

● 회귀분석

○ 회귀분석 > 상관분석

- 수치자료들에 대한 상관분석 결과를 제공함
- 분석품
 - 변수목록에서 분석할 변수들을 '분석변수' 목록으로 전달 후 분석하면 해당변수들에 대해 상관행렬과 함께 유의성 검정에 대한 p값을 출력함
 - '기준변수'에 변수가 있는 경우 기준변수와 분석변수들 간의 상관분석을 진행함
 - '상관분석방법'으로 Pearson의 상관분석, Spearman의 순위상관, Kendall의 tau를 모두 선택할 수 있음
 - '언어변환'은 같은 결과를 얻기 위한 'R' 또는 'SAS' 명령어를 출력해 주고 "출력옵션"을 통해 결과시트, 출력형태, 분석표의 결과를 좌우/상하로 할 것인지를 지정할 수 있음

상관분석 V1.0

변수목록

분석변수

기준변수

상관분석

☒ Pearson ☒ Spearman ☐ Kendall tau

대립가설 ☐ $p < 0$ ☐ $p > 0$ ☒ $p \neq 0$

언어변환

☐ R ☐ SAS ☐ Python

분석
출력옵션
재설정
도움말
종료

[그림 1] 상관분석 분석품

[기준변수가 없는 경우]

상관분석: Stackloss				
※ 대립가설: $p \neq 0$				
Pearson 상관분석				
	표본 크기	21	21	21
공기흡입량	상관계수	1.0000	0.7819	0.5001
	p값	0.0000	0.0000	0.0209
냉각수온도	표본 크기	21	21	21
	상관계수	0.7819	1.0000	0.3909
	p값	0.0000	0.0000	0.0797
질산농도	표본 크기	21	21	21
	상관계수	0.5001	0.3909	1.0000
	p값	0.0209	0.0797	0.0000

[기준변수가 있는 경우]

상관분석: Stackloss							
※ 대립가설: $p \neq 0$							
Pearson 상관분석				Spearman 상관분석			
기준변수	분석변수	표본 크기	상관계수	p값	표본 크기	상관계수	p값
손실	공기흡입량	21	0.9197	0.0000	21	0.9180	0.0000
	냉각수온도	21	0.8755	0.0000	21	0.8521	0.0000
	질산농도	21	0.3998	0.0725	21	0.4974	0.0218

○ 회귀분석 > 선형회귀모형

- 수치자료들에 대해 선형회귀분석 결과를 제공함
- 분석품
 - '변수목록'에서 '반응변수'와 '설명변수'를 전달함.
모든 변수는 수치자료이며 요인이 있는 경우에는 공분산분석모형에서 분석할 수 있음
 - '모형설정 및 결과출력'에서 회귀모형에 절편 포함 여부, 반응변수를 변환(자연로그, 상용로그, 제곱근)할 것인지를 결정함
 - '다중공선성'을 선택하면 VIF를 계산하기 위한 결정계수를 같이 출력하고 $1/VIF$ 인 공차한계를 출력함
 - '표준화회귀계수'를 선택하면 각 변수들을 표준화하여 회귀계수를 구한 것으로 그 크기로 반응변수에 대한 상대적 영향력을 확인할 수 있는데 회귀계수추론부분에 출력됨
 - '예측값 출력', '잔차출력', '잔차검진'의 상세 옵션은 "출력옵션"에서 지정할 수 있으며 자세한 내용은 아래의 그림을 통해 확인할 수 있음. 사용불가인 체크박스는 이후 버전에 추가될 예정임. 예측값은 설명변수가 있는 모든 값에 대해 출력하고 잔차의 경우 반응변수나 설명변수가 없으면 출력하지 못함
 - '변수선택'은 '전진선택', '후진제거', '단계적 선택' 3가지를 모두 선택할 수 있으며 기준은 F-검정, AIC, BIC 중 하나를 선택해 적용함
 - F검정의 경우 전진선택은 모형에 추가한 변수의 회귀계수의 p-값이 추가유의수준보다 낮은 변수가 없을 때까지, 후진제거는 모형에 포함된 변수의 회귀계수의 p-값이 제거유의수준보다 높은 변수가 없을 때까지, 단계적 선택의 전진과 후진의 기준을 모두 적용하여 유의한 변수가 없을 때까지 진행함
 - "언어변환"은 같은 결과를 얻기 위한 'R' 또는 'SAS' 명령어를 출력해 줌
- 출력옵션
 - 출력옵션에는 '결과출력', '예측값', '잔차', '잔차검진' 관련 옵션을 설정할 수 있음

선형회귀분석 V1.0

변수목록 | 반응변수

손실 | 설명변수

공기흡입량
냉각수온도
질산농도

분석
출력옵션
재설정
도움말
종료

모형설정 및 결과출력

☒ 절편포함 ☐ 반응변수변환 자연로그

☒ 예측값 출력 ☒ 잔차출력 ☒ 잔차검진

☒ 다중공선성 ☒ 표준화회귀계수 ☐ 영향관측값(특이값)

변수선택

☒ 전진선택법 ☐ 후진제거법 ☐ 단계적 선택

선택기준

☒ F-검정 ☐ AIC ☐ BIC

추가유의수준 5% 제거유의수준 5%

언어변환

☐ R ☐ SAS ☐ Python

[그림 2] 선형회귀모형 분석품

출력옵션 V1.0

결과출력 | 예측값 | 잔차 | 잔차검진

☒ 분산분석표 ☒ 회귀계수추론 ☒ 결정계수

☐ 기술통계 ☐ 상관분석

출력시트

결과출력시트 _KESS분석결과시트_

☒ 표준액셀형식 ☐ Gridline삭제

소수점표시 자리수 4

예측잔차출력시트 신체측정자료

변환시트명 _KESS변환시트_

그림크기 폭 400 높이 400

확인
취소

출력옵션 V1.0

결과출력

예측값

잔차

잔차검진

예측값 출력

☒ 예측값

변수명

반응변수(예측)

☒ 기댓값 신뢰구간

반응변수(ECI)

☐ 관측값 예측구간

반응변수(OPI)

신뢰수준

95 %

산점도

☐ 설명변수 vs 예측값

☐ + 관측값

☐ 관측순서 vs 예측값

☐ + 관측값

☐ 관측값 vs 예측값

확인

취소

출력옵션 V1.0

결과출력

예측값

잔차

잔차검진

☒ 정규성검정(Shapiro-Wilk 검정)

☒ 정규확률그림(Q-Q plot)

☒ 등분산성

☐ Durbin-Watson 검정(관측순)

☐ Durbin-Watson 검정(예측크기순)

☒ 잔차그림(예측값 vs 잔차)

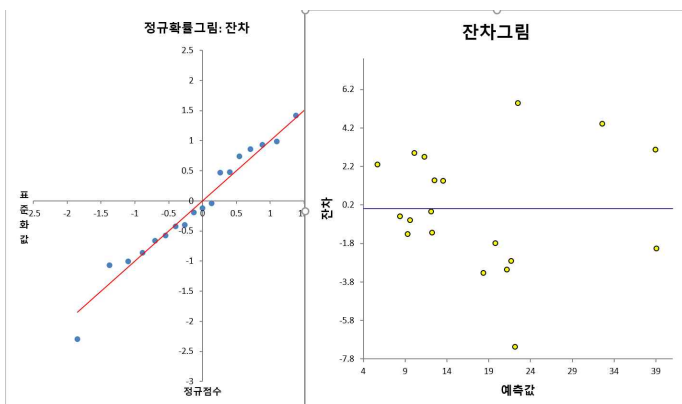
☐ Spread-Level 그림
(예측값 vs |스튜던트화 잔차|)

확인

취소

【회귀분석 결과】

회귀분석결과: Stackloss					
분석모형: 손실 = 절편 + 공기흡입량 + 냉각수온도 + 질산농도					
분산분석표					
소스	자유도	제곱합(SS)	평균제곱(MF-통계량)	p-값(F >)	
모형	3	1773.2997	591.0999	50.4789	0.0000
오차	15	175.6477	11.7098		
전체	18	1948.9474			
결정계수	0.9099				
수정결정계수	0.8919				
회귀계수추론					
설명변수	회귀추정계수	표준화회귀	표준오차	T-통계량	p-값
절편	-41.8900	0.0000	13.3353	-3.1413	0.0067
공기흡입량	0.7018	0.6290	0.1526	4.5999	0.0003
냉각수온도	1.3529	0.4132	0.4075	3.3197	0.0047
질산농도	-0.1332	-0.0693	0.1765	-0.7548	0.4620
다중공선성 검진					
변수	결정계수	VIF	공차한계		
공기흡입량	0.6787	3.1126	0.3213		
냉각수온도	0.6122	2.5785	0.3878		
질산농도	0.2869	1.4023	0.7131		
잔차검진					
가정	검정방법	검정통계량	p-값		
정규성	Shapiro-Wilk	0.9756	0.8800		
정규성	Jarque-Bera	0.2298	0.8915		
등분산성	Breusch-Pagan	4.4855	0.2136		



【변수선택 예제】

변수선택: 전진선택법(Forward Selection) & 기준: F-검정					
기저모형: 절편					
설명변수	자유도	추정회귀계수	표준오차	p-값	
+ 공기흡입	17	1.0205	0.1094	0.0000	<---
+ 냉각수온	17	2.8775	0.3789	0.0000	
+ 질산농도	17	0.7950	0.4247	0.0785	
기저모형: 절편+공기흡입량					
설명변수	자유도	추정회귀계수	표준오차	p-값	
+ 냉각수온	16	1.3824	0.4002	0.0033	<---
+ 질산농도	16	-0.1893	0.2241	0.4106	
기저모형: 절편+공기흡입량+냉각수온도					
설명변수	자유도	추정회귀계수	표준오차	p-값	
+ 질산농도	15	-0.1332	0.1765	0.4620	
※ 최종선택모형: 절편+공기흡입량+냉각수온도					

【예측값, 잔차 출력 예제】

	A	B	C	D	E	F	G	H
1	공기흡입량	냉각수온도	질산농도	손실	손실(예측)	손실(ECI:관측순)	손실(ECI:예측크기순)	잔차
2	80	27	89	42	38.9238	34.8394	43.0082	3.0762
3	62	24	87	28	22.4991	19.6471	25.3511	5.5009
4	62	24	93	19	21.6996	17.7730	25.6263	-2.6996
5	58	18	80	14	12.5072	9.1067	15.9076	1.4928
6	58	18	82	11	12.2407	9.2452	15.2361	-1.2407
7	50		86	7				
8	50	20	80	9	9.5986	6.4491	12.7480	-0.5986
9	80	27	88	37	39.0571	34.8654	43.2487	-2.0571
10	62	22	87	18	19.7933	18.0009	21.5857	-1.7933
11	62	24	93		21.6996	17.7730	25.6263	
12	58	18	89	14	11.3079	8.3372	14.2786	2.6921
13	58	19	93	12	12.1278	8.5647	15.6910	-0.1278
14	50	19	72	8	9.3116	4.5482	14.0751	-1.3116
15	56	20	82	15	13.5429	11.4313	15.6544	1.4571
16	75	25	90	37	32.5757	29.4799	35.6716	4.4243
17	62	23	87	18	21.1462	18.9283	23.3640	-3.1462
18	58	23	87	15	18.3390	15.2897	21.3882	-3.3390
19	58	17	88	13	10.0883	6.6109	13.5657	2.9117
20	50	18	89	8	5.6935	2.1469	9.2401	2.3065
21	50	19	79	8	8.3789	5.3852	11.3726	-0.3789
22	70	20	91	15	22.1689	18.2678	26.0700	-7.1689

○ 회귀분석 > 공분산분석 모형

- 설명변수에 수치자료(공변량, covariate)와 범주형자료(요인, factor)가 동시에 있는 선형모형 분석을 제공함
- 사용법은 분산분석 > 공분산분석 모형을 참조