

Pengategorian Data dengan Menggunakan Statistik Hipotetik dan Statistik Empirik

Wahyu Widhiarso

(Fakultas Psikologi Universitas Gadjah Mada)

Di dalam pengembangan alat ukur ada dua pendekatan dalam menafsirkan skor, yaitu pendekatan dengan menggunakan referensi berupa norma (norma referenced) maupun referensi berupa kriteria (criterion referenced; Kaplan & Saccuzzo, 2009). Kedua pendekatan ini memiliki kegunaan yang berbeda yang implikasinya juga berbeda. Mengembangkan alat ukur berbasis norma sedikit berbeda dengan alat ukur yang berbasis kriteria. Contoh tes berbasis norma adalah tes inteligensi. Tes inteligensi yang baik harus mampu membedakan individu dengan baik, baik membedakan antar individu yang memiliki inteligensi yang rendah maupun yang tinggi.

Alat ukur berbasis pada norma berusaha untuk membedakan antara orang dengan orang lainnya dalam populasi. Semakin alat ukur ini mampu membedakan orang dengan baik semakin dia menjalankan fungsinya. Dalam pendekatan ini semua orang harus kelihatan berbeda, hingga nilai di kelas pun harus berbeda. Di dalam kelas harus ada siswa yang memiliki nilai A, B, C, D dan E karena ini merepresentasikan adanya perbedaan individual. Dalam tes ini kategori tinggi dan rendah ditentukan oleh posisi relatifnya di dalam populasi. Meskipun tergolong kurus (berat badan 50 kg), anda terlihat paling gemuk jika di populasi kebanyakan orang-orang ternyata lebih kurus dari anda.

Alat ukur berbasis pada kriteria membedakan orang berdasarkan performansi atau penguasaan individu berdasarkan kriteria-kriteria yang di tentukan. Tes ini berusaha membedakan individu yang telah memenuhi kriteria maupun tidak memenuhi kriteria. Melalui tes ini semua siswa di dalam kelas dapat memperoleh nilai A jika semuanya telah mendemonstrasikan penguasaan atau performansi yang melebihi kriteria yang ditentukan. Pada tes ini tinggi rendah sebuah skor ditentukan oleh titik potong (*cut off*). Berat badan anda yang sebesar 50 kg itu terlihat kurus karena batas (atau kriteria) berat badan orang yang gemuk adalah 80 kg.

Dampak Penggunaan Referensi Sebuah Tes

Pengembangan tes dengan menggunakan norma seringkali melibatkan individu berukuran besar karena tes ini memerlukan norma tes untuk memaknai skor yang dihasilkan dari tes. Misalnya, untuk memaknai apakah skor tes kreativitas si Badu sebesar 50 itu termasuk skor yang tinggi atau tidak, pengembang tes harus mengadministrasikan tes pada sampel dengan ukuran yang besar. Sebagai catatan proses penormaan pada tes Woodcock-Johnson IV menggunakan ukuran sampel lebih dari 7.000 orang (McGrew, LaForte, & Schrank, 2014). Dari tabel penormaan ini kita akan mendapatkan informasi mengenai posisi skor tes kreativitas sebesar 50 itu masuk ke dalam kategori rendah, sedang atau tinggi. Biasanya penentuan ini didasarkan pada plus-minus simpangan baku skor dari rerata. Misalnya jika rerata di populasi adalah $M = 70$ dan deviasi standarnya adalah $SD = 10$, maka batas skor dengan kategori rendah adalah $X < 70 - 10$ atau $X <$

60. Dari informasi ini dapat disimpulkan bahwa skor Badu ($X = 50$) termasuk dalam kategori rendah.

Dua Strategi Pengategorian Data

Beberapa peneliti terkadang melakukan kategorisasi pada data skor hasil pengukuran mereka dengan cara membagi kategori skor menjadi rendah, tinggi rendah. Tujuannya bermacam macam, antara lain untuk mempresentasikan distribusi skor subjek penelitian secara umum, melihat kecenderungan skor subjek atau membandingkan skor antara subjek. Ada pengategorian yang menggunakan bahan berupa statistik berasal dari data empirik (**statistik empirik**). Pada teknik ini rerata dan deviasi standar yang didapatkan dari data empirik dipakai sebagai referensi. Sama seperti contoh di atas, jika rerata dan deviasi di dalam populasi adalah $M = 50$ dan $SD = 5$, maka skor Anda (misalnya $X = 60$) masuk dalam kategori tinggi (atau di atas rerata) karena melebihi $M + 1SD$ atau $X > 55$.

Sebagai alternatif, kita dapat menggunakan statistik dari alat ukur kita (**statistik hipotetik**). Dasar teori ini dapat adalah tulisan Azwar (1993). Pada teknik ini, rerata dan deviasi standar yang dipakai sebagai bahan penyusunan titik kategori didapatkan dari alat ukur. Misalnya anda memiliki skala berisi 25 butir yang menggunakan format Likert dengan 4 opsi yang diskor (0 hingga 4). Berdasarkan informasi ini didapatkan rerata skor hipotetik skala adalah $M = 50$. Didapatkan dari nilai tengah skor maksimal yang dapat diraih subjek pada skala, yaitu $X = 100$ (25 butir x 4), dengan skor minimal yang dapat diraih subjek, yaitu $X = 0$ (25 butir x 0). Dari informasi ini juga didapatkan rentang hipotetik skor diperoleh subjek dari $100 - 0 = 100$ (skor maksimal – skor minimal), sehingga deviasi standar (SD) skor skala adalah $100 : 6 = 16,67$. Berdasarkan informasi mengenai rerata dan deviasi standar yang sudah ditemukan ini maka didapatkan bahwa batas skor yang termasuk di di atas rerata adalah $X = 50 + 16,67 = 66,67$; Berdasarkan hal ini maka skor Anda (misalnya 60) masuk dalam kategori rata-rata karena kriteria di atas rata-rata adalah $X > 66,67$. Penjelasan mengenai prosedur ini dapat dilihat di Widhiarso (2010)

Perbedaan Kedua Strategi

Berdasarkan contoh di atas, pengategorian data dapat dibagi menjadi dua yaitu pengategorian berdasarkan statistik empirik dan statistik hipotetik. Berikut ini perbedaan antara keduanya.

- Pengategorian dengan menggunakan statistik empirik analog dengan pengembangan tes dengan menggunakan referensi norma sedangkan pengategorian dengan menggunakan alat ukur analog dengan pengembangan tes dengan menggunakan referensi kriteria.
- Oleh karena menggunakan referensi berupa norma, maka kategorisasi berdasarkan statistik empirik akan selalu menghasilkan proporsi kategori yang menggambarkan kurva normal (lihat Tabel 1), kategori di tengah (misalnya sedang) persentasenya selalu lebih besar dibanding dengan kategori di ujung (misalnya rendah atau tinggi).

Tabel 1. Dua contoh hasil kategorisasi berdasarkan statistik empirik

Kategori	Frekuensi	Persentase	Kategori	Frekuensi	Persentase
Rendah	59	27%	Rendah	45	15%
Sedang	102	47%	Sedang	201	69%
Tinggi	57	26%	Tinggi	47	16%

- c) Pengategorian dengan menggunakan statistik hipotetik dapat menghasilkan proporsi yang tidak selalu mengikuti kurva normal (lihat Tabel 2).

Tabel 2. Dua contoh hasil kategorisasi berdasarkan statistik hipotetik

Kategori	Frekuensi	Persentase	Kategori	Frekuensi	Persentase
Rendah	10	11%	Rendah	98	62%
Sedang	25	27%	Sedang	5	3%
Tinggi	56	62%	Tinggi	56	35%

Syarat Penggunaan Strategi

Setiap strategi memiliki persyaratan yang berbeda untuk dipakai. Berikut ini persyaratannya.

A. Penggunaan Statistika Empirik

Penggunaan statistika empirik menggunakan acuan pada subjek di populasi karena tinggi rendahnya sebuah makna skor tergantung dari populasi. Sama dengan halnya proses penormaan yang membutuhkan ukuran sampel besar dan heterogen, maka penggunaan statistik empirik dalam kategorisasi harus menggunakan ukuran sampel yang besar. Kita tidak dapat mengatakan skor Budi ($X = 20$) itu tergolong sangat tinggi jika skor tersebut melebihi rata-rata belasan orang. Untuk mengatakan skor Budi sangat tinggi kita harus mendapatkan informasi dari ribuan orang.

B. Penggunaan Statistik Hipotetik

Penggunaan statistika hipotetik menggunakan alat ukur sebagai acuan. Tinggi rendahnya skor subjek tergantung dari posisinya pada rentang skor yang memungkinkan diperoleh pada sebuah alat ukur. Skor $X = 9$ termasuk kategori sangat tinggi jika didapatkan dari hasil pengukuran oleh skala 10 butir dengan opsi ($Ya = \text{skor } 1 \ \& \ \text{Tidak} = \text{skor } 0$), namun skor $X = 9$ itu termasuk kategori sangat rendah jika didapatkan dari skala yang berisi 10 butir dengan empat opsi respons (*sangat sesuai* hingga *sangat tidak sesuai*) dengan skor butir antara 0 hingga 5. Penggunaan statistik ini mensyaratkan alat ukur yang dipakai adalah alat ukur yang sudah terstandar yang telah divalidasi melalui banyak penelitian.

Referensi

- Azwar, S. (1993). "Kelompok subjek ini memiliki harga diri yang rendah"; kok, tahu...? *Buletin Psikologi*, 1(2), 13-17.
- Kaplan, R. M., & Saccuzzo, D. P. (2009). *Psychological Testing Principles, Applications, and Issues*. Belmont, CA: Wadsworth, Cengage Learning.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Woodcock-Johnson IV Technical Manual*. Rolling Meadows, IL: Riverside.
- Widhiarso, W. (2010). *Membuat Kategori Skor Hasil Pengukuran dari Skala*. Fakultas Psikologi. Universitas Gadjah Mada. Yogyakarta.