 남. 분산 검정 수행. 유의확률 0.05

데이터 : 금속 성분 함유량 데이터





최대 수익을 낼 수 있을 제품 1과 제품2의 개수 구하기

데이터 없음

- 상품 a와 b가 있을 때 구매 패턴이 aa bb aaaa bbbb a b 등으로 나타날 때 두 상품의 연관성 유무를 검정할 것

데이터 없음

출처 : <https://lovelydiary.tistory.com/381>

[23회, 2021-12-18]

- A공장에서 만든 진공관 수명이 10,000 시간이라고 주장하여 관리팀에서 12개 샘플을 뽑았을때, 유의수준 5%에서 부호검정을 하시오.

1. 연구가설과 귀무가설 작성하여라.(5점)
2. 유효한 샘플의 수를 제시하여라. (5점)
3. 검정통계량을 제시하고 연구가설 채택 여부를 작성하여라.(5점)

- 문제 : 문제 3. 코로나 시계열 데이터 6만 관측치 가량. 날짜, 코로나 누적 확진자 등 변수 3개.

1. ACF로 distance를 계산하여라. (10점)
2. (3-1을 이용하여) 계층적 군집 분석을 위한 덴드로그램을 작성하여라. (10점)

- 문제 : 학과와 성적이 관계있는지 검정하시오

사회과학, 자연과학, 공학, 총 3 개 학과의 평점 조사 표(문제지에만 제공)

3.5-4.5, 2.5-3.5, 1.5-2.5 3개 점수구간이 row index이며 학과가 컬럼이고 값으로는 사람 수가 들어가있

1. 연구가설과 귀무가설을 작성하여라. (5점)
2. 학과와 성적이 독립일 때, 기댓값 구하시오.(5점)
3. 검정통계량 구하고 연구가설 채택여부를 작성하여라. (5점)

출처 : <https://9566.tistory.com/57>





참고 자료

기본 분석

- R을 사용한 t-test - 두 그룹 간 평균 차이가 유의미 한 지를 비교해 보자.
- R을 사용한 일원배치 분산분석 (ANOVA in R) - 집단 별 차이가 통계적으로 의미가 있는지 검증해 보자.

상관 & 회귀 분석

- 회귀분석 lm()함수의 결과 해석
- R을 활용한 통계분석 - 4.상관분석 단순회귀분석
- [R] 회귀분석(Regression Analysis) [회귀식 추정, 회귀모형 검정, 적합도 파악]
- 로지스틱 회귀분석을 통한 유방암 예측(분류) - 398, 399, 400, 401
- 회귀분석에서의 교호작용 Interaction
- R을 사용한 다중회귀분석 (Multiple regression in R)
- R을 사용한 로지스틱 회귀분석 (Logistic regression in R)

주성분 분석

- 고유값, 고유벡터 구하기 (calculation of eigenvalue and eigenvector) : 고유값(eigenvalue), 고유 벡터(eigenvector) 계산
- R을 활용한 주성분 분석 : 표준화, 방향변화, scree plot, biplot, 분석 실습, 결과 해석
- 주성분분석(PCA: Principal component analysis) with R : 주성분 분석 다양한 분석 사례
- R 데이터 변환 : (5) 차원 축소 - (5-1) 주성분분석 (PCA) : 표준화, 방향변화, 상관계수(다중공선성), 주성분 분석, scree plot, biplot, 누적기여율(80%이상 선택), 결과해석





- 부동산 매매가격 추이 분석

데이터 처리

- 클래스 불균형을 다루는 방법 (Class Imbalance)
- 클래스 불균형(Class imbalance)
- 성능지표 계산을 위한 Confusion Matrix
- R 데이터마이닝 dplyr

ADP 시험 응시자 분들 모두 화이팅입니다!!! ^^

ES식품원료 농축액

농축액 원료, 농축과즙, 한방농축등 맛소재 일체 취급, 식품개선, 원가절감, 개발 원료

<http://esfood.kr>

5

구독하기

관련 글

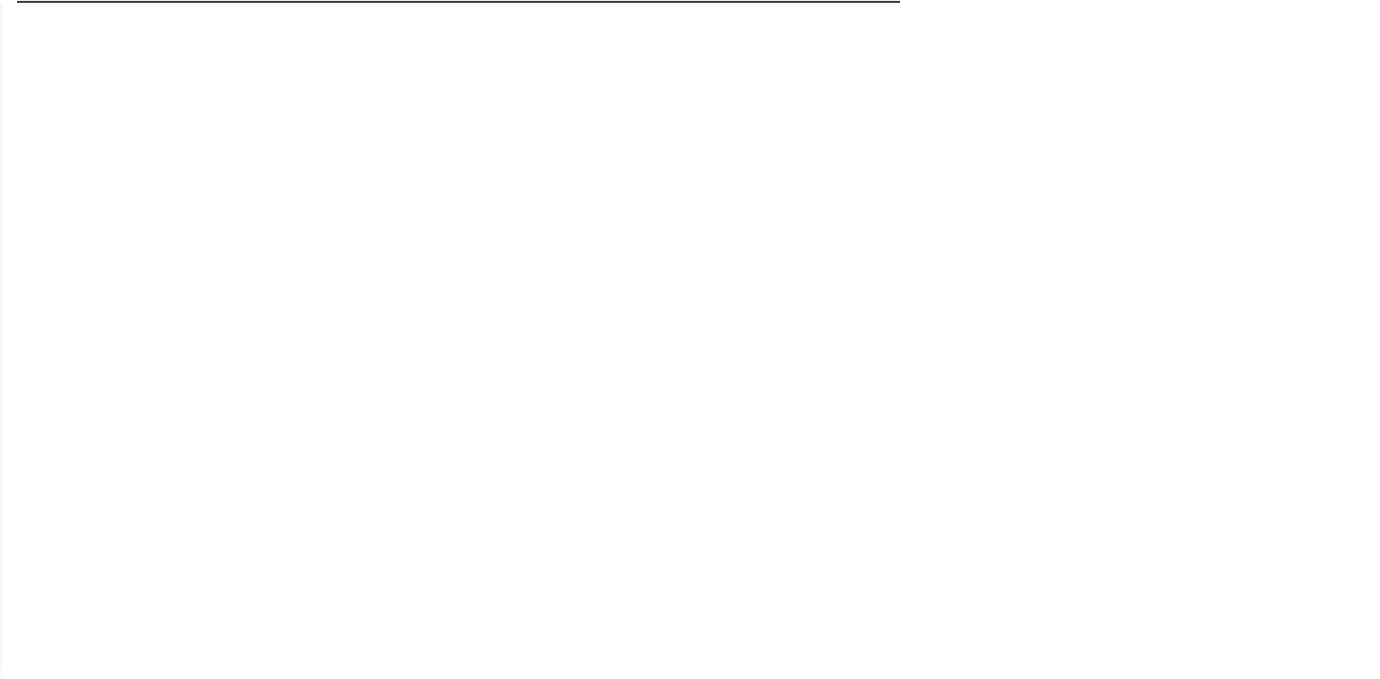
- ▶ 데이터분석전문가(ADP) 실기시험 기출문제 조사 - 3. 텍스트 마이닝
- ▶ 데이터분석전문가(ADP) 실기시험 기출문제 조사 - 2. 데이터 마이닝 & 기계 학습
- ▶ 데이터분석전문가(ADP) 실기시험 기출문제 조사 - 1. 통계 분석
- ▶ 데이터분석전문가(ADP) 19회 실기시험 준비
- ▶ 데이터분석전문가(ADP) 17회 실기시험 후기
- ▶ 데이터분석전문가(ADP) 15회 실기시험 후기

태그

#ADP, #ADsP, #데이터분석, #데이터분석자격증, #데이터분석전문가, #빅데이터, #시험후기, #자격증

추천 글





댓글 4

연어킬러

2021.04.30 13:13 신고



안녕하세요. 혹시 포스팅 출처남기고 퍼가도될까요!?!?

답글

분석하는다니엘

2021.04.30 13:18 신고



네~ 가능합니다~ 꼭 출처 남겨주세요~ ^^

leafjjang

2021.12.07 23:15



혹시 풀이과정은 볼 수 없을까요? ^^

답글

leafjjang

2021.12.10 22:15



풀이과정도 좀 올려주세요~

답글





☐비밀글

등록

