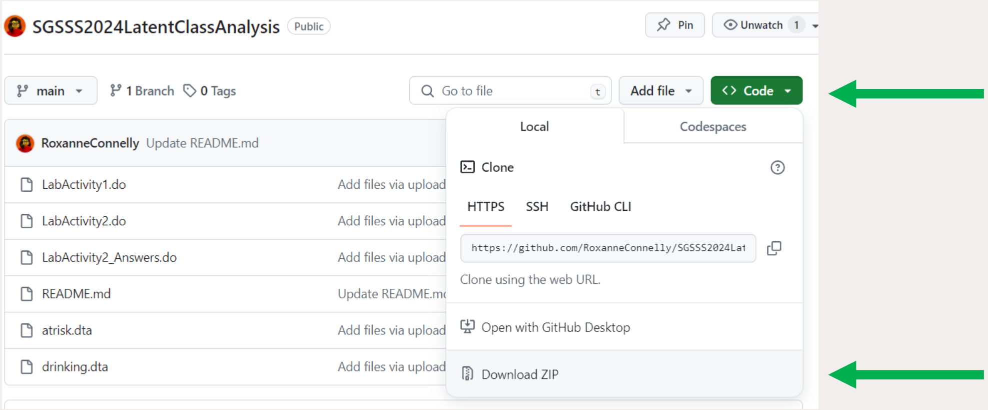
# Latent Class Analysis

June 2024

Dr Roxanne Connelly

Materials are available here:

https://github.com/RoxanneConnelly/SGSSS2024LatentClassAnalysis




Lecture Slides: **lectureslides.pdf**

# Workshop Aims

This workshop will provide you with:

1. An introduction to what latent class analysis is and how it is used in the social sciences;

2. An introduction to how to carry out a latent class analysis and how to interpret the results.

# Timetable

| | |
|---|---|
| 1000 – 1100 | Talk 1 |
| 1100 – 1130 | Lab Activities |
| 1130 – 1200 | Talk 2 |
| 1200 – 1230 | Lab Activities |
| 1230 – 1300 | Final Advice and Questions |

# Talk 1

# Operationalising Social Science Concepts

Operationalisation is the process of turning abstract concepts into measurable observations

Many concepts in the social sciences cannot be observed or measured directly

Authoritarianism
Social Exclusion
Cultural Capital
Sexism
Economic Development

Neoliberalism
Racial Prejudice
Social Capital
Anomie
Social Class

# Operationalising Social Science Concepts

Religiosity

'The quality or state of being religious'

# Latent Class Analysis

# Latent Class Analysis

- Latent class analysis is a method used to extract meaningful groups (i.e. latent classes) from data.
- The concept of latent classes was first introduced by Paul Lazarsfeld (1950) and the implementation of the method was substantially developed by Leo Goodman (1974).


Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis & the interpretation and mathematical foundation of latent structure analysis. In S. A.Stouffer, L.Guttman, E. A.Suchman, P. F.Lazarsfeld, S. A.Star, J. A.Clausen (Eds.), *Measurement and prediction* (pp. 362–472). Princeton, NJ: Princeton University Press.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I—A modified latent structure approach. *American Journal of Sociology*, 79, 1179–1259.

Magidson, J., Vermunt, J. K., & Madura, J. P., (2020). Latent Class Analysis, In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), Sage Research Methods Foundations. Sage.

# Latent Class Analysis

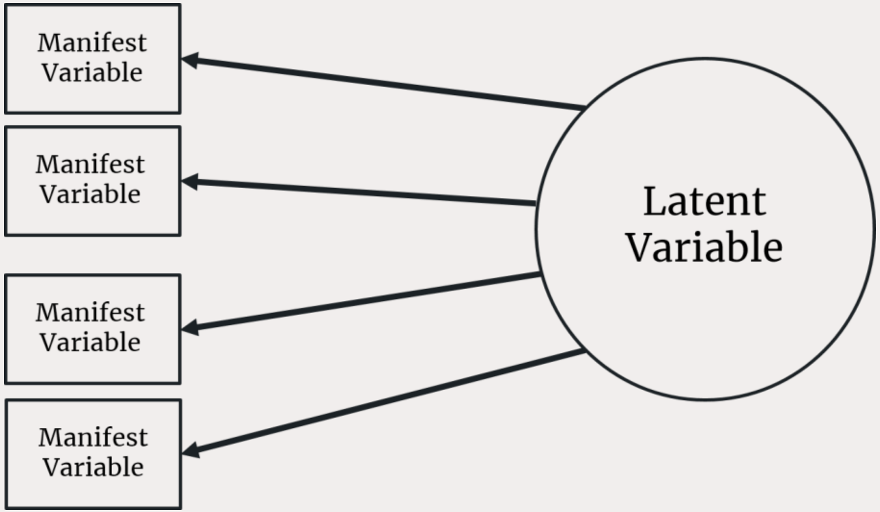
- Covariation in the observed variables (i.e. manifest variables) is theorised to be due to the latent variable.
- The latent variable ‘explains’ the relationship between the observed variables.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis & the interpretation and mathematical foundation of latent structure analysis. In S. A.Stouffer, L.Guttman, E. A.Suchman, P. F.Lazarsfeld, S. A.Star, J. A.Clausen (Eds.), *Measurement and prediction* (pp. 362–472). Princeton, NJ: Princeton University Press.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I—A modified latent structure approach. *American Journal of Sociology*, 79, 1179–1259.

# Latent Class Analysis

Latent class analysis (LCA) is a statistical procedure used to identify **qualitatively different** subgroups.



# Latent Class Analysis

Using latent class analysis we can fit a statistical model to try to determine which individuals are likely to belong to each latent class group based on the patterns in the manifest variables.

# Example Applications

# Cultural Consumption

- Chan, T. W., & Goldthorpe, J. H. (2007). **Social stratification and cultural consumption: The visual arts in England.** Poetics, 35(2), 168-190.
- The aim was to examine the relationship between social stratification and cultural consumption. Specifically the consumption of the ‘visual arts’.
- This analysis used the Arts in England Survey.

# Cultural Consumption

Table 1
Percentage of respondents who have visited various visual arts events or establishments in the past 12 months

| | |
|---|---|
| Event involving video or electronic art | 7.7 |
| Cultural festival | 11.0 |
| Craft exhibition | 18.5 |
| Exhibition or collection of art, photography or sculpture | 21.0 |
| Museum or art gallery | 38.7 |

Chan, T. W., & Goldthorpe, J. H. (2007). Social stratification and cultural consumption: The visual arts in England. Poetics, 35(2), 168-190.

# Cultural Consumption

Potential patterns of response:

- 5 manifest variables
- 2 possible outcomes (yes/no)

| Video / EA | Cultural Festival | Craft Exhib. | Art / Photo | Museum / Gallery |
|---|---|---|---|---|
| Yes | Yes | Yes | Yes | Yes |
| Yes | Yes | Yes | Yes | No |
| Yes | Yes | Yes | No | Yes |
| Yes | Yes | Yes | No | No |
| Yes | Yes | No | Yes | Yes |
| Yes | Yes | No | Yes | No |
| Yes | Yes | No | No | Yes |
| Yes | Yes | No | No | No |

Chan, T. W., & Goldthorpe, J. H. (2007). Social stratification and cultural consumption: The visual arts in England. Poetics, 35(2), 168-190.

# Cultural Consumption

Latent Class 1: Inactives

Latent Class 2: Paucivores

Latent Class 3: Omnivores

Table 3
Estimated relative sizes of latent classes and conditional probabilities of different forms of visual arts consumption under a three-class model

| | 1 | 2 | 3 |
|---|---|---|---|
| Relative size (%) | 58.6 | 34.4 | 7.0 |
| Event involving video or electronic art | 0.092 | 0.252 | 0.632 |
| Cultural festival | 0.040 | 0.120 | 0.644 |
| Craft exhibition | 0.035 | 0.067 | 0.478 |
| Exhibition or collection of art, photography or sculpture | 0.004 | 0.416 | 0.922 |
| Museum or art gallery | 0.071 | 0.809 | 0.966 |

# Cultural Consumption

Latent Class 1: Inactives

Latent Class 2: Paucivores

Latent Class 3: Omnivores

Latent group membership is more strongly differentiated by social status than by social class or by income.

Education is the most important determinant of latent group membership.

# Caregiver Coping Strategies

- Lin, I. F., & Wu, H. S. (2014). **Patterns of coping among family caregivers of frail older adults.** Research on aging, 36(5), 603-624.
- Previous studies have extensively examined the coping strategies of caregivers, but they have not considered the possibility that distinct groups of caregivers may use different coping strategies.
- This analysis used the National Long Term Care Survey.

**Table 1.** Percentages of Family Caregivers Who Said How They Coped.

| | % |
|---|---|
| Emotion-focused coping | |
| Talk with friends or relatives | 77.39 |
| Pray or meditate | 71.82 |
| Watch TV | 71.49 |
| Read | 67.15 |
| Spend time alone | 61.08 |
| Spend time on exercise or hobbies | 55.80 |
| Eat | 44.33 |
| Take medication to calm | 14.61 |
| Drink alcohol | 13.22 |
| Smoke | 12.52 |
| Problem-focused coping | |
| Obtain assistive devices such as wheelchairs or walkers for care recipient | 58.16 |
| Use services for personal or nursing care at care recipient's home | 29.72 |
| Make modifications in care recipient's home | 28.56 |
| Use services to help with housework at care recipient's home | 18.90 |
| Use services delivering meals to care recipient's home | 11.04 |
| Use transportation services for care recipient | 10.80 |
| Use respite or caregiver support services | 10.80 |
| Request information on financial help for care recipient | 10.59 |
| *N* | 1,552 |

**Table 3.** Three Latent Class Model ($N$ = 1,552).

| | Latent Class | | |
|---|---|---|---|
| | Unpatterned Coping | Emotional Coping | Hybrid Coping |
| Probability of membership | .20 | .46 | .33 |
| Conditional probability of a *Yes* response | | | |
| Emotion-focused coping | | | |
| Talk with friends or relatives | .22 | .91 | .92 |
| Pray or meditate | .22 | .85 | .85 |
| Watch TV | .08 | .90 | .86 |
| Read | .04 | .86 | .81 |
| Spend time alone | .14 | .72 | .75 |
| Spend time on exercise or hobbies | .07 | .67 | .70 |
| Eat | .05 | .50 | .61 |
| Take medication to calm | .02 | .13 | .24 |
| Drink alcohol | .01 | .15 | .18 |
| Smoke | .03 | .16 | .14 |
| Problem-focused coping | | | |
| Obtain assistive devices | .45 | .41 | .89 |
| Use personal or nursing care | .21 | .08 | .63 |
| Make home modifications | .16 | .15 | .55 |
| Use services to help with housework | .14 | .08 | .36 |
| Use meal deliver services | .05 | .04 | .24 |
| Use outside services for transportation | .07 | .04 | .22 |
| Use respite or caregiver support services | .04 | .02 | .27 |
| Request information on financial help | .03 | .07 | .20 |

# A Bourdieusian Social Class Measure

- Connelly, R., Gayle, V., & Playford, C. (2021). **Social class inequalities in educational attainment: measuring social class using capitals, assets and resources.** Contemporary Social Science, 16(3), 280–293.
- Is it possible to create a social class measure based on indicators of capitals, assets and resources using existing social survey data resources?
- This analysis used the United Kingdom Household Longitudinal Study (UKHLS, Understanding Society).

**Table 3.** Prior probabilities for the six latent class solution.

| Ascribed label | 'Elite' | 'The established middle class' | 'The socially engaged middle class' | 'The culturally engaged middle class' | 'The traditional working class' | 'The precariat' |
|---|---|---|---|---|---|---|
| % Allocated to Latent Class | – | 31 | – | 11 | 35 | 19 |
| Prior Probabilities | | | | | | |
| *Economic Capital* | | | | | | |
| Equivalised Net Monthly Household Income (mean) | 7298.64 | 1559.42 | 1568.76 | 1494.50 | 1232.39 | 1348.61 |
| Own Home | 1.00 | 0.91 | 0.86 | 0.95 | 0.75 | 0.52 |
| *Social Capital* | | | | | | |
| Number of Friends (mean) | 6.14 | 5.16 | 25.03 | 7.45 | 4.91 | 3.96 |
| Active in Trade Union | 0.15 | 0.09 | 0.05 | 0.09 | 0.02 | 0.06 |
| Active in Professional Organisation | 0.68 | 0.16 | 0.37 | 0.37 | 0.04 | 0.00 |
| *Highbrow Cultural Capital* | | | | | | |
| Classical Music Concerts | 0.57 | 0.23 | 0.23 | 0.34 | 0.03 | 0.00 |
| Historic Buildings | 0.42 | 0.86 | 0.61 | 0.95 | 0.12 | 0.13 |
| Museums / Galleries | 0.85 | 0.84 | 0.78 | 0.98 | 0.21 | 0.29 |
| Plays/Theatre/ Pantomime | 1.00 | 0.88 | 0.80 | 0.77 | 0.33 | 0.20 |
| Ballet Performance | 0.14 | 0.13 | 0.23 | 0.14 | 0.00 | 0.00 |
| *Emerging Cultural Capital* | | | | | | |
| Social Media | 0.58 | 0.64 | 0.42 | 0.88 | 0.83 | 0.28 |
| Regularly use the Internet | 1.00 | 0.92 | 1.00 | 1.00 | 1.00 | 0.24 |
| Gym/Fitness Classes | 0.60 | 0.45 | 0.55 | 0.93 | 0.44 | 0.22 |
| Rock/Pop/Jazz Concert | 0.82 | 0.59 | 0.58 | 0.69 | 0.22 | 0.10 |
| Cycle | 0.42 | 0.46 | 0.43 | 0.99 | 0.28 | 0.12 |
| Racquet Sports | 0.13 | 0.24 | 0.66 | 0.66 | 0.09 | 0.06 |
| Golf | 0.27 | 0.19 | 0.19 | 0.28 | 0.07 | 0.03 |

Note: – indicates that the values in the cell have been suppressed for statistical disclosure control purposes. $n = 616$. The sample size reduces by one case in the weighted analyses as one sample member has a weight of zero (unweighted $n = 617$).

# Latent Class Analysis

Latent class analysis is a method used to extract meaningful (qualitatively different) groups (i.e. latent classes) from data.

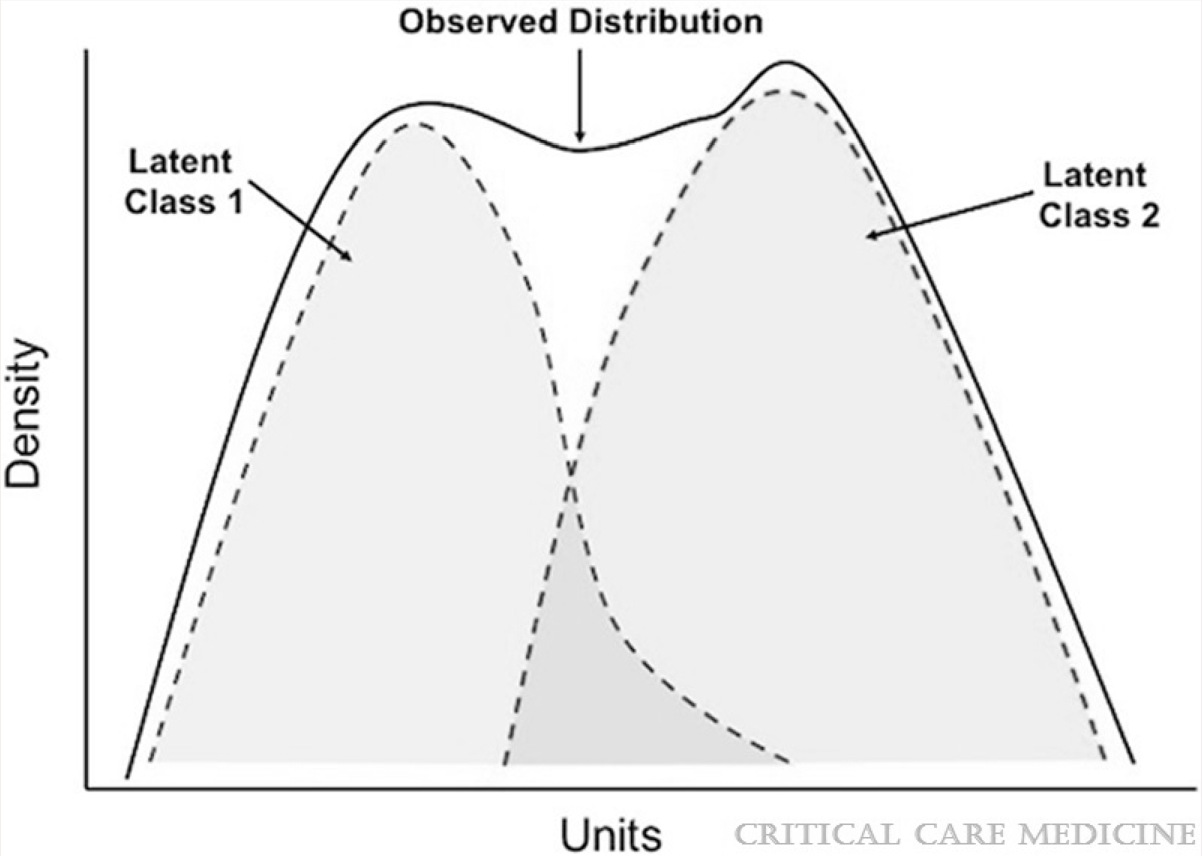
| | | Latent Variables | |
|---|---|---|---|
| | | **Continuous** | **Categorical** |
| **Manifest (i.e. observed) Variables** | **Continuous** | Factor Analysis | Latent Profile Analysis |
| | **Categorical** | Latent Trait Analysis (aka Item Response Theory) | Latent Class Analysis |

Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2008). Analysis of Multivariate Social Science Data. CRC Press. (p.178)

# Latent Class Analysis

- Theorise your latent concept and identify manifest variables.
- Fit a series of latent class models.
- Select the most appropriate latent class model.
- Describe the patterns in your latent classes.
- Use your latent classes in further analysis (i.e. as an explanatory variable or an outcome variable).

# Under the Hood




Sinha, P., Calfee, C. S., & Delucchi, K. L. (2020). Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Critical care medicine*, *49*(1), e63-e79.

# Latent Class Analysis

$$P(y_1, y_2, y_3, y_4) = \sum_{k=1}^{K} P(X = k)\, P(y_1, y_2, y_3, y_4 \mid X = k)$$

Each latent class corresponds to one of k categories that comprise the underlying categorical latent variable (X).

LCA will extract that latent class variable (X) with the fewest number of classes to explain the associations between the observed variables (y1, y2, y3... $y_k$).

Magidson, J., Vermunt, J. K., & Madura, J. P., (2020). Latent Class Analysis, In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), Sage Research Methods Foundations. Sage.

# Latent Class Analysis

$$P\left(y_1, y_2, y_3, y_4 \mid X=k\right) = P\left(y_1 \mid X=k\right) P\left(y_2 \mid X=k\right) P\left(y_3 \mid X=k\right) P\left(y_4 \mid X=k\right)$$

A key assumption of LCA is that the latent variable explains all the associations between the manifest variables (i.e. the assumption of local independence).

Magidson, J., Vermunt, J. K., & Madura, J. P., (2020). Latent Class Analysis, In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), Sage Research Methods Foundations. Sage.

# Latent Class Analysis

$$P\left(y_1, y_2, y_3, y_4\right) = \sum_{k=1}^{K} P\left(X = k\right) P\left(y_1 \middle| X = k\right) P\left(y_2 \middle| X = k\right) P\left(y_3 \middle| X = k\right) P\left(y_4 \middle| X = k\right)$$

More technical introductions:


- McCutcheon, A. L. (1987). Latent Class Analysis. Sage.
- Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). Applied latent class analysis. Cambridge University Press.
- Collins, L. M., & Lanza, S. T. (2009). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. John Wiley & Sons.

# Maximum Likelihood Estimation

- Maximum likelihood estimation is a method used to obtain parameter estimates for statistical models.
- Maximum likelihood estimation consists of procedures for finding estimates of unknown parameters that maximise the likelihood of the observed data, the resulting parameter estimates are called maximum likelihood estimates.
- Maximum likelihood follows an iterative process known as convergence.
- Sometimes models 'fail to converge' (more on that later…)

# Running Latent Class Analysis in Stata

# Software

General Purpose Statistical Software:

- Stata (gsem)
- R (poLCA)
- Mplus

Specialist Software:

- LatentGold
- LEM

# Running Latent Class Analysis in Stata

- From Stata 16 onwards LCA can be undertaken using the gsem suite of commands.

STATA
STRUCTURAL EQUATION MODELING
REFERENCE MANUAL
RELEASE 17

https://www.stata.com/manuals/sem.pdf

# Example: Drinking Behaviour

I like to drink

I drink hard liquor

I have drank in the morning

I have drank at work

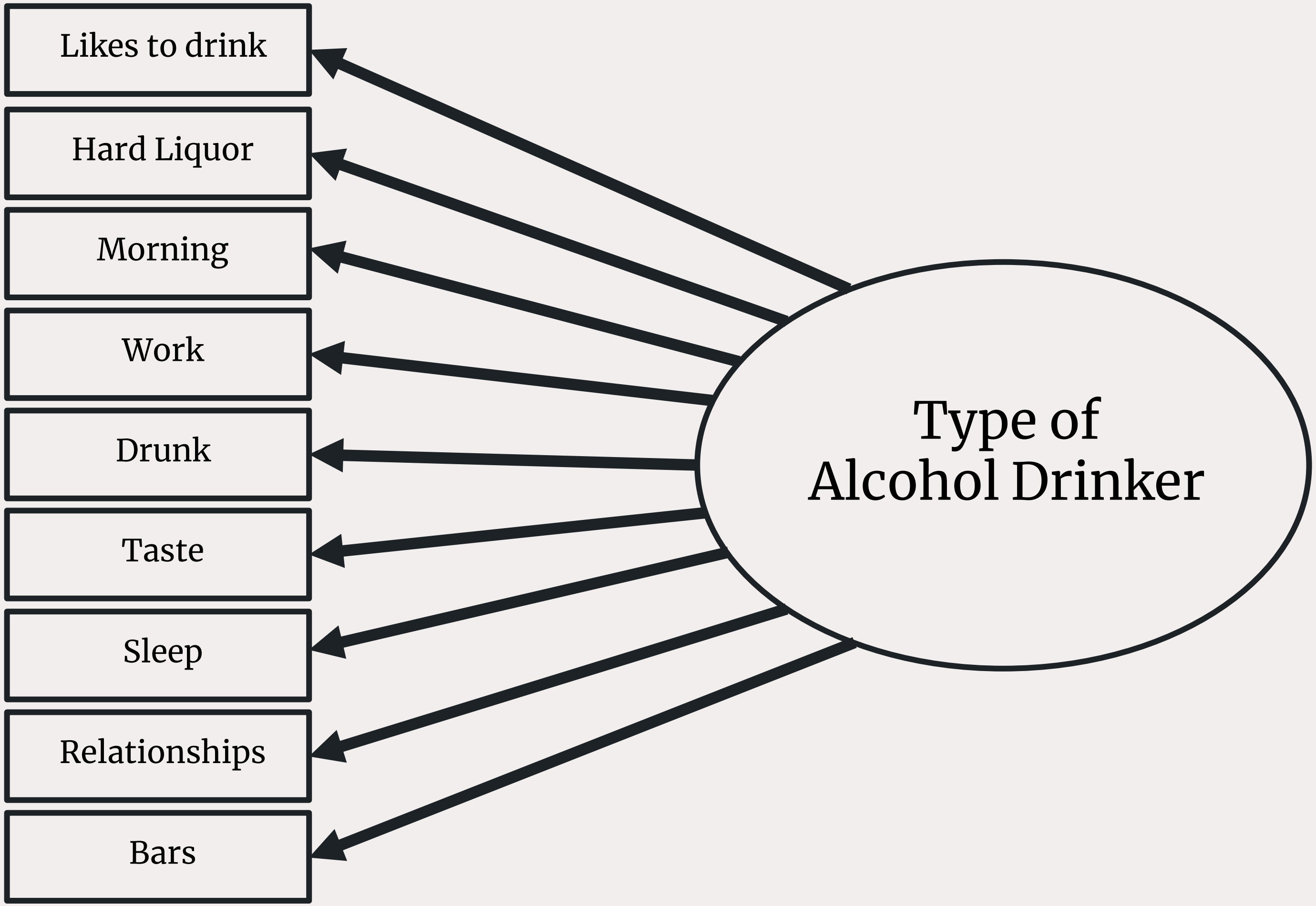
I drink to get drunk

I like the taste of alcohol

I drink to help me sleep

Drinking interferes with my relationships

I frequently visit bars


Likes to drink
Hard Liquor
Morning
Work
Drunk
Taste
Sleep
Relationships
Bars
Type of
Alcohol Drinker

# Fitting a Latent Class Analysis Model

We use the **gsem** command to fit a latent class model

```
gsem (var1 var2 var3 var4 var5 var6 <- ), logit lclass(C 3)
```

Our observed variables are all binary so we use the **logit** option

The **lclass (C 3)** option specifies that we want to allow for three categories in a latent class variable named 'C'

```
gsem (like hard morning work drunk taste sleep rel bar <- ), logit lclass(C 3)
```

| | Coefficient | Std. err. | z | P>\|z\| | [95% conf. interval] | |
|---|---|---|---|---|---|---|
| 1.C | (base outcome) | | | | | |
| 2.C | | | | | | |
| _cons | .4287798 | .8185974 | 0.52 | 0.600 | -1.175642 | 2.033201 |
| 3.C | | | | | | |
| _cons | -1.521675 | .6923291 | -2.20 | 0.028 | -2.878615 | -.1647353 |

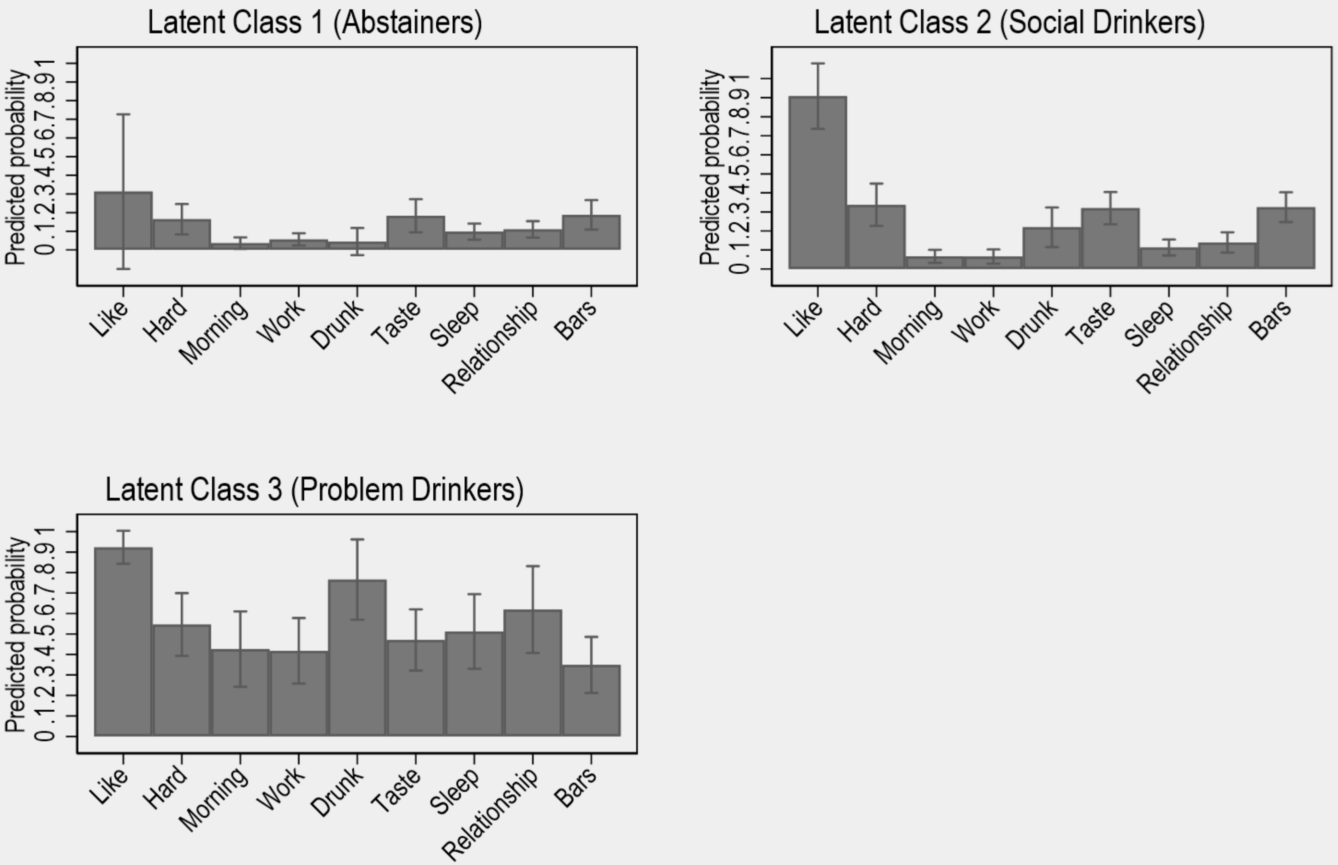
```
gsem (like hard morning work drunk taste sleep rel bar <- ), logit lclass(C 3)
```

| | Coefficient | Std. err. | z | P>\|z\| | [95% conf. interval] | |
|---|---|---|---|---|---|---|
| like | | | | | | |
| _cons | -.7899339 | .9859737 | -0.80 | 0.423 | -2.722407 | 1.142539 |
| hard | | | | | | |
| _cons | -1.628514 | .3027822 | -5.38 | 0.000 | -2.221956 | -1.035071 |
| morning | | | | | | |
| _cons | -3.295288 | .4608422 | -7.15 | 0.000 | -4.198523 | -2.392054 |
| work | | | | | | |
| _cons | -2.823975 | .3176524 | -8.89 | 0.000 | -3.446562 | -2.201388 |
| drunk | | | | | | |
| _cons | -3.069835 | .8750441 | -3.51 | 0.000 | -4.78489 | -1.35478 |
| taste | | | | | | |
| _cons | -1.496508 | .3030282 | -4.94 | 0.000 | -2.090432 | -.9025834 |
| sleep | | | | | | |
| _cons | -2.223255 | .2496319 | -8.91 | 0.000 | -2.712524 | -1.733985 |
| rel | | | | | | |
| _cons | -2.091981 | .2295699 | -9.11 | 0.000 | -2.541929 | -1.642032 |
| bar | | | | | | |
| _cons | -1.464306 | .2636135 | -5.55 | 0.000 | -1.980979 | -.9476332 |

```
estat lcmean, nose
```

| | Margin |
|---|---|
| 1 | |
| like | .3121829 |
| hard | .1640341 |
| morning | .0357332 |
| work | .0560423 |
| drunk | .0443688 |
| taste | .182947 |
| sleep | .0976816 |
| rel | .1098787 |
| bar | .1878096 |
| 2 | |
| like | .9082864 |
| hard | .3374061 |
| morning | .0667436 |
| work | .0654803 |
| drunk | .2191517 |
| taste | .3196319 |
| sleep | .1128118 |
| rel | .1398925 |
| bar | .3248739 |
| 3 | |
| like | .9232913 |
| hard | .5461479 |
| morning | .4263958 |
| work | .4179356 |
| drunk | .7654785 |
| taste | .471019 |
| sleep | .5123812 |
| rel | .6192121 |
| bar | .3488301 |

# Latent Class Analysis of Drinking Behaviour



| 3 Class model - modal LC | Freq. | Percent | Cum. |
|---|---|---|---|
| Class 1 Abstainer | 288 | 28.80 | 28.80 |
| Class 2 Social Drinker | 646 | 64.60 | 93.40 |
| Class 3 Problem Drinker | 66 | 6.60 | 100.00 |
| Total | 1,000 | 100.00 | |

```
. list id class3post1 class3post2 class3post3 expclass3 in 4/29
```

| | id | class3~1 | class3~2 | class3~3 | expclass3 |
|---|---|---|---|---|---|
| 4. | 4 | .1773417 | .7967343 | .025924 | Class 2 Social Drinker |
| 5. | 5 | .0407349 | .9341407 | .0251244 | Class 2 Social Drinker |
| 6. | 6 | .6856872 | .3119247 | .0023881 | Class 1 Abstainer |
| 7. | 7 | .0917532 | .9033101 | .0049367 | Class 2 Social Drinker |
| 8. | 8 | .2176675 | .7657966 | .0165359 | Class 2 Social Drinker |
| 9. | 9 | .2897012 | .6958461 | .0144527 | Class 2 Social Drinker |
| 10. | 10 | .850441 | .149076 | .000483 | Class 1 Abstainer |
| 11. | 11 | .922773 | .0770952 | .0001318 | Class 1 Abstainer |
| 12. | 12 | .1111802 | .8849409 | .003879 | Class 2 Social Drinker |
| 13. | 13 | .1737878 | .822176 | .0040362 | Class 2 Social Drinker |
| 14. | 14 | .850441 | .149076 | .000483 | Class 1 Abstainer |
| 15. | 15 | .1737878 | .822176 | .0040362 | Class 2 Social Drinker |
| 16. | 16 | .2082288 | .7899443 | .0018269 | Class 2 Social Drinker |
| 17. | 17 | .0193921 | .933044 | .0475639 | Class 2 Social Drinker |
| 18. | 18 | .0198259 | .6659695 | .3142046 | Class 2 Social Drinker |
| 19. | 19 | .6856872 | .3119247 | .0023881 | Class 1 Abstainer |
| 20. | 20 | .922773 | .0770952 | .0001318 | Class 1 Abstainer |
| 21. | 21 | .1239127 | .6570181 | .2190692 | Class 2 Social Drinker |
| 22. | 22 | .8212255 | .1780553 | .0007192 | Class 1 Abstainer |
| 23. | 23 | .0917532 | .9033101 | .0049367 | Class 2 Social Drinker |
| 24. | 24 | .4565689 | .5076751 | .035756 | Class 2 Social Drinker |
| 25. | 25 | .922773 | .0770952 | .0001318 | Class 1 Abstainer |
| 26. | 26 | .0000985 | .010796 | .9891055 | Class 3 Problem Drinker |
| 27. | 27 | .922773 | .0770952 | .0001318 | Class 1 Abstainer |
| 28. | 28 | .3537591 | .6449012 | .0013397 | Class 2 Social Drinker |
| 29. | 29 | .3537591 | .6449012 | .0013397 | Class 2 Social Drinker |

# Class Enumeration

- The model won’t tell you how many latent groups there are.
- The process of deciding how many classes to include is called ‘class enumeration’.
- To decide how many latent classes to include in your model you should consider statistical fit and substantive interpretability.

Akaike's information criterion and Bayesian information criterion

| Model | N | ll(null) | ll(model) | df | AIC | BIC |
|---|---|---|---|---|---|---|
| class1 | 1,000 | . | -4348.878 | 9 | 8715.755 | 8759.925 |
| class2 | 1,000 | . | -4251.208 | 19 | 8540.416 | 8633.664 |
| class3 | 1,000 | . | -4231.696 | 29 | 8521.392 | 8663.716 |
| class4 | 1,000 | . | -4225.508 | 39 | 8529.016 | 8720.419 |

Note: BIC uses N = number of observations. See [R] BIC note.

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are indexes of how well a model fits, which seek to balance the complexity of the model against the sample size.

AIC and BIC are calculated from maximum likelihood estimates.

Akaike's information criterion and Bayesian information criterion

| Model | N | ll(null) | ll(model) | df | AIC | BIC |
|---|---|---|---|---|---|---|
| class1 | 1,000 | . | -4348.878 | 9 | 8715.755 | 8759.925 |
| class2 | 1,000 | . | -4251.208 | 19 | 8540.416 | $\boxed{8633.664}$ |
| class3 | 1,000 | . | -4231.696 | 29 | $\boxed{8521.392}$ | 8663.716 |
| class4 | 1,000 | . | -4225.508 | 39 | 8529.016 | 8720.419 |

Note: BIC uses N = number of observations. See [R] BIC note.

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are indexes of how well a model fits, which seek to balance the complexity of the model against the sample size.

AIC and BIC are calculated from maximum likelihood estimates.

# Entropy

Entropy is a diagnostic statistic which measures the certainty of a posterior classification.

Entropy can help you consider how ‘crisp’ or ‘fuzzy’ your model solution is.

Values approaching 1 indicate that there is a clear delineation classes, values approaching 0 indicate that there is not a clear separation between the classes.

```
.
.           lcaentropy
Entropy =  .547709
```

Wang, M.-C., Deng, Q., Bi, X., Ye, H., & Yang, W. (2017). Performance of the entropy as an index of classification accuracy in latent profile analysis: A Monte Carlo simulation study. Acta Psychologica Sinica, 49(11), 1473-1482.

# Convergence

```
Iteration 281: log likelihood =  -4225.364  (not concave)
Iteration 282: log likelihood =  -4225.364  (not concave)
Iteration 283: log likelihood =  -4225.364  (not concave)
Iteration 284: log likelihood =  -4225.364  (not concave)
Iteration 285: log likelihood =  -4225.364  (not concave)
Iteration 286: log likelihood =  -4225.364  (not concave)
Iteration 287: log likelihood =  -4225.364  (not concave)
Iteration 288: log likelihood =  -4225.364  (not concave)
Iteration 289: log likelihood =  -4225.364  (not concave)
Iteration 290: log likelihood =  -4225.364  (not concave)
Iteration 291: log likelihood =  -4225.364  (not concave)
Iteration 292: log likelihood =  -4225.364  (not concave)
Iteration 293: log likelihood =  -4225.364  (not concave)
Iteration 294: log likelihood =  -4225.364  (not concave)
Iteration 295: log likelihood =  -4225.364  (not concave)
Iteration 296: log likelihood =  -4225.364  (not concave)
Iteration 297: log likelihood =  -4225.364  (not concave)
Iteration 298: log likelihood =  -4225.364  (not concave)
Iteration 299: log likelihood =  -4225.364  (not concave)
Iteration 300: log likelihood =  -4225.364  (not concave)
convergence not achieved
```

# Convergence

| startvalues() option | Description |
|---|---|
| factor | runs a factor analysis on all observed variables to obtain preliminary class predictions |
| randomid, draws(#) | randomly assigns observations to initial classes |
| randompr, draws(#) | randomly assigns initial class probabilities |
| jitter, draws(#) | randomly perturbs starting values from a Gaussian approximation to each outcome |
| classid *varname* | specifies a variable that identifies the initial class membership for each case |
| classpr *varlist* | specifies a list of variables that give the probability of membership in each class |

```
gsem (like hard morning work drunk taste sleep rel bar <- ), logit lclass(C 4) ///
        startvalues(randomid, draws(15) seed(1583))
```

# Convergence

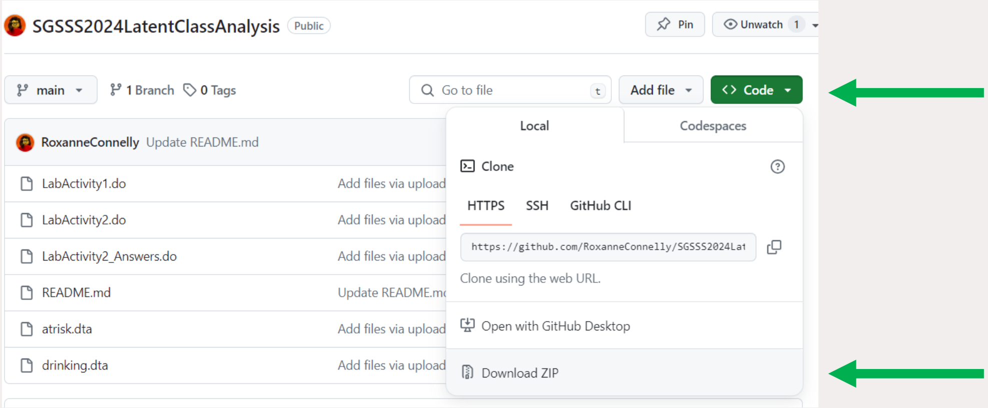
- In practice, researchers often find that their hypothesised LCA models do not converge.
- If this happens you should try different sets of starting values and make use of additional optional commands to help the model converge.
- Sometimes, you need to simply accept that the model will not converge (and report this transparently).

# Computer Lab Activity 1

# Computer Lab Activity 1

Materials are available here:

https://github.com/RoxanneConnelly/SGSSS2024LatentClassAnalysis




File: **LabActivity1.do**

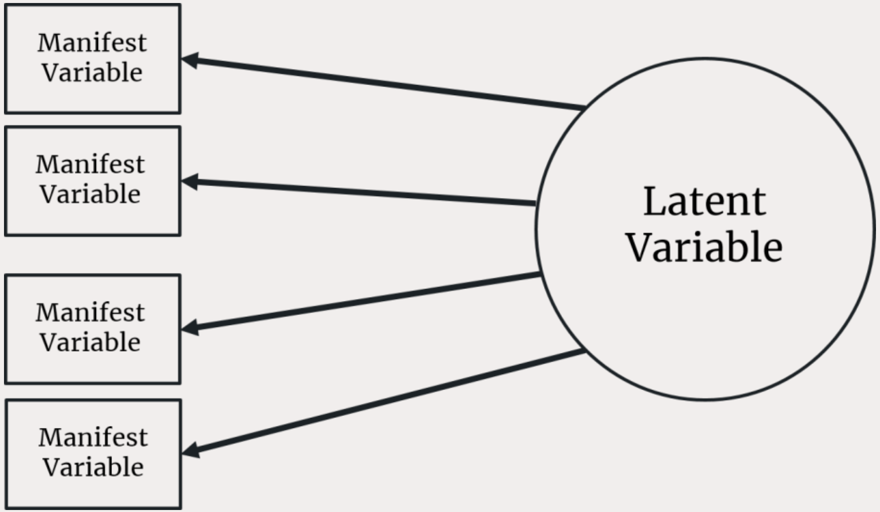
# Timetable

| | |
|---|---|
| 1000 – 1100 | Talk 1 |
| 1100 – 1130 | Lab Activities |
| 1130 – 1200 | Talk 2 |
| 1200 – 1230 | Lab Activities |
| 1230 – 1300 | Final Advice and Questions |

# Talk 2

# Latent Class Analysis

Latent class analysis (LCA) is a statistical procedure used to identify **qualitatively different** subgroups.



# Sample Size

‘more is better, but it depends’

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis: a guide to best practice. Journal of Black Psychology, 46(4), 287-311.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science, 4(4), 440- 461.

# Selecting Indicator Variables

More indicators and indicators which effectively differentiate between groups will lead to better results

Wurpts, I. C., & Geiser, C. (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a Monte Carlo study. Frontiers in Psychology, 5, 1-15. https://doi.org/10.3389/fpsyg.2014.00920

# Selecting the Number of Classes

Statistical Model Evaluation + Theoretical Interpretability

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling, 14(4), 535-569.

Vermunt, J. K. (2002). Latent class analysis of complex sample survey data: Application to dietary data: Comment. Journal of the American Statistical Association, 97(459), 736-737.

Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable centered analyses: Growth mixture modeling with latent trajectory classes. Alcoholism: Clinical & Experimental Research, 24(6), 882-891.

# Allocation to Latent Classes

Individuals are not placed in a latent class group with 100% certainty:

- You might consider an alternative to modal class allocation (e.g. analyse the probability of class membership)
- You might undertake some sensitivity analysis

```
. list id class3post1 class3post2 class3post3 expclass3 in 4/29
```

| | id | class3~1 | class3~2 | class3~3 | expclass3 |
|---|---|---|---|---|---|
| 4. | 4 | .1773417 | .7967343 | .025924 | Class 2 Social Drinker |
| 5. | 5 | .0407349 | .9341407 | .0251244 | Class 2 Social Drinker |
| 6. | 6 | .6856872 | .3119247 | .0023881 | Class 1 Abstainer |
| 7. | 7 | .0917532 | .9033101 | .0049367 | Class 2 Social Drinker |
| 8. | 8 | .2176675 | .7657966 | .0165359 | Class 2 Social Drinker |
| 9. | 9 | .2897012 | .6958461 | .0144527 | Class 2 Social Drinker |
| 10. | 10 | .850441 | .149076 | .000483 | Class 1 Abstainer |
| 11. | 11 | .922773 | .0770952 | .0001318 | Class 1 Abstainer |
| 12. | 12 | .1111802 | .8849409 | .003879 | Class 2 Social Drinker |
| 13. | 13 | .1737878 | .822176 | .0040362 | Class 2 Social Drinker |
| 14. | 14 | .850441 | .149076 | .000483 | Class 1 Abstainer |
| 15. | 15 | .1737878 | .822176 | .0040362 | Class 2 Social Drinker |
| 16. | 16 | .2082288 | .7899443 | .0018269 | Class 2 Social Drinker |
| 17. | 17 | .0193921 | .933044 | .0475639 | Class 2 Social Drinker |
| 18. | 18 | .0198259 | .6659695 | .3142046 | Class 2 Social Drinker |
| 19. | 19 | .6856872 | .3119247 | .0023881 | Class 1 Abstainer |
| 20. | 20 | .922773 | .0770952 | .0001318 | Class 1 Abstainer |
| 21. | 21 | .1239127 | .6570181 | .2190692 | Class 2 Social Drinker |
| 22. | 22 | .8212255 | .1780553 | .0007192 | Class 1 Abstainer |
| 23. | 23 | .0917532 | .9033101 | .0049367 | Class 2 Social Drinker |
| 24. | 24 | .4565689 | .5076751 | .035756 | Class 2 Social Drinker |
| 25. | 25 | .922773 | .0770952 | .0001318 | Class 1 Abstainer |
| 26. | 26 | .0000985 | .010796 | .9891055 | Class 3 Problem Drinker |
| 27. | 27 | .922773 | .0770952 | .0001318 | Class 1 Abstainer |
| 28. | 28 | .3537591 | .6449012 | .0013397 | Class 2 Social Drinker |
| 29. | 29 | .3537591 | .6449012 | .0013397 | Class 2 Social Drinker |

# The Naming Fallacy

Researchers usually assign names to the identified classes, and because of the complexity of the classes, may advertently engage in '**naming fallacy**', wherein the name of the class does not accurately reflect the class membership.

These limitations should be appreciated when reporting latent class analyses.

# Including Covariates in Latent Class Analysis

Does the composition of the latent class groups vary by socio-demographic characteristics?

Is membership of a latent class group associated with an outcome of interest?

# Including Covariates in Latent Class Analysis

Should you include the covariates in the same model used to identify the latent class solution?

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved threestep approaches. Political Analysis, 18(4), 450-469.

Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. Structural Equation Modeling: A Multidisciplinary Journal, 23(6), 782-797.

# Three-Step Approach

The three step approach:

- Develop your latent class model solution
- Add covariates to this model
- Fix the measurement parameters of the latent class model to those without the covariates included

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: A 3-step approach using Mplus. Structural Equation Modeling: A Multidisciplinary Journal, 21(3), 329-341.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. Political Analysis, 12(1), 3-27.

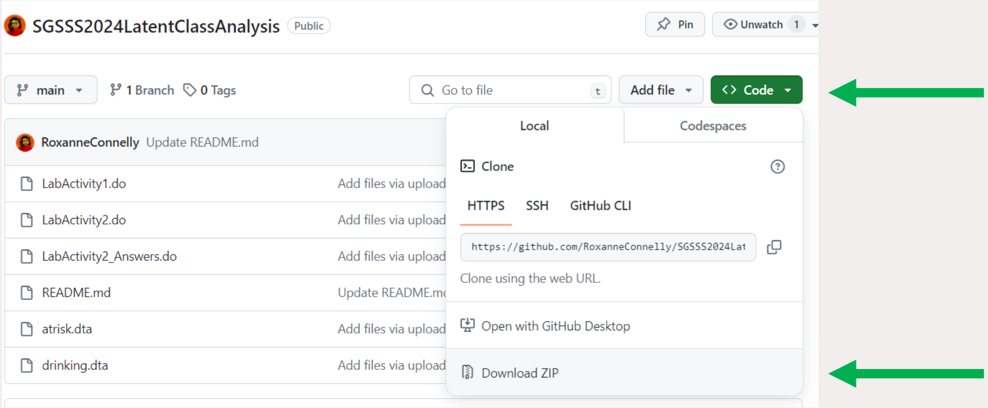
 Dr Roxanne Connelly

# Computer Lab Activity 2

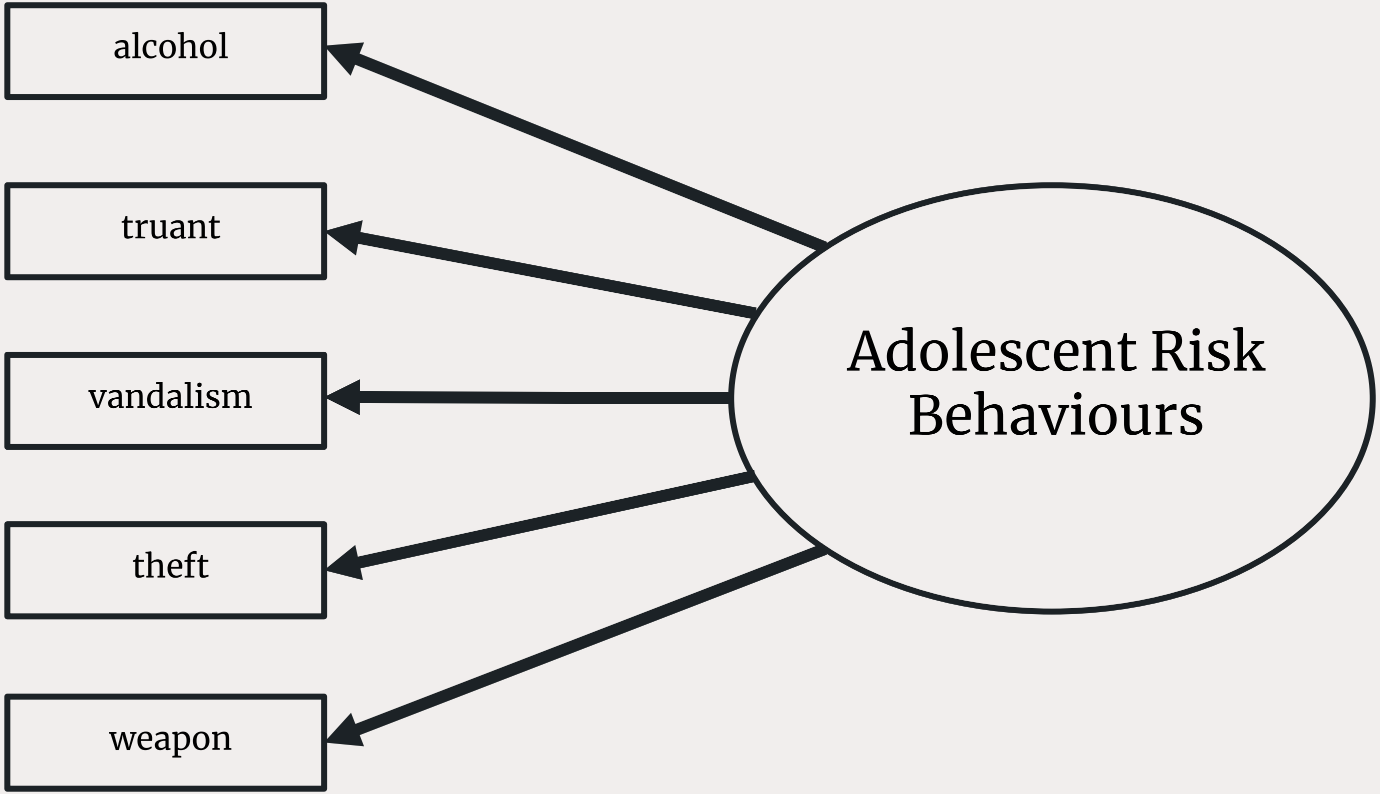
# Computer Lab Activity 2

Materials are available here:

https://github.com/RoxanneConnelly/SGSSS2024LatentClassAnalysis




File: **LabActivity2.do** or **LabActivity2_Answers.do**


alcohol
truant
vandalism
theft
weapon
Adolescent Risk
Behaviours

# Timetable

| | |
|---|---|
| 1000 – 1100 | Talk 1 |
| 1100 – 1130 | Lab Activities |
| 1130 – 1200 | Talk 2 |
| 1200 – 1230 | Lab Activities |
| 1230 – 1300 | Final Advice and Questions |

# Final Advice

# Final Advice: Share your Code

WORLD VIEW · 24 MAY 2018

## Before reproducibility must come preproducibility

Instead of arguing about whether results hold up, let's push to provide enough information for others to repeat the experiments, says Philip Stark.

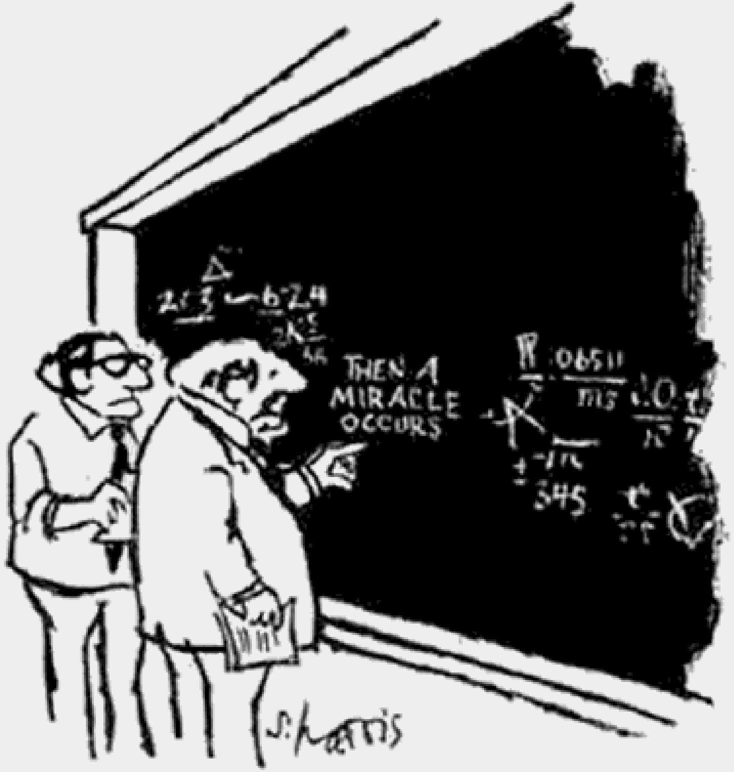
Philip B. Stark

Stark, P. B. (2018). Before reproducibility must come preproducibility. Nature, 557(7706), 613-614.

"A published paper should be considered as an advertisement of the research."

(Claerbout 1994)


Claerbout, J. (1994). Seventeen years of super computing and other problems in seismology [Paper presentation]. National Research Council Meeting on High Performance Computing in Seismology. http://sepwww.stanford.edu/sep/jon/nrc.html




"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

# Final Advice: Best Practice In Reporting LCA

Make your code available for others to view the full details of your analysis.

# Final Advice: Best Practice In Reporting LCA

Provide a rationale for the manifest variables selected.

# Final Advice: Best Practice In Reporting LCA

Provide a table of model fit statistics and report how the preferred model was selected.

**Table 2.** Model of fit statistics of the latent class models.

| Model | n | ll(model) | df | AIC | BIC | Entropy |
|---|---|---|---|---|---|---|
| Two Class | 616 | −12000.03 | 37 | 24074.06 | 24237.72 | 0.81 |
| Three Class | 616 | −11927.50 | 54 | 23962.99 | 24201.85 | 0.83 |
| Four Class | 616 | −11834.21 | 68 | 23804.42 | 24105.20 | 0.86 |
| Five Class | 616 | −11739.05 | 85 | 23648.10 | 24024.08 | 0.94 |
| Six Class | 616 | −11693.06 | 103 | 23592.12 | 24047.71 | 0.96 |
| Seven Class | 616 | −11670.82 | 120 | 23581.65 | 24112.43 | 0.95 |
| Eight Class | 616 | −11659.57 | 135 | 23589.13 | 24186.27 | 0.94 |
| Nine Class | 616 | −11605.72 | 142 | 23495.43 | 24123.54 | 0.93 |

Note: The sample size reduces by one case in the weighted analyses as one sample member has a weight of zero (unweighted $n = 617$).

# Final Advice: Best Practice In Reporting LCA

Provide a table summarising the characteristics of each latent class group.

**Table 3.** Prior probabilities for the six latent class solution.

| Ascribed label | 'Elite' | 'The established middle class' | 'The socially engaged middle class' | 'The culturally engaged middle class' | 'The traditional working class' | 'The precariat' |
|---|---|---|---|---|---|---|
| % Allocated to Latent Class | – | 31 | – | 11 | 35 | 19 |
| Prior Probabilities | | | | | | |
| *Economic Capital* | | | | | | |
| Equivalised Net Monthly Household Income (mean) | 7298.64 | 1559.42 | 1568.76 | 1494.50 | 1232.39 | 1348.61 |
| Own Home | 1.00 | 0.91 | 0.86 | 0.95 | 0.75 | 0.52 |
| *Social Capital* | | | | | | |
| Number of Friends (mean) | 6.14 | 5.16 | 25.03 | 7.45 | 4.91 | 3.96 |
| Active in Trade Union | 0.15 | 0.09 | 0.05 | 0.09 | 0.02 | 0.06 |
| Active in Professional Organisation | 0.68 | 0.16 | 0.37 | 0.37 | 0.04 | 0.00 |
| *Highbrow Cultural Capital* | | | | | | |
| Classical Music Concerts | 0.57 | 0.23 | 0.23 | 0.34 | 0.03 | 0.00 |
| Historic Buildings | 0.42 | 0.86 | 0.61 | 0.95 | 0.12 | 0.13 |
| Museums / Galleries | 0.85 | 0.84 | 0.78 | 0.98 | 0.21 | 0.29 |
| Plays/Theatre/Pantomime | 1.00 | 0.88 | 0.80 | 0.77 | 0.33 | 0.20 |
| Ballet Performance | 0.14 | 0.13 | 0.23 | 0.14 | 0.00 | 0.00 |
| *Emerging Cultural Capital* | | | | | | |
| Social Media | 0.58 | 0.64 | 0.42 | 0.88 | 0.83 | 0.28 |
| Regularly use the Internet | 1.00 | 0.92 | 1.00 | 1.00 | 1.00 | 0.24 |
| Gym/Fitness Classes | 0.60 | 0.45 | 0.55 | 0.93 | 0.44 | 0.22 |
| Rock/Pop/Jazz Concert | 0.82 | 0.59 | 0.58 | 0.69 | 0.22 | 0.10 |
| Cycle | 0.42 | 0.46 | 0.43 | 0.99 | 0.28 | 0.12 |
| Racquet Sports | 0.13 | 0.24 | 0.66 | 0.66 | 0.09 | 0.06 |
| Golf | 0.27 | 0.19 | 0.19 | 0.28 | 0.07 | 0.03 |

Note: – indicates that the values in the cell have been suppressed for statistical disclosure control purposes. $n = 616$. The sample size reduces by one case in the weighted analyses as one sample member has a weight of zero (unweighted $n = 617$).

# Final Advice: Best Practice In Reporting LCA

Don’t forget about the characteristics of your data (e.g. complex samples and missing data).

# Final Advice: Extensions to LCA

This is only the beginning, Latent Class Analysis can do much more...

- You can extend Latent Class Analysis to include different types of variables.
- You can undertake LCA separately by group (e.g. men and women).
- There might be more than one latent variable.

# Workshop Summary

You should now know:

- What latent class analysis is.
- How latent class analysis is used in the social sciences.
- How to undertake a latent class analysis.
- How to select the number of latent classes.
- How to allocate sample members to latent classes using the modal allocation method.
- How to overcome (some) model convergence issues.
- How to report a latent class analysis.

# Questions

# Latent Class Analysis

June 2024

Dr Roxanne Connelly

roxanne.connelly@ed.ac.uk