기초통계연습

기술통계에서 일원분산분석까지

김현우, PhD[1]

[1] 충북대학교 사회학과 조교수

July 4, 2023

진행 순서

1. 기술통계
2. 교차표와 χ^2 독립성 검정
3. 평균비교와 t 검정
4. 일원분산분석

기술통계

기술통계

기술통계를 제시하여 자료를 요약한다.

- 기술통계(descriptive statistics)는 자료를 요약하는 목적을 가지고 있으므로 요약통계(summary statistics)라고도 부를 수 있다.
- "자료를 요약하라" 라는 말은 곧 기술통계를 제시하고 이를 언어적으로 표현하라는 말과도 같다.
- 자료를 요약하라고 했는데 자료를 그대로 복사해서 붙여넣으면 멍청한 짓이다(Why?).

기술통계

무엇보다, 가장 먼저, 자료의 척도를 옳게 식별해야 한다!

- 변수가 비율척도에 근접할수록 보다 다양한 기술통계를 활용할 수 있다.

변수유형	척도	추가되는 연산	요약 통계량
질적변수	명목척도	셈	빈도(frequency)와 구성비(percentage) 최빈값(mode)
	서열척도	순위 측정	중앙값(median) 범위(range)
양적변수	등간척도	덧셈/뺄셈	산술평균(arithmetic mean) 분산(variance)과 표준편차(standard deviation)
	비율척도	곱셈/나눗셈	기하평균(geometric mean) 조화평균(harmonic mean)

기술통계

기술통계는 주로 중심성향과 산포성향으로 나뉘어 살펴본다.

- 중심성향(central tendency)는 자료를 대표하는(representative) 값이 자료의 가운데 어딘가에 위치해 있다라는 아이디어에 기반한다.
- 주로 세 가지 통계가 중심성향을 파악하기 위해 사용된다: (1) 평균(mean), (2) 중앙값(median), (3) 최빈값(mode)
- 산포성향(dispersion tendency)는 자료를 대표하는(representative) 값이 자료의 흩어진 정도에서 나타난다는 아이디어에 기반한다.
- 주로 세 가지 통계가 산포성향을 파악하기 위해 사용된다: (1) 범위(range), (2) 사분위수간 범위(interquartile range; IQR), (3) 분산(variance) 또는 표준편차(standard deviation)

기술통계

연습 8. income.sav 자료를 SPSS에서 불러들여 모든 변수들의 기술통계를 적절히 보고하시오.

기술통계

- Jamovi에서는 [Exploration]에서 [Descriptives]를 선택하여 기술통계를 산출할 수 있다.
- 이때 "Frequency tables"를 체크하면 빈도분포표를 함께 보여주므로 편리하다.
- 기술통계표를 꾸밀 때는 엑셀을 활용하는 것이 바람직하다. 기존 문헌에서 표를 어떻게 꾸몄는가를 보고 흉내내서 연습해야 한다.
- 표 안에 들어가야 하는 정보는 대체로 정해져 있다: (1) 평균(또는 비율), (2) 표준편차(또는 분산), (3) 최소값 및 최대값.
- 먼저 표를 깔끔하게 만들어 보고하고 이를 언어적으로 표현한다. "이 변수의 평균은 얼마이고 표준편차는 얼마이다. 최소값과 최대값은 각각 얼마와 얼마이다."
- 리커트 척도의 평균이나 표준편차 따위를 그대로 보고하는 것은 다소 좋지 않다 (Why?).

교차표와 χ^2 독립성 검정

교차표와 χ^2 독립성 검정

교차표는 어쩌면 가장 중요한 분석기법이다.

- 원칙적으로 교차표는 두 개의 질적 변수 사이의 관계를 분석하는데 사용한다.
- 하지만 일정한 정보의 손실을 감수한다면 양적 변수를 질적 변수로 얼마든지 변환할 수 있다(e.g., 숫자형 연령에서 범주형 연령으로).
- 그러므로 교차표는 (1) 두 개의 질적 변수, (2) 하나의 질적 변수와 하나의 양적 변수, (3) 두 개의 양적 변수 등 자료유형과 무관하게 다 사용할 수 있다.
- 즉 교차표는 둘 이상의 변수 간 관계를 분석하는데 있어 매우 탄력적이고 강력한 도구이다.

교차표와 χ^2 독립성 검정

- 교차표는 각 셀에는 빈도(frequency)를 보고하는 것에서 종종 출발한다. 하지만 사실 비율(percentage)을 보고하는 쪽이 해석에 훨씬 편리하다.
- 그런데 상대비율을 구할 때는 세 가지 방법을 상상해 볼 수 있다!

  (1) 행 합계(row total)로 표준화하는 방법
  (2) 열 합계(column total)로 표준화하는 방법
  (3) 총 합계(grand total)로 표준화하는 방법
- 그런데 우리는 관습에 따라 종종 독립변수 X에 해당하는 부분을 행(row)에 놓고, 종속변수 Y에 해당하는 부분을 열(column)에 놓는 경향이 있다.
- 이 경우 행 합계로 표준화하는 편이 해석에 편리하다(Why?).

교차표와 χ^2 독립성 검정

연습 9. religious.csv 자료를 SPSS에서 불러들여 종교 유무의 성별의 연관성을 드러내는 표를 작성하고 이를 적절히 해석하시오.

교차표와 χ^2 독립성 검정

- 좌측은 빈도를, 우측은 행 합계 비율을 나타낸다.

	비종교인	종교인	합계
남자	232	68	300
여자	124	97	221
합계	356	165	521

	비종교인	종교인	합계
남자	0.77	0.23	1
여자	0.56	0.44	1
합계	0.68	0.32	1

- 좌측은 열 합계 비율을, 우측은 총 합계 비율을 나타낸다.

	비종교인	종교인	합계
남자	0.65	0.41	0.58
여자	0.35	0.59	0.42
합계	1	1	1

	비종교인	종교인	합계
남자	0.45	0.13	0.58
여자	0.24	0.19	0.42
합계	0.68	0.32	1

교차표와 χ^2 독립성 검정

교차표 내용에 대해서도 유의성 검정을 수행할 수 있다.

- χ^2 분포를 따르는 확률밀도함수를 정의하기 위해서는 독립성 가정(independence assumption)이 필요하다.
- Karl Pearson의 이 가정을 절묘하게 이용하여 χ^2 독립성 검정(chi-square test of independence)을 개발했다.
- 독립성을 가정하고 이론적으로 계산한 기대빈도(expected frequency) E와 실제 교차표의 관찰빈도(observed frequency) O를 비교하여 너무 큰 차이가 나는지 살펴보는 방식이다.

교차표와 χ^2 독립성 검정

- 연습 10에서 사용한 자료로 χ^2 독립성 검정의 가설을 다음과 같이 세울 수 있다.

$$H_0 : \text{성별과 종교인 여부는 서로 독립적이다.}$$
$$H_a : \text{성별과 종교인 여부는 서로 독립적이지 않다.}$$

- 만일 귀무가설을 기각할 수 있었다고 할지라도 여성이 더 종교적인지 여부 등은 독립성 검정을 통해서는 알 수 없다.
- 단지 "성별과 종교성이 서로 독립적"이라는 귀무가설을 통계적으로 유의하게 기각할 수 있을 뿐이다.
- 대립가설 이상의 해석을 χ^2 독립성 검정에 멋대로 덧붙이지 않도록 꼭 주의해야 한다!

교차표와 χ^2 독립성 검정

연습 11. finedust.SAV 자료를 SPSS에서 불러들여 미세먼지로 인한 영향이 응답자 거주지에 따라 연관성이 있는지 여부를 나타내고 해석하시오.

교차표와 χ^2 독립성 검정

- SPSS에서는 [분석(A)]-[기술통계량(E)]-[교차분석(C)]를 선택한다.
- Jamovi에서는 [Frequencies]-[Independent Samples χ^2 test of association]를 선택한다.
- 교차표를 만들고 χ^2 독립성 검정을 수행할 때는 (1) 행과 열의 결정, (2) 표준화 방식, (3) χ^2 값의 적절한 해석에 주의를 기울이자.

교차표와 χ^2 독립성 검정

- 배경 변수별로 어학연수 및 영어 사교육 경험에는 어떤 차이가 있을까?

(단위: 명(%))

성별				
활동	구분	남	여	비고
어학연수 경험	유	900(16.1)	866(18.8)	$X^2 = 13.332$ df = 1 $p = .000$
	무	4692(83.9)	3728(81.2)	
	전체	5592(100.0)	4595(100.0)	
영어 사교육 경험	유	1451(25.9)	1473(32.1)	$X^2 = 45.996$ df = 1 $p = .000$
	무	4141(74.1)	3122(67.9)	
	전체	5592(100.0)	4595(100.0)	

아버지 교육수준					
활동	구분	중졸 이하	고졸	전문대졸	비고
어학연수 경험	유	283(11.8)	622(14.5)	861(24.5)	$X^2 = 200.628$ df = 2 $p = .000$
	무	2110(88.2)	3660(85.5)	2651(75.5)	
	전체	2393(100.0)	4282(100.0)	3512(100.0)	
영어 사교육 경험	유	599(25.0)	1233(28.8)	1092(31.1)	$X^2 = 25.588$ df = 2 $p = .000$
	무	1794(75.0)	3049(71.2)	2420(68.9)	
	전체	2393(100.0)	4282(100.0)	3512(100.0)	

가계소득						
활동	구분	200만 미만	200~400만	400~500만	500만 이상	비고
어학연수 경험	유	276(12.0)	671(16.3)	280(17.7)	532(24.4)	$X^2 = 124.259$ df = 3 $p = .000$
	무	2015(88.0)	3443(83.7)	1303(82.3)	1647(75.6)	
	전체	2291(100.0)	4114(100.0)	1583(100.0)	2179(100.0)	
영어 사교육 경험	유	528(23.0)	1170(28.4)	458(28.9)	766(35.2)	$X^2 = 80.236$ df = 3 $p = .000$
	무	1763(77.0)	2944(71.6)	1125(71.1)	1413(64.8)	
	전체	2291(100.0)	4114(100.0)	1583(100.0)	2179(100.0)	

김정숙. 2009. "대졸자들의 취업준비 활동의 차이 및 직업이행 효과." 교육과학연구 40(1): 141-165.

평균비교와 t 검정

평균비교와 t 검정

두 개 집단이 주어졌을 때 그 평균의 차이에 관해 가설검정할 수 있다.

- 남성의 연 평균소득과 여성의 연 평균소득의 차이
- 전업 공연예술인의 예술지원정책 만족도과 비전업인의 정책 만족도의 차이
- 진중문고 독서지도 프로그램을 이수한 군장병의 복무의욕과 그렇지 않은 군장병의 복무의욕의 차이
- 가출경험이 있는 중학생이 가진 비행경험 친구의 수와 가출경험이 없는 중학생이 가진 비행경험 친구의 수의 차이
- 집중강화 트레이닝을 받기 전 2군 선수들의 이전 평균기록과 이후 평균기록의 차이
- 처방약을 복용한 환자의 혈압과 위약을 복용한 환자의 혈압의 차이

평균비교와 t 검정

- 평균비교(mean comparison)는 양측검정과 단측검정으로 수행될 수 있다. 그런데 압도적으로 양측검정이 많이 사용된다(Why?).

$$H_0 : \mu_1 - \mu_2 = 0 \qquad H_0 : \mu_1 - \mu_2 \geq 0 \qquad H_0 : \mu_1 - \mu_2 \leq 0$$
$$H_a : \mu_1 - \mu_2 \neq 0 \qquad H_a : \mu_1 - \mu_2 < 0 \qquad H_a : \mu_1 - \mu_2 > 0$$

- 사실 0이 아닌 숫자로 가설검정을 할 수 있다. 하지만 0이 실무나 연구에서 많이 쓰이므로 이렇게 알아두어도 괜찮다.
- 이제 귀무가설을 세우고 적절하게 가설검정을 수행할 수 있다. 다만 자료구조에 따라 (2) 독립표본 t 검정(t test for independent samples) 또는 (3) 쌍체표본 t 검정(t test for paired samples)을 사용해야 한다.

평균비교와 t 검정

독립표본 t 검정과 쌍체표본 t 검정은 어떻게 다른가?

- 왼쪽은 쌍체표본(paired samples)으로 같은 사람에 대해 처방 전후(before and after)로 기록이 짝지어(paired) 있는 반면, 오른쪽은 독립표본(independent samples)은 처방(treatment) 실시 여부를 말해주는 더미변수(dummy variable)가 있다.

ID	BEFORE	AFTER
1	35	27
2	31	39
3	46	33
4	39	40
5	31	31

ID	TREATED	RECORD
1	0	35
2	0	31
3	0	46
4	0	39
5	0	31
1	1	27
2	1	39
3	1	33
4	1	40
5	1	31

평균비교와 t 검정

- 먼저 독립표본과 쌍체표본을 정확하게 식별해야 한다. 실무적으로는 독립표본이 압도적으로 많이 사용된다.
- 쌍체표본 t 검정은 그다지 사용되지 않는데 괜히 사람 헷갈리게 한다.
- 두 가지 자료 구조는 본질적으로 같다. 다만 구조화의 방식이 다를 뿐이다.
- 지난 시간에 설명한 자료의 재배열(data reshaping)을 통해 마음대로 자료 구조를 바꿀 수 있다.

평균비교와 t 검정

연습 12. self_ind.sav 파일을 SPSS로 불러오시오. 초등학교 3학년 학생의 네 가지 자기 개념들에 성차(gender difference)가 있을 것이라고 연구가설을 세우고 이를 95% 신뢰수준에서 검정하시오.

평균비교와 t 검정

- 통계적으로 적절한 가설은 다음과 같으며 모두 4개의 양측검정이 필요하다(Why?)

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

- 연구가설(research hypothesis)과 귀무가설(null hypothesis)을 혼동하지 말것.
- 자료를 열어보면 독립표본이므로 독립표본 t 검정이 적합하다.
- SPSS에서는 [분석]-[평균 비교]-[독립표본 T검정]을 선택하면 된다. 단 "집단변수"를 적절히 정의하자.
- Jamovi에서는 [T-Tests]-[Independent Samples T-Test]를 선택하면 된다.

평균비교와 t 검정

- 등분산(equal variances) 가정의 성립 여부에 따라 결과표를 적절히 해석하자.
- 사실 표본 크기가 조금만 커져도 등분산 가정 성립 여부와 무관하게 결과는 대체로 강건하다(robust).
- $|t|$ 값이 크다는 것은 그만큼 "귀무가설이 옳다는 전제 아래 그려진 표집분포" 위에 그런 극단적인 표본평균의 차이가 나올 확률, 즉 유의확률(p-value)이 작다는 것을 의미한다(Why?).
- "Sig. (2-tailed)" 부분에 우선 주목하자.

평균비교와 t 검정

- t 검정을 수행하고 결과표를 맵시있게 만드는 연습을 해야 한다.

	가톨릭 평균(SD)	개신교 평균(SD)	t-value	Sig. (2-tailed)
인구학적 요인				
성별	0.69 (0.46)	0.61 (0.49)	−1.66	0.10
연령	47.99 (16.91)	43.86 (16.47)	−2.49	.013
교육년수	12.84 (4.35)	12.87 (4.28)	0.07	0.94
주관적 계층의식	1.80 (0.50)	1.77 (0.50)	−0.50	0.62
소득수준 (월평균 가구소득:만원)	372.31 (264.5)	369.11 (347.2)	−0.097	.923
종교적 태도	**8.28 (1.84)**	**9.53 (2.14)**	**5.863**	**.000**
신앙심	6.18 (2.98)	7.37 (3.20)	3.734	.000
신(God)에 대한 믿음	9.45 (2.77)	10.23 (2.37)	3.149	.002
초월적 세계에 대한 믿음	8.46 (2.40)	9.61 (2.46)	4.668	.000
종교 절대주의	9.00 (2.14)	10.75 (2.79)	6.589	.000
배타성	**2.53(0.51)**	**3.00(0.61)**	**7.883**	**.000**
타종교인과의 결혼 반대	1.79 (0.80)	2.24 (1.01)	4.682	.000
유교전통에 대한 반-친화성	2.72 (0.89)	3.30 (0.80)	7.020	.000
토속신앙(부적, 사주)과의 반-친화성	3.09 (0.74)	3.45 (0.65)	5.308	.000
종교적 행위	**40.13 (58.17)**	**69.36 (69.18)**	**4.356**	**.000**
종교예식참여빈도(1년 기준)	52.52 (69.13)	91.41 (84.21)	4.800	.000
종교예식외참여빈도(1년 기준)	28.49 (61.45)	47.82 (72.79)	2.754	.006

조세희·이재혁. 2013. “왜 ‘느슨한’ 교회가 성장하는가?” 한국사회학 47(2): 61-100.

평균비교와 t 검정

유의성 검정에서 α 값을 미리 정해놓는 경우는 그다지 없다.

- 실무나 연구 상황에서는 먼저 t 값을 써놓고, 유의성 검정을 수행한 다음에 독특한 표식을 남긴다.
- 그 대신 신뢰수준이 99.9%일 때는 별 3개(***), 99%일 때는 별 2개(**), 95%일 때는 별 1개(*)를 통계량 뒤에 덧붙여 표기한다.
- 좀 더 구차해지고 싶을 때는 90% 신뢰수준에 대해 대거 1개(†)를 붙인다.
- 주의할 것은 이것이 관례에 지나지 않는다는 점이다!

평균비교와 t 검정

연습 13. medcond.SAV 자료를 SPSS에서 불러들여 각종 제한사항의 비율을 성별에 따라 비교하는 표를 작성하고 유의성 검정 결과를 보고하시오. 이때 표에는 (노령을 제외한) 모든 제한사항의 총합계 역시 성별로 비교하는 내용을 추가하시오.

평균비교와 t 검정

변수	여성			남성			격차	t값
	표본	평균/비율	표준편차	표본	평균/비율	표준편차		
골절, 관절부상	829	0.08	0.27	769	0.08	0.27	0.00	0.01
관절염, 류머티즘	828	0.16	0.37	769	0.08	0.27	0.08	5.09***
심장질환	828	0.05	0.21	769	0.03	0.17	0.02	1.64
호흡문제, 폐질환, 천식	828	0.04	0.21	768	0.05	0.21	-0.00	-0.09
뇌졸중	828	0.01	0.11	768	0.01	0.11	-0.00	-0.17
당뇨병	828	0.05	0.22	769	0.04	0.20	0.01	1.00
고혈압	828	0.10	0.30	769	0.10	0.30	0.00	0.17
등, 목의 문제	827	0.17	0.38	769	0.11	0.31	0.06	3.74***
암	828	0.01	0.12	769	0.01	0.10	0.01	1.00
치아 및 구강질환	829	0.18	0.39	768	0.21	0.41	-0.02	-1.13
시력문제	828	0.22	0.42	768	0.22	0.42	0.00	0.09
청각문제	828	0.04	0.21	768	0.07	0.25	-0.02	-2.09*
치매	827	0.01	0.08	768	0.00	0.05	0.00	1.05
우울/불안/정서상 문제	829	0.09	0.29	768	0.04	0.18	0.06	4.78***
정신지체	828	0.00	0.05	768	0.01	0.07	-0.00	-0.91
비만	828	0.10	0.30	768	0.08	0.27	0.02	1.36
기타	827	0.05	0.22	768	0.03	0.17	0.02	1.98*
총합계	825	1.38	1.94	766	1.16	1.66	0.22	2.43*

주: †p<.1, *p<.05, **p<.01, ***p<.001

일원분산분석

일원분산분석

분산분석의 기초적인 용어에 먼저 친숙해질 필요가 있다.

- 분산분석의 맥락에서 독립변수는 요인(factor)이라고 불리운다.
- 요인은 실험처리 그룹의 식별자(identifiers)가 되므로 종종 그룹화 요인(grouping factor)이라고도 불리운다. 물론 반드시 범주형 변수(주로 명목척도)이다(Why?).
- 종속변수는 결과(outcome) 또는 반응 변수(response variable)이라고 불리운다. 이것은 반드시 양적변수가 된다.
- 일원분산분석(one-way ANOVA)은 분산분석에 속한 여러 기법들의 가장 기초가 된다.
- 일원분산분석의 경우 분석에 사용되는 요인과 반응변수는 각각 하나씩이다.

일원분산분석

분산분석은 2개 이상의 모평균에 관한 가설검정에 사용된다.

- t 검정을 통해 두 모평균의 차이 $\mu_1 - \mu_2$에 관한 가설검정을 수행할 수 있었다.
- 분산분석을 통해서는 둘 이상의 모평균이 같은가 $\mu_1 = \mu_2 = ... = \mu_j$에 관한 가설검정도 수행할 수 있다.
- 분산분석은 t 검정보다 훨씬 일반화된 분석기법인 셈이다.
- 그러나 연구나 실무에서는 범주형 독립변수에서 주어진 범주의 수가 2개 일 때 (예컨대 성별, 성인 여부, 대졸 여부 등)는 주로 t 검정을, 주어진 범주의 수가 3개 이상일 때(예컨대 최종학력별, 지역별, 고용조건별 등)는 주로 분산분석을 사용한다.

일원분산분석

일원분산분석의 검정통계량은 F 값이다.

- 일원분산분석은 F 분포를 사용하며, 그 근본 원리는 (앞서 설명한) 두 모집단 분산 비율에 대한 가설검정과 같다.

$$\begin{aligned} F_{(k-1,n-k)} &= \frac{Variance_{between}}{Variance_{within}} \\ &= \frac{SS_{between}/(k-1)}{SS_{within}/(n-k)} \\ &= \frac{MS_{between}}{MS_{within}} \end{aligned}$$

- 분자는 실험처리에 의한 효과(treatment effect)로 볼 수 있고, 분모는 무작위오차(random error)로 볼 수 있다(Why?).

일원분산분석

- 계산을 진행하면서 분산분석표(ANOVA table) 안에 각각의 숫자를 채워넣으면 된다. 아! 물론 계산은 컴퓨터가 한다.

Variance	SS	df	MS	F
Between groups	$SS_{between}$	$(k-1)$	$MS_{between}$	F
Within groups	SS_{within}	$(n-k)$	MS_{within}	
Total	SS_{total}	$(n-1)$	MS_{total}	

- 실제 표에서는 F 옆에 유의확률(p-value)이나 임계값(critical value) 등도 추가해서 써넣는 경우도 있다.

일원분산분석

일원분산분석 절차는 굉장히 정형화되어 있다.

- 귀무가설은 "모든 그룹에 걸쳐 평균이 같다"이다. 예를 들면 다음과 같다.

  (1) 인지된 기후변동의 심각성 점수의 평균에 최종학력별 차이는 없다.
  (2) 출신지역별로 정치적 보수주의 점수 차이가 없다.
  (2) 기업내 직급 수준에 따른 인적자원개발 프로그램 만족도 차이는 없다.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$

- 그 대립가설은 "적어도 하나의 그룹에서 평균값이 다르다"이다.

$$H_a : (\mu_1 \neq \mu_2) \text{ or } (\mu_1 \neq \mu_3) \text{ or } ... \text{ or } (\mu_{j-1} \neq \mu_j)$$

- "모든 그룹에 있어 평균값이 다르다($H_a : \mu_1 \neq \mu_2 \neq \mu_3 = ... \neq \mu_j$)"가 아님에 주의할 것(Why?)!

일원분산분석

- 당연히 “최종학력이 높아질수록 인지된 기후변동의 심각성 점수도 함께 높아진다”와 같은 해석은 본래 일원분산분석으로 할 수 없는 과잉해석이 된다.
- 이른바 다중비교검정(multiple comparison test)은 임의로 선정된 두 그룹 중 어느 그룹에서 평균값이 더 높은가를 살펴보는 기법이다. 교과서에 따라서는 사후비교(post hoc comparison)라고 부르기도 한다(우리는 다루지 않는다).
- 일원분산분석은 실제 분석상 분산의 비율을 비교하고 있음에도 불구하고, 가설 설정은 평균에 대해 이루어진다는 점도 주의해야 한다.
- SPSS에서는 [분석]-[평균 비교]-[일원배치 분산분석]을 선택한다. 이때 “기술통계”를 옵션에서 선택하는 쪽이 좋다.
- Jamovi에서는 [ANOVA]-[One-Way ANOVA]를 선택한다.

일원분산분석 연습

연습 14. vulnerab.sav 자료를 SPSS에서 불러들여 인지된 한국사회의 위험 취약성이 정치적 진보·보수 성향에 따라 다른지 여부를 확인하시오. 적절한 유의성 검정을 수행하고 표를 만들어 설명하시오.

일원분산분석

- 여성의 가사노동시간은 연도 별로 달라졌을까?

		시간(일/분)							
								통계적 검증	
		1999 (t1)	2004 (t2)	2009 (t3)	2014 (t4)	2019 (t5)	t5-t1	F	Scheffe, α=0.05 동일집단
	가사노동시간	239.3	223.7	218.1	209.8	194.5	-44.8	177.0**	1/23/4/5
	가정관리시간	190.0	175.5	168.9	166.4	158.6	-31.5	171.8**	1/2/34/5
	음식준비	101.6	94.1	90.8	90.0	82.7	-19.0	154.0**	1/2/34/5
	의류관리	28.8	24.4	21.8	15.6	17.7	-11.0	467.3**	1/2/3/5/4
	청소및정리	36.8	34.2	32.0	37.7	32.1	-4.6	75.6**	41/2/53
	주거및가정용품관리	3.8	3.9	4.6	0.6	0.7	-3.1	274.7**	3/21/54
여	차량관리유지	0.1	0.1	0.1	0.1	0.2	0.0	0.9	-
성	반려동식물관리	-	-	-	3.2	5.2	2.0	102.6**	-
	상품및서비스구입	16.0	15.4	16.7	16.8	17.6	1.6	12.5**	543/431/2
	기타 가정관리	3.0	3.3	2.9	2.3	2.4	-0.5	22.6**	21/13/54
	가족돌봄시간	49.3	48.2	49.2	43.4	35.9	-13.4	59.4**	132/4/5
	미성년가구원돌보기	43.1	41.1	42.6	37.2	32.8	-10.3	38.5**	132/4/5
	그외가족돌보기	6.2	7.1	6.6	6.2	3.1	-3.1	47.4**	2314/5
	n	22837	17063	10486	14947	14930			

서지원·기은광·고선강. 2021. "한국 여성과 남성의 세대별 가사노동시간의 변화(1999-2019): 가정관리 및 가족돌봄 시간을 중심으로." 가족자원경영과 정책 25(2): 53-78.

일원분산분석

분산분석 계열의 기법은 사실 굉장히 다양한 테크닉들을 광범위하게 포괄한다.

- 일원분산분석에서 일원(one-way)은 독립변수가 하나임을 뜻한다. 독립변수가 두 개인 경우 이원분산분석(two-way ANOVA)이 되고, 그보다 많은 경우 다원분산분석(multi-way ANOVA)이 된다.
- 그 밖에도 다변량분산분석(multivariate ANOVA 또는 MANOVA), 공분산분석(analysis of covariance 또는 ANCOVA)이나 다변량공분산분석(multivariate analysis of covariance 또는 MANCOVA) 등 다양하게 있으나 이것들은 이 수업의 수준을 한참 벗어난 것이므로 다루지 않는다.