


Workshop Latent Variable Modeling (LVM)

Menggunakan JASP: Bagian 2 - Korelasi

Rizqy Amelia Zein

- Dosen, Fakultas Psikologi, Universitas Airlangga
- Anggota, #SainsTerbuka Airlangga 
- Relawan, INA-Rxiv
- Researcher-in-training, Institute for Globally Distributed Open Research and Education (IGDORE)

Mengapa memulai dari korelasi?

LVM-SEM merupakan teknik yang digunakan untuk mengestimasi korelasi antar-variabel

Untuk melakukan SEM, peneliti tidak harus menyediakan data kasar (*raw data*), tetapi ada pilihan untuk menginput *correlation* atau *variance-covariance matrix*.



Jenis-jenis korelasi

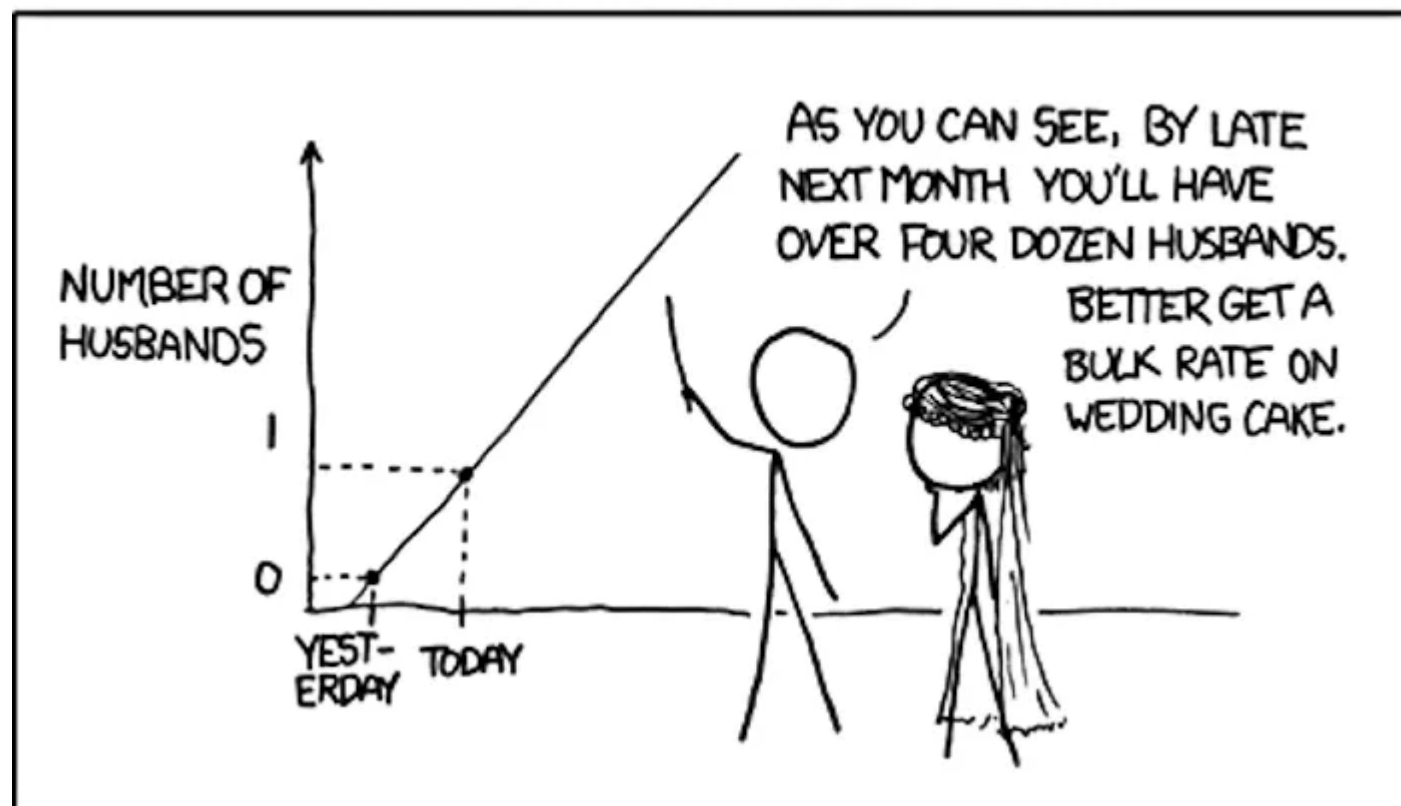
Koefisien Korelasi	Level Pengukuran
<i>Pearson's product moment</i>	Kedua variabel setidaknya interval
<i>Spearman's rank</i> dan <i>Kendall's tau</i>	Kedua variabel ordinal
<i>Phi, contingency table</i>	Kedua variabel nominal
<i>Point biserial</i>	Variabel interval dengan nominal
<i>Gamma, rank biserial</i>	Variabel ordinal dengan nominal
<i>Biserial</i>	Variabel interval dengan <i>dummy</i>
<i>Polyserial</i>	Variabel interval dengan variabel <i>underlying continuity</i>
<i>Tetrachoric</i>	Kedua variabel <i>dummy</i> (dikotomis)
<i>Polychoric</i>	Kedua variabel ordinal (dengan kontinuitas implisit)

Faktor-faktor yang mempengaruhi korelasi

- **Level pengukuran** (apakah variabel tersebut nominal, ordinal, interval, atau rasio)
 - Sehingga berdampak pada **variabilitas** (*restriction range*) dan **normalitas data**
- **Linearitas**
 - Semua teknik korelasi mengasumsikan korelasi antar-variabel linear, sehingga korelasi yang tidak linear akan memberikan informasi **tidak adanya korelasi** (padahal tidak selalu).
- **Adanya data *outlier***
- **Koreksi atenuasi**
- **Jumlah sampel**
 - Jumlah sampel yang terlalu sedikit akan memberikan estimasi yang kurang akurat (karena *standard error*nya besar)
- ***Sampling variance***
 - Yang kemudian berefek pada *confidence interval*, *effect size* (koefisien korelasi itu sendiri), dan *statistical power*
- ***Missing data***
 - Kalau data tidak lengkap, estimasi koefisien korelasi akan langsung terdampak.
 - Ada beberapa pilihan yang bisa dilakukan, yaitu ➡ *listwise deletion*, *pairwise deletion*, dan melakukan *data imputation*.
 - *Listwise deletion* tidak disarankan karena membuat jumlah sampel turun drastis ➡ mengurangi *statistical power*.

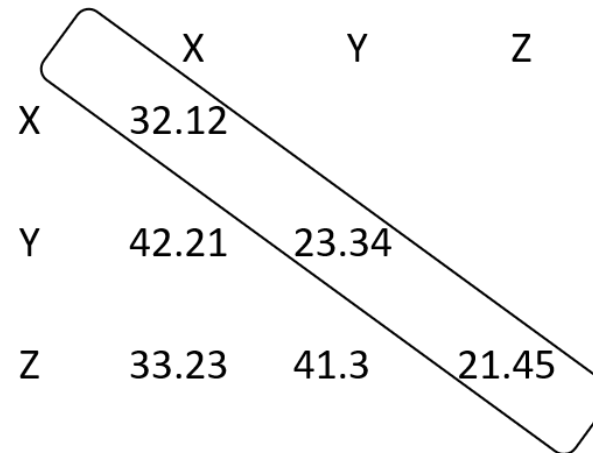
Silahkan unduh dan buka **Dataset Contoh Korelasi**, untuk melihat contoh.

MY HOBBY: EXTRAPOLATING



Variance-covariance dan correlation matrix (1)

- Untuk melakukan SEM, maka perangkat lunak membutuhkan *variance-covariance matrix* untuk mengestimasi parameter model
- Pada bagian diagonal *variance-covariance matrix* menunjukkan varians, sedangkan sisanya adalah *covariance*



	X	Y	Z
X	32.12		
Y	42.21	23.34	
Z	33.23	41.3	21.45

Varians
..dan sisanya adalah *covariance*

- Jumlah nilai unik (*non-redundant information*) dalam *variance-covariance matrix* adalah $p(p+1)/2$
 - dimana p adalah jumlah *observed variable*
 - Sehingga dengan contoh di atas maka jumlah nilai unik (*non-redundant information*) di *variance-covariance matrix* adalah $3(3+1)/2=6$, yaitu **3 varians (diagonal)** dan **3 covariance (sisanya)**

Variance-covariance dan correlation matrix (2)

- Sebagian besar perangkat lunak SEM menggunakan **variance-covariance matrix** bukan *correlation matrix*
 - Ingat ! korelasi adalah *standardised covariance*.
- Menggunakan *correlation matrix* biasanya lebih sering menghasilkan **parameter yang statistically significant** tapi *standard error* yang tidak akurat.
- Oleh karena itu, meskipun *user* menginput *correlation matrix*, maka perangkat lunak akan mengubahnya dulu menjadi *variance-covariance matrix* dulu, baru parameter model dapat diestimasi.



Koreksi Atenuasi

- Asumsi dasar dalam Psikometri adalah skor kasar (*observed score*) mengandung skor murni (*true score*) dan *measurement error*, sehingga dalam mengestimasi korelasi, *measurement error* perlu dibuang agar estimasi lebih akurat.
- Dengan teknik *koreksi atenuasi*, kita dapat 'membuang' *measurement error*, sehingga kita dapat mengestimasi korelasi antar-variabel menggunakan *true score*-nya.
- Tetapi apabila reliabilitas skala kita kurang baik, maka setelah dikoreksi **koefisien korelasi bisa lebih dari 1** !
- Misalnya diketahui bahwa korelasi *observed scores* antar dua variabel (r_{ab}) adalah 0.9 dan reliabilitas skala a (Cronbach's α) adalah 0.6 dan skala b adalah 0.7, maka:

$$r_{ab}^* = \frac{r_{ab}}{\sqrt{r_{aa} * r_{bb}}}$$

$$r_{ab}^* = \frac{0.9}{\sqrt{0.7(0.6)}}$$

$$r_{ab}^* = 1.389$$

WARNING! Covariance/correlation matrix is not positive definite



Apa yang terjadi?

Perangkat lunak akan menghentikan proses estimasi

...dan memberikan pesan *non-positive definite*



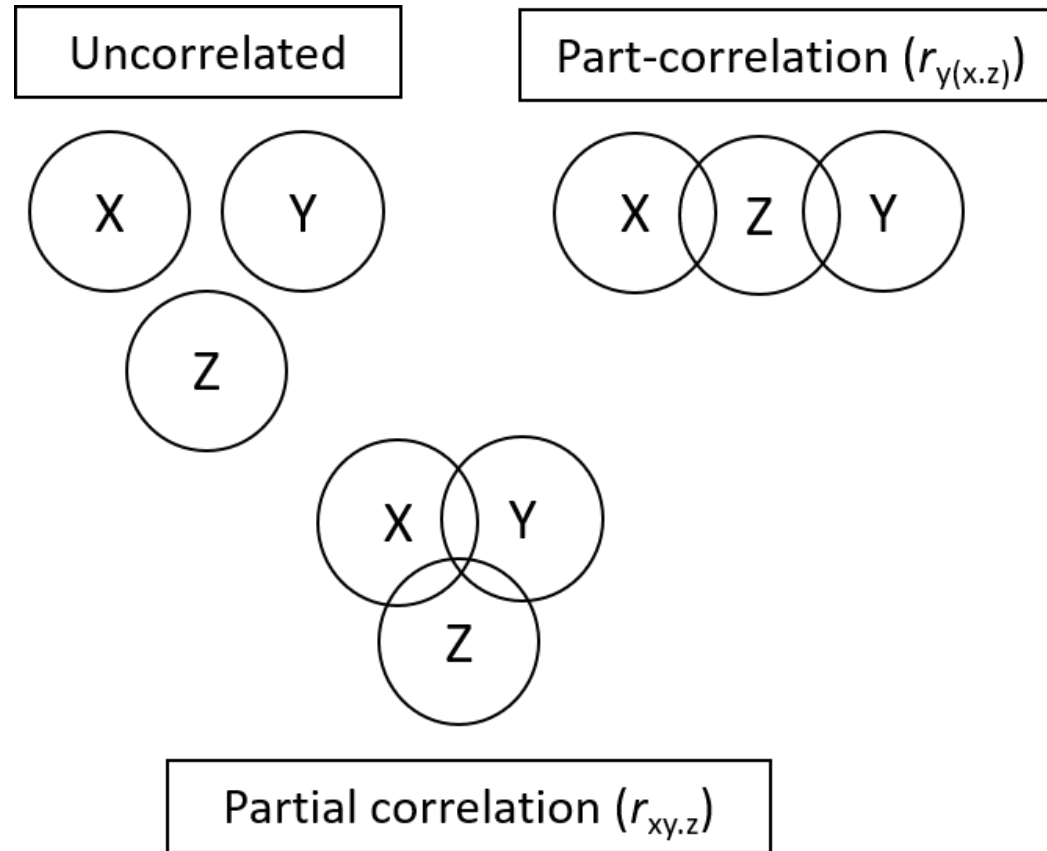
Matrik korelasi dengan *non-positive definite*

- Koefisien korelasi yang nilainya ≥ 1 menyebabkan matriks korelasi menjadi *non-positive definite*
 - Artinya, parameter model tidak mungkin diestimasi
- Mengapa terjadi?
 - Data didapatkan dari observasi yang **tidak independen** (*linear dependency*)
 - Terjadi **multikolinearitas**
 - **Jumlah sampel** lebih **sedikit** dari jumlah variabel yang diuji dalam model
 - Sepasang variabel berbagi **varians negatif** atau **tidak sama sekali** (0) ➔ *Heywood case*
 - Varians, kovarians, dan korelasi nilainya diluar batas kewajaran
 - Kesalahan mengatur pembatasan (*constraint*) pada parameter tertentu yang dilakukan oleh peneliti (*user-specified model*)

Heywood dan ultra-Heywood case

- Terjadi ketika *communalities* = 1 (Heywood) atau ≥ 1 (ultra-Heywood), atau terjadi ketika varians *measurement error* bernilai negatif
 - *Communalities* adalah kuadrat dari koefisien korelasi (R^2)
 - Apabila terjadi, maka ada yang salah dengan spesifikasi model (hipotesis)
- Terjadi karena
 - *Common factor* terlalu banyak/terlalu sedikit
 - Ukuran sampel tidak memadai
 - Model SEM (*common factor model*) bukan model yang cocok untuk menguji hipotesis hubungan antar-variabel (alternatifnya *Principal Component Analysis - PCA*)
- Yang bisa dilakukan
 - Tinjau kembali hipotesis modelnya
 - Kurangi jumlah faktor laten dengan 'membuang' jalur/korelasi yang bermasalah
 - Identifikasi variabel yang terlibat multikolinearitas. Masukkan salah satu saja dalam model, sisanya sisihkan

Korelasi Bivariat: *Part* dan *partial correlation*



Metrik variabel (*standardised vs unstandardised*)

- *Unstandardised solution/estimates*
 - Dapat **dibandingkan** antar kelompok sampel
 - Merupakan parameter yang digunakan oleh perangkat lunak untuk menghitung *standard error* dan *taraf signifikansi (p-value)*
 - Membandingkan *unstandardised factor loading* **harus melihat *standard error*nya juga** karena mereka seharusnya sepaket
- *Standardised solution/estimates*
 - Hanya *interpretable* untuk kelompok sampel yang diuji - tidak bisa dibandingkan dengan kelompok sampel yang lain.
 - Berguna untuk membandingkan *factor loading* antar-variabel di dalam model - peneliti dapat **mengidentifikasi variabel mana yang paling berkontribusi** menjelaskan *dependent variable*
 - Apabila variabel dalam model memiliki unit pengukuran yang berbeda, maka *standardised estimates* akan sangat membantu untuk menstandarisasi unit antar-variabel tersebut
- Ada banyak perbedaan pendapat mengenai metrik mana yang harus dilaporkan, tetapi...
 - **Selalu laporkan *unstandardised solution/estimates* dan *standard error*nya** (boleh juga dengan *standardised solution/estimates*, boleh juga tidak)

Terima kasih banyak! 😊



Paparan disusun dengan menggunakan  *package* **xaringan** dengan *template* dan *fonts* dari **R-Ladies**.

Chakra dibuat dengan **remark.js**, **knitr**, dan **R Markdown**.