政策特論（ノンパラメトリック統計学）
第2回宿題解答例

2009年1月23日

Q. 3.1

$S = S_+$ のとき　$s_i^+ = \max\{s_i, 0\}$　とすると、

$$S_+ = \sum_{i=1}^{n} s_i^+$$

と表すことができ、帰無仮説の下で $rank$ の符号が正値をとる確率も負値をとる確率も、どちらも $\frac{1}{2}$ であるから、

$$\begin{aligned} \mathrm{E}(S_+) &= \mathrm{E}(\sum_{i=1}^{n} s_i^+) = \sum_{i=1}^{n} \mathrm{E}(s_i^+) \\ &= \sum_{i=1}^{n} \left[\frac{1}{2} \mid s_i \mid + \frac{1}{2} 0\right] = \frac{1}{2} \sum_{i=1}^{n} \mid s_i \mid \end{aligned} \tag{1}$$

$$\begin{aligned} \mathrm{Var}(S_+) &= \mathrm{Var}(\sum_{i=1}^{n} s_i^+) \qquad \text{(i.i.d より)} \\ &= \sum_{i=1}^{n} \mathrm{Var}(s_i^+) = \sum_{i=1}^{n} \{\mathrm{E}({s_i^+}^2) - [\mathrm{E}(s_i^+)]^2\} \\ &= \sum_{i=1}^{n} \left[\frac{1}{2}|s_i|^2 - \frac{1}{4}|s_i|^2\right] = \frac{1}{4} \sum_{i=1}^{n} \mid s_i \mid^2 \end{aligned} \tag{2}$$

$S = S_-$ のとき　$s_i^- = \min\{s_i, 0\}$　とすると、

$$S_- = \sum_{i=1}^{n} \mid s_i^- \mid$$

と表すことができるので

$$\begin{aligned}\mathrm{E}(S_-) &= \mathrm{E}(\sum_{i=1}^{n} \mid s_i^- \mid) = \sum_{i=1}^{n} \mathrm{E} \mid s_i^- \mid \\ &= \sum_{i=1}^{n} \left(\frac{1}{2} \mid s_i \mid + \frac{1}{2} 0 \right) = \frac{1}{2} \sum_{i=1}^{n} \mid s_i \mid \end{aligned} \tag{3}$$

$$\begin{aligned}\mathrm{Var}(S_-) &= \mathrm{Var}(\sum_{i=1}^{n} \mid s_i^- \mid) = \sum_{i=1}^{n} \mathrm{Var}(\mid s_i^- \mid) \\ &= \frac{1}{4} \sum_{i=1}^{n} s_i^2 \end{aligned} \tag{4}$$

(1)(3) より

$$\mathrm{E}(S) = \frac{1}{2} \sum_{i=1}^{n} \mid s_i \mid$$

(2)(4) より

$$\mathrm{Var}(S) = \frac{1}{4} \sum_{i=1}^{n} s_i^2$$

■

Q. 3.8

1) Wilcoxon Signed rank test

$$\begin{cases} H_0 : \mu = 400 & (\mu\text{は平均値（＝中央値）}) \\ H_1 : \mu \neq 400 \end{cases}$$

設問のサンプルから帰無仮説のもとでの S_+ を求めると、表 1 のとおり、

$$S_+ = 19$$

となる。

$$P(S_+ \leq 19) \simeq 0.06$$

なので両側検定を行うと有意水準 10% でも帰無仮説を棄却できない。したがってページ数の平均値が 400 という仮説は棄却できない。

2) Wilcoxon の 95% 信頼区間

$n = 12$ のときの Wilcoxon の signed rank test 統計量の分布（テキスト p50, table3.2）より

$$P(S_+ \leq 14) \simeq 0.026$$

なので S_+ の 95% 信頼区間は (14, 64) となる。（78-14=64 であるので）

Walsh average を用いると、この S_+ の信頼区間に相当する μ（平均値）の区間は (200.5, 433) になることがわかる（表 2 を見よ）。

3) t 分布に基づく 95% 信頼区間

ページ数の標本平均 $\bar{\mu}$ とその標準誤差を求めると、

$$\bar{\mu} \simeq 316.67$$

$$\sqrt{\mathrm{Var}(\bar{\mu})} = \sqrt{\hat{\sigma^2}/12} \simeq 40.56$$

ここで $\hat{\sigma^2}$ はページ数の分散の推定量である。

自由度 11(=12-1) の t 分布の上側 2.5% 点は 2.201 であるから、ページ数 μ の 95% 信頼区間は

$$\begin{aligned} \Pr\left(-2.201 \leq \frac{\mu - 316.7}{40.56} \leq 2.201\right) &= \Pr(227.43 \leq \mu \leq 405.97) \\ &= 0.95 \end{aligned}$$

となり、2) で求めた信頼区間より狭いことがわかる。

■

表1 Wilcoxonのsigned rank

観測値	126	142	156	228	245	246	370	419	433	454	478	503
観測値－４００	-274	-258	-244	-172	-155	-154	-30	19	33	54	78	103
符号付順位 signed rank	-12	-11	-10	-9	-8	-7	-2	1	3	4	5	6

表2 Walsh average

	126	142	156	228	245	246	370	419	433	454	478	503
126	126	134	141	177	185.5	186	248	272.5	279.5	290	302	314.5
142		142	149	185	193.5	194	256	280.5	287.5	298	310	322.5
156			156	192	200.5	201	263	287.5	294.5	305	317	329.5
228				228	236.5	237	299	323.5	330.5	341	353	365.5
245					245	245.5	307.5	332	339	349.5	361.5	374
246						246	308	332.5	339.5	350	362	374.5
370							370	394.5	401.5	412	424	436.5
419								419	426	436.5	448.5	461
433									433	443.5	455.5	468
454										454	466	478.5
478											478	490.5
503												503