2008-02-29

統計レポート課題２：解答例

１．A 大学の入学試験では、1200 人の入学定員に対して 4540 人の受験者があった。入学試験問題は 800 点満点で、受験者全体の成績の分布は、平均 395 点、標準偏差 130 点の正規分布とみなしてよいとする。このとき、合格者数を 1200 人として、次の問いに答えよ。試験の成績は整数値とする。

(1)B 君が自己採点をしたところ 616 点であった。B 君は上位何パーセント以内に入ると予想されるか、小数第 1 位未満を四捨五入して答えよ。

$N(395, (130)^2)$ の正規分布なので、$Z = \frac{X-395}{130}$ は標準正規分布 N(0, 1) に従う。ゆえに、

$$
\begin{aligned}
P\{X \geqq 616\} &= P\{Z \geqq \frac{616-395}{130}\} \\
&= P\{Z \geqq 1.7\} \\
&= 0.5 - P\{0 \leqq Z \leqq 1.7\} \\
&= 0.5 - 0.4554 \\
&= 0.0446
\end{aligned}
$$

A. 4.46 %の小数第 2 位を四捨五入し、上位 4.5 %以内

(2) 合格するには少なくとも何点以上の成績であればよいか

まず合格率を求める。

$$1200 \div 4540 = 0.2643 \cdots$$

約 26.43 %が合格できる

正規分布表で上位 26.43 %のところを見ると

$$
\begin{aligned}
0.5 - 0.2643 &= 0.2357 \\
I(z) &= 0.2357 \\
z &= 0.63
\end{aligned}
$$

したがって、

$$
\begin{aligned}
\frac{X-395}{130} &= 0.63 \\
X &= 476.9
\end{aligned}
$$

A. 476.9 以上の整数値なので、477 点以上

(3)C 高校からの受験者 300 人の成績の分布は、平均 463 点、標準偏差 100 点の正規分布とみなしてよいとする。この 300 人の何パーセントが合格できると予想されるか、小数第１位未満を四捨五入して答えよ。

C 高校の成績は、$N(463, (100)^2)$ の正規分布なので、$Z = \frac{X-463}{100}$ は標準正規分布 N(0, 1) に従う。ゆえに、

$$P\{Z \geqq \frac{477-463}{100}\} = P\{Z \geqq 0.14\}$$

正規分布表でみると

$$\begin{aligned} &= 0.5 - P\{0 \leqq Z \leqq 0.14\} \\ &= 0.5 - 0.0557 \\ &= 0.4443 \end{aligned}$$

A. 44.43 %の小数第 2 位を四捨五入し、44.4 %

2. 平均 100、分散 2 で製造されている部品９ヶを抜き取ったところ 102, 101, 104, 98, 102, 96, 106, 97, 103 だった。この部品の標本平均を用いて有意水準 5 %で検定を行う。

(1) 帰無仮説 H_0

H_0 :　$\mu = 100$

これら 9 標本は N(100, 2) となる母集団に属する。

(2) 検定

部品は大きくても小さくてもいけないので、有意水準 5 %の両側検定となり、正規分布の片側 2.5 %のところを見て、

$$t \leqq -1.96, \quad t \geqq 1.96$$

が成り立つかどうかを検定する。

まず、標本平均を求める。

$$\bar{x} = \frac{1}{9}\sum_{k=1}^{9} x_k = 101$$

$$n = 9, \quad \bar{x} = 101, \quad \mu = 100, \quad \sigma = \sqrt{2},$$

統計量は以下の式によって求められる。

$$t = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{101-100}{\sqrt{2}/\sqrt{9}} = \frac{3}{\sqrt{2}} \risingdotseq 2.1216 > 1.96$$

したがって、H_0 は棄却される。

(3) 結論

これらはN(100, 2) となる母集団に属さず、機械は正常に動作していないと考えられる。

3．予防注射の有効性を５％で検定したい

(1) 帰無仮説 H_0

H_0：予防注射と罹病とは無関係。２つの事象は独立。

(2) 検定

帰無仮説 H_0 に基づいて、周辺確率 (注射あり 0.6；かかった 0.7) から、それぞれの理論値を求める。

注射あり＋かかった　：　$12 \div 20 \times 14 = 8.4$
注射あり＋かからない：　$12 \div 20 \times 6 = 3.6$
注射なし＋かかった　：　$8 \div 20 \times 14 = 5.6$
注射なし＋かからない：　$8 \div 20 \times 6 = 2.4$

実現値

	かかった	かからない	計
注射あり	8	4	12
注射なし	6	2	8
計	14	6	20

理論値

	かかった	かからない	計
注射あり	8.4	3.6	12
注射なし	5.6	2.4	8
計	14	6	20

自由度 f = (縦の枠数 - 1)(横の枠数 - 1) = 1、χ^2 分布の有意水準 0.05 の点は 3.84 である。

統計量 t は $\frac{(\text{実現値}-\text{理論値})^2}{\text{理論値}}$ の総和は

$$t = \frac{(8-8.4)^2}{8.4} + \frac{(4-3.6)^2}{3.6} + \frac{(6-5.6)^2}{5.6} + \frac{(2-2.4)^2}{2.4} = 0.159 < 3.84 = \chi_1^2(0.05)$$

したがって、H_0 を採択する。

(3) 結論

・「予防注射と罹病に関連がある」とはいえない（何もいえない)。
・「予防注射と罹病は独立と考えても矛盾はない。

4．以下のようなサイコロはイカサマか？

(1) 帰無仮説 H_0

H_0：サイコロはイカサマではない。
　　つまり、すべての目は等確率で i の目が出る確率 $P(i) = \frac{1}{6}$

(2) 検定

　実現値と理論値は以下のとおり。
　実現値

1の目	2の目	3の目	4の目	5の目	6の目	計
7	11	6	14	8	14	60

　理論値

1の目	2の目	3の目	4の目	5の目	6の目	計
10	10	10	10	10	10	60

自由度5、有意水準0.05で χ^2 検定を行う。

統計量tは $\frac{(\text{実現値}-\text{理論値})^2}{\text{理論値}}$ の総和となる。

$$\begin{aligned} t &= \frac{(7-10)^2}{10}+\frac{(11-10)^2}{10}+\frac{(6-10)^2}{10}+\frac{(14-10)^2}{10}+\frac{(8-10)^2}{10}+\frac{(14-10)^2}{10} \\ &= 6.2 < 11.07 = \chi_5^2(0.05) \end{aligned}$$

したがって、帰無仮説 H_0 を採択する。

(3) 結論

・「サイコロはイカサマである」とはいえない(何もいえない)。
・サイコロはイカサマでないと考えても矛盾はない

5．平均値の差の検定の例題を作り、検定方法を明快に説明する模範解答を作れ

例題：バーガー・ハシモトのレタスバーガーにはレタスが30g入っているという。それを確かめるために、毎日１個ずつ１週間、計７個のレタスバーガーについてレタスの量を調べた。結果は以下の表のとおり。これについて、有意水準５％で検定を行う。

　各バーガーのレタスの量

1日目	2日目	3日目	4日目	5日目	6日目	7日目
26	35	33	24	25	34	26

(1) 帰無仮説 H_0

H_0：バーガー・ハシモトのレタスバーガーのレタスの量は 30g である。

(2) 検定

母分散が未知なので t 検定を用いる。標本数 n = 7 なので、自由度 $f = (n-1)$ は 6 となり、自由度 6 の t 分布に従う。レタスは多くても少なくてもいけないので両側検定と考え、$t_6(0.025) = 2.447$ がさかいめとなる。

まず、標本平均を求める。

$$\bar{x} = \frac{1}{7}\sum_{k=1}^{7} x_k = 29$$

次に、母分散が未知なので不偏分散 U^2 を求める。

$$U^2 = \frac{1}{n-1}\sum_{n}^{k=1}(x_k - \bar{x})^2$$

$$U^2 = \frac{1}{6}\{(26-29)^2 + (35-29)^2 + (33-29)^2 + (24-29)^2 + (25-29)^2 + (34-29)^2 + (26-29)^2\}$$

$$= 22.67$$

統計量 t は以下の式によって求められる。

$$t = \frac{\bar{x} - \mu}{U/\sqrt{n}} = \frac{29 - 30}{\sqrt{22.67}/\sqrt{7}} = -0.56$$

$-2.447 \leqq -0.56 \leqq 2.447$ より、H_0 は採択される。

(3) 結論

・「バーガー・ハシモトのレタスバーガーのレタスの量は 30g でない」とはいえない（何もいえない）。
・「バーガー・ハシモトのレタスバーガーのレタスの量は 30g と考えても矛盾はない。