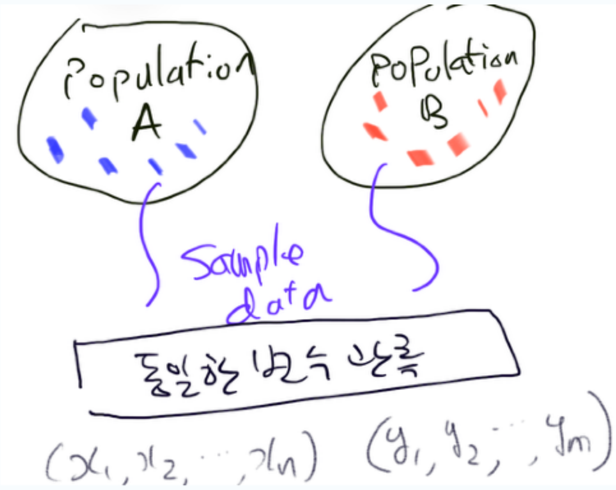
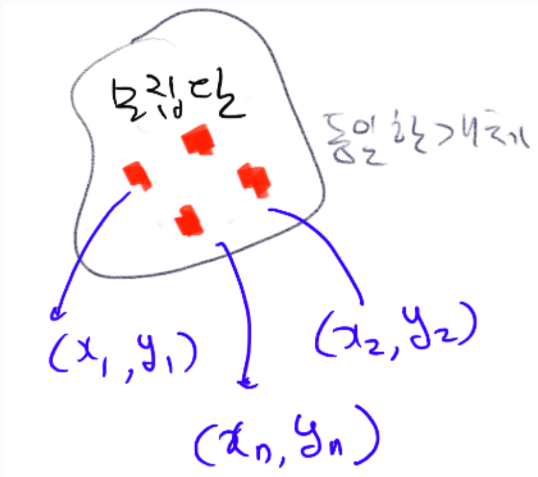
1. 개념

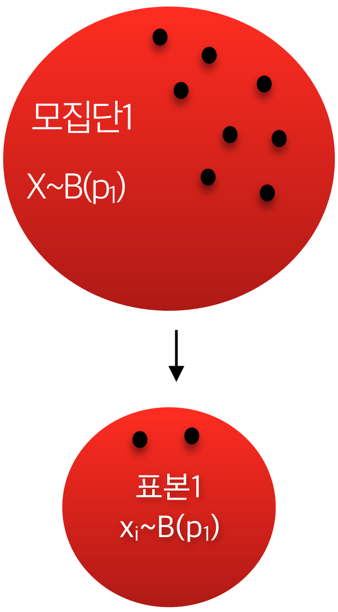
- 일변량 두집단 분석은 측정변수가 1개이고, 이를 2개 집단의 차이가 유의한지 significant 검정하게 된다. -> 관심 측정 변수는 1개임
  - 측정변수는 정량적 데이터, 정성적 데이터로 나뉨
  - 정성적 데이터 관심 모수 : 모비율 -> 두 집단 모비율 차이 검정
  - 정량적 데이터 관심 모수 : 모평균, 모분산 -> 두 집단 모평균/모분산 차이 검정

- 비교하는 2개의 집단은 "독립" independent, "짝진" paired 으로 나뉜다. 독립과 짝진의 구별은 관측 대상이 동일개체인지 상이한 개체인지에 의해 결정된다.

독립집단 경우	짝진집단 경우
	
관측변수 X=Y 동일함	
서로 다른 개체로 부터 관측한다.	동일개체로부터 관측된다.
관측치 개수 n, m 크기는 같을 필요 없음	관측치 개수는 n개 항상 동일하다.
경영학 전공자와 통계학 전공자의 일주일 공부시간 차이를 알고자 한다. 경영학 전공자 30명, 통계학 전공자 20명을 임의 추출(확률표본)하여 공부시간을 측정하였다.	경영학 전공자와 통계학 전공자의 일주일 공부시간 차이를 알고자 한다. 그런데. 학점 군에 따른 공부시간 차이를 고려하여 (4.5~4.3), (4.3~4.1), (4.1~3.9), ... 각 구간에서 한 명씩 임의 추출하였다.

2. 독립 2집단 모비율 차이 검정

1) 시나리오 (연구문제)

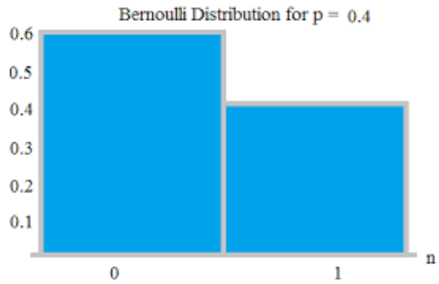


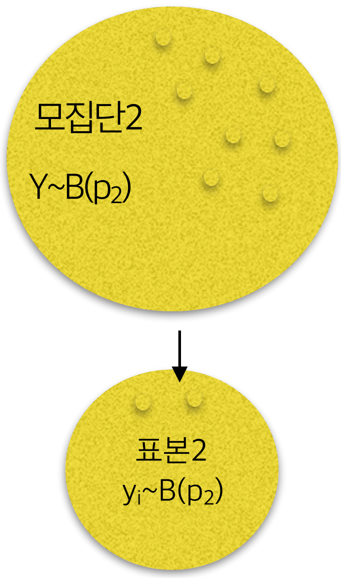
두 모집단 서로 독립

개체에 의해 관측되는 값은 이진형 binary

(0=실패, 1=성공)

Bernoulli 베르누이 분포





확률표본 (iid) = 서로 독립이고 동일분포에서 추출

$x_i \sim B(p_1)$ $y_i \sim B(p_2)$

$x_1 = 0, x_1 = 1, x_3 = 1, \ldots, x_n = 0$ 데이터 $y_1 = 1, y_2 = 1, y_3 = 0, \ldots, y_m = 0$

2) 모수와 점추정치(MVUE)

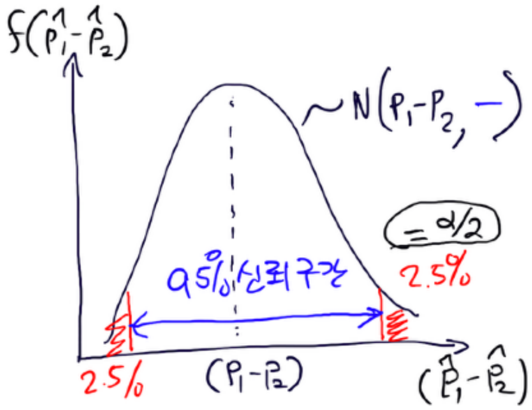
- 모수 : p_1, 모집단1 비율, p_2 모집단2 비율
- 점추정치 (MVUE) : $\hat{p}_1 = \dfrac{\#of1's\_sample1}{n}, \hat{p}_2 = \dfrac{\#of1's\_sample2}{m}$
- 샘플링 분포 (대표본) $min(np_1, mp_2) > 5$, 모비율에 대한 정보가 없다면 0.5 사용

  * $\hat{p}_1 \sim N(p_1, \dfrac{p_1(1-p_1)}{n})$ (중심극한정리-대표본이론 Large sample theory)
  * $\hat{p}_2 \sim N(p_2, \dfrac{p_2(1-p_2)}{m})$: 동일한 이유

✳ 정규분포의 가법성에 의해 $\hat{p}_1 - \hat{p}_2 \sim N(p_1 - p_2, \frac{p_1(1-p_1)}{n}) + \frac{p_2(1-p_2)}{m}$

3) 신뢰구간 (대표본)

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}$$



예) 95% 신뢰구간의 $z_{1-\alpha/2} = z_{0.975} = 1.96$

A1 =NORM.S.INV(0.975)

	A	B	C	D	E
1	1.95996				

4) 검정통계량, 가설검정

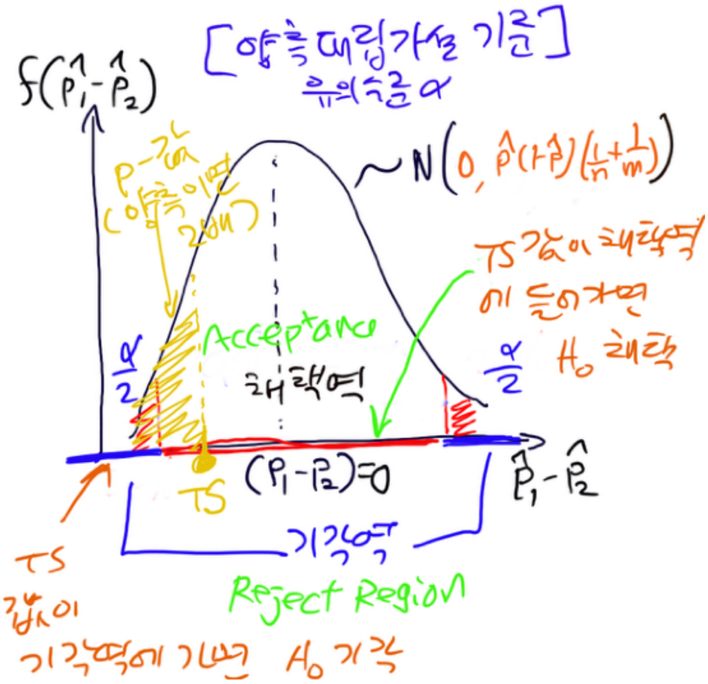
- 귀무가설 : 두 모집단 비율은 동일하다. $p_1 = p_2$ -> 모수로 전환 $\theta = p_1 - p_2 = 0$
- 대립가설 : 두 모집단 비율은 다르다. (양측검정) $p_1 \neq p_2$ -> $\theta = p_1 - p_2 \neq 0$
- 검정통계량 test statistic :

$$TS = \frac{\hat{\theta} - \theta}{s(\hat{\theta})} = \frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{(\hat{p}_1(1-\hat{p}_2))/(1/n + 1/m)}} \sim N(0,1),$$

통합 비율 추정 (귀무가설이 맞다면)

$$\hat{p} = \frac{\#of1's\_S1 + \#of1's\_S2}{n+m}$$

- 결론 : 검정통계량(TS) 값이 기각역(유의수준에 의해 결정)에 들어가면 귀무가설을 기각하고 그렇지 않으면 귀무가설 채택
- 대립가설이 양측이면 양쪽에 유의수준 α를 양쪽에 균등하게 배분하나 대립가설이 단측이면 대립가설의 끝 부분 한 쪽에 α를 모두 배분한다. - 통계소프트웨어는 항상 양측검정을 기준하여 출력하는 것이 원칙



5) 사례연구

세미소사와 맥과이어 홈런 경쟁으로 인하여 여성 팬이 증가하였다고 주장한다. 이를 알아보기 위하여 CNN/ USA이 다음 조사를 하였다. 1995년 1008명 여성 중 413이 팬이라고 대답했고, 홈런 경쟁이 있은 1998년에는 1082 여성 중 681명이 팬이라고 답하였다. 이를 이용하여 주장에 대해 답하시오.

(1) 연구문제 : 1995년에 비해 1998년 야구 여성 팬 비율이 높은가?

- ✳ 독립인 2집단(연도) 비율 차이 검정
- ✳ (데이터) 1995년 여성 1,008명 중 413 야구 팬이라고 응답, 1998년 여성 1,082명 중 681 야구 팬이라고 응답 (USA/CNN 조사)

(2) 통계적 가설

- ✳ 귀무가설 : 1995년, 1998년 여성 야구 팬 비율은 동일하다. $p_1 = p_2$
- ✳ 대립가설 : 1998년 여성 야구팬 비율이 1995년보다 높다. $p1 < p2$

(3) 추정값과 검정통계량 계산

연도	표본크기(n)	야구팬(x)	p_hat(x/n)	z-통계량	p-값
1995년	1008	413	0.4097	-10.0472	0.000
1998년	1082	681	0.6294		
		통합비율	52.3%		

통합 비율 : $\hat{p} = \dfrac{(413+681)}{(1008+1082)}$ =0.523 =(C2+C3)/(B2+B3)

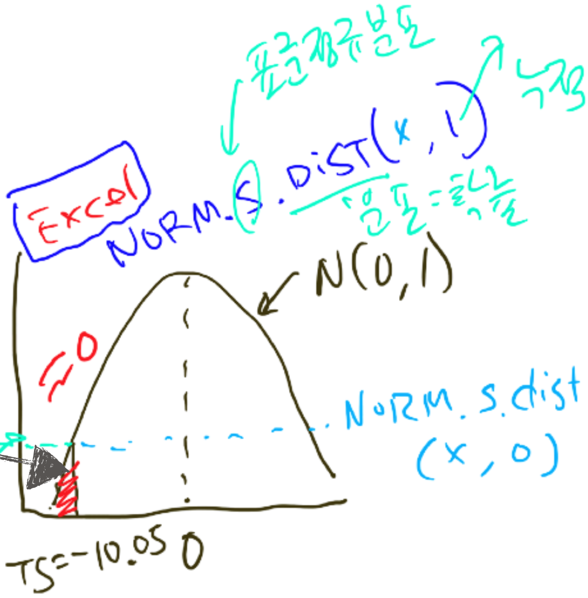
$$TS = \frac{0.4097 - 0.6294}{\sqrt{(0.523(1-0.523)(1/1008+1/1082)}} = -10.05 \sim N(0,1)$$

=(D2-D3)/SQRT(D4*(1-D4)*(1/B2+1/B3))

(결론)

유의확률 p-값 (붉은 부분 확률) = 0.0000 (매우 유의) <<< 유의수준 1% => 귀무가설 기각, 그러므로 1998년 여성 야구팬 비율은 1995년에 비해 유의적으로 높으므로 홈런 경쟁이 여성 팬을 증가시켰다고 할 수 있다.

=NORM.S.DIST(E2,1)



100(1-α)% 신뢰구간 계산하기

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}$$

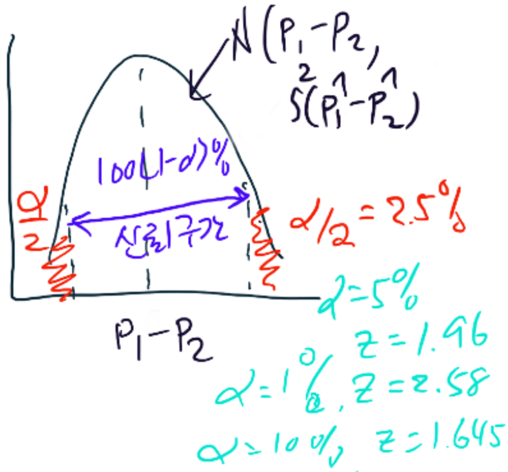
z-값 = D6

z-값	1.959964
하한	-0.2615
상한	-0.17784

=NORM.S.INV(0.975)

=(D2-D3)-D6*SQRT(D2*(1-D2)/B2+D3*(1-D3)/B3)

두 모집단 비율 차이 95% 신뢰구간 = (-26.2%, - 17.7%) - 95% 신뢰구간이 0%를 포함하고 있지 않으므로 5% 양측대립 가설검정을 기각 -> 양측 5% 가설검정 결과와 95% 신뢰구간 결과는 동일



연도	표본크기(n)	야구팬(x)	p_hat(x/n)	z-통계량	p-값
1995년	1008	413	0.4097	-10.0472	0.000
1998년	1082	681	0.6294		
		통합비율	52.3%		

6) 연습문제 : Keller 9판 12.5 예제문제

비누 회사에서 포장지 디자인을 2개 개발하였다. 고객들이 어느 디자인을 선호하는지 알아보기 위하여 자회사 비누 판매율이 비슷한 두 수퍼마켓을 임의 선택하여, 수퍼마켓에 각각 서로 다른 디자인의 비누를 진열하여 판매비율을 조사하였다. 이 회사 비누의 바 코드 “9077”이다. 데이터에는 각 슈퍼마켓에서 팔린 비누의 코드가 있다. package.csv

(1) 고객은 어느 디자인을 선호하는가? 유의수준 = 5%

(2) 슈퍼마켓 1의 디자인이 슈퍼마켓 2의 디자인보다 3% 이상 선호되고 있는가? 유의수준 5%

7) 사례연습문제 : 은행문제 내용

표본크기 : 남성CEO 1,050, 여성 CEO 115명 중 대출 승인 받은 CEO 데이터 BANK.csv

은행은 승인율 측면에서 여성CEO을 차별하는지 분석하시오. 유의수준 5%

3. 짝진 2집단 비율 차이 검정 McNemar 검정

1) 시나리오 (연구문제)

동일한 개체로부터 이진형(성공, 실패) 변수를 서로 다른 기간(before - after)에 측정하여 프로그램 효과가 있는지 알아보는 방법이다.

Bland (2000) 1319명 어린이, 12살에 독감에 걸릴 가능성은 나이가 14살이 되면 높아지는지 낮아지는지 알아보기 위하여 조사한 결과이다. https://www.medcalc.org/manual/mcnemartest2.php

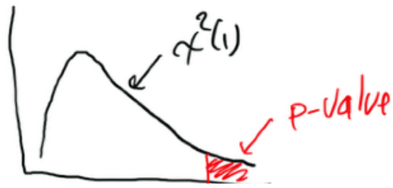
Severe colds at age 12	Severe colds at age 14		Total
	Yes	No	
Yes	212 A	144 B	356
No	256 C	707 D	963
Total	468	851	1319

2) 모수와 점추정치(MVUE)

- 모수 : p_1, p_2 (예) p_C=12살 독감 걸릴 비율, p_B = 14살 독감 걸릴 비율
- 점추정치 (MVUE) : $\hat{p}_A = (356/1319) = 0.354, \hat{p}_D = 144/1319 = 0.27$
- 총 표본크기 : $n = A + B + C + D$

3) 검정통계량, 가설검정

- 귀무가설 : $p_1 = p_2$, 대립가설 : $p_1 \neq p_2$
- 검정통계량 : $TS = \dfrac{(B-C)^2}{B+C} \sim asymtotic\chi^2(1)$
  단, [근사 조건] B, C > 20, (B+C) >25,
- (예) $TS = \dfrac{(144-256)^2}{144+256}$=31.36
- A, D를 사용해도 동일한 검정 통계량을 얻는다.
- 유의확률 < 0.001 - 매우 유의



=CHISQ.DIST.RT(31.36,1)

D	E
0.00000002	

4) 결론

- 귀무가설이 기각되어 나이가 올라가면 독감 걸릴 가능성이 높아진다. 12살 독감 걸릴 확률은 35.4%, 14살에는 27%로 청소년의 경우 나이가 들면 유의적으로 낮아진다.

표본의 크기가 충분히 크지 않으면(근사 조건 위반) 검정통계량이 카이제곱분포에 근사하지 않는다. 이런 경우 Fisher's Exact 검정을 사용한다. 초기하(Hyper Geometric) 분포 이용.

(1) 유의확률 : $$\sum_{a}^{n} \frac{(a+b)Ca \times (c+d)Cc}{nC(a+c)} \sim HG(n, a+b, b+d)$$

(2) 근사 유의확률(초기하분포의 이항분포 근사 활용) : $$2\sum_{i=a}^{n} nCi0.5^{i} \times (1-0.5)^{(}n-i)$$

- 12살 35.4%에서 14살이 되면 27%로 낮아진다. (대립가설이 양측이나, 귀무가설이 기각되면 단측 대립가설도 선택되기 때문에 이런 해석 가능하다)

집단	독감 확률	카이제곱-통계량	유의확률
12살	35.40%	31.3	<0.001
14살	27%		

	Children's seat belt status		
	Unbelted	Belted	Total
Drivers' seat belt status:			
Unbelted	116	19	135
Belted	73	262	335
Total	189	281	470

5) 사례연구

부모의 안전벨트 착용여부가 아이(4~14세)들의 안전벨트 착용여부에 영향을 미치는지 알아보기 위하여 부모-어린이 함께 탄 차 470 차량을 조사한 결과이다. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC107902/

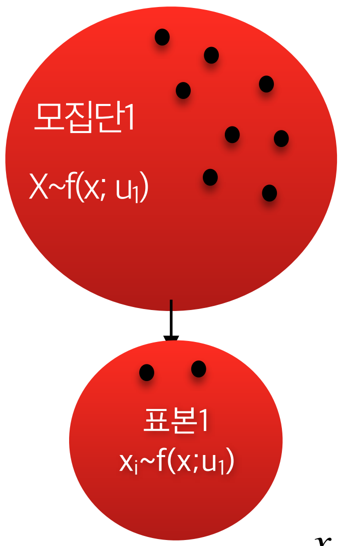
	After: present	**After:** absent	Row total
Before: present	101	121	222
Before: absent	59	33	92
Column total	160	154	314

6) 사례연구2

H병을 앓고 있는 환자 중 S 증상에 효과가 있는 K-신약이 개발되었다고 하자. H병 환자 314명 중 H 증상여부를 조사하고 K-신약 투여 후 H 증상여부를 다시 조사하여 얻은 결과이다. K-신약의 효과를 분석하시오. https://en.wikipedia.org/wiki/McNemar%27s_test

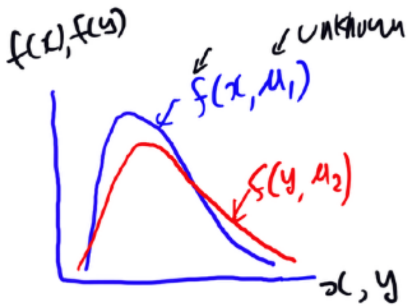
4. 독립 2집단 모평균 차이 검정

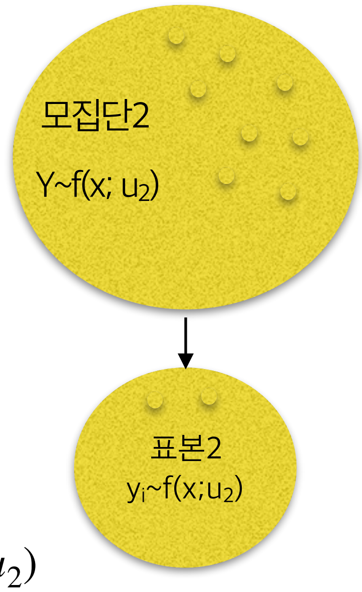
1) 시나리오 (연구문제)



두 모집단 서로 독립

개체에 의해 관측되는 값은 연속형분포





$x_i \sim f(x;\mu_1)$ $\quad$ $y_i \sim f(y;\mu_2)$

$x_1 = 11.2, x_2 = 15.3, \ldots, x_n = 11.1$ **데이터** $y_1 = 12.1, y_2 = 12.3, \ldots, y_m = 11.11$

확률표본 (iid) = 서로 독립이고 동일분포에서 추출

2) 모수와 점추정치(MVUE)

- 모수 : μ_1, μ_2 모집단1, 2 평균
- 점추정치 (MVUE) : $\hat{\mu}_1 = \bar{x} = \dfrac{\sum x_i}{n}$, $\hat{\mu}_2 = \bar{y} = \dfrac{\sum y_i}{m}$
- 샘플링 분포 (대표본) $n, m > 20 \sim 30$

  * $\hat{\mu}_1 \sim N(\mu_1, s(\hat{\mu}_1)^2)$ (중심극한정리- Large sample theory), $s(\hat{\mu}_1)^2 = \dfrac{\hat{\sigma}_1^2}{n}$
  * $\hat{\mu}_2 \sim N(\mu_2, s(\hat{\mu}_2)^2)$: 동일한 이유, $s(\hat{\mu}_2)^2 = \dfrac{\hat{\sigma}_2^2}{m}$

s_1^2, s_2^2은 각각 표본1, 표본2의 표본분산, $\hat{\sigma}_1^2 = s_1^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$, $\hat{\sigma}_2^2 = \dfrac{\sum (y_i - \bar{y})^2}{m-1}$

- 정규분포의 가법성에 의해 $\hat{\mu}_1 - \hat{\mu}_2 \sim ???(\mu_1 - \mu_2, \frac{s_1^2}{n} + \frac{s_2^2}{m})$
- 만약 모집단의 분산(σ_1^2, σ_2^2)을 알고 있다면 $\hat{\mu}_1 - \hat{\mu}_2 \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m})$
- 일반적으로 모집단의 분산은 모르므로
- W.S. Gosset이 발견한 t-분포 정규분포(모수 2개, μ, σ)와 동일하며 자유도 1개 모수를 가지는 분포

두 모집단 분산 차이 검정

- 귀무가설 : $\sigma_1^2 = \sigma_2^2$
- 대립가설 : $\sigma_1^2 \neq \sigma_2^2$
- 검정통계량 : $\frac{(n-1)s_1^2}{\sigma_1^2} \sim \chi^2(n-1)$ $\quad \frac{(m-1)s_2^2}{\sigma_2^2} \sim \chi^2(m-1)$

$$TS = \frac{max(s_1^2, s_2^2)}{min(s_1^2, s_2^2)} \sim F(df_1, df_2)$$

- 만약, 등분산이 가정되면 $\frac{(\hat{\mu}_1 - \hat{\mu}_2) - (\mu_1 - \mu_2)}{s_p^2(\frac{1}{n} + \frac{1}{m})} \sim t(df = n + m - 2)$, where

  s_p^2 = 통합분산 pooled variance

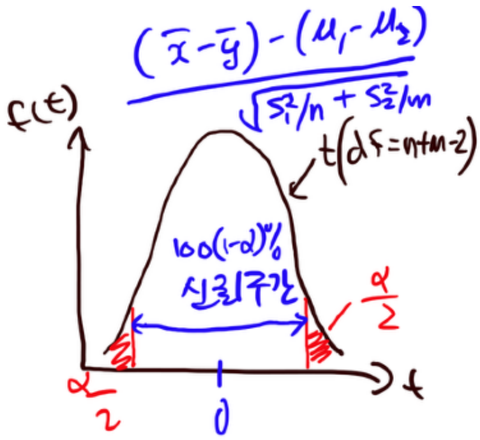
- 등분산이 무너지면 $\frac{(\hat{\mu}_1 - \hat{\mu}_2) - (\mu_1 - \mu_2)}{(\frac{s_1^2}{n} + \frac{s_2^2}{m})} \sim t(Welch - Satterthwaite's)$
- $$(W - S)df = \frac{(\frac{s_1^2}{n-1} + \frac{s_2^2}{m-1})^2}{\frac{{s_1^2}^2}{n^2(n-1)} + \frac{{s_2^2}^2}{m^2(m-1)}}$$

3) 신뢰구간

$$(\bar{x}-\bar{y}) \pm t(1-\alpha/2, df=n+m-2)\sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}}$$

(대표본)

소표본인 경우는 모집단의 정규분포 가정이 필요하다. 소표본, 모집단 정규분포 가정이 무너지면 비모수검정 방법 (Non-parametric 검정, Wilcoxon Rank Sum Test) 적용



4) 통계적 가설검정

- 귀무가설 : 두 모집단 평균은 동일하다. $\mu_1 = \mu_2$ -> $\mu_1 - \mu_2 = 0$
- 대립가설 : 두 모집단 비율은 다르다. (양측검정) $\mu_1 \neq \mu_2$
- 검정통계량 test statistic

  ✳ (등분산) $TS = \dfrac{\bar{x}-\bar{y}-(\mu_1-\mu_2)}{s_p^2(\frac{1}{n}+\frac{1}{m})} \sim t(n+m-2)$

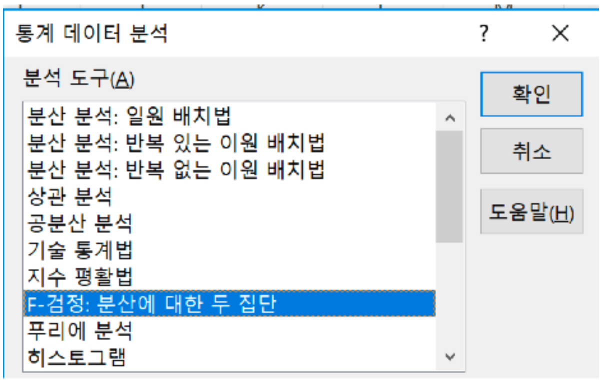
  ✳ (이분산) $TS = \dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{(\frac{s_1^2}{n}+\frac{s_2^2}{m})} \sim t(Welch-Satterthwaite's)$

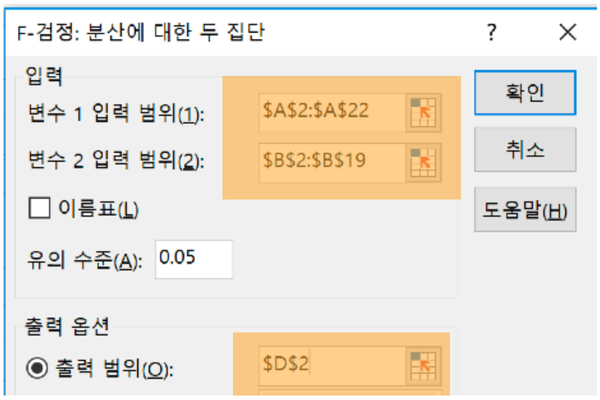
- 결론 : 검정통계량(TS) 값이 기각역(유의수준에 의해 결정)에 들어가면 귀무가설을 기각하고 그렇지 않으면 귀무가설 채택
- 대립가설이 양측이면 양쪽에 유의수준 α를 양쪽에 균등하게 배분하나 대립가설이 단측이면 대립가설의 끝 부분 한 쪽에 α를 모두 배분한다. - 통계소프트웨어는 항상 양측검정을 기준하여 출력하는 것이 원칙

5) 사례연구 : Keller “Managerial Statistics” 9th edition Milk.csv

커피로 인해 우유 소비가 많은 시애틀(n=21)이 다른 대도시(아틀란타,n=18)에 비해 높은지 알아보기 위하여 다음 조사를 하 였다. 각 도시의 K-mart에서 팔리는 우유 값을 관찰하였다.

	A	B
1	Seattle	Atlanta
2	2.55	2.25
3	2.67	2.3
4	2.5	2.49
5	2.61	2.41
6	2.43	2.39
7	2.36	2.26
8	2.5	2.4
9	2.36	2.33
10	2.54	2.29
11	2.54	2.48
12	2.8	2.59
13	2.61	2.38
14	2.56	2.39
15	2.64	2.4
16	2.43	2.23
17	2.43	2.29
18	2.38	2.53
19	2.49	2.45
20	2.57	
21	2.71	
22	2.27	

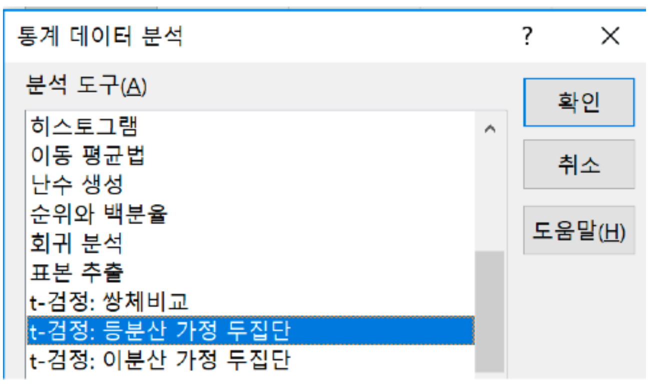




F-검정: 분산에 대한 두 집단

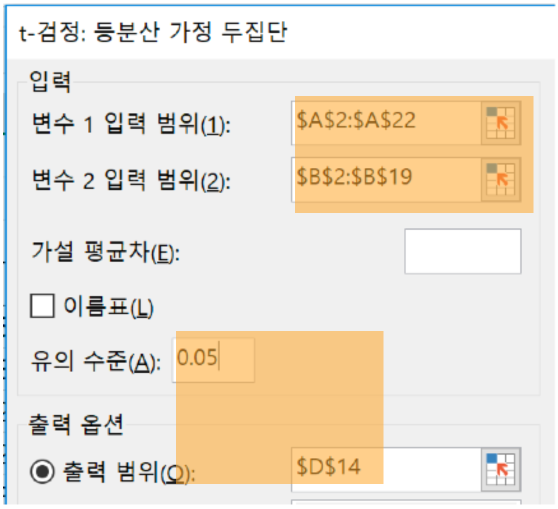
	변수 1	변수 2
평균	2.52142857	2.38111111
분산	0.01663286	0.01025752
관측수	21	18
자유도	20	17
F 비	1.6215287	
P(F<=f) 단측 검정	0.1590983	
F 기각치: 단측 검정	2.23035428	

유의확률이 0.15로 유의수준 5%보다 높아 귀무가설(등분산 가정 만족)을 기각할 수 없어 "등분산 가정"이 만족된다. 표본1(시애틀) 분산 = 0.017, 표본2(아틀란타) 분산=0.0103으로 유의적 차이가 없다.



등분산 가정이 만족되는 경우 독립인 두 모집간 차이 검정

대립가설 : 시애틀 우유 값이 아틀란타 우유 값보다 높다.



t-검정: 등분산 가정 두 집단		
	변수 1	변수 2
평균	2.52142857	2.38111111
분산	0.01663286	0.01025752
관측수	21	18
공동(Pooled) 분산	0.01370365	
가설 평균차	0	
자유도	37	
t 통계량	3.73170443	
P(T<=t) 단측 검정	0.0003184	
t 기각치 단측 검정	1.68709362	
P(T<=t) 양측 검정	0.0006368	
t 기각치 양측 검정	2.02619246	

단측검정이므로 유의확률은 <0.001이므로 귀무가설이 기각된다. 시애틀 우유 평균 가격은 2.52$, 아틀란타는 2.38$로 커피 소비가 많은 시애틀 우유 가격이 유의적으로 높다.

도시	평균	표준편차	검정통계량	유의확률
시애틀	2.52	0.129	3.73	<0.001
아틀란타	2.38	0.101		

6) 사례연구 2 Keller “Managerial Statistics” 9th edition

표본크기 : 남성CEO 1,050, 여성 CEO 115명 중 대출 승인 받은 CEO 데이터 BANK.csv

은행은 이자율 측면에서 여성CEO을 차별하는지 분석하시오. 유의수준 5%

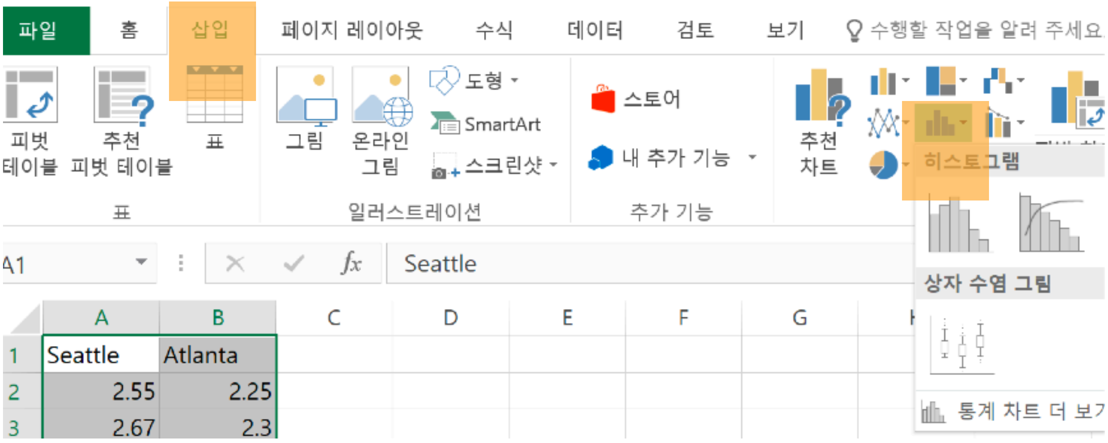
7) 사례연구 3 Keller “Managerial Statistics” 9th edition

직접(direct) 투자하는 수익률과 투자전문가(broker) 활용하는 수익율의 차이가 있는지 유의수준 5% 검정하시오. FUNDS.csv

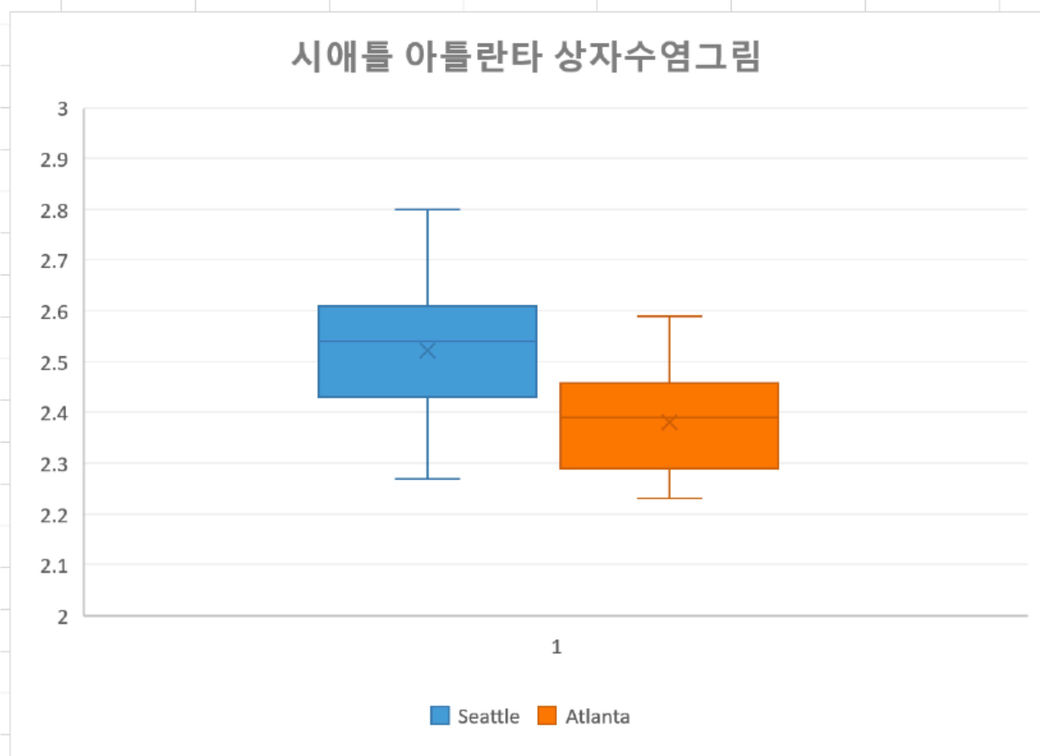
8) 사례연구 4 Keller “Managerial Statistics” 9th edition

MBA 전공 재무, 마케팅 연봉을 조사한 자료이다. 유의수준 5%에서 재무전공, 마케팅 전공의 연봉 차이가 있는지 검정하시오. MBA.csv

| 엑셀 상자-수염 그리기 | 버전 2016 이상 가능



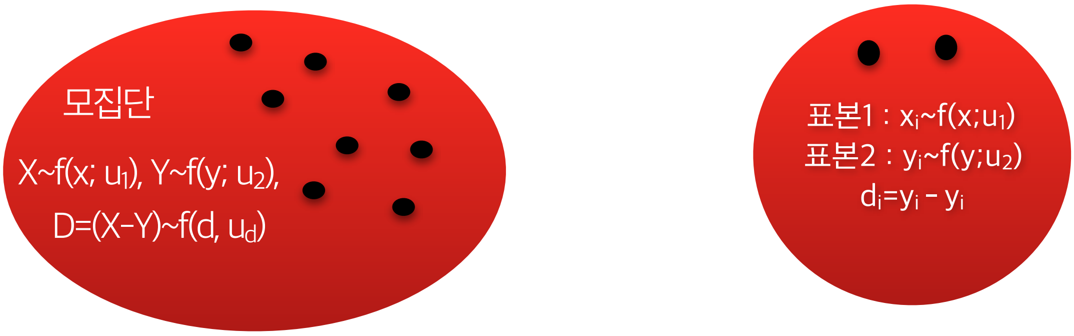
삽입 -> 통계 차트 메뉴 -> 작성 후 차트도구->디자인 선택



이상치는 존재하지 않으며 시애틀의 우유가격 평균, 중앙값은 아틀란타보다 높다. 통계적 가설 검정 결과와 일치

5. 짝진 2집단 모평균 차이 검정

1) 시나리오 (연구문제)



$$d_i \sim f(d;\mu_d)$$

1표본 데이터 $d_1 = -1.1, d_2 = 0.3, \ldots, d_n = -0.11$

표본크기 n인 1집단 평균 분석과 동일

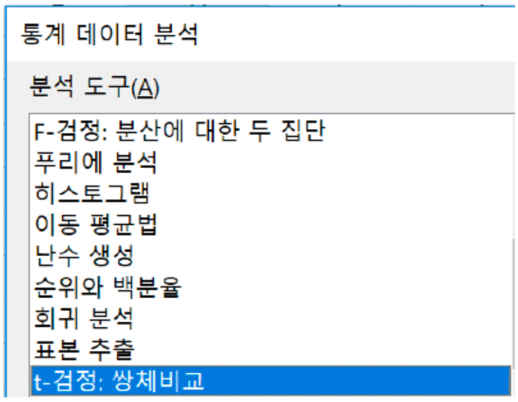
2) 모수와 점추정치(MVUE)

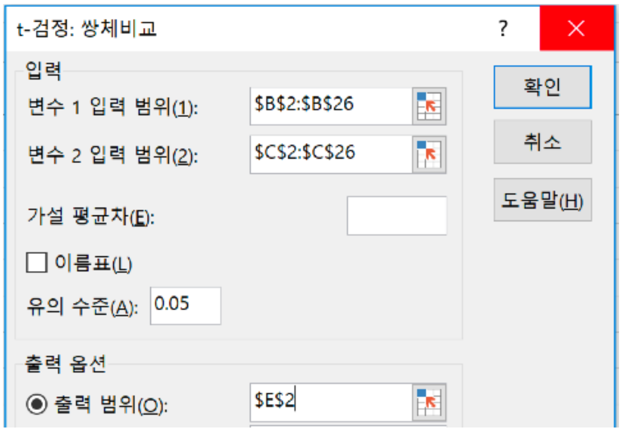
- 모수 : μ_d 모집단 차이 평균
- 점추정치 (MVUE) : $\hat{\mu}_d = \bar{d} = \dfrac{\sum d_i}{n}$
- 신뢰구간 및 가설 검정은 1집단 모평균 검정과 동일함 : 귀무가설 $H_0 : \mu_d = 0$

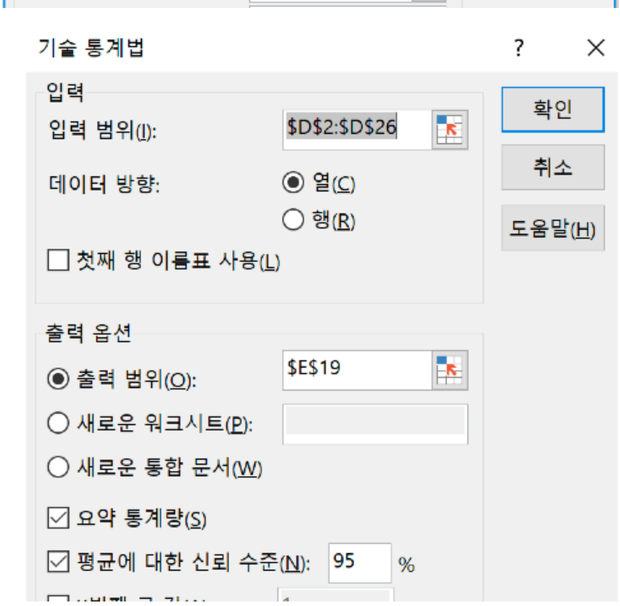
3) 사례연구 Keller "Managerial Statistics" 9th edition

MBA 전공 재무, 마케팅 연봉을 조사한 자료이다. 성적(GPA)에 따른 차이가 있을 가능성을 고려하여 (4, 3.92)=1그룹, (3.92, 3,84)=2그룹2, ..., 총 25개 그룹에서 한 명씩 임의 추출하여 연봉을 조사하였다. 유의수준 5%에서 재무전공, 마케팅 전공의 연봉 차이가 있는지 검정하시오. MBA2.csv

Group	Finance	Marketing	D
1	95171	89329	5842
2	88009	92705	-4696
3	98089	99205	-1116







t-검정: 쌍체 비교

	변수 1	변수 2
평균	65438.2	60373.68
분산	444981809.5	469441784.6
관측수	25	25
피어슨 상관 계수	0.952024734	
가설 평균차	0	
자유도	24	
t 통계량	3.809688414	
P(T<=t) 단측 검정	0.000425543	
t 기각치 단측 검정	1.71088208	
P(T<=t) 양측 검정	0.000851086	
t 기각치 양측 검정	2.063898562	

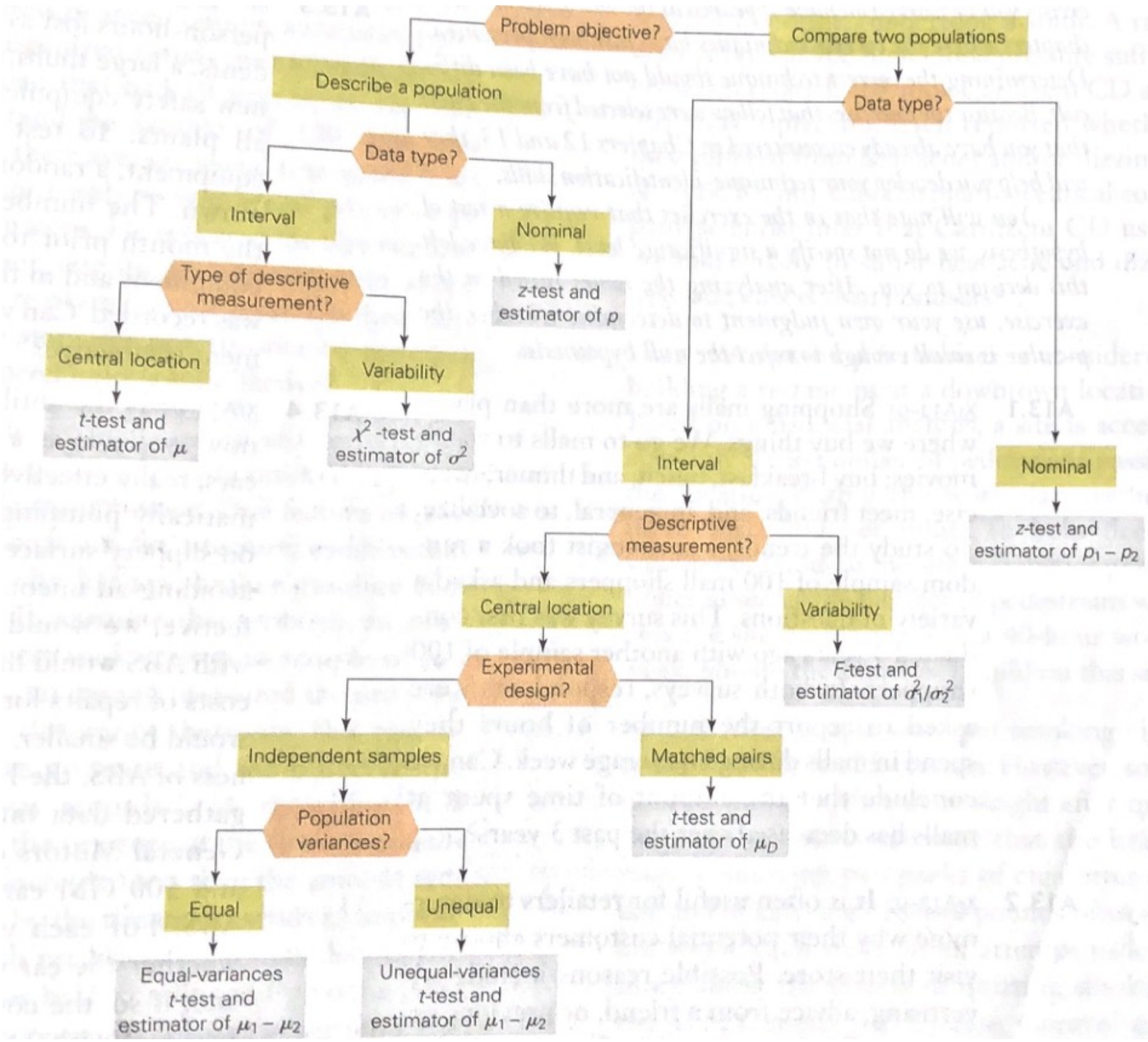
Column1	
평균	5064.52
표준 오차	1329.379059
중앙값	3285
최빈값	#N/A
[illegible]	[illegible]
범위	23533
최소값	-5721
최대값	17812
합	126613
관측수	25
신뢰 수준(95.0%)	2743.703529

전공	평균	표준편차	검정통계량	유의확률
재무	65,438	21,667	3.81	<0.001
마케팅	60,374	21,095		

변수1=재무, 변수2=마케팅 -> 재무 전공 평균초봉 = 65,438$, 마케팅 전공은 60,374$로 재무 전공자들의 초봉이 유의적으로 높다. (유의확률 0.0004 < 유의수준 5%)

차이에 의한 1집단 검정 결과도 동일하다. (5,065 ± 2,744)는 0을 포함하고 있지 않으므로 재무 전공의 초봉이 높다는 이전 결과와 동일하다.

6. 사례연구



사례연구 1 : Keller's 9 edition Emergecy.csv

소방차에 응급장비를 장착하는 문제에 대한 논쟁이 벌어져, (1) 콜센터 전화를 받고 소방차가 응급차보다 1분 먼저 도착하고 (2) 8분 이내 도착하는 비율이 소방차가 높다면 응급장비를 소방차에 장착하기로 시의회에서 결정하였다. (Cambridge, Waterloo, Kitchner) 3개 지역 중 장착 가능한 도시?

사례연구 2 : Keller's 9 edition HRT.csv

인지능력 장애를 가진 75세 여성의 경우 호르몬 대체 치료방법(HRT)의 효과가 있는지 알아보기 위한 조사를 실시하였다. HRT 방법으로 치료하고 있는 58명(user=1), 그렇지 않은 여성 47명(user=2) 임의 추출하여 캘리포니아 언어 습득 점수(100점 만점)와 로직 기억력 점수(100점 만점)를 측정한 결과이다. 효과를 분석하시오.