

FAQ

- Answers and explanations

삼성의료원
삼성생명과학연구소
통계지원팀

통계지원팀에서는 지난해에 이어 2009년 한 해 동안 원내 연구자들이 문의한 통계적 사항에 대하여 답변한 내용 중 일부를 발췌하여 정리하였습니다.

이 내용들은 질문과 답변 형식으로 기술되었으며, 등장하는 주요 용어들은 마지막 쪽에 가나다순으로 정리되어 있습니다. 또한 관심 있는 용어를 PDF 파일 내에서 찾기 도구를 사용하여 빠르게 찾을 수 있습니다.

연구진행시 궁금하신 사항들에 대해서는 주저하지마시고 문의해주시면 답변과 함께 향후 FAQ 에 포함되어 다른 연구자들에게도 도움이 될 수 있는 기회가 될 것입니다. 또한 FAQ 게재에 대한 건의가 있으시면 언제든지 통계지원팀으로 알려주시면 감사하겠습니다.

2010.02.03

통계지원팀

(T. 02.3410-6760/3410-6499)

Contents

1. 자료 탐색 - 기술통계량	1
2. 두 변수 간 연관성 검정	3
2.1 범주형(순서, 명목) 자료 vs 범주형(순서, 명목) 자료	3
2.2 범주형(순서, 명목) 자료 vs 연속형 자료	9
2.3 연속형 자료 vs 연속형 자료	15
3. 공분산 분석	16
4. 회귀 분석	17
5. 로지스틱 회귀분석, 오즈비	22
6. 생존분석	27
6.1 Median survival time	27
6.2 생존율 추정 및 구간 생존 곡선 비교 검정	28
6.3 Cox proportional hazard model (Cox regression)	32
6.4 Time dependent Cox proportional hazard model	34
7. 반복측정 자료 분석	35
8. 진단법 평가(민감도, 특이도, ROC), 경계값(Cut-off value)	38
9. 타당도와 신뢰도	42
10. 유의확률(P-value), 본페로니 보정(Bonferroni's correction)	43
11. 표본의 크기 (sample size)	46
12. 기타	50
13. 주요 용어	55

1. 자료 탐색 - 기술통계량

Q. 측정된 변수 값(양의 값)의 평균보다 표준편차(SD)가 더 큰 경우, 예를 들어 평균이 5이고, 표준편차가 15인 경우 $\text{mean} \pm \text{SD} = 5 \pm 15$ 로 기술하기가 적절하지 않은 것 같은데, 어떻게 하면 좋을까요?

A. 측정된 변수 값이 양인데도 $\text{mean}-\text{SD}$ 가 음의 값이 나오는 현상은 그 변수의 분포가 한쪽으로 많이 치우쳤을 때(skewed) 생길 수 있습니다. 이러한 자료의 기술통계는 대표 값으로서 중간값(median)을, 퍼진 정도로서 사분위 범위(Interquartile range: IQR)로 기술하는 것이 적절합니다. 그러나 자료가 평균을 중심으로 대칭(symmetric)인 경우 $\text{mean} \pm \text{SD}$ 로 기술하는 것이 좋습니다.

Q. 원 자료의 변수 값이 아닌 로그 변환한(log-transformed) 변수를 통계분석 하였습니다. 이 때, 기술통계를 논문 Table에 기술하고자 하면, 로그 변환한 변수의 $\text{mean} \pm \text{SD}$ 를 기술해야 하는지요?

A. 의학논문의 Table에는 통계분석 이전의 각 변수별 분포를 확인하는 차원에서 보통 원 자료의 분포를 써 넣기 때문에 원 자료의 $\text{mean} \pm \text{SD}$ 를 쓰는 것이 좋습니다(일반 독자들은 log를 취한 수치에 대하여 분포를 확인하는 것이 익숙하지 않기 때문입니다). 그러나 skewed 분포 자료를 로그 변환한 경우는 원 자료의 $\text{mean} \pm \text{SD}$ 를 제시하는 것보다는 median과 사분위 범위를 제시하는 것이 좋습니다. 그리고 Table 상의 P-value에 대하여 Table 각주(footnote)에서 로그 변환한 변수에 대하여 분석하였다는 것을 제시하는 것이 좋겠습니다.

Q. 일반적으로 median, 25percentile, 75 percentile, 최소값, 최대값을 이용하여 그리는Box-plot을 가지고 자료를 설명하거나 분포를 판단할 수 있는지 알고 싶습니다.

A. Box plot은 자료의 중심을 median으로 나타내면서 자료가 퍼진 양상도 함께 보여주므로 자료 그 자체가 어떻게 생겼는지를 보여주는 좋은 plot입니다. 따라서 box plot으로 자료를 설명하는 것에는 특별히 문제는 없습니다. 또한 각 군의 box plot을 한 그래프에 제시함으로써 두 군 간의 분포 차이를 대략적으로 확인하는데 유용하게 사용됩니다. 즉, box-plot을 이용한 자료의 설명은 max와 min에 분포하는 자료의 극단치와 median 주위의 50% 정도를 간단히 확인하는 것과 더불어 두 group 간의

분포 차이가 있을지 시각적으로 짐작을 할 수 있습니다. 그러나 일반적으로 도표는 자료의 전체적인 경향이나 구조 이해에 활용하고 객관적인 결론은 통계적인 분석을 이용하게 됩니다.

Q. 연속형 변수에 대해 기술통계량을 제시할 때 평균과 SD만 제시하면 되나요? 중앙값과 사분위 범위를 같이 제시해 주어야 하나요?

A. 자료가 대칭이면 평균과 SD로 자료의 분포 특성을 설명해 줄 수 있습니다. 하지만, 자료가 비대칭이면 평균이 대표값으로 적절하지 않습니다. 이와 같은 경우에는 평균과 표준 편차 대신에 중앙값과 사분위 범위를 제시해 주어야 합니다. 그런데 기술통계량 표를 만들다보면 어떤 변수는 정규성 가정을 만족하는데 다른 변수는 정규성 가정을 만족하지 못하여 표를 일관되게 만들지 못할 수 있으므로 이런 경우에는 평균, 표준편차, 중앙값, 사분위 범위를 함께 제시해 주세요.

Q. 연구기간을 제시할 때 median으로 제시하려고 합니다. 그런데 대부분의 환자들이 초반에 사망하여, 전체 연구기간의 median이 너무 짧게 나옵니다. 이런 경우에는 초기 사망자들을 제외하여 median을 제시해야 되나요?

A. 연구기간은 전체 연구시작시점부터 전체 연구 종료시점까지의 기간을 말합니다. 따라서 아마도 연구기간의 median이 아니라 환자들의 추적기간의 median을 의미하는 듯합니다. 그렇다면 각 환자들마다 추적시간의 median을 제시할 수 있습니다. 사망이라는 event가 초기에 발생되어 추적기간이 너무 짧은 subject 라고 할지라도, 해당 subject에 대한 추적기간은 사망하기까지의 시간이 맞습니다. 이를 삭제해서는 안 됩니다. median을 제시할 경우 전체 연구기간에 비해 짧다고 생각된다고 하더라도, median 의 값은 그대로 제시해야 합니다. 다만, 사분위수나, 최대, 최소값도 함께 제시하여, 추적기간의 median 이 짧게 나온 이유에 대해서 코멘트를 해주면 될 것 같습니다. 예를 들면 연구대상 환자들은 2000년 1월 1일부터 2004년 12월 31일까지 위암으로 수술 받은 환자들로 이들은 2009년 12월 31일까지 추적 조사되었다. Median 추적기간은 32개월이었으며, 사분위 범위 및 범위는 각각 (12개월, 40개월)과 (3개월, 96개월)이었다라고 기술하면 되겠습니다.

2. 두 변수 간 연관성 검정

2.1 범주형(순서, 명목) 자료 vs 범주형(순서, 명목) 자료

Q. 범주 명목형 변수와 순서형 변수 간 비교 검정하려면 어떤 것을 사용하는 것이 좋을까요?

A. 무엇을 분석하시고자 하는지 반드시 알고 정의한 분석 변수의 속성에 맞게 분석 방법을 찾고 분석하셔야 됩니다. 예를 들어, 치료군 유무를 나타내는 명목형 변수와 연령범주(20대, 30대, 40대, 50대)를 나타내는 순서형 범주를 가지고 두 군간 연관성을 보시고 하는 경우 다음과 같은 가설들을 세워 볼 수 있고 각각의 가설에 대한 분석 방법들은 다음과 같습니다.

No	분석방법	가설
1	Fisher's exact test 혹은 Chi-square test (이 경우 연령 범주는 순서형 변수이지만 명목형 범주로 취급하여 분석하는 방법입니다.)	치료 유무 두군 간 연령 범주 비율이 다른가. (즉 치료 유무 간 20대 상대적 비율, 30대 상대적 비율, 40대 상대적 비율, 50대 상대적 비율이 다른가.)
2	Fisher's exact test 혹은 Chi-square test (이 경우 연령 범주는 순서형 변수이지만 명목형 범주로 취급하여 분석하는 방법입니다.)	연령군 (연령 범주) 간 치료 유무 비율이 다른가.
3	Cochran-Armitage test	연령이 높을수록 “치료=유”일 비율이 높을가. (Trend test)
4	Mann-Whitney test	두 군간 연령이 다른가. (중간값 비교)

Q. 2범주 이상의 명목형 변수와 순서형 변수를 비교 검정하려면 어떤 방법을 사용하는 것이 좋을까요?

A. 예를 들어 설명 드리도록 하겠습니다. 만일 세 군(치료A, 치료B, 치료C)과 연령 범주(20대, 30대, 40대, 50대)간 연관성을 보시고자 한다면 다음과 같은 가설들을 생각해 볼 수 있습니다. 이들 중에서 해당하는 연구 가설을 택하여 분석하시면 됩니다.

No	분석방법	가설
1	Fisher's exact test 혹은 Chi-square test (이 경우 연령 범주는 순서형 변수이지만 명목형 범주로 취급하여 분석하는 방법입니다.)	치료 세 군간 연령 범주 비율이 다르다.
2	Fisher's exact test 혹은 Chi-square test (이 경우 연령 범주는 순서형 변수이지만 명목형 범주로 취급하여 분석하는 방법입니다.)	연령 범주간 치료 세 군 비율이 다르다.
3	Jonckheere's test (존키어 검정)	A → B → C로 갈수록 연령이 높다.(Trend test)
4	Kruskal-Wallis test(크루스칼 월리스 검정)	세 군간 연령이 다르다. (중간값 비교)

Q. 독립 두 군 간에 breast metastasis 유무 비율이 다른지를 살펴보고자 합니다. 그런데 전체 raw data은 없고 2X2 table에 대한 빈도수만 알고 있습니다. 어떻게 분석해야 되나요?

A. 빈도수만 알고 있어도 카이제곱 검정 방법을 이용할 수 있는데 통계 SW에서(ex. SAS, SPSS) 빈도수만큼 가중치를 주어 분석할 수 있습니다. (단, 기대 빈도수가 5보다 작은 칸이 하나라도 있으면 Fisher's exact test를 이용하여 분석하세요.) 다만, 이 경우 주의할 점은 primary tumor 진단 후 전이 유무에 대한 추적기간이 환자마다 상이할 경우 단순히 전이 유무 자료만을 가지고 분석할 수 없고 추적시점까지 전이 유무 및 추적시점까지 기간 값을 가지고 생존분석을 수행해야 합니다.

Q. 2008년 3.4.5월 3개월 동안 중환자실에 입원하여 사망한 비율과(A군) 2009년 3.4.5월 3개월 동안 중환자실에 입원하여 사망한 비율 간(B군) 차이에 대해서 검정하고자 합니다.

어떻게 분석하면 되나요? 각 군 모두 추적 기간이 동일하고 follow up loss된 자료도 없었습니다.

A. 독립된 두 군간 비율 검정 시 카이제곱 검정 방법을 이용하여 분석 할 수 있습니다. A, B를 입력한 그룹 변수와 중환자실에 입원하여 사망한 여부를 나타내는 변수를 이용하여 카이제곱 검정을 이용하여 분석하세요. 단, 기대빈도가 5보다 작은 셀이 하나라도 있으면 Fisher's exact test를 수행해야 합니다.

Q. Fisher's exact test를 한 후 결과에서 다음과 같은 message가 뜨면 그 결과는 빈칸으로 두는 것이 좋을 까요? 아님 뭔가를 잘못된 것일까요? (SPSS을 이용하여 분석하였습니다.)

- Cannot be computed because there is insufficient memory

- Cannot be computed because the time limit has been exceeded

A. 위와 같이 message가 떴다고 한다면 자료 구조상 분할표에서 적은 도수를 가지는 cell들이 너무 많기 때문입니다. 이런 경우 이론상으로는 분석할 수 있지만 실제적으로는 컴퓨터 메모리가 불충분하여 분석할 수 없는 경우에 해당됩니다. 논문에 그 결과를 빈칸으로 두거나 message써서 분석이 불가능했다고 쓰는 것은 안 좋아 보입니다. 만일, 같은 범주로 묶어 볼 수 있다면 묶어서 다시 Fisher's exact test 혹은 Chi-square test를 하는 방법을 생각해 볼 수 있습니다.

예를 들어, 하루에 피우는 담배량 그룹 간 뇨의 코티닌 수준 비율이 다른지를 볼 때 다음과 같이 자료가 구성되어 있는 경우를 생각해 보겠습니다.

담배량(3범주): 담배량<5, 5≤담배량<10, 10≤담배량

코티닌(5범주): 코티닌<5, 5≤코티닌<10, 10≤코티닌<20, 20≤코티닌<30, 30≤코티닌

그러면, 3 X 5 table을 만들어 볼 수 있을 텐데, 이것을 다음과 같이 2 X 3 table로 다시 re-categorization 해서 분석해 볼 수 있습니다.

담배량(2범주) - 담배량<10, 10≤담배량

코티닌(3범주) - 코티닌<10, 10≤코티닌<20, 20≤코티닌

이때 주의하실 점은 같은 범주로 묶을 시에는 임상적으로도 해석이 가능하게 유사한 항목끼리 범주로 묶으셔야 됩니다.

Q. N X M 테이블($N \geq 3$ 또는 $M \geq 3$ 인 경우)에서 N 구간 M 범주 비율이 다른지 비교 검토하려고 합니다. 어떻게 분석하면 되나요?

A. 두 가지 예를 들어 설명하겠습니다.

(Ex1) 만일 2군을 나타내는 변수 A와 명목형 세 범주를(a, b, c) 나타내는 변수 B를 가지고 2X3 table을 만들어 2 구간 a, b, c 비율이 다른지 비교 검토할 경우, 기대 빈도수를 체크하여 5보다 작은 칸이 하나라도 있으면 Fisher's exact test를 이용하고 그렇지 않으면, Chi-square test를 이용하여 분석하면 됩니다. 이때, 유의수준을 5%로 하는 경우 P-value가 0.05보다 작으면 2 구간 a, b, c 범주 비율이 다르다고

할 수 있고, 어느 범주에서 다른지 알기 위해 사후 검정을 시행합니다. (물론, P-value가 0.05보다 크면 2 군간 a, b, c 범주 비율이 다르다고 할 수 없으므로 사후 검정도 할 필요가 없습니다.) 사후 검정은 세 범주를(a, b, c)을 2 범주씩 짝지어 (a vs b, a vs c, b vs c) 세 개의 2X2 table로 만들어 Chi-square test 혹은 Fisher's exact test를 이용하여 비교 검정 하시면 됩니다. 이때, 주의하실 점은 3 범주 중 2 범주씩 짝지어 3번 비교 검정했으므로 다중 비교 검정 시 발생 될 수 있는 Type I error가 최대 3배까지 커지므로 P-value를 보정해 주어야하는 점입니다. 방법은 산출된 P-value에 다중 검정 회수인 3만큼 곱하여 보정합니다. (이것을 Bonferroni's method에 의해 보정되었다라고 합니다.) (Figure 1)

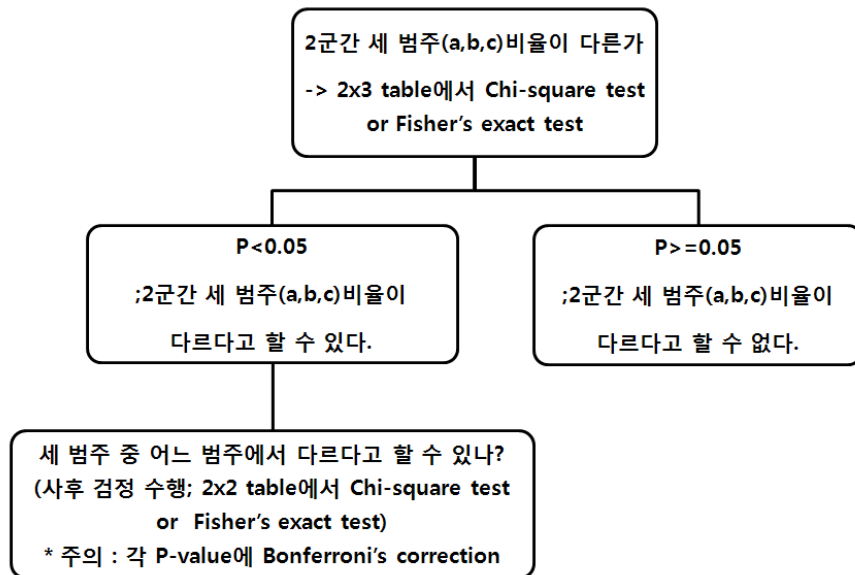


Figure 1. 2X3 table에서 군간 비율 검정

(Ex2) 이번에는 조금 더 복잡한 경우에 대해서 생각해 보겠습니다. 세군을 나타내는 변수 A와 명목형 세 범주를(a, b, c) 나타내는 변수 B를 가지고 3X3 table을 만들어 세 군간 a, b, c 비율이 다른지 비교 검정을 할 경우입니다. 우선 기대 빈도수를 체크하여 5보다 작은 칸이 하나라도 있으면 Fisher's exact test를 이용하고 그렇지 않으면, Chi-square test를 이용하여 분석하세요. 이때, P-value가 0.05보다 작으면 세 군간 a, b, c 비율이 다르다고 할 수 있으므로 어느 군간 다른지 사후 검정을 시행해야 합니다. 사후 검정은 3군을 2군씩 짝지어 세 개의 2X3 table로 만들어 Chi-square test나 Fisher's exact test를 이용하여 비교 검정 하시면 됩니다. 이때, 주의하실 점은 2군씩 짝지어 3번 비교 검정했으므로 다중 비교 검정 시 발생 될 수 있는 Type I error가 최대 3배까지 커지므로 P-value를 보정해 주어야하는 점입니다. 산출된 P-value에 곱하기 3을 하여 보정합니다. (이것을 Bonferroni's method

에 의해 보정되었다고 합니다.) 그리고, $P\text{-value} < 0.05$ 이면 두 군간 a, b, c 비율이 다르다고 할 수 있으므로 두 군간 어느 범주에서 다른지 사후 검정을 또 시행해야 됩니다. 사후 검정은 3 범주를 2 범주씩 짝지어 세 개의 2X2 table로 만들어 Chi-square test나 Fisher's exact test를 이용하여 비교 검정 하시면 됩니다. 역시 주의하실 점은 2범주씩 짝지어 3번 비교 검정했으므로 P-value를 보정해 주어야 한다는 점입니다. 이때에는, 앞서서 2X3 table에 대한 P-value를 보정 할 때 곱하기 3을 했기 때문에 2X2 table에 대한 다중 비교 검정 시 발생 될 수 있는 Type I error가 최대 9배까지 커지므로 산출된 P-value에 곱하기 9를 하여 보정합니다. (참고로, 사후 검정 후 Bonferroni's correction외에 다른 보정방법을 적용할 수도 있는데 예를 들어 Permutation method 등이 있습니다.) (Figure 2)

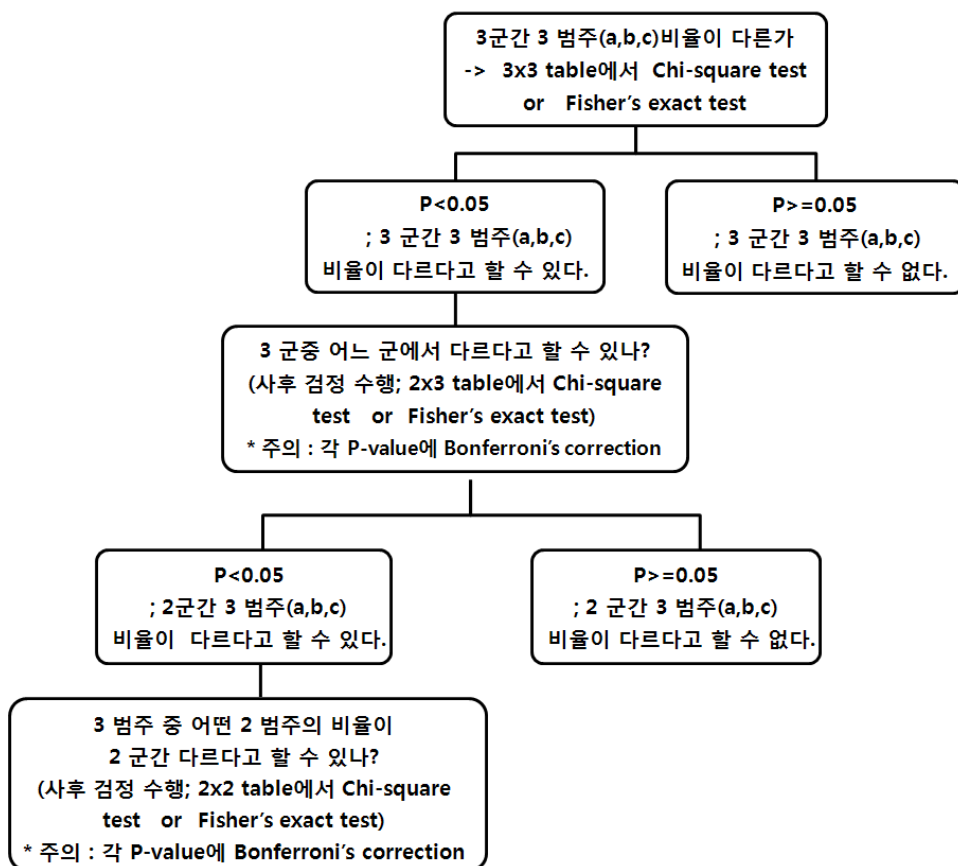


Figure 2. 3X3 table에서 군간 비율 검정

Q. 등급(혹은 학력, 근무지)간 행위 점수가 다른지를 살펴보고자 합니다. 등급, 학력, 근무지는 범주형 변수들이고, 행위점수는 (아주 나쁘다, 나쁘다, 보통, 좋다, 아주 좋다)와 같이 5점 척도로 측정하였습니다. 이와 같은 경우 등급(혹은 학력, 근무지)간 행위점수(median)가 다른지 비교 검정 시에는 어떤 분석 방법을 써야 되나요?

A. Kruskal-Wallis test (크루스칼 왈리스 검정 방법)로 분석하시면 됩니다. 그리고 P-value가 유의수준 (일반적으로 0.05) 보다 작아 유의하다고 할 수 있으면 전체 행위 점수에 대해 rank를 매겨서 다중비교법을 수행하면 어느 등급 간 차이가 있는지를 알 수 있습니다.

Q. 세 그룹(1, 2, 3)간에 사망률이 서로 차이가 있음을 증명하려면 어떻게 해야 할까요?

A. 사망 여부에 대하여 모든 피험자가 추적 기간이 동일하여 특정 시점에서의 사망 여부 관측이 모두 가능하다면 Chi-square test 또는 Fisher's exact test를 이용하여 분석할 수 있습니다.

(주의: 특정 시점에서의 사망 여부에 대해 이 특정 시점만큼 F/u이 되지 않은 생존자가 있다면 특정 시점에서 이 생존자의 사망 여부를 알 수 없으므로 survival analysis을 해야 됩니다.)

2.2 범주형(순서, 명목) 자료 vs 연속형 자료

Q. T-test와 Mann Whitney test의 차이점이 무엇인가요?

A. 독립 두 군간 연속형 수치가 다른지 비교 검정을 하고자 할 때 T-test 혹은 Mann Whitney test을 이용하여 분석할 수 있습니다. 이때, 두 군 모두 정규성 가정을 만족하면 T-test을 사용하시고 한 군이라고 정규성 가정을 만족하지 못하면 Mann Whitney test을 이용하여 분석하세요. 이렇게 분석할 변수가 어떤 분포를 하고 있어야한다는 가정을 요구하는 통계적 방법을 모수적 방법, 그렇지 않은 경우를 비모수적 방법이라고 합니다. 따라서 T-test는 모수적 방법, Mann-Whitney test는 비모수적 방법입니다.

Q. 12시간과 24시간 두 시점에서 측정한 자료가 있습니다. 두 시점 간 차이가 실험군과 대조군 간에 차이가 있는지 보기 위해서는 어떻게 해야 되나요?

A. 두 시점 간 차이를 계산하고 두 군 각각에 대해 정규성 가정을 검정한 후, 두 군 모두 정규성 가정을 만족하면 차이값을 가지고 T-test을 이용하여 두 군간 비교 검정하시면 됩니다. 하지만, 한 군이라도 정규성 가정을 만족하지 못하면 Mann-Whitney 검정을 수행 하세요.

Q. 분석변수의 자료 값에서 특정 상수만큼 뺀 후, 다시 분석하고자 합니다. 이때, 기존에 분석했던 T-test 결과와 달라지나요?

A. 표본의 크기가 이전 자료와 달라지지 않았다면 특정 상수만큼 빼서(혹은 더해서) 다시 분석하여도 T-test의 P-value에는 변화가 없습니다. 즉 결과는 달라지지 않습니다. 물론 SD(표준편차)에도 변화는 없습니다. 다만, 평균(mean), 중앙값(median), 최소값(min), 최대값(max) 등은 특정 상수를 뺀 만큼 수치가 달라집니다.

Q. Wilcoxon's two sample test, Wilcoxon's signed rank test, Wilcoxon's rank sum test 이 세 방법의 차이가 뭔가요?

A. Wilcoxon's two sample test은 Wilcoxon's rank sum test(윌콕슨 순위합 검정)와 같은test입니다. 이 test 방법은 Mann-Whitney test와 동일한 검정 결과를 제공합니다.

Mann-Whitney test은 윌콕슨의 순위합 통계량과 동치인 Mann Whitney U 통계량을 연구하여 Mann-Whitney test가 제안되었기 때문에 서로 독립인 두 그룹간의 연속형 변수 혹은 순서형 변수를 비교 검정하는 비모수적인 방법으로 동일한 검정 결과를 제공해 주는 방법이라고 할 수 있습니다.

Wilcoxon's signed rank test (윌콕슨의 부호 순위합 검정)는 서로 짝지은 두 그룹 간 연속형 변수를 비교하는 검정이라고 할 수 있습니다. 예를 들어 해열제 효과를 연구하기 위해 열이 있는 환자를 대상으로 해열제 투여 전과 투여 후간 체온을(연속형 변수) 비교 시, 측정값들의 차이가 정규성 가정을 만족한다고 할 수 없을 때 고려할 수 있는 비모수적 검정 방법이라고 할 수 있습니다.

Q. 독립 세군인 control, A, B의 각 그룹마다 연속 변수를 측정하여 control vs A, control vs B 군 간 비교 검정을 하고자 합니다. 각각 T-test를 수행하면 되나요?

A. 각각 T-test를 통해 control vs A, control vs B간 비교 검정을 할 수는 있습니다. 하지만, 각 test마다 동일한 "control" 그룹을 이용하여 검정한 두 검정 간에는 서로 독립이라고 할 수 없기 때문에 Type I error가 최대 2배까지 증가하므로 이를 보정해주어야 합니다. 보정해주는 방법은 다음과 같습니다. T-test를 통해 산출된 P-value에 독립이 아닌(서로 연관된) 검정 회수인 2를 곱하여 P-value를 다시 계산하면 됩니다. 이를 Bonferroni's correction이라고 합니다. 그런데 이렇게 P-value를 보정해주면 너무 conservative하게 결과를 보정하게 되는 단점이 있으므로, 다중비교법 중 control 대비 비교법인 Dunnett's test로 control vs A, control vs B 간 비교 검정을 수행하는 대안이 있습니다. 이것은 T-test with Bonferroni's correction보다 검정력이 더 우수한 장점이 있습니다. 그러므로 control vs A와 control vs B 비교 검정시 Dunnett's test를 사용하는 것을 추천합니다. 방법은 control, A, B 군 모두 정규성 가정을 만족하면 다중 비교 검정 방법 중의 하나인 Dunnett's test를 이용하여 분석하는 것입니다. 만일, 어느 한 군이라도 정규성 가정을 만족하지 않으면 세군 모두 합쳐 순위를 매긴 후, 순위를 매긴 자료를 가지고 Dunnett's test를 이용하여 분석할 수 있습니다.

Q. 세 군간 연속형 변수의 수술전후 차이 값이 다른지를 비교하고자 합니다. 무슨 방법을 사용하면 되나요? 이때, 각 그룹 마다 수술전후 차이 값은 정규성 가정을 만족 하지 않았습니다.

A. 세 군간 비교 검정에는 Kruskal-Wallis test(크루스칼 왈리스)로 검정한 결과를

보시고 P-value가 유의수준(일반적으로 0.05) 보다 작아 유의하다면 (세 군간 차이값이 모두 같지는 않다고 할 수 있으면) 사후검정을 통해 어느 군간 다른지 비교하셔야 합니다. 사후 검정 분석 방법은 다음과 같습니다. 세 군을 모두 합친 차이값 자료를 가지고 rank를 매긴 후, rank를 매긴 자료를 가지고 다중비교법을 수행하면 됩니다.

Q. 검사결과 세 군간(반응 없음, 부분반응, 양성) 단백질 수치가 다른지 보고 싶습니다. 어떤 분석 방법을 이용해야 되나요?

A. One-way ANOVA 혹은 Kruskal-Wallis test을 이용하면 3군간 단백질 수치가 다른지 검정 할 수 있습니다. 이때, 등분산 가정을 만족하고 세 군 각각 마다 정규성 가정을 만족하면 One-way ANOVA를 용하여 검정을 하고, 가정이 하나라도 만족되지 않은 경우에는 Kruskal-Wallis test를 이용하여 분석합니다. 만일, One-way ANOVA로 분석한 후 P-value가 유의수준(일반적으로 0.05) 보다 작아 유의하다면 (세 군간 단백질 수치가 모두 같지는 않다고 할 수 있으면) 사후검정을 통해 어느 군간 다른지 비교할 수 있습니다. 만일, Kruskal-Wallis test로 분석한 결과를 보시고 P-value가 유의수준(일반적으로 0.05) 보다 작아 유의하다면 (세 군간 단백질 수치가 모두 같지는 않다고 할 수 있으면) 세 군을 모두 합친 자료에 대해 순위를 매긴 후, 순위를 매긴 자료를 가지고 다중비교법을 수행합니다.

Q. 서로 독립인 세 군(A, B, C)간 평균비교 검정(ANOVA)을 했는데(혹은 median비교; Kruskal Wallis test) overall P-value가 유의수준 0.05보다 크게 나와 세 군간 평균이(혹은 중간값) 다르다고 할 수 없었습니다. 세 군간 평균이(혹은 중간값) 다르다고 할 수 없으면 어느 군에서 평균(혹은 중간값)이 다른지 살펴보지 않고 분석을 끝낼 수 있지만, 그래도 다중비교법을 이용하여 각 두 군의 쌍(pair)별 평균비교(혹은 중간값 비교) 검정을 해 보았습니다. 그런데 다른 두 군간 비교 검정에서는 유의한 결과가 없었으나 A vs B 간 비교 검정 시 P-value가 0.05보다 작아 유의한 결과가 나왔습니다. 이것을 어떻게 이해해야 될까요?

A. (서로 독립인 세 군(A, B, C)간 평균비교 검정(ANOVA)을 기준으로 말씀 드리겠습니다. 중간값 비교도 같은 방식으로 이해할 수 있습니다.)

네 그럴 수 있습니다. 또한 반대의 경우도 있을 수 있습니다. 왜냐하면 ANOVA 검정의 가설과 가정은 사후 검정에 이용되는 다중비교법에 따라 가설과 가정과 동일하지 않기 때문입니다. 즉 ANOVA 검정과 다중비교검정은 서로 다른 목적을 가지고 있으

며, 알고리즘 또한 다르다는 것입니다. 세 군 간의 overall ANOVA F test의 귀무가설(H_0)은 “세 군의 평균이 차이가 없다($\mu_1 = \mu_2 = \mu_3$)” 이고, 우리가 보이고 싶은 대립가설(H_A)은 “세 군 중 적어도 두 군 간 평균의 차이가 있다($\mu_1 \neq \mu_2$ or $\mu_1 \neq \mu_3$ or $\mu_2 \neq \mu_3$)” 입니다. Overall ANOVA F test는 이 귀무가설에 대해 기각할 만큼 충분한 증거가 있는지를 test하고자 합니다. ANOVA test의 일반적인 사후 검정은 다중 비교 검정(multiple comparison test)을 통해 세부적으로 어느 두 군 쌍의 평균이 다른지를 결정하고, 그들이 얼마나 다른지를 추정하고자 합니다. 사후 검정의 다중 비교 검정방법은 크게 두 가지로서, 모든 가능한 두 군 쌍에 대하여 평균을 비교하는 방법(LSD, Bonferroni, Tukey, Scheffe, Duncan 등)과, 하나의 대조(control)군에 대해 다른 나머지 군들의 평균을 비교하는 방법(Dunnett 등)이 있습니다. 첫 번째 사후 검정방법의 귀무가설은 $H_{01}: \mu_1 = \mu_2$, $H_{02}: \mu_1 = \mu_3$, $H_{03}: \mu_2 = \mu_3$ 이고, 두 번째 사후 검정방법의 귀무가설은(ex. 1군이 대조군이라 가정했을 때) $H_{01}: \mu_1 = \mu_2$, $H_{02}: \mu_1 = \mu_3$ 이며, 각각에 대하여 (Adjusted) T-test, Multiple range test 등의 다양한 방법으로 Type I error(실제 비교하는 두 군의 평균 차이가 없음에도 불구하고 두 군의 평균 차이가 부풀려져서 유의성을 갖게 되는 통계적 오류)를 보정하는 다중비교 검정을 수행합니다. 또한, Bonferroni, Tukey와 Scheffe 방법(특히, Bonferroni와 Scheffe 방법)은 experimentwise error rate(전체 군 간 분석에 대해서 적어도 하나의 Type I error가 생길 확률) 보정을 하므로, 이를 보정하지 않는 Fisher's LSD나 Duncan 방법보다 더 보수적인(즉, 사후 검정 시 잘못된 유의성을 갖게 되는 통계적 오류를 더욱 엄격하게 줄이는) 다중비교 검정방법으로 알려져 있습니다. 이와 같이 overall F test와 사후 검정은 분석하고자 하는 가설 및 분석방법이 다르고, 사후 검정 시 어떤 다중비교 검정방법을 적용하느냐에 따라 다중비교로 인한 Type I error를 보정하는 정도가 달라서 분석결과가 조금씩 다르게 나올 수는 있습니다.

최초 연구계획 단계에서 세 군 간의 평균의 차이가 있는지 (즉 모든 평균들이 같거나 적어도 한 그룹의 평균이 다른 평균들과 다른가)를 test하고자 하였고, overall ANOVA F test에서 유의한 결과가 나오지 않는다면, 평균차이의 충분한 증거가 없기 때문에 세 군의 평균이 차이가 있다고 할 수 없다고 해석합니다. 이 경우, 세 군 중 어느 두 군 pair의 평균이 다른지를 추정하는 다중비교검정에 의한 사후검정을 수행하는 것은 의미가 없습니다. Fisher's Protected LSD 방법에 따르면, 세 군 간 overall F test에서 유의한 결과가 나오지 않으면, 더 이상 사후분석을 수행하지 않습니다. 그러나 만약 overall ANOVA F test 결과 유의성이 검출되면 “평균이 다른 그룹이 적어도 하나가 있다”가 이 검정의 결론입니다. 그런 다음 어느 그룹이 다른지를 알아보기 위해 다중비교법으로 넘어가는 것입니다. 그러나 이 때 overall ANOVA F test 결과 유의하였더라도 다중비교법 적용 결과 차이가 있는 그룹이 나오지 않을 수도 있고, 그 반대의 경우도 있을 수 있습니다. 이는 두 검정 방법의 목적

이 다르고 요구하는 가정이 다르기 때문입니다. 즉 서로 다른 검정법 이므로 결과가 반드시 일치할 필요는 없다는 것입니다.

하지만, 최초 연구계획 단계에서 어떤 그룹 간 평균이 차이가 있는지를 알아보는 것이 목적이라면 ANOVA 검정을 수행하지 않고 바로 다중비교법을 수행합니다. 또한, 연구계획 단계에서부터 여러 군 pair 중 일부 특정 pair만을 선택하여 평균차이를 확인하고자 하는 연구목적이 있었다면, 다중비교에 의한 Type I error를 보정하는 정도가 달라져서 검정의 유의성이 달라질 수도 있습니다. 따라서 연구계획 단계에서의 주 연구 목적과 가설에 따라 다른 방법으로 분석하고 연구결과를 해석하는 것이 좋겠습니다.

Q. kruskal-Wallis test로 분석한 후 수행한 Sigma-Stat 과 SPSS 의 다중비교 검정 결과가 다르게 나옵니다. 예를 들어 LSD을 이용한 SPSS에서는 차이가 있다고 나오는데 Bonferroni's method을 이용한 Sigma-Stat에서는 차이가 없다고 나옵니다. 어떤 패키지가 맞는 것이지요?

A. 다중비교 시 1종 오류 상승을 보정하기 위한 방법들(LSD, Bonferroni, etc)을 적용하게 되는데, Sigma-stat 에서는 Bonferroni's method를 사용하였고, SPSS 에서는 LSD method를 이용했기 때문에 각 통계패키지에서 사용된 multiple comparison method 가 다르기 때문입니다. 서로 다른 다중비교법은 서로 다른 power를 가지고 있습니다. LSD method 에 비해 Bonferroni's method는 실험 오류율 (experimentwise error rate, 전체검정의 오류율)를 더 잘 관리하는 대신 검정의 power가 떨어지는 방법입니다. LSD 는 비교별 오류율 만을 관리합니다.

만약, 1종 오류 즉, 네 그룹 간에 차이가 실제로 없는데도 불구하고 차이가 있다고 결론을 내릴 오류율을 낮은 수준으로 일정하게 관리하는 것이 매우 중요하다면, Bonferroni's method 의 결과를 사용해야 합니다. LSD 방법은 전체 실험의 오류율을 관리하지 못하고 각 쌍별 비교에서 오류율만 관리하므로, 유의수준을 조정하여, 추가로 분석을 해야 합니다.

결론적으로 Bonferroni's correction으로 제1종 오류를 보정할 경우 Sigma Stat 의 결과가 올바르게 나온 것이나, 검정력(power)이 떨어져서, 실제로 존재하는 차이를 통계적 검정으로 밝혀내지 못한 것일 수 있습니다. sample size가 작을 경우에는 그 정도가 더 심해집니다. 따라서 3군을 비교할 때 보다 4군을 비교할 때 더 많은 sample size가 확보되어야 할 필요가 있습니다.

Q. SPSS로 T-test를 수행하면 95% CI of difference가 나오는데 어떤 의미를 갖는지요?

A. 신뢰구간(confidence interval)이란 어떤 값을 신뢰할 수 있는 범위를 나타내는데, T-test후 산출되는 95% CI of difference는 두 독립 집단의 평균값 차이에 대한 95% 신뢰구간을 말합니다. 이것은 t 분포를 이용하여 계산되며, 95% 신뢰구간이라는 것은 우리가 제시하는 신뢰 구간에 실제 평균 차이 값이 포함될 가능성이 95%라는 뜻이 됩니다.

2.3 연속형 자료 vs 연속형 자료

Q. 변수들 간의 연관성을 보고자 하는데 변수들 간의 원인-결과를 명확하게 정의할 수 없는 경우에 (로지스틱)회귀분석을 수행할 수 있는지요?

A. 변수들 간의 원인-결과를 명확하게 정의할 수 없는 경우, 회귀분석은 역학적인 해석이 어렵기 때문에 상관분석이나 혼란변수 효과를 보정한 편상관분석(partial correlation coefficient analysis)을 수행하는 것이 좋습니다. 이 때 만약 범주의 수가 세 개 이상인 명목형 변수가 (예를 들어 4개 범주) 있다면 (범주 수 - 1)만큼의 (예를 들어 3개의) 더미변수를 만들어 분석해야 합니다. 더미변수는 a, b, c, d와 같이 4범주에 대하여 a인 경우를 기준이라고 한다면 세 개의 더미 변수(dummy_var1, dummy_var2, dummy_var3)에서, a는 (0, 0, 0), b는 (1, 0, 0), c는 (0, 1, 0) 그리고 d는 (0, 0, 1)과 같이 변수화 할 수 있습니다. 추가적으로 일부 실행문 또는 package에서는 범주를 자동으로 인식하는 경우도 있습니다.

Q. 0~30점 scale로 측정한 변수와 0~10점 scale로 측정한 변수 간 연관성을 보고 싶습니다. 어떤 방법을 이용하여 분석해야 되나요?

A. 원칙적으로는 이 두 변수가 이변량 정규분포(bivariate normal distribution)을 따르면 피어슨 상관분석을 그렇지 않으면 스피어만 상관분석을 사용합니다. 그러나 이변량 정규분포에 대한 check가 쉽지 않은 경우 우선 각 연속형 변수에 대해 정규성 검정을 한 후, 두 변수 모두 정규성 가정을 만족하면 피어슨 상관분석방법을 이용하시고 한 변수라도 정규성 가정을 만족하지 못하면 스피어만 상관 분석방법을 이용하세요.

Q. 연속형 변수로 측정한 두 변수가 있습니다. 그런데 두 연속형 변수 중 한 변수가 정규분포를 따르지 않아서 두 연속형 변수 간 상관성을 스피어만 상관분석(Spearman correlation analysis)으로 확인을 했습니다. 그런데, Spearman이나 Pearson correlation에서는 R^2 를 구할 수 없는 것이 맞는 건가요? R^2 는 정규분포를 따르는 경우에 선형회귀분석(Linear regression analysis)으로 구하는 것이 맞는 건가요?

A. 네, 맞습니다. 부연하자면, R^2 는 회귀분석에서 결과변수에 대한 회귀식의 설명력을 의미합니다. 그러나 상관분석은 단지 변수 간 상관관계를 분석하는 것이므로 R^2 는 산출하지 않습니다.

3. 공분산 분석

Q. ANCOVA(analysis of covariance)가 무엇인지요?

A. ANCOVA(공분산 분석)은 그룹 간 연속형 반응 변수에 대한 차이를 보고자 할 때 교락 요인을 보정하여 분석하는 것을 말합니다. 이때, 교락 요인은 연속형 변수 혹은 범주형 변수 모두 해당 될 수 있습니다.

예를 들어, 성별, 나이가 그룹을 나타내는 독립변수와 연관성이 있고, 결과변수에도 영향을 끼치는 경우 성별, 나이와 같은 변수는 교락 변수라고 할 수 있는데, 교락 변수인 성별, 나이를 보정하여 그룹 간 연속형 반응 변수 값이 다른지를 분석해야 순수하게 그룹 간 비교 검정을 할 수 있습니다. 이와 같이 수행하는 분석 방법을 ANCOVA(공분산 분석) 라고 합니다.

Q. 독립된 두 군간 기억력에 관련된 여러 개의 반응 변수들(모두 연속형 수치)을 비교하는 연구를 하고 있습니다. 반응변수들이 기억력에 관련된 변수들이라 모두 연령과 연관이 있을 수 있어서 반응변수와 연령(연속적 수치)간 상관성을 검정해 보니, 또한 두 군간 연령이 유의하게 달랐습니다. 그러면 각 반응변수가 두 군간 다른지를 검정할 때 연령을 보정한 다음에 분석해야 하나요?

A. 두 군간 연령이 유의하게 달랐기 때문에 수집된 환자들의 연령이 두 군간 다르다고 할 수 있습니다. 또한 반응변수들과 연령간 연관이 있을 수 있다고 하셨기 때문에, 각 반응변수들의 두 군간 비교검정에서 연령을 보정하는 방법인 ANCOVA를 적용하셔야 합니다.

ANCOVA를 적용하실 때 각 반응변수가 정규성을 만족하지 않는다면 정규변환(로그, 로그 로그 변환, 제곱근변환, 지수변환, 제곱변환, 세제곱변환, 역수변환 등)을 적용한 뒤 ANCOVA를 적용해야 합니다.

Q. 다른 변수들을 보정하여 두 군간 CRP 수치가 다른지 보고자 합니다. CRP 수치에 대해 정규성 가정을 체크 해 보니 정규성 가정을 만족하지 않다고 할 수 있어서 로그 변환을 하였습니다. 그런데 CRP 수치의 대부분이 0 이상 1 미만이라, 로그 변환을 취하니 계산이 안 되는 것도 있고 음수로 산출 되는 것도 있어서 분석하기가 어렵습니다. 이럴 때는 어떻게 해야 되나요?

A. CRP 수치에 상수를 더하여 양수(+)로 만든 후 로그 변환을 하면 됩니다. 로그 변환 후 다시 정규성 가정을 체크를 하여 정규성 가정이 만족하는지 살펴보시고 만족하면 변환된 자료를 가지고 다중 회귀분석을 이용하여 다른 독립 변수들이 일정할 때(보정하여) 두 군간 CRP 수치가 다른지 살펴 볼 수 있습니다

4. 회귀 분석

Q. 단변량 분석 후 다중 회귀분석 시 어떤 순서로 진행해야 하나요? 그리고, 다중회귀분석(multiple regression)을 수행할 때 어떤 변수를 모형에 넣고 다중회귀분석을 수행해야 하는지요?

A. 분석 시 고려해야 되는 사항과 진행 순서는 다음과 같습니다. 연구를 진행하기에 앞서 연구목적에 고려해 볼 때 어떤 인자가 종속변수(결과변수, 반응변수)와 연관이 있는지 살펴 보기위해서 꼭 보고 싶은 또는 영향을 끼치리라 생각되는 독립변수(설명변수, 위험인자)들을 먼저 선정합니다. 그리고 이 독립변수들과 종속변수 간 true relationship 을 보는데 영향을 줄 수 있는 (즉, 독립변수와 연관성이 있고, 결과변수에도 영향을 끼치는) 제3의요인인 교란변수(confounding factor)가 있는지를 생각하여 교란변수(들)을 결정합니다.

그렇게 되면 독립변수에는 원래 선정한 독립변수들과 교란변수 들이 들어가는 것입니다. 이렇게 한 후 각 독립변수와 종속변수간 연관성을 보는 단변량 분석을 시행하고, 다중회귀분석(multiple regression)에서는 단변량 분석에서 사용된 모든 독립변수들을 모두 넣고 분석합니다.

그래야 다중회귀분석 결과 각 변수의 결과가 다른 독립변수들의 영향을 모두 보정했다고 할 수 있습니다. 그러나 만약 꼭 넣어야할 독립변수들이 너무 많다면 먼저 variable selection과정을(후진 선택 방법 및 stepwise방법) 통해 변수들을 selection할 수 있습니다. 그러나 이 방법은 임상적 관점과는 전혀 관계없이 순전히 computer based method에 의해서 뽑혀지는 방법이기 때문에 임상적으로 중요한 변수들이 선택이 안 될 수도 있고 또 변수 선택 과정이 다른 변수들의 영향을 다 보정하여 선택되었다고 볼 수도 없기 때문에 꼭 이 방법을 사용해야 할 때만 사용하는 것을 권합니다.

변수 선택을 하는 방법 중에 P-value를 이용하는 방법도 있는데 $P\text{-value} < 0.05$ 인 변수만 뽑게 되면 너무 보수적으로 variable selection 하고 임상적으로 중요한 변수들이 선택이 안 될 수도 있으므로, 이런 경우에 P-value가 약 0.2 보다 작은 변수들을 선택하는 방법을 생각해 볼 수 있습니다. (Ref. Fisher B, Anderson S, Fisher ER, Redmond C, Wickerham DL, Wolmark N, Mamounas EP, Deutsch M, Margolese R. Significance of ipsilateral breast tumour recurrence after lumpectomy. Lancet. 1991;338(8763):327-31). P-value가 약 0.2 보다 작은 변수들을 선택하고 걸러진 변수들을 모두 포함하여 다중회귀분석을 시행하세요. 하지만, 이경우도 임상적으로 중요한 변수들이 선택이 안 될 수도 있고 또 변수선택과정이 다른 변수들의 영향을 다

보정하여 선택되었다고 볼 수도 없습니다. 꼭 이 방법을 사용해야 할 때만 사용하는 것을 권합니다.

Q. 회귀분석 식에서 종속변수가 정규성분포의 가정을 만족하지 않는 경우는 어떻게 해야 하는지요?

A. log와 square root 등의 변수변환방법을 이용하여 종속변수를 변환하여 정규화하여 회귀분석 합니다. 하지만, 변수변환방법을 통하여도 종속변수가 정규화 되지 못하는 경우에는 종속변수의 순위를 이용한 비모수방법인 Spearman (편)상관관계분석을 수행합니다. 일반적으로 비모수적 방법은 모수적 방법에 비해 통계적 타당도 (validity)와 검정력(power)이 더 떨어지는 경향은 있습니다.

Q. 공변량을 보정하여 회귀분석을 적용하고자 하였는데 등분산 가정이 만족하지 않아 비모수 검정을 적용해야 합니다. 그런데 공변량들을 보정하는 비모수 검정방법은 어떤 것이 있는지요?

A. 등분산성이 만족하지 않을 경우에는 우선 Variance-stabilizing transformation (log, square root transformation, etc) 을 통해 회귀분석을 적용할 수 있습니다. transformation 을 통해서도 분산이 안정화되지 않는다면, 자료의 순위에 대한 분석 방법인 Partial spearman correlation analysis을 수행할 수 있습니다. 이는 보정해야 할 confounder 들을 부분 변수로 설정하여 보정한 채, 관심 있는 두 변수간의 상관관계를 살펴보는 방법입니다.

Q. 단순 회귀분석과 다중 회귀분석을 시행하였습니다. 다중 회귀분석에서는 의미가 없었고 단순 회귀분석에서는 의미 있게 나온 변수가 있다면 논문작성 시 어떻게 해석을 해야 되나요?

A. 논문 작성 시 단순회귀 분석 결과와 다중 회귀 분석결과 모두 제시하시되 논문의 주된 결과는 다중 회귀 분석 결과를 이용하여 해석하셔야 합니다. 왜냐하면, 다중회귀 분석은 다른 독립 변수들의 영향을 모두 보정했을 때 결과를 보여주기 때문입니다.

그러므로 단순회귀 분석에서 의미 있게 나왔지만 다중회귀 분석에서 의미 없게 나온 변수 는 다른 독립 변수들을 보정하지 않았을 때는 의미 있게 나왔지만, 최종적으로 보정했을 때는(다른 변수들이 일정할 때) 의미가 있다고 할 수 없으므로 실제로 의

미 있는 변수로 해석될 수 없습니다.

단순회귀 분석에서는 의미가 있다고 할 수 있지만, 다중 회귀 분석에서는 의미가 있다고 할 수 없는 이유는, 실제로는 유의하지 않는 변수임에도 불구하고 단순회귀 분석 시, false positive하게 confounded되어 유의한 결과가 나오게 된 것이기 때문입니다. 그러므로 단순회귀분석의 결과가 다중 회귀 분석의 결과와 반대로 나오게 된 것입니다. 따라서 논문 작성 시에는 단변량 분석 결과는 preliminary result로 보여 주고 다중회귀분석 결과를 연구의 주 결과로 기술합니다.

Q. 다중회귀분석 수행 시 SPSS에서 enter 방식이나 backward wald 방식으로 한 결과의 경우 어느 것이든 P-value 가 의미 있다면 multivariate analysis 결과로 의미가 있는 건가요?

A. enter 방식과 backward 방식은 반응변수에 대한 독립변수들의 regression model을 구축할 때, 어떤 변수들을 포함하여 분석할 것인가를 기준으로 생각하는 변수선택 방법입니다. enter 방식은 confounder(반응변수에 대한 관심변수의 영향에 혼란을 줄 수 있는 변수)라고 판단되는 모든 변수들을 포함(Full model)하여 분석하는 것이고, backward 방식은 Full model에서 통계적으로 의미가 없다고 판단되는 변수들을 하나씩 제거하는 변수선택방법입니다. enter방식은 모든 변수들을 고려할 수 있으나 분석이 불가능한 경우도 있고, 모형의 설명이 까다롭다(많은 변수들과 교호작용 고려시)는 단점이 있으며, Backward 방식은 순전히 통계적인 유의성을 기준으로 구축된 모형이므로 임상적으로 의미 있는 변수들이 선택되지 않을 수도 있어서 임상적인 실제(예상) 모형과 일치하지 않을 수 있습니다.

예를 들면, 만약 임상적으로 confounder 라고 판단되는 모든 변수들은 보정하여 관심 변수들 간 관계를 분석하기 위해 regression 을 적용하는 것이라면, 예상되는 confounder 들을 모두 보정한 enter 방식의 모형을 해석하는 것이 타당합니다.

만약 의미 있는 위험 인자를 탐색하여, 통계적으로 의미 있는 변수들을 선별해서 예측모형을 구축하고자 한다면, backward 방식이나 stepwise 방법으로 변수들을 선택한 모형도 고려해볼 수 있습니다.

Q. R-square = 0.036 이라고 되어 있어서, 논문에 수록 시 예측력이 3.6%다 라고 말할 수 있는 건가요? 설명력이 좀 낮은데 보통 어느 정도면 높다고 하는지요?

A. 예측력과 설명력은 분명히 다른 개념입니다. 예측력은 현재 추정된 모형이 다른 sample에 적용했을 경우에 얼마나 잘 맞는지를 나타내며, 설명력은 현재의 자료에

대한 적합도를 나타냅니다. 따라서 예측력이 아니라 설명력(explanatory power, 또는 R-square) 이라고 표기하는 것이 좋겠습니다.

설명력은 높을수록 좋은 척도이지만, 실제분석 시 의미 있는 모형의 경우에도 (즉, 유의한 변수들이 있는 경우) 설명력이 낮게 나오는 경우가 많아서, 어떤 경계값을 두고 설명력이 그보다 낮다고 해서 좋지 않s다고 말하기보다는 “현재 구축된 모형의 설명력이 이정도 수준이다” 라는 관점으로 보는 것이 적당할 것 같습니다. 하지만 어쨌든 3.6% 설명력은 매우 낮은 수준입니다.

Q. PAI-1이라고 하는 inflammation maker가 대장선종 발생의 risk factor인지 그냥 수치상으로 T test를 돌렸을 때는 의미 있게 나왔으나, Logistic regression을 했을 때는 의미가 없어서 사분위수로 PAI-1을 나눠서 선종의 크기별로 또는 조직검사의 악성도 정도에 따라 분류를 한 후 회귀분석을 해도 되는지 궁금합니다.

A. 이렇게 사분위로 나누어 분석을 많이 하기도 합니다. 이 뜻은 상대적 크기에 대해 분석하고 해석하겠다는 의미입니다. 즉 하위 25%인 사람에 비해 25%~50%인 사람은 위험도가 높고 또는 낮고, 하위 25% 사람에 비해 50%~75%인 사람은 위험도가 높고 또는 낮고, 하위 25%인 사람에 비해 상위 25%인 사람은 위험도가 높고 또는 낮다 인지를 분석하고 해석하겠다는 것입니다. 만약 이것이 실제적, 임상적으로 의미가 있다면 그렇게 하셔도 됩니다.

그러나 만약 사분위로 나누어 PAI-1이 얼마이하인 사람에 비해 ... ~ ... 에 있는 사람은 위험도가 높다 또는 낮다. 등등과 같이 분석 및 해석하는 것은 좀 위험합니다. 왜냐하면 사용되는 사분위라는 것이 자료에 따라 변하기 때문입니다.

Q. 다중회귀분석을 수행할 때, 결측치를 가지는 변수들을 분석에 포함할 경우 n 수가 지나치게 작아졌습니다. 왜 이러죠?

A. 다중회귀분석에 포함되는 변수들 중 하나라도 missing value를 가지는 subject는 분석대상에서 제외되기 때문입니다. 이런 경우 missing value를 많이 포함하고 있는 변수를 분석대상에서 제외하는 것을 추천합니다. 또 자료구조와 연구목적에 따라서 Imputation(대체) method를 적용하여, missing value를 대체하는 방법도 생각해볼 수 있습니다. 그러나 이 경우 자료값의 대체로 인한 bias의 증가 정도를 가능하기가 까다롭고, imputation method를 결정하는 것도 신중해야하기 때문에 자주 사용되지 않습니다.

Q. 포아송 회귀분석 결과를 작성하는 과정에서 P-value와 $\exp(\text{Estimate})$ 만 작성하여도 되는가요?

A. 일반적으로 회귀분석 (로지스틱, 포아송 포함) 후 결과는 Estimate, Std.Error, z.value, P-value 및 신뢰구간 등을 Table로 구성합니다. 이 중 Estimate와 P-value는 반드시 작성하여야만 자신의 결과에 대한 정보를 제공할 수 있습니다. 그런데 Poisson (or logistic) regression은 Estimate 대신에 $\exp(\text{Estimate})$ 인 지수값으로 결과를 작성하기도 합니다. 이는 결과를 해석할 때 좀 더 쉽게 해석하기 위함입니다. 예를 들어 Poisson regression에 Estimate가 0.13682라고 한다면 이는 추정치의 변수가 한 단위 증가할 때 0.13682씩 증가한다고 해석하게 됩니다. 그런데 $\exp(0.13682)=1.14662$ 는 한 단위 증가할 때 평균적으로 1.14622배 증가한다고 해석할 수 있습니다.

5. 로지스틱 회귀분석, 오즈비

Q. 연령을 보정하여 두 군간 발생 비율이 다른지를 보고 싶습니다. 어떻게 하면 되는지요?

A. 결과 변수(=발생, 발생 안함)가 0, 1로 코딩 되어 있으므로 로지스틱 회귀 분석을 이용하여 분석 할 수 있습니다. 이 때 독립변수는 연령과 두 군을 나타내는 그룹 변수로 지정하고, 결과 변수는 발생 유무로 지정합니다.

Q. Urin acid(세 그룹으로 나눔)와 HTN 발생 비율과 연관성을 보고자 하는 것이 논문의 주제입니다. 이때, 로지스틱 회귀 분석을 이용하여 HTN 발생에 영향을 끼치는 risk factor가 무엇인지 알아보려고 합니다. 성별="여"를 기준 범주(reference category)로 하여 분석해 보니 P-value가 0.05 보다 작아 유의하였습니다. 그런데 회귀계수가 음수인데 어떻게 해석해야 되나요?

A. 다른 변수들이 일정할 때 (보정하여) 여자일 때보다 남자일 경우 HTN가 발생할 가능성이 적다고 할 수 있다고 해석하시면 됩니다.

Q. tumor size, stage를 보정하여 남자일 때 보다 여자일 경우 location(왼쪽 vs 오른쪽) 간 MSI 유무가 차이가 있는지 보고 싶습니다. 한 사람당 왼쪽 오른쪽 모두 MSI여부를 측정한 것은 아닙니다. 왼쪽을 측정한 그룹과 오른쪽을 측정한 그룹은 서로 다른 피험자로 구성되어 있습니다. 어떤 분석 방법을 이용하면 되나요?

A. Multiple logistic regression analysis를 하시면 됩니다. 이때, 남자일 때 보다 여자일 경우 location(왼쪽 vs 오른쪽) 간 MSI 유무가 차이가 있는지 보고 싶다고 하셨기 때문에 성별과 location간 교호작용을 통해서 이를 살펴 볼 수 있습니다. 즉, 보정변수, 성별, location, 성별*location을 독립변수로 넣고 분석합니다.

Q. 질병 유무에 영향을 끼치는 Risk factors가 무엇인지 알아보려고 합니다. Multivariable analysis를 수행 시 SPSS에서 이분형 로지스틱과 다항 로지스틱 중 어떤 것을 이용해서 분석해야 되나요?

A. 이분형 로지스틱을 이용해서 분석하시면 됩니다. 만일 종속 변수가 질병 유무처럼 이분형 범주라면 이분형 로지스틱을 이용해서 분석하면 되고 종속변수가 3개 이

상의 명목형 범주형인 경우엔 다항 로지스틱 회귀 분석(multinomial logistic regression)을 이용하면 됩니다.

반면, 종속변수가 3개 이상의 순서형 범주형인 경우 순서를 고려한 분석을 수행한다면 순서 회귀 분석(Ordinal regression)을 이용하여 분석할 수 있습니다. (예를 들어 유방암진단에 대한 세 가지 태도를(전혀 아니다, 다소 그렇다, 매우 그렇다) 조사해서 이 세 가지 태도를 반응변수로 하여 ‘검사시간이 경과 할수록 진단을 잘한다고 할 때 보다 잘하지 못한다고 할 경향이 높다’ 라는 것을 검정 시(추세를 검정) 순서 회귀 분석을 이용할 수 있습니다.)

만일, 순서를 고려하지 않은 분석을 수행한다면 다항 로지스틱 분석을 수행할 수 있습니다.(예를 들어 유방암진단을 잘하는지에 대해 세 가지 태도를(전혀 아니다, 다소 그렇다, 매우 그렇다) 반응변수로 하여 검사 시기에 따라 세 가지 태도 비율이 다른지 검정을 할 경우에 사용할 수 있습니다.)

Q. 다른 변수들이 일정할 때(보정하여) 임신주수(1,2,3,4,5; 순서형 범주)가 긴 그룹으로 갈수록 질병이 일어날 가능성이 높은지 보기 위해서는 어떻게 해야 되나요? (SPSS를 이용하여 분석하고 있습니다.)

A. 로지스틱 회귀 분석 시 임신 주수에 대한 Linear Trend test를 하시면 됩니다. SPSS에서 일반화 선형 모형(일반선형분석이 아님)을 이용하여 분석할 수 있습니다. 회귀 분석처럼 종속 변수에는 질병유무와 같은 결과변수를 넣어주고, 모수요인에는 임신 주수(순서형 범주형 변수), 보정할 범주형 변수들을, 공변량에는 보정할 연속형 변수를 넣어줍니다. 그리고 나서 대비를 누르고 임신주수를 다항으로 넣고 분석 하시면 됩니다.

Q. 종속변수가 연속변수이지만, 이 종속변수를 이분변수화한 후 로지스틱회귀분석을 수행해도 되는지요?

A. 임상적으로 의미 있거나 알려져 있는 종속변수의 경계 값을 이용하여 이분변수화한 변수에 대하여 로지스틱회귀분석을 수행할 수 있습니다. 다만, 종속변수를 어떻게 이분화 했는지에 대한 임상적 타당성 또는 통계적 타당성이 있어야합니다.

Q. Odds ratio(오즈비; OR), 오즈비에 대한 95% 신뢰구간에 대해 알고 싶습니다. 그리고 P-value와 관련해서 오즈비를 어떻게 해석해야 되는지도 궁금합니다.

A1. Odds ratio 해석 - Odds ratio는 위험군에서 case:control의 비인 odds와 비위험군에서 case:control의 비인 odds간 ratio입니다 (즉, $OR = [(case비율/control비율) \text{ in 위험군}] / [(case비율/control비율) \text{ in 비위험군}]$). $OR=1$ 을 기준으로 연관성 정도를 나타내는데, 위험군에서 case:control의 비가 비위험군에서 case:control의 비보다 크다면 (즉 $OR>1$) 위험군에서 case 가능성이 더 높다는 것이고 (반대의 경우도 마찬가지로 성립), 위험군에서 case:control의 비와 비위험군에서 case:control의 비가 같다면 (즉 $OR=1$) 관심 위험인자는 case와 연관이 없음을 나타냅니다.

A2. 95% 신뢰구간 - 실제 오즈비를 포함할 확률을 95%로 정해 놓고 계산된 구간으로, 오즈비 95% 신뢰구간은 이 구간 안에 실제 오즈비가 포함될 가능성이 95%라는 뜻이 됩니다. 간혹 로지스틱 회귀 분석 결과 특정 연속형 독립변수의 오즈비의 95% 신뢰구간이 좁거나 넓은 경우가 존재하여 결과 해석이 어려운 경우가 발생하는데, 이는 이 연속형 독립변수의 수치가 모형 내의 다른 독립변수들에 비해서 상대적으로 클 때 혹은 작을 때 이런 결과가 나타날 수 있습니다. 이 경우, 독립변수의 scale을 조정한 후(ex. g \rightarrow kg, 즉 변수값/1000 또는 반대로) 로지스틱 회귀분석 하시기 바랍니다. 다른 방법으로는, 독립변수의 표준편차(SD)로 나누고 난 후, 로지스틱 회귀분석 하는 것인데, 이는 그 독립변수의 분포가 어느 정도 대칭(symmetric)일 때 의미가 있습니다.

(참고: 범주형 독립변수인 경우에도 어느 범주의 빈도가 매우 크거나 혹은 매우 작은 경우에 이런 현상이 나타날 수 있습니다.)

A3. OR과 P-value - OR과 P-value는 서로 직접적인 연관성이 있다고는 할 수 없습니다. OR은 연관정도를 나타내는 척도일 뿐이고 이 수치의 대소를 가지고 유의성 여부를 판단하기는 어렵습니다. 반면에 P-value는 가설 $OR=1$ 에 대한 기각 여부 (즉 통계적 유의성 여부)를 알려주는 수치이므로 유의성 여부는 P-value를 가지고 판단합니다. 종종 OR값이 크더라도 P-value가 유의수준 0.05보다 커서 가설 $OR=1$ 을 기각할 수 없는 경우가 있습니다. 이 때 OR의 95% 신뢰구간을 들여다보면 구간의 폭이 매우 넓고 $OR=1$ 을 포함하고 있음을 볼 수 있습니다. 즉 자료로부터 산출한 OR값은 크지만 실제 OR값은 1.00이 될 가능성이 95%나 되기 때문에 가설 $OR=1$ 을 기각할 수가 없어서 연관성이 있다고 할 수 없게 됩니다. 이와 유사한 경우가 상관분석 결과에서 상관계수와 P-value가 있습니다.

Q. 로지스틱 회귀 모형을 적용하는데, 증상의 발생여부에 대한 연령의 오즈비가 0.712, 95% CI는 (0.560, 0.904)로 나왔습니다. 또한 회귀계수는 -0.340이고, 회귀계수의 표준오차(=se)는 0.122 입니다. 어떻게 해석해야 하는지요?

A. 연령이 1세 높아질 때 증상이 일어날 가능성(오즈)이 0.712 배 증가하는 것입니다.

- 오즈비를 1보다 큰 값으로 나타내고 싶으시다면 오즈비와 신뢰구간에 역수를 취하시면 됩니다.

: 오즈비=1/0.712=1.404, 95% CI=(1/0.904, 1/0.561)=(1.106, 1.786)

- 그리고 연령의 증가단위를 1세가 아니라 10세로 하고 싶으시다면 회귀계수와 표준오차를 10을 곱한 다음에 오즈비와 신뢰구간을 구합니다.

: 오즈비= $\exp(\text{회귀계수}) = \exp(-0.340 \times 10) = 0.033$

95%신뢰구간의

하한= $\exp(\text{회귀계수} - 1.96 \times \text{표준오차}) = \exp(-0.340 \times 10 - 1.96 \times (0.122 \times 10)) = 0.003$

상한= $\exp(\text{회귀계수} + 1.96 \times \text{표준오차}) = \exp(-0.340 \times 10 + 1.96 \times (0.122 \times 10)) = 0.365$

혹은, 연령/10한 자료를 로지스틱 회귀 분석하여 추정된 오즈비와 95% CI를 이용할 수도 있습니다.

Q. 주 위험요인 변수(main risk factor)가 case 발현에 미치는 영향을 보고자 사례-대조(case-control) 로지스틱 회귀분석을 수행하려 합니다. 여기에 특정 변수를 혼란변수로 작용할 경우, 이 혼란변수에 따른 주 위험요인 변수의 case 발현에 미치는 영향을 어떻게 확인할 수 있는지요?

A. 로지스틱 회귀분석 모형에 주 위험요인변수와 혼란변수의 교호작용(interaction) 변수를 포함하여 분석합니다. 이 교호작용이 통계적으로 유의한 경우, 이 혼란변수의 값에 따라 주 위험요인변수의 case 발현에 미치는 영향이 다르다고 해석합니다. 예를 들어, 총 흡연량이 폐암 발병에 미치는 영향 연구 시, 나이가 혼란변수로 작용할 수 있습니다. 이 때, 나이에 따라 흡연량이 폐암에 미치는 영향이 다른지를 확인하고자 한다면 나이와 흡연량의 교호작용을 모형에 넣어 유의성을 확인합니다. 이 효과가 통계적으로 유의하다면, 나이에 따라 흡연량이 폐암 발병에 미치는 영향이 다르다고 해석할 수 있습니다.

Q. 연구결과를 작성하는데 공간이 부족하여 다중 로지스틱 회귀분석(Multiple logistic regression analysis) 결과 중 유의한 것만 table화 하였는데 괜찮은지요?

A. Multivariable analysis를 하는 경우 사용한 변수에 대한 결과를 모두 작성해야 합니다. 그런데 만약 아주 예외적으로 제시 공간이 부족해서 유의한 결과만 작성하

여야 한다면, table 아래에 Multivariable 분석에 사용된 변수는 이런 것이 있고, 그 결과 유의했던 것이 위와 같다는 식으로 footnote를 작성하는 것이 좋을 것 같습니다. 만약 그렇지 않으면 다른 사람이 table에 작성된 변수만으로 Multivariable analysis를 한 것으로 오해 할 수 있습니다. 그러나 이 방법은 가능한 사용하지 않는 것이 좋습니다.

Q. Multiple logistic regression 에서 상관성이 높은 독립변수들은 빼주는 것이 좋다고 하는데 0.75 이상이면 빼주는 것인가요?

A. 상관계수가 0.75 이상이면 높은 상관계수가 있다고 보지만, Multiple logistic regression 에서는 특별히 기준이 되는 숫자가 있지는 않습니다. 더 높은 상관계수를 가지는 경우에도 분석에 포함시킬 경우도 있습니다.

상관계수보다 Multi-collinearity (다중공선성)를 살펴보는 것이 더욱 reasonable한 방법이라고 볼 수 있습니다. 그러나 SAS나 SPSS에서는 Logistic regression 에서 이러한 Multi-collinearity를 검정하는 옵션이 없으므로, Regression(회귀모형)을 사용하여 살펴볼 수 있습니다.

방법

1. Logistic regression 은 종속변수(결과변수)가 범주형 변수이고 Linear regression에서 종속변수는 연속형 변수임
2. Multi-collinearity 는 독립변수간의 상호 연관성을 살펴보는 것이므로, SAS나 SPSS에서 linear regression 을 구동함 - 종속변수를 이분형 범주형 변수로 설정.
3. Multi-collinearity에 대한 검정 결과에서 각 독립변수마다 VIF(Variance inflation factor) 또는 Tolerance 값이 산출됨. - 종속변수가 이분형 범주형 변수일 경우 linear regression의 결과는 종속변수에 대한 분석결과로는 적절하지 않으나 Multi-collinearity 의 검정결과는 살펴볼 수 있음.

보통 수행결과 VIF가 10보다 큰 경우 변수를 제거하지만, 특별히 정해진 기준이 있는 것은 아닙니다. Tolerance가 0.4보다 작은 변수를 제거하기도 합니다.

Ref: Allison, Paul D. 1999. logistic regression Using The SAS system: Theory and Application. Cary, NC: SAS Institute Inc. p50

6. 생존분석

6.1 Median survival time

Q. 생존 분석 시 median survival time을 구하고자 합니다. Excel을 이용하여 survival time에 대한 median값을 구할 수 있나요?

A. 생존 자료의 경우에는 최종 추적 시간 때까지 사건(ex. 사망)이 일어나지 않아서 사건이 일어날 때까지의 정확한 시점을 알 수 없거나, 연구 기간 중도에 follow up loss가 일어나서 역시 사건이 일어날 때까지의 정확한 시점을 알 수 없는 경우가 있습니다. 이런 자료를 중도 절단된 자료(censored data)라고 합니다. 중도 절단된 자료(censored data)가 하나도 없는 경우엔 사건(ex, 사망)이 일어날 때까지의 정확한 시간을 모두 알고 있는 경우이므로 이 time 변수를 가지고 엑셀에서 [수식 -> 함수 마법사 -> 통계 -> median]을 이용하여 time에 대한 중앙값을 계산하면 이것을 median survival time라고 할 수 있습니다. 하지만, 중도 절단된 자료(censored data)가 있는 경우엔 사건(ex, 사망)이 일어날 때까지의 정확한 시간을 알지 못하는 자료도 섞여 있으므로, 사건이 일어날 때까지의 시간을 알고 있는 경우와 모르는 경우를 합하여 단순히 time의 중앙값을 계산해서는 안 됩니다. 왜냐하면 중도 절단된 측정값은 측정된 시간까지 사건이 일어나지 않았다는 정보만 줄 뿐 실제 사건이 발생한 시간은 그 시점 이후이나 해당 연구에서는 알 수 없기 때문입니다. 이와 같은 경우엔, 중도 절단된 자료의 정보를 이용하여 생존율을 계산한 후, 생존율 50%에 해당되는 시점을 찾으면 median survival time을 구할 수 있습니다.

Q. 환자의 50%가 생존하는 시점 알기 위해서는 median 값을 봐야 되나요? 아니면 mean 값을 봐야 되나요? median값을 보니 "."으로 출력이 안 되었습니다.

A. median survival time은 환자의 50%가 생존하는 시점을 나타내므로 median을 보시면 됩니다. 그런데, median값이 출력이 안 된 이유는 선생님 자료와 같이 event 가 많지 않을 경우 생존 곡선(survival curve)이 생존율 50% 위에서 그려져 생존율 50%에 해당하는 시점을 구할 수 없기 때문입니다. 이런 경우 f/u 기간 동안 event가 몇 명밖에 안 되어 생존 커브가 생존율 50% 위에서 그려져 median값을 구할 수 없다고 설명할 수 있습니다.

6.2 생존율 추정 및 구간 생존 곡선 비교 검토

Q. 생존 분석 시 카플란 마이어(Kaplan-Meier) 방법으로 생존율을 추정 하지 않고 생명표(Life table) 방법을 사용하였습니다. 생명표(Life table) 방법을 사용한 것이 맞는 방법인지요?

A. 만일 event수 및 환자의 수가 상당히 많고 추적기간이 긴 경우, 생존 기간을 일정한 간격으로 나누어 분석하는 생명표(Life table) 방법을 이용하여 분석할 수 있겠지만, 그렇지 못한 경우 사용하시면 안 됩니다. 왜냐하면 생명표(Life table) 방법은 생존 기간을 몇 개의 구간으로 나눈 후 분석하는데 추적기간이 짧은 경우 일정한 간격으로 나누기 어렵고, event가 일어난 시점의 생존율을 계산할 수 없기 때문입니다. 그러므로, event수가 작은 경우라면 event가 일어난 시점마다 생존율을 계산할 수 있는 카플란 마이어(Kaplan-Meier) 방법으로 분석하시는 것이 좋습니다.

(참고) 생명표 방법을 이용할 때 다음과 같이 주의할 점

1) 구간을 나눌 때에도 어느 정도의 구간으로 나누는 것이 연구자가 임상적인 의미로 볼 때

유용한지에 대한 기준도 필요합니다.

2) 생명표 방법을 사용할 때에는 "survival experience가 study기간 동안 변치 않아야 한다" 와 같은 기본 가정을 만족해야 됩니다. 예를 들어, 처음에는 치료 방법이 없었는데 연구기간 중간에 치료 방법이 개발되어 처음 시작 시점에 비해 생존율이 달라지는 경우라면 생명표 방법을 사용할 수 없습니다.

Q. 생명표법을 사용하여 1994년부터 2007년까지의 환자군을 대상으로 1994년부터 2001년까지의 군과 2002년부터 2007년까지의 군을 나누어서 stage별 5년 생존률을 비교했습니다. 그런데 1994년부터 2001년까지의 군에서 60개월째 생존률이 똑 떨어지는 것을 확인할 수 있었습니다. 이상해서 24개월을 했더니 마찬가지로였습니다. 무엇이 잘못 되었는지요?

A. 24개월째 혹은 60개월째 생존율이 똑 떨어지는 이유는 생명표 방법으로 분석 시 time에 대한 끝점을 time 자료의 최대값으로 지정해주지 않았기 때문입니다. 그러므로 Life table (생명표 방법)으로 분석하실 때 주의하실 점은 분석하시고자 하는 time에 대한 끝점과 구간 폭을 지정해 주는 일이라고 할 수 있습니다. 끝점은 time의 최대값을 적어 주시면 됩니다.

예를 들어, 94년 ~ 01년까지의 환자군과 02년~07년까지의 환자군의 time에 대한 최

대값은 각각 96개월, 60개월입니다. 그러므로 두 군의 survival curve를 비교 분석하실 때 time에 대한 끝점을 96으로 기입하셔야 됩니다.

만일, 96보다 작은 24 개월로 time의 끝점을 기입하고 구간 폭을 1로 하셨다면 (0-1], (1-2], (2-3],.....,(24개월 이상)까지 각 구간에 대한 생존율 결과를 제시합니다. 그리고 24개월 이상에서 event가 일어난 경우는 모두 마지막 구간의 event로 count가 되기 때문에 24개월 이상에서 event가 많이 일어났다면 Figure 3과 같이 생존율이 뚝 떨어지는 그래프를 보여 주게 됩니다.

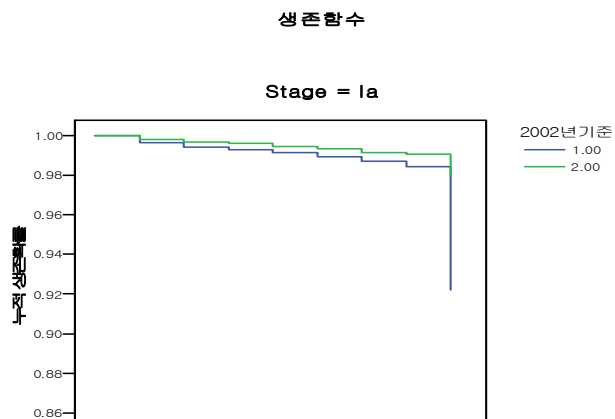


Figure 3. 생명표 방법으로 분석 시 time에 대한 끝점을 time의 최대값으로 지정해 주지 않아 생기는 그래프

Q. 5년 생존율을 구하고자 합니다. 실제로 follow up은 60개월 이상을 했지만, 5년 생존율을 구하려고 합니다. 그럼 만약에 70개월 생존한 사람은 60개월을 생존시간으로 보고 사망하지 않은 것으로 해야 되는 건가요?

A. 전체 생존 곡선(전체 follow up time에 걸친)에서 5년 생존율을 구하게 되면, 70개월을 60으로 수정하지 않아도 5년 생존율을 구할 수 있습니다. 70개월을 생존한 사람의 경우 5년째(60개월째) 순간에는 생존으로 간주됩니다.

Q. 두 그룹 간에 recurrent rate에 대해 비교를 하고자 합니다. 어떻게 해야 되나요?

A. 우선 어떤 시점에서 재발율을 산출하고 비교할 것인지를 결정합니다. 만약 재발이 일어나지 않은 환자 중 이 시점까지 F/U이 되지 않은 환자가 있다면, 생존분석을 수행해야 합니다. 이 경우 모든 환자들에 대한 추적기간을 알아야 합니다. 그러나 이 시점내 재발 여부를 모두 알고 있다면, chi-square test 같은 completed data analysis로 분석할 수 있습니다.

Q. 두 군 간 생존분포 곡선분포 비교 시, 두 생존곡선이 여러 번 교차하는 경우, 어떤 분석을 해야 하는지요?

A. 두 생존곡선이 교차하는 경우, 전 생존기간에 걸쳐 log-rank test나 Breslow test (Wilcoxon test)를 이용하여 두 군 간 생존분포 곡선분포를 비교하는 것은 의미가 없습니다. 이러한 경우, 그래프의 양상을 확인하여 두 생존곡선의 교차점의 생존시간 구간 별로 나눈 후, 생존시간 구간 별 두 군 간 생존분포 곡선분포를 비교분석하고 해석하거나, 또는 Time-dependent Cox regression을 적용합니다.

Q. 두 군간 생존곡선이 다른지 검정하기 위해 Log rank test 및 Breslow test를 이용하였고, 이때, Breslow test가 약간 더 유의하게 나와서 이 방법을 이용하려고 합니다. 이렇게 해도 괜찮은지요? 만일 사용할 수 없다면 왜 그런지요?

A. 군 간(2군 이상) 생존 곡선이 다른지 검정하기 위해 사용하는 방법에는 Log-Rank test, Breslow test 등이 있습니다. Log-rank test는 각 시점에서 위험에 처해 있는 인원수가 동일한 비중을 갖는다는 가정 하에 분석하는 방법입니다. 각 군의 생존 곡선이 서로 교차 되지 않고 생존 시점에 따라 전체적으로 평행하게 그려져 있다면 Log-Rank test 방법으로 분석하는 것이 좋습니다. 반면에, Breslow test는 각 시점에서 위험에 처해 있는 인원수에 비례하도록 가중치를 주어 분석하는 방법입니다. 만일, 연구 초반에 처리의 효과가 두드러지게 나타나는 연구라면(생존 곡선을 그렸을 때 초반에 생존 곡선이 크게 벌어지다가 시간에 따라 점점 좁혀지는 양상) 각 시점마다 위험에 처해 있는 인원수가 동일한 비중을 갖는다고 할 수 없고 연구 초반에 더 많은 가중치를 주는 것이 효과적일 수 있으므로 이 경우 Log-Rank test 방법 대신 Breslow test로 분석해야 됩니다.

그러므로 유의성 여부에 따라 분석 방법을 선택하지 마세요. 적절한 분석 방법을 선택하기 위해서는 생존 곡선 그래프 양상을 살펴보는 작업이 선행되어야 합니다.

Q. 아래 그림(Figure 4)과 같이 favorable variant 는 119명 중 관찰기간 내 2명이 사망하였고, aggressive는 41명 중 2명이 사망하였습니다. 그러나 10-year survival rate 는 favorable variant 가 89% 로 aggressive 의 94% 에 비해 왜 적게 나오는지 궁금합니다.

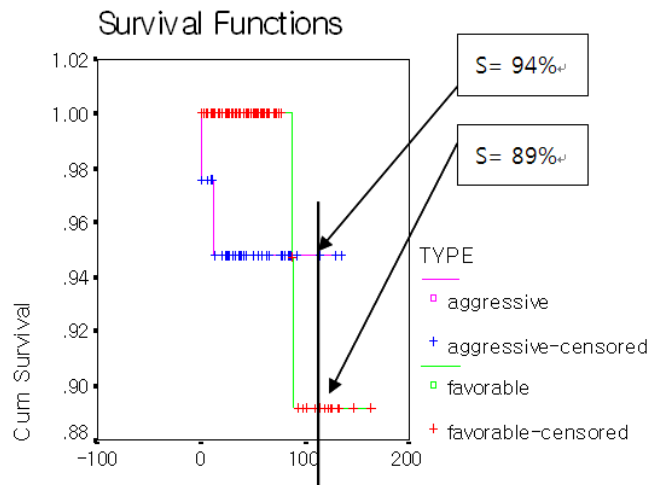


Figure 4. favorable variant type과 aggressive type의 생존 곡선

A. Survival rate는 전체 표본 크기와 event 수만으로 계산할 수 없습니다. 중도 절단이 일어난 시점, 중도절단 수, event가 일어난 시점, event 수 모두를 고려하여 계산합니다. 그러므로 favorable variant 그룹의 10yr survival rate(=89%)가 aggressive 그룹의 10yr survival rate(=94%)보다 작게 추정되는 이유는 중도 절단자료의 분포와 event(=사망) 자료 분포가 서로 다르기 때문인데, Figure 4 를 통해 살펴보면favorable 그룹의 경우 event가 일어나기 전 연구 초기에 중도 절단된 자료가 aggressive 그룹보다 많기 때문입니다.

6.3 Cox proportional hazard model (Cox regression)

Q. 두 처리군과 어떤 질환의 수술 후 재발 여부간 연관성이 있는지를 검정할 때, 어떤 경우에 카이제곱 검정이 아니라 생존분석인 cox proportional hazard analysis를 사용해야 하나요?

A. 예를 들어 각 환자에서 질환의 수술 후 재발여부를 알기 위해 추적, 관찰한 시간이 모두 1년으로 같거나 최소 1년이라고 한다면 모든 환자에서 1년 내 재발여부를 모두 알 수 있기 때문에 두 처리군간 수술 후 1년 내 재발여부가 다른지를 카이제곱 검정으로 볼 수 있습니다. 이처럼 모든 환자가 1년 이상 추적, 관찰되어서 수술 후 1년 내 질환의 재발여부에 대해 모두 조사되었다면, 생존분석을 적용하지 않고도 카이제곱 검정 또는 피셔의 정확검정으로 처리군과 수술 후 1년 내 재발 여부간 연관성을 검정할 수 있습니다. 그러나 어떤 환자에서 추적기간이 1년 미만이고 추적 관찰될 동안 질환이 재발하지 않았다면, 추적 관찰하는 동안은 재발하지 않았지만 그 이후에 재발하였는지를 알 수 없게 됩니다. (이를 중도 절단되었다(Censored)고 합니다.) 이렇게 특정 시점 내에 재발 여부를 알아보는데 이 특정 시점내 중도 절단된 자료가 있어서 이 특정 시점 내 재발여부를 확실히 알 수 없는 환자들 이 있다면 단순히 카이제곱 검정을 사용하는 것이 아니라 중도절단을 고려하는 생존분석의 방법을 사용하여야 합니다. (이 경우 Kaplan-Meier 방법으로 각 처리군에서 특정 시점에서의 생존율을 산출하고 이 생존율과 생존율의 standard error(SE)를 이용한 Z-값을 산출하여 비교 검정합니다.)

또한 환자마다 추적기간이 상이한 경우 특정 시점 내 재발여부 (즉 특정시점에서의 재발율) 비교가 아니라 두 처리군간 recurrence-free curve를 비교한다면 (즉 전 추적기간에 걸쳐 recurrence-free 분포가 차이가 있는지 또는 recurrence rate가 차이가 있는지를 비교) log-rank test, Breslow test 등으로 검정할 수 있고 여러 다른 요인들을 보정하여 두 처리군간 recur에 대한 hazard rate을 비교한다면 Cox proportional hazard analysis를 적용할 수 있습니다.

Q. ADL 측정수치를 독립변수로, 재발율에 대한 cox-regression 결과, hazard ratio가 0.92 정도로 나왔습니다. 그럼 해석을 어떻게 해야 하나요?

A. ADL 측정수치가 한 단위 증가할 때, 재발에 대한 위험율(hazard rate)이 감소한다고 해석하면 됩니다. hazard ratio를 해석할 때에는 반드시 신뢰구간도 함께 봐야 하며, 신뢰구간이 너무 넓지 않고, 1을 포함하고 있지 않다면, 이때 hazard ratio의 point estimate 를 가지고 위와 같이 해석할 수 있습니다.

Q. 총 91명의 data 중 17명만이 event가 일어났습니다. 이때, 11개의 risk factors를 가지고 Cox regression을 이용하여 분석하였는데 분석이 안 되었습니다. 왜 그런 결과가 나오는 지요?

A. 11개의 risk factors 중 범주형 변수의 경우 해당 그룹에 event가 있는 피험자가 한명도 없거나 두 개 이상의 범주형 변수의 범주간 조합에서 event가 한 명도 없는 조합이 있다면 그 분석 변수는 분석이 안 됩니다. 만일, 분석을 가능하게 하기 위해서는 risk factors를 살펴보고 event에 영향을 끼치는데 있어서 임상적으로 중요하지 않은 인자가 들어갔으면 제외하시거나 범주를 묶을 수 있는 factor가 있으면 묶어서 분석해 볼 수 있습니다. 참고로, 연구에 따라 event가 적게 나올 수밖에 없는 연구가 있을 수 있습니다. 이런 경우에는 비교 검정을 하기 보다는 descriptive 하게 설명 하시는 것이 더 좋습니다. 또는 다기관 연구 등을 통하여 event 수가 많이 나올 만큼 충분한 환자 수를 확보하는 것이 좋겠습니다.

Q. SPSS로 분석하고 있습니다. Cox 회귀 분석 시 다음과 같은 경고 messages가 떠서 분석이 되지 않았습니다. 분석을 실행하려면 어떻게 해야 되나요?

"8바이트보다 긴 문자 변수 값은 사용할 수 없습니다. 이 명령문은 실행되지 않습니다."

A. 변수보기 sheet에서 분석 변수의 너비를 8바이트 미만으로 해 주세요

6.4 Time dependent Cox proportional hazard model (Time dependent Cox regression)

Q. 생존 분석에서 Time dependent covariate란 무엇인지요?

A. 생존 분석에서 Time dependent covariate란 event 발생 위험에 대한 어떤 covariate의 영향이 시간에 따라 달라지는 것을 말합니다.

예를 들어 subject A의 smoking status를 1년마다 조사해보니, smoking status가 시간에 따라 다르다고 한다면 smoking 변수를 time dependent covariate로 처리하여 분석해야 됩니다. 왜냐하면 smoking status가 시간에 따라 다르다면 시점마다 survival에 대한 smoking status의 효과가 시간에 따라 달라지기 때문입니다.

참고)

Definition : A time-dependent variable is defined as any variable whose value for a given subject may differ over time (t)

Ref. David G. Kleimbaum, Mitchel Klein, Survival analysis: A Self-Learning 2nd, 1996, Springer

7. 반복측정 자료 분석

Q. 동일한 환자를 대상으로 농도 0mg, 20mg, 40mg별로 처리를 다르게 하여 연속형 수치로 측정되는 자료를 얻었습니다. 농도 간 연속형 수치가 다른지 검정하기 위해서는 어떤 분석 방법을 이용해야 되나요?

A. 동일한 환자를 대상으로 처리만 다르게 하여 (0mg, 20mg, 40mg 순서로 처리)자료를 얻었기 때문에 농도간, 연속형 수치가 다른지 검정하기 위해서는 Repeated Measures ANOVA을 이용하여 분석해야 됩니다. 이 경우에 연속형 수치값들은 정규분포를 따라야 합니다. 만일 정규성 가정을 만족하지 않으면 로그와 제곱근변환과 같은 변환을 수행해 볼 수 있습니다. 변환 후 정규성 가정을 만족하면 변환된 자료를 가지고 Repeated Measures ANOVA 분석할 수 있습니다.

그럼에도 불구하고 정규성 가정을 만족하지 않으면 차선택으로 0mg vs 20mg, 0mg vs 40mg, 20mg vs 40mg로 두 군씩 짝지어 Paired t-test 혹은 Wilcoxon's signed rank test로 검정할 수 있습니다. 그런데 서로 독립이 아닌 군을 두 군씩 짝지어 비교 검정할 경우 Type I error가 최대 3배까지 늘어날 수 있으므로 독립이 아닌 검정 회수인 3만큼 곱하여 각 P-value를 보정해 주어야 합니다.

Q. 각 장기별로 좌/우 마다 높이를 다르게 해서 (좌 우 순서로, 0cm, 3cm, 6cm 순서로 측정) X-ray을 측정한 값이 있습니다. 높이 간(0cm, 3cm, 6cm) X-ray 측정값이 다른지 살펴보고 싶습니다. 어떤 분석 방법을 이용해야 되나요?

A. 좌, 우 마다 높이를 다르게 해서 측정(좌 우 순서로, 0cm, 3cm, 6cm 순서로 측정)하였기 때문에 반복 요인이 2개인 자료 구조를 고려하여 Repeated measures analysis를 가지고 분석할 수 있습니다. 이 경우에 X-ray 측정값들은 정규분포를 따라야 합니다. 만일 정규성 가정을 만족하지 않으면 로그와 제곱근변환과 같은 변환을 수행해 볼 수 있습니다. 변환한 후 정규성 가정을 만족하면 변환된 자료를 가지고 Repeated Measures ANOVA 분석합니다. 그럼에도 불구하고 정규성 가정을 만족하지 않으면 차선택을 생각해 볼 수 있는데 방법은 다음과 같습니다.

방법1. 좌우 별로 0cm vs 3cm vs 6cm간 X-ray 측정값이 다른지를 반복 요인(높이) 1개인 자료 구조를 고려하여 Repeated Measures ANOVA로 분석할 수 있습니다. 이때에도 좌우 별로 정규성 가정을 체크한 후 정규성 가정을 만족한다고 할 수 있으면 Repeated Measures ANOVA로 분석합니다. 그런데, 서로 독립이 아닌 좌우 군 마다 분석할 경우 Type I error가 최대 2배까지 늘어날 수 있으므로 독립이 아닌 검정 회수

2만큼 곱하여 각 P-value를 보정해 주어야 합니다.

방법2. 방법1에서 정규성 가정을 만족하지 못하여 분석 할 수 없을 경우에는 좌우 별로 0cm vs 3cm, 0cm vs 6cm, 3cm vs 6cm간 X-ray 측정값이 다른지 각각 Paired t-test 혹은 Wilcoxon's signed rank test 으로 검정할 수 있습니다. 그런데, 좌우 별로 서로 독립이 아닌 군을 두 군씩 짝지어 비교 검정할 경우 Type I error가 최대 6배까지 늘어날 수 있으므로 독립이 아닌 검정 회수인 6만큼 곱하여 각 P-value를 보정해 주어야 합니다.

<참고> 만약 순서를 고려하지 않고 좌우를 random하게 측정하고, 0cm, 3cm, 6cm도 순서를 고려하지 않고 random하게 측정하였다면 Repeated measures ANOVA가 아니라 Two-way ANOVA로 분석합니다. 물론 이 경우에도 X-ray 측정값들은 정규분포를 따라야 합니다. 만일 정규성 가정을 만족 하지 못하면 좌 우 별로 0cm vs 3cm vs 6cm간 다른지 Friedman's test로 검정할 수 있습니다. 그리고 서로 독립이 아닌 좌우 군마다 분석했기 때문에 Type I error가 최대 2배까지 늘어날 수 있으므로 독립이 아닌 검정 회수 2만큼 곱하여 각 P-value를 보정해 주어야 합니다.

Q. 한 사람 당 특정 부위의 수치가 두 가지(ex. 오른쪽, 왼쪽 수치)인 경우 이 두 수치의 평균값을 이용하여 분석해도 될 지요?

A. 평균을 산출해서 사용하는 것이 임상적으로 의미가 있고 실제로 그렇게 하여 진단, 치료 등을 하고 있다면 그렇게 할 수도 있습니다. 하지만, 부위 수치내의 분산을 고려하지 않았으므로, 평균값을 이용하는 것은 통계학적으로 좋지 않은 방법입니다. 또한 두 환자의 평균값이 동일하다고 해도 각 환자별 실제 측정값들은 크게 다를 수 있기 때문입니다 (예: 환자 A의 측정값 1과 10, 환자 B의 측정값 5와 6). 두 가지 수치 중 하나만을 고려하거나, 두 곳 모두 고려해야 한다면 부위를 보정한 반복측정 자료 분석 방법을 이용하여 분석하는 것이 좋습니다.

Q. 치료 전후 간 차이가 두 군간 다른지 비교 시 Repeated Measures ANOVA로 분석하면 되나요?

A. 시점이 2개만 있을 경우(치료전, 치료후), 치료 전후 간 차이가 두 군간 다른지 비교 검정 시 Two sample T-test 혹은 Wilcoxon's Rank Sum test을 쓰시고 치료 전

후 간 다른지 비교 검정 시에는 Paired T-test or Wilcoxon's signed rank test를 시행하시면 됩니다.

이렇게 하는 이유는 다음과 같습니다.

Repeated Measures ANOVA로 분석하기 위해서는 종속 변수의 등분산성 가정 및 정규성 가정을 전제로 합니다. 만일 정규성 가정을 만족하지 못하거나 등분산성 가정을 만족하지 못하면 Repeated Measures ANOVA를 사용할 수 없습니다.

치료 전 후 간 차이가 두 군간 다른지 비교 검정 시 각 군별로 치료 전후 차이가 정규성 가정을 만족하지 못하면 비모수적 검정 방법인 Wilcoxon's rank sum test을 이용하여 분석할 수 있습니다. 마찬가지로 치료 전 후 간 다른지 비교 검정시에는 치료 전 후 간 차이가 정규성 가정을 만족하지 못하면 비모수적 검정 방법인 Wilcoxon's signed rank test를 이용하여 분석할 수 있습니다.

또한, Repeated Measures ANOVA은 등분산성 가정시에만 분석할 수 있지만, Two sample T-test는 등분산성 가정을 만족하지 못할 경우에도 분석할 수 있습니다. 정규성, 등분산성이 만족할 경우 각 군에서 두 시점 간 차이를 보는 경우, Paired t-test 결과와 Repeated measures ANOVA 결과가 동일하고, 두 시점 간 차이가 두 군간 다른가를 보는 경우, Two-sample t-test, Repeated Measures ANOVA 결과가 동일합니다.

따라서 동일한 결과를 주면 누구나 알고 있는 방법을 사용하여 기술하는 것이 더 좋습니다.

참고로, 시점이 일정할 때(보정하여) 그룹 간 (실험군 vs 대조군)다른지, 아니면 그룹이 일정할 때(보정하여) 치료 전후 간 우울점수가 다른지를 Repeated Measures ANOVA를 이용하여 분석할 수는 있습니다.

Q. SPSS로 분석하였습니다. GEE(Generalized estimation equation)로 분석했더니 출력창에 "개체 및 개체 내 변수에 값이 2개 이상 있어서 분석할 수 없습니다." 라고 경고가 뜨고 분석이 안 됩니다. 왜 그런지요?

A. 예를 들어 설명하겠습니다. 30명을 가지고 8번 반복하여 측정한 자료를 가지고 분석할 경우 피험자는 30명이고 각 피험자에는 8개의 자료로 이루어져 있다는 정보를 알려 주어야 합니다. 그런데, 30명 중에 "홍길동" 이라는 피험자가 2번 들어간 있다면 홍길동이라는 피험자는 8개의 반복 측정된 자료로 구성되어 있는 것이 아니라 8개의 자료가 2번씩 더 측정된 자료로 읽히게 되므로 위와 같은 경고 message가 뜨게 되는 것입니다.

8. 진단법 평가(민감도, 특이도, ROC), 경계값(Cut-off value)

Q. 두 진단 도구 간 민감도(=sensitivity), 특이도(=specificity)는 어떻게 비교해야 하나요 ?

A. 동일한 환자를 대상으로 두 진단 도구를 적용한 경우 민감도(sensitivity)와 특이도(specificity)는 McNemar test로 분석하면 됩니다. 민감도를 비교할 때에는 실제로 양성인 사람들만을 대상으로, 두 진단도구의 McNemar test를 하면 되고 특이도를 비교할 때에는 실제로 음성인 사람들만을 대상으로, 두 진단도구에 대한 McNemar test를 수행하면 됩니다.

Q. ROC 커브는 어떻게 만들 수 있나요?

A. outcome variable이 이분형(binary)일 경우에는 outcome을 예측하기 위해 구축된 모형 (logistic regression)으로부터 예측인자(독립변수)들의 각 측정값에서 민감도와 특이도가 계산될 수 있습니다. 그리고 이렇게 산출된 여러 개의 민감도와 특이도를 가지고 ROC 커브를 만들 수 있습니다.

Q. Area under the ROC (AUC)의 의미는 무엇인가요?

A. AUC는 특정 질병에 대한 overall한 진단력을 나타내는 measure로 ROC을 그린 후에 ROC 아래 면적을 계산하여 나타냅니다. 예를 들어 AUC=1이면 perfect test라고 보고 AUC=0.5이면 worthless test라고 평가합니다.

평가 기준은 다음과 같습니다.

.90-1 = excellent (A)

.80-.90 = good (B)

.70-.80 = fair (C)

.60-.70 = poor (D)

.50-.60 = fail (F)

Ref. Thomas G. Tape. MD, Interpreting Diagnostic tests, University of Nebraska Medical Center

Q. 진단력 평가시 AUC로 살펴봐야 하나요? 아니면 sensitivity, specificity로 살펴봐야 하나요?

A. AUC는 전체적인 진단력을 말합니다. 그러나 민감도와 특이도는 특정 cutoff에 대해서 평가됩니다. 따라서 두 AUC가 차이가 없더라도 민감도와 특이도는 차이가 있을 수 있습니다. 그러므로 아직 cutoff가 정해지지 않은 경우에는 AUC로 전체적인 진단력을 평가 또는 비교하고, 실제 임상에서 사용하기 위해서는 cutoff가 필요하므로 cutoff를 결정해서 이때의 민감도와 특이도가 받아들일만한 것인지를 평가합니다. 따라서 AUC 및 민감도, 특이도를 다 같이 평가함이 좋습니다.

Q. 두 군에 대한 어떤 연속적인 수치들의 cut-off point 값을 결정할 때, (1) 일반적으로 규칙이 있는지요? 또한 (2) NPV+PPV가 최고점이 되는 경우와 민감도+특이도 최고점이 되는 곳의 의미가 어떻게 다른지 궁금합니다.

A. cut-off value를 정하기 위한 general rule은 없지만 몇 가지 방법들은 있습니다. 먼저 cut-off value의 결정에는 크게 두 가지 연구목적이 있는데, 이 중 어떤 목적이 선생님의 연구목적과 일치하는지를 파악하신 다음에 cut-off value를 정하는 것이 좋습니다.

- ① 군 변수와 연속적 수치를 두 범주로 나눈 이분 변수(Binary variable)간 연관성
 - 연속적 수치로 측정된 변수를 cut-off value로 나누어 이분 변수(Binary Variable)로 만든 다음 군 변수와의 연관성을 검정합니다(예, 카이제곱 검정).
 - 이 때 군 변수는 질환 발생에 대한 Gold Standard가 아니어도 됩니다. 왜냐하면 연구목적이 두 변수 간 연관성을 보는 것이기 때문입니다. 만약 군 변수를 Gold Standard 라고 볼 수 있다면 정해진 cut-off value에 대한 민감도와 특이도를 구할 수는 있습니다.
 - cut-off value는 minimum P-value(최소유의확률) 방식으로 정하게 됩니다. 먼저 연속적 수치를 가진 변수에서 cut-off value가 있을 것으로 예상되는 범위를 정합니다. 만약 예상범위가 없다면 측정된 값들의 분포를 보고 결정하는데 5th percentile에서 95th percentile내 값들을 흔히 이용합니다. cut-off value의 예상범위 내에 있는 모든 가능한 값들에 대해 군변수와 이분변수 간 연관성 검정의 P-value를 구하고 그 중에서 가장 작은 P-value를 가지는 값으로 cut-off value를 결정합니다. 이 때 모든 가능한 cut-off value에 대해서 연관성 검정을 하였기 때문에 늘어난 Type 1 error를 보정해야 합니다. 최적의 cut-off value에 대한 P-value를 보정하는 방법은 다음의 참고문헌에 정리되어 있습니다. (참고문헌 : 김선우, 박철근(2002), 간세포암종의 재발 위험과 관련된 한 예후인자로서의 종양의 크기와 범주화, 응용통계연구, 15, 1-8)
- > 위 방법은 군 변수가 Gold standard가 아니어도 P-value라는 객관적인 기준에 의해서 cut-off value를 구할 수 있습니다.

② 진단도구의 개발

- 연속적 수치를 가진 검사방법으로 어떤 질환의 발생을 예측하거나 진단하고자 할 때, 검사수치의 cut-off value를 결정합니다.
- 이 때 질환 발생에 대한 Gold Standard가 있고 이 방법으로 진단한 자료가 있어야 합니다. 따라서 결정된 cut-off value의 민감도(Sensitivity)와 특이도(Specificity)를 구할 수 있습니다.
- 민감도와 특이도는 검사가 실제 환자나 정상인을 얼마나 잘 진단하는가를 나타낸 비율로 false positive 오류와 false negative 오류에 관계된 척도입니다. 특정 질환을 confirm하는데 필요한 검사는 이 질환에 매우 specific해야 하므로 specificity가 높은 검사가 필요합니다. 반대로 질환 발생 여부를 일차적으로 걸러내는 스크리닝 검사라면 덜 specific해도 false negative에 주의하여야하므로 false negative가 낮은 검사 (즉 sensitivity가 높은 검사)가 필요합니다. 만약 sensitivity를 높이려고 cutoff를 변경하면 sensitivity는 올라가지만 specificity는 내려갑니다. 또는 specificity를 높이려고 cutoff를 변경하면 specificity는 올라가지만 sensitivity는 내려갑니다. 따라서 동시에 둘 다 높일 수 없기 때문에 종종 민감도+특이도를 최대로 하는 cutoff를 이용합니다. 그러나 앞서 기술한 것처럼 특이도 보다는 민감도가 중요한 검사라면 특이도는 최소값 (연구나 검사 방법에 따라 연구자가 결정)만 유지하는 범위 내에서 민감도를 최대로 하는 cutoff를 결정하고, 특이도가 중요한 검사라면 민감도는 최소값만 유지하는 범위 내에서 특이도를 최대로 하는 cutoff를 결정합니다.
- Predictive value는 검사의 결과가 주어져 있을 때 실제로 질환 여부를 얼마나 정확히 예측 또는 진단하는가 대한 척도입니다. 따라서 질병 진단에 있어서 실제적으로 더 유용한 척도가 될 수 있습니다. 그러나 이 척도는 질병의 유병률 특히 자료에서 실제 환자의 비율에 따라 많이 변하므로 주의해야 합니다. 따라서 자료로부터 cutoff를 결정하는데 잘 이용되지 않고 있습니다.

Q. 진단도구 개발시, 연속형 변수를 범주화하는 경계 즉, A와 B group으로 나누는 기준을 찾는 방법에 대해 알고 싶습니다.

A. 연속형 측정 변수를 이용한 범주화하여 (A군과 B군으로 나눔) 두 군간 outcome variable이 차이가 난다고 볼 수 있는 경계값을 찾는 방법으로 최소 유의확률 방법 (Minimum P-value approach)이 있습니다. 이는 최적의 cutoff (연속형 변수의 경계)를 찾는 방법으로 모든 가능한 연속적인 경계값을 검토하여 가장 작은 유의확률을 구하는 것입니다. 그런데 이 방법은 다중검정 문제를 야기하기 때문에 즉, 제 1종 오류를 증가시키기 때문에 유의확률을 보정하여야 합니다. 이를 보정하는 방법으로 Bonferroni's correction, Miller and Siegmund method (1982), and Lausen and Schumacher method (1996)가 있습니다.

Q. 실제로 양성인 4개 표본 중에서 3개를 정확하게 예측했다는 사실을 논문에 기술 하였습니다. 그런데 리뷰어가 양성이라고 예측된 7개 중에서 실제로 3개만이 양성 이었다 라는 사실을 기술하라고 하였습니다. 이 사실을 어떻게 기술해야 되나요? 용어 가 따로 있나요?

A. 실제로 양성일 경우에 양성이라고 판단한 확률 즉, sensitivity 에 대해서 기술 하였고, 리뷰어는 양성이라고 예측된 경우 실제로 양성일 확률 즉, Positive predicted value를 제시하라고 말한 것입니다. 보통 진단도구의 특성을 판단할 경우 sensitivity, specificity(실제로 음성일 때 음성이라고 예측할 확률), Positive predicted value, Negative predicted value(음성이라고 예측하였을 때 실제로 음성 인 확률), accuracy(양성을 양성으로 음성을 음성으로 예측한 확률) 등을 보고 판단 하며, 연구의 특성에 따라 선호되는 또는 중요한 속성들이 달라질 수 있습니다. $1 - \text{sensitivity}$ 는 false negative, $1 - \text{specificity}$ 는 false positive 와 같습니다

9. 타당도와 신뢰도

Q. "10-fold Cross-Validation and Bootstrapping Procedure"에 대해 문의드립니다. Multivariate model의 validation이라는데 무슨 의미인지요?

A. validation 은 자료로부터 구축된 모델이 적절하게 구축되어 있는 지에 대하여 평가하는 방법입니다. 일반적인 validation 방법은 자료를 training set과 test set 으로 나누어 training set에서 구축한 모델을 test set 에서 테스트하는 것입니다 (hold-out method). cross validation 은 이러한 validation 을 두 번(교차) 수행하여 - 2fold cross validation: 자료를 test set 과 training set 두 부분으로 나누었으므로- 산출된 결과를 종합하는 것을 말합니다.

기본적인 Leave one out Cross validation(L00CV) 의 경우 데이터를 하나씩 제거하여 나머지 자료들로 모형을 구축한 후, 제거된 개체의 데이터를 모형에 집어넣어 산출된 예측값과 실제값의 차이를 구합니다. 이런 작업을 데이터 내의 모든 개체들에게 시행하는 것을 L00CV라고 합니다. 즉, k-fold cross validation에서 k가 n이 되는 것을 말합니다. 10-fold cross 는 자료를 10 개의 fold로 나누어 9개의 그룹에서 적합한 모형을 나머지 한 그룹으로 예측력을 테스트하는 것을 말합니다. fold의 수가 작을수록 추정치의 bias와 variance 가 증가할 수 있습니다.

Bootstrapping method는 대표본 방법으로 샘플로부터 통계량을 모수적인 방법으로 추정할 수 없을 때, 여러 번(많게는 천 번 이상) 유사한 가상의 샘플을 반복 추출하여, 통계량의 점추정과 구간추정을 시행하는 방법을 말합니다. 자료구조에 따라 간단할 수도 복잡할 수도 있습니다.

Cross Validation 방법과 Bootstrapping 방법은 둘 다 자료에 적합한 모형을 추정하고 평가할 때 사용할 수 있는 방법입니다.

Q. The index of content validity 가 무엇인가요?

A. Content validity(내용타당성)이란 작성된 설문지의 타당성 평가 항목 중 하나로 이 설문지가 이 연구에서 보려는 목적(주제)을 잘 반영하고 있는가를 나타내는 것으로 보통 이 분야 전문가들에 의해 평가됩니다.

따라서 이 전문가들이 각 item에 대해 이 item이 representative(주제에 대해)하다 또는 약간의 minor revision으로 representative하다라고 평가한 비율이 content validity index입니다.

10. 유의확률(P-value), 본페로니 보정(Bonferroni's correction)

Q. 분석결과를 받아 보면 Bonferroni correction을 하는 경우가 발생하는데, 이를 어떻게 해야 하고 왜 하는지 알고 싶습니다.

A. 통계적 검정결과에서 보통 유의수준 0.05를 기준으로 결과를 해석하는데, 통계검정 과정에서 유의수준이 0.05가 되지 않는 경우가 있습니다. 예를 들어 두 검정이 서로 independent 하지 않은 경우입니다. 즉, 서로 독립적이지 않은 검정을 2번 한 경우 유의수준이 각각의 경우에 0.05라 할 때, 최대 0.10이 될 수 있습니다. 이는 Bonferroni inequality에 근거한 것입니다. 즉 False positive rate이 최대 2배 증가한다고 할 수 있습니다. 그러므로 0.05라는 기준을 맞추기 위해 $(0.05/2+0.05/2)=0.05$ 와 같이 correction을 하는 것이 Bonferroni correction입니다. 예를 들어 case(심장질환)와 control(비심장질환)의 수술 전 심박동과 수술 후 심박동 차이에 대하여 수술 전과 수술 후에 Two sample test 결과 P-value가 0.1224와 0.1028이라고 할 때, 첫 번째로 유의수준을 보정하여 해석할 수 있는데 서로 독립적이지 않은 검정이 2번 수행되었으므로 $0.05/2=0.025$ 를 기준으로 분석에서 도출된 P-value에 대한 유의성을 판단하는 것입니다. 두 번째 방법으로 0.05라는 통상적인 유의수준을 이용하기 위해 P-value를 correction하는 방법으로 Bonferroni correction P-value $0.1224*2=0.2448$, $0.1028*2=0.2056$ 을 0.05 기준에서 해석하는 것입니다.

Q. 같은 사람을 대상으로 N 시점마다 측정하여($N \geq 3$, ex. 8시간, 12시간, 24시간) N 시점 각각에서 두 군간 다른지 비교 하였습니다.

이런 경우 P-value에 Bonferroni's correction을 하라고 하는데 해야 되는 것이 맞는지요? 맞다면 어떻게 하는지요?

A. N 개의 각 시점마다 검정하였을 경우, N 개의 검정 결과들 간에는 서로 독립이라고 볼 수 없습니다. 그러므로 P-value는 Bonferroni's method에 의해 보정된 값을 이용하셔야 합니다. 방법은 산출된 P-value에 N 을 곱하여 새로 계산된 값을 쓰셔야 됩니다.

Ref. Bland M. *An introduction to medical statistics*: Oxford University Press
Oxford, UK, 2000, pp144-147

Q. 한 변수에 대한 분석을 하였습니다. 이에 더하여, 이 변수에 대하여 범주화한 후 재분석을 하고자 하는데, 가능한지요?

A. 동일한 변수에 대하여 다르게 범주화 또는 변환하여 분석하고, 이를 모두 연구결과로서 기술한다면, 다중비교 검정(Multiple Comparison test)으로 인한 통계적 오류중 하나인 Type I error (False positive error)가 발생합니다. 이를 보정하기 위하여 동일한 변수에 대한 분석 검정 수만큼 Bonferroni's correction을 해야 합니다 (즉, $P\text{-value} \times \text{분석 검정 수}$).

Q. 보통 유의수준 0.05를 기준으로 유의하다 혹은 유의하지 않다를 해석하는데 현재 살펴보고 있는 논문에서는 Mann-Whitney test를 수행할 때는 유의수준을 0.0001을 기준으로 하고, Fisher's exact test를 수행할 때는 유의수준을 0.0013로 해석을 하고 있습니다. 왜 분석 방법마다 기준이 다른지요?

A. 유의수준이 변경된 이유는 Bonferroni's correction 과 같은 유의수준의 보정을 수행했기 때문일 것입니다. 가령 어떤 변수를 술전, 술후에 측정하여 술전과 술후에 두 그룹 간의 차이가 있음을 보인다고 할 때 만약 술전 test 만을 수행하게 된다면, 1) 술전 test가 올바르게 수행됨 (95%)

2) 술전 test가 올바르게 수행되지 않음. (5%)

술전 test를 한번만 수행하게 되므로 1), 2) 와 같이 두 가지 상황이 발생할 수 있고, 2) 처럼 잘못 판단할 확률(1종 오류)의 수준을 0.05 로 고정시키게 됩니다. (= 유의수준을 0.05로 설정하는 것과 같음)

만약 술전 test와 술후 test를 수행하게 된다면, 다음과 같은 네 가지 경우가 발생할 수 있습니다.

3) 술전 test가 올바르게 수행됨. & 술후 test가 올바르게 수행됨. ($95\% \times 95\% = 90.25\%$)

4) 술전 test가 올바르게 수행됨. & 술후 test가 올바르게 수행되지 않음.

($95\% \times 5\% = 4.75\%$)

5) 술전 test가 올바르게 수행되지 않음. & 술후 test가 올바르게 수행됨.

($5\% \times 95\% = 4.75\%$)

6) 술전, 술후 test가 모두 올바르게 수행되지 않음. ($5\% \times 5\% = 0.25\%$)

술전 술후 test를 수행할 때에는 적어도 한번 잘못된 판단을 하게 될 확률이 9.75%로 증가하게 됩니다. 따라서 이러한 1종 오류가 증가하는 것을 보정해주는 것을 Bonferroni's correction 이라고 합니다.

Q. 3*3 table 에서 기대빈도가 5보다 작은 셀이 여러 개 있어서 Fisher's exact test를 수행한 후에 차이가 있다고 나오면 사후검정을 수행해서 Bonferroni's correction *9를 해줘야 한다고 하였습니다. 근데 제 빈도표에는 빈도가 0인 셀들이 많아서 나누고 보면 결국 안나오는 게 생겨요 그래도 *9를 해줘야 하나요?

A. *9 를 하는 이유는 그렇게 3*3 table로부터 2*2 table로 쪼갤 수 있는 모든 가짓수가 9개이기 때문입니다. 따라서 선생님이 사후검정으로 각 범주들을 나누어 가능한 모든 2*2 table 을 산출해보고 그 개수만큼 곱해주면 됩니다.

Q. case와 control에 대하여 특정 변수를 분석하고 이 변수를 남녀를 나누어 다시 case와 control에 대하여 각각 분석하는 경우에 P-value는 어떻게 되나요?

A. case와 control 전체 자료에 대한 특정 변수에 대한 분석 결과는 P-value를 그대로 사용하시면 됩니다. 그러나 남녀를 나누어(subgroup) 그 변수에 대하여 다시 분석하시게 되면 이 경우는 각 성별 결과에 대하여 P-value를 correction 하여야 하는데 남과 여 2 group으로 나누어 분석하였기 때문에 (즉, 검정한 subgroup 수 = 2) 각 성별 결과 P-value에 *2 correction 하시면 됩니다. 즉, 원래 분석에서 남자의 경우 P-value 0.1224, 여자의 경우 P-value 0.1013이면 각각 남: $p=0.1224*2=0.2448$ 이고 여: $p=0.1013*2=0.2026$ 이 됩니다. 이와 같이 하는 이유는 전체 자료에서 검정한 후 각 subgroup에 대한 검정이 전체 자료에서의 검정과 연관되어 (독립적이지 않은) 있어서 부분군 분석에서 false positive error가 나올 가능성이 커지기 때문에 이를 보정해주어야 하기 때문입니다. 즉 각 subgroup에 대한 검정의 유의수준은 0.05가 아니라 최대 $0.05*(\text{subgroup 수})$ 가 될 수 있으므로 각 subgroup 검정에 대한 유의수준을 $0.05/(\text{subgroup 수})$ 로 변경하거나, 또는 각 subgroup 검정 결과 P-value에 $*(\text{subgroup 수})$ 를 해줍니다.

Ref. Bland M. An introduction to medical statistics: Oxford University Press Oxford, UK, 2000, pp144-147

11. 표본의 크기 (sample size)

Q. 전향적(prospective) 연구 입니다. 60세 이상과(고위험군) 60세 미만(저위험군) 간에 질병(양성, 음성) 비율이 다른지 살펴보고자 합니다. 카이제곱 검정을 통해 연령 두 군간 질병 유무 비율이 다르다고 할 수 있었습니다. 그런데, 문제는 질병이 양성으로 진단된 환자는 전체 4939명중에 27명밖에 안됩니다. 이와 같은 경우, 60세 이상과(고위험군) 60세 미만(저위험군)간에 질병(양성, 음성) 비율이 다르다고 하는 것은 통계적으로 문제가 없나요?

A. 결론에 대한 신뢰성 문제가 발생할 수 있습니다. 모든 연구에서도 그렇지만 특히 이렇게 발생율이 낮은 연구에서는 고위험군과 저위험군 각각에 대한 적절한 sample size를 먼저 산출한 후 연구를 진행하는 것이 좋습니다.

Q. 두 집단 간 비교를 하고자 하는데 평균이나 다른 기초통계들은 눈으로 보기에 차이가 있는 것처럼 보이지만, 실제로 P-value는 높게 나왔습니다. 각 집단 간 n수가 8명 6명인데요. power가 좀 떨어질 수 있다고 하는데 이게 무슨 말인가요?

A. n 수가 8, 6 명이면 sample size가 작아서 실제로 차이가 있음에도 불구하고 통계적으로 그 차이를 찾아내지 못할 수도 있다고 의심할 수 있다는 의미입니다. 이것이 power(실제로 차이가 있다면 차이가 있다고 해주는 힘)가 떨어질 수 있다는 의미입니다. 왜냐하면 적은 n수로 두 집단 간 차이가 있는 결과가 나왔다고 해서 두 집단 간에 실제로 차이가 있다고 결론을 내리는 것은 위험(즉 실제로 차이가 있다는 것을 입증하기에는 n수가 부족)하기 때문입니다. 그러나 실제 n수가 늘어나면 이러한 차이가 유지될 것이라는 보장이 없기도 합니다.

Q. regression 을 수행하였는데 sample size가 1100이고 독립변수의 수가 13개 입니다. 제가 다른데서 수업을 듣고 와 보니 변수가 하나 있을 때마다 sample size가 20개씩은 있어야 된다는 소릴 들어서, 독립변수가 13개면 최소 260개는 있어야 되는 건가요? sample size가 부족한 것이 아닌지요.

A. 특별히 한 변수 당 20개정도는 있어야 된다는 규칙 같은 것은 없습니다. 다시 말하면 선생님의 자료의 경우에는 자료의 수가 110 개이기 때문에 독립변수의 숫자에 비해 sample size가 작다고 말할 수 없습니다.

다만, 현재 분석대상자 수(sample size)에 비해 설정한 변수의 수가 너무 많거나 혹

은 이분형 종속 변수의 경우 event의 수에 비해 설정한 변수의 수가 너무 많거나 하면 회귀분석에서 회귀 계수 추정 시 문제가 생길 수 있습니다. 또한 설정한 설명 변수(독립변수, 위험인자)들이 많으면 이중에는 서로 높은 상관관계를 가지는 변수들도 있을 가능성이 크기 때문에 높은 상관관계를 가지는 변수들이 있을 경우 역시 회귀계수 추정 시 문제가 생길 수 있습니다. 그러므로 현재 분석대상자 수(sample size)에 비해 설명 변수의 수가 너무 많으면 안 되고 이분형 종속 변수의 경우 event 수에 비해 설명 변수(독립변수, 위험인자)의 수가 너무 많아도 안 되고, 서로 높은 상관관계를 가지는 변수들이 설명 변수로 고려 되서도 안 됩니다.

따라서 연구를 진행하기에 앞서 설명 변수와 반응 변수간의 관계를 파악해서 연구 목적을 고려해 볼 때 어떤 인자가 종속변수(결과변수, 반응변수)와 연관이 있는지 꼭 보고 싶은 또는 영향을 끼치리라 생각되는 독립변수(설명변수, 위험인자)들 만으로 선정하는 것이 중요합니다.

Q. T-test의 경우 정규성 가정을 만족해야 한데 저의 샘플은 한 그룹이 9개, 나머지 한 그룹이 20개 입니다. 10개보다 작다고 해서 꼭 비모수 검정을 해야 하는 건가요? 정확한 기준이 뭐지요?

A. 정규성을 만족하는 샘플의 최소 기준에 대해 정확한 기준이 있는 것은 아닙니다. 다만 정규성 test도 sample size가 작을 경우 정규성에 대한 신뢰성이 보장되지 않습니다. 이와 같이 정규성 test 의 신뢰성이 확보되지 않으면 정규성 가정을 만족해야 하는 T-test보다 정규성 가정이 필요하지 않는 비모수 검정을 적용하는 것이 보다 “안전” 하다고 볼 수 있습니다.

Q. 비교하고자 하는 그룹 간 age나 weight 와 같은 baseline characteristics 들을 비교하고자 합니다. 그런데 한 그룹은 61명이고 다른 그룹은 2000 명이 넘습니다. 이렇게 샘플사이즈가 크게 차이가 나면 비교하는데 문제가 있는 것은 아닌가요?

A. 한 그룹이 60명이고 다른 그룹이 2000명이라도 비교하고자 하는 변수를 검정하는데 적절한 power를 갖는다면 문제는 없습니다. 그러나 대부분 한 그룹의 sample size가 61명으로 고정되어 있을 때 적절한 power를 갖기 위해 다른 한 군에서 필요한 sample size를 61명 이상으로 계속 늘리면 power는 점점 증가하다가 어느 점에서 더 이상 power가 증가하지 않게 됩니다. 예를 들면 500명이상으로 가면, 예를 들면 500명이상으로 가면 2000명으로 분석하는 것이 불필요하기는 합니다. sample size가 작을 경우에는 위와 같은 그룹 간 sample size의 imbalance 가 더욱 큰 문제

가 될 수 있습니다.

예) 남녀 간 평균의 차이가 있어 보이지만 (2.47 vs 4.73), black, other 군의 수가 적기 때문에, 이를 감안하여 weighted mean을 산출한다면 남녀의 평균은 각각 6.88, 6.967로 나타나게 됩니다. 따라서 이 경우에는 반드시 인종을 보정하여 다중 분석을 수행해야 합니다.

HADS				
Gender of subject	Ethnicity	Mean	N	Std. Deviation
Male	White	1.4800	133	1.6300
	Black	6.6000	10	1.7800
	Other	12.5800	9	2.7400
	Total	2.4729	152	3.3121
Female	White	2.7100	114	1.9600
	Black	6.2600	19	1.2400
	Other	11.9300	28	4.1100
	Total	4.7324	161	4.2419
Total	White	2.0477	247	1.8889
	Black	6.3772	29	1.4262
	Other	12.0832	37	3.7964
	Total	3.6351	313	3.9770

Q. 두 집단 간 평균 비교를 수행하고자 합니다. 이때, 나이로 일대일 매칭을 시켜서 비교를 할 수도 있나요? 할 수 있다면, 한 그룹의 sample size가 다른 그룹보다 3개 정도 많은데, 어떻게 해야 되나요? 3개는 빠지게 되나요?

A. age로 매칭을 실행 시에는 가령 case-control study에서는 case 의 subject 와 나이가 같은 control 의 subject를 찾아서 1대1로 매칭을 시켜줍니다.

따라서 age를 기준으로 Exact individual matching 을 실시한다면, age가 같지 않은 control 그룹의 환자들은 선택되지 않으며, 따라서 결과적으로 매칭 된 페어의 수가 작아질 수 있습니다.

특히 선생님의 자료에서 한 그룹이 3개 정도 많다면, 1대1 매칭을 수행 시에는 정확하게 일치하는 age를 가지는 다른 그룹의 subject가 충분하지 않을 수 있습니다.

즉, 기준 그룹의 모든 subject 들과 다른 그룹에서 3개만 제외한 모든 subject 들이 1대1로 매칭이 안 될 수도 있습니다. 비슷한 페어를 만들다 보면 더욱 작은 수의 매칭만이 만들어집니다.

이렇게 해서 매칭 된 쌍의 수가 너무 적게 되면, 즉, 기존의 sample 에서 누락되는 subject 들이 많다는 것은, 분석의 신뢰성이 떨어져 문제가 됩니다.

보통 matching 은 observational study에서 다수의 control 그룹과 control 군보다는 작은 수의 case 그룹의 subject 들을 적절한 비율에 따라 의미 있는 confounder 들의 balance를 달성하여 새로운 dataset을 만들기 위해 시행합니다. design stage에서의 조정은 small sample size에서는 위험할 수 있습니다. 따라서 matching을 적용하기 위해서는 되도록 많은 sample population을 수집해야 합니다.

예) 5명의 case 군과 20명의 control 군을 age로 matching 시켜준다면,

case's age= 20, 15, 31, 70, 55

control's age= 21, 33, 55, 56, 48, 29, ..., control 군의 나이 분포가 20대~50대라고 가정함.

이때 1:1 matching 을 시행할 경우 (동일한 나이 대 끼리 매칭)

case 군에서 20, 31, 55살의 환자만이 매칭 되고, 15살, 70살 환자들은 matched dataset에서 제외됨. 따라서 전체 sample size 25 명중 단 6명만이 분석대상으로 선정됨. 이는 신뢰성에 심각한 위험이 있으며, 이런 경우 matching 을 시행하는 것은 적절치 않음.

case가 5명이 아니라 200명이고, control 집단이 300명이라면, 위와 같이 matching 을 시행할 경우에는 좀 더 많은 쌍이 생성될 수도 있음. 따라서 통계적으로 산출된 적정 sample size를 고려하여 Matched Data analysis 의 타당함을 말해줄 수 있음.

12. 기타

Q. 귀무가설(H_0 , Null hypothesis)은 어떻게 써야하는지요?

A. Null hypothesis는 연구에서 주장(입증)하고자하는 가설 (이를 H_A , 대립가설이라 부릅니다)의 반대가설을 말합니다. 예를 들어 특정 치료군과 대조군간 회복시간이 차이가 있다는 것을 이 연구에서 입증하고자 한다면 null hypothesis는 회복시간이 두 군간 차이가 없다가 됩니다. 참고로 가설이 여럿일 경우 여러 개의 가설들을 기술하면 됩니다. 그리고 자료에 의한 가설 검정 결과는 원칙적으로는 reject the null hypothesis (귀무가설을 기각할 수 있다) 또는 do not reject the null hypothesis (귀무가설을 기각할 수 없다)가 됩니다.

Q. 결측치(missing value)가 있는 변수를 포함하는 자료는 어떻게 분석하는 것이 좋을까요?

A. 모든 연구의 시작과 끝에 동일한 개체(subjects)를 포함하는 것이 좋기 때문에, 연구 상 중요하지 않은 특정 변수에서 결측치의 수가 많은 경우 그 변수를 제외하여 분석하는 것이 좋습니다. 혹은 결측치를 포함한 개체를 연구에서 처음부터 제외시키는 방법도 생각할 수 있는데, 이 방법은 표본의 크기(sample size)가 적어져 분석의 검정력(power)을 감소시킬 수 있습니다. 뿐만 아니라 결측치가 있는 개체들이 그렇지 않은 개체들과 특정 성질이 다르다면 이 경우 biased 결과를 산출할 수 있습니다. 통계적인 방법으로 대치법(imputation)을 할 수는 있으나 연구목적, 자료의 특성 등에 따라서 대치법 결과의 민감도(sensitivity)가 매우 다를 수 있어 많이 사용되지는 않습니다. 따라서 연구를 수행함에 있어서 연구계획을 먼저 세워서 필요한 변수는 무엇이고, 이 필요한 변수들에 대해 충분히 측정값들을 얻을 수 있는지를 먼저 고려함이 좋겠습니다.

Q. 시점마다 산소 주입군과 비주입군 간 비교를 하고자 합니다. 결측치 처리는 어떻게 해야 되나요?

A. 연구목적과 관련된, 즉, 실험 결과 예측이 좋지 않아 산소 비주입군에서 어쩔 수 없이 다시 주입하게 된 사람의 경우 이 환자의 예측을 결과에 포함시켜야 합니다. 이 경우 보수적(conservative)인 방법으로 결측된 값들은 환자의 좋지 않은 예측들을 나타내는 결과를 LOCF 방법으로 대체하여 분석에 포함할 수 있습니다.

LOCF(Last Observation Carried Forward) 방법은 이전 시점에 관측된 값을 대체하는 방법입니다. 하지만, 이 방법도 LOCF 이후의 변수 값 간의 분산을 고려하지 않아 over-estimation될 수 있으므로 주의해서 사용해야 합니다.

Q. Single arm study 와 Double arm study는 어떻게 다른가요?

A. Single arm study는 임상시험에 참여한 모든 피험자가 동일한 처리를 받은 것을 말합니다. 대조군이 없이 진행되기 때문에 적용질환이 안정적이고 위약 효과나 자연적인 변화를 배제할 수 있는 경우에만 제한적으로 선택됩니다. Double arm study(=Two arm study)는 처리 그룹(intervention arm)과 대조 그룹(control arm)간 비교 연구를 생각해 볼 수 있습니다.

Q. Confidence interval(신뢰구간)이 의미하는 바는 무엇인지요?

A. 예를 들어, 만일 어떤 집단의 평균 혈압에 대해 이야기를 하고자 할 때 평균만 제시한다면, 과연 이 평균이 모수(미지의 수, ex 실제 평균)와 얼마나 가까운지 알 수 없습니다. 그러므로 실제 평균 혈압이 속할 것으로 기대되는 범위를 제시하게 되는데 이 범위가 실제 평균을 포함하고 있을 지를 통계적으로 얼마나 믿을 수 있는지 신뢰도의 아이디어가 첨가된 구간을 신뢰구간이라고 합니다.

보통 95% 신뢰구간이라 함은 우리가 제시하는 신뢰 구간에 모수(미지의 수, ex 실제 평균)가 포함될 가능성이 95%라는 뜻이 됩니다. 그렇기 때문에 신뢰 구간에 포함되지 않을 가능성 또한 5%가 됩니다. 신뢰도를 높이면 구간의 폭이 넓어져 모수에 대한 정보의 요약이 어렵고, 반대로 신뢰도를 낮추면 구간의 폭이 좁아져 판단 착오를 할 위험이 매우 커지게 됩니다. 이런 이유로 신뢰도를 분석자 마음대로 높이거나 낮출 수 없습니다. 대신, 실제 데이터 분석을 통한 과거 경험에 비추어 볼 때 일반적으로 95% 신뢰도를 사용하여 신뢰구간을 구합니다.

Q. 단변량과 다변량 분석은 어떻게 다른가요?

A. 보통 선생님께서 말씀하시는 단변량 분석(Univariable analysis)과 다변량 분석(Multivariable analysis)은 다음과 같습니다. 단변량 분석(Univariable analysis)은 종속변수와 연관이 있는 독립변수 하나를 가지고 분석하는 방법을 말합니다. 반

면 Multivariable analysis는 종속변수와 연관이 있으리라 생각되는 독립 변수들(2 이상)을 가지고 분석하는 방법을 말합니다.

그러므로, Univariable analysis에는 Chi-square test, T-test, Mann-Whitney test, Simple linear regression 등을 생각해 볼 수 있고, Multivariable analysis에는 Multiple linear regression, Multiple logistic regression과 같은 분석을 할 경우 처럼 관심변수인 종속변수에 영향을 끼치리라 생각되는 독립변수 2개 이상을 가지고 다중회귀 분석하는 것들을 생각해 볼 수 있습니다.

Q. ITT(Intent to treat) 분석에서 집단 간 성별 차이가 있었는데, 추후 분석에서 성별변인을 통제하지 않아도 되는지요?

A. confounder 는 결과변수에도 영향을 끼치고, 독립변수 간에도 차이가 있는 변수를 말합니다. 다시 말하면, 각 endpoint 들에 대하여 성별이 risk factor 가 될 수 있다면, 보정해줘야 합니다. risk factor 가 된다는 의미는 통계학적으로 비율의 차이가 있을 때 risk factor로 간주하는 것이 아니라, 임상적으로 의미가 있는지에 대해 연구자가 판단해야 합니다.

Q. Mediator 가 무엇인가요?

A. moderator , 엄밀히 mediator 라고 하는 변수들은 다음과 같은 조건이 있습니다. mediating variable의 정의

A predictor hypothesized to lie on the causal pathway between a predictor of interest and the outcome, and thus to mediate the predictor's effects.

1. the predictor of interest also predicts the mediator
2. the mediator predicts the outcome in a model controlling for the predictor of interest.
3. addition of the mediator to a amutipredictor model for the outcome attenuates the estimated coefficient for the predictor of interest.

예를 들어 statin drugs의 사용은 LDL cholesterol 수치를 낮출 수 있으며, 이는 곧 heart attack 의 risk를 감소시킬 수 있습니다. 여기서 LDL cholesterol 변수는 drug>heart attack 의 causal pathway 의 중간에 위치하는 mediation effect를 가지고 있습니다. 이러한 mediator 들은 path analysis 나 사회과학에서 많이 사용되는 SEM(structural equation modeling) 을 적용할 수 있습니다. 대체로 이러한 방법

들은 여러 번 분석을 수행하여 해당 변수들의 유의성을 살펴보고 임상적인 의미와 함께 고려하여 인과관계에 놓여있는지를 판단합니다.

path analysis 의 경우, outcome :C, interested predictor = A, mediator =B 라고 한다면

1. A → B
 2. A → C
 3. B → C (A, B를 보정하여)
- 에 대한 regression 을 수행합니다.

만약 1이 성립하지 않는다면, mediator라고 볼 수 없습니다. 1이 성립한다면, 2와 3의 분석 결과를 비교하여, mediator의 효과에 대해 살펴볼 수 있습니다.

구조방정식(SEM) 방법은 인과관계가 존재할 것으로 예상되는 구성들 간의 회귀분석과 요인분석을 통해서 구성개념간의 상호관계를 분석하는 방법으로, 회귀분석과 요인분석(측정 변수들의 상관관계를 고려하기 위한 다변량 분석방법)을 수행하여 이론적인 구조를 설명하는 방법입니다.

구조방정식 방법을 구현하는 프로그램으로 AMOS, LISREL 등이 있다고 합니다.

Q. 전체자료에서 두 그룹의 평균과 표준편차를 각각 표시한 후, P-value는 매칭 된 자료에서 산출된 P-value를 제시하였습니다. 이렇게 하면 안 되나요?

A. 먼저 전체자료의 기초통계를 기술하고, 전체자료의 분석 결과를 가능하다면 기술한 후, 새로 table을 만들어 matched data 에 대한 기초통계와 검정 결과를 기술하세요. matched dataset 은 중요한 confounder 들의 balance를 달성시켜주기 위해서 가공한 dataset 이므로 마치 새로 design 을 수행한 것과 같습니다. 특히 Observational study에서 원래의 sample과 matched sample 에 대한 분석결과를 모두 수록해주어서, matching 으로 인해 변경된 정도가 어느 정도인지를 살펴봐야 합니다. matched data set 의 분석결과를 마치 전체 데이터에 대한 분석결과처럼 사용하면 안 됩니다.

Q. 제가 1대1 매칭을 수행해서 분석을 했는데 이것 뭐라고 부르나요?

A. 논문에 기술할 때 그냥 1:1 matching 어떤 기준으로 했는지를 달아주시면 됩니다. (1:1 matching for age, sex). 또 1:1 matching 을 수행했을 경우에는

individual matching 이라고 부르고 n:m matching 을 수행했을 경우에는 frequency matching 이라고 부릅니다. 그러나 논문에 기술할 때에는 꼭 이런 이름들을 사용하지 않아도 무방합니다.

Q. 설문지 분석하는 경우에 요인 분석 (factor analysis) 후 factor를 이용하여 각 설문을 group화 하는데, 이 경우 grouping하는 것과 그렇지 않는 경우를 어떤 기준으로 하는지 알고 싶습니다.

A. 우선 요인 분석으로 생성된 각 Factor에서 설문의 문항에 잘 설명할 수 있는 요인(factor)의 수를 결정하게 됩니다. 이 경우 요인의 수는 요인 분석 수행에서 수반되는 Eigen value를 이용하게 됩니다. 일반적으로 Eigen value가 1보다 크거나 같은 요인을 선택하여 group화하는 요인으로 이용됩니다. 따라서 전체 설문 n개에 대한 요인 분석 결과 설문과 동일한 수인 n개요인(factor 1, factor 2, ..., factor n)이 생성되고 요인을 선택하게 되는데, 요인을 선택하는 방법으로 eigen value 1 이상인 $m(m \leq n)$ 개의 요인이 선택하기도 하고, 다른 방법으로 factor 1부터 k까지 $k(k \leq n)$ 개 요인들이 자료를 설명할 수 있는 누적비율이 70~80%이상인 것을 선택하기도 합니다. 여기서 누적비율이라 함은 Factor 1부터 k까지 Eigen value의 합을 전체 Eigen value의 합으로 나눈 것을 의미합니다. 그리고 인자의 회전방법(예를들어, Varimax rotation)으로 group하기 위해 loading을 계산합니다. 이렇게 계산된 loading이 각 요인마다 존재하게 되는데, 각 설문의 문항에서 각 요인의 loading 값 중 가장 큰 값의 요인에 설문을 대응하여 각 요인에 속하는 설문을 grouping 할 수 있습니다. 그리고 설문 1의 loading 값 중 가장 큰 것이 Factor 1에서 0.3210, Factor 2에서 0.3190이라고 한다면 loading 값으로는 Factor 1에 속하지만 Factor 2에도 비슷한 영향을 미치기 때문에 grouping에서 제외하기도 합니다.

13. 주요 용어

가변수	dummy variable
검사특성곡선	ROC curve
검정력	power
결정 계수	R-square
결측 대체 방법	missing imputation method
경계값	cut-off point
공분산 분석	Analysis Of Covariance(ANCOVA)
관측연구	observational study
교락 변수	confounding factor
교차 타당법	Cross-Validation
귀무가설	null hypothesis
내용타당성	content validity
다변수 모형 분석	Multivariable model analysis
다중 로지스틱 회귀분석	Multiple logistic regression analysis
다중공선성	multi-collinearity
다중비교 종류	Fisher's LSD, Bonferroni, Tukey, Scheffe
다중비교	multiple comparison
다중회귀분석	multiple regression
던넛검정	Dunnett's test
로그 순위 검정	Log rank test
매칭 자료분석	Matched data analysis
맥네마 검정	McNemar test
맨-휘트니 검정법	Mann-Whitney test
민감도	sensitivity
반복측정 자료 분석	Repeated Measures ANOVA
변수변환	variable transformation
변수선택	variable selection
본페로니 보정법	Bonferroni's correction
분산 안정화 변환	variance-stabilizing transformation
분산 팽창 인자	Variance Inflation Factor(VIF)
부스트랩 방법	Bootstrapping method
브레슬로 검정	Breslow test
비례위험 모형	Cox proportional hazard model (Cox regression)
사례-대조	case-control

사분위 범위	interquartile range
생존 분석	Survival analysis
선형회귀분석	Linear regression analysis
순서 회귀 분석	Ordinal regression analysis
스피어만 상관분석	Spearman correlation analysis
시간의존 비례위험 모형	Time dependent Cox proportional hazard model (Time dependent Cox regression)
신뢰구간	confidence interval
오즈비	Odds Ratio(OR)
요인분석	Factor analysis
윌콕슨 부호순위합 검정	Wilcoxon's signed rank test
윌콕슨 이표본 검정	Wilcoxon's two sample test
윌콕슨 순위합 검정	Wilcoxon's rank sum test
유의확률	P-value
이변량 정규분포	Bivariate normal distribution
이요인분산분석	Two-way ANOVA
일요인분산분석	One-way ANOVA
정규성	normality
제 1종 오류	Type I error
존키어 검정	Jonckheere's test
중도 절단된 자료	censored data
카이제곱 검정법	Chi-square test
카플란 마이어 방법	Kaplan-Meier method
코크란 아미티지 검정	Cochran-Armitage trend test
크루스칼 월리스 검정	Kruskal-Wallis test
통계적 타당도	validity
특이도	specificity
티검정	T-test
편상관분석	Partial correlation coefficient analysis
표본의 크기	sample size
피셔의 정확 검정	Fisher's exact test
피어슨 상관 분석	Pearson correlation analysis