2009 年度前期　データ解析序説　第１５回

総合演習 (2)　問題

１．次の各項は，統計学の観点からみて正しい（あるいは適切である）かどうかを答えよ．正しくない（適切でない）ときは，どういう点がどのように正しくない（適切でない）かを説明せよ．

1. ある年の総選挙で選ばれた衆議院議員の，血液型の分布を調べたところ，その分布は，日本人全体の分布からの有意な偏りがあった．したがって，「血液型と『国会議員になる資質』の間には関連がある」といえる．
2. 母集団の平均の区間推定を行なう問題で，取り出した標本を使って計算した結果，「母平均の 95 パーセント信頼区間は 20 から 30 である」という結論を得た．このとき，母平均が 20 以上 30 以下である確率は 95 パーセントである．

２．さいころを 50 回投げたところ，１の目が 13 回出た．このさいころは，正しいさいころに比べて 1 が出やすいといえるかどうかを考える．

1. この問題に，講義で説明した知識を用いて答えるには，どのような前提がなりたつ必要があるか．
2. 上の前提が成り立つとする時，有意水準 5%の検定を行なって上の問題に答えよ．

３．X 薬品の「Y」という薬は，１つ１グラムの錠剤となっている．いま，10 個の錠剤を無作為抽出し，各々の錠剤に含まれる物質 P の量を調べた．その結果，各錠剤の物質 P の量（ミリグラム）は次の通りであった．

1.2　0.8　0.9　0.9　1.0　1.3　1.2　1.0　0.8　0.9

1. 講義で説明した知識を使って，「薬 Y に含まれる物質 P の割合」を区間推定するには，この測定がどのようなものであると仮定できる必要があるか．
2. 1. で答えた仮定が正しいとして，信頼係数 95% で 1. の区間推定を行なえ．

解答例

1．

1. このことが言えるには，「国会議員になる資質がある人たち」という母集団が存在し，現実の国会議員が，その母集団から無作為抽出された標本になっていなければならない．しかし，「国会議員になる資質がある人たち」という母集団はおそらく存在しない．
2. 母平均は，人が知らないだけで，実際にはひとつに決まっている．だから，「20 以上 30 以下」という具体的区間を求めた段階で，母平均が「20 以上 30 以下」であるかどうかはすでに決まっている．「95% 信頼区間」とは，この方法で信頼区間を何度も求めると，そのうち 95% は母平均を本当に含んでいる区間である，という意味である．

2．

1. 「さいころを投げて 1 の目が出る」という事象が，ベルヌーイ試行によって起きていること．「正しいさいころ」では，1 の目が出る確率は 1/6 であること．
2. 問題のさいころで，1 の目が出る確率を p とする．また，さいころを投げる回数を n，うち 1 の目が出る回数を S とする．正しいさいころでは，1 の目が出る確率は 1/6 であるから，「p が 1/6 よりも大きい」といえるかどうかを調べるために，帰無仮説「$p = 1/6$」，対立仮説「$p > 1/6$」という検定を行なう．「1 の目が出る」という事象は，確率 p で成功するベルヌーイ試行と考えられるので，さいころを n 回投げたときに 1 の目が出る回数 S は，２項分布 $B(n, p)$ にしたがう．よって，n が大きいとき，S は概ね正規分布 $N(np, np(1-p))$ にしたがうから，
$$Z = \frac{S - np}{\sqrt{np(1-p)}} \tag{1}$$
とおくと，Z は標準正規分布 $N(0, 1)$ にしたがう．
問題では $S = 13$ で，そのとき
$$Z = \frac{13 - 50 \cdot 1/6}{\sqrt{50 \cdot 1/6 \cdot (1 - 1/6)}} = 1.77 \tag{2}$$
となる．この問題では，対立仮説は「$p > 1/6$」であるから，「帰無仮説が正しいとしたとき，1 の出る回数 $S = 13$ は多すぎる」かどうかを調べる必要があるので，$S \geq 13$ である確率が小さいかどうかを調べる必要がある．$S \geq 13$ である確率は (1)(2) 式から $Z \geq 1.77$ となる確率と同じである．数表から，$Z \geq 1.64$ となる確率が 5%なので，$Z \geq 1.77$ となる確率は 5%よりも小さい．したがって，有意水準 5%で，帰無仮説は棄却され，「このさいころは，正しいさいころよりも 1 が出やすい」といえる．

3．

1. 薬Ｙの各錠に含まれる物質Ｐの割合が，正規分布にしたがっていること．
2. 薬Ｙ全体での 1 グラムあたりの物質Ｐの量を μ，各錠剤での物質Ｐの量の平均を $\bar{X}$，不偏分散を s^2，標本サイズを n とする．このとき，問題文から，各錠剤が含む物質Ｐの量は，平均 μ の正規分布にしたがうと考えられる．したがって，
$$t = \frac{\bar{X} - \mu}{\sqrt{\frac{s^2}{n}}} \tag{3}$$

という値（t 統計量）は自由度 $n-1$ の t 分布 $t(n-1)$ にしたがうので，$t_{0.025}(n-1)$ を「t 統計量がその値以上になる確率が 0.025 であるような値」とし，$-t_{0.025}(n-1)$ を「t 統計量がその値以下になる確率が 0.025 であるような値」とすると

$$P\left(\overline{X} - t_{0.025}(n-1)\sqrt{\frac{s^2}{n}} \le \mu \le \overline{X} + t_{0.025}(n-1)\sqrt{\frac{s^2}{n}}\right) = 0.95 \tag{4}$$

となるので，μ の 95% 信頼区間の下限・上限は (4) 式のかっこ内の不等式で与えられる．

$X = (1.2 + 0.8 + \cdots + 0.8 + 0.9)/10 = 1.0$，$s^2 = \{(1.2 - 1.0)^2 + \cdots + (0.9 - 1.0)^2\}/(10 - 1) = 0.031$，$n = 10$，$t_{0.025}(9) = 2.262$ を (4) 式に代入して，信頼区間の下限・上限を求めると，μ の 95%信頼区間は [0.87, 1.13]（ミリグラム）となる．したがって，物質 P の割合は [0.087, 0.113](%) となる．