

WOMBBaGS: West Virginia Old Men, Bicycling, Beer and Glucose Study

SDC

March 2015

<http://bendixcarstensen.com/WOMBBaGS>

Version 0

Compiled Sunday 22nd March, 2015, 12:00
from: /home/bendix/sdc/proj/WV/WV

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk
<http://BendixCarstensen.com>

Contents

1	Protocol	1
1.1	Background	1
1.2	Purpose	1
1.3	Setting	1
1.3.1	Shortcomings	1
1.4	Data collection	1
1.4.1	Data structure	2
1.4.2	Practical data entry	2
2	Analysis outline	4
2.1	Statistical model	4
2.2	Practical analytics	4

Chapter 1

Protocol

1.1 Background

The interaction of age, bicycling and beer consumption in mountainous environment and their effect on glucose metabolism has long been an area with profound lack of insight.

1.2 Purpose

The study aims at elucidating the complex relationships between primarily beer and bicycling as determinants and glucose levels as outcome.

1.3 Setting

The study is conducted in the West Virginia Appalachians in June 2015 in a group of grown-up men.

1.3.1 Shortcomings

Initially we aimed at investigating the effect of sex as well, but we were not able to obtain informed consent from any eligible women.

1.4 Data collection

The primary outcome measure is blood glucose as measured in capillary blood from a finger-prick. The secondary outcome is weight as measured on a simple scale (the same for all measurements).

The explanatory variables are the fixed covariates:

- height
- date of birth

and the time-varying covariates recorded at different times:

- time of measurement

- preceding activity: sleep / breakfast / lunch / dinner / cycling / cooking / beer

Additionally we will attempt to collect times of (start of) ingestion of beer / wine.

1.4.1 Data structure

We shall collect a dataset with the following variables:

- participant ID
- time
- glucose
- weight
- preceding activity
- distance ridden (if relevant)
- beers ingested (if relevant)

1.4.2 Practical data entry

We create a skeleton data frame

```
> WV <- data.frame( id = "BxC",
+                   date = 10,
+                   time = "20:14",
+                   glucose = 0,
+                   weight = 0,
+                   prec = "sleep",
+                   dist = 0,
+                   beer = 0 )
> str( WV )

'data.frame':      1 obs. of  8 variables:
 $ id      : Factor w/ 1 level "BxC": 1
 $ date    : num 10
 $ time    : Factor w/ 1 level "20:14": 1
 $ glucose : num 0
 $ weight  : num 0
 $ prec    : Factor w/ 1 level "sleep": 1
 $ dist    : num 0
 $ beer    : num 0

> WV

   id date  time glucose weight  prec dist beer
1 BxC  10 20:14         0        0 sleep    0    0
```

Data will be entered using the `edit` function:

```
> d0 <- WV
> d1 <- edit( d0 )
```

1.4.2.1 Post-processing data

Eventually we will want to carry last date / time forward, using `na.locf` from the `zoo` package

```
> library(zoo)
> ( DF <- data.frame( A = rep("a",4),
+                    B = c(1:2,NA,4),
+                    C = 1:4,
+                    D = c(1,NA,NA,NA) ) )
```

	A	B	C	D
1	a	1	1	1
2	a	2	2	NA
3	a	NA	3	NA
4	a	4	4	NA

```
> na.locf( DF )
```

	A	B	C	D
1	a	1	1	1
2	a	2	2	1
3	a	2	3	1
4	a	4	4	1

Chapter 2

Analysis outline

The basic structure of data is a set of time-series (one for each person) of glucose measurements. What we aim to learn is how the recorded time-series of covariates influence the measured time series of glucose.

2.1 Statistical model

The basic analytic model will model the glucose level by a fixed mean defined from the covariates and a set of random effects for each covariate. In this way we obtain the mean effect of covariates as well as the between persons-variation of these.

Specifically the model for the measurement of glucose in individual i at time t looks like this:

$$y_{it} = \mu + \beta_{\text{beer}} + \gamma_{\text{cycle}} + a_i + b_i \text{I}\{\text{beer}\}_{it} + c_i \text{I}\{\text{cycle}\}_{it} + e_{it}$$

where we have defined beer and cycle for each measurement point suitably — these need not necessarily be the indicators of whether the activity was associated with the the time in question, they might be the cumulative exposure during the last 10 hours (for cycling) or 3 hours (for beers).

The fixed effects β and γ shows the overall effects of the activities, whereas the variation in b_i and c_i give the between-person variation in these, whereas the posterior estimates $\hat{\beta} + \hat{b}_i$ $\hat{\gamma} + \hat{c}_i$ will be the prediction of the effects in individual i .

2.2 Practical analytics

The model as stated above is a standard random effects model and could be fitted using `lme` from the the standard `nlme` package or simple yet `lmer` from the `lme4` package.

However, it might not be feasible to stick to the classical linear structure of the model in which case it might be preferable to use Bayesian estimation (that is MCMC-simulation), which will provide not only estimates of the parameters, but a simulated sample from the joint posterior distribution of all parameters enabling much more detailed inference. The downside is that programming the model is a bit more cumbersome.