The 'regmed' package for Regularized Mediation Analysis

Jason Sinnwell

MAYO CLINIC

Outline

MAYO CLINIC

Intro

MAYO CLINIC

Biography: Jason Sinnwell

- Mayo Clinic, Biostatistician. MS from Iowa State University (2002)
- Research agreas: cancer genomics, GWAS, DNA/RNA Sequencing
- CRAN package co-author: haplo.stats, kinship2, arsenal, pedgene, regmed

MAYO CLINIC

Regularized Mediation, background

MAYO CLINIC

Sources

- Most of this talk is from two papers:
  - 2020, Penalized models for analysis of multiple mediators
  - 2021, Penalized Mediation models for multivariate data
- regmed on CRAN

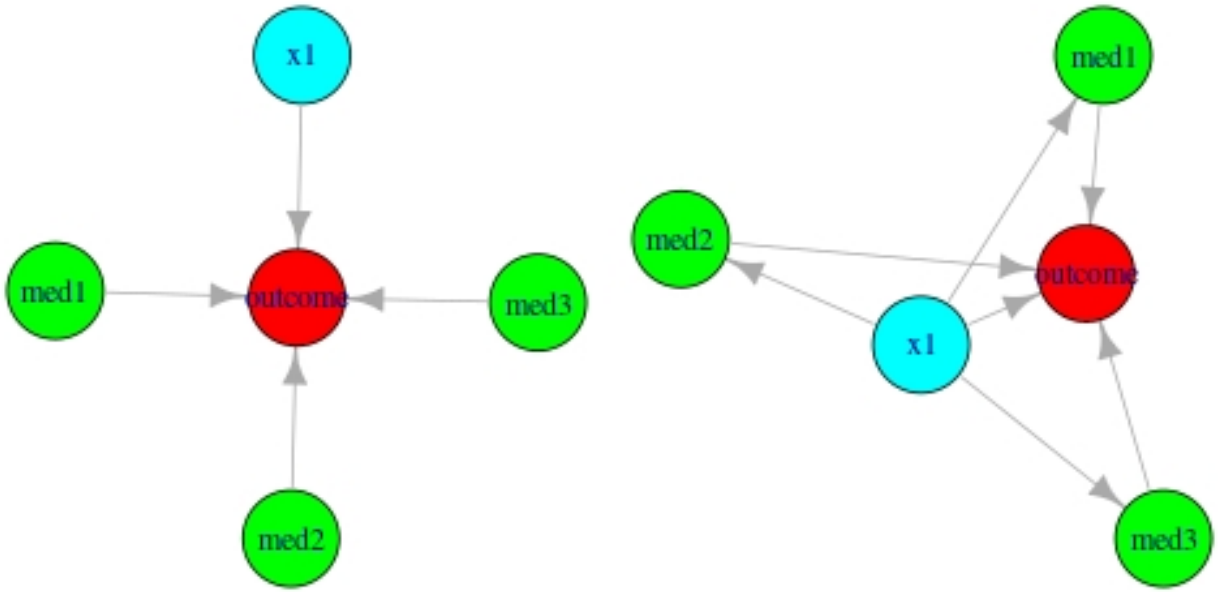
MAYO CLINIC

Mediation: Motivating experiments

- Quantitative exposure variable (x) has an effect on continuous outcome (y)
- Variability in outcome could be explained by one or more biomarkers
- Biomarkers are not independent of exposure variable
- Models to impose linked association between exposure, biomarkers, and outcome

MAYO CLINIC

Mediation: Graphical Representation



MAYO CLINIC

Mediation Analysis: simple version

- Exposure associated with mediator: $m_i = \alpha_0 + \alpha_i x + \epsilon$
- Outcome associated with both: $y = \beta_0 + \delta x + \beta_i m_i + \epsilon$
- Where α_i and β_i both need to be non-zero for mediation to exist
- Formal test for multiple mediators, H0 : $\Sigma_i(|\alpha_i * \beta_i|) = 0$
- What to do when many (> 10) mediators?
  - Pre-filter using covariances (Sobel 1975, Fan and Lv, 2008)
  - Penalized approach

MAYO CLINIC

Mediation with many mediators

- Structural Equation Models (SEMs) do both regressions in single model
- Allow directed relationships from exposure to mediators to outcome (also x->y)
- Allow correlations among mediators
- Model Metrics assessed over multiple models
  - Comparable Fit Index (CFI). Between (0,1); higher is better
  - Root MSE of Approximation (RMSEA). Between (0,1); lower is better
  - BIC, uses ln-likelihood, lower is better

MAYO CLINIC

The regmed method

MAYO CLINIC

Regmed methodology

- If structure or grouping of mediators, use SEMs with latent vars (lavaan)
- *regmed*: penalized SEMs to select from measured mediators for best model
- Similar approach to *regsem* package; improvement on speed, convergence, penalty options
- Key: log-likelihood of a model, given the coefficients (for BIC)

MAYO CLINIC

Model Penalties

- Combination of L1 ($|\alpha|$, aka 'lasso') and L2 (α^2, aka 'ridge') penalty
- Sparse group lasso penalty in regmed:
- $P(\alpha, \beta, \delta, w, \lambda, f) =$
  $\lambda w|\delta| + \lambda(1-f)\Sigma_q\sqrt{(}\alpha^2+\beta^2) + \lambda f \Sigma_q(|\alpha| + |\beta|)$
- Where f is fraction of penalty assigned to L1 portion, and $1-f$ applied to L2 portion
- And w weights how much to penalize δ, the direct x->y effects.

MAYO CLINIC

Simulated dataset

```
data(medsim, package="regmed")
dim(x); dim(med)
```

```
## [1] 100  10
```

```
## [1] 100 200
```

MAYO CLINIC

Analysis Steps

- Too many mediators, pre-filter
- Find best lambda penalty by BIC
- Apply fit with best penalty
- Select final model
- Fit final model without penalties

MAYO CLINIC

Pre-filter mediators, k=7

- For simple example, use first x, y, and 7 most likely mediators
- With *sure-select* (Fan and Lv, 2008) steps (covariances)
- Standardizes (center+scale), and removes any rows with missings

```
filter7 <- regmed.prefilter(x[,1], med, y[,1], k=7)
colnames(filter7$mediator)
```

```
## [1] "med.1"   "med.2"   "med.74"  "med.88"  "med.99"  "r
```

MAYO CLINIC

Regmed Grid of Penalties (frac.lasso=f, wt.delta=w)

```
x1 <- filter7$x; y1=filter7$y; med=filter7$mediator
gridfit <- regmed.grid(x=x1, mediator=med, y=y1,
                       lambda.vec=seq(1, 0, by=-.33),
                       frac.lasso=.8, wt.delta=.5)
```

fitting grid lambda [1] = 1 fitting grid lambda [2] = 0.67 fitting grid lambda [3] = 0.34 fitting grid lambda [4] = 0.01

MAYO CLINIC

Regmed Grid Results

lambda	converge	iter	df	bic
1.00	TRUE	96	4	858.5203
0.67	TRUE	6	4	820.7717
0.34	TRUE	7	8	803.1092
0.01	TRUE	20	16	818.1816

MAYO CLINIC

Regmed Fit: best penalty, default w,f

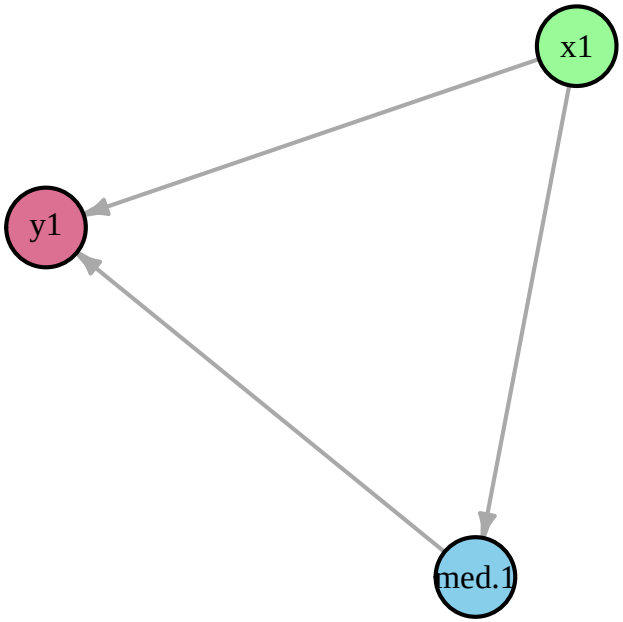
```
fit.lam3 <- regmed.fit(x=x1, mediator=med, y=y1,
                       lambda = 0.3)
kable(with(fit.lam3, cbind(alpha, beta)))
```

	x1	y1
med.1	0.4104309	0.1439518
med.2	0.3993711	0.0000000
med.74	0.0000000	0.0000000
med.88	-0.0663993	0.0000000
med.99	-0.0121319	-0.0035687
med.112	-0.1233457	0.0000000
med.129	0.0918760	0.0000000

MAYO CLINIC

Edges Graph: mediators only

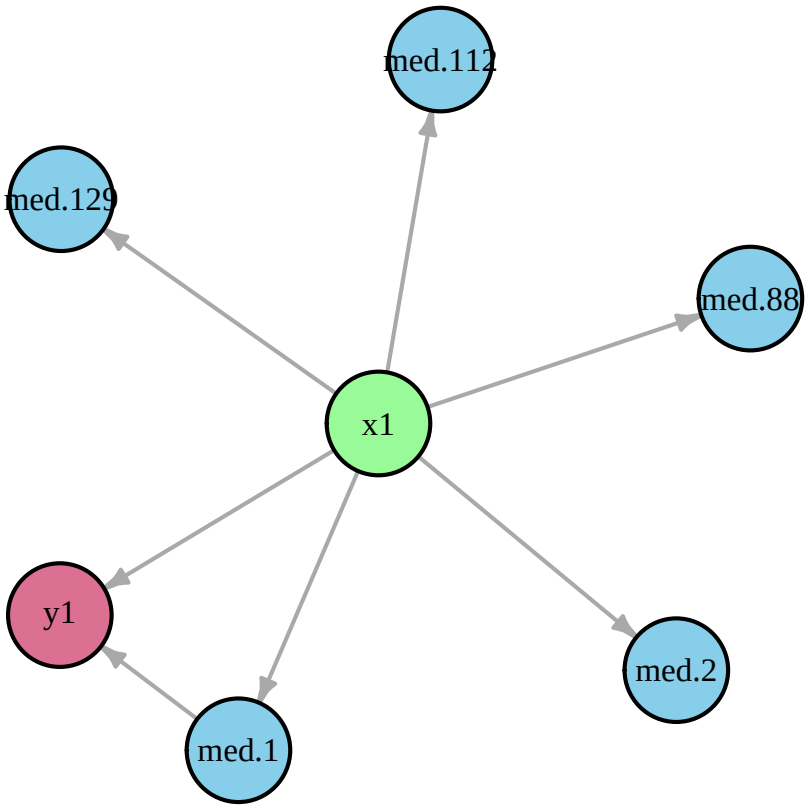
```
edges.meds <- regmed.edges(fit.lam4, eps=0.01,
                           type="mediators")
plot(edges.meds)
```



MAYO CLINIC

Edges Graph: Any non-zero coefficient

```
edges.any <- regmed.edges(fit.lam4, eps=0.01, type="any")
plot(edges.any)
```



MAYO CLINIC

Re-fit best model with lavaan for coefficients

Tools for multi-level model formula for *lavaan*.

```
lmodel.meds <- regmed.lavaan.model(edges.meds, fit.lam3)
lavaandat <- regmed.lavaan.dat(x1, med, y1)
lav.med <- sem(model=lmodel.meds, data = lavaandat)
```

MAYO CLINIC

Lavaan model summary

lhs	op	rhs	est	se	z	pvalue
med.1	~	x1	0.5515	0.0838	6.5772	0.0000
y1	~	med.1	0.2927	0.0930	3.1474	0.0016
y1	~	x1	0.4149	0.0933	4.4448	0.0000
y1	~~	y1	0.6021	0.0852	7.0711	0.0000

MAYO CLINIC

Multi-variate extension

MAYO CLINIC

M-V Extention, Prefilter

- Not always single exposure/outcome, extend to M-V
- Pre-filtered for all mediators with y1 (not shown)
- Total of 12 in top 5 for any of the first 5 exposures (not shown)
- Fit with grid of lambda 0 to 0.6, by 0.1

MAYO CLINIC

MV-Regmed Grid results

	lambda	df	df.alpha	df.beta	df.delta	df.vary	bic
3	0.4	2	0	0	1	1	-2891.432
4	0.3	2	0	0	1	1	-2897.307
5	0.2	5	0	2	2	1	-2889.847
6	0.1	18	4	10	3	1	-2881.722
7	0.0	155	125	24	5	1	-2487.752

MAYO CLINIC

MV-Regmed Fit: best penalty, default w,f

- The *best* lambda at 0.3, but few parameters
- Try lambda=0.1, and keep edges with coeff > 0.05

```
mvfit.lam <- mvregmed.fit(x=xmv, mediator=medmv, y=ymv,
                          lambda = 0.1)
mvedges <- mvregmed.edges(mvfit.lam, eps=0.05)
```

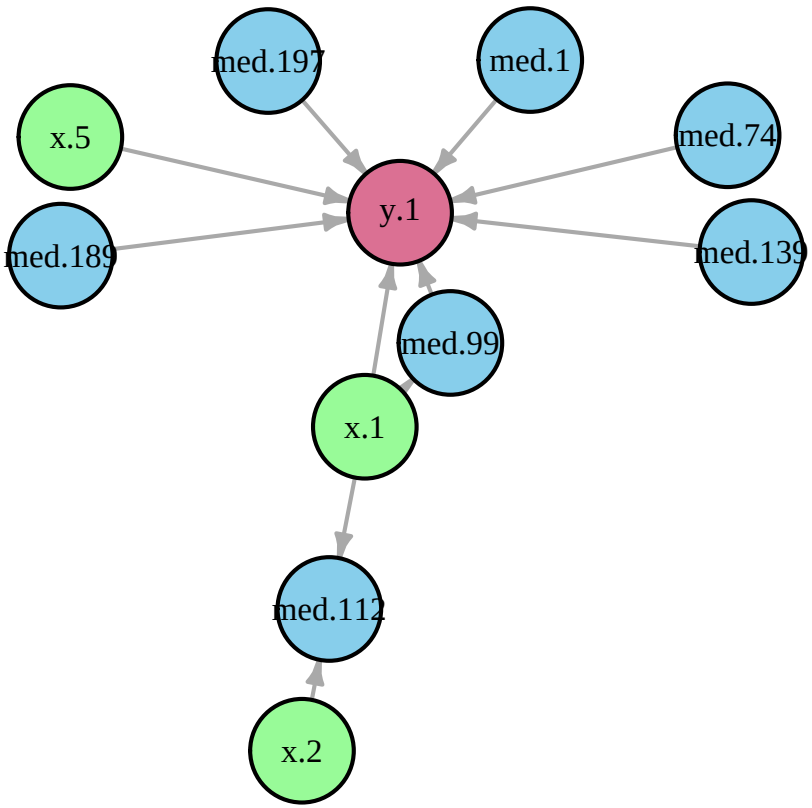
MAYO CLINIC

MV Fit Edges

vertex1	vertex2	coefftype
x.1	med.99	alpha
x.1	med.112	alpha
x.2	med.112	alpha
x.1	y.1	delta
x.5	y.1	delta
med.1	y.1	beta
med.74	y.1	beta
med.99	y.1	beta
med.197	y.1	beta
med.189	y.1	beta
med.139	y.1	beta

MAYO CLINIC

MV Edges Graph

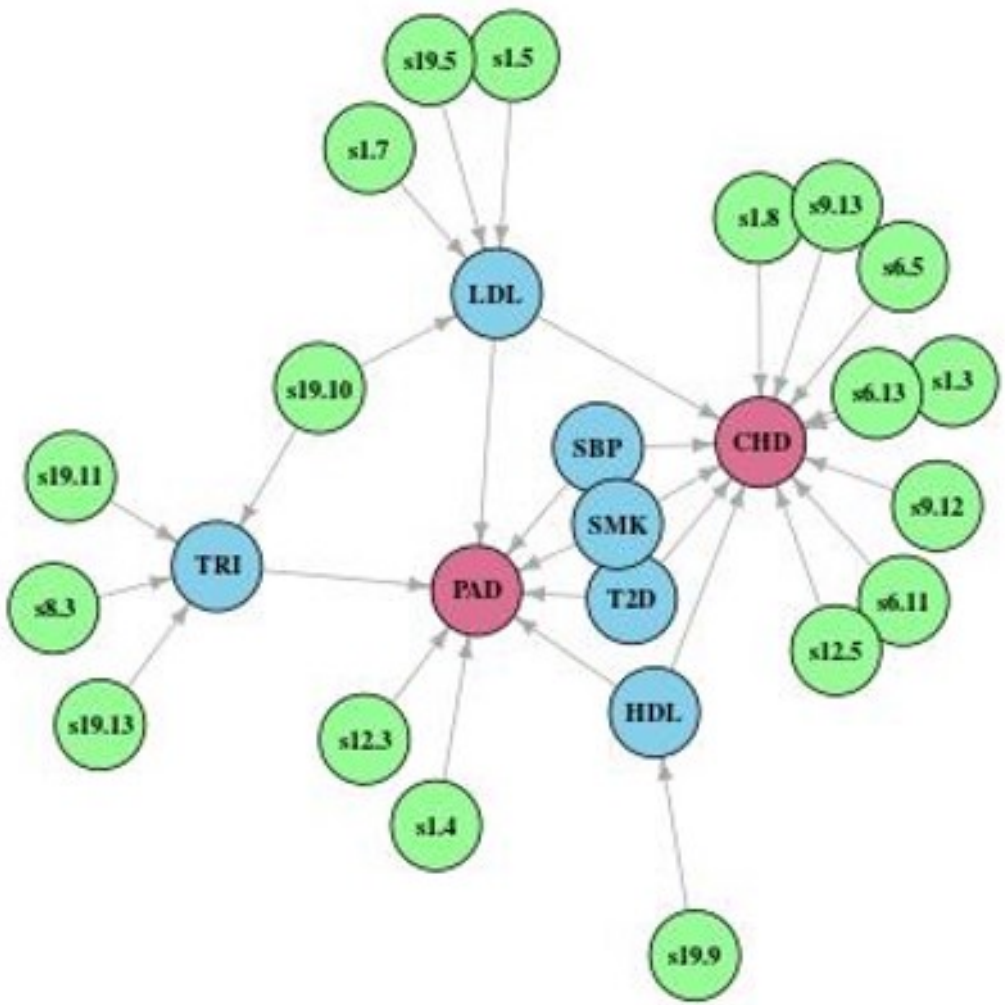


MV-Regmed: Results in Review

- Coronary Heart Disease (CHD, binary) and Peripheral Artery Disease (PAD, continuous)
- Biomarkers, risk factors play a role (LDL, HDL, etc)
- Multiple SNP hits
- Research question: How do SNPs and biomarkers play role in either endpoint?

MAYO CLINIC

MV-Regmed: Published Figure



MAYO CLINIC

Conclusion

MAYO CLINIC

Review

- Experiments that fit the mediator framework
- Flexible to allow just α_i without $\beta_i > 0$, and vice-versa
- Performance with other quantitative variables, e.g., resids from coxph().
- Useful connections with graphs, and lavaan for SEM.
- Sample size limitations
- What questions do you have?

MAYO CLINIC