3-4. 비율(proportion)검정

3-4-1. 단일표본 비율검정

① 모집단을 대표하는 표본을 추출하여 어떤 비율이 통계적으로 유의한지를 검정한다.

〚예〛 신생아의 남자비율이 0.6이라면 통계적으로 유의한지를 알아봄

〚예〛 국가 정책을 결정하는데 있어 국민들의 찬성비율이 통계적으로 유의한지를 검정

② 단일표본 비율을 위한 Z검정의 Z통계값 $Z_{\hat{p}} = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$

〚예〛 중학교 무시험 진학에 대한 중학교 교사들의 찬성비율이 0.5이었다. 올해도 중학교 무시험 진학에 대한 찬성비율이 0.5인지를 유의수준 0.01에서 검정하여라. 단, 서울시 중학교에 근무하는 교사 300명을 무선추출하여 중학교 무시험 진학에 대한 찬성을 물은 결과 올해는 90명이 찬성하였다.

[풀이]

㉠ 귀무가설과 대립가설을 세운다.

H_0 : $P = 0.5$

H_1 : $P \neq 0.5$

㉡ 표집분포의 비율=0.5, 표준오차=$\dfrac{(0.5)(0.5)}{300}$. 따라서 Z통계값은

$$Z = \frac{0.3 - 0.5}{\sqrt{\dfrac{(0.5)(0.5)}{300}}} = -6.928$$

㉢ 유의수준 0.01에서 양방적 검정이므로 기각값은 ± 2.58이고 Z통계값 -6.928이 -2.58보다 작으므로 귀무가설을 기각한다.

㉣ 그러므로 '유의수준 0.01에서 올해 중학교 교사들이 무시험 진학에 찬성하는 비율이 0.5가 아니다' 라고 결론을 내릴 수 있다.

『예』 나이와 상관없이 미국성인의 20%는 일주일에 적어도 2번 이상 운동을 하고 있다. A지방의 40세 이상의 성인 100명을 대상으로 일주일에 적어도 2번 이상 운동을 하는 사람이 15명으로 조사됐다. 이 자료를 통해 40세 이상의 성인들 중 운동하는 성인의 비율이 20%보다 적다고 할 수 있을까? $p-$값을 계산하고, 이를 이용하여 적당한 결론을 내리시오.

[풀이]

㉠ 귀무가설과 대립가설을 세운다.

$H_0 : p = 0.2$

$H_1 : p < 0.2$

㉡ H_0이 참이라고 가정하여 시작한다. 즉, 모비율 p의 참값이 $p_0 = 0.2$일 때, $\hat{p} = X/n$은 근사적으로 평균이 p_0이고 표준오차가 $\sqrt{\dfrac{p_0 q_0}{n}}$ 인 정규분포를 따를 것이다.

따라서 $\hat{p}$의 관측값은 $\hat{p} = \dfrac{X}{n} = \dfrac{15}{100} = 0.15$이고, 검정통계량은

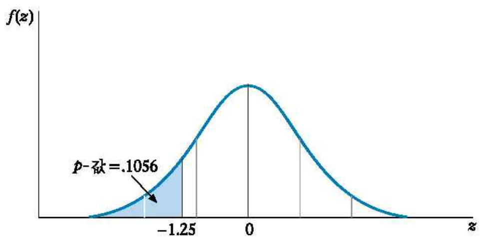
$$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}} = \frac{0.15 - 0.20}{\sqrt{\dfrac{(0.20)(0.80)}{100}}} = -1.25$$

이다.

㉢ 이 검정의 $p-$값은 표준정규분포곡선에서 $z = -1.25$의 왼쪽 꼬리의 넓이가 된다. 그러므로,

$$p-\text{값} = P(Z < -1.25) = 0.1056$$

이다.



㉣ $p-$값 0.1056은 0.10보다 크기 때문에 H_0를 기각하지 않는다. 즉, 40세 이상의 성인 중 일주일에 2번 이상 운동을 하는 사람의 비율이 20%보다 적다고 할 충분한 증거가 없다.

3-4-2. 두 독립표본 비율검정

① 독립적으로 추출된 두 독립표본들간의 측정비율의 차이가 있는지를 검정하기 위한 통계적 방법이다.

㉠ 어떤 사건이나 정책에 대하여 집단 간의 견해 차이 검정

㉡ 어떤 행위의 수행에 있어 집단간 성공률에 차이가 있는지에 대한 검정

② 두 모집단의 비율간의 차 $(p_1 - p_2)$에 대해 유의한 차이가 있는지 검정한다.

㉠ $(p_1 - p_2)$의 추정량인 표본비율의 차 $(\hat{p_1} - \hat{p_2})$

㉡ 두 집단간 비율의 표준오차 $SE = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$, $\left(\text{단, } \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}\right)$

㉢ 가설설정

H_0 : 두 집단간의 비율은 같다. $p_1 = p_2$ 또는 $p_1 - p_2 = 0$

H_1 : 두 집단간의 비율은 같지 않다. $p_1 \neq p_2$ 또는 $p_1 - p_2 \neq 0$

㉣ 두 독립표본 비율검정을 위한 Z통계값

$$Z = \frac{(\hat{p_1} - \hat{p_2}) - 0}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

③ 두 집단간 비율의 표준오차 계산 과정

㉠ 두 집단 비율의 차이에 대한 분산 $Var(p_d)$은

$$\begin{aligned} Var(p_d) &= Var(p_1 - q_2) = (+1)^2 Var(p_1) + (-1)^2 Var(p_2) \\ &= \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2} \end{aligned}$$

㉡ 만약 H_0이 참이면 $p_1 = p_2$이므로 귀무가설 아래서 비율인 p_0로 치환되어

$$Var(p_d) = \frac{p_0 q_0}{n_1} + \frac{p_0 q_0}{n_2}$$

이다. 그러나 p_0값이 얼마인지 모르므로 이를 추정해야 한다.

㉢ 두 모집단의 비율이 같으며 이를 대표할 비율은 첫 번째 표본에서 나타난 비율과 두 번째 표본에서 나타난 비율의 평균 비율(합성, pooled)이며, p에 대한 추정량은

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

이다.

㉣ 따라서 두 집단간 비율의 표준오차 $SE = \sqrt{\frac{p_0 q_0}{n_1} + \frac{p_0 q_0}{n_2}} = \sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ 이다.

〚예〛 부부가 결혼하여 모은 모든 재산은 부부의 공동명의로 소유해야 한다는 주장에 대하여 남녀 집단 간의 찬성비율의 차이가 있는지 알아보고자 한다. 결혼한 30대 성인 여자 200명, 남자 250명을 추출하여 부부 재산공동명의에 대한 찬성을 물은 결과 여성은 140명, 남성은 75명이었다. 여성들이 이 주장을 보다 지지한다는 이론적, 경험적 배경이 있을 때, 유의수준 0.05에서 검정하여라.

[풀이]

부부 재산공동명의에 찬성한 여성의 비율 p_1, 부부 재산공동명의에 찬성한 남자의 비율 p_2

㉠ 귀무가설과 대립가설을 세운다.

H_0 : $p_1 \leq p_2$

H_1 : $p_1 > p_2$

㉡ $p_1 = 0.7$, $p_2 = 0.3$

㉢ $\hat{p} = \dfrac{200 \times 0.7 + 250 \times 0.3}{200 + 250} = \dfrac{140 + 75}{450} = 0.477$

㉣ $Z = \dfrac{(0.07 - 0.03) - 0}{\sqrt{(0.477)(0.523)\left(\dfrac{1}{200} + \dfrac{1}{250}\right)}} = \dfrac{0.4}{0.05} = 8.441$

㉤ 유의수준 0.05에서 일방적 검정이므로 기각값은 +1.645이다. 통계값 8.441은 기각값보다 크므로 영가설을 기각한다. 그러므로 '부부의 공동명의에 의한 재산소유에 대한 찬성비율이 유의수준 0.05에서 여성들이 남성들보다 높다' 고 결론 내린다.