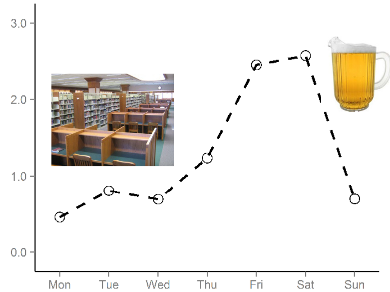


Modeling Cyclical Patterns in Daily College Drinking Data with Many Zeroes

David Huh, Ph.D., Debra Kaysen, Ph.D. & David Atkins, Ph.D.
Department of Psychiatry & Behavioral Sciences, University of Washington

Modern Modeling Methods Conference
May 21, 2013
Storrs, Connecticut

Supported by NIAAA grant
T32 AA007455

From simulated to real data…

- Now we transition from simulation to an applied case where cycles are relevant: College Drinking Data
- The backdrop: Current approaches to modeling alcohol consumption may be missing rising and falling patterns across days of the week
- In addition to cycles, drinking data has the added complication of huge stacks of zeroes.
- This talk will focus on modeling cyclical patterns with a specific type of zero-altered model: a Hurdle model

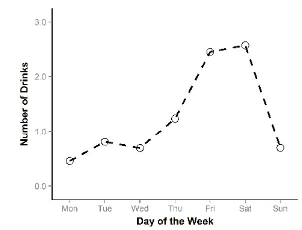
Why do people drink to excess?

- Let's back up to the big picture. One of the broad questions alcohol researchers care about:
  - Why do people, including college students, drink to excess?
  - Important because of the negative consequences[1,2]:
    - Greater alcohol-related morbidity and mortality
    - Greater interpersonal violence
    - Greater suicide risk
    - Poorer educational attainment
- Interest in clarifying factors that predict drinking has driven substance use researchers to pursue intensive longitudinal designs.
  - e.g., Participants report alcohol use one or more times a day for a set time frame (e.g., 30 days)[3].

[1]Hingson, Heeren, Winter, & Wechsler, 2005; [2]Perkins, 2002; [3]Kaysen et al., in press

Drinking changes predictably over the week

- Not surprisingly, daily drinking data shows a predictable pattern over days of the week
  - Greater drinking on weekends as opposed to weekdays
  - What is the best way to incorporate this rising and falling rhythm into a statistical model?



- In the alcohol literature, the most common approach is some type of dummy coding.
  - Most common is a single dummy variable for weekend vs. weekday.(e.g., 4)
  - Also seen: dummy codes for individual days of the week(e.g. 5)

[4]Neighbors et al., 2011; [5]Simons, Dvorak, Batien, & Wray, 2010

Dummy variables are easy, but problematic

- Advantage
  - Dummy variables approaches are simple to implement
- Disadvantages
  - Single dummy variable approaches imply an abrupt transition across days of the week.
  - Multiple dummy variables can precisely capture shifts, but are unwieldy, especially with covariates.
- An attractive alternative is to model data with periodicity as a sinusoidal function

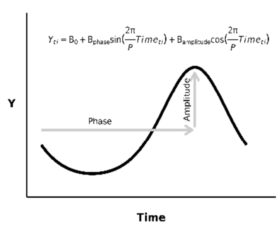
Saturated time models are unwieldy

	Single Dummy	Saturated Dummy	Cyclical terms
Time Predictor(s)	1. Weekend vs. weekday	1. TUE (vs. Monday) 2. WED 3. THU 4. FRI 5. SAT 6. SUN	1. Amplitude 2. Phase
Total	= 1	= 6	= 2
With a covariate	1. Covariate 2. Weekend x Covariate	1. Covariate 2. TUE x Covariate 3. WED x Covariate 4. THU x Covariate 5. FRI x Covariate 6. SAT x Covariate 7. SUN x Covariate	1. Covariate 2. Amplitude x Covariate 3. Phase x Covariate
Total w/ Covariate	= 3	= 13	= 5

Cyclical predictors are straightforward

- Simple transformation of the linear time predictor (e.g., day of the week) into sine and cosine terms to represent[6,7]:
  - The magnitude (amplitude)
  - A location of a regular peak (phase)

1. Multiply the TIME variable by 2π
2. Divide by the PERIOD (P = 7 days)



- The Amplitude term is the cosine of the above value
- The Phase term is the sine of the above value

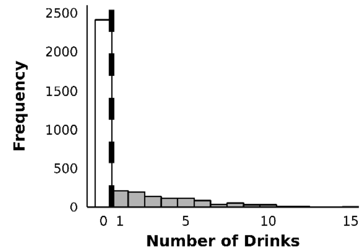
[6]Fluri & Levri, 1999; [7]Pinheiro & Bates, 2000

Cyclical Models have a long history

- Cyclical models of time series data back more than 4 decades
  - In the biomedical literature, known as "cosinor analysis."
  - Early example by Tong (1974) with circadian (i.e., 24 hour) rhythms
  - Commonly used to model physiological processes.
- Also adopted within the ecology field
  - Flury and Levri (1999) examined 24-hour foraging patterns of snails with cyclical logistic regression
- Pinherio & Bates' (2000) classic mixed effects modeling book showed the use cyclical terms in random effects models.
- To date, rarely used in psychology, but they have attractive features that make them suited for behavioral outcome data.

Important to attend to excess zeroes…

- Distribution of the data is another important consideration
  - Behavioral outcomes assessing short intervals will often contain a lot of zeroes.
    - Substance use
    - Sexual behavior[8]
- Zeroes may be a key feature of the phenomena of interest and not just a nuisance of the data.

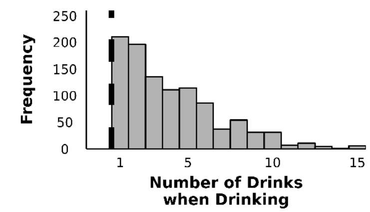


  - In the context of alcohol use, the processes that predict…
    - **<u>the decision to drink at all</u>** may be quite different than
    - **<u>how much one drinks one they start</u>**

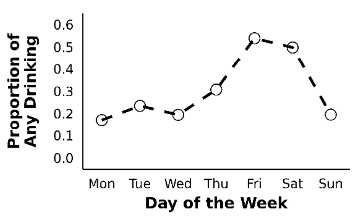
[8]Bodenmann, Atkins, Schär, & Poffet, 2010

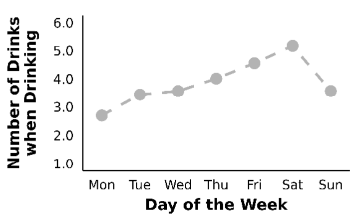
Hurdle models give meaning to zeroes

- Hurdle models, a type of two-part model are a practical approach
- A threshold must be crossed from zero into positive counts.
- As illustrated with the DASH data, the outcome is effectively divided into two parts.
  - No drinking vs. any drinking:
    - Logistic regression
  - Amount of drinking when drinking:
    - Truncated count regression



Hurdle and cyclical models can be combined





- Cyclical parameters can be used to model both the binary and positive count models.
- Can we model trends in <u>any drinking</u> and the <u>amount of drinking when drinking</u> with cyclical models as a sinusoidal function?
- Plots of mean drinking across days in our recent study of college women suggest cyclical parameters are a reasonable candidate for both.

An example with longitudinal data

- Project DASH
  - Intensive longitudinal study on the association of PTSD and drinking[9]
    - 172 female undergraduates
    - Baseline assessment followed by a 30-day monitoring period
    - On each monitored day, participants completed two PDA assessments

[9]Kaysen et al., in press

The longitudinal drinking outcome

- PDA-assessed daily number of standard drinks **(Outcome)**[10]
  - "How many standard drinks have you had in the past 24 hours?"
    - Participants provided with the definition of a standard drink:
      - **Equivalent to:**
        - **12 oz. can of beer**
        - **5 oz. glass of wine**
        - **1.5 oz. shot of liquor**

- [10]Grant, Stewart, O'Connor, Blackwell & Conrad, 1999

A covariate to predict drinking patterns

- Self-reported Social Drinking Motives
  - Five items from the Drinking Motives Questionnaire-Revised (DMQ-R[11])
  - Example item:
    - "Because it is what most of my friends do when we get together."
      - 1 = *never/almost never*
      - 5 = *almost always/always*

- [11]Grant, Stewart, O'Connor, Blackwell & Conrad, 1999

The regression approach used

- Hurdle negative binomial mixed effect regression
  - Maximum likelihood estimation
    - glmmADMB package in R[12,13]
  - Two separate regressions:
    - Binary logistic regression
      - no drinking vs. any drinking
    - Truncated negative binomial regression
      - number of drinks when drinking
  - Random effects for each model determined by likelihood ratio tests.

- [12]Skaug, Fournier, Nielsen, Magnusson, & Bolker, 2012; [13]R Core Team, 2013

Two sets of cyclical vs. dummy variable comparisons

- Drinking Trends Only (Models 1-3)
  - Baseline models to evaluate the suitability of the cyclical versus dummy variable approaches to modeling drinking data.
- Prediction of Drinking Trends (Models 4-6):
  - Extend each baseline model with social drinking motives as a moderator of time to assess their performance when evaluating a covariate.
- Model's evaluated using BIC and AIC
  - BIC's goal is identifying the true model[14]
  - AIC's goal is the prediction of new data[15]

- [14]O'Connell & McCoach, 2008; [15]Kuha, 2004

A split decision for the cyclical model

	Overall		Binary		Count	
Baseline models	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**
1. Single dummy	+40	+78	+40	+72	+49	+59
2. Cyclical terms	✓	+40	+49	+75	✓	✓
3. Saturated dummy	+10	✓	✓	✓	+60	+50

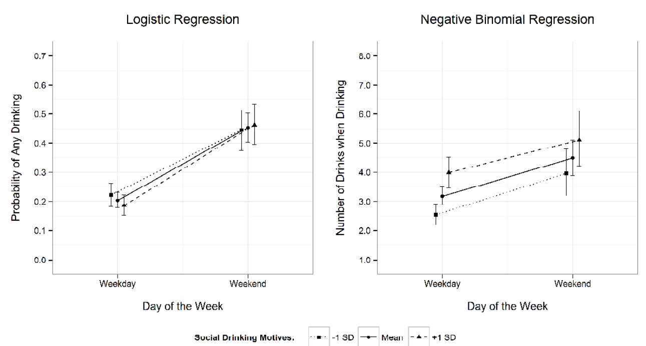
"Best" model
second
third

- Overall, strong evidence per BIC that the cyclical model was the better model of drinking.
  - AIC preferred the saturated model, but cyclical model better predicted the data than a single dummy variable
- However, this obscures differences by sub-model…
  - Cyclical model was a better model for the amount of drinking
  - Saturated model was a better model for the probability of any drinking

Cyclical and dummy models all evidenced significant moderation

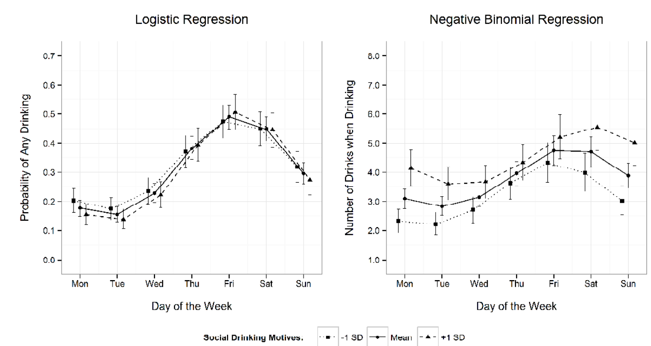
- The next set of models added a covariate (social motives)
- The weekend, cyclical, and weekday models all detected statistically significant moderation effects in both the probability and amount of drinking
  - Skipping the parameter-by-parameter breakdown, but the complete regression tables are on supplementary slides.
- The key difference is how informative a picture each model paints about
  - Trends over the week.
  - Differences in those trends by level of social motives

Lack of specificity in weekend moderation models…



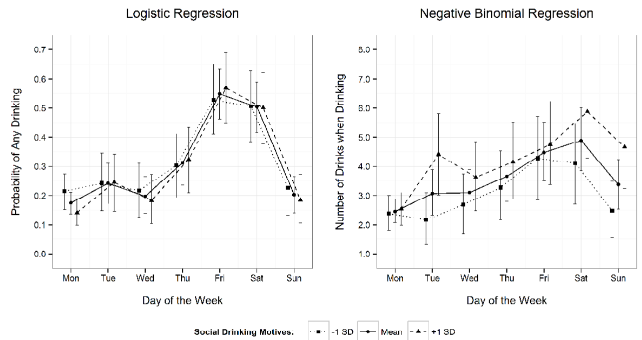
- <u>Logistic model:</u> Greater weekend rise in the probability of any drinking among those with higher social drinking motives.
- <u>Count model:</u> Smaller weekend rise in the amount of drinking from among those with higher social drinking motives.

Finer-grained predictions in a cyclical model



- <u>Logistic model:</u> A more similar probability of any drinking across the week among those with higher motives
- <u>Count model:</u> Higher and more consistent number of drinks when drinking among those with higher motives

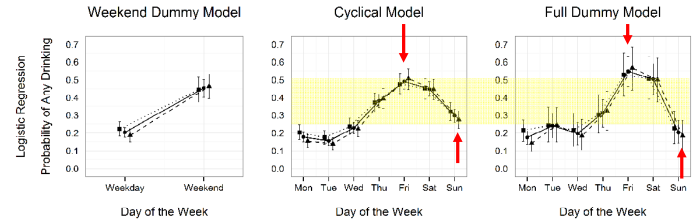
Piecewise predictions in a full dummy model



- Logistic model: Pairwise elevations on Thu and Fri in the probability of drinking compared with Mon among those with higher social motives.

- Count model: Pairwise elevations on Tue and Sun in the number of drinks on compared with Mon among those with higher social motives.

Cyclical model errs in late week prediction of drinking probability



- Poorer fit of the cyclical logistic model (seen earlier) coincides with cyclical/full model divergence in late week predictions. Eg.,
  - Friday/Saturday predictions about 5% too low
  - Sunday predictions about 10% too high.

Cyclical regression covariates a qualified success

- Cyclical regression covariates were a practical alternative to full dummy variables for modeling drinking patterns
  - More elegant interpretation that focuses on the magnitude of the peak.
  - Introducing a covariate for time added far fewer parameters in the cyclical model.
  - More difficult to understand because time divided into many pieces
  - Unable to estimate random slopes in the saturated model.

- Cyclical regression parameters were easily combined with hurdle regression
  - Modeling zeroes versus non-zeroes as a separate process led to richer picture of drinking as a two-part process.

Interesting insights from comparing approaches

- In particular, not all aspects of drinking were perfectly sinusoidal
  - The number of drinks when drinking had a rhythmic pattern that was reasonably approximated by cyclical terms.
  - However, either of the dummy variable approaches were a better model for the probability of any drinking

- That the very simple weekend model fit better than the cyclical model provides insight into day-to-day differences in the decision to drink.
  - Suggests there is some homogeneity in the probability of drinking during weekdays, rather than a continuous rise and fall implied by a cyclical model.

Questions?

- For post-conference questions, contact:
  - David Huh (dhuh@uw.edu).

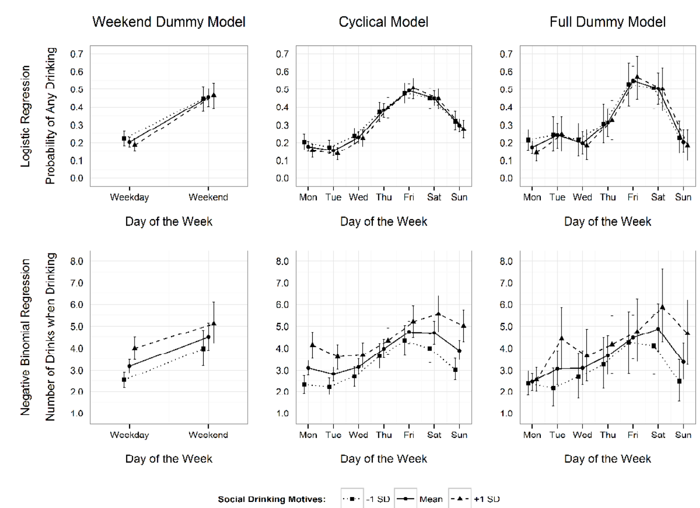
Recommended Reading

- Example of cyclical models using mixed effect logistic regression:
  - Bodenmann, G., Atkins, D. C., Schär, M., & Poffet, V. (2010). The association between daily stress and sexual activity. Journal of Family Psychology, 24, 271–279. doi:10.1037/a0019365
- Tutorial on longitudinal count regression methods (including zero-altered models):
  - Atkins, D. C., Baldwin, S. A., Zheng, C., Gallop, R. J., & Neighbors, C. (2013). A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of Addictive Behaviors*, *27*, 166–177. doi:10.1037/a0029508

References

- Bodenmann, G., Atkins, D. C., Schär, M., & Poffet, V. (2010). The association between daily stress and sexual activity. *Journal of Family Psychology*, *24*, 271–279. doi:10.1037/a0019365
- Flury, B. D., & Levri, E. P. (1999). Periodic logistic regression. *Ecology*, *80*, 2254–2260. doi:10.1890/0012-9658(1999)080[2254:PLR]2.0.CO;2
- Hingson, R., Heeren, T., Winter, M., & Wechsler, H. (2005). Magnitude of alcohol-related mortality and morbidity among U.S. college students ages 18-24: Changes from 1998 to 2001. *Annual Review of Public Health*, *26*, 259–279. doi:10.1146/annurev.publhealth.26.021304.144652
- Kaysen, D. L., Atkins, D. C., Simpson, T. L., Stappenbeck, C. A., Blaynew, J. A., Lee, C. M., & Larimer, M. E. (in press). Proximal relationships between PTSD symptoms and drinking among female college students: Results from a daily monitoring study. *Psychology of Addictive Behaviors*.
- Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research, 33,* 188–229.
- Neighbors, C., Atkins, D. C., Lewis, M. A., Lee, C. M., Kaysen, D., Mittmann, A., … Rodriguez, L. M. (2011). Event-specific drinking among college students. Psychology of Addictive Behaviors, 25, 702–707. doi:10.1037/a0024051
- O'Connell, A. A., & McCoach, D. B. (2008). Multilevel Modeling of Educational Data. IAP.
- Perkins, H. W. (2002). Surveying the damage: A review of research on consequences of alcohol misuse in college populations. *Journal of Studies on Alcohol and Drugs, Sup(14),* 91.
- Pinheiro, J. C., & Bates, D. M. (1995). Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics, 4,* 12–35.
- Simons, J. S., Dvorak, R. D., Batien, B. D., & Wray, T. B. (2010). Event-level associations between affect, alcohol intoxication, and acute dependence symptoms: Effects of urgency, self-control, and drinking experience. *Addictive Behaviors, 35,* 1045–1053. doi:10.1016/j.addbeh.2010.07.001
- Skaug, H., Fournier, D., Nielsen, A., Magnusson, A., & Bolker, B. (2012). glmmADMB: Generalized linear mixed models using AD model builder. *R package version 0.7.3, 4*.
- Tong, Y. L. (1976). Parameter Estimation in Studying Circadian Rhythms. *Biometrics, 32,* 85–94. doi:10.2307/2529340

Comparing all models simultaneously



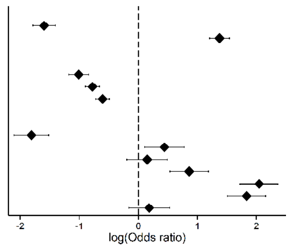
Regression Table of Baseline Models

Logistic Regression

Weekend Dummy Model	B	SE	p
Intercept	-1.60	0.10	<.001
Weekend vs. Weekday	1.38	0.09	<.001

Cyclical Model	B	SE	p
Intercept	-1.01	0.09	<.001
Amplitude (cosine)	-0.78	0.06	<.001
Phase (sine)	-0.61	0.06	<.001

Full Dummy Model	B	SE	p
Intercept	-1.81	0.15	<.001
Tuesday vs. Monday	0.44	0.17	.01
Wednesday vs. Monday	0.15	0.17	.40
Thursday vs. Monday	0.86	0.17	<.001
Friday vs. Monday	2.04	0.16	<.001
Saturday vs. Monday	1.83	0.17	<.001
Sunday vs. Monday	0.18	0.18	.30

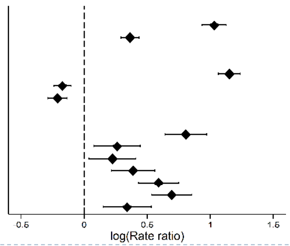


Truncated Negative Binomial Regression

Weekend Dummy Model	B	SE	p
Intercept	1.03	0.05	<.001
Weekend vs. Weekday	0.36	0.04	<.001

Cyclical Model	B	SE	p
Intercept	1.15	0.04	<.001
Amplitude (cosine)	-0.17	0.04	<.001
Phase (sine)	-0.21	0.04	<.001

Full Dummy Model	B	SE	p
Intercept	0.81	0.08	<.001
Tuesday vs. Monday	0.26	0.09	.01
Wednesday vs. Monday	0.22	0.10	.02
Thursday vs. Monday	0.39	0.09	<.001
Friday vs. Monday	0.59	0.08	<.001
Saturday vs. Monday	0.69	0.08	<.001
Sunday vs. Monday	0.34	0.10	<.001



Regression Table of Moderation Models

Logistic Regression / **Truncated Negative Binomial Regression**

Weekend Dummy Model	B	SE	p	B	SE	p
Intercept	-1.60	0.10	<.001	1.07	0.04	<.001
Social Motives	-0.13	0.10	.18	0.20	0.05	<.001
Weekend vs. Weekday	1.38	0.09	<.001	0.38	0.03	<.001
Social Motives × Weekend	0.17	0.08	.04	-0.09	0.04	.02

Cyclical Model	B	SE	p	B	SE	p
Intercept	-1.01	0.09	<.001	1.15	0.04	<.001
Social Motives	-0.05	0.09	.52	0.18	0.04	<.001
Amplitude (cosine)	-0.78	0.06	<.001	-0.18	0.03	<.001
Phase (sine)	-0.61	0.06	<.001	-0.20	0.04	<.001
Social Motives × Amplitude	-0.13	0.06	.03	0.10	0.04	<.01
Social Motives × Phase	-0.02	0.06	.78	-0.01	0.04	.86

Full Dummy Model	B	SE	p	B	SE	p
Intercept	-1.83	0.15	<.001	0.81	0.08	<.001
Social Motives	-0.28	0.15	.06	0.03	0.08	.70
Tuesday vs. Monday	0.46	0.17	.01	0.22	0.10	.02
Wednesday vs. Monday	0.16	0.18	.36	0.23	0.10	.02
Thursday vs. Monday	0.88	0.17	<.001	0.39	0.09	<.001
Friday vs. Monday	2.07	0.17	<.001	0.60	0.08	<.001
Saturday vs. Monday	1.85	0.17	<.001	0.68	0.08	<.001
Sunday vs. Monday	0.20	0.18	.26	0.31	0.10	<.01
Social Motives × Tuesday	0.29	0.17	.09	0.33	0.10	.001
Social Motives × Wednesday	0.17	0.18	.35	0.12	0.10	.21
Social Motives × Thursday	0.34	0.17	.04	0.09	0.09	.32
Social Motives × Friday	0.38	0.17	.02	0.02	0.08	.80
Social Motives × Saturday	0.27	0.17	.10	0.15	0.08	.08
Social Motives × Sunday	0.13	0.18	.45	0.29	0.10	<.01