

소비자연구방법

.13주차. 회귀분석



01. 단순회귀분석

- 단순회귀분석은 독립변수가 1개인 경우에 분석하는 방법
- 단순회귀분석의 회귀모형 등에 대해서는 앞서 회귀분석의 개념과 기본원리에서 단순회귀분석을 토대로 설명되었기 때문에 이 장에서는 SPSS를 활용한 단순회귀분석 방법에 대해서 설명하고자 함

SPSS 프로그램을 사용한 단순회귀분석방법 예시 및 결과

① 연구문제 및 가설의 설정

[연구문제] 광고비 지출은 매출에 영향을 미치는가?

[연구가설] 광고비 지출은 매출에 영향을 미칠 것이다.

-귀무가설(H_0) : 광고비 지출은 매출에 영향을 미치지 않는다.

-연구가설(H_1) : 광고비 지출은 매출에 영향을 미친다.

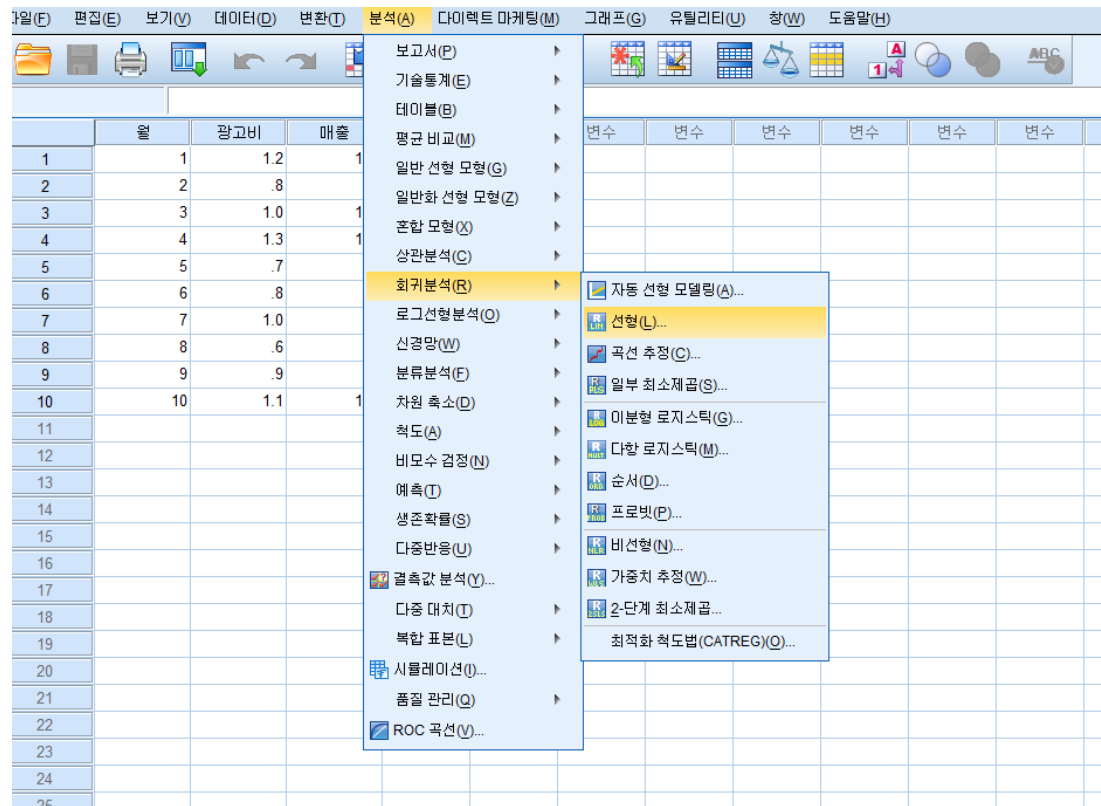
→독립변수: 광고비

→종속변수: 매출

SPSS 프로그램을 사용한 단순회귀분석방법 예시 및 결과

② 분석

단순회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭



The image shows the SPSS software interface with the '분석(A)' menu open. The '회귀분석(R)' option is selected, and its submenu is displayed, showing '선형(L)...' as the next step in the process. The background data table is visible, showing columns for '월' (Month), '광고비' (Advertising Cost), and '매출' (Sales).

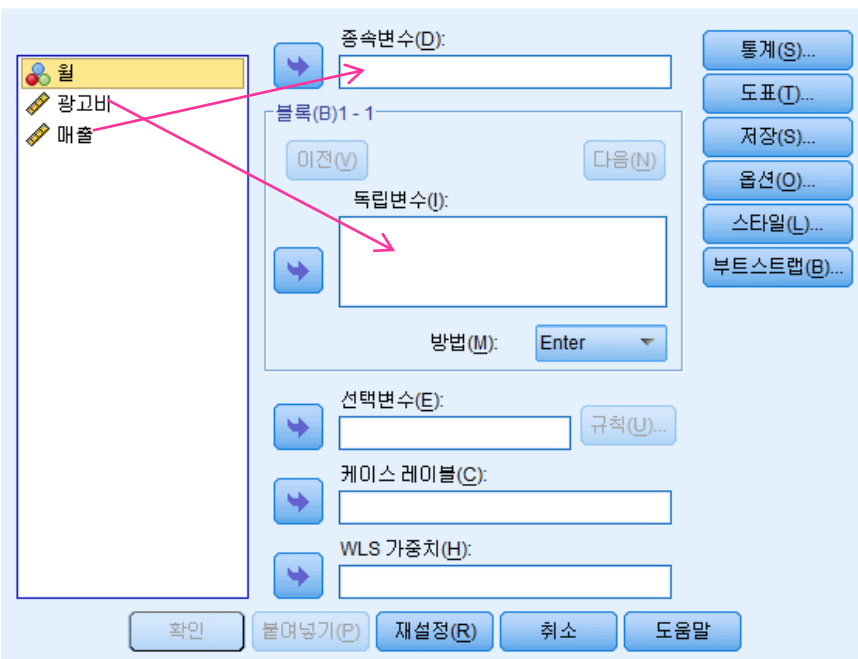
	월	광고비	매출
1	1	1.2	1
2	2	.8	
3	3	1.0	1
4	4	1.3	1
5	5	.7	
6	6	.8	
7	7	1.0	
8	8	.6	
9	9	.9	
10	10	1.1	1
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			

SPSS 프로그램을 사용한 단순회귀분석방법 예시 및 결과

② 분석

단순회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭

선형 회귀분석



The dialog box for Linear Regression in SPSS. The dependent variable (종속변수(D)) is '매출' (Sales). The independent variable (독립변수(I)) is '광고비' (Advertising Cost). The method (방법(M)) is set to 'Enter'. The selection variable (선택변수(E)) is empty. The case labels (케이스 레이블(C)) and WLS weights (WLS 가중치(H)) are also empty. The '확인' (OK) button is highlighted.

선형 회귀분석



The dialog box for Linear Regression in SPSS, showing the same settings as the previous image. The dependent variable (종속변수(D)) is '매출' (Sales). The independent variable (독립변수(I)) is '광고비' (Advertising Cost). The method (방법(M)) is set to 'Enter'. The selection variable (선택변수(E)) is empty. The case labels (케이스 레이블(C)) and WLS weights (WLS 가중치(H)) are also empty. The '확인' (OK) button is highlighted.

SPSS 프로그램을 사용한 단순회귀분석방법 예시 및 결과

② 분석

단순회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭

선형 회귀분석



The image shows the 'Linear Regression' dialog box in SPSS. On the left, a list of variables includes '월' (Month) and '광고비' (Advertising Cost). The 'Dependent Variable' field contains '매출' (Sales). The 'Block(s)' section shows '광고비' as the independent variable. The 'Method' dropdown is set to 'Enter' and is highlighted with a pink box. Other options like 'Stepwise', 'Forward', and 'Backward' are available. On the right, there are buttons for 'Statistics...', 'Plots...', 'Save...', 'Options...', 'Style...', and 'Bootstrap...'. At the bottom, there are buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'.

- 입력(Enter): 모든 독립변수들이 동시에 투입되는 동시입력방식을 의미(기본설정)
- 그 밖에 단계선택, 제거, 후진 및 전진 방식이 있는데, 단순회귀분석에서는 입력 방식만이 적용

SPSS 프로그램을 사용한 단순회귀분석방법 예시 및 결과

② 분석

단순회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭



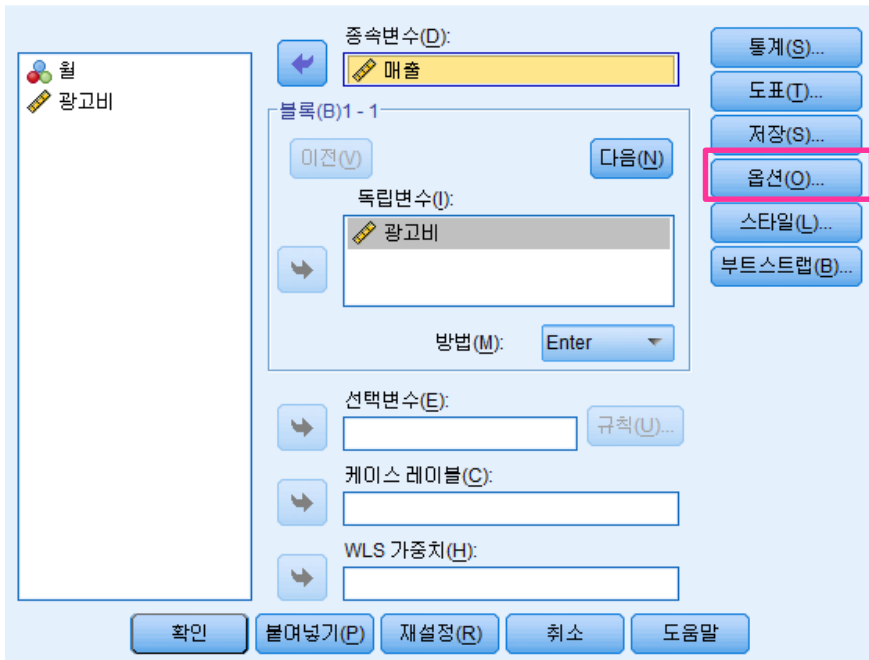
- 추정값: 회귀계수 추정치를 표시, 즉 회귀계수, 베타, 표준오차, t-value, 유의확률 등을 표시(기본설정)
- 신뢰구간: 회귀계수에 대한 95% 신뢰구간을 표시
- 공분산 행렬: 공분산 및 상관행렬을 표시, 대각선에는 분산이 표시되고 대각선의 위와 아래는 공분산이 표시
- 모형적합: 다중 R, R^2 , 수정된 R^2 , 표준오차 등을 표시, 분산분석표에 자유도, 제곱합, 평균제곱, F값 등이 표시(기본설정)
- R제곱 변화량: 단순회귀분석에서는 의미가 없으며 다중회귀분석의 단계선택방식에서 활용
- 기술통계: 각 변수의 평균과 표준편차, 모든 입력변수들 간의 상관관계를 나타냄
- 부분상관 및 편상관계수: 각 독립변수와 종속변수 간의 부분상관계수와 편상관계수를 나타내며, 단순회귀분석에서는 의미가 없으며 다중회귀분석에서 활용
- 공선성진단: 단순회귀분석에서는 의미 없음

SPSS 프로그램을 사용한 단순회귀분석방법 예시 및 결과

② 분석

단순회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭

선형 회귀분석



The 'Linear Regression' dialog box in SPSS. The 'Dependent Variable' is '매출' (Sales). The 'Independent Variable' is '광고비' (Advertising Cost). The 'Method' is set to 'Enter'. The 'OK' button is highlighted with a red box.

종속변수(D): 매출

블록(B) 1 - 1

이전(V) 다음(N)

독립변수(I): 광고비

방법(M): Enter

선택변수(E):

케이스 레이블(C):

WLS 가중치(H):

통계(S)...

도표(T)...

저장(S)...

옵션(O)...

스타일(L)...

부트스트랩(B)...

확인

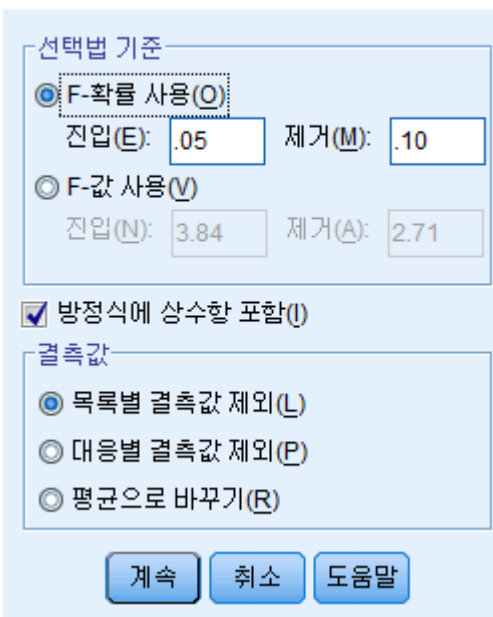
불여넣기(P)

재설정(R)

취소

도움말

선형 회귀분석: 옵션



The 'Linear Regression: Options' sub-dialog box. The 'F-statistic' is selected. The 'Confidence Interval' is set to 0.05. The 'Significance Level' is set to 0.10. The 'Include Constant in Model' checkbox is checked. The 'Display' section is empty.

선택법 기준

☒ F-확률 사용(O)

진입(E): .05 제거(M): .10

☐ F-값 사용(V)

진입(N): 3.84 제거(A): 2.71

☒ 방정식에 상수항 포함(I)

결측값

☐ 목록별 결측값 제외(L)

☐ 대응별 결측값 제외(P)

☐ 평균으로 바꾸기(R)

계속

취소

도움말

- 옵션 대화상자가 나타나면 그림과 같이 선택법 기준, 방정식에 상수항 포함, 결측값 처리방법을 기본설정으로 유지

SPSS 프로그램을 사용한 단순회귀분석방법 예시 및 결과

③ 단순회귀분석 결과

입력된/제거된 변수^a

모형	입력된 변수	제거된 변수	방법
1	광고비 ^b	.	Enter

a. 종속 변수: 매출

b. 모든 요청된 변수가 입력되었습니다.

- 종속변수가 매출, 독립변수는 광고비이고 방법은 동시입력방식을 사용했음이 나타남

모형 요약

모형	R	R 제곱	조정된 R 제곱	표준 추정값 오류
1	.875 ^a	.766	.737	6.837

a. 예측변수: (상수), 광고비

- 모형요약표에서 R제곱은 결정계수라 부르는데, 종속변수의 분산의 몇%가 독립변수에 의해 설명되는가를 나타냄
- 회귀식이 자료에 적합한 정도를 나타냄
- R제곱값은 .766으로서 종속변수 분산의 76.6%가 독립변수에 의해 설명됨을 알 수 있음

분산 분석^a

모형		제곱합	df	평균 제곱	F	유의수준
1	회귀분석	1226.927	1	1226.927	26.246	.001 ^b
	잔차	373.973	8	46.747		
	총계	1600.900	9			

a. 종속 변수: 매출

b. 예측변수: (상수), 광고비

SPSS 프로그램을 사용한 단순회귀분석방법 예시 및 결과

③ 단순회귀분석 결과

계수^a

모형		비표준 계수		표준 계수	t	유의수준
		B	표준 오차	베타		
1	(상수)	46.486	9.885		4.703	.002
	광고비	52.568	10.261	.875	5.123	.001

a. 종속 변수: 매출

- B는 비표준화계수를, 베타(β)는 표준화회귀계수를 의미함
 - 비표준화 회귀계수: 회귀식 구성, 회귀계수간 비교 불가(독립변수의 절대적 영향력)
 - 표준화 회귀계수: 표준화된 점수를 투입, 회귀계수 간 상호 비교 가능하며, 다른 독립변수들을 통제한 상황에서 해당 독립변수가 종속변수를 설명해주는 수치로 부분상관계수와 밀접한 관계가 있음(독립변수의 상대적 영향력)
- 회귀계수의 유의도는 회귀계수 값의 t값이 기각영역에 있는가의 여부를 가지고 판정함
 - 위 회귀계수표를 보면, $t=5.123$, $P=.001(p<.01)$ 로 유의미한 것으로 나타남
- 따라서 귀무가설은 기각되고 연구가설은 지지됨.
 - 즉, 광고비 지출은 매출에 정적인 영향을 미치는 것으로 나타냄을 알 수 있음

02. 다중회귀분석

- 다중회귀분석은 독립변수가 2개 이상인 경우 독립변수들과 종속변수 간의 선형적 인과관계를 조사하기 위한 통계기법임
- 다중회귀분석은 회귀식추정, 기본가정 등 여러 측면에서 단순회귀분석과 같은 원칙에 의해 이루어져 있으며 단순회귀분석의 확대라고 볼 수 있음
- 다중회귀분석은 단순회귀분석보다 종속변수를 더 정확히 예측할 수 있고, 다른 독립변수들을 통제된 상태에서 해당 독립변수와 종속변수 간의 부분적 관계를 분석할 수 있음

① 회귀모형

- ✓ 다중회귀모형은 단순회귀분석에서처럼 선형적 관계를 나타내는 1차식으로 여러 개의 설명변수(독립변수)에 의해 설명되는 부분과 설명되지 못하는 부분으로 나타나게 됨
- ✓ 단순회귀모형에서와 마찬가지로 설명되지 않는 부분은 회귀오차로서 ϵ 로 표기하는 것이 보통임
- ✓ X_1, X_2, X_3 , 3개의 독립변수와 한 개의 종속변수, Y 의 회귀모형은 다음과 같음

모집단회귀모형 $Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i + \epsilon_i$

모집단회귀식 $Y_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_i X_i$

표본회귀모형 $\hat{Y}_i = \alpha + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \dots + \hat{\beta}_i X_i + \epsilon_i$

표본회귀식 $\hat{Y}_i = \alpha + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \dots + \hat{\beta}_i X_i$

→ 변수들 간의 선형적 1차 함수관계를 나타내고 있으며 인과관계를 파악할 수 있음. 다른 독립변수들의 값을 통제된 상태에서 특정 독립변수가 종속변수에 어떤 영향을 미치는지를 예측할 수 있게 설정하고 있음

02. 다중회귀분석

② 회귀선 도출

- ✓ 다중회귀분석에서 회귀식의 추정치는 단순회귀에서와 마찬가지로 최소자승법으로 추정함
- 독립변수가 X_1, X_2 2개 인 경우 추정된 회귀식

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$$

- 위 식에서 기울기 값 $\hat{\beta}_1, \hat{\beta}_2$ 를 구하는 공식과 절편 $\hat{\alpha}$ 값을 구하는 일반적인 공식

$$\hat{\beta}_1 = \frac{(\sum X_1^2 \sum X_2 y) - (\sum X_1 X_2 \sum X_1 y)}{(\sum X_1^2 \sum X_2^2) - (\sum X_1 \sum X_2^2)}$$

$$\hat{\beta}_2 = \frac{(\sum X_2^2 \sum X_1 y) - (\sum X_1 X_2 \sum X_2 y)}{(\sum X_1^2 \sum X_2^2) - (\sum X_1 \sum X_2^2)}$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

02. 다중회귀분석

③ 중회귀 모형의 가정

✓ 앞서 회귀분석의 이해에서 설명한 회귀모형의 기본가정은 모두 중회귀분석에 적용됨

✓ 중회귀분석에서 필요한 추가적 가정은 다음과 같음

① 독립변수들 간에 상관관계가 높지 않아야 함

-독립변수들 간에 서로 높은 상관관계가 존재하지 않아야 함은 다중공선성이 존재하지 않아야 함을 의미

② 독립변수 간에 상호작용효과가 존재하지 않아야 함

-특정 독립변수가 종속변수에 미치는 영향력이 다른 독립변수와 복합하여 일어나지 않아야 함을 의미

③ 표본의 수가 독립변수 개수보다 커야 함

-독립변수가 늘어나서 종속변수에 대한 설명력이 높아질 수는 있으나, 정보의 개수인 자유도 역시 늘어나게 되어 회귀모형의 설명력이 낮아지게 됨

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

- 다중회귀식을 추정하는 방식에는 여러 가지가 있는데 대표적인 것인 **동시입력방식**과 **단계선택방식**임
- **동시입력방식**
 - 연구자가 고려하는 모든 독립변수들을 한꺼번에 포함하여 분석하는 방법
 - 이 방식을 이용하면 다른 독립변수들이 통제된 상태에서 특정 독립변수의 영향력을 알 수 있으며, 연구자가 고려하는 모든 독립변수들이 동시에 종속변수를 설명하는 정도를 알 수 있음
- **단계적 선택방식**
 - 다른 변수들이 회귀식에 존재할 때 종속변수에 영향력이 있는 변수들만을 회귀식에 포함시키는 방식

동시입력방식

① 연구문제 설정

[연구문제] 동거여부, 연령, 교육기간은 TV 시청시간에 영향을 미치는가?

*동거여부는 더미변수로 변경

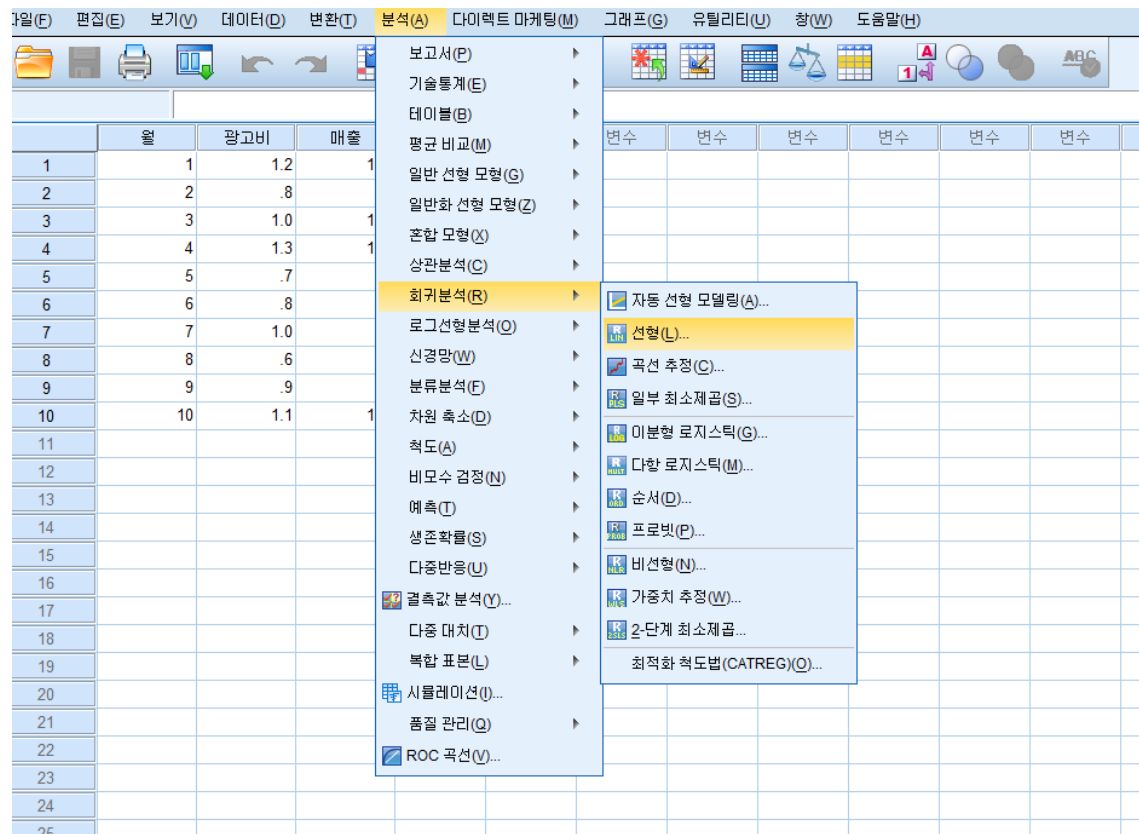
-동거안함=0

-동거함=1

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 분석

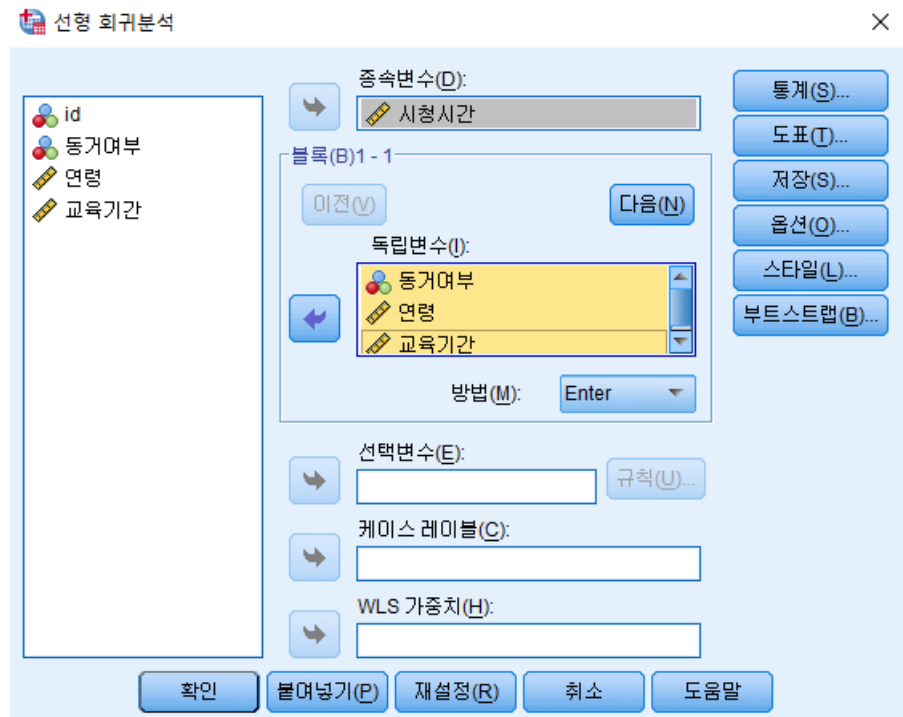
다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭



SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 분석

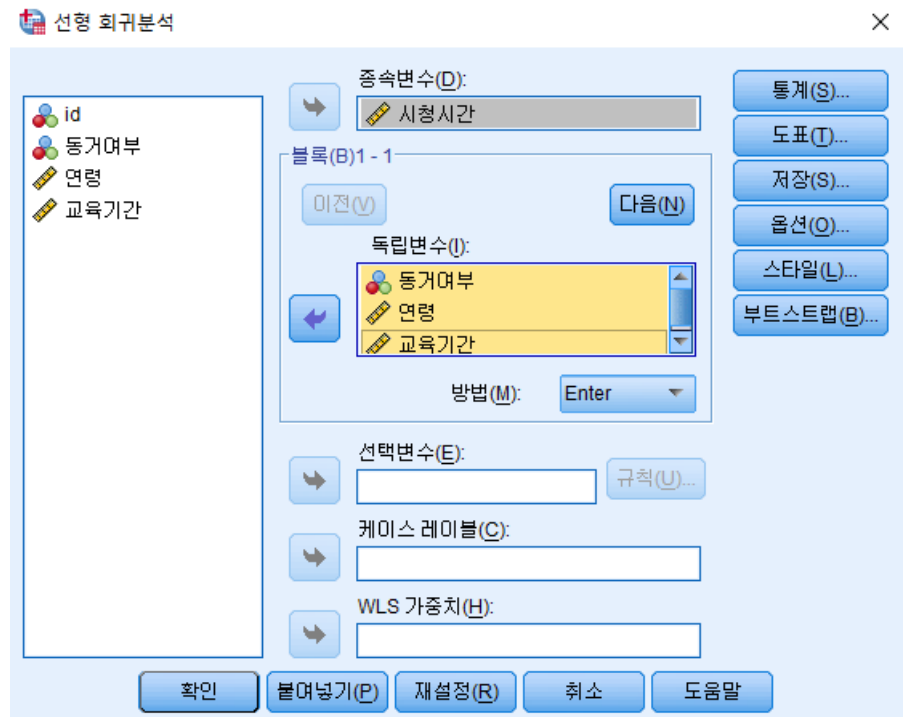
다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭



SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 분석

다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭



SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 분석

다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭

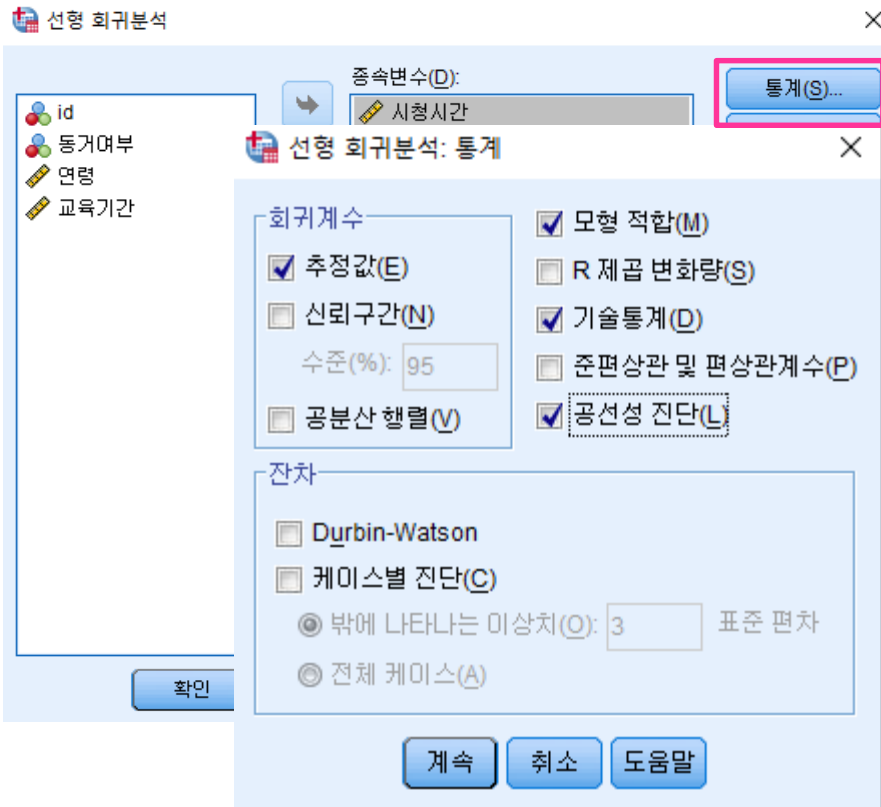


- '방법[M]'으로는 기본설정된 '입력'을 유지
- 입력: 모든 독립변수들이 동시에 투입(기본설정)
- 단계선택: 여러 개의 독립변수들 중에서 설명력이 어느 정도 높은 변수들로만 회귀모형을 구성하기 위한 것
 - 첫단계: 종속변수와 상관관계가 가장 높은 변수가 회귀식에 들어감
 - 두번째 단계부터: 전 단계에서 들어가지 않은 변수들 중에 종속변수와 가장 높은 편상관관계를 갖는 변수가 들어감
 - 단계선택방식에서는 진입조건으로 F값 혹은 p-value 를 지정하는데 진입조건을 충족시키는 변수가 더 이상 없는 단계에서 분석이 종료
- 전진: 단계선택과 마찬가지로 상관관계(혹은 편상관관계)가 높은 독립변수의 순서로 회귀식에 진입
- 후진: 모든 독립변수들이 한꺼번에 들어가서 각 단계에서 종속변수에 대한 설명력이 낮은 순서로 제거

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 분석

다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭



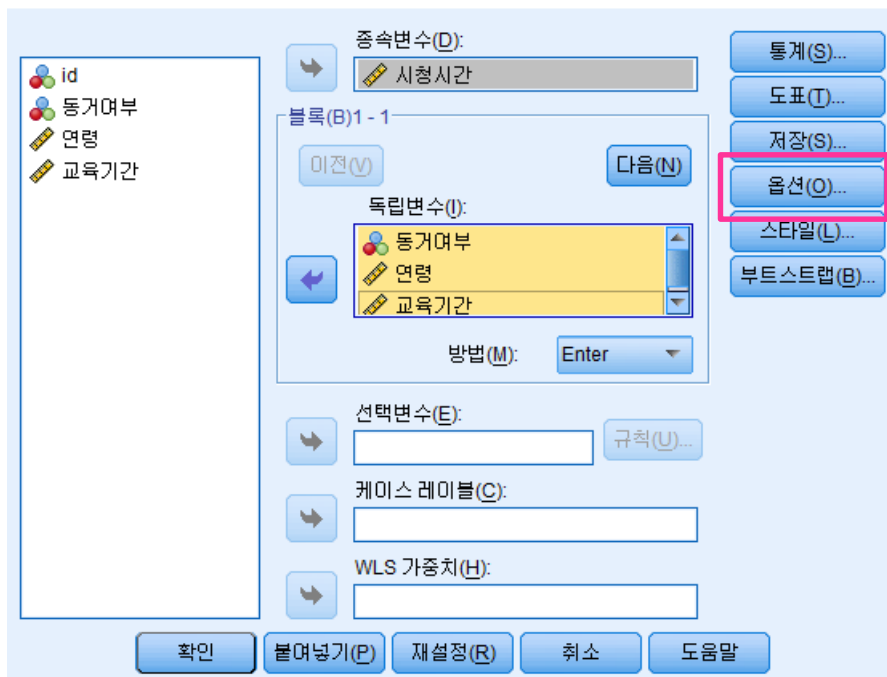
- 통계 대화상자가 나타나면, 회귀계수에서는 추정값을 선택하고, 모형적합, 기술통계, 공선성진단을 선택함
- 다중회귀분석에만 적용되는 R제곱 변화량과 공선성 진단
 - R제곱 변화량: 변수를 추가하거나 삭제하는데 따르는 R2 변화정도를 의미, 동시입력방식에는 적용되지 않음
 - 공선성 진단: 개별 변수에 대한 공차와 공선성 문제 진단을 위한 다양한 통계량을 표시하며, 다중회귀분석을 실행할 때 지정하는 기능으로서 한 독립변수가 다른 독립수들과 선형관계를 갖고 있을 때 발생하는 공선성의 문제를 검증하기 위함
- '잔차' 선택사항의 의미
 - Durbin-Watson: 시계열자료를 회귀분석한 경우 오차항의 값들이 서로 상관관계가 있는지를 조사하는데 사용되는 값
 - 케이스별 진단: 분석의 대상이 된 케이스들의 회귀식에 의한 예측값, 예측값과 실제값의 차이(잔차)를 나타냄

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 분석

다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭

선형 회귀분석



The 'Linear Regression' dialog box in SPSS. The 'Dependent Variable' is '시청시간' (Viewing Time). The 'Independent Variable(s)' are '동거여부' (Living together), '연령' (Age), and '교육기간' (Education period). The 'Method' is set to 'Enter'. The 'OK' button is highlighted with a red box.

종속변수(D): 시청시간

블록(B) 1 - 1

이전(V) 다음(N)

독립변수(I): 동거여부, 연령, 교육기간

방법(M): Enter

선택변수(E): 규칙(U)...

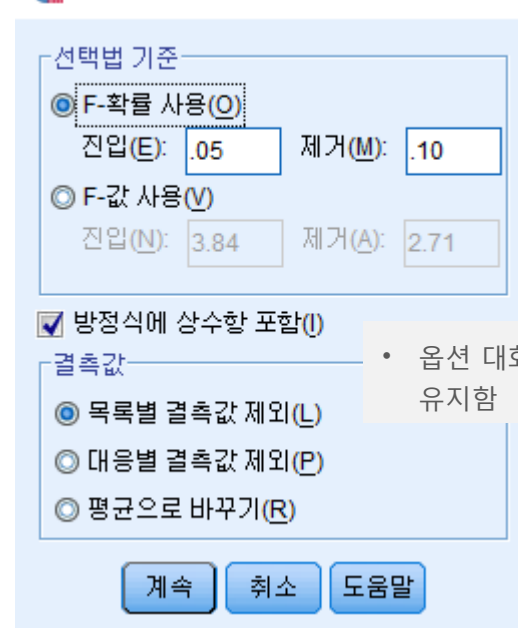
케이스 레이블(C):

WLS 가중치(H):

통계(S)... 도표(T)... 저장(S)... 옵션(O)... 스타일(L)... 부트스트랩(B)...

확인 불여넣기(P) 재설정(R) 취소 도움말

선형 회귀분석: 옵션



The 'Linear Regression: Options' sub-dialog box. The 'Selection Criteria' section has 'F-to-change' selected with an entry value of .05 and a removal value of .10. The 'Change Statistics' section has 'Display R-squared change for each step' selected. The 'Continue' button is highlighted with a red box.

선택법 기준

☒ F-확률 사용(Q)

진입(E): .05 제거(M): .10

☐ F-값 사용(V)

진입(N): 3.84 제거(A): 2.71

☒ 방정식에 상수항 포함(I)

결측값

☒ 목록별 결측값 제외(L)

☐ 대응별 결측값 제외(P)

☐ 평균으로 바꾸기(R)

계속 취소 도움말

• 옵션 대화상자에서는 기본설정을 유지함

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

③ 다중회귀분석(동시입력방식)결과

기술통계

	평균	표준 편차	N
시청시간	2.340	1.1522	25
등거여부	.40	.500	25
연령	73.84	5.691	25
교육기간	10.16	3.602	25

상관

		시청시간	등거여부	연령	교육기간
Pearson 상관계수	시청시간	1.000	-.521	.358	-.612
	등거여부	-.521	1.000	.141	.079
	연령	.358	.141	1.000	-.501
	교육기간	-.612	.079	-.501	1.000
유의수준 (한쪽)	시청시간	.	.004	.039	.001
	등거여부	.004	.	.251	.354
	연령	.039	.251	.	.005
	교육기간	.001	.354	.005	.
N	시청시간	25	25	25	25
	등거여부	25	25	25	25
	연령	25	25	25	25
	교육기간	25	25	25	25

- 기술통계 결과에서는 변수들의 평균과 표준편차가 제시되어 있음
- 상관계수 결과표를 살펴보면, 각각의 독립변수는 종속변수와 모두 유의적인 상관관계가 있는 것으로 나타남
 - 연령과 교육기간 간에는 유의적인 부적(-) 상관관계가 있는 것으로 나타나는데, 이 결과는 연령과 교육기간 간에 약간의 공선성이 존재할 가능성이 있음을 알 수 있게 해줌

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

③ 다중회귀분석(동시입력방식)결과

입력된/제거된 변수^a

모형	입력된 변수	제거된 변수	방법
1	교육기간, 등거여부, 연령 ^b	.	Enter

a. 종속 변수: 시청시간

b. 모든 요청된 변수가 입력되었습니다.

- 종속변수는 시청시간, 회귀모형개발을 위한 독립변수 투입방법은 동시입력방식(Enter)으로서 교육기간, 등거여부, 연령 세 개의 독립변수가 모두 투입되었음을 나타냄

모형 요약

모형	R	R 제곱	조정된 R 제곱	표준 추정값 오류
1	.791 ^a	.626	.572	.7536

a. 예측변수: (상수), 교육기간, 등거여부, 연령

- 모형요약 표를 살펴보면, 교육기간, 등거여부, 연령의 세 개의 독립변수들이 투입된 결과 R^2 은 .626으로서 종속변수(시청시간)를 62.6% 설명하고 있음을 알 수 있음 -조정된(수정된) R제곱은 .572로 나타남

분산 분석^a

모형	제공함	df	평균 제곱	F	유의수준
1 회귀분석	19.933	3	6.644	11.698	.000 ^b
잔차	11.927	21	.568		
총계	31.860	24			

a. 종속 변수: 시청시간

b. 예측변수: (상수), 교육기간, 등거여부, 연령

- 분산분석표를 살펴보면, $F=11.698$, 유의수준(p-value)는 .000으로 나타나 회귀식(모형요약)의 설명력이 유의한 것으로 나타남

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

③ 다중회귀분석(동시입력방식)결과

계수^a

모형		비표준 계수		표준 계수	t	유의수준	공선성 통계	
		B	표준 오차	베타			허용 오차	VIF
1	(상수)	1.495	2.637		.567	.577		
	동거여부	-1.176	.316	-.510	-3.726	.001	.951	1.052
	연령	.039	.032	.191	1.214	.238	.717	1.395
	교육기간	-.152	.050	-.476	-3.039	.006	.727	1.376

a. 종속 변수: 시청시간

- 각 독립변수 계수의 유의성 검증한 결과는 다음과 같음
- 동거여부(1=동거, 0=비동거)는 TV시청시간에 유의한 부(-)적영향을 미치는 것으로 나타남($t=-3.726$, $p=.001$)
-배우자와 동거하는 사람의 경우 그렇지 않은 사람에 비해 TV시청시간이 적다고 해석할 수 있음
- 연령은 TV시청시간에 유의한 영향을 미치지 않는 것으로 나타남($t=1.214$, $p=.238$)
- 교육기간은 TV시청시간에 유의한 부(-)적영향을 미치는 것으로 나타남($t=-3.039$, $p=.006$)
-교육기간이 길수록 TV시청시간이 적다고 해석할 수 있음

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

③ 다중회귀분석(동시입력방식)결과

계수^a

모형	비표준 계수		표준 계수	t	유의수준	공선성 통계	
	B	표준 오차	베타			허용 오차	VIF
1 (상수)	1.495	2.637		.567	.577		
등거여부	-1.176	.316	-.510	-3.726	.001	.951	1.052
연령	.039	.032	.191	1.214	.238	.717	1.395
교육기간	-.152	.050	-.476	-3.039	.006	.727	1.376

a. 종속 변수: 시청시간

- 공선성은 두 개의 독립변수들 간의 관계를 의미함
- 두 개의 독립변수들간의 상관계수가 1이면 완전한 공선성을 보인다 하고, 상관계수가 0이면 전혀 공선성이 없음을 의미함
- 세 개 이상의 변수들 간의 관계를 다중공선성이라 하는데, 한 독립변수가 종속변수에 대한 설명력이 높더라도 공선성이 높으면 설명력이 낮은 것처럼 나타남
- 공선성을 보다 엄격하게 점검하려면 공차와 VIF를 보아야 함
 - 공차(허용오차, 공차한계): 공선성을 점검하기 위한 지표
 - VIF: 공차의 역수로 표시되며, VIF 값이 클수록 독립변수들 간의 공선성이 높음
 - 공차는 0.1보다 크고, VIF는 10보다 작아야 다중공선성의 문제가 없는 것으로 봄
- 위 결과표를 보면, 모든 독립변수의 공차가 0.1보다 크고, VIF는 10보다 작으므로 다중공선성의 문제가 없는 것으로 파악할 수 있음

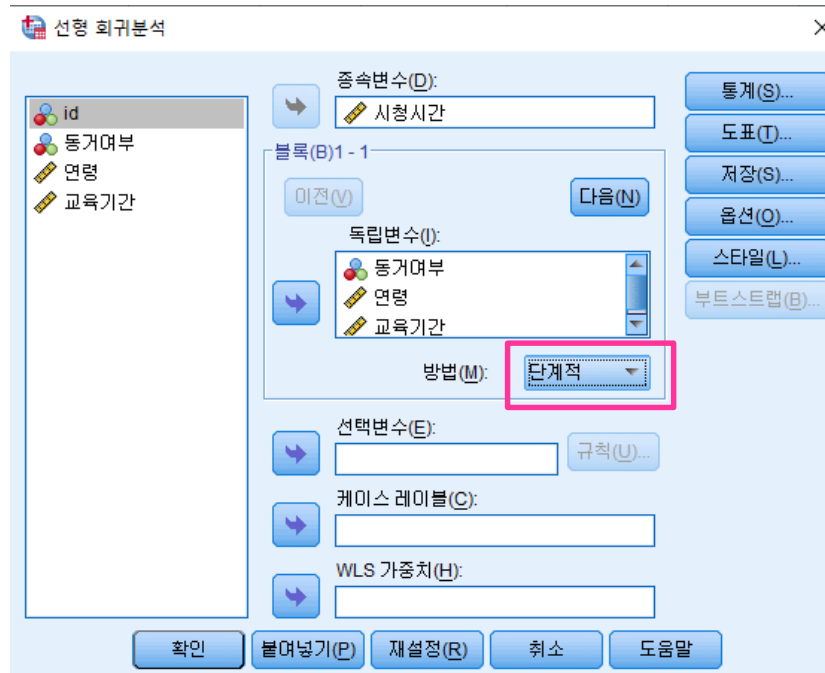
SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

단계선택방식

① 분석

다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭

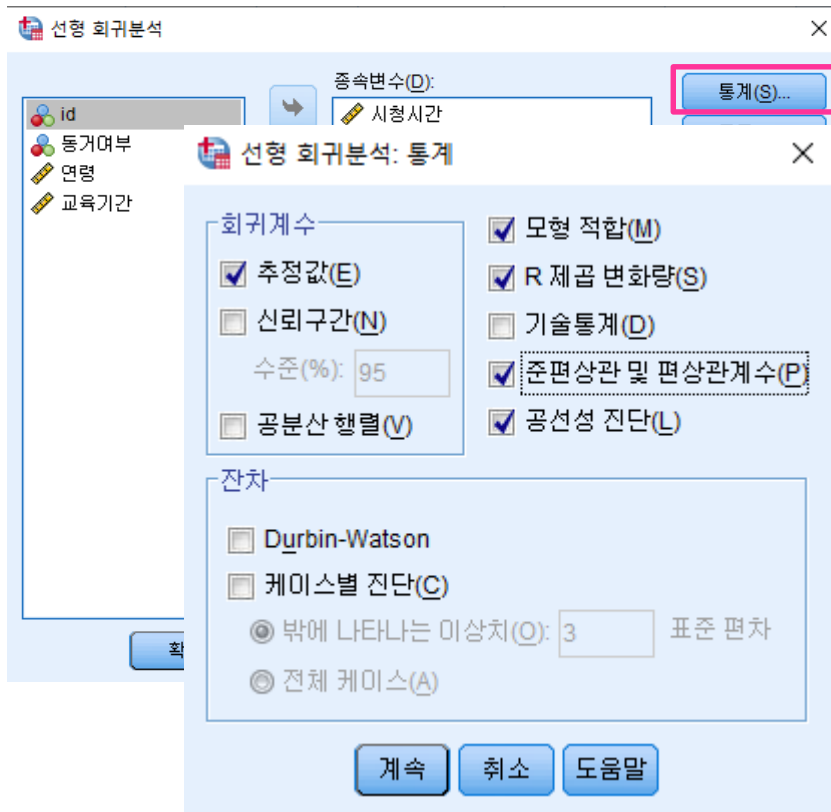
*앞 순서는 같아 생략



SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

① 분석

다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭

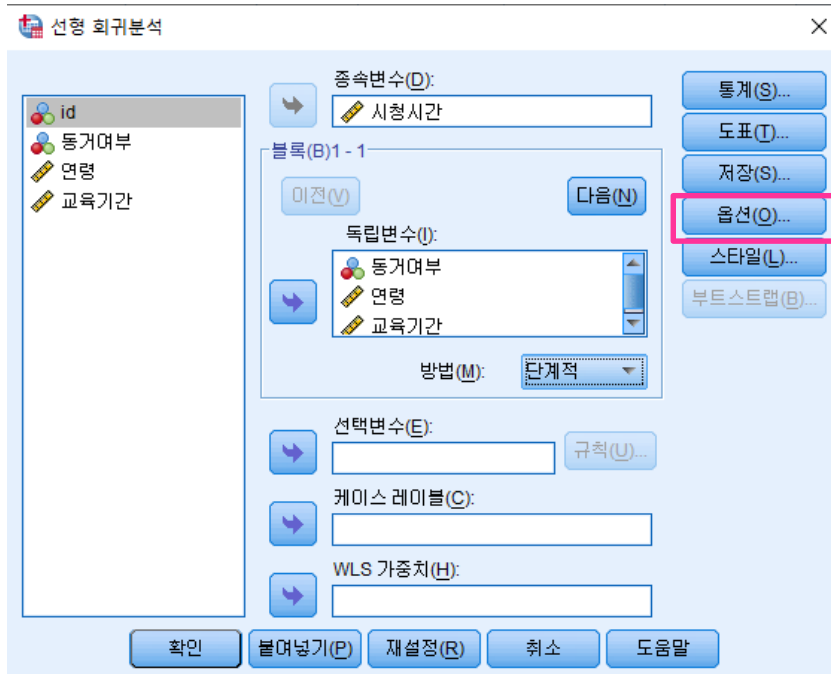


- 통계 대화상자가 나타나면, 회귀계수에서는 추정값을 선택하고, 모형적합, R 제곱 변화량, 준편상관(부분상관) 및 편상관계수, 공선성 진단을 선택
- R제곱 변화량
- 독립변수가 추가됨으로서 R제곱 값이 얼마만큼 커지는지 나타냄
- R제곱 변화량이 크다는 것은 그 변수가 종속변수에 대한 설명력이 높음을 의미함
- R제곱 변화량은 다중회귀분석에서 단계선택방식에서만 의미를 가짐

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

① 분석

다중회귀분석 수행과정 : [분석(A)] → [회귀분석(R)] → [선형(L)...] → 클릭



선형 회귀분석

종속변수(D): 시청시간

블록(B) 1 - 1

미전(V) 다음(N)

독립변수(I): 동거여부, 연령, 교육기간

방법(M): 단계적

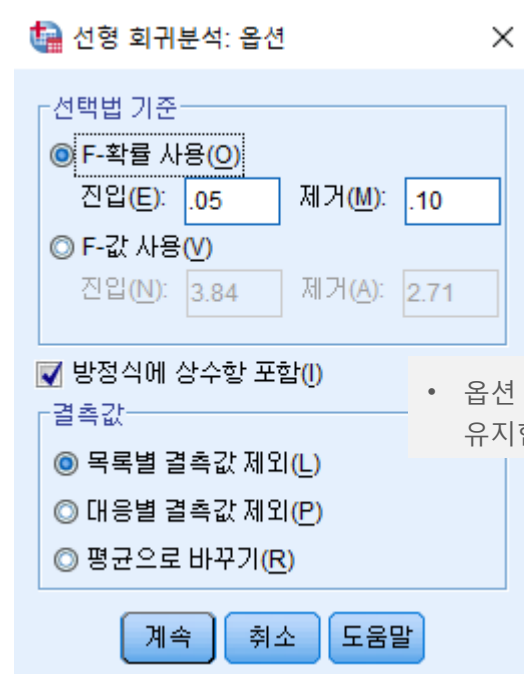
선택변수(E): 규칙(U)...

케이스 레이블(C):

WLS 가중치(H):

통계(S)...
도표(T)...
저장(S)...
옵션(O)...
스타일(L)...
부트스트랩(B)...

확인 불여넣기(P) 재설정(R) 취소 도움말



선형 회귀분석: 옵션

선택법 기준

☒ F-확률 사용(O)
진입(E): .05 제거(M): .10

☐ F-값 사용(V)
진입(N): 3.84 제거(A): 2.71

☒ 방정식에 상수항 포함(I)

결측값

☒ 목록별 결측값 제외(L)
☐ 대응별 결측값 제외(P)
☐ 평균으로 바꾸기(R)

계속 취소 도움말

• 옵션 대화상자에서는 기본설정을 유지함

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 다중회귀분석(단계선택방식) 결과

입력된/제거된 변수^a

모형	입력된 변수	제거된 변수	방법
1	교육기간	.	단계적(기준: F-to-enter의 확률 $\leq .050$, F-to- remove의 확률 $\geq .100$).
2	동거여부	.	단계적(기준: F-to-enter의 확률 $\leq .050$, F-to- remove의 확률 $\geq .100$).

a. 종속 변수: 시청시간

- 종속변수는 시청시간이고, 모형개발을 위한 독립변수 투입방식은 단계선택방식으로서 1단계에서는 교육기간만이 독립변수로 투입, 2단계에서는 동거여부가 추가적인 독립변수로 투입되었음을 알 수 있음
- 여기에서 교육기간이 먼저 투입된 것은 시청시간과의 상관관계가 세 개의 독립변수들 중 교육기간이 가장 높기 때문임

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 다중회귀분석(단계선택방식) 결과

모형 요약

모형	R	R 제곱	조정된 R 제곱	표준 추정값 오류	통계 변경				
					R 제곱 변화량	F 변화량	df1	df2	유의수준 F 변화량
1	.612 ^a	.375	.347	.9308	.375	13.777	1	23	.001
2	.774 ^b	.599	.563	.7617	.225	12.342	1	22	.002

a. 예측변수: (상수), 교육기간

b. 예측변수: (상수), 교육기간, 동거여부

- 1단계에서 교육기간만 투입되어 회귀분석한 결과 R제곱 값은 .375로 나타났고, 2단계에서 동거여부가 추가적으로 투입된 결과 R제곱 값은 .559로 R제곱 변화량은 .225만큼 증가된 것을 확인할 수 있음
- 이때 R제곱 증가량은 유의한 것으로 나타남

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 다중회귀분석(단계선택방식) 결과

분산 분석^a

모형	제공합	df	평균 제공	F	유의수준
1 회귀분석	11.935	1	11.935	13.777	.001 ^b
잔차	19.925	23	.866		
총계	31.860	24			
2 회귀분석	19.096	2	9.548	16.456	.000 ^c
잔차	12.764	22	.580		
총계	31.860	24			

a. 종속 변수: 시청시간

b. 예측변수: (상수), 교육기간

c. 예측변수: (상수), 교육기간, 동거여부

- 각 단계별 회귀식의 유의성을 보여주고 있는데, 교육기간만을 투입한 회귀모형1과 동거여부를 추가한 회귀모형 2 모두 유의한 것으로 나타남

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 다중회귀분석(단계선택방식) 결과

계수^a

모형		비표준 계수		표준 계수	t	유의수준	상관			공선성 통계	
		B	표준 오차	베타			0차	편	준편 상관	허용 오차	VIF
1	(상수)	4.329	.567		7.631	.000					
	교육기간	-.196	.053	-.612	-3.712	.001	-.612	-.612	-.612	1.000	1.000
2	(상수)	4.646	.473		9.823	.000					
	교육기간	-.184	.043	-.575	-4.245	.000	-.612	-.671	-.573	.994	1.006
	동거여부	-1.096	.312	-.476	-3.513	.002	-.521	-.599	-.474	.994	1.006

a. 종속 변수: 시청시간

- 위 표는 각 단계 모형별 계수 값들을 보여주고 있음
- 단계선택방식에 의해 독립변수들을 입력한 결과 얻어진 회귀모형의 각 계수값은 모두 유의한 것으로 나타남
- 모형2의 경우, 0차 상관은 그대로 -.612데, 편상관과 준편상관(부분상관) 값은 변화였음. 이는 모형에 포함된 다른 변수인 동거여부에 기인함
- 두 변수의 공차는 0.1보다 크고, VIF는 10보다 작으므로 다중공선성의 문제는 없는 것으로 파악할 수 있음

SPSS 프로그램을 사용한 다중회귀분석방법 예시 및 결과

② 다중회귀분석(단계선택방식) 결과

제외된 변수^a

모형	변수	베타 IN	t	유의수준	편상관	공선성 통계	
						허용 오차	최소 허용 오차
1	동거여부	-.476 ^b	-3.513	.002	-.599	.994	.994
	연령	.069 ^b	.355	.726	.075	.749	.749
2	연령	.191 ^c	1.214	.238	.256	.717	.717

a. 종속 변수: 시청시간

b. 모형의 예측변수: (상수), 교육기간

c. 모형의 예측변수: (상수), 교육기간, 동거여부

- 위 표는 각 단계별 모형에서 제외된 변수들에 대한 통계량이 제시되어 있음
- 교육기간만을 투입한 회귀모형 1에서 제외된 동거여부와 연령에 대한 통계량이 제시되어 있는데, 여기서 베타 입력, t 및 유의확률은 그 변수가 회귀모형에 추가적으로 투입되는 경우에 해당하는 표준화계수 베타, t 및 유의확률을 의미함
- 회귀모형 1에서 동거여부가 추가적으로 투입되는 경우에 그 계수값이 유의하므로 이를 추가적으로 투입하여 회귀모형 2를 구성하게 됨
- 회귀모형 2에서 제외된 변수인 연령에 대한 통계이 유의하지 않아 더 이상 단계적 투입이 진행되지 않고 교육기간과 동거여부만을 가지고 최종 회귀모형이 구성