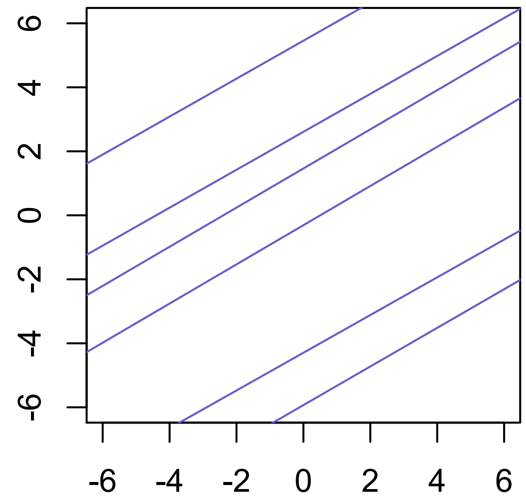
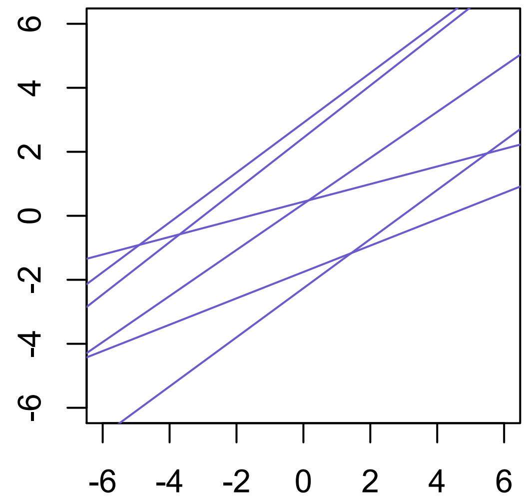
Kinds of varying effects

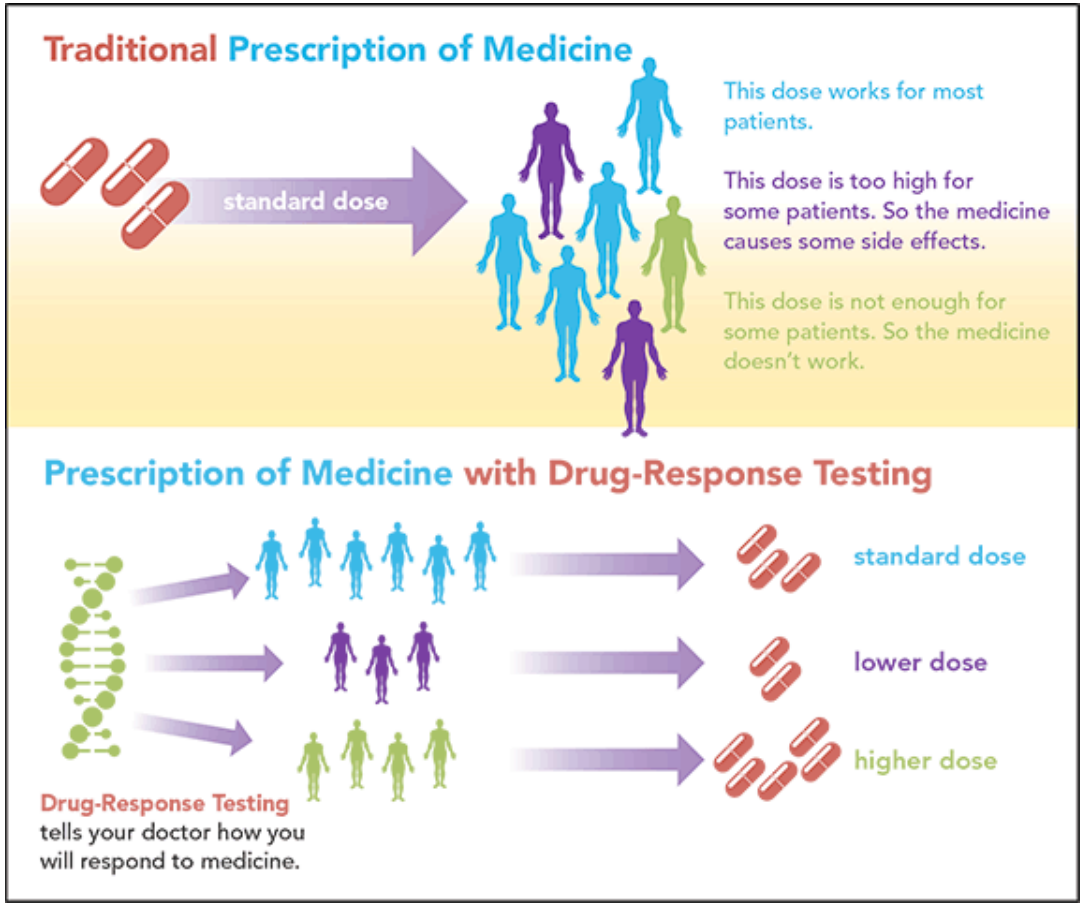
- *Varying intercepts*: means differ by cluster
- *Varying slopes*: effects of predictors vary by cluster
- Any parameter can be made into a varying effect
  - (1) split into vector of parameters by cluster
  - (2) define population distribution





Varying slopes

- Why varying slopes?
  - drugs affect people differently
  - after school programs don't work for everyone
  - not every unit has same relationship to predictor
  - variation is important, whether for intervention or inference
- *Average* effect misleading?
- Pooling, shrinkage, mnesia



Population pooling

- Major innovation with varying slopes is **pooling across parameters (intercepts & slopes)**
- Features of units have correlation structure
- Learn one feature —> info about other features
- e.g. Higher intercepts associated with smaller slopes

Varying slopes model

$$
\begin{aligned}
W_i &\sim \text{Normal}(\mu_i, \sigma) \\
\mu_i &= \alpha_{\text{CAFÉ}[i]} + \beta_{\text{CAFÉ}[i]} A_i \\
\begin{bmatrix} \alpha_{\text{CAFÉ}} \\ \beta_{\text{CAFÉ}} \end{bmatrix} &\sim \text{MVNormal}\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \mathbf{S} \right) \\
\mathbf{S} &= \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \\
\alpha &\sim \text{Normal}(5, 2) \\
\beta &\sim \text{Normal}(-1, 0.5) \\
\sigma &\sim \text{Exponential}(1) \\
\sigma_\alpha &\sim \text{Exponential}(1) \\
\sigma_\beta &\sim \text{Exponential}(1) \\
\mathbf{R} &\sim \text{LKJcorr}(2)
\end{aligned}
$$

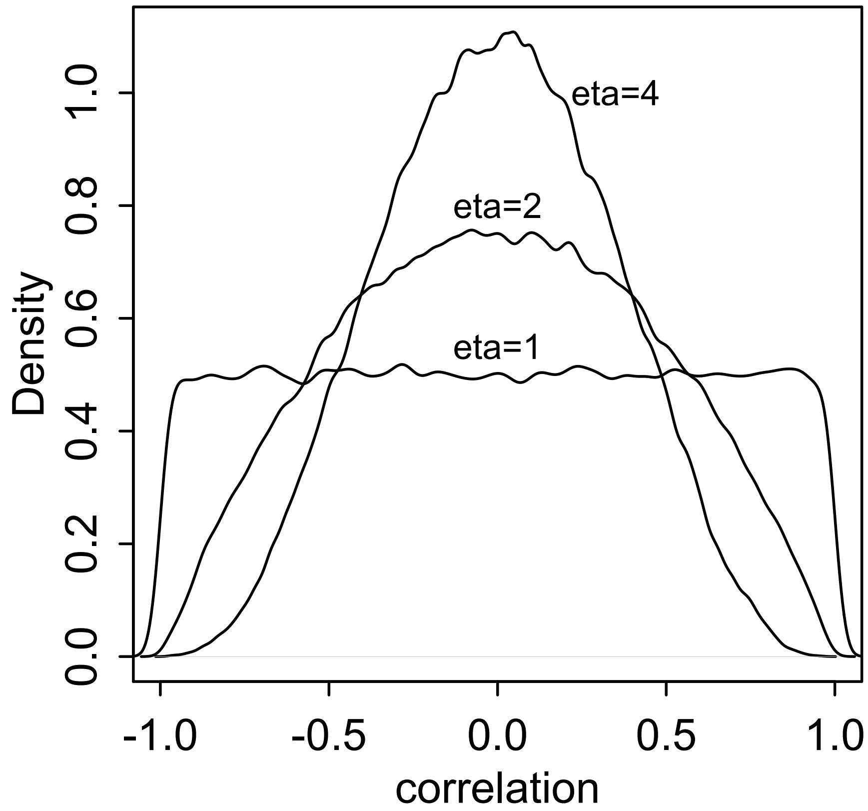
Covariance matrix shuffle

- m-by-m covariance matrix requires estimating
  - m standard deviations (or variances)
  - $(m^2 - m)/2$ correlations (for covariances)
  - total of $m(m + 1)/2$ parameters
- Several ways specify priors
  - Conjugate: inverse-Wishart (`inv_wishart`)
  - inverse-Wishart cannot pull apart stddev and correlations
  - Better to decompose:

$$\mathbf{S} = \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} = \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix} \underbrace{\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}}_{\mathbf{R}} \begin{pmatrix} \sigma_\alpha & 0 \\ 0 & \sigma_\beta \end{pmatrix}$$

LKJ Correlation prior

- After Lewandowski, Kurowicka, and Joe (LKJ) 2009
- One parameter, *eta*, specifies concentration or dispersion from *identity matrix* (zero correlations)
  - *eta* = 1, uniform correlation matrices
  - *eta* > 1, stomps on extreme correlations
  - *eta* < 1, elevates extreme correlations

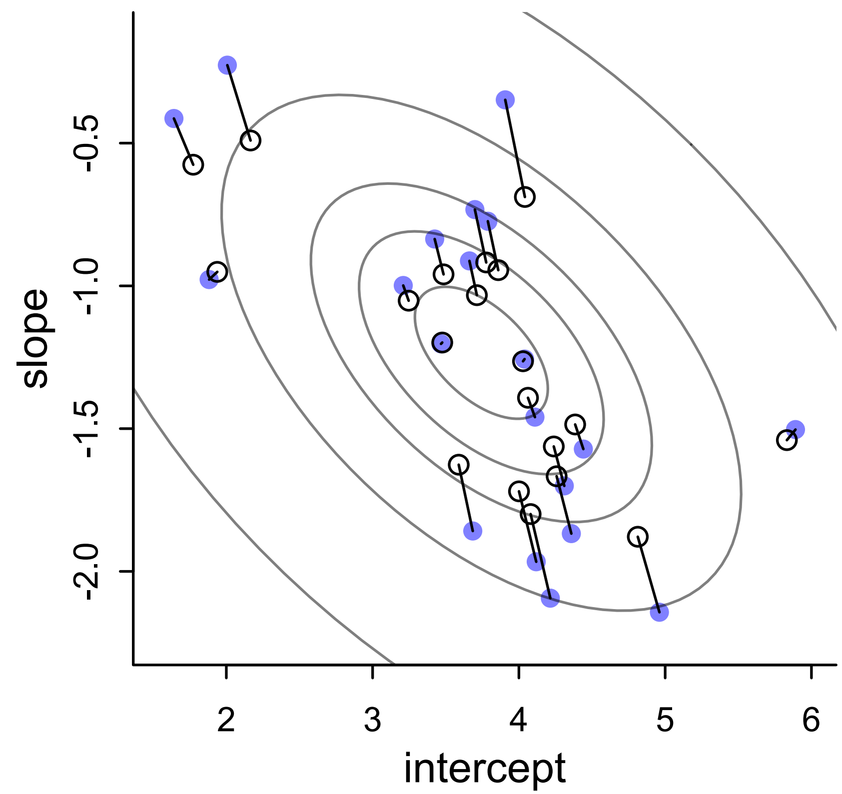


Varying slopes estimation

```
m14.1 <- ulam(
    alist(
        wait ~ normal( mu , sigma ),
        mu <- a_cafe[cafe] + b_cafe[cafe]*afternoon,
        c(a_cafe,b_cafe)[cafe] ~ multi_normal( c(a,b) , Rho , sigma_cafe ),
        a ~ normal(5,2),
        b ~ normal(-1,0.5),
        sigma_cafe ~ exponential(1),
        sigma ~ exponential(1),
        Rho ~ lkj_corr(2)
    ) , data=d , chains=4 , cores=4 )
```

Multi-dimensional shrinkage

- Joint distribution of varying effects pools information across intercepts & slopes
- Correlation btw effects => shrinkage in one dimension induces shrinkage in others
- Improved accuracy, just like varying intercepts



Multilevel horoscopes

- **Think about the causal model first**
- Begin with "empty" model with varying intercepts on relevant clusters
- Standardize predictors
- Use regularizing priors (simulate)
- Add in predictors and vary their slopes
- Can drop varying effects with tiny sigmas
- Consider two sorts of posterior prediction
  - Same units: What happened in these data?
  - New units: What might we expect for new units?
- **Your knowledge of domain trumps all**

ARIES (March 21–April 19)
You have more than one fresh start ahead of you—don't be afraid to reboot Vista often. Your lucky numbers for today are: 3.428, 1.417, 1.155, 1.096, and 1.043.

TAURUS (April 20–May 20)
There is harmony in the universal machinery that regulates the heavens. Get that filing in now!

GEMINI (May 21–June 21)
You learn that your coworkers are more or less of one mind—that you need to get the team moving and on to new projects. Focus them on personal hygiene.

CANCER (June 22–July 22)
CFOs are reawakening their chakras. Channel this energy to strengthen reserves.

LEO (July 23–August 22)
Your negative energy is blocking your ability to utilize MS Office fully. Look to leverage pre-existing analyses and presentations. Don't forget to update those headers and footers!

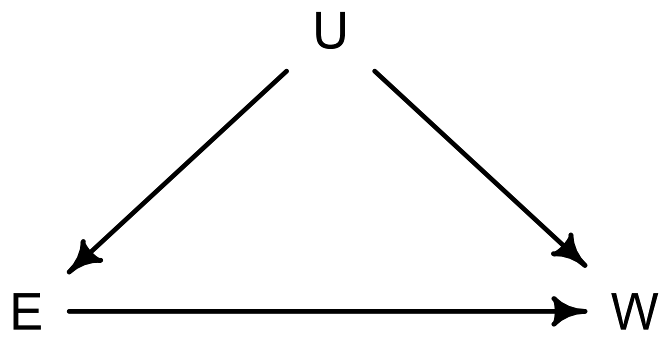
VIRGO (August 23–September 22)
Triangles are aligning with the 5th moon of Neptune. Multiplicative methods beware! Cape Cod is more than a summer vacation destination!

Adventures in covariance

- Many possibilities arise from using multi-variate Gaussian distributions
- Models of unobserved confounds: Instrumental variables, Mendelian randomization
- Models of social relations, networks
- Factor analysis (item-response theory)
- “Animal model” — heritability of phenotype
- Phylogenetic regressions
- Spatial autocorrelation

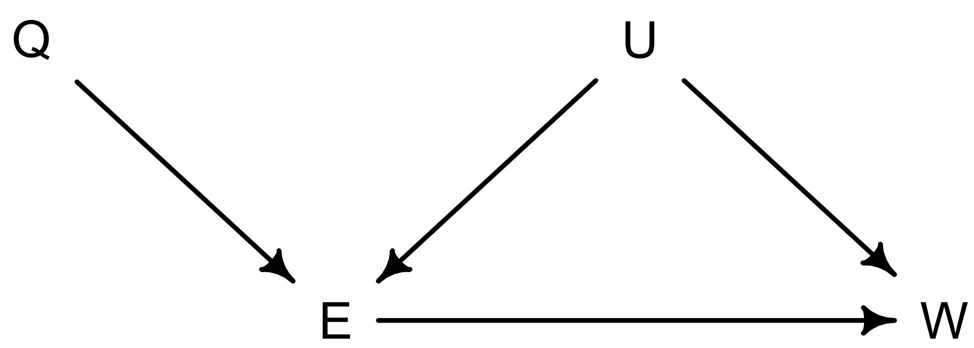
Instrumental variables

- Imagine trying to estimate influence of education on wages — lots of unmeasured confounds.



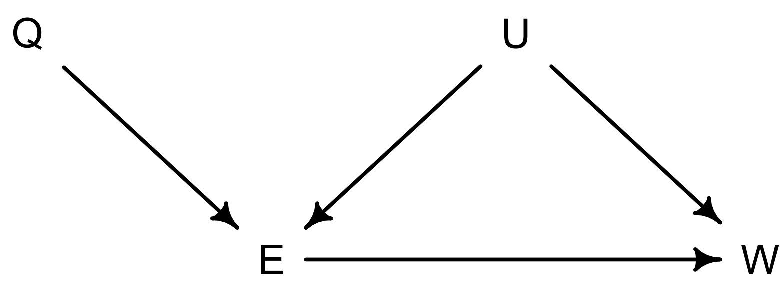
Instrumental variables

- Instrument: A variable that influences exposure (E) but not outcome (W)
- How could this help us?
- Gives us information about U
- E and W correlated, due to U
- Q helps us measure that correlation

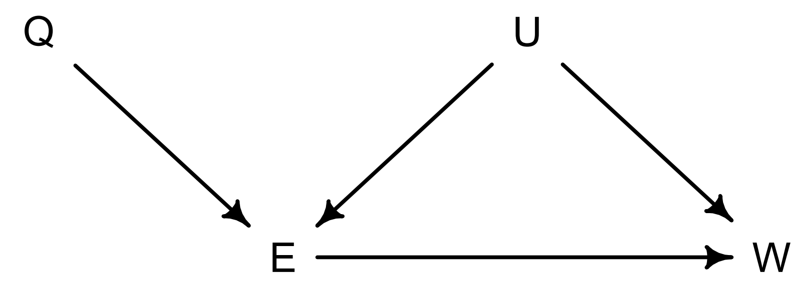


Instrumental variables

- Another perspective:
- Q is a “natural experiment”
- Q assigns E, as if by experimenter giving education pills
- But individuals are uncooperative and don’t always take their pills => imperfect randomization
- Many (most?) real “experiments” are actually like this, have *intent to treat*



Simulated instrument



$$W_i \sim \text{Normal}(\mu_{\text{W},i}, \sigma_{\text{W}})$$ [Wage model]

$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i + U_i$$

$$E_i \sim \text{Normal}(\mu_{\text{E},i}, \sigma_{\text{E}})$$ [Education model]

$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i + U_i$$

$$Q_i \sim \text{Bernoulli}(0.25)$$ [Birth model]

$$U_i \sim \text{Normal}(0, 1)$$ [Confound model]

Instrumentality

- Think of pairs of (W,E) values as sampled from a common distribution with some covariance structure:

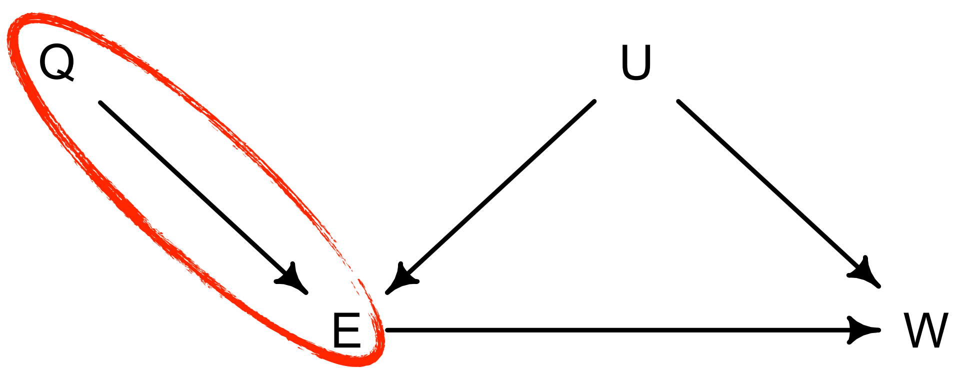
$$
\begin{aligned}
\begin{pmatrix} W_i \\ E_i \end{pmatrix} &\sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right) \\
\mu_{\text{W},i} &= \alpha_{\text{W}} + \beta_{\text{EW}} E_i \\
\mu_{\text{E},i} &= \alpha_{\text{E}} + \beta_{\text{QE}} Q_i
\end{aligned}
$$

Instrumentality

$$\begin{pmatrix} W_i \\ E_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right)$$

$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i$$

$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i$$

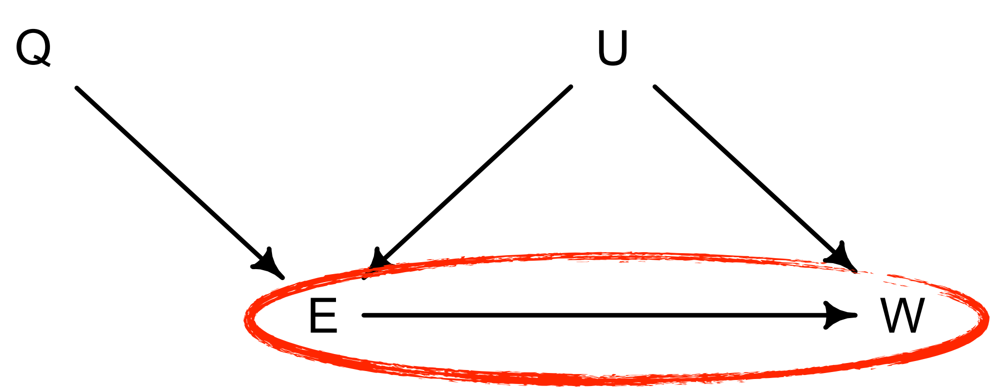


Instrumentality

$$\begin{pmatrix} W_i \\ E_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right)$$

$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i$$

$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i$$

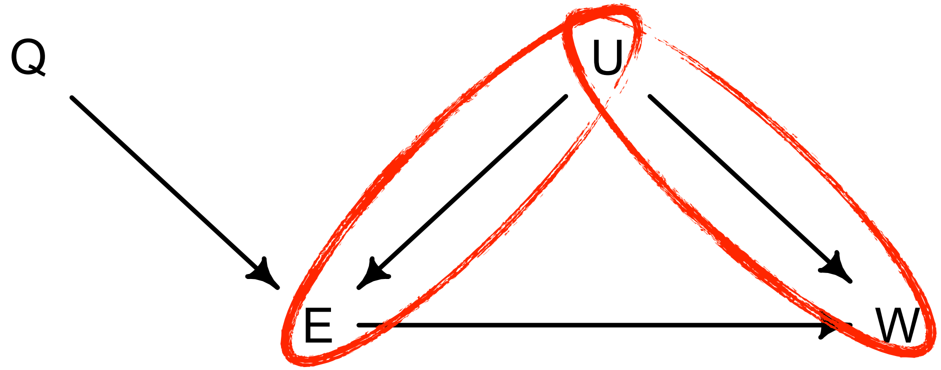


Instrumentality

$$\begin{pmatrix} W_i \\ E_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right)$$

$$\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i$$

$$\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i$$



$$
\begin{pmatrix} W_i \\ E_i \end{pmatrix} \sim \text{MVNormal}\left(\begin{pmatrix} \mu_{\text{W},i} \\ \mu_{\text{E},i} \end{pmatrix}, \mathbf{S} \right)
$$

$$
\mu_{\text{W},i} = \alpha_{\text{W}} + \beta_{\text{EW}} E_i
$$

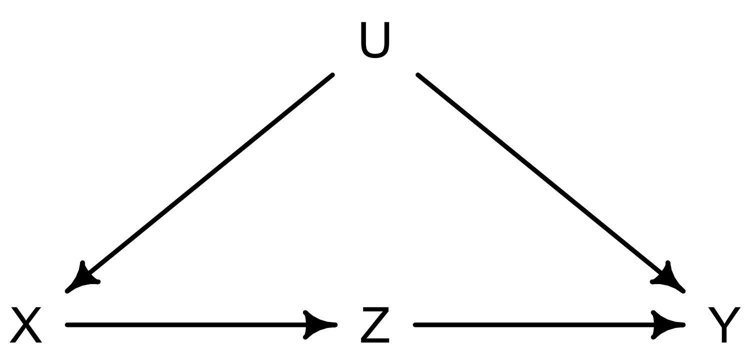
$$
\mu_{\text{E},i} = \alpha_{\text{E}} + \beta_{\text{QE}} Q_i
$$

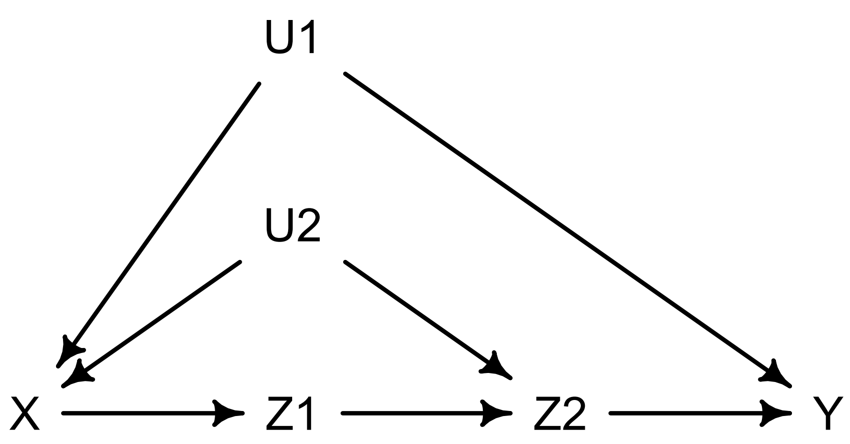
R code 14.25

```
m14.5 <- ulam(
    alist(
        c(W,E) ~ multi_normal( c(muW,muE) , Rho , Sigma ),
        muW <- aW + bEW*E,
        muE <- aE + bQE*Q,
        c(aW,aE) ~ normal( 0 , 0.2 ),
        c(bEW,bQE) ~ normal( 0 , 0.5 ),
        Rho ~ lkj_corr( 2 ),
        Sigma ~ exponential( 1 )
    ), data=dat_sim , chains=4 , cores=4 )
precis( m14.5 , depth=3 )
```

Other doors

- In principle, many idiosyncratic ways to de-confound inference, if you analyze the graph correctly ("do-calculus")
- Another well-known tool: **Front-door criterion**





Continuous categories

- Traditional clusters discrete, unordered => every category equally different from all others (in prior)
- Continuous dimensions of difference:
  - Age, income, location, phylogenetic distance, social network distance, many others
  - No obvious cut points in continuum, but close values share common exposures/covariates/interactions
- Would like to exploit pooling in these cases as well
- Common approach: **Gaussian process regression**