論文紹介 1

On The use of The Sample Range In an Analogue of Student's t-Test

By Joseph F. Daly
Bureau of Ships, Naoy Department

The Annals of Mathematical Statistics, Vol. 17, No. 1. 1946.

$x_1, x_2, \cdots\cdots, x_N$ を平均 μ, 分散 σ^2 なる正規母集団から抽出された大さ N の任意標本として，μ, σ が未知である場合に平均 μ がある特定の値 μ_0 に等しいと云ふ統計的假説 H_0： $\mu=\mu_0$ を片側の対立假説（例へば $\mu<\mu_0$）に対して検定する場合に普通用ひられるのは Student の t－検定であつて，それは

$$t=\sqrt{N}\,(\bar{x}-\mu_0)\Big/\sqrt{\frac{\sum(x-\bar{x})^2}{N-1}}$$

但し $\bar{x}=\frac{1}{N}\sum x$（標本平均）

なる統計量を用い，その棄却域は $t > t_\varepsilon$ で定められる。

ところが毎日毎日の製品検査のやうな場合には標本標準偏差の計算が大変なので，もっと簡単な方法として，Dodge[2]及び Knudsen [3] は，w を標本の Range として，t の代りに

$$G = \frac{\bar{x} - \mu_0}{w}$$

なる統計量を採用すべきことを提案してゐる。この論文ではこの G－検定の検定力を調べようと云ふのであるが，その結果は $N' = 10$ のときは大体実用上は t－検定と同じであると云ふことになる。

先づ次の補題から証明しよう。

補題．正規分布をする確率変数 x の独立な N 個の測定値 $x_1, x_2, \text{-----}, x_N$ の標本平均 $\bar{x}$，標本の Range w は独立である．

証明．$\mu = 0,\ \sigma^2 = 1$ として一般性を失はない。$\bar{x}$ と $x_j - x_k\ (j < k)$ な $\frac{1}{2}N(N-1)$ 個の差との同時の特性函数は

$$\varphi(t, t_{jk}) = (2\pi)^{-N/2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\sum_j x_j^2 + i\frac{t}{N}\sum_j x_j + i\sum_{j,k} t_{jk}(x_j - x_k)} dx_1 \cdots dx_N$$

ここで $j \geqq k$ なら $t_{jk} \equiv 0$ としておく。

$$\varphi(t,t_{jk}) = e^{-\frac{1}{2}\sum_j\left[\frac{t}{N}+\sum_k(t_{jk}-t_{kj})\right]^2}$$

$$\times(2\pi)^{-N/2}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\sum_j\left\{x_j-i\left[\frac{t}{N}+\sum_k(t_{jk}-t_{kj})\right]\right\}^2}dx_1\cdots dx_N$$

$$= e^{-\frac{1}{2}\sum_j\left[\frac{t}{N}+\sum_k(t_{jk}-t_{kj})\right]^2}$$

$$= e^{-\frac{t^2}{2N}}\cdot e^{-\frac{1}{2}\sum_j\left[\sum_k(t_{jk}-t_{kj})\right]^2}$$

$\bar{x}$ の特性函数 $\varphi_1(t)$ は

$$\varphi_1(t) = e^{-\frac{t^2}{2N}}$$

$x_j - x_k$, $(j<k)$ なる $\frac{1}{2}N(N-1)$ 箇の変数の特性函数は

$$\varphi_2(t_{jk}) = e^{-\frac{1}{2}\sum_j\left[\sum_k(t_{jk}-t_{kj})\right]^2}$$

であるから

$$\varphi(t,t_{jk}) = \varphi_1(t)\cdot\varphi_2(t_{jk})$$

依って，$\bar{x}$ と x_j-x_k, $(j<k)$ は独立である。

ところで，$w=\max|x_j-x_k|$ は x_j-x_k の可測函数であるから，$\bar{x}$ と w とは独立である。

上の補題は，もっと一般化して次の様にも云へる。

補題．$g(x_1,x_2,\cdots\cdots x_N)$ は

$$g(x_1+a,x_2+a,\cdots\cdots,x_N+a)\equiv g(x_1,x_2,\cdots\cdots,x_N)$$

なる性質を有し，$\bar{x}$ と g との同時特性函数

$$\varphi(t,\lambda)=e^{-\frac{t^2}{2N}}\cdot(2\pi)^{-\frac{N}{2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\sum(x-i\frac{t}{N})^2+i\lambda g(x)}dx_1\cdots dx_N$$

$$=\varphi_1(t)\cdot\psi(t,\lambda)$$

とおくとき，$\psi(t,\lambda)$ が t に関して解析的ならば，$\bar{x}$ と g とは独立である。

証明．何者，上式で $t=iNa$ とおくと

$$\psi(iNa,\lambda)=(2\pi)^{-\frac{N}{2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\sum(x+a)^2+i\lambda g(x)}dx_1\cdots dx_N$$

$$=(2\pi)^{-\frac{N}{2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\sum(x+a)^2+i\lambda g(x+a)}dx_1\cdots dx_N$$

$$=(2\pi)^{-\frac{N}{2}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\sum z^2+i\lambda g(z)}dz_1\cdots dz_N$$

$$=\varphi_2(\lambda)$$

依って t－平面上の虚軸上では $\psi(t,\lambda)$ は一定で，従って $\psi(t,\lambda)$ は全平面で一定であるから，$\psi(t,\lambda)$ は λ のみの函数 $\varphi_2(\lambda)$ である．依って $\bar{x}$ と g とは独立である。

$G_\varepsilon>0$ が與えられたとき，$G>G_\varepsilon$ となる確率は

$$P\left\{\frac{\bar{x}-\mu}{w}>G_\varepsilon\right\}=P\left\{\frac{\sqrt{N}(\bar{x}-\mu)/\sigma}{\sqrt{N}\,G_\varepsilon}>\frac{w}{\sigma}\right\}$$

$$= \int_{z=0}^{\infty} \int_{w=0}^{z/\sqrt{N}G_\varepsilon} f(z) h(w)\, dw\, dz$$

但しここに，$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$，又 $h(w)$ は Range w の密度函数である。

Pearson. Hartley [4] の表を用ひる為に

$$P(u) = \int_0^{u\sigma} h(w)\, dw$$

なる記号を用ひると，

$$P\left\{ \frac{\bar{x}-\mu}{w} > G_\varepsilon \right\} = \int_0^{\infty} f(z) P\left(\frac{z}{\sqrt{N}\, G_\varepsilon} \right) dz$$

$\mu = \mu_0$ なる仮説の下で，$G_{.05}$ を計算すると次のやうになる

$$P_N \left\{ \frac{\bar{x}-\mu_0}{w} > G_{.05} \,\middle|\, \mu = \mu_0 \right\} = .05$$

第 I 表

G の分布の右の5%点

N	$G_{.05}$
3	.88
5	.39
7	.26
10	.19

$\mu \neq \mu_0$ の場合に $G > G_\varepsilon$ となる確率を求めるには

$$\frac{\bar{x}-\mu_0}{w} = \frac{\sqrt{N}(\bar{x}-\mu)/\sigma + \sqrt{N}(\mu-\mu_0)/\sigma}{\sqrt{N}\,w/\sigma}$$

$$= \frac{z+a}{\sqrt{N}\,w/\sigma}$$

と書直すと，求める確率は

$$\int_{z=-a}^{\infty} f(z)\,P\left(\frac{z+a}{\sqrt{N}\,G_\varepsilon}\right)dz,\quad a=\sqrt{N}(\mu-\mu_0)/\sigma$$

で與えられる。　$N=3$ として，$\frac{\mu-\mu_0}{\sigma}$ の色々な値に対して $G > G_{.05}$ となる確率と $t > t_{.05}$ となる確率を比較すると次表のやうである。

第　II　表

N＝3のときのG及びtに対する棄却確率

$\frac{\mu-\mu_0}{\sigma}$	$P\{G > .88\}$	$P\{t > 2.92\}$
.00	.050	.050
.50	.151	.151
.75	.229	.230
1.00	.322	.322

$N = 10$ の場合に $\frac{\mu - \mu_0}{\sigma} = .383$（従って $a=1.21$）

のときは

$$P\{G > G_{.05}\} = .296$$

$$P\{t > t_{.05}\} = .30^{+}$$

$N' \equiv 10$ のときは $G_{.05}$ は

$$t_{.05} \times k_N$$

として良く近似される。 但し

$$k_N = \frac{E\left[\sqrt{\frac{\Sigma(x-\bar{x})^2}{N-1}}\right]}{\sqrt{N}\ E[w]}$$

$E[w]$ は Tippett の表［6］から求められる.

第 III 表

G 分布の右の5%点の推定値

N	$G_{.05}$
3	.882
4	.526
5	.385
6	.309
7	.260
8	.227
9	.202
10	.183

+ Neyman, Tokarska (1) 参照

参　考　文　献

[1] J. Neyman and B. Tokarska, "Errors of the second kind in testing 'Student's' hypothesis." Am. Stat. Assn. Jour., Vol. 31 (1936). pp. 318-326.

[2] H. F. Dodge, "Statistical control in sampling inspection." American Machinist. Vol. 76 (1932), p. 1130

[3] Lila F. Knudsen, "A method for determining the significance of a shortage." Am. Stat. Assn. Jour., Vol. 38 (1943), pp. 466-470.

[4] E. S. Pearson and H. D. Hartley, "The probability integral of the range in samples of n observations from a normal population." Biometrika. Vol. 32 (1942), pp. 301-310.

[5] N. L. Johnson and B. L. Welch, "Applications of the non-central distribution," Biometrika, Vol. 31 (1940), pp. 362-389.

[6] L. H. C. Tippett, "On the extreme individuals and the range of samples taken from a normal population." Biometrika, Vol. 17 (1925). pp. 364-387.

（小川潤次郎紹介）