3. Z분포에 의한 검정

3-1. Z검정

① Z검정(Ztest)은 Z분포(표준정규분포)에 의하여 가설을 검정하는 통계적 방법으로 다음의 네 가지 조건을 충족시킬 때 사용한다.
- ㉠ 종속변수가 양적변수일 때
- ㉡ 모집단의 분산을 알고 있을 때
- ㉢ 모집단의 분포가 정규분포일 때(정규분포 가정(normality assumption)을 충족)
- ㉣ 두 모집단을 비교할 경우 두 모집단의 분산이 같을 때(등분산 가정(equal variance assumption)이 충족)

② Z검정은 집단비교를 위한 추측통계의 기본 개념과 절차를 설명하는 데 기본이 되는 검정 방법임에 비하여 실제 연구에서 사용되는 예를 흔하지 않다.
- ㉠ Z검정을 사용하기 위해서는 모집단의 분산 또는 표준편차를 알아야 하는데 실제 연구상황에서 모집단의 분산을 아는 경우는 매우 드물다.
- ㉡ 표본의 크기가 클 때 표본으로 얻은 표준편차나 분산이 모집단의 표준편차나 분산과 유사하므로 Z검정을 쓸 수 있다고 주장하는 경우도 있으나, 표본의 크기가 크다는 기준이 모호하므로 Z검정을 사용하는 것은 무리이다.
- ㉢ 모집단 분포의 정규성 여부는 이론적 배경이나 경험적 배경에 의하여 규명될 수 있으나, 만약 모집단의 분포가 정규분포를 유지하지 못한다면 비모수 통계(nonparametric statistics)를 사용하여야 한다.
- ㉣ 등분산 가정 여부는 Z검정, t검정, F검정의 필수적인 요소이며, 만약 모집단분포의 흩어진 정도가 모집단마다 다를 때 각기 다른 집단들을 비교한다면 잘못된 결론을 유도할 수 있다.
- ㉤ 만약 두 집단 비교시 두 모집단의 평균은 같고 분산의 크기가 다르면, 이 두 모집단이 같은 집단이라고 말할 수 없으며 비교가 쉽지 않으므로, 이런 경우는 비모수통계를 사용하여야 한다.

③ 검사의 목적에 따라 단일표본 Z검정, 두 독립표본 Z검정, 두 종속표본 Z검정, 비율검정이 있으며, 상관계수의 유의도 검정도 시행 가능하다.

3-2. 단일표본 Z검정

① 단일표본 Z검정(one-samlple Ztest)이란 한 모집단의 속성을 알기 위하여 모집단을 대표하도록 추출된 한 표본의 통계값인 평균과 연구자가 이론적 혹은 경험적 배경에서 얻은 특정값을 비교하는 통계적 방법이다.

② Z통계값에 의한 검정

㉠ 점추정에 의한 방법이며, 표본으로부터 얻은 통계값이 표집분포의 어느 곳에 위치하느냐에 의하여 가설을 검정한다.

㉡ 일반적으로 연구보고에서는 점추정에 의한 통계치를 제공하여 귀무가설의 기각 여부를 검정한다.

③ 신뢰구간에 의한 검정

㉠ 구간추정에 의한 방법이며, 귀무가설 하에서 모수치의 범위에 의해 가설을 검정한다.

㉡ 구간추정 방법은 귀무가설 하에서 모집단 평균의 범위를 알려준다.

[예] A 교육구 고등학교 3학년생들의 학업적성 점수가 전국 고등학교 3학년 학생들의 평균과 같은지를 검정하고자 한다. 이 때, 전국 고등학교 3학년생의 학업적성 평균점수가 300점이고 표준편차는 20점임을 알고 있다. A 교육구 고등학교 3학년생들 중 100명을 무선표집하여 학업적성 검사를 실시하였더니 평균이 306점이 나왔다. 유의수준 0.05에서 A 교육구의 모집단의 학업적성 능력이 300점인지 검정하여라.

[풀이]

① Z통계값에 의한 검정

㉠ 귀무가설과 대립가설을 세운다.

H_0 : A 교육구 고교 3학년생들의 학업적성 평균점수는 300점이다. $\mu = 300$

H_1 : A 교육구 고교 3학년생들의 학업적성 평균점수는 300점이 아니다. $\mu \neq 300$

㉡ 귀무가설 하에서 표집분포를 그린다. 이 때, 중심극한정리에 의하여

$$\text{표집분포의 평균} = 300,\ \text{표준오차} = \frac{20}{\sqrt{100}} = 2$$

㉢ 100명 고등학교 3학년생 학업적성검사 점수의 Z 통계값을 계산하고, Z분포에 의하여 유의수준 0.05에 해당하는 기각역과 통계값을 그린다.

$$Z = \frac{306 - 300}{2} = 3$$

㉣ Z통계값은 3이고 1.96보다 크므로 귀무가설을 기각하게 된다. 즉, 유의수준 0.05에서 A 교육구 고등학교 3학년생들의 학업적성 점수는 300점이 아니다.

② 신뢰구간에 의한 검정

㉠ A 교육구 고교 3학년생들의 학업적성 평균점수가 300점인지를 유의수준 0.05에서 검정할 때 100명의 평균점수가 $300 \pm 1.96SE$ 사이에 있으면 귀무가설을 채택하고 그 밖에 있으면 귀무가설을 기각한다.

㉡ 채택역: $300 - (1.96)(2) \leq \overline{X} \leq 300 + (1.97)(2)$

$$296.08 \leq \overline{X} \leq 303.92$$

㉢ A 교육구 고교 3학년생들의 학업적성 평균점수가 306점이고 이는 채택역 밖에 있으므로 귀무가설을 기각한다. 즉, 유의수준 0.05에서 A 교육구 고등학교 3학년생들의 학업적성 점수는 300점이 아니다.

3-3. 두 표본 Z검정

3-3-1. 두 독립표본 Z검정

① 알지 못하는 두 모집단의 속성, 즉 평균을 비교하기 위하여 각기 모집단을 대표하도록 추출된 독립적인 두 표본을 가지고 두 모집단의 유사성을 검정하는 통계적 방법이다.

② 두 모집단의 분산이 이론적 배경이나 경험적 배경에 의하여 같음을 알고 있어야 한다.

③ 단, Z검정을 위한 네 가지 기본 가정을 충족시켜야 한다.

④ 전국 중학교 3학년 남녀학생의 수학능력 비교 ; 이때, 모집단을 대표하기 위하여 추출된 중학교 3학년 남학생의 표본과 중학교 3학년 여학생의 표본이 상호 독립적이어야 한다.

⑤ 모평균의 차 $(\mu_1 - \mu_2)$를 위해 표본으로부터 통계량 $(\overline{X_1} - \overline{X_2})$를 구하여 두 표본 평균 간에 차이가 있는지를 검정한다.

㉠ 표준오차= $\sqrt{\dfrac{\sigma_1{}^2}{n_1} + \dfrac{\sigma_2{}^2}{n_2}}$

㉡ Z통계값 $Z = \dfrac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1{}^2}{n_1} + \dfrac{\sigma_2{}^2}{n_2}}}$

[예] 전국 30세 남녀 100명을 각기 무작위 추출하여 체중을 측정하였더니 남자의 평균 체중이 68kg 이었고, 여자의 평균 체중은 60kg이었다. 또한 이론적 배경에 의하여 30세 남자모집단의 체중의 표준편차는 10kg이고 여자 체중의 표준편차는 9kg임을 알고 있다. 30세 성인 남녀의 체중이 차이가 있는지의 여부를 유의수준 0.05에서 검정하여라.

[풀이]

㉠ 귀무가설과 대립가설을 세운다.

H_0 : 30세 성인 남(1)녀(2)체중에 차이가 없다. $\mu_1 = \mu_2$ 또는 $\mu_1 - \mu_2 = 0$

H_1 : 30세 성인 남(1)녀(2)체중에 차이가 있다. $\mu_1 \neq \mu_2$ 또는 $\mu_1 - \mu_2 \neq 0$

㉡ 통계량 $\overline{X_1} - \overline{X_2} = 68 - 60 = 8$, 표준오차 $\sqrt{\dfrac{10^2}{100} + \dfrac{9^2}{100}} = 1.345$

에 대하여 남녀 30세 성인의 체중비교를 위한 Z 통계값은

$$Z = \frac{8-0}{1.345} = 5.948$$

㉢ 유의수준 0.05에서의 기각값 ± 1.96보다 크므로 귀무가설을 기각한다.

㉣ 따라서 '유의수준 0.05에서 30세 성인남녀의 체중에 유의한 차이가 있다'

⑥ 두 모집단의 분산을 모른다면

$$\sqrt{\frac{\sigma_1{}^2}{n_1} + \frac{\sigma_2{}^2}{n_2}} \text{ 에 대한 추정량 } \Rightarrow SE = \sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}$$

『예』 차량의 소유가 학생들의 학업 성취도에 영향을 미치는지의 여부를 알기 위해 학생들 중, 임의로 남학생 100명씩 두 그룹의 표본을 추출하였다.

차량을 소유하지 않은 남학생 집단	차량을 소유한 남학생 집단
평균 2.70, 분산 0.36	평균 2.54, 분산 0.40

차량의 소유가 학생들의 학업성취도에 영향을 미친다는데 충분한 근거가 될 수 있는지를 유의수준 0.05에서 가설검정 하여라.

[풀이]

차량을 소유하지 않은 남학생들의 학점의 평균 μ_1

차량을 소유한 남학생들의 학점의 평균 μ_2

㉠ 귀무가설과 대립가설을 세운다.

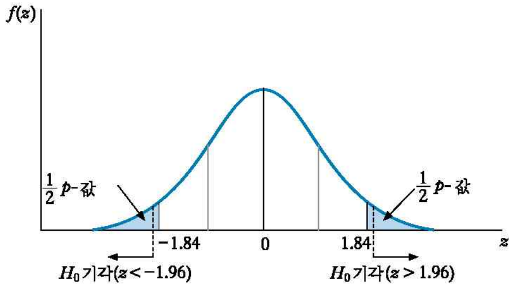
H_0 : 이들 평균 간에 차이가 없다. $\mu_1 = \mu_2$ 또는 $\mu_1 - \mu_2 = 0$

H_1 : 이들 평균 간에 차이가 있다. $\mu_1 \neq \mu_2$ 또는 $\mu_1 - \mu_2 \neq 0$

㉡ Z 통계값

$$Z = \frac{(2.70 - 2.54) - 0}{\sqrt{\frac{0.36}{100} + \frac{0.40}{100}}} = 1.84$$

㉢ 두 그룹간의 학업 성취도의 평균이 차이가 난다고 할 만한 충분한 근거가 없다.



『예』 위 문제에서 차량을 소유하고 있는 남학생들과 그렇지 않은 남학생들의 학업 성취도의 평균차에 대한 95% 신뢰구간을 구하여라. 신뢰구간을 이용하여 두 학생 그룹 간의 모평균 차이가 있다고 할 수 있겠는가?

[풀이]

㉠ 두 모평균 차에 대한 신뢰구간은

$$(2.70 - 2.54) \pm 1.96\sqrt{\frac{0.36}{100} + \frac{0.40}{100}} = 0.16 \pm 0.17$$

즉, $-0.01 < \mu_1 - \mu_2 < 0.33$이 된다.

㉡ 이 구간은 두 모평균 차가 가질 수 있는 값들의 범위이다. 두 모평균의 차를 0으로 가정했고, 신뢰구간이 이 값을 포함하고 있기 때문에 H_0를 기각해서는 안 된다. 그리고 신뢰구간은 0을 포함하고 있으므로 차량을 소유하고 있는 남학생들과 그렇지 않은 남학생들의 학업 성취도의 평균 차이가 있다고 할 수 없다.

3-3-2. 두 종속표본 Z검정

① 알지 못하는 두 모집단의 속성을 비교하는 통계적 검정 방법이며, 추출된 두 표본이 독립적이지 않을 때 사용된다.

② 기혼남녀의 체중을 비교하기 위하여 결혼한 부부 남자집단과 여자집단의 체중을 추정할 때, 부부를 추출하여 기혼남자와 기혼여자의 표본을 설정하여 두 표본의 비교로 모집단의 속성을 비교할 수 있다. ; 집단비교를 위하여 추출된 부분의 체중은 관계가 있으며 기혼남자의 표본과 기혼여자의 표본은 상호독립적이 아니라고 가정되어야 한다.

③ 두 독립표본 Z검정과 기본적인 개념과 절차는 동일하다. 다만 표준오차를 계산하는 공식이 다르다. 즉, 두 표본이 상호독립적이면 두 표본평균의 상관이 0이었으나 두 표본이 상호독립적이 아니기 때문에 두 표본이 상관이 있어 다음의 공식을 사용한다.

$$\text{표준오차}=\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}-2\rho\frac{\sigma_1}{\sqrt{n_1}}\cdot\frac{\sigma_2}{\sqrt{n_2}}}$$

④ 두 모집단의 상관계수인 ρ를 이론적 혹은 경험적으로 알 수 없을 때에는 두 표본으로 얻은 상관계수 r로 대치할 수 있다.

$$\text{표준오차}=\sqrt{\frac{{\sigma_1}^2}{n_1}+\frac{{\sigma_2}^2}{n_2}-2r\frac{\sigma_1}{\sqrt{n_1}}\cdot\frac{\sigma_2}{\sqrt{n_2}}}$$