همگنی واریانس در آزمون میانگینهای دو نمونه‌ای

باری ک. موزِر   گری ر. اِستیونز

مترجم: معصومه ذکاوت[۱]

چکیده

در آزمون میانگینهای دو نمونه‌ای، نیاز به یک آزمون واریانس مقدماتی و تأکید خاص بر فرض برابری واریانسها را مورد سؤال قرار می‌دهیم. نتایج بدست آمده، می‌توانند به آسانی در تدریس آزمون میانگینهای دو نمونه‌ای در کلاس مورد استفاده قرار گیرند.

۱ مقدمه

مسأله آزمون برابری میانگینها از دو جامعه مستقل، با توزیع نرمال در هر متن آماری مقدماتی آمده است. تحت فرض تساوی واریانسهای دو جامعه σ_1^2 و σ_2^2 اکثر مؤلفان، آزمون t را توصیه کرده‌اند. اگر واریانسها برابر نباشند، غالباً روشی دیگر همچون روش پیشنهاد شده توسط اسمیت[۲] (۱۹۳۶)، ولش[۳] (۱۹۳۷) و بعداً توسط سترتویت[۴] (۱۹۴۶) توصیه می‌شود. برای بکار بردن این روش نسبت واریانسها، $\theta = \sigma_2^2/\sigma_1^2$ باید معلوم باشد. بهرحال در اغلب کاربردها این نسبت نامعلوم است. در چنین مواردی، بکاربردن آزمون مقدماتی برای فرض $\sigma_1^2 = \sigma_2^2$ یک تمرین متداول است. اگر فرض $\sigma_1^2 = \sigma_2^2$ در آزمون اولیه پذیرفته شد، آنگاه آزمون t برای تساوی میانگینها انجام می‌شود. در غیر اینصورت آزمون اسمیت/ولش/سترویت (SWS) برای تساوی میانگینها انجام می‌شود.

در بسته‌های نرم‌افزاری آماری مانند SPSS , SAS , BMDP آزمون اولیه واریانس با آزمون میانگینهای دو نمونه‌ای ادغام شده‌است. تمام این نرم‌افزارها آزمون میانگینهای دو نمونه‌ای را که آزمون مقدماتی تساوی واریانسها را دربر دارد، شامل هستند. اول آزمون و سپس آزمون تساوی میانگینها با انتخاب یکی از آزمونهای t یا SWS دنبال می‌شود. برای تشریح اینکه چگونه این نرم‌افزارهای آماری استفاده می‌شوند، مثالی که

---

[۱] معصومه ذکاوت، دانشکده علوم ریاضی، دانشگاه شهید بهشتی

[۲] Smith

[۳] Welsh

[۴] Satterthwaite

توسط آرنولد و میلتون [5] (مثال ۱۰.۶.۱ ص.۳۱۲ ، ۱۹۸۶) ارایه شده را بررسی می‌کنیم. در این مثال دو نوع بخاری نفتی آزمایش شده است. مشاهدات، زمان مورد نیاز برای بالابردن دمای اتاق به اندازه ده درجه فارنهایت توسط دو نوع بخاری است. ما لگاریتم زمان را با استفاده از رویه TTEST در SAS تحلیل می‌کنیم. خروجی SAS [6] در جدول ۱ آورده شده است. برای هریک از دو نوع بخاری، روش موجود در SAS ، مقادیر تعداد نمونه، میانگینها، انحراف معیارها، انحراف خطاها ، مینیمم و ماکزیمم را لیست می‌کند. سپس آزمون مساوی بودن واریانسها انجام شده، نسبت F ، درجات آزادی و مقدار P معرفی می‌شوند. آماره‌های SWS و t به ترتیب با تیترهای سطری واریانسهای «نامساوی» و «مساوی» لیست شده‌اند.

بنابراین ابتدا کاربر، برابری واریانسها را چک می‌کند. اگر برابری پذیرفته شد، اقدام به آزمون t می‌کند و اگر فرض برابری واریانسها پذیرفته نشد از آزمون SWS استفاده می‌کند. از این بحث آشکار می‌شود که در آزمون میانگینهای دو نمونه‌ای تأکید بسیار بر فرض برابری واریانسها است. مؤلفان متن‌های آماری مقدماتی بر اهمیت این فرض به قدری تأکید می‌کنند که حوزه متنهای درسی را به دو بخش تقسیم کرده‌اند، در یک بخش آزمون t ، در صورتیکه برابری واریانسها مفروض باشد، ارایه می‌شود و در بخش دیگر آزمون SWS ، در صورتیکه برابری واریانسها نقض شده باشد، ارایه می‌گردد. بطور مشابه، نرم‌افزارها آماری رایج آزمونهای مقدماتی واریانس را در برنامه آزمون میانگینهای دو نمونه‌ای خود درج کرده‌اند. سؤالهای زیر مطرح می‌شود: آیا تمرین رایج بر آزمونهای مقدماتی واریانس مناسب است؟ و آیا تأکید بر همگنی واریانس دلیلی دارد؟

هدف این مقاله پاسخ به این سؤالات است. تا پایان این بخش، ارتباط بین آزمون t و آزمون SWS وقتی انتخاب بین این دو روش بر اساس آزمون واریانس مقدماتی است، بررسی می‌شود. در بخشهای بعدی، مسأله بطور مختصر شرح داده شده است، اندازه‌ها و توانهای آزمونهای میانگین خلاصه شده‌اند و دو پرسش ذکر شده فوق، جواب داده‌شده‌اند.

۲ مسأله

$X_1,\ldots,X_{n_1}$ و $Y_1,\ldots,Y_{n_2}$ را به عنوان دو نمونه تصادفی از دو جامعه با توزیع نرمال در نظر بگیرید بطوریکه برای $i=1,2,\ldots,n_1$ و $j=1,2,\ldots,n_2$ ؛
$x_i \sim N(\mu_1,\sigma_1^2)$ ، $y_j \sim N(\mu_2,\sigma_2^2)$ باشند.

برای آزمون مقدماتی فرض $H_0 : \sigma_1^2 = \sigma_2^2$ در مقابل $H_1 : \sigma_1^2 \neq \sigma_2^2$ آماره

$$F' = s_2^2/s_1^2 \qquad (1)$$

باید محاسبه شود.

اگر $F' > F^{\alpha/2}_{n_2-1,n_1-1}$ یا $F' < F^{1-\alpha/2}_{n_2-1,n_1-1}$ باشد، H_0 رد می‌شود که در آن α سطح معنی‌داری آزمون واریانس است، $F^{\alpha^*}_{(a,b)}$ صدک $100(1-\alpha^*)$ از توزیع F با درجات آزادی a و b است،

$$\bar{x} = \sum_{i=1}^{n_1} \frac{x_i}{n_1} \quad , \quad \bar{y} = \sum_{j=1}^{n_2} \frac{y_j}{n_2}$$

$$s_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_i-\bar{x})^2 \quad , \quad s_2^2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(y_j-\bar{y})^2$$

است. معیار رد بالا معادل معیار مقدار P است وقتی که H_0 برای تمام مقادیر P کمتر از α، رد شود.

اگر H_0 رد نشود، آنگاه در آزمون t-test برای $H_0^* : \mu_1 = \mu_2$ در مقابل $H_1^* : \mu_1 \neq \mu_2$ (یا در مقابل $H_2^* : \mu_1 > \mu_2$) باید

$$t^* = \frac{(\bar{x}-\bar{y})}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \qquad (2)$$

محاسبه شود و اگر $t^{*2} > F^{\delta}_{1,n_1+n_2-2}$ (یا اگر $t^* > t^{\delta}_{n_1+n_2-2}$)، H_0^* رد می‌شود، که در آن δ سطح معنی‌داری

---
[5] Arnold and Milton

[6] یکی از نرم‌افزارهای مورد استفاده در تجزیه و تحلیل داده‌های آماری

از پیش تعیین شده برای آزمون میانگینهای دو نمونه‌ای است و

$$s_p^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2]/(n_1 + n_2 - 2)$$

اگر H_0 رد شود، آنگاه معیار SWS برای آزمون H_0^* در مقابل H_1^* (یا H_2^*) عبارت است از محاسبه

$$t^{**} = \frac{(\bar{x} - \bar{y})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (3)$$

اگر $t^{**2} > F^{\delta}_{1,\nu}$ (یا اگر $t^{**} > t^{\delta}_{\nu}$)، H_0^* با همان سطح معنی‌داری از قبل تعیین شده δ و با درجه آزادی زیر رد می‌شود

$$\nu = \frac{(1/n_1 + u/n_2)^2}{1/n_1^2(N_1 - 1) + u_2^2/n_2^2(n_2 - 1)} \qquad (4)$$

که در آن $u = s_2^2/s_1^2$ است.

لازم به تذکر است که اگر سطح معنی‌داری از قبل تعیین شده آزمون واریانس، α، به صفر نزدیک باشد، H_0 پذیرفته شده (رد نمی‌شود) و همیشه آزمون t (۲) انجام می‌شود. بطور مشابه اگر α نزدیک یک در نظر گرفته شود، همیشه تست SWS (۳) برای H_0^* انجام می‌شود. بنابراین سطح معنی‌داری صفر (یا یک) معادل انجام آزمون t (یا SWS) بدون انجام آزمون مقدماتی برای H_0 است. به هدف شناسایی، آزمون مقدماتی برای H_0 با $1 > \alpha > 0$ که با آزمون t یا SWS برای H_0^* دنبال می‌شود، آزمون گاهگاهی t(ST[7]) اطلاق می‌شود. این آزمون برای $\alpha = 0$ منجر به آزمون همیشگی T(AT[8]) و برای $\alpha = 1$ منجر به آزمون همیشگی SWS(ASWS) می‌شود.

۳ اندازه و توان آزمون

احتمال رد $H_0^* : \mu_1 = \mu_2$ به نفع $H_1^* : \mu_1 \neq \mu_2$ یا $H_2^* : \mu_1 > \mu_2$ تابعی از α, δ, n_2, n_1 و

$$\lambda = (\mu_2 - \mu_1)^2/2(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

است. وقتی $\lambda = 0$ است احتمال رد H_0^* را اندازه آزمون گویند. توان آزمون برابر با احتمال رد H_0^* برای هر $\lambda > 0$ است. در هر حالت احتمال رد H_0^* مجموع دو شق مجزا است.

$$Pr(\text{رد } H_0^*) = Pr(\text{رد } H_0^*, \text{رد } H_0) + Pr(\text{رد } H_0^*, \text{پذیرش } H_0)$$

اندازه‌ها و توانها برای ترکیبات مختلفی از $\lambda, n_1, n_2, \theta, \delta, \alpha$ بوسیله موزر [9]، استیونز [10] و واتس [11] (۱۹۸۹) محاسبه شده‌اند. نتایج آنها به صورت زیر خلاصه شده‌است.

الف) اگر حجم نمونه‌ها برابر باشد، آزمونهای ASWS ، ST و AT برای هر n_1 و n_2 و $0 < \lambda$ و $10 \leq \theta \leq 1$ تقریباً اندازه‌ها و توانهای مساوی دارند.

ب) اگر حجم نمونه‌ها نابرابر باشد، اما نسبت واریانس (θ) نزدیک یک باشد، در حین نگهداشتن سطح معنی‌داری تعیینی نزدیک δ، آزمون AT بیشترین توان را بدست می آورد.

ج) اگر حجم نمونه‌ها نابرابر و θ نزدیک یک نباشد و کمترین واریانس مربوط به بیشترین حجم نمونه باشد، آزمون AT هنوز بیشترین توان را برای همه $0 < \lambda$ دارد و بعد از آن آزمونهای ST و ASWS بیشترین توان را دارند. به هر حال آزمون AT این توان زیاد را با مقدار بزرگی از اندازه آزمون بدست می‌آورد. برای مقدارهای معینی از θ، آزمون θ توان بیشتری را با اندازه آزمون زیاد فراهم می‌آورد در هر صورت آزمون ASWS برای همه $10 \leq \theta \leq 1$ سطح آزمون معقولی را نزدیک δ حفظ کرده‌است.

د)اگر حجم نمونه‌ها برابر نباشد و θ نزدیک یک نبوده و واریانس کمتر مربوط به حجم نمونه کمتر باشد، در اینصورت آزمون ASWS بیشترین توان را دارد و بعد از آن به ترتیب آزمونهای ST و AT، بیشترین توان را دارند. آزمون ASWS، به حفظ سطح معنی‌داری قابل قبول نزدیک δ برای همه $10 \leq \theta \leq 1$ ادامه می‌دهد.

---

[7] Sometimes t
[8] Always t
[9] Moser
[10] Stevens
[11] Watts

۴ بحث و پیشنهاد

اکنون ما سؤالی را مطرح می‌کنیم: آیا تمرین رایج آزمون مقدماتی واریانس مناسب است؟ جواب خیر است. ما پاسخ را در بندهای بعد توجیه می‌کنیم.

هر وقت اندازه نمونه‌ها نابرابر و نسبت واریانس نزدیک یک باشد، آزمون t در حین حفظ اندازه نزدیک سطح تعیینی δ،بیشترین توان را فراهم می‌آورد. بنابراین اگر حجم نمونه‌ها نابرابر و نسبت واریانس معلوم باشد که نزدیک یک است، آزمون t مناسب می‌باشد.

آزمونهای SWS و ST و t وقتی حجم نمونه‌ها برابر باشد، همگی اندازه و توان یکسانی دارند. بنابراین آزمون مقدماتی واریانس غیر ضروری است، بطوریکه به راحتی یک گام آماری زائد را ایجاد می‌کند. بنابراین در مورد حجم نمونه‌های برابر، هریک از دو آزمون t یا SWS مناسب است.

تنها شرایط باقی‌مانده زمانی اتفاق می‌افتد که حجم نمونه‌ها نابرابر و نسبت واریانس نامعلوم باشد و یا معلوم باشد که یک نیست. هر وقت حجم نمونه‌ها نامساوی است، آزمون SWS در حین نگهداشتن اندازه، نزدیک δ ، توان خوبی را فراهم می‌آورد. به هر حال ، هر دو آزمون ST و t زمانی که نسبت واریانس یک نیست، می‌توانند اندازه‌های بزرگی را تولید کنند. بنابراین در این حالت آزمون SWS مناسب است.

توجه کنید که آزمون ST با تست واریانس مقدماتی هرگز توصیه نمی‌شود. برای همه مقادیر نسبت واریانس، هیچ ترکیبی از اندازه نمونه‌ها وجود ندارد که در آنجا ST به خوبی عمل کند در حالی که SWS و t ضعیف عمل می‌کنند. این ناشی از این واقعیت است که اندازه (توان) آزمون ST متوسط وزنی اندازه‌ها (توانهای) آزمونهای SWS و t است. بنابراین آزمون مقدماتی (۱) هرگز برای این مسأله مناسب نیست.

بالاخره آیا تأکید بر همگنی واریانس دلیلی دارد؟ دوباره پاسخ خیر می‌باشد. نسبت واریانس، فقط وقتی معلوم باشد، روی انتخاب مناسب روش آزمون تأثیر می‌گذارد. به هر حال، در عمل، نسبت واریانس، به ندرت معلوم است. از بحث فوق در می‌یابیم که آزمون SWS در همه مواردی که واریانس نامعلوم باشد، مناسب است. بنابراین هنگام تدریس آزمون میانگینهای دو نمونه‌ای، با تأکید کمتر بر فرض همگنی واریانس، باید تلاش بیشتر صرف آموزش کیفیتهای آزمون SWS شود.

جدول ۱: خروجی SAS Proc TTEST برای مثال بخاری نفتی

روش آزمون

بخاری	N	میانگین	انحراف استاندارد	خطای استاندارد	مینیمم	ماکزیمم	واریانسها	T	DF	prob > $\lvert T \rvert$
B	۱۲	۳/۶۸۵	۰/۴۸۷	۰/۱۴۱	۲/۶۲	۴/۲۴	نامساوی	۱/۹۳	۱۶/۴	۰/۰۷۱۰
A	۱۵	۳/۳۸۱	۰/۲۷۳	۰/۰۷۱	۲/۶۳	۳/۷۳	مساوی	۲/۰۵	۲۵	۰/۰۵۰۸

مراجع

[1] Milton, J. S., and Arnold, J. C.(1986), *Probability and Statistics in the Engineering and Computing Sciences,* New York: McGraw-Hill.

[2] Moser, B. k., Stevens, G. R., and Watts, C. L. (1989), "The Two Sample tTest Versus Satterthwaite's Approximate FTest," *Communications in Statistics-Theory and Methods, 18.*

[3] Satterthwaite, F. E. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin, 2, 110-114.*

[4] Smith, H. F. (1936), "The Problem of Comparing the Results of Two Experiments With Unequal Errors," *Journal of the Council for Scientific and Industrial Research, 9, 211-212.*

[5] Welch, B. L. (1937), "The Significance of the Difference Between Two Means When the Population Variances are Unequal," *Biometrika, 29, 350-362.*

اصل این مقاله با عنوان "Homogeneity of Variance in the Two-Sample Means Test"

نوشته Stevens, G. R. و Moser, B. k. است که در

The American Statistician, February 1992, VOl. 46, No.1

چاپ شده است.
