2021.11.10(水 13:10 ～)

兵庫高校

（演習問題 8a）「母比率の差の検定・推定」

ある２つの交差点で、信号を守るか守らないかについて、人数を調査した。一つ目の交差点 A は比較的狭い交差点で、２つ目の交差点 B は比較的広い交差点である。信号を順守（守る）人の割合において、広い交差点の方が、狭い交差点よりも守る比率が高いかどうかを、有意水準 5% で検定せよ。

表１ 交差点信号遵守データ

交差点	信号無視	信号順守	計
狭い交差点 A	36	4	40
広い交差点 B	24	36	60
計	60	40	100

(1) 信号順守率の違いに関する検定（正規近似を用いた解析）

〔手順１〕帰無仮説と対立仮説

交差点 A における順守率を P_A、交差点 B における順守率を P_B とする。

$$H_0\colon P_A \boxed{} P_B$$

$$H_1\colon P_A \boxed{} P_B$$

〔手順２〕有意水準と棄却域

$$\alpha = \boxed{}$$

$$R:\ u_0 \ \boxed{} \ \boxed{}$$

（ ≦ , ≠, ≧ ）

〔手順３〕検定統計量の計算

まず、$m = \boxed{}$, $n = \boxed{}$ であって、

$$p_A = \frac{x}{m} = \frac{\boxed{}}{\boxed{}} = \boxed{},$$

$$p_B = \frac{y}{n} = \frac{\boxed{}}{\boxed{}} = \boxed{},$$

$$\overline{p} = \frac{x+y}{m+n} = \frac{\boxed{} + \boxed{}}{\boxed{} + \boxed{}} = \frac{\boxed{}}{\boxed{}} = \boxed{}$$

となるから、u_0 は以下のように計算できる。

番　号					氏　名	

$$u_0 = \frac{p_A - p_B}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)\overline{p}(1-\overline{p})}}$$

$$= \frac{\boxed{} - \boxed{}}{\sqrt{\left(\frac{\boxed{}}{\boxed{}} + \frac{\boxed{}}{\boxed{}}\right) \times \boxed{}(\boxed{} - \boxed{})}} = \frac{\boxed{}}{\sqrt{\boxed{}}}$$

$$= \boxed{}$$

〔手順４〕判定と結論

統計量 $u_0 = \boxed{}$　$\boxed{}$ ゆえ、有意水準 5 % で帰無仮説は棄却 { される, されない } 。
したがって、交差点 B による順守率は、交差点 A の順守率よりも高いと { 言える, は言えない }。

(2) 順守率の違いに関する推定
〔手順１〕点推定

$$\widehat{P_A - P_B} = p_A - p_B = \frac{x}{m} - \frac{y}{n} = \frac{\boxed{}}{\boxed{}} - \frac{\boxed{}}{\boxed{}} = \boxed{} - \boxed{} = \boxed{}$$

〔手順２〕区間推定（信頼率 95% の信頼区間）

$$p_A - p_B \pm 1.960\sqrt{\frac{p_A(1-p_A)}{m} + \frac{p_B(1-p_B)}{n}}$$

$$= \boxed{} - \boxed{} \pm 1.960\sqrt{\frac{\boxed{}(1-\boxed{})}{\boxed{}} + \frac{\boxed{}(1-\boxed{})}{\boxed{}}}$$

$$= \boxed{} \pm \boxed{} = \boxed{}, \boxed{}$$

参考：ただし、$\sqrt{10} = 3.16$ として計算してよい。）

結論：したがって、交差点 B による順守率は、交差点 A の順守率よりも点推定値で $\boxed{}$% 高く、95% 信頼区間として $\boxed{}$% から $\boxed{}$% まで高いと言える。

2021.11.10(水 13:10 ～)

兵庫高校

（演習問題 8b）「分割表による検定」

3 つのドラマの「第 1 話」視聴世帯数は、表 1 の通りであった（実際のデータを一部改変)。有意水準 5% で、ドラマ間の視聴率に差があるかどうかを検定せよ。

表 1　ドラマごとの視聴世帯数データ

ドラマ名	見てた	見てない	計	視聴率
99, 9%	89	511	600	14.8%
せかむず	77	523	600	12.8%
できしな	62	538	600	10.3%
計	228	1572	1800	−−−

補足：ビデオリサーチ社によると、関東地区での第 1 話視聴率は、15.5%, 12.8%, 10.3% である。
関東地区（600 世帯）における、見ていた世帯数は、「93 世帯、77 世帯、 62 世帯」と考えられる。

分割表の検定

〔手順 1〕帰無仮説　H_0：ドラマ毎の視聴率には、差が { ある，ない }
　　　　　対立仮説　H_1：ドラマ毎の視聴率には、差が { ある，ない }

〔手順 2〕有意水準

$\alpha = \boxed{}$

〔手順 3〕棄却域の設定

$R : \chi_0^2 \ \boxed{} \ \chi^2(\boxed{}, \boxed{}) = \boxed{}$
（ ≦ , ≠, ≧ ）

〔手順 4〕検定統計量の計算

1) 期待度数の計算

$t_{ij} = \dfrac{T_i. \times T._j}{T}$ を計算する。

表 2　期待度数の計算（ t_{ij} 表）

ドラマ名	見てた	見てない	計
99.9%	76.0	524.0	600.0
せかむず			
できしな			
計			

番　号					氏　名	

表３　$x_{ij}-t_{ij}$ の計算表

ドラマ名	見てた	見てない	計
99.9%	13.0	−13.0	0.00
せかむず			
できしな			
計			

表４　$(x_{ij}-t_{ij})^2/t_{ij}$ の計算表

ドラマ名	見てた	見てない	計
99.9%	2.22	0.32	2.54
せかむず	0.01	0.00	0.01
できしな			
計	−−−−−−−−−−−		

$$\chi_0^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(x_{ij}-t_{ij})^2}{t_{ij}}=\boxed{}$$

〔手順５〕判定と結論

$$\chi_0^2=\boxed{}\qquad\boxed{}$$

となり、有意と { なる，はならない }。したがって、有意水準 5% で、ドラマ毎に視聴率が異なると { 言える，は言えない }。

2021.11.10(水 13:10 ～)

兵庫高校

（演習問題 8a）「母比率の差の検定・推定」

ある 2 つの交差点で、信号を守るか守らないかについて、人数を調査した。一つ目の交差点 A は比較的狭い交差点で、2 つ目の交差点 B は比較的広い交差点である。信号を順守（守る）人の割合において、広い交差点の方が、狭い交差点よりも守る比率が高いかどうかを、有意水準 5% で検定せよ。

表 1　交差点信号遵守データ

交差点	信号無視	信号順守	計
狭い交差点 A	36	4	40
広い交差点 B	24	36	60
計	60	40	100

(1) 順守率の違いに関する検定（正規近似を用いた解析）

〔手順 1〕帰無仮説と対立仮説

交差点 A における順守率を P_A、交差点 B における順守率を P_B とする。

$$H_0 \colon P_A = P_B$$

$$H_1 \colon P_A < P_B$$

〔手順 2〕有意水準と棄却域

$$\alpha = 0.05$$

$$R : u_0 \leqq -1.645$$

〔手順 3〕検定統計量の計算

まず、$m = 40,\ n = 60$ であって

$$p_A = \frac{x}{m} = \frac{4}{40} = 0.100,\ p_B = \frac{y}{n} = \frac{36}{60} = 0.600,$$

$$\overline{p} = \frac{x+y}{m+n} = \frac{4+36}{40+60} = \frac{40}{100} = 0.400$$

となるから、u_0 は以下のように計算できる。

$$u_0 = \frac{p_A - p_B}{\sqrt{\left(\dfrac{1}{m} + \dfrac{1}{n}\right)\overline{p}(1-\overline{p})}}$$

$$= \frac{0.100 - 0.600}{\sqrt{\left(\dfrac{1}{40} + \dfrac{1}{60}\right) \times 0.400(1-0.400)}} = \frac{-0.500}{\sqrt{0.0100}} = -5.000$$

〔手順４〕判定と結論

統計量 $u_0 = -5.000 < -1.645$ ゆえ、有意水準 5 % で帰無仮説は棄却される。

したがって、交差点 B による順守率は、交差点 A の順守率よりも <u>高いと言える</u>。

(2) 順守率の違いに関する推定

〔手順１〕点推定

$$\widehat{P_A - P_B} = p_A - p_B = \frac{x}{m} - \frac{y}{n} = \frac{4}{40} - \frac{36}{60} = 0.100 - 0.600 = -0.500$$

〔手順２〕区間推定（信頼率 95% の信頼区間）

$$\begin{aligned}
&p_A - p_B \pm 1.960\sqrt{\frac{p_A(1-p_A)}{m} + \frac{p_B(1-p_B)}{n}} \\
&= 0.100 - 0.600 \pm 1.960\sqrt{\frac{0.100(1-0.100)}{40} + \frac{0.600(1-0.600)}{60}} \\
&= -0.500 \pm 0.155 = -0.345,\ -0.655
\end{aligned}$$

したがって、交差点 B による順守率は、交差点 A の順守率よりも点推定値で 50.0% 高く、95% 信頼区間として 34.5% から 65.5% まで高いと言える。

創造応用 IS（稲葉）　　No.k8b

2021.11.10(水 13:10 ～)

兵庫高校

（演習問題 8b）「分割表による検定」（解答例）

3 つのドラマ（第 1 話）の視聴率は、表 1 の通りであった（実際のデータを一部改変）。有意水準 5% で、ドラマ間の視聴率に差があるかどうかを検定せよ。

表 1　ドラマごとの視聴率データ

ドラマ名	見てた	見てない	計
99, 9%	89	511	600
せかむず	77	523	600
できしな	62	538	600
計	228	1572	1800

分割表の検定

〔手順 1〕帰無仮説　H_0：ドラマ毎の視聴率には、差が ない

　　　　対立仮説　H_1：ドラマ毎の視聴率には、差が ある

〔手順 2〕有意水準

$\alpha = 0.05$

〔手順 3〕棄却域の設定

$R : \chi_0^2 \geqq \chi^2(\phi, \alpha) = \chi^2(2, 0.05) = 5.99$

ただし、$\phi = (a-1)(b-1) = (3-1)(2-1) = 2$

〔手順 4〕検定統計量の計算

1) 期待度数の計算

$t_{ij} = \dfrac{T_{i\cdot} \times T_{\cdot j}}{T}$ を計算する。

表 2　期待度数の計算（ t_{ij} 表）

ドラマ名	見てた	見てない	計
99.9%	76.0	524.0	600.0
せかむず	76.0	524.0	600.0
できしな	76.0	524.0	600.0
計	228.0	1572.0	1800.0

表 3　$x_{ij} - t_{ij}$ の計算表

ドラマ名	見てた	見てない	計
99.9%	13.0	−13.0	0.0
せかむず	1.0	−1.0	0.0
できしな	−14.0	14.0	0.0
計	0.0	0.0	0.0

表 4　$(x_{ij} - t_{ij})^2 / t_{ij}$ の計算表

ドラマ名	見てた	見てない	計
99.9%	2.22	0.32	2.54
せかむず	0.01	0.00	0.01
できしな	2.58	0.37	2.95
計	−−−−−−−		5.50

$$\chi_0^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{(x_{ij}-t_{ij})^2}{t_{ij}} = \boxed{5.50}$$

〔手順５〕判定と結論

$$\chi_0^2 = 5.50 < 5.99 = \chi^2(2, 0.05)$$

となり、有意と はならない。したがって、有意水準 5% で、ドラマ間に視聴率の違いがあると はいえない。

（補足：93 世帯のときは、$\chi_0^2 = 7.14 \geqq 5.99$ より、有意となり、視聴率に違いがあるといえる。）

参考：各セルにおける、観測度数と期待度数のズレ

カイ 2 乗検定では、全体として、行と列に関係があるかどうかを判定することができる。しかし、特定の行や、特定の列における値が、統計的に判断して大きいとか小さいとか、判断できると、有意義なアクションに繋がる可能性がある。

例えば、この演習問題の例であれば、特定のドラマにおいて、見ている人が多いとすれば、これは、（視聴率の意味で）良いドラマであったと考えられるからである。

このような目的に対して、残差、とか、基準化残差、とか、が提案されている。

残差 とは、$e_{ij} = x_{ij} - t_{ij}$

基準化残差 とは、$e'_{ij} = \dfrac{x_{ij}-t_{ij}}{\sqrt{t_{ij}}}$

いずれも、データが、期待値から、どの程度、ずれているかを判断できる。

（特に、基準化残差は、平均が 0 で、分散が 1 に基準化されているから、解釈ができる。）

表５　残差の表（ $e_{ij} = x_{ij} - t_{ij}$ 表）

ドラマ名	見てた	見てない	計
99.9%	13.0	– 13.0	0.0
せかむず	1.0	–1.0	0.0
できしな	– 14.0	14.0	0.0
計	0.0	0.0	0.0

表６　基準化残差の表（$e'_{ij} = \dfrac{x_{ij}-t_{ij}}{\sqrt{t_{ij}}}$表）

ドラマ名	見てた	見てない
99.9%	1.491	–0.568
せかむず	0.115	–0.044
できしな	–1.606	0.612

表７　93 世帯のときの基準化残差表

ドラマ名	見てた	見てない
99.9%	1.782	–0.685
せかむず	–0.038	0.015
できしな	–1.744	0.671

注：基準化残差は、その絶対値が、1.96 を超えると、注意し、2.50 を超えると、そのことをコメントすることができる。

補足：検定の多重性について

検定は、繰り返すと、そのうちの１つ以上が間違う確率（これを、**Type I FWE**(Family-Wize-Error) という。）が、大きくなる。そこで、全体として、１つ以上間違う確率を、コントロールする手法がある。これを、**多重比較法** という。今回の解釈で、1.960 を主に判断しないのは、このためである。