R101: A practical guide to making R your everyday statistical tool (PSY532)

Lecture 4

Working with skewed, categorical and clustered data

Programme (lecture and seminar)

- **Skewed** outcome variable: generalized linear modelling
- **Categorical** outcome variable:

		Outcome	
		Two categories	*More than two categories*
Predictor(s)	*One categorical*	Chi-squared test	Chi-squared test
	More than one categorical and/or continuous	Logistic regression	Multinomial logistic regression^ Ordered logistic regression^

^ Covered in readings (Baguley) but not in lecture

Red denotes version of generalized linear modelling

- **Clustered** outcome variable: zero-inflated (mixture) modelling; multilevel modelling
- For each analysis type: descriptive statistics, running the analysis, diagnostics, and reporting/article examples

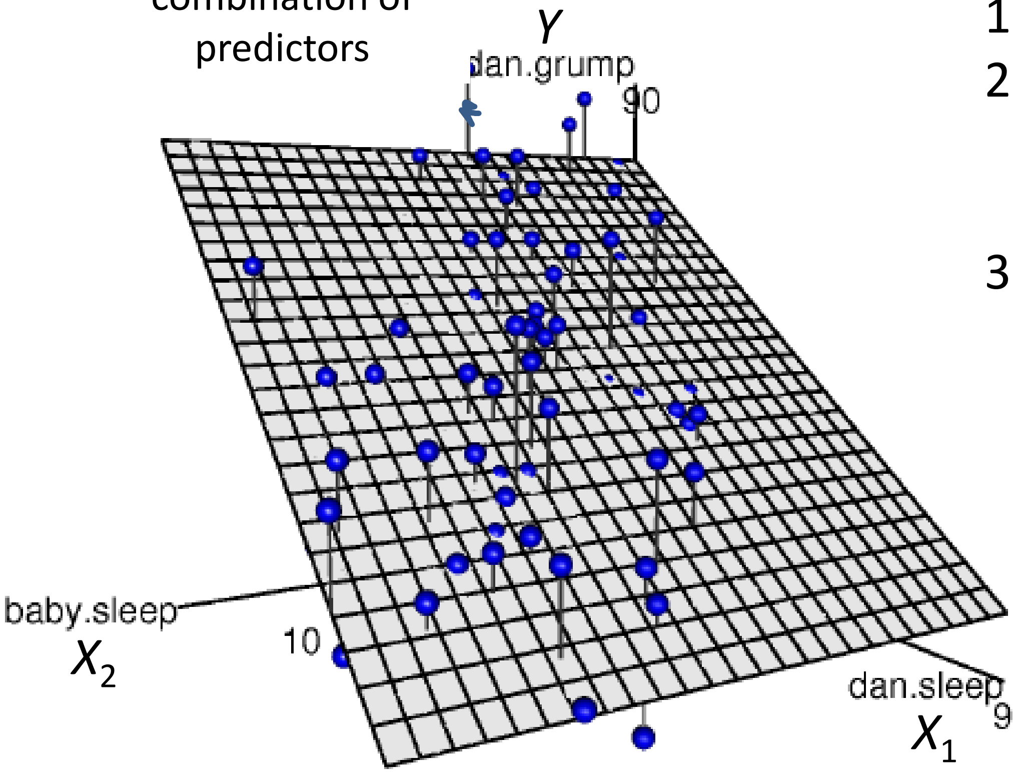
Skewed outcome variable

Generalized linear models

Reading: Baguley Ch 17, Gelman Ch 6

- A principled alternative to transformation when the outcome variable is constrained – e.g., when the outcome variable is:
  - a count of members in a category (this is where chi-squared tests and logistic regression come in, and these will be covered later in the lecture)
  - a count of occurrences in a calendar year or some other period of time (this is where Poisson regression is useful)
  - highly skewed, and therefore better represented by a Poisson or negative binomial distribution (this is the case we will focus on in this section)
- If your outcome variable has a highly skewed distribution (e.g., life satisfaction scores or people's estimates of a count, as in our SS data), these models are worth fitting after the initial ANOVA or regression to see if they fit the data better.
- Terminology: the ANOVA and regression techniques we have discussed so far are instances of general linear modelling, a special case of *generalized* linear modelling

Additive combination of predictors



Three components:

1. Additive combination of predictors
2. Random component/ family of distribution for *Y* (e.g., Poisson, negative binomial, binomial)
3. The link function (e.g., logarithm, logistic). Analysts typically use canonical (default) link functions as listed in the R help file. Logarithmic function is canonical for Poisson and negative binomial random components. Logistic function is canonical for binomial distributions.

$$Y_i = b_2 X_{i2} + b_1 X_{i1} + b_0 + \epsilon_i$$

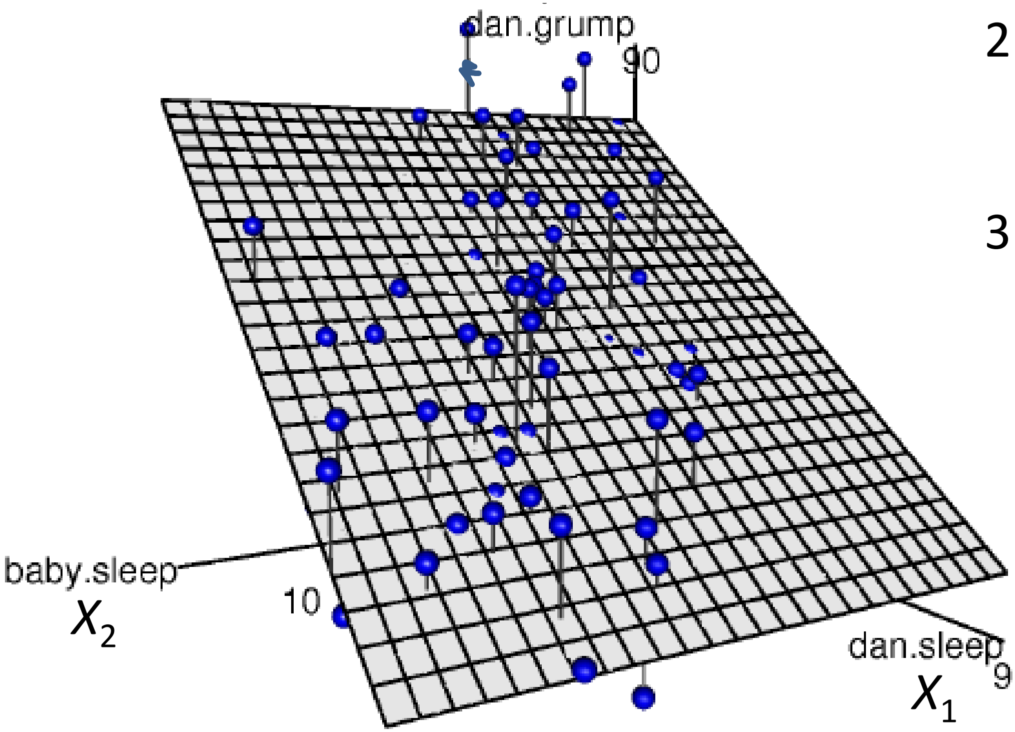
implies...

$Y_i \sim$ Normal(mean = $b_2 X_{i2} + b_1 X_{i1} + b_0$, variance = σ^2 which does not depend on X_1 or X_2)

General linear model

2. Normally distributed *Y (*~ means “distributed as”)
3. Identity link function (Y *is* the additive comb.)

Additive combination of predictors



Three components:

1. Additive combination of predictors
2. Random component/ family of distribution for *Y* (e.g., Poisson, negative binomial, binomial)
3. The link function (e.g., logarithm, logistic). Analysts typically use canonical (default) link functions as listed in the R help file. Logarithmic function is canonical for Poisson and negative binomial random components. Logistic function is canonical for binomial distributions.

$Y_i \sim$ Norm(mean = $b_2X_{i2} + b_1X_{i1} + b_0$, variance = σ^2)

2. Normally distributed *Y*
3. Identity link function (Y *is* the additive comb.)

$Y_i \sim$ Poisson(mean = variance = exp($b_2X_{i2} + b_1X_{i1} + b_0$))

2. Poisson-distributed *Y*
3. Logarithmic link: hence the exponentiation

$Y_i \sim$ NegBin(mean = log(exposure parameter) $\cdot$ exp($b_2X_{i2} + b_1X_{i1} + b_0$), overdispersion = *w*)

2. Negative-binomial-distributed *Y*
3. Logarithmic link: hence the exponentiation

Generalized linear modelling to test Hypothesis 2 from our dataset

Hypothesis

In the literature on short-term memory, the first few words in word lists are consistently found to be remembered better than the other words. Thus, more wins should be remembered in the *descending condition* relative to the others.

Descriptive statistics

Same as for ANCOVA performed in Lecture/Seminar 2

Running the analysis

`glm` in `base` package and, for negative binomial random components, `glm.nb` in `MASS` package. By default, non-sequential sums of squares. Can use `anova.glm` function to obtain sequential sums of squares.

Diagnostics: Does the model fit well? Does it fit better than the original ANOVA/regression?

- Residual deviance: looking for lower deviance values
- AIC: the smaller this is, the better the model
- Cook's distances: view plot to make sure there are not a few stand-out influential points. Cook's distances three times greater than the mean of the Cook's distances are a cause of concern, especially if there are only one or two (easily deletable) associated cases.

Go to script

Reporting the analysis – as in Results section

- Table 1 (or very clear graph) showing means and SDs of outcome variable across levels of the categorical predictors.
- In text: To test Hypothesis 2, a generalized linear model was fitted using the glm.nb package in *R* Version 3.1.0 with percentage of remembered wins as the outcome variable, success-slope and question wording as the predictors, and background beliefs (*Drake Beliefs About Chance* total score) as a covariate. The analysis (with Type II sums of squares) revealed significant effects of success-slope (LR χ^2 (3) = 11.56, *p*=.01) and question wording (LR χ^2 (1) = 48.95, *p*<.001). There was also a significant interaction (LR χ^2 (3) = 15.90, *p*=.001), together with a significant effect of the covariate (LR χ^2 (1) = 14.81, *p*<.001). Planned comparisons of the Descending condition's mean to those of other groups revealed a significant difference between the U-shaped and Descending groups (*p* = .01), and the Flat and Descending groups (*p* = .02).

Categorical outcome variable

Chi-square tests

Reading: LSR Ch 12

- Used for determining whether two categorical variables are significantly associated, based on counts displayed in a frequency (or “contingency”) table.

	Desc.	U-shaped	Asc.	Flat
Strategy “no”	74	72	57	65
Strategy “yes”	11	14	22	19

- Logic: Compare observed frequencies to what would be expected under the null hypothesis in light of your degrees of freedom.
- Effect size: Cramer’s *V*, ranging from 0 (no association) to 1 (perfect association)
- Assumptions:
  - The two variables in the table are independent (otherwise, for a “repeated measures” design, use McNemar test)
  - The expected frequency in each cell is sufficiently large (a rough guide is that, if the expected frequency in any cell is less than 5, use Fisher exact test)

Chi-squared test of a version of Hypothesis 1 in our dataset

Hypothesis

If people perceive themselves to be problem-solving (learning a strategy) in games of chance, the illusion of natural control should be greater in the *Ascending slope* condition relative to the *Descending slope* condition. This implies that the number of people saying “Yes, I had a strategy” in response to the PostStrategyPresent question should be associated with success-slope condition (SeqCond).

Descriptive statistics

Frequency table as on previous slide. `xtabs` in `base` package.

Running the analysis

`associationTest` in `lsr` package (`fisher.test` and `mcnemar.test` are in the `stats` package)

Go to script

Reporting the analysis

- Frequency table
- Text: A chi-square test indicated that the association between success-slope and reporting of a strategy approached significance ($\chi^2(3) = 6.84$, $p = .07$, Cramer’s $V = 0.14$). This might reflect the slightly higher incidence of strategy reports in the *Ascending* condition, and possibly the *Flat* condition also. [Notice that it is useful to comment on what the association implies - where were the differences?]

Logistic regression

Reading: Baguley Ch 17

- In its basic form, can be used only with a binary outcome variable (e.g., alive or dead; ill or not). However, can include any number of categorical and continuous predictors.
- Uses: hypothesis tests and prediction. There is an example of each in the Study Materials Lecture 4 folder. For another good example of prediction, search Masaryk University Catalogue or Google Scholar for: “Prediction of probable Alzheimer's disease in memory-impaired patients: A prospective longitudinal study” (article available only in HTML). Here, we will focus on hypothesis testing.
- A generalized linear model:
  - Binomial random component
  - Logit link function

Hypothesis from a new dataset: a nationally representative Czech survey on drug use and gambling (CG1 in workspace)

Survey questions

- What type of game/s have you played in the last 12 months? *Select all that apply:*
  - ❑ slot machines
  - ❑ online slot machines
  - ❑ virtual gaming machines (e.g. virtual roulette)
  - ❑ casino games (e.g. roulette, cards, dice)
  - ❑ card tournaments outside of casinos (e.g. poker)
  - ❑ sports and non-sports betting at betting offices/bookmakers
  - ❑ online betting at registered Czech operators
  - ❑ other online betting (e.g. online poker, roulette)
  - ❑ lotteries
  - ❑ I did not play on any of these

Over 2000 people answered the survey but they were excluded from this analysis if they answered lotteries only or did not play.

- In the last 12 months, how often have you played any of the games listed above? *Response options:* (0) only once, (1) less than once a month, (2) once a month, (3) several times a month (2-3 times), (4) at least once a week (1-2 times), (5) several times a week (3-4 times), (6) every day or almost every day (5-7 times per week)

Hypothesis from a new dataset: a nationally representative Czech survey on drug-use and gambling

Survey questions (cont.)

- Problem Gambling Severity Index (Ferris & Wynne, 2001)

Thinking about the last 12 months...

- have you bet more than you could really afford to lose?
  - 0: never
  - 1: sometimes
  - 2: most of the time
  - 3: almost always
- have you needed to gamble with larger amounts of money to get the same feeling of excitement?
- did you go back another day to try to win back the money you lost?
- ... 9 questions total
- Interpretation of total score:
  - 0-2: no risk
  - 3-7: at risk
  - 8 or more: pathological gambler

Hypothesis from a new dataset

Research question: Is online gambling a particularly dangerous type of gambling?
Past studies have found pathological gambling to be associated with a higher reported frequency of play, and also with play on a wider variety of gambling types. If online gambling leads to an increased probability of gambling pathology, respondents' pathological gambling scores (scores on the PGSI) should relate not only to the frequency and variety of gambling activity, but also to whether at least one of the gambling activities was performed online.

Analysis plan
Hierarchical logistic regression with pathological gambling status (no risk vs. at risk/pathological) as the binary dependent variable.

- Step 1: Frequency of play (ordered), ranging from 0 (once only) to 6 (almost every day). Entered first.
- Step 2: Variety of games played (categorical): one vs. more than one
- Step 3: Whether one of the played games was online (categorical): yes vs. no

Descriptive statistics and running the analysis
Means and frequency counts of outcome and predictors; then `glm` in `base`

Script

Hypothesis from a new dataset

Hierarchical logistic regression

Use the `anova` function at each step (Steps 1, 2 and 3 in our example) to assess the contribution of each predictor to the model: is the reduction in deviance significant with the introduction of the predictor at each step? The displayed analysis of deviance test will provide this information through a deviance statistic and associated *p*-value, determined based on a chi-square distribution.

This approach can have some problems when there are continuous predictors in the model. In such a case, you can conclude that a predictor is significant if its inclusion in the model reduces the deviance and AIC. See Baguley p. 683.

Hypothesis from a new dataset

Interpretation of coefficients

If performing a hierarchical logistic regression, it makes sense to do this for the final model, with all the predictors included.

Some helpful plots are also shown in the script.

Three sources of information:

- coefficients expressed as log odds using `summary(modelname)`: multiply coefficient by .25 to obtain a measure of percentage change in probability of moving from 0 (absence) to 1 (presence) on the outcome variable with each one unit change in the predictor (or with a shift from the reference category to the listed category if the predictor is categorical, as in our example)
- Wald *z*-tests shown as part of `summary(modelname)`: tell us whether the listed predictor is a significant predictor in that its slope in the model is significantly greater than 0
- coefficients expressed as odds ratios using `exp(modelname$coefficients)`: tell us the odds of moving from 0 (absence) to 1 (presence) on the outcome variable as a proportion of the odds of doing so when the predictor is one unit less or is at the reference level (see further explanation on next slide and in the script)

Hypothesis from a new dataset

Interpretation of coefficients (cont.):

The odds ratio

- If equal to 1, the odds are the same, so there is no change in the outcome variable with changes in the predictor
- If greater than 1, the odds increase with the predictor’s unit increase or change from reference category
- If less than 1, the odds decrease with the predictor’s unit increase or change from reference category

A mixture of categorical and continuous predictors

The log odds and odds ratios are likely to be much smaller for continuous predictors than for categorical ones. This is because unit changes are generally smaller than entire category shifts. To make different types of predictors comparable in the same model, it is therefore useful to calculate the coefficient for more than one unit change (e.g., some meaningful number or two standard deviations). This is just a case of multiplying the predictor by the chosen number of units.

The difference between “odds” and “probabillity”

In a race, you may see the odds for your horse, Camilla, are 8 to 1, which are the odds AGAINST winning. This means in nine races Camilla would be expected to win 1 and lose 8. In probability terms, Camilla has a probability of winning of 1/9, or 0.111. But the odds of winning are 1/8, or 0.125. Odds are actually the ratio of two probabilities...

$$\frac{\text{probability of event}}{1 - \text{probability of event}}$$

Diagnostics

- Collinearity: Use chi-square tests and/or ANOVAs to determine whether any predictors are related to each other. Relationships between predictors make the odds ratios for individual coefficients less interpretable, since each ratio expresses the effect of a unit change in the associated predictor when all others remain constant. But what if the predictor can only change when another factor changes? Of course, the same considerations apply in linear regression as well, but to a lesser degree. An alternative model should be considered with one of the two related predictors removed. We do this here (model4).
- Influential points: If any are detected, try running the regression without them.
- Sparse data: If you have relatively few people in one of the two categories of the outcome variable, your model will perform little better (if not worse) than a model that places everyone into one group (e.g., “no risk”). This situation is very common, especially if you are dealing with illnesses, which tend to affect the minority (e.g., Alzheimer’s, pathological gambling). Try to make the counts in the categories of the outcome variable as equal as possible through data collection or careful selection of the two categories in the existing data set (e.g., in our analysis here we look only at 206 people who reported gambling in the preceding 12 months, not the whole survey sample of 2000 people).
- No normality of residuals or homogeneity of variance assumption.

Reporting the analysis

- Frequency table and/or description based on descriptive statistics:
  - 197 people who reported gambling in the last 12 months answered all three relevant questions: gambling types selection, playing frequency and pathological gambling (PGSI)
  - Among these participants, the play frequency variable was restructured to contain the following categories: once total (*N* = 40), less than once per month (*N* = 31), once per month (*N* = 30), 2-3 times per month (*N* = 43), and weekly (*N* = 53).
  - 108 of the 197 people did not report gambling online.
  - etc.
- Table 1 (next slide): Slope coefficients, Wald tests, standard errors and odds ratios.
- Text: In a hierarchical logistic regression with PGSI category as the outcome variable, playing frequency was entered into the model first, followed by game variety, and online experience in the third step. Results are presented in Table 1. It can be seen that online experience did not account for PGSI category membership over and above the marginally significant effects of playing frequency and game variety. Since chi-square tests revealed all three predictors to be related, we conducted a logistic regression that included only game variety and online experience and considered the effect of their interaction. Likelihood ratio tests following a Type II Sums of Squares ordering revealed the effect of online gambling to not be significant (LR $\chi^2(1) = .38$, $p = .53$), while game variety emerged again as a marginally significant predictor (LR $\chi^2(1) = 3.58$, $p = .06$). There was also no significant interaction effect (LR $\chi^2(1) = .55$, $p = .46$).

Table 1. Final model of a hierarchical logistic regression in which playing frequency, game variety and online gambling experience were entered in that order as predictors of pathological gambling risk

	Slope	*S.E.*	Odds ratio
Intercept	-2.06	0.49	0.13
Less than once a month (reference: only once)	-0.77	0.88	0.46
Once a month (reference: only once)	0.11	0.70	1.12
2-3 times per month (reference: only once)	0.89	0.60	2.44
Weekly (reference: only once)	1.07*	0.59	2.93
Game variety (reference: one)	0.72^	0.42	2.06
Online gambling experience (reference: no)	-0.08	0.43	0.92

* *p* = .06, ^ *p* = .08 (Wald *z*-test)

Clustered outcome variable

Zero-inflated (mixture) modelling

Reading: Baguley Ch 17

- Useful when the data has a high proportion of 0s (e.g., “Not at all”; “Completely disagree”; “Never played”)
- Two stages:
  - Logistic regression to estimate the effect of one or more predictors on the outcome variable being “0” or “other”
  - Generalised linear modelling (Poisson, negative binomial) to estimate the effect of one or more predictors (not necessarily the same as in the logistic regression) on the values of the outcome variable in the “other” category. If trying a number of random components (e.g., Poisson and negative binomial), choose the model with Log likelihood closer to zero.

Examples based on success-slope Hypothesis 1 extension

- The Extensions section in *Assignment 2 answers* (Study Materials)shows that many people in the SS dataset scored zero on the illusion of natural control measure.The stem-and-leaf plot for PostNaturalIoC with its many zeroes is shown on the next slide.
- The SS data also includes responses (0-10) expressing degree of agreement with whether “It was all chance” is an accurate description of how goals were achieved in the soccer-themed gambling game (Lecture 1 Slide 13). The histogram on the next slide shows that, as might be expected, many people agreed fully (10) with this statement. 10s can be considered “zeroes”.
- In Assignment 2, we also discovered a significant interaction between prior beliefs and success-slope in predicting the illusion of natural control, except that the assumptions for the associated ANOVA were not met. Here, we use success-slope and prior beliefs as predictors in a zero-inflated model.
- Research questions: To what extent are responses to “It was all chance” (PostListCHANCE) and the illusion of natural control (PostNaturalIoC) influenced by success-slope (SeqCond), prior beliefs (PreDBC_Total) and their interaction? In relation to success-slope, Hypothesis 1 predicts:
  - Higher illusion of control in the Ascending condition, compared to Descending
  - Less agreement with the “It was all chance” statement in Asc. condition

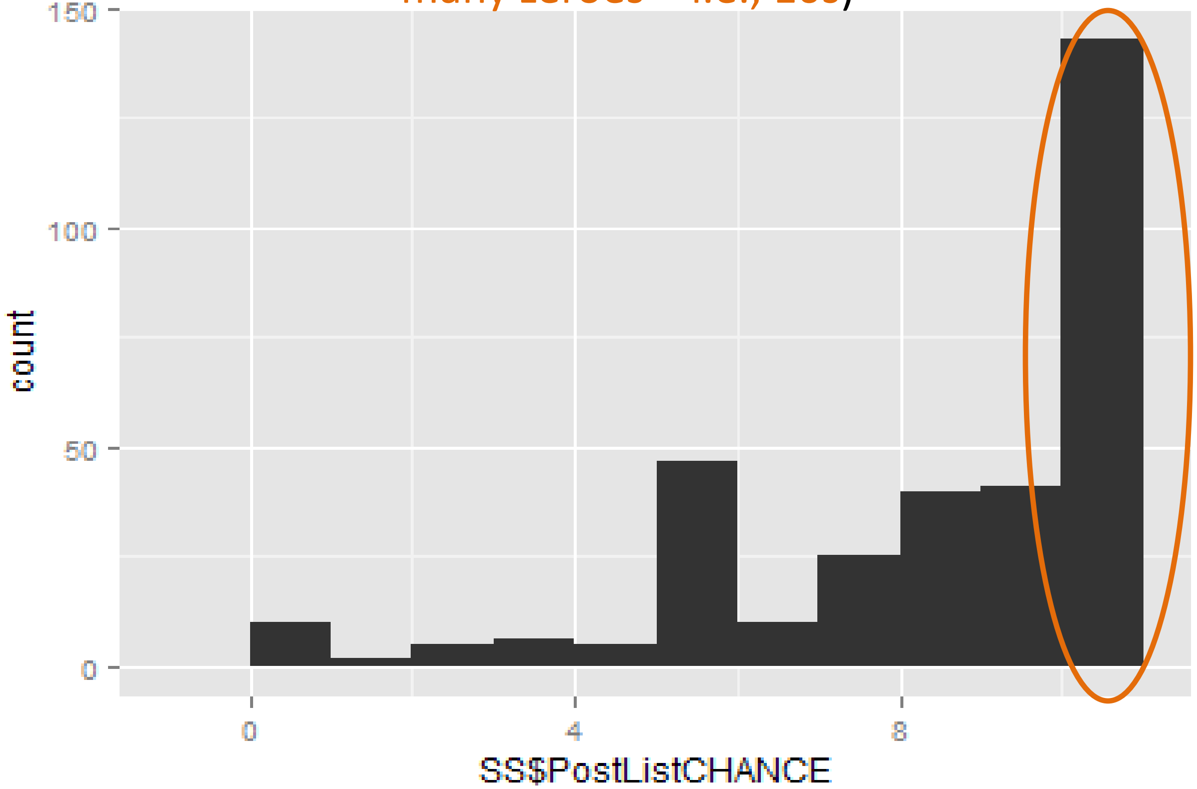
Descriptive statistics

PostNaturalIoC stem-and-leaf plot (many zeroes visible)

The decimal point is at the |

```
0 | 00000000000000000000000000000000000000000001111333344444444444
0 | 555666666666666888888899999999
1 | 00000000111113333333333444444444
1 | 555555566666688888889999999
2 | 00000111113333334444444444
2 | 55555555555566666688899999999
3 | 00000111133333334444
3 | 55566666688888889999
4 | 000000001133444444444
4 | 5555566666888999
5 | 000001114444
5 | 555566666888999
6 | 0000344
6 | 555699
7 | 00013344
7 | 9
8 | 33
```

PostListCHANCE histogram (also many zeroes – i.e., 10s)



Running the analysis

- `zeroinfl` function in the `pscl` package
- Outcome variable must be in integer form, with a clearly defined zero (see script for examples: with "It was all chance", for example, we reverse and categorise the responses)
- Random components in the generalized linear models can be Poisson, negative binomial, etc.
- Interpretation:
  - "Count model coefficients (negbin with log link):" to be interpreted as with generalized linear models. For NaturalIoC, we see a significant effect of prior beliefs, but not success-slope.
  - "Zero-inflation model coefficients (binomial with logit link):" To be interpreted as with logistic regression. For "It was all chance", we see a significant effect of prior beliefs on whether a person fully agreed (0) with the statement.

Reporting the analysis

- Table 1: Descriptive statistics table showing not only means and standard deviations, but also the frequency of zeroes across levels of the predictor(s)
- Table 2: Model fit results
- Text: A zero-inflated mixture model with a Poisson random component for the count model was fitted using the `pscl` package in R Version 3.1.0. Prior beliefs, success-slope, and their interaction were the predictors for both the count and zero-inflated parts of the model. As the model fit results in Table 2 show, the only observed significant effect was that of prior beliefs on non-zero illusion-of-control scores.

Table 2. Estimated effects in the count and logistic regression parts of the zero-inflated model. The Descending condition is the reference category for success-slope.

	Estimate	*S.E.*
Count model		
Intercept	-0.31	0.43
U-shaped minus Descending	0.12	0.57
Ascending minus Descending	0.10	0.60
Flat minus Descending	0.13	0.55
Prior beliefs	0.02*	0.01
Prior beliefs: U-shaped minus Descending	0.004	0.01
Prior beliefs: Ascending minus Descending	0.01	0.01
Prior beliefs: U-shaped minus Descending	0.001	0.01
Logistic regression		
Intercept	0.51	1.49
U-shaped minus Descending	1.03	1.79
etc.		

* *p* = .01 (Wald *z*-test)

Multilevel modelling

Reading: Gelman & Hill Ch 11-13

- Variables manipulated or gathered to represent a theoretically meaningful range in the population (e.g., full range of success-slope conditions) are modelled as “fixed” effects.
- Categorical variables potentially influencing the outcome variable but showing variability just in the sample (e.g., the range of prior beliefs in the sample) are modelled as “random” effects.
- Random variables can affect the intercept, the slope, or both.
- The resultant analysis is a compromise between “complete pooling” (investigation of a manipulated variable only) and “no pooling” (investigation of only sample-specific effects). Regression coefficients for the sample-specific variable(s) are pulled towards their mean (completely pooled) level. This is known as “shrinkage”. If a cluster within the sample is very small (e.g., if a school has only two respondents in a survey where there are evident school clusters), the coefficient for that group is pulled *further* towards the mean to adjust for the uncertainty arising from the small number of people in the cluster.

Shrinkage towards completely pooled regression slope (dotted line) across clusters (Minnesota counties) with different sample sizes: Counties (the random variable) affects the intercept, but not the slope of the regression line. More shrinkage is evident with smaller sample size (N = 2), and the intercepts in counties with larger sample sizes are generally closer to the completely pooled line anyway (examples).

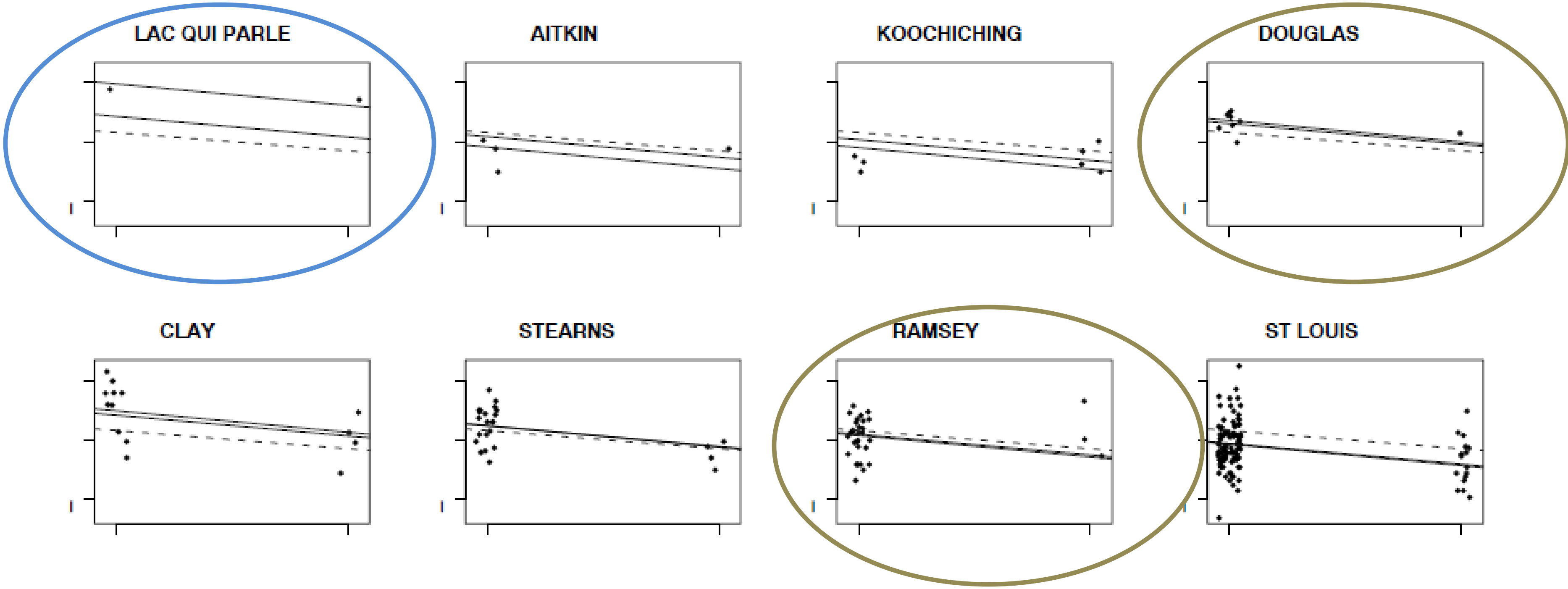
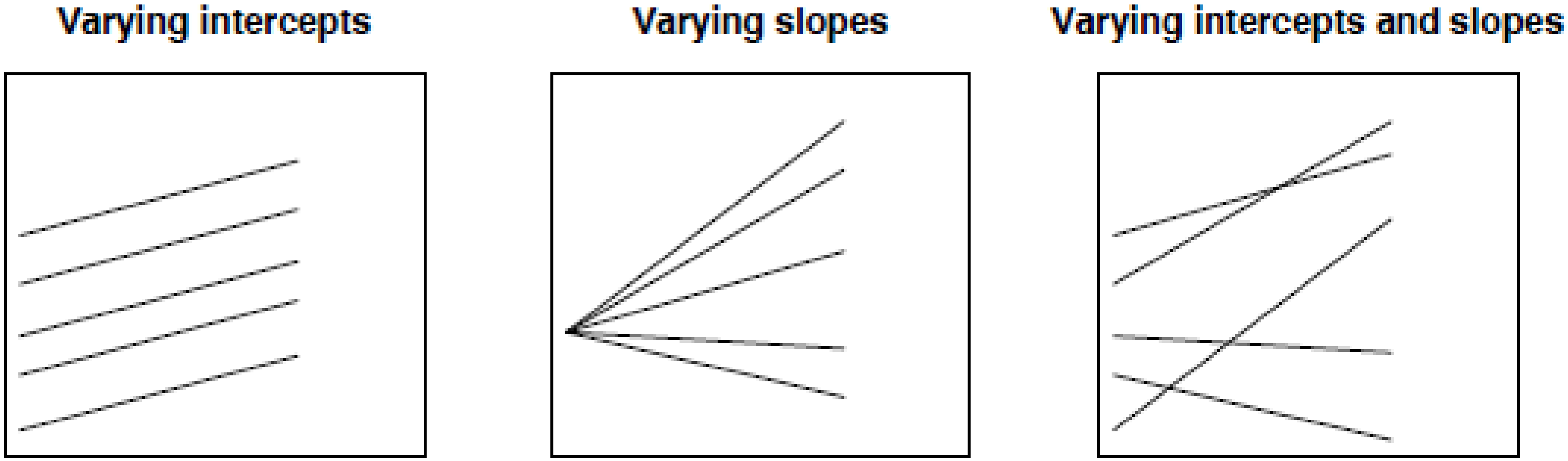


Figure 12.4 *Multilevel (partial pooling) regression lines $y = \alpha_j + \beta x$ fit to radon data from Minnesota, displayed for eight counties. Light-colored dashed and solid lines show the complete-pooling and no-pooling estimates, respectively, from Figure 12.3a.*

Gelman & Hill Chapter 12 p257

Illustration of random effects on:

- The intercept (example: Gelman text p. 259)
- The slope (example: Gelman text p. 284)
- Intercept and slope (example: Gelman text p. 279)
  - can be thought of as an interaction between the fixed and random variables



Varying intercepts

Varying slopes

Varying intercepts and slopes

x-axis: fixed variable
y-axis: outcome variable
separate lines: random variable

Example based on success-slope Hypothesis 1

- In Assignment 2 (see *Assignment 2 answers* in Study Materials), we discovered a significant interaction between prior beliefs and success-slope in predicting the illusion of natural control, except that the assumptions for the associated ANOVA were not met. Here, we split prior beliefs into “low” and “high” categories and model them as a random variable (i.e., a variable whose range is not representative of the population the study is investigating). The outcome variable is the illusion of natural control (PostNaturalIoC), and the fixed predictor is success-slope.
- The descriptive statistics obtained through describeBy suggest that a varying intercept, varying slope model could be appropriate. Indeed, this is also suggested by our finding in Assignment 2 of an interaction between success-slope and prior beliefs. in the terminology of the `lmer` function in the `lme4` package, the notation for a varying intercept, varying slope model is:

```
Hyp1mm <- lmer(PostNaturalIoC ~ SeqCond + (1 + SeqCond|CatPreDBC_IOC)
```

- After `summary(Hyp1mm)` reveals a prefect correlation between slopes and intercepts, we switch to a simpler varying intercept model:

```
Hyp1mm <- lmer(PostNaturalIoC ~ SeqCond + (1|CatPreDBC_IOC)
```

- The effect of success-slope was found to be significant in this model, with the assumptions of normality and homogeneity of variance met.

Reporting the analysis

- References to articles using `lme4`: http://lme4.r-forge.r-project.org/bib/lme4bib.html
- Description of category boundaries for the newly-created prior belief categories; obtainable through:

```
describeBy(x = SS$PreDBC_IOC, group = SS$CatPreDBC_IOC)
```

- Table 1: Table of descriptive statistics based on `describeBy`
- The fitting algorithm (ML, REML, FML) needs to be mentioned. The default (used here) is REML.
- Table 2: Table showing coefficients and random effects as on next slide.
- Text: A multilevel varying intercept model was fitted using REML in the `lme4` package in R Version 3.1.0. Prior belief category (low vs. high) was the random variable, while success-slope was a fixed predictor. Model coefficients are shown in Table 2. A Wald Chi-square test showed the effect of success-slope to be significant (Wald χ^2 (3) = 9.79, *p*=.02).

Table 2. Estimated fixed and random effects in the multilevel model. The Descending success-slope condition serves as the reference category in fixed effects.

	Coefficient: estimated group difference	Wald CI (95%)
Intercept	1.47 (low prior illusion); 3.26 (high prior illusion)	
U-shaped minus Descending	0.49	-0.07-1.04
Ascending minus Descending	0.90	0.34-1.47
Flat minus Descending	0.43	-0.13-1.0
Random effect variance estimate: 0.89		

Reading

Navarro, D. J. (2014). *Learning statistics with R: A tutorial for psychology students and other beginners*. Available online: http://health.adelaide.edu.au/psychology/ccs/teaching/lsr/. Chapter 12.

Baguley, T. *Serious Stats: A Guide to Advanced Statistics for the Behavioural Sciences.* Palgrave Macmillan: UK. Chapter 17 “Modelling discrete outcomes” (pdf in Study Materials/Readings).

Gelman, A., & Hill, J. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press: New York. Chapters 11-13 (pdf in Study Materials/Readings).