北陸大学　紀要
第22号（1998）
pp. 53～61

棄却検定法について

小　嶋　迪　孝*

On Rejection Tests

Michitaka Kojima *

Received October 30, 1998

§１．はじめに

棄却検定法とは，ある実験における測定値のうち，明らかに実験上の過失があったものは最初から除外するのは当然であるが，特に原因が考えられないのに他の値と大きくかけ離れている測定値があった場合，その測定値がその集団に属するものであるか否かをみる方法である．その方法は，多くの測定値の分布が正規型であることを利用して，他の値と著しく離れていると思われるものについて，それを含めた（あるいは含めない）測定値の平均値からの離れ具合を，それを含めた（あるいは含めない）測定値の標準偏差の尺度で計り，その出現の確率によってその値を集団の要素としてよいかどうかを検討するものである．帰無仮説H_0，対立仮説H_1を各々

H_0：得られた測定値は全て同一の正規母集団からの値とみなす，

H_1：ある特定な測定値はこの正規母集団からかけ離れた値とみなす，

と設定し，$\alpha=0.05$ または $\alpha=0.01$ を有意水準とする．

この検定には，(Ⅰ)増山元三郎博士の棄却検定法，(Ⅱ)Thompsonの棄却検定法，(Ⅲ)Smirnoff-Grubbsの棄却検定法 の３つが特に知られているが，これらの持つ意味はそれぞれ異なるものである．これらの棄却検定法についての陳述は，数学の立場での一般的な統計学の書物には載っていない場合が多く，また応用する立場での推計学の書物にはその手順のみが載っていて，その背後にある標本分布やその証明についての記載があるものは皆無であると言っても過言ではない．この稿での目的は，これらを[2, pp. 230-233]および[3, pp. 52-55]に従って定理の形で公式化し，[1, pp. 168-169, 176-180]を参考にして厳密な証明を与えることである．

以下，正規母集団$N(\mu, \sigma^2)$からの標本変量$X_1, X_2, \ldots, X_n$に対して，その標本平均を$\overline{X}$，標本分散をS^2で表す：

---

*薬　学　部
Faculty of Pharmaceutical Sciences

$$\overline{X}:=\frac{1}{n}\sum_{i=1}^{n}X_i,\qquad S^2:=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\qquad(S=\sqrt{S^2}).$$

また $n\geqq 3$ とする.

§2. 定理

(Ⅰ)増山元三郎博士の棄却検定法

定理1. 正規母集団からn個の標本$X_1, X_2, ..., X_n$およびもう1つの標本X_0をとるとき,

$$\text{統計量}\,T:=\sqrt{\frac{n-1}{n+1}}\,\frac{\overline{X}-X_0}{S}\sim t_{n-1}$$

が成立する.

この事実により $P(|T|\geqq t_{n-1}(\alpha))=\alpha$ に従ってX_0を棄却検定するのが増山博士の棄却検定法である.

(Ⅱ)Thompsonの棄却検定法

定理2. 正規母集団からn個の標本$X_1, X_2, ..., X_n$をとるとき,これらの中からランダムに1つのX_iを選んで,

$$\text{統計量}\,T_i:=\frac{X_i-\overline{X}}{S}$$

を作ると,T_iの分布の確率密度関数は,iについて一様に,次の $f(x)=f_n(x)$ となる:

$$f(x)=\frac{1}{\sqrt{n-1}}\,\frac{\Gamma((n-1)/2)}{\sqrt{\pi}\,\Gamma((n-2)/2)}\left\{1-\frac{x^2}{n-1}\right\}^{(n-4)/2}I_{\{|x|\leqq\sqrt{n-1}\}}(x)\quad\cdots\cdots(*)$$

ここに$I_A(x)$は集合Aの特性関数.

$\alpha=0.05$ あるいは $\alpha=0.01$ に対して,$T=T_i$について

$$\alpha=P(|T|\geqq\tau_n(\alpha))=1-P(|T|\leqq\tau_n(\alpha))=1-2\int_0^{\tau_n(\alpha)}f(x)\,dx$$

となる$\tau_n(\alpha)$がThompsonの棄却限界値として次のように知られている.この表の源は[4]である.

Thompson の棄却限界値

標本数 n	α 0.05	α 0.01	標本数 n	α 0.05	α 0.01	標本数 n	α 0.05	α 0.01
3	1.4099	1.41404	12	1.910	2.348	21	1.936	2.451
4	1.6454	1.7147	13	1.915	2.368	22	1.937	2.460
5	1.757	1.9175	14	1.919	2.385	23	1.938	2.465
6	1.814	2.051	15	1.923	2.399	24	1.940	2.470
7	1.848	2.142	16	1.926	2.411	25	1.941	2.475
8	1.870	2.207	17	1.928	2.422	30	1.944	2.493
9	1.885	2.256	18	1.931	2.432	∞	1.960	2.576
10	1.895	2.294	19	1.932	2.440			
11	1.904	2.324	20	1.934	2.447			

(Ⅲ)Smirnoff-Grubbsの棄却検定法

定理3． 正規母集団からｎ個の標本をとり，大きさの順に並べて$X_1 \leqq X_2 \leqq \cdots \leqq X_{n-1} \leqq X_n$としたとき，最大変量$X_n$に対して

$$\text{統計量}\,T := \frac{X_n - \overline{X}}{S}$$

の確率密度関数 $g(t)=g_n(t)$ は

$$g(t)=n\left(\int_{-\infty}^{t} f(x)\,dx\right)^{n-1} f(t) \quad (-\infty<t<\infty)$$

によって与えられる．ここに$f(x)=f_n(x)$は(＊)で定義された関数である．

$\alpha=0.05$ あるいは $\alpha=0.01$ に対して，

$$\alpha=P(T \geqq \tau_n(\alpha))=\int_{\tau_n(\alpha)}^{\infty} g(t)\,dt$$

となる$\tau_n(\alpha)$がSmirnoff-Grubbsの棄却限界値として次のように知られている．この表の源は[5]である．

Smirnoff-Grubbs の棄却限界値

標本数 n	α 0.05	α 0.01	標本数 n	α 0.05	α 0.01	標本数 n	α 0.05	α 0.01
3	1.412	1.414	11	2.343	2.606	19	2.600	2.932
4	1.689	1.723	12	2.387	2.663	20	2.623	2.959
5	1.869	1.955	13	2.426	2.714	21	2.644	2.984
6	1.996	2.135	14	2.461	2.759	22	2.664	3.008
7	2.093	2.265	15	2.493	2.800	23	2.683	3.030
8	2.172	2.374	16	2.523	2.837	24	2.701	3.051
9	2.237	2.464	17	2.551	2.871	25	2.717	3.071
10	2.294	2.540	18	2.577	2.903			

§3．定理の証明

定理1の証明．この証明は簡単である．正規母集団を$N(\mu,\sigma^2)$とするとき，

$$\overline{X}\sim N(\mu,\sigma^2/n),\qquad X_0\sim N(\mu,\sigma^2)$$

であり，$\overline{X}$とX_0は互いに独立であるから

$$\overline{X}-X_0\sim N\left(0,\frac{n+1}{n}\sigma^2\right)$$

であり，よって

$$Z:=\frac{\overline{X}-X_0}{\sqrt{\frac{n+1}{n}\sigma^2}}=\sqrt{\frac{n}{n+1}}\frac{\overline{X}-X_0}{\sigma}\sim N(0,1)$$

が成立する．また

$$\chi^2:=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{nS^2}{\sigma^2}\sim\chi^2_{n-1}$$

であり，Zとχ^2は互いに独立であるから

$$T=\sqrt{\frac{n-1}{n+1}}\frac{\overline{X}-X_0}{S}=\frac{Z}{\sqrt{\chi^2/(n-1)}}\sim t_{n-1}$$

が得られる．

定理2の証明．証明は技巧的である．$f(x)\geqq 0$であり，また$\int_{-\infty}^{\infty}f(x)\,dx=1$であることは簡単な変数変換によって容易に確かめられる．選ばれたX_iを例えばX_nとして，$X_1, X_2,\dots,X_{n-1}$より

$$\overline{X'}:=\frac{1}{n-1}\sum_{i=1}^{n-1}X_i,\qquad (S')^2:=\frac{1}{n-1}\sum_{i=1}^{n-1}(X_i-\overline{X'})^2$$

とおく．

先ず，$\overline{X}$およびS^2を$X_n, \overline{X'}, (S')^2$でもって表現する．

$$n\overline{X}=\sum_{i=1}^{n-1}X_i+X_n=(n-1)\overline{X'}+X_n$$

であるから

(1)　$$\overline{X}=\frac{n-1}{n}\overline{X'}+\frac{1}{n}X_n$$

と表される．これを用いることにより

$$nS^2=\sum_{i=1}^{n-1}(X_i-\overline{X})^2+(X_n-\overline{X})^2$$

$$=\sum_{i=1}^{n-1}\{X_i-(\frac{n-1}{n}\overline{X'}+\frac{1}{n}X_n)\}^2+\{X_n-(\frac{n-1}{n}\overline{X'}+\frac{1}{n}X_n)\}^2$$

$$=\sum_{i=1}^{n-1}\{(X_i-\overline{X'})-\frac{1}{n}(X_n-\overline{X'})\}^2+\{\frac{n-1}{n}(X_n-\overline{X'})\}^2$$

$$=\sum_{i=1}^{n-1}(X_i-\overline{X'})^2+\frac{n-1}{n^2}(X_n-\overline{X'})^2+\frac{(n-1)^2}{n^2}(X_n-\overline{X'})^2$$

$$(\because \quad \sum_{i=1}^{n-1}(X_i-\overline{X'})=0 \quad)$$

$$=\sum_{i=1}^{n-1}(X_i-\overline{X'})^2+\frac{n-1}{n}(X_n-\overline{X'})^2$$

$$=(n-1)(S')^2+\frac{n-1}{n}(X_n-\overline{X'})^2$$

が得られ，その結果

(2)　$$S^2=\frac{n-1}{n}(S')^2+\frac{n-1}{n^2}(X_n-\overline{X'})^2$$

が得られる．

次に，S'と$X_n-\overline{X'}$は互いに独立な統計量であることに注意して，次の統計量Vとその分布を考える：

(3)　$$V:=\sqrt{\frac{n-2}{n}}\,\frac{X_n-\overline{X'}}{S'}.$$

$X_n \sim N(\mu,\sigma^2)$，$\overline{X'}\sim N(\mu,\sigma^2/(n\text{-}1))$ で，X_nとX'は互いに独立であるから，

$$X_n-\overline{X'}\sim N(0,\ \frac{n}{n-1}\sigma^2)$$

であり，それ故

$$Z:=\frac{X_n-\overline{X'}}{\sqrt{\frac{n}{n-1}\sigma^2}}=\sqrt{\frac{n-1}{n}}\,\frac{X_n-\overline{X'}}{\sigma}\sim N(0,1)$$

である．また更に

$$\chi^2:=\frac{1}{\sigma^2}\sum_{i=1}^{n-1}(X_i-\overline{X'})^2=\frac{(n-1)(S')^2}{\sigma^2}\sim\chi^2_{n-2}$$

であるから

(4)　$$V=\sqrt{\frac{n-2}{n}}\,\frac{X_n-\overline{X'}}{S'}=\frac{Z}{\sqrt{\chi^2/(n-2)}}\sim t_{n-2}$$

が成り立つ．

TとVとの関係式を求める．(1)より

$$X_n-\overline{X}=X_n-(\frac{n-1}{n}\overline{X'}+\frac{1}{n}X_n)=\frac{n-1}{n}(X_n-\overline{X'})$$

であるから

$$T=\frac{X_n-\overline{X}}{S}=\frac{n-1}{n}\ \frac{X_n-\overline{X'}}{S}$$

である．ここで今$T\neq 0$ i.e. $X_n\neq\overline{X}$ i.e. $X_n\neq\overline{X'}$とする．そうすると

$$S=\frac{n-1}{n}\ \frac{X_n-\overline{X'}}{T}$$

であるから，これを(2)に代入して両辺を$(n-1)/n^2$で割ると

$$\frac{(n-1)(X_n-\overline{X'})^2}{T^2}=n(S')^2+(X_n-\overline{X'})^2$$

を得る．そして(3)より

$$n(S')^2=\frac{(n-2)(X_n-\overline{X'})^2}{V^2}$$

であるから，これを上の式に代入して両辺を$(X_n-\overline{X'})^2(\neq 0)$で割ると

$$\frac{n-1}{T^2}=\frac{n-2}{V^2}+1=\frac{(n-2)+V^2}{V^2}$$

即ち

$$T^2=\frac{(n-1)V^2}{(n-2)+V^2}$$

が得られる．TとVは同符号であるから関係式

$$T=\frac{\sqrt{n-1}\ V}{\sqrt{(n-2)+V^2}} \tag{5}$$

が成立する．$T=0$と$V=0$とは同値であるから(5)は$T=0$の時でも成り立つ．

(4)により$V^2\sim F^1{}_{n-2}$であるから，この事実と(5)を利用して，Tの分布関数$P(T\leqq x)$ $(-\infty<x<\infty)$を求める．$F^1{}_m$の確率密度関数を$f^1{}_m(y)$と書くと，その定義式は

$$f^1{}_m(y)=\frac{\Gamma((1+m)/2)}{\Gamma(1/2)\Gamma(m/2)}(\frac{1}{m})^{1/2}\ y^{-1/2}\ (1+\frac{y}{m})^{-(1+m)/2}\ I_{[0,\infty)}(y)$$

である．またTのとる値の範囲は(5)より

$$-\sqrt{n-1}\leqq T\leqq\sqrt{n-1}$$

である．結論を示すには $x\in(-\infty,\infty)$ に対して等式

$$P(T\leqq x)=\int_{-\infty}^{x} f(t)\ dt \tag{6}$$

が成り立つことを示せばよい．

① $x\geqq\sqrt{n-1}$ のとき．

$$P(T \leqq x)=1=\int_{-\infty}^{\infty} f(t)\ dt=\int_{-\infty}^{x} f(t)\ dt.$$

② $x \leqq -\sqrt{n-1}$ のとき.

$$P(T \leqq x)=0=\int_{-\infty}^{-\sqrt{n-1}} f(t)\ dt=\int_{-\infty}^{x} f(t)\ dt.$$

③ $0 \leqq x \leqq \sqrt{n-1}$ のとき.

$$P(T \leqq x)=1-P(T \geqq x)=1-P\left(\frac{\sqrt{n-1}\ V}{\sqrt{(n-2)+V^2}} \geqq x\right)$$

$$=1-\frac{1}{2}P\left(\frac{(n-1)V^2}{(n-2)+V^2} \geqq x^2\right)$$

であるが，ここで

$$\frac{(n-1)V^2}{(n-2)+V^2} \geqq x^2 \Leftrightarrow \quad V^2 \geqq \frac{(n-2)x^2}{(n-1)-x^2} (\fallingdotseq \alpha_x \geqq 0)$$

であるから

$$P(T \leqq x)=1-\frac{1}{2}P(V^2 \geqq \alpha_x)$$

である．ところで

$$\frac{1}{2}P(V^2 \geqq \alpha_x)=\frac{1}{2}\int_{\alpha_x}^{\infty} f^1{}_{n-2}(y)\ dy$$

$$=\frac{1}{2}\ \frac{\Gamma((n-1)/2)}{\sqrt{\pi}\ \Gamma((n-2)/2)}\ \frac{1}{\sqrt{n-2}}\int_{\alpha_x}^{\infty} y^{-1/2}\ \left(1+\frac{y}{n-2}\right)^{-(n-1)/2} dy$$

において，

$$t=\frac{\sqrt{n-1}\ \sqrt{y}}{\sqrt{(n-2)+y}} \fallingdotseq \phi(y)$$

と変数変換すると，$\phi(\alpha_x)=|x|(=x)$，$\lim_{y\to\infty}\phi(y)=\sqrt{n-1}$ であり，

$$y=\frac{(n-2)t^2}{(n-1)-t^2}, \qquad \frac{dy}{dt}=\frac{2(n-2)(n-1)t}{\{(n-1)-t^2\}^2}$$

であるから，

$$\frac{1}{2}P(V^2 \geqq \alpha_x)$$

$$=\frac{1}{2}\ \frac{\Gamma((n-1)/2)}{\sqrt{\pi}\ \Gamma((n-2)/2)}\ \frac{1}{\sqrt{n-2}}\int_{x}^{\sqrt{n-1}} \left\{\frac{(n-2)t^2}{(n-1)-t^2}\right\}^{-1/2}\cdot$$

$$\cdot\left\{\frac{n-1}{(n-1)-t^2}\right\}^{-(n-1)/2}\cdot\frac{2(n-2)(n-1)t}{\{(n-1)-t^2\}^2}\ dt$$

$$=\frac{\Gamma((n-1)/2)}{\sqrt{\pi}\,\Gamma((n-2)/2)}\frac{1}{\sqrt{n-2}}\int_x^{\sqrt{n-1}}\{1-\frac{t^2}{n-1}\}^{(n-4)/2}\,dt=\int_x^{\sqrt{n-1}}f(t)\,dt$$

を得る．故に

$$P(T\leqq x)=1-\int_x^{\sqrt{n-1}}f(t)\,dt=\int_{-\infty}^{x}f(t)\,dt$$

が成立する．

④ $-\sqrt{n-1}\leqq x\leqq 0$ のとき．

$$P(T\leqq x)=P(\frac{\sqrt{n-1}\;V}{\sqrt{(n-2)+V^2}}\leqq x)=\frac{1}{2}P(\frac{(n-1)V^2}{(n-2)+V^2}\geqq x^2)$$

$$=\frac{1}{2}P(V^2\geqq\alpha_x)$$

であるが，上と同じ計算を $\phi(\alpha_x)=|x|(=-x)$ であることに注意して辿れば，

$$P(T\leqq x)=\int_{|x|}^{\sqrt{n-1}}f(t)\,dt=\int_{-\sqrt{n-1}}^{x}f(t)\,dt=\int_{-\infty}^{x}f(t)\,dt$$

が得られ，以上で(6)が示され，定理の証明が終る．

定理３の証明． この証明は最も簡単な順序統計量の分布を考えることに基づき，それ故多項分布の求め方と同様の計算をすることによって可能となる．定理２に従えば，各 i について $T_i=\dfrac{X_i-\overline{X}}{S}$ は i に無関係な確率密度関数f(x)をもつ統計量である．X_nを $X_i(i=1,2,\ldots,n)$の最大変量とすると，T_nは$T_i(i=1,2,\ldots,n)$の最大変量である．T_nの確率密度関数を $g(t)$ とおくと

$$\text{確率}P\equiv P(t\leqq T_n\leqq t+dt)=g(t)dt$$

と表されるが，この確率は

$$T_1,T_2,\ldots,T_{n-1}\leqq t\leqq T_n\leqq t+dt$$

となる事象の同時確率に他ならない．この事象を言葉で表せば，n個の標本の内，

(n−1)個がt以下の値をとり，

1個がtとt+dtの間の値をとる，

と云うことになり，そして1つの標本が，

t以下の値をとる確率は $\int_{-\infty}^{t}f(x)\,dx$，

tとt+dtの間の値をとる確率は $f(t)dt$，

であるから，多項分布の求め方と同様の計算をすれば，求める確率Pは

$$P=\frac{n!}{(n-1)!1!}\left(\int_{-\infty}^{t} f(x)\ dx\right)^{n-1}\cdot f(t)dt=n\left(\int_{-\infty}^{t} f(x)\ dx\right)^{n-1}\cdot f(t)dt$$

となり，結論が示されたことになる.

§4．注意

1．　3つの検定の利用の仕方は次の通りであろう.

増山博士の検定法(Ⅰ)は今得られた1つの測定値が過去の測定値を基準にして異常であるか否かを検討する場合に便利であると思われる.

Thompsonの検定法(Ⅱ)は，得られた測定値のいずれのものに対しての標本分布が与えられた結果から出てきたものであるから，棄てられないであろうと思われる資料を保留するときに利用すると都合がよい.

Smirnoff-Grubbsの検定法(Ⅲ)は最大値（あるいは最小値）の資料が棄てられるであろうと思われるとき棄てるのに利用すると都合がよい．しかし資料を棄てる場合，安易に棄てるのではなく，棄てられるその原因を発見しなければならないであろう.

2．　これらの棄却検定法は母集団が正規分布をなすと仮定できる場合に限って適用できるものである．従って時には例えば適合度のχ^2検定法によりその正規性を確かめた上で適用しなければならないこともある．また，自然現象の中には対数用量－反応曲線が正規曲線をなすことがよくあるので，母集団が正規分布とかけ離れている時でも，適当な変数変換を行なって正規分布に近づけてからこれらの方法を用いることもある.

3．　これらの棄却検定法はいずれも1つの異常値に対するものであるから，異常値が2つ以上あった時は，逐次1つづつその資料を棄ててもよいかどうかを決め，その都度，平均値や分散値を計算し直さなければならない.

引用文献

[1] 国沢清典編，確率統計演習2(統計)，培風館，1988.
[2] 石川栄助編著，実務家のための新統計学，槙書店，1994.
[3] 木村都，砂田久一，辻和男共著，生物検定のための応用推計学，広川書店，1997.
[4] W.R.Thompson, Annals of Mathematical Statistics, 6(1935), 214-219.
[5] F.E.Grubbs, Annals of Mathematical Statistics, 21(1950), 27-58.