

S_05 Choosing Statistical Test

Statistical Analysis

Statistical Assumptions

- 일반적으로 Statistical Test들의 공통적인 가정들:
 1. Independence of observations (a.k.a. no autocorrelation)
 - Test에 포함된 variable (또는 observation) 들은 서로 관련이 없다
 - 여러가지 측정치 (Predictor variable) 들이 서로 독립적이다
 2. Homogeneity of variance
 - 비교되는 여러 group들의 variance가 비슷하다
 - 어떤 group의 variation이 다른 것들에 비해 매우 크면 test의 유효성에 문제가 된다
 3. Normality of data: the data follows a normal distribution
 - 제시된 data는 normal distribution을 따른다

Parameters for Selecting Statistical Test

- Statistical Test 선택 시 고려할 parameter들:
 - Types of variables
 - Categorical (nominal, ordinal, binary)
 - Quantitative (continuous, discrete)
 - Test에 고려되는 Group (Sample) 의 개수
 - Group들 간의 관계 (independent vs dependent)
 - Normality Condition: Data의 normality 여부 (quantitative data일 경우만)
 - Homogeneity of Variance Condition: 여러 Group들의 분산이 비슷해야 (quantitative data일 경우만)

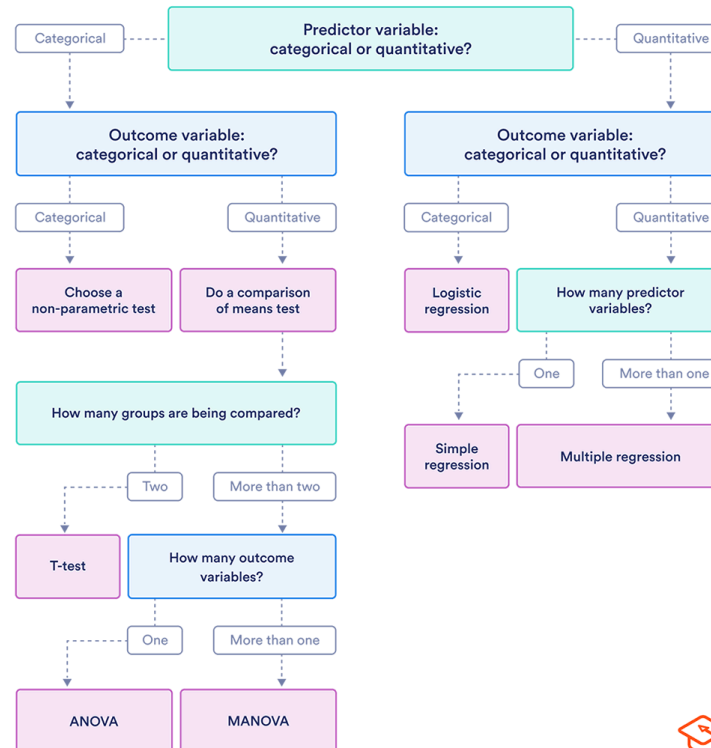
Parametric? or Nonparametric?

- Parametric test는 일반적으로 data가 가져야 할 조건이 더 엄격함
- Nonparametric test를 해야 하는 경우 (예외 있음)
 - Homogeneity of Variance Condition을 위배하는 경우
 - Normality Condition을 위배하는 경우
 - Predictor variable과 Outcome variable이 모두 Categorical type 일 경우

Flowchart (1/3)

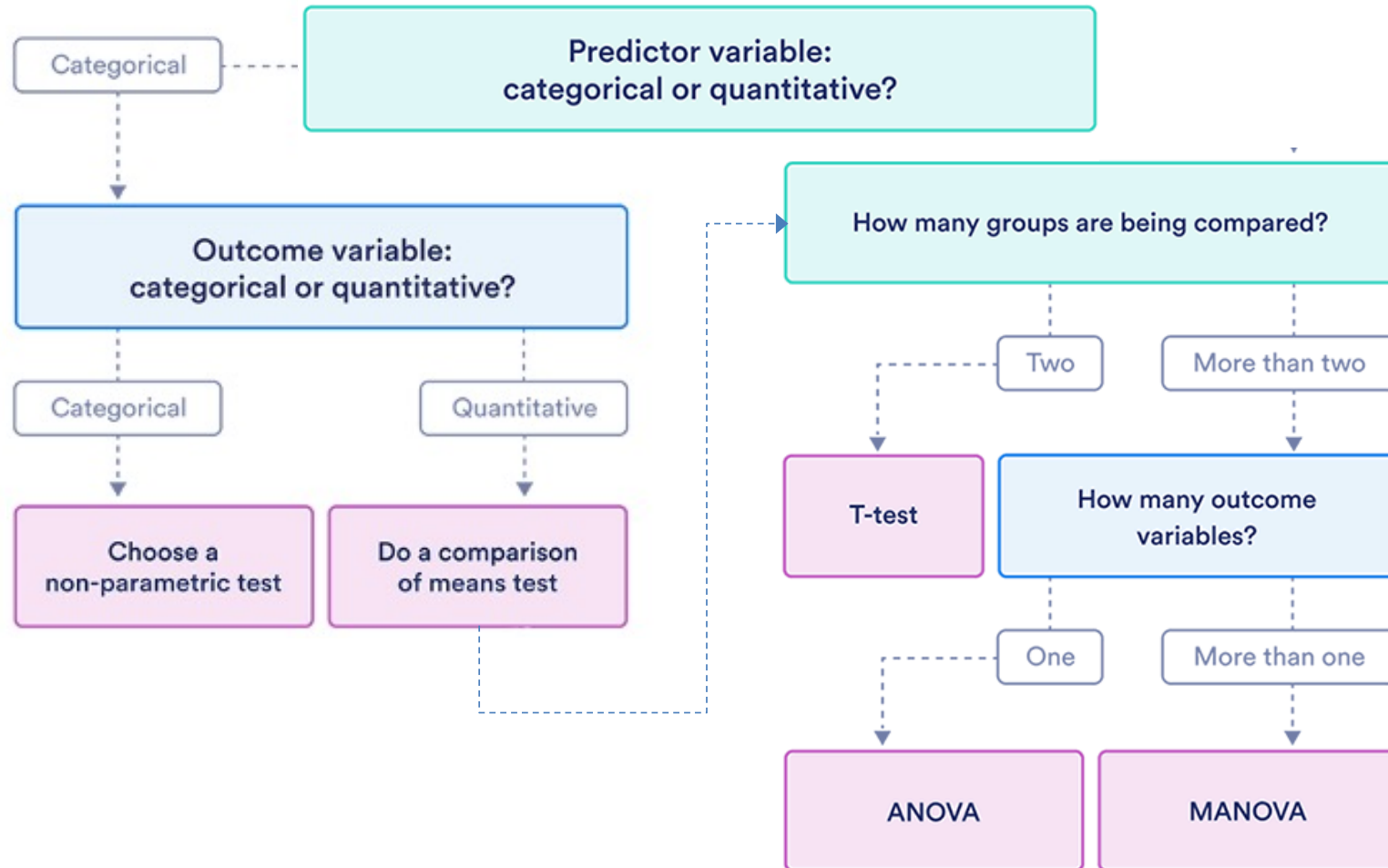
Choosing a statistical test

This flowchart helps you choose among parametric tests

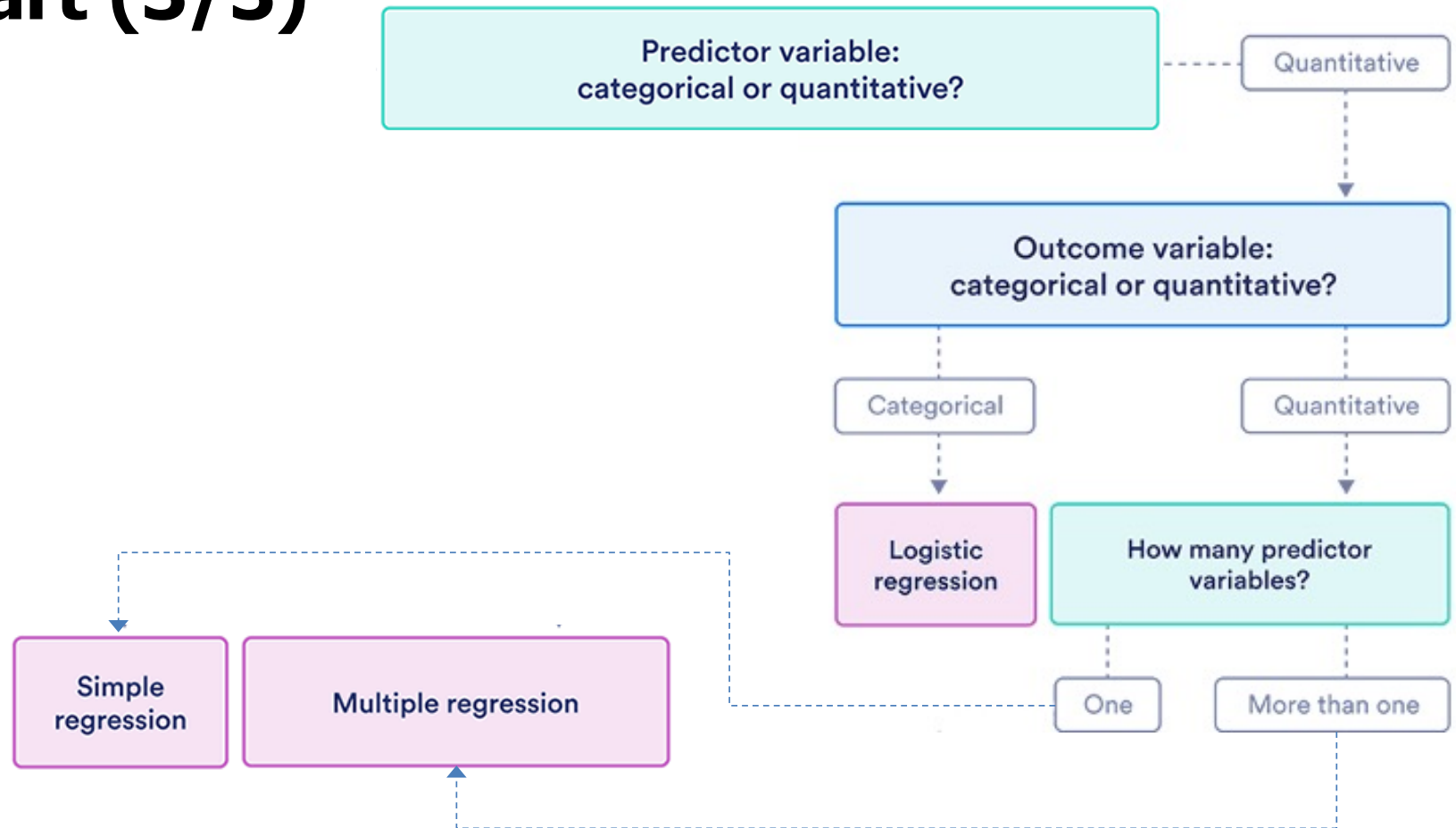


variable type과
group의 개수를
기준으로 함

Flowchart (2/3)



Flowchart (3/3)



Comparison Test (Parametric)

- Group mean의 차이를 test

Test	Predictor Variable	Outcome Variable	Example RQ
Paired t-test	Categorical 1 Predictor	Quantitative groups come from the same population	두 가지 다른 시험 준비 프로그램이 같은 반 학생의 평균 시험 점수에 미치는 영향?
Independent t-test	Categorical 1 Predictor	Quantitative groups come from different populations	서로 다른 두 학교 학생의 평균 시험 점수의 차이는 어떻게 되나요?
ANOVA	Categorical 1 or more Predictor	Quantitative 1 outcome	수술 후 세 가지 진통제를 투여한 환자들의 평균 통증 수준에는 어떤 차이가 있을까요?
Welch's ANOVA	Categorical 1 or more Predictor	Quantitative 1 outcome	Variance 조건 위배 시 ANOVA 대신 사용 가능
MANOVA	Categorical 1 or more Predictor	Quantitative 2 or more outcome	꽃의 종류가 꽃잎 길이, 꽃잎 너비, 줄기 길이에 미치는 영향은 무엇인가요?

Regression Test (Parametric)

- Cause-and-effect relationships

Test	Predictor Variable	Outcome Variable	Example RQ
Simple Linear Regression	Continuous 1 Predictor	Continuous 1 Outcome	소득이 장수에 미치는 영향
Multiple Linear Regression	Continuous $n \geq 2$ Predictor	Continuous 1 Outcome	서로 다른 두 학교 학생의 평균 시험 점수의 차이는 어떻게 되나요?
Logistic Regression	Continuous	Binary	수술 후 세 가지 진통제를 투여한 환자들의 평균 통증 수준에는 어떤 차이가 있을까요?
Multinomial Logistic Regression	Continuous	3 or more category	직원의 근무 연수에 따라 직무 만족도가 우수, 보통, 미흡 중 어디에 속하는지에 차이가 있는가?
Ordinal Logistic Regression	Continuous	Ordinal	환자의 운동 시간에 따라 건강 상태가 매우 좋음, 좋음, 보통, 나쁨 중 어디에 속하는지에 차이가 있는가?

Correlation Test (Parametric)

- Check whether variables are related

Test	Variables	Example RQ
Pearson r	2 Continuous Variables	위도와 기온의 관계는?

Nonparametric Test (1/3)

Test	Conditions	Predictor Variable	Outcome Variable	Instead of
Wilcoxon Rank-Sum test (Mann-Whitney U test)	Normality 위반	Categorical independent 2 groups	Quantitative groups come from different populations	Independent t-test
Welch's t-test	Variance 위반	Categorical independent 2 groups	Quantitative groups come from different populations	Independent t-test
¹⁾ Wilcoxon Signed-rank test	Normality, Variance 위반	Categorical 2 groups	Quantitative groups come from the same population	Paired t-test
Sign test	Normality 위반	Categorical	Quantitative	One-sample t-test

1) variance만 위반시에는 Paired t-test를 사용 가능

Nonparametric Test (2/3)

Test	Conditions	Predictor Variable	Outcome Variable	Instead of
Kruskal–Wallis H	Normality, Variance 위반	Categorical 3 or more groups	Quantitative	ANOVA
Friedman test	Normality, Variance 위반	Categorical 3 or more groups	Quantitative 2 or more outcome variables	MANOVA
Multivariate Rank test (ART ANOVA)	Normality, Variance 위반	Categorical 3 or more groups	Quantitative 2 or more outcome variables	MANOVA
ANOSIM	Normality, Variance 위반	Categorical 3 or more groups	Quantitative 2 or more outcome variables	MANOVA

Nonparametric Test (3/3)

Test	Conditions	Predictor Variable	Outcome Variable	Instead of
Spearman's r	Normality, Variance 조건 위반	Quantitative	Quantitative	Pearson's r
Chi square test of independence	이 경우는 대부분 Nonparametric 을 씀	Categorical	Categorical	Pearson's r
Mann-Whitney U test (= Wilcoxon Rank-Sum test)	독립 변수와 로그 오즈 비율 간에 선형 관계가 아닐 때 관측치가 독립적이지 않을 때	Quantitative	Categorical (Binary) independent 2 groups	Logistic Regression
Kruskal-Wallis H test	독립 변수와 로그 오즈 비율 간에 선형 관계가 아닐 때 관측치가 독립적이지 않을 때	Quantitative	Categorical (Multipartite) independent 3 groups	Multinomial Logistic Regression
Cochran-Armitage Trend Test	독립 변수와 로그 오즈 비율 간에 선형 관계가 아닐 때 관측치가 독립적이지 않을 때	Quantitative	Categorical (Binary or Multipartite)	Multinomial Logistic Regression

Example) Therapy

- Predictor variable (1): Therapy
 - A, B, C
- Outcome variables (3)
 - Before, Middle, After (Blood Pressure)
- Test Procedure
 - a. csv 데이터파일 load
 - b. variance similarity test
 - c. regularity test
 - d. 위의 b, c를 통과했을 경우 MANOVA 실행
 - e. 위의 b, c를 통과하지 못했을 경우 Friedman test 실행
 - f. 위의 d 또는 e의 결과를 바탕으로 post-hoc test 진행 (Bonferroni correction 포함)

Therapy	Before	Middle	End
A	165	145	140
A	155	142	133
A	138	135	140
A	150	147	145
A	149	148	149
B	135	130	115
B	145	150	132
B	170	166	150
B	138	139	130
B	144	143	132
C	165	145	118
C	139	140	125
C	141	140	126
C	149	146	125
C	135	134	118

Code and Data

- <https://github.com/iklee99/StatCode>
- 05_MANOVA_EX.py
- datasets/medical_data_for_manova3.csv
- datasets/medical_data_for_manova4.csv

Example 1 (1/4)

- datasets/medical_data_for_manova4.csv

Variance Test Results: {'Before': 0.9469256208780882, 'Middle': 0.2294124482689248, 'End': 0.530342792699958}

Normality Test Results: {'Before': 0.08428268879652023, 'Middle': 0.14866092801094055, 'End': 0.5409590601921082}

Failed Variance Tests: [] # Variance Test Passed

Failed Normality Tests: [] # Normality Test Passed

Example 1 (2/4)

MANOVA Results:

Multivariate linear model

Intercept	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.0071	3.0000	10.0000	463.2512	0.0000
Pillai's trace	0.9929	3.0000	10.0000	463.2512	0.0000
Hotelling-Lawley trace	138.9753	3.0000	10.0000	463.2512	0.0000
Roy's greatest root	138.9753	3.0000	10.0000	463.2512	0.0000

Intercept Table은 모델 자체의 유의미성을 평가. Intercept는 모델이 종속 변수들의 평균 값에 영향을 미치는지를 나타냄. Intercept가 의미 있는 값이라면 전체 모델이 통계적으로 유의미하다는 것을 보여줌. Intercept의 p-value < 0.05 이므로, 모델이 종속 변수들에 대해 통계적으로 유의미한 설명력을 가진다는 것을 의미

Therapy	Value	Num DF	Den DF	F Value	Pr > F
Wilks' lambda	0.0966	6.0000	20.0000	7.3923	0.0003
Pillai's trace	1.0342	6.0000	22.0000	3.9263	0.0081
Hotelling-Lawley trace	7.9994	6.0000	11.7143	12.8618	0.0001
Roy's greatest root	7.8264	3.0000	11.0000	28.6968	0.0000

Therapy 테이블은 독립 변수(치료법)가 종속 변수들에 미치는 영향을 평가. p-value < 0.05 이므로 치료법 간에 유의미한 차이가 있다는 결론

Example 1 (3/4)

Post-hoc test for Before:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A	B	-5.0	0.7788	-25.1626	15.1626	False
A	C	-5.6	0.7342	-25.7626	14.5626	False
B	C	-0.6	0.9	-20.7626	19.5626	False

Post-hoc test for Middle:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A	B	2.2	0.9	-12.6499	17.0499	False
A	C	-2.4	0.9	-17.2499	12.4499	False
B	C	-4.6	0.6863	-19.4499	10.2499	False

Post-hoc test for End:

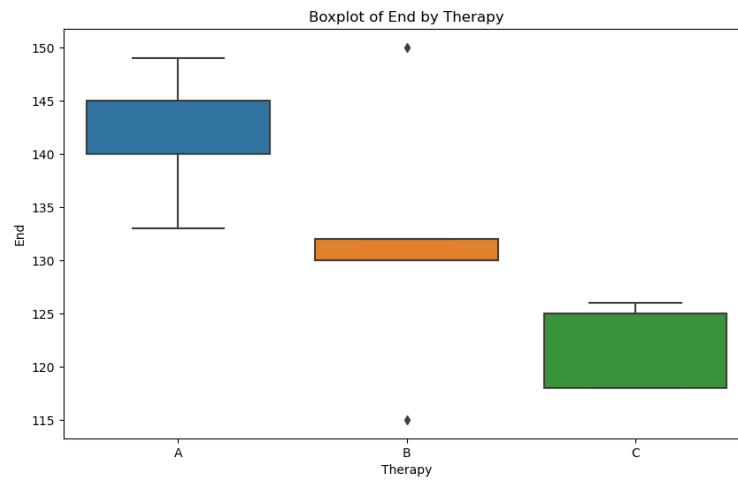
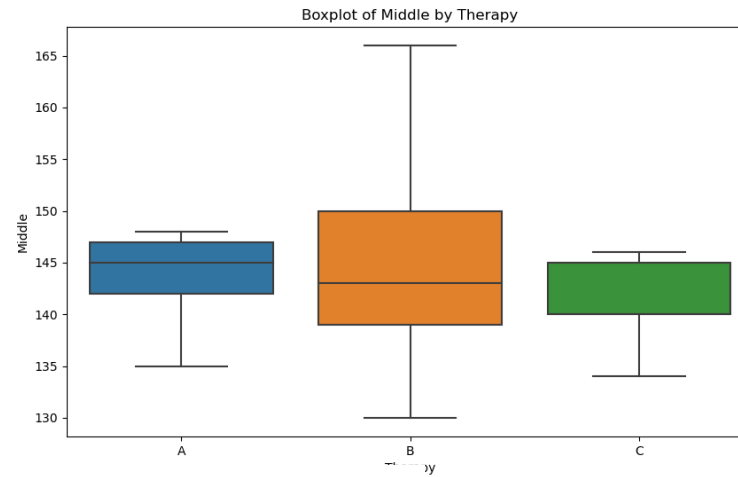
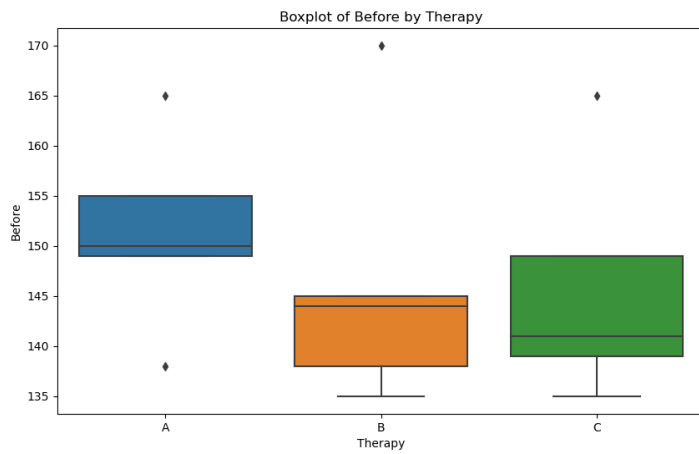
Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A	B	-9.6	0.202	-23.6021	4.4021	False
A	C	-19.0	0.0091	-33.0021	-4.9979	True
B	C	-9.4	0.2141	-23.4021	4.6021	False

Post-hoc test

p-value < 0.05, 즉, Therapy A와 C의 차이 End group에 대해서만 유의미 하다.

Example 1 (4/4)



Example 2 (1/3)

Variance Test Results: {'Before': 0.04910363798530076, 'Middle': 0.5900278357791713, 'End': 0.0017611953781425309}

Normality Test Results: {'Before': 2.1902758362557506e-06, 'Middle': 4.5216318540042266e-05, 'End': 0.057111285626888275}

Failed Variance Tests: ['Before', 'End']. Variance Test Fail

Failed Normality Tests: ['Before', 'Middle'] Normality Test Fail

Friedman Test Results:

(FriedmanchisquareResult(statistic=1.73333333333332, pvalue=0.42035038450868467),

FriedmanchisquareResult(statistic=0.9333333333333371, pvalue=0.6270890852730551))

모든 pvalue > 0.05 이므로 Therapy A, B, C 간에 효과의 차이는 없다.

Example 2 (2/3)

Post-hoc test for Before:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A	B	-26.219	0.2782	-67.2766	14.8387	False
A	C	-3.6206	0.9	-44.6782	37.4371	False
B	C	22.5984	0.3842	-18.4592	63.6561	False

Post-hoc test for Middle:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A	B	-8.3623	0.8825	-52.2953	35.5708	False
A	C	-15.5862	0.6551	-59.5192	28.3468	False
B	C	-7.224	0.9	-51.157	36.7091	False

Post-hoc test for End:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
A	B	88.551	0.1409	-22.4987	199.6007	False
A	C	55.5773	0.4525	-55.4724	166.627	False
B	C	-32.9737	0.7351	-144.0234	78.076	False

- Post-hoc test
- 모든 case의 p-value > 0.05
- 모든 Therapy 방식이 Before, Middle, End 모두에게 효과가 없다.
- H0를 reject할 수 없다.

Example 2 (3/3)

