統計　宿題第 3 回　解答例

（by TA 各務和彦 ＆ 竹内惠行）

問 1

【お詫び】解答例作成過程で、問１を見直したところ、作問にミスがあることが判明した。ミスの箇所は、
「分散は各都道府県で一定の 400、標本数は各都道府県 1000 であるものとします。」
という部分である。推定した普及率 $\hat{p}$ の分散は、真の視聴率（普及率）を p $(0 \le p \le 1)$、標本数を n と表すと、二項分布の性質より $\mathrm{Var}(\hat{p}) = p(1-p)/n$ となるから、「分散は各都道府県で一定の 400」という箇所は不要である。<u>以下は、この不要の部分が削除された問題についての解答例を与えることにする。</u>十分に注意して欲しい。

(1) 解答例

表より東京都のインターネット普及率は 50%を超えており、全国では高いほうである。そこで「東京はインターネット過疎である」という主張を、「東京は大阪よりもインターネットの普及率が低い」という仮説で置き換えることにしよう。

このとき、帰無仮説を「東京都と大阪府のインターネット普及率は同等」、対立仮説を「東京都は大阪府よりもインターネットの普及率が低い」として、統計的仮説検定に持ち込むことができる。これは、東京都と大阪府の普及率について差の検定をすることと同じである。標本数は両都府ともに 1000 で同じと仮定しているので、中心極限定理より、推定した（＝標本調査から得られた）大阪府の普及率を $\hat{p}_o$、東京都の普及率を $\hat{p}_T$ とすると、帰無仮説 $p_o = p_T$ の下では、

$$Q = \frac{\hat{p}_o - \hat{p}_T}{\sqrt{\left(\frac{p_o(100-p_o)}{1000} + \frac{p_T(100-p_T)}{1000}\right)}} \sim \mathrm{N}(0,1)$$

となる。<u>（普及率がパーセント表示になっていることに注意。）</u>これによって、東京都と大阪府の普及率に差があるかを検定することができる。上式に推定した普及率を代入し、分母の分散部の計算に必要な普及率に推定した普及率を代入すると、

$$Q = \frac{59.2 - 53.4}{\sqrt{\left(\frac{59.2(100-59.2)}{1000} + \frac{53.4(100-53.4)}{1000}\right)}} = \frac{5.8}{\sqrt{2.47456 + 2.48844}} = \frac{5.8}{2.23} = 2.60$$

であり、また対立仮説は $p_o > p_T$ であるから片側検定であるから、正規分布表より

$P(Q>2.60)<0.0047$。したがって有意水準 5%で（そして 1%でも）帰無仮説は棄却される。このことから、大阪と比べて「東京はインターネット過疎」であることを主張できる。

ちなみに関西圏で普及率の低い兵庫県と東京都について差の検定を行うと、

$$Q=\frac{56.8-53.4}{\sqrt{2.45376+2.48844}}=\frac{3.4}{2.22}=1.53$$

から、正規分布表より $P(Q>1.53)=0.063$。したがって有意水準 5%で帰無仮説は棄却されない。このことから、東京は兵庫と同等の普及率という主張にならざるを得ない。

(2) 解答例

問題より全国平均が $\mu=50$ と与えられているので、大阪の普及率が 50%であるかかどうかの検定を行うことにする。この検定の帰無仮説は $p_o=50$、対立仮説は $p_o>50$。帰無仮説の下では、

$$Q=\frac{\hat{p}_o-50}{\sqrt{\dfrac{50(100-50)}{1000}}}\sim \mathrm{N}(0,1)$$

である。推定した大阪の普及率を上式に代入すると、$Q=(59.2-50)\big/\sqrt{2.5}=5.82$ であり、標準正規分布表から $P(Q>5.82)<0.001$。したがって有意水準 5%で（1%でも）大阪のインターネット普及率が全国平均 50%と等しいという仮説は棄却され、東京出身の友人の反論に根拠はない。

問 2

(1) 母平均 μの推定量として、標本平均を用いる。

$$\hat{\mu}=\frac{\begin{array}{l}850+900+870+950+770+900+760+800+800+800\\ \qquad +820+850+800+900+830+800+850+800+800+800\end{array}}{20}$$
$$=832.5$$

(2) 正規分布の再生性より、標本平均もまた正規分布に従う。標本平均の分散は、母分散 σ^2 を標本数 n で割ったものに等しいから、求める分散は 3380/20=169。よって、標本平均は平均 μ、分散 169 の正規分布に従う、または $\overline{X}\sim N(\mu,169)$。

(3) 標本分散の不偏推定量を $\hat{\sigma}^2$ で表すと、

$$\hat{\sigma}^2 = \frac{\begin{aligned}&(850-832.5)^2+(900-832.5)^2+(870-832.5)^2+(950-832.5)^2\\&+(770-832.5)^2+(900-832.5)^2+(760-832.5)^2+(800-832.5)^2+(800-832.5)^2\\&+(800-832.5)^2+(820-832.5)^2+(850-832.5)^2+(800-832.5)^2+(900-832.5)^2\\&+(830-832.5)^2+(800-832.5)^2+(850-832.5)^2+(800-832.5)^2+(800-832.5)^2\\&+(800-832.5)^2\end{aligned}}{20-1}$$

$$= \frac{47575}{19} = 2503.95$$

(4) $\dfrac{\overline{X}-\mu}{\sqrt{\hat{\sigma}^2/20}} \sim t(19)$、または自由度 19 の t 分布。

(5) 母分散が既知の場合

$\overline{X} \sim N(\mu,\,169)$ に従うので、標準化すると $Z = \dfrac{\overline{X}-\mu}{13} \sim N(0,\,1)$ となる。

信頼係数は 0.95 とするので、

$P(-c < Z < c) = 0.95$ を求めればよい。

$$\begin{aligned}0.95 &= P(-1.96 \le Z \le 1.96)\\ &= P\left(-1.96 \le \frac{\overline{X}-\mu}{13} \le 1.96\right)\\ &= P(\overline{X} - 1.96\times 13 \le \mu \le \overline{X} + 1.96 \times 13)\end{aligned}$$

ここで、 $\overline{X} = 832.5$ を代入すると、

$807.02 \le \mu \le 857.98$ を得ることができる。

母分散が未知の場合

$T = \dfrac{\overline{X}-\mu}{\sqrt{\hat{\sigma}^2/20}} \sim t(19)$ に従うので、信頼係数 0.95 のとき

$$\begin{aligned}0.95 &= P(-t \le T \le t)\\ &= P\left(-t \le \frac{\overline{X}-\mu}{\sqrt{\hat{\sigma}^2/20}} \le t\right)\\ &= P(\overline{X} - 2.093\times 11.19 \le \mu \le \overline{X} + 2.093\times 11.19)\end{aligned}$$

$\hat{\sigma}^2 = 2503.95$、 $t = 2.093$ を代入して、

ここで、 $\overline{X} = 832.5$ を代入すると、

$809.08 \le \mu \le 855.92$ を得ることができる。