# What are Mixed Models?
# Why are they Used?
# An Introduction

Ed Stanek

# Example and Context

- Seasons Study
  - Purpose:  To evaluate seasonal changes in cholesterol
  - Subjects: Volunteer's age 20-75 from HMO
  - Design:  Every 3 months, measure SC, diet, etc
- Quarterly Diet Assessment
  - Three 24-hour phone interviews by dietician
  - data coded and run through nutrient data bank
  - output- kcal/day, protein, vitamins, **alcohol**, etc.

2

# Questions, Problems

- Design an Intervention Study on Alcohol
- Analyze the Study Results
- Predict Alcohol Consumption for a Subject

- First, we describe Alcohol Consumption using Season's Study Data

# Describing the 24-Hr Alcohol Data

- 1 Beer = 13.2 grams of Alcohol

| Quantile | Estimate |
|---|---|
| 100% Max | 35.4 |
| 99% | 6.8 |
| 95% | 3.2 |
| 90% | 2.1 |
| 75% Q3 | 0.6 |
| 50% Median | 0 |
| 25% Q1 | 0 |
| 10% | 0 |
| 5% | 0 |
| 1% | 0 |
| 0% Min | 0 |

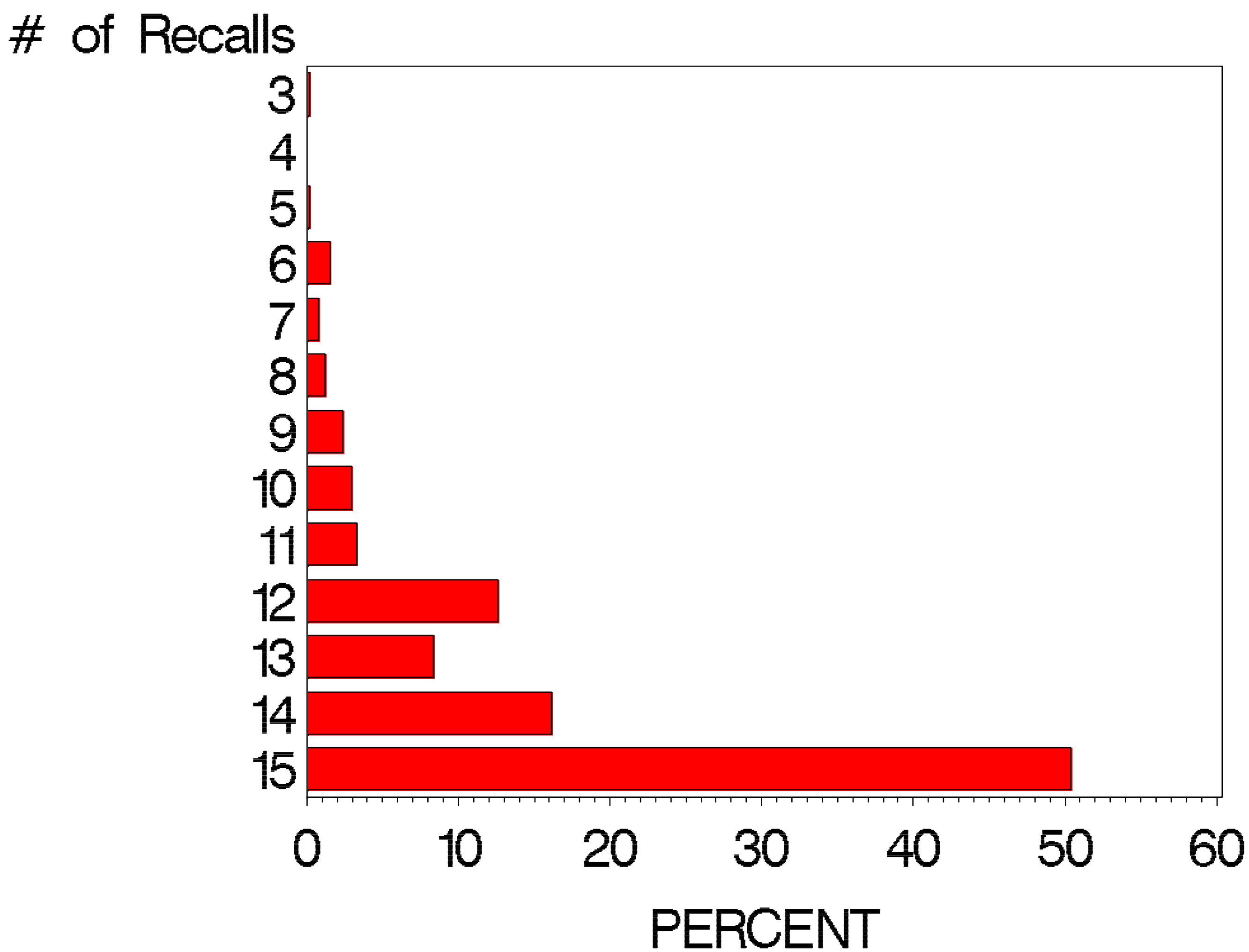
- (Based on over 7000 24-hour diet recalls)

- Table 2. List of Reports of More than 10 Beers/day

| ID | QUARTER | day | BEER |
|---|---|---|---|
| 278 | 3 | Sunday | 11.8 |
| 319 | 1 | Monday | 11.5 |
| 319 | 1 | Tuesday | 10.3 |
| 319 | 5 | Thursday | 10.3 |
| 482 | 3 | Sunday | 10.8 |
| 505 | 3 | Saturday | 10.5 |
| 505 | 3 | Sunday | 10.5 |
| 505 | 3 | Tuesday | 10.5 |
| 579 | 5 | Wednesday | 11.9 |
| 585 | 3 | Wednesday | 35.4 |
| 597 | 5 | Friday | 10.9 |
| 631 | 5 | Saturday | 10.7 |
| 631 | 5 | Wednesday | 12.2 |
| 653 | 4 | Wednesday | 11.5 |
| 773 | 4 | Thursday | 11.3 |
| 854 | 3 | Wednesday | 19.1 |
| 895 | 2 | Friday | 11.8 |
| 895 | 3 | Friday | 12.7 |
| 989 | 3 | Saturday | 17.3 |
| 989 | 3 | Wednesday | 14.9 |
| 989 | 3 | Wednesday | 10.6 |
| 989 | 4 | Tuesday | 12.6 |

- (Source: sne02p12.sas)

Figure 1a. Dist. of Recall Interviews




6

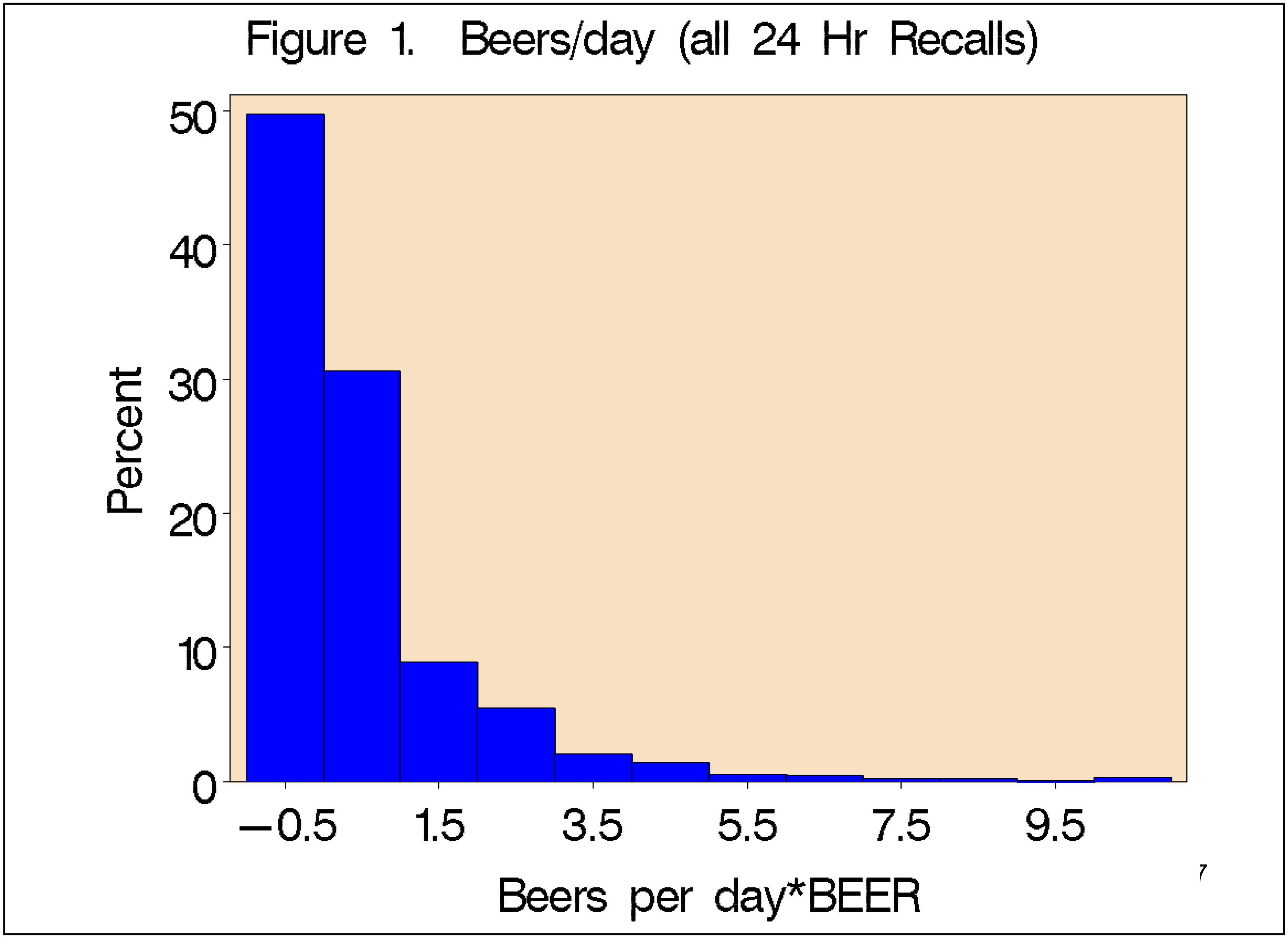
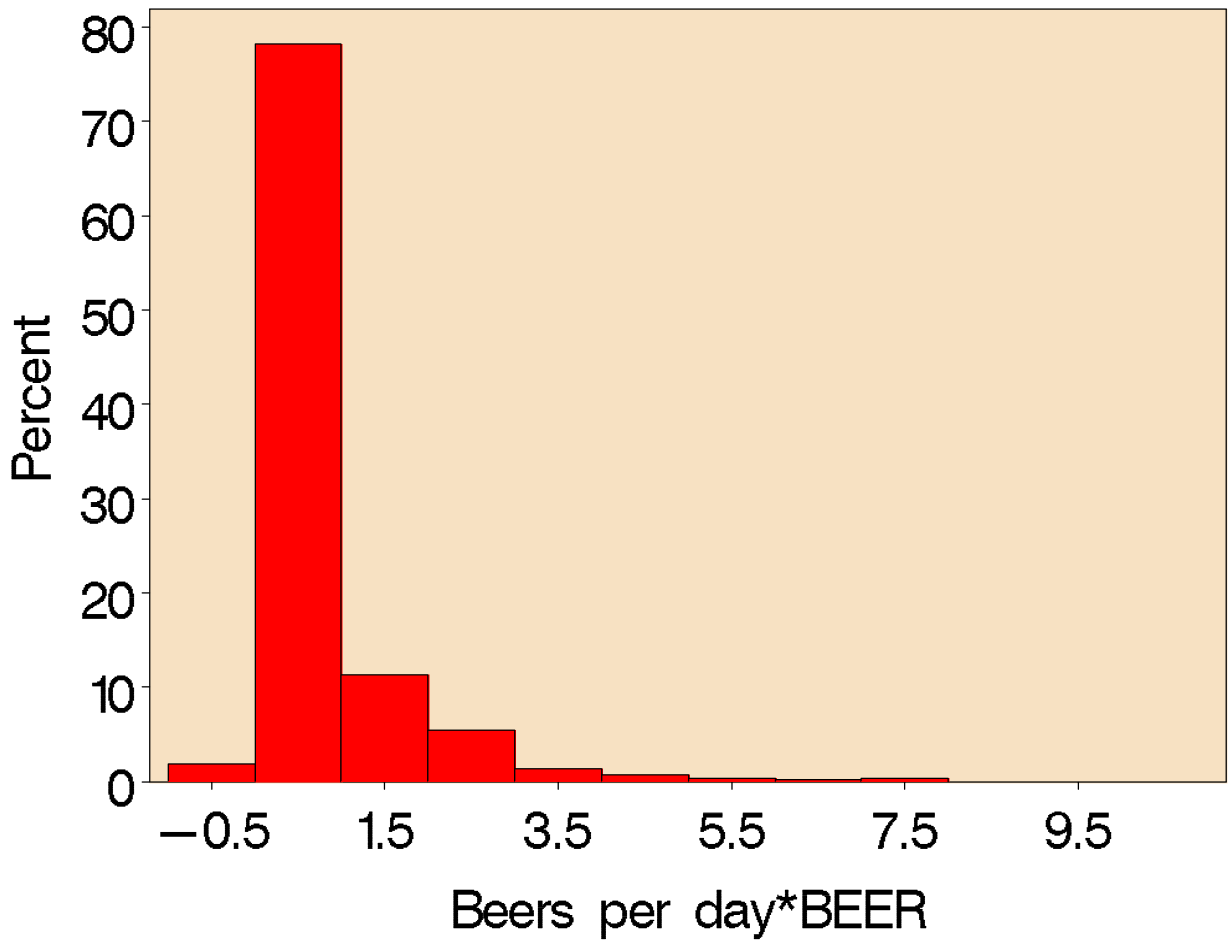


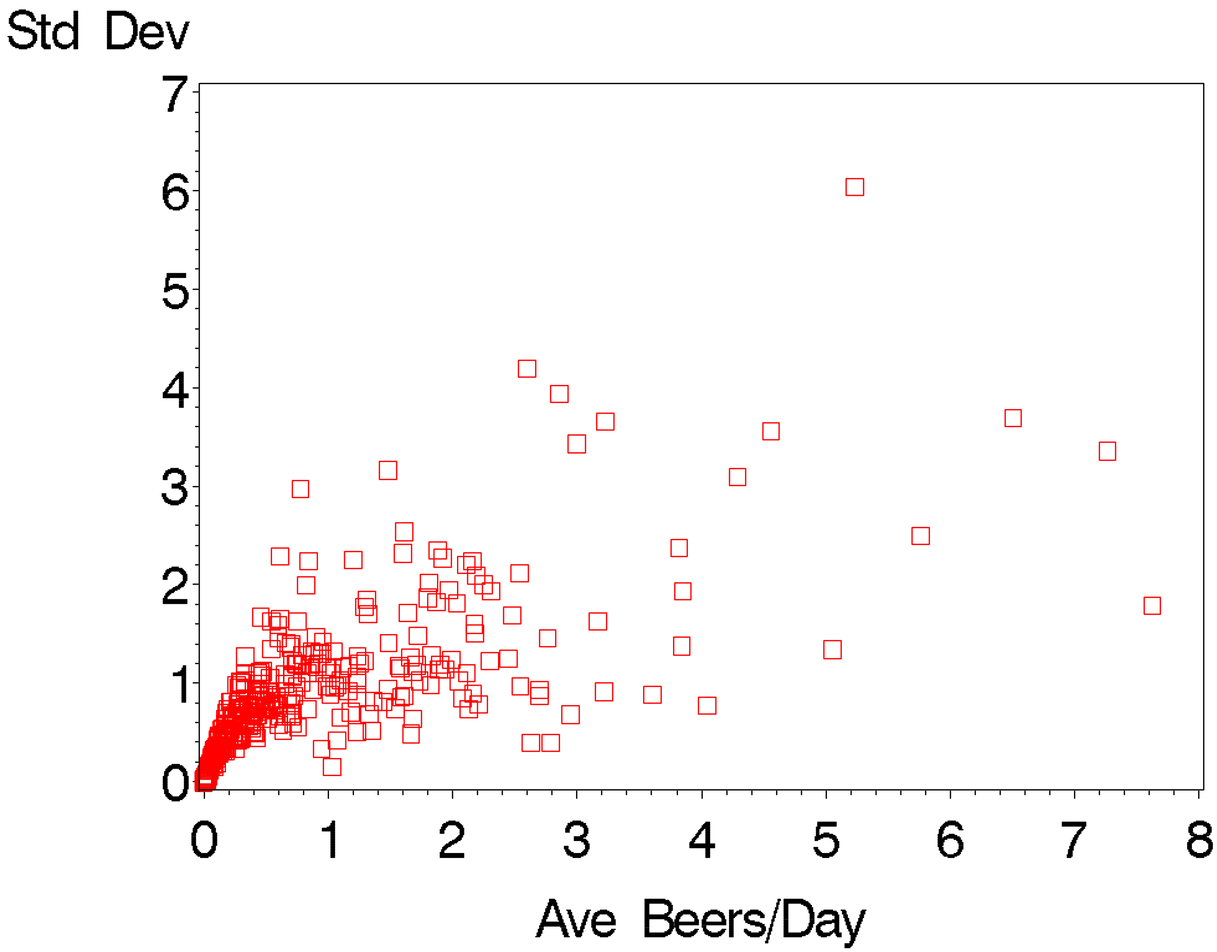
Figure 1. Beers/day (all 24 Hr Recalls)
Percent
0
10
20
30
40
50
−0.5
1.5
3.5
5.5
7.5
9.5
Beers per day*BEER

Figure 2. Beers/day (per Subject)




8

Figure 3. Scatter Plot of SD vs Mean Beers/day (per subj)




9

# A Descriptive Model for a subject-day

Average Beers/Day $$\mu_s = \frac{\sum_{t=1}^{30} \mu_{st}}{30}$$

or $$\mu_{st} = \mu_s + \varepsilon_{st}$$

For a Population: $$\mu = \frac{\sum_{s=1}^{N} \mu_s}{N}$$

or

$$\mu_{st} = \mu + \beta_s + \varepsilon_{st}$$

# A Descriptive Model for Selected subject on a selected day

Beers on j-th selected day for i-th selected subject:

$$Y_{ij} = \mu + B_i + E_{ij}$$

This is a “mixed” model.

Average Beers/day for i-th selected subject:

$$M_i = \mu + B_i$$

This is a Random Effect



11

In the Population:

$$E_S\left(M_i\right) = \mu$$

$$\text{var}_S\left(M_i\right) = \left(\frac{N-1}{N}\right)\sigma^2 = \frac{\sum_{s=1}^{N}\left(\mu_s - \mu\right)^2}{N}$$

Also:

$$E_T\left(E_{ij} \mid i = s\right) = 0$$

$$\text{var}_T\left(E_{ij} \mid i = s\right) = \left(\frac{M-1}{M}\right)\sigma_{se}^2 = \frac{\sum_{t=1}^{M}\left(\mu_{st} - \mu_s\right)^2}{M}$$

12

To summarize, the mixed model is given by:

Beers on jth selected day for ith selected subject

$$Y_{ij} = \mu + B_i + E_{ij}$$

where

$$E_{ST}\left(Y_{ij}\right) = \mu$$

$$\text{var}_{ST}\left(Y_{ij}\right) = \sigma^{2} + \sigma_{e}^{2}$$

Note that

$$\text{cov}_{ST}\left(Y_{ij}, Y_{ij*}\right) = \sigma^{2}$$



13

## Use the Mixed Model to Summarize Beers/day

```
PROC MIXED;
  CLASS id;
  MODEL beer= /solution;
  RANDOM id;
  RUN;
```

**Results:**

Covariance Parameter Estimates

| Cov Parm | Estimate |
|---|---|
| ID | 0.97 |
| Residual | 0.94 |

Solution for Fixed Effects

| Effect | Estimate | Standard Error |
|---|---|---|
| Intercept | 0.61 | 0.045 |

- (Source: sne02p15.sas)

# Gender Specific Results

| Beers Per Day | | Estimate | | Variance | |
|---|---|---|---|---|---|
| | n | Mean | Std Err | Subject | Day |
| Male | 262 | 0.87 | 0.079 | 1.52 | 1.41 |
| Female | 252 | 0.33 | 0.033 | 0.25 | 0.47 |

15

## Design a Study to Reduce Alcohol Consumption in Men by 25%

$$H_0: \mu = 0.87$$

$$H_A: \mu \le 0.65$$

Assume:

- completely randomized study
- 80% power to detect alternative hypothesis

$$n = \frac{\left(z_{1-\alpha} + z_{1-\beta}\right)^2}{\left(\mu_0 - \mu_1\right)^2}\sigma^{*2} = \frac{\left(1.645 + 0.84\right)^2}{\left(0.32\right)^2}\sigma^{*2}$$

or $n = 60.3\ \sigma^{*2}$ where $\sigma^{*2} = \sigma^2 + \left(1 - f_m\right)\frac{\sigma_e^2}{m}$ and $f_m = \frac{m}{M}$

16

| Subj. Var | Day Var | M | m | n | nm |
|---|---|---|---|---|---|
| 1.52 | 1.41 | 30 | 1 | 174 | 174 |
| 1.52 | 1.41 | 30 | 2 | 131 | 263 |
| 1.52 | 1.41 | 30 | 3 | 117 | 351 |
| 1.52 | 1.41 | 30 | 4 | 110 | 440 |
| 1.52 | 1.41 | 30 | 5 | 106 | 529 |
| 1.52 | 1.41 | 30 | 6 | 103 | 618 |
| 1.52 | 1.41 | 30 | 7 | 101 | 707 |



17

With m=3 Recalls, for the i-th subject, we'll observe: $Y_{i1} \quad Y_{i2} \quad Y_{i3}$

where $$E_{ST}\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix} = \begin{pmatrix} \mu \\ \mu \\ \mu \end{pmatrix}$$

and $$\mathrm{var}_{ST}\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix} = \begin{pmatrix} \sigma^2+\sigma_e^2 & \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2+\sigma_e^2 & \sigma^2 \\ \sigma^2 & \sigma^2 & \sigma^2+\sigma_e^2 \end{pmatrix}$$

18

Compound Symmetry:

$$\text{var}_{ST}\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix} = \left(\sigma^2 + \sigma_e^2\right)\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

Intra-class Correlation:

$$\rho = \frac{\sigma_e^2}{\sigma^2 + \sigma_e^2}$$

$$\rho = \frac{1.41}{1.52 + 1.41} = 0.48$$

Design effect:

$$DEFF = \left(1 + \left(m - 1\right)\rho\right)$$

$$DEFF = \left(1 + \left(3 - 1\right)0.48\right) = 1.96$$

19

# Analyze Study Results

Beers on jth selected day for ith selected subject in Treatment Group k

$$Y_{ijk} = \left(\mu + \alpha_k\right) + B_i + E_{ij}$$

where $$\alpha_k = \mu_k - \mu$$

Use Proc Mixed.

# Predicting a Random Effect

A model for the Random Effect:

$$M_i = \mu + B_i$$

How would you predict the mean for a realized selection?

$$\overline{Y}_i = \frac{\sum_{j=1}^{m} Y_{ij}}{m} \cdots\cdots\cdots \frac{1.6 + 3.2 + 3.2}{3} = 2.7$$

For ID 294 in Seasons Study

BLUP： $\hat{P}_i = \bar{\bar{Y}} + k\left(\bar{Y}_i - \bar{\bar{Y}}\right)$

Shrinkage factor: $k = \dfrac{\sigma^2}{\sigma^2 + \dfrac{\sigma_e^2}{m}}$

Regression to the mean $k = 0.76$

For ID=294, rather than $\bar{Y}_i = 2.7$ Beers/day

$$\hat{P}_i = \bar{\bar{Y}} + k\left(\bar{Y}_i - \bar{\bar{Y}}\right) = 0.87 + 0.76(2.7 - 0.87)$$
$$= 2.3$$

# Promise of Prediction

Predict response for a realized Subject

- a clinic
- a class-room
- a town, or region
- a family

Not simply predicting response for the 'average'

# Link with Bayesian Thinking

- Assume a prior distribution for Random Effects.
- Express the joint distribution of Data and Random effects
- The posterior distribution is found by conditioning on the data.
- The mean of the posterior distribution is the Bayes predictor.



24

# New Insights

To Predict 30-day average Beer Drinking

- Example: For ID 294 in Seasons Study

$$\overline{Y}_i = \frac{\sum_{j=1}^{m} Y_{ij}}{m} \cdots\cdots\cdots \frac{1.6 + 3.2 + 3.2}{3} = 2.7 \qquad BLUP = \hat{P}_i = 2.3$$

- No need to predict these 3 days. Just predict the remaining 27 days in the month!
- New Predictor: $\left(\frac{3}{30}\right)\overline{Y}_i + \left(\frac{27}{30}\right)\hat{P}_i = 2.34$

25

# Basic Problem

Realized Random Effect = Latent Value for Realized Unit

$$M_i = \mu + B_i$$

PROBLEM: Units are not 'identifiable' in typical Mixed Model

Example: pick a coin (dime, nickel, penny) , flip once, and observed Head (or tail). Can this be used to predict Prob(Dime=Head)?

26

# Summary and Conclusions

- Mixed Models are Fundamental
- For Random Effects
  - There are Problems with the theory.
  - Problems with interpretation
  - These problems are not yet solved!!
- Stay Tuned!