Modelling alcohol-consumption in New Zealand: A Bayesian conditional copula-based regression approach

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

Jose Romeo, Taisia Huckle & Sally Casswell

SHORE & Whariki Research Centre, College of Health

e-mail: j.romeo@massey.ac.nz

Introduction

Most regression models focus on explaining distributional aspects of one single outcome variable. Modelling multiple outcome variables and their dependence structure have become very popular.

Examples:

- Medicine - Time to relapse of a disease and time of death
- Marketing - purchase of different products
- Actuarial sciences - age of death of husbands and wives ("widowhood effect")
- Finance - understanding the relationship between some financial stocks (time series)
- Ecological community data - species display association with one another
- Alcohol research - quantity on a typical occasion and frequency of drinking

➪ Alcohol consumption can be measure in two dimensions: the amounts consumed in a typical occasion and the frequency of drinking (Casswell et al., 2016).

Alcohol consumption IAC study New Zealand data: The International Alcohol Control (IAC) study is an international cohort study of alcohol use and alcohol policy relevant behaviours coordinated at SHORE (Huckle et al., 2018). In 2011 a national stratified sample of households was surveyed in NZ:

- Alcohol consumption data were collected using a beverage- and location-specific measure in last 6 months
- For each place, they were asked how often they drank there and what they would drink on a typical occasion at that location
- This information was used to calculate the outcomes:

  Y_1 : quantity on a typical occasion (standard drinks)
  Y_2 : frequency of drinking
- Covariate information: Gender, age, ethnicity, level of education, log equivalised income and poverty line.

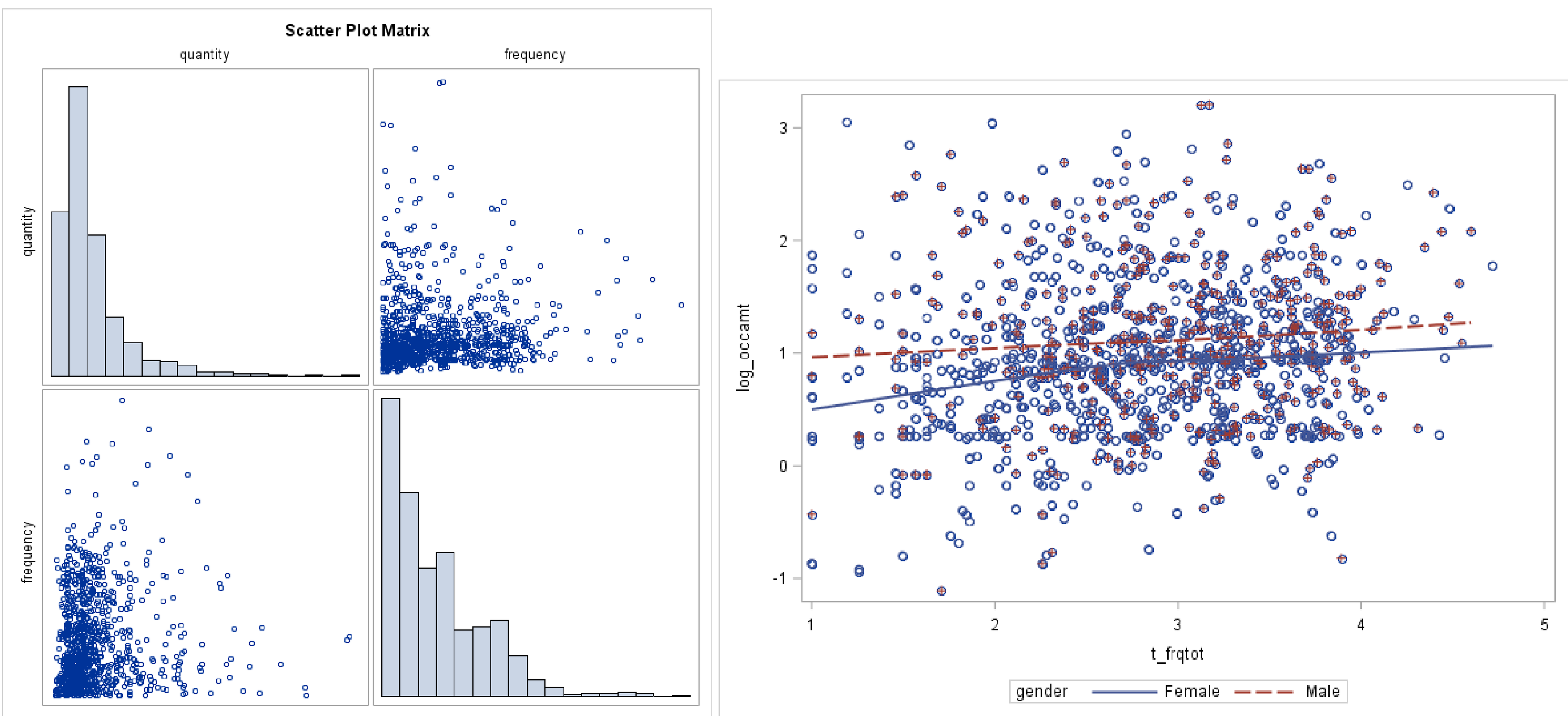

Figure 1: Quantity and frequency of drinking, Kendall's tau = 0.12

- Some previous studies on the factors affecting quantity and frequency of alcohol consumption have treated these two outcomes independently.
- We fit conditional copula-based regression models for explaining the joint distribution of the typical amount and the frequency of alcohol consumption.

➪ Copula-based regression models allow such an analysis since they enable the separation of the marginal outcome distributions and the dependence structure modelled by a specific copula function.

➪ We include socio-demographic factors not only in the marginal distributions through the mean and the dispersion but also in the copula parameter allowing a direct modelling of the association and flexibility in model specification.

Copulas

(See e.g. Mai & Scherer, 2012)
Suppose we have two continuous variables of interest Y_1 and Y_2 where

- $Y_1 \sim F_1(\cdot)$ and $Y_2 \sim F_2(\cdot)$, marginal distributions
- $F(y_1, y_2) = \Pr\{Y_1 \le y_1, Y_2 \le y_2\}$, joint distribution of the pair (Y_1, Y_2)
- Then the joint cdf can be written as a copula function relating the marginal distributions as

$$F(y_1, y_2) = C_\alpha(F_1(y_1), F_2(y_2)), \quad \alpha \in \mathcal{A}$$

- The dependence parameter α measures the strength of association between Y_1 and Y_2
- The joint pdf is given by

$$f(y_1, y_2) = c_\alpha(F_1(y_1), F_2(y_2)) \cdot f_1(y_1) \cdot f_2(y_2)$$

Based on a specific copula C_α we can compute Kendall's tau coefficient that is a measure of concordance. Considering $u_j = F_j(y_j)$, $j = 1, 2$,

$$\tau_\alpha = 4 \int\int_{[0,1]^2} C_\alpha(u_1, u_2) dC_\alpha(u_1, u_2) - 1$$

Conditional copula-based regression model

(See e.g. Nikoloulopoulos & Karlis, 2010, Klein & Kneib, 2016)

- Dependence parameter α is mostly treated as constant
- We would like to explain the association between outcomes not just the marginals

Suppose we have two outcomes of interest Y_1 and Y_2 given covariate information x where

- $Y_1|x \sim F_1(\cdot|x)$ and $Y_2|x \sim F_2(\cdot|x)$, conditional marginal distributions
- The conditional joint cdf can be written as a conditional copula function relating the conditional marginal distributions as

$$F(y_1, y_2|x) = C_{\alpha(x)}(F_1(y_1|x), F_2(y_2|x)|x)$$

- Copula parameter depends on covariates $\alpha(x)$ through an appropriate link function $h(\cdot)$ as in GLM

Suppose (y_{i1}, y_{i2}) independent realizations of (Y_1, Y_2), and covariate information x_i, with $i = 1, \ldots, n$.

Then we consider

- $Y_1|x \sim F_1(y_1|\mu_1(x), \sigma_1(x))$, $\quad Y_2|x \sim F_2(y_2|\mu_2(x), \sigma_2(x))$
- $(Y_1, Y_2)|x \sim C_{\alpha(x)}(F_1(y_1|x), F_2(y_2|x)|x)$

and for $j = 1, 2$, linear predictors

- $h_j(\mu_j(x_i)) = \beta_{j0} + \beta_{j1}x_{1i} + \beta_{j2}x_{2i} + \ldots + \beta_{jp}x_{pi}$
- $h_{\sigma_j}(\sigma_j(x_i)) = \gamma_{j0} + \gamma_{j1}x_{1i} + \gamma_{j2}x_{2i} + \ldots + \gamma_{jp}x_{pi}$
- $h(\alpha(x_i)) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \ldots + \alpha_p x_{pi}$

Application: Alcohol consumption IAC study NZ data

Marginals modelling:

- Quantity: $\quad \log y_{i1} \, |x_i \sim \text{Normal}\left(\mu_{i1}(x_i), \sigma^2_{i1}(x_i)\right)$
- Frequency: $\quad \sqrt[4]{y_{i2}} \, |x_i \sim \text{Normal}\left(\mu_{i2}(x_i), \sigma^2_{i2}(x_i)\right)$, where $\mu_{ij}(x_i) = \beta' x_i$ and $\sigma_{ij}(x_i) = \exp(\gamma' x_i)$

Copula model	DIC	Copula model	DIC
Independent	4343.1	PVF†	4341.5
Gaussian	4346.5	Inverse Gaussian	**4340.1**
Clayton	4341.1	Frank	4357.1

Table 1: Model selection criterion

Modelling:	*Mean*				*Dispersion*				*Dependence*	
	Y_1		Y_2		Y_1		Y_2		α	
Effect	Median	SD	Median	SD	Median	SD	Median	SD	Median	SD
Intercept	0.865	0.028	2.877	0.025	-0.962	0.045	-0.675	0.042	0.417	0.391
Gender: Male vs Female	0.119	0.019	0.108	0.023	0.083	0.044			0.551	0.319
Poverty line: Under vs Over	-0.046	0.067								
Age	-0.012	0.002	0.004	0.002	-0.015	0.004			0.035	0.022
Education: Low vs High	0.377	0.075								
Medium vs High	0.169	0.038								
Ethnicity: Maori vs NZ Euro	0.353	0.067	-0.155	0.079						
Pacific vs NZ Euro	0.492	0.132	-0.547	0.147						
Asian vs NZ Euro	-0.280	0.089	-0.415	0.099						
Log equivalised income			0.206	0.032	-0.301	0.059	-0.202	0.057	-0.761	0.466
Age × Under poverty line	-0.012	0.005								

Table 2: Parameter estimates, Inverse Gaussian copula model

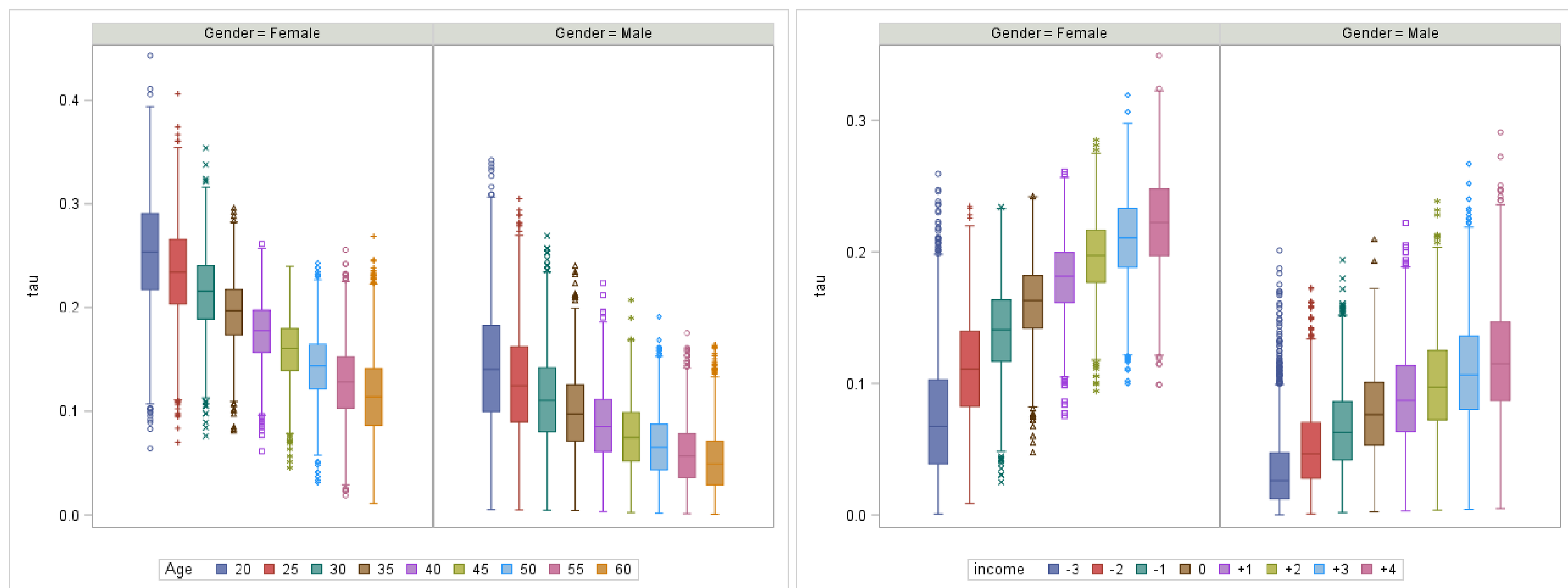

Figure 2: Estimated Kendall's tau between Quantity and Frequency by gender, age and income

➪ **Interpretation**: Young women with high income: as they drink more frequently they also drink more quantities. Older men with low income: quantity and frequency are almost independent.

Discussion and future work

- Overall, the association between quantity and frequency of drinking is not strong: Kendall's $\tau = 0.15$ (95% CI: $0.11 - 0.19$).
- Flexibility of models based on copulas: dependence structure and marginals distributions

Future work

- About the marginals (e.g. skew-normal, negative binomial distributions)
- A more efficient posterior computation (see e.g. Wichitaksorn et al. 2018)
- To include more countries members of the IAC study, e.g., Australia, Thailand, Vietnam, England, Scotland, South Africa: ➪ Hierarchical model

Main References

Casswell, S., Huckle, T., Wall, M. & Parker, K. (2016). Policy relevant behaviours mediate the relationship between socio-economic status and alcohol consumption - analysis from the IAC study. Alcoholism: Clinical and Experimental Research, 40(2), 385-392.

Huckle, T., Casswell, S., Mackintosh, A.-M. et al. (2018). The International Alcohol Control Study: methodology and implementation. *Drug and Alcohol Review*, 37(2), S10–S17.

Klein, N. & Kneib, T. (2016). Simultaneous inference in structured additive conditional copula regression models: a unifying Bayesian approach. *Statistics and Computing*, 26(4), 841–860.

Mai, J. & Scherer, M. (2012). Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications. Imperial College, Boca Raton.

Nikoloulopoulos, A.K. & Karlis, D. (2010). Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, 37(9), 1555–1568.

†Romeo, J.S., Meyer, R. & Gallardo, D.I. (2018). Bayesian bivariate survival analysis using the power variance function copula. *Lifetime Data Analysis*, 24, 355–383.

Wichitaksorn, N., Gerlach, R. & Choy, B. (2019). Efficient MCMC estimation of some elliptical copula regression models through scale mixtures of normals. *Applied Stochastic Models in Business and Industry*, 35(3), 808–822.

Modelling alcohol-consumption in New Zealand: A Bayesian conditional copula-based regression approach.

Romeo Nunez, J

2019-11-25