75 Year Anniversary
NC State University Department of Statistics
—Biostatistics—

Geert Molenberghs
geert.molenberghs@uhasselt.be & geert.molenberghs@kuleuven.be

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)
UHasselt & KU Leuven, Belgium

www.ibiostat.be
Theme I

Some Historical Reflections
Sampling

- Example
- Population → sample
- Variability
Average age in a population

Average age in 21 units

Random Statistics
\[ Y_i \sim N(\mu, \sigma^2) \]
$$Y_i \sim N(\mu, \sigma^2)$$

<table>
<thead>
<tr>
<th>Bernie Sanders</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\[ Y_i \sim N(\mu, \sigma^2) \]

<table>
<thead>
<tr>
<th>Bernie Sanders</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Simon</td>
<td></td>
</tr>
<tr>
<td>Art Garfunkel</td>
<td></td>
</tr>
<tr>
<td>Neil Diamond</td>
<td></td>
</tr>
</tbody>
</table>
\[ Y_i \sim N(\mu, \sigma^2) \]

<table>
<thead>
<tr>
<th>Bernie Sanders</th>
<th>Martha Reeves</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Simon</td>
<td>David Crosby</td>
</tr>
<tr>
<td>Art Garfunkel</td>
<td></td>
</tr>
<tr>
<td>Neil Diamond</td>
<td></td>
</tr>
<tr>
<td>Bob Dylan</td>
<td></td>
</tr>
<tr>
<td>Paul Anka</td>
<td></td>
</tr>
<tr>
<td>Chubby Checker</td>
<td></td>
</tr>
</tbody>
</table>
\[ Y_i \sim N(\mu, \sigma^2) \]

<table>
<thead>
<tr>
<th>Bernie Sanders</th>
<th>Martha Reeves</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Simon</td>
<td>David Crosby</td>
</tr>
<tr>
<td>Art Garfunkel</td>
<td>Joan Baez</td>
</tr>
<tr>
<td>Neil Diamond</td>
<td>David Crosby</td>
</tr>
<tr>
<td>Bob Dylan</td>
<td>Placido Domingo</td>
</tr>
<tr>
<td>Paul Anka</td>
<td></td>
</tr>
<tr>
<td>Chubby Checker</td>
<td></td>
</tr>
</tbody>
</table>
\[ Y_i \sim N(\mu, \sigma^2) \]

<table>
<thead>
<tr>
<th>Bernie Sanders</th>
<th>Martha Reeves</th>
<th>Fred West</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Simon</td>
<td>David Crosby</td>
<td>Charles Whitman</td>
</tr>
<tr>
<td>Art Garfunkel</td>
<td>Joan Baez</td>
<td>Jesse Jackson</td>
</tr>
<tr>
<td>Neil Diamond</td>
<td>David Crosby</td>
<td>George Clinton</td>
</tr>
<tr>
<td>Bob Dylan</td>
<td>Placido Domingo</td>
<td>Desmond Dekker</td>
</tr>
<tr>
<td>Paul Anka</td>
<td>Kim Jong Il</td>
<td>Vivienne Westwood</td>
</tr>
<tr>
<td>Chubby Checker</td>
<td>Dick Cheney</td>
<td></td>
</tr>
</tbody>
</table>

75 Year Anniversary NC State University Department of Statistics
$Y_i \sim N(75,0)$

<table>
<thead>
<tr>
<th>Bernie Sanders</th>
<th>Martha Reeves</th>
<th>Fred West</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paul Simon</td>
<td>David Crosby</td>
<td>Charles Whitman</td>
</tr>
<tr>
<td>Art Garfunkel</td>
<td>Joan Baez</td>
<td>Jesse Jackson</td>
</tr>
<tr>
<td>Neil Diamond</td>
<td>David Crosby</td>
<td>George Clinton</td>
</tr>
<tr>
<td>Bob Dylan</td>
<td>Placido Domingo</td>
<td>Desmond Dekker</td>
</tr>
<tr>
<td>Paul Anka</td>
<td>Kim Jong Il</td>
<td>Vivienne Westwood</td>
</tr>
<tr>
<td>Chubby Checker</td>
<td>Dick Cheney</td>
<td>NCSU Stat Dept!!</td>
</tr>
</tbody>
</table>
A Brief Time Line for Our Field

- 1857 Karl Pearson
- 1890 Sir R.A. Fisher
- 1894 Jerzy Neymann
- 1900 Gertrude Cox
- 1917 David Finney
- 1924 David Cox
- 1941 NC State Statistics Department
- 1947 International Biometric Society
- 1976 SAS Institute Inc.
A Brief Time Line for Our Field

- Experimental design & linear models
- Epidemiology, observational studies, & GLM
- Statistical genetics
- Bioinformatics
- Big data, data science, & data analytics
A Virtuous Triangle

Research & Education \uparrow \downarrow \rightarrow Research & Consulting

Education \leftarrow Consulting

Education & Consulting

Research & Education & Consulting
Biostatistics in North America

- Strong focus on medical statistics
- Do not forget the agriculture/USDA roots

- **First generation**
  - U. Buffalo, NY

- **Second generation**
  - Harvard U.
  - U. Washington
  - U. Waterloo

- **Third generation**
  - U. Michigan
  - Johns Hopkins U.
  - NC State
Education

- Here is to the very popular and successful statistics programs and biostatistics concentration at NC State!

- Formal programs:
  - Undergraduate level
  - Master level
  - PhD level

- On site & distance

- Summer schools and short courses

- For statisticians and our partner fields
Collaboration and Institutes

- Gertrude Cox’s determination to build institute and department!
- Regional densification:

\[
\begin{align*}
\text{NC State} \\
\text{RTI} & \leftrightarrow \text{UNC} & \leftrightarrow \text{Duke} \\
\text{SAS} & \leftrightarrow \text{NISS}
\end{align*}
\]
An International Society Devoted to the Mathematical and Statistical Aspects of Biology

The Biometric Society is an international society for the advancement of quantitative biological science through the development of quantitative theories and the application, development and dissemination of effective mathematical and statistical techniques. To this end the society welcomes to membership biologists, mathematicians, statisticians and others interested in applying similar techniques.

Gertrude Cox: preferred primary training in substantive field!
<table>
<thead>
<tr>
<th>Volume</th>
<th>Name</th>
<th>Nationality</th>
<th>Years</th>
<th>Name</th>
<th>Nationality</th>
</tr>
</thead>
<tbody>
<tr>
<td>47-49</td>
<td>R.A. Fisher</td>
<td>British &amp; Irish</td>
<td>84-85</td>
<td>Pierre Dagnelie</td>
<td>Belgian</td>
</tr>
<tr>
<td>50-51</td>
<td>A. Linder</td>
<td>at large</td>
<td>86-87</td>
<td>Geoffrey H. Freeman</td>
<td>British &amp; Irish</td>
</tr>
<tr>
<td>52-53</td>
<td>G. Darmois</td>
<td>French</td>
<td>88-89</td>
<td>Jonas H. Ellenberg</td>
<td>ENAR</td>
</tr>
<tr>
<td>54-55</td>
<td>W.G. Cochran</td>
<td>ENAR</td>
<td>90-91</td>
<td>Richard Tomassone</td>
<td>French</td>
</tr>
<tr>
<td>56-57</td>
<td>E.A. Cornish</td>
<td>Australasian</td>
<td>92-93</td>
<td>Niels Keiding</td>
<td>Nordic</td>
</tr>
<tr>
<td>58-59</td>
<td>C.H. Goulden</td>
<td>ENAR</td>
<td>94-95</td>
<td>Lynne Billard</td>
<td>ENAR</td>
</tr>
<tr>
<td>60-61</td>
<td>L. Martin</td>
<td>Belgian</td>
<td>96-97</td>
<td>Byron J.T. Morgan</td>
<td>British &amp; Irish</td>
</tr>
<tr>
<td>62-63</td>
<td>Chester I. Bliss</td>
<td>ENAR</td>
<td>98-99</td>
<td>Susan Wilson</td>
<td>Australasian</td>
</tr>
<tr>
<td>64-65</td>
<td>David J. Finney</td>
<td>British &amp; Irish</td>
<td>00-01</td>
<td>Nanny Wermuth</td>
<td>German</td>
</tr>
<tr>
<td>66-67</td>
<td>L.L. Cavalli-Sforza</td>
<td>Italian</td>
<td>02-03</td>
<td>Norman E. Breslow</td>
<td>WNAR</td>
</tr>
<tr>
<td>68-69</td>
<td>Gertrude M. Cox</td>
<td>ENAR</td>
<td>04-05</td>
<td>Geert Molenberghs</td>
<td>Belgian</td>
</tr>
<tr>
<td>70-71</td>
<td>Berthold Schneider</td>
<td>German</td>
<td>06-07</td>
<td>Thomas A. Louis</td>
<td>ENAR</td>
</tr>
<tr>
<td>72-73</td>
<td>Peter Armitage</td>
<td>British &amp; Irish</td>
<td>08-09</td>
<td>Andrew Mead</td>
<td>British &amp; Irish</td>
</tr>
<tr>
<td>74-75</td>
<td>C.R. Rao</td>
<td>ENAR</td>
<td>10-11</td>
<td>Kaye Basford</td>
<td>Australasian</td>
</tr>
<tr>
<td>76-77</td>
<td>Henry L. Le Roy</td>
<td>Austro-Swiss</td>
<td>12-13</td>
<td>Clarice Demétrio</td>
<td>Brazilian</td>
</tr>
<tr>
<td>78-79</td>
<td>John A. Nelder</td>
<td>British &amp; Irish</td>
<td>14-15</td>
<td>John Hinde</td>
<td>British &amp; Irish</td>
</tr>
<tr>
<td>80-81</td>
<td>Richard M. Cormack</td>
<td>British &amp; Irish</td>
<td>16-17</td>
<td>Elizabeth Thompson</td>
<td>WNAR</td>
</tr>
<tr>
<td>82-83</td>
<td>Herbert A. David</td>
<td>ENAR</td>
<td>18-19</td>
<td>Louise M. Ryan</td>
<td>Australasian</td>
</tr>
</tbody>
</table>
BIOMETRICAL CLINIC ON
ENTOMOLOGICAL PROBLEMS

New York City December 13, 1948

Proceedings of a joint meeting of the
American Association of Economic Entomologists
and the
Biometric Society, Eastern North American Region

Edited by G. J. Bliss
Obtainable from the
Biometric Society,
C. I. Bliss, Secretary
Box 1106, New Haven, Conn.

Price to members of sponsoring organizations $.50,
to all others $.75.

February 1950

Foreword

The advances in biometry have developed through
the close cooperation of biologists confronted with
problems and of statisticians who develop methods for
solving them. Both have gained from this collabora-
tion. One medium for maintaining contact is the
"biometrical clinic", in which questions are asked by
the biologist and answered informally by the statistic-
ian. The meeting recorded here followed this pattern.
Even when the answers can be found in textbooks or in
scientific journals, the method which is most relevant
to a specific problem may not be apparent to the
uninitiated. If a question cannot be answered, its
asking may guide the statistician into a new and
interesting field of study. Hence the session is of
interest to both the biologist and the statistician.

The proceedings of the present "clinic" were
recorded electronically and then transcribed and
edited. In view of the demand, the Council of the
Biometric Society has agreed to their publication in
the present form. Although a few references have been
inserted, there has been no attempt to document each
statement as in a scientific paper. Each participant
on the panel has checked his own remarks and the
proceedings have been read critically by several
biometricians not on the panel. Responsibility for
the final form, however, rests with the editor.
The advances in biometry have developed through the close cooperation of biologists confronted with problems and of statisticians who develop methods for solving them. Both have gained from this collaboration. One medium for maintaining contact is the "biometrical clinic", in which questions are asked by the biologist and answered informally by the statistician. The meeting recorded here followed this pattern. Even when the answers can be found in textbooks or in scientific journals, the method which is most relevant to a specific problem may not be apparent to the uninitiated. If a question cannot be answered, its asking may guide the statistician into a new and interesting field of study. Hence the session is of interest to both the biologist and the statistician.
Acknowledgements

The Biometric Society is indebted to Professor C. E. Pals and other members of the program committee of the American Association of Economic Entomologists for their cooperation in scheduling this joint meeting, to the SoundScriber Corporation for the loan of microphones and other equipment for recording the proceedings of the meeting, and to the Connecticut Agricultural Experiment Station for its cooperation, especially through two members of its staff, Miss Nancy Woodruff, who operated the SoundScriber during the session, and Mrs. Anna Franchini, who transcribed the original records and prepared the copy.

QUESTIONS

1. What is the value of replication and randomization if later the data are not analyzed statistically?
2. Why should entomologists not use systematic designs?
3. How best to set up large airplane plots for statistical interpretation?
4. How many replicates are there in a randomized block experiment with six blocks?
6. The calculation and limitation of least significant differences from the statistics obtained in the analysis of variance.
7. In applying the analysis of variance to the results of experiments on insect control, what is the importance of the observed distribution of insects?
8. What units should be used in analyzing data on the number of rat ectoparasites on individual rats?
9. What is the value of pretreatment estimates of an insect population in relation to the analysis of covariance and Abbott's formula?
10. What is a suitable design and analysis for experiments on off-flavor in potatoes and other vegetables resulting from insecticidal treatments?
QUESTIONS

1. What is the value of replication and randomization if later the data are not analyzed statistically?
Members of the Panel

Frank Wilcoxon, Chairman, American Cyanamid Company, Stanford, Connecticut
C. I. Bliss, Connecticut Agricultural Experiment Station, New Haven, Connecticut
Walter C. Jacob, Long Island Vegetable Research Farm, Riverhead, New York
John W. Tukey, Princeton University, Princeton, New Jersey

Questioners and Participants in the Discussion

J. W. Apple, University of Wisconsin, Madison
L. A. Carruth, New York Agricultural Experiment Station, Geneva
C. R. Cutright, Ohio Experiment Station, Wooster
M. M. Darley, General Chemical Co., Long Island City
M. W. Farrar, University of New Hampshire, Durham
H. C. Fryer, Kansas Agricultural Experiment Station, Manhattan
Phil Garman, Connecticut Agricultural Experiment Station, New Haven
J. L. Horsfall, American Cyanamid Company, New York
L. B. Reed, Bureau of Entomology and Plant Quarantine, Washington, D. C.
Neely Turner, Connecticut Agricultural Experiment Station, New Haven
W. M. Ucholt, C. D. C., Savannah, Georgia
F. W. Wadley, 3125 W. Albermarle Street, Arlington
R. H. Wellman, Boyce Thompson Institute, Yonkers

BIOLOGICAL CLINIC ON ENTOMOLOGICAL PROBLEMS

Wilcoxon: This meeting is a joint session of the American Association of Economic Entomologists and the Biometric Society, Eastern North American Region. When the question of this meeting first came up there was a good deal of discussion as to the form which it should take. We finally decided to send out a circular letter to a number of entomologists, asking them to submit questions regarding the statistical treatment of certain types of experimental data, and in response to this questionnaire a number of replies were received. These questions were then forwarded to the members of our statistical panel, to select the particular ones which they wanted to consider. The program was made up from the questions selected.

I might take a minute to introduce the members of the statistical panel. On my extreme left is Dr. C. I. Bliss of the Connecticut Agricultural Experiment Station; next to him is Dr. A. E. Brandt of the U. S. Atomic Energy Commission; immediately on my right is Professor W. C. Jacob of the Long Island Vegetable Research Farm, while on my extreme right is Professor John Tukey of the Princeton mathematics faculty. I will read each question aloud and if the man who offered the question is in the room and wishes to amplify it, he will have a chance to do so. Then members of the panel will attempt to answer the question. Following that there may be comments, criticisms or suggestions from any one in the audience and whatever discussion seems necessary.

Question 1.

Wilcoxon: The first question on our list goes like this: For what it might be worth in improving experimental technique among entomologists who do not now and probably never will employ statistical analyses
Theme II

Incomplete Data: A Case in Point
Ignorable Likelihood/Bayesian Inference

\[ \text{MAR} : f(Y^0_i | X_i, \theta) f(r_i | X_i, Y^0_i, \psi) \]

Mechanism is MAR

\( \theta \) and \( \psi \) distinct

Interest in \( \theta \)

(Use observed information matrix)

\[ \Rightarrow \text{Lik./Bayes inference valid} \]

<table>
<thead>
<tr>
<th>Outcome type</th>
<th>Modeling strategy</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>Linear mixed model</td>
<td>SAS MIXED</td>
</tr>
<tr>
<td>Non-Gaussian</td>
<td>Gen./Non-linear mixed model</td>
<td>SAS GLIMMIX, NLMIXED</td>
</tr>
</tbody>
</table>
Rubin, 1976

- Ignorability: Rubin (Biometrika, 1976): 35 years ago!

- Little and Rubin (1976, 2002)

- Why did it take so long?
A Vicious Triangle

- **Academe**: The $R^2$ principle

- **Regulatory**: Rigid procedures $\longleftrightarrow$ scientific developments

- **Industry**: We cannot / do not want to apply new methods
Terminology & Confusion

• The Ministry of Disinformation:

  All directions

  Other directions

• MCAR, MAR, MNAR: “What do the terms mean?”

• MAR, random dropout, informative missingness, ignorable, censoring, . . .

• Dropout from the study, dropout from treatment, lost to follow up, . . .

• “Under MAR patients dropping out and patients not dropping out are similar.”
A Virtuous Triangle

Industry

Academe ↔ Regulatory

- FDA/Industry Workshops
- DIA/EMA Meetings
- The NAS Experience
The NAS Experience: A Wholesome Product

- FDA → NAS → the working group

- Composition

- Encompassing:
  - terminology/taxonomy/concepts
  - prevention
  - treatment
Taxonomy

- **Missingness pattern:** complete — monotone — non-monotone
- **Dropout pattern:** complete — dropout — intermittent
- **Model framework:** SEM — PMM — SPM
- **Missingness mechanism:** MCAR — MAR — MNAR
- **Ignorability:** ignorable — non-ignorable
- **Inference paradigm:** frequentist — likelihood — Bayes
# The NAS Panel

<table>
<thead>
<tr>
<th>Name</th>
<th>Specialty</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rod Little</td>
<td>biostat</td>
<td>U Michigan</td>
</tr>
<tr>
<td>Ralph D’Agostino</td>
<td>biostat</td>
<td>Boston U</td>
</tr>
<tr>
<td>Kay Dickerson</td>
<td>epi</td>
<td>Johns Hopkins</td>
</tr>
<tr>
<td>Scott Emerson</td>
<td>biostat</td>
<td>U Washington</td>
</tr>
<tr>
<td>John Farrar</td>
<td>epi</td>
<td>U Penn</td>
</tr>
<tr>
<td>Constantine Frangakis</td>
<td>biostat</td>
<td>Johns Hopkins</td>
</tr>
<tr>
<td>Joseph Hogan</td>
<td>biostat</td>
<td>Brown U</td>
</tr>
<tr>
<td>Geert Molenberghs</td>
<td>biostat</td>
<td>U Hasselt &amp; K.U.Leuven</td>
</tr>
<tr>
<td>Susan Murphy</td>
<td>stat</td>
<td>U Michigan</td>
</tr>
<tr>
<td>James Neaton</td>
<td>biostat</td>
<td>U Minnesota</td>
</tr>
<tr>
<td>Andrea Rotnitzky</td>
<td>stat</td>
<td>Buenos Aires &amp; Harvard</td>
</tr>
<tr>
<td>Dan Scharfstein</td>
<td>biostat</td>
<td>Johns Hopkins</td>
</tr>
<tr>
<td>Joseph Shih</td>
<td>biostat</td>
<td>New Jersey SPH</td>
</tr>
<tr>
<td>Jay Siegel</td>
<td>biostast</td>
<td>J&amp;J</td>
</tr>
<tr>
<td>Hal Stern</td>
<td>stat</td>
<td>UC at Irvine</td>
</tr>
</tbody>
</table>
Frameworks and Their Methods

\[ f(y_i, r_i | X_i, \theta, \psi) \]

Selection Models: \[ f(y_i | X_i, \theta) f(r_i | X_i, y_i^o, y_i^m, \psi) \]

\begin{align*}
\text{MCAR/simple} & \quad \rightarrow \quad \text{MAR} & \rightarrow \quad \text{MNAR} \\
\text{CC?} & \quad \text{direct likelihood!} & \quad \text{joint model!?} \\
\text{LOCF?} & \quad \text{direct Bayesian!} & \quad \text{sensitivity analysis?!} \\
\text{single imputation?} & \quad \text{multiple imputation (MI)!} \\
\text{:} & \quad \text{IPW } \supset \quad \text{W-GEE!} \\
\text{d.l. + IPW } = \text{ double robustness! (consensus)}
\end{align*}
Frameworks and Their Methods: Start

\[ f(y_i, r_i | X_i, \theta, \psi) \]

Selection Models: \[ f(y_i | X_i, \theta) f(r_i | X_i, y_i^0, y_i^m, \psi) \]

- **MCAR/simple** \( \rightarrow \) **MAR** \( \rightarrow \) **MNAR**

- Direct likelihood!
- Direct Bayesian!
- Multiple imputation (MI)!
- IPW \( \supset \) W-GEE!
- d.l. \( \oplus \) IPW = double robustness!
Frameworks and Their Methods: Next

\[ f(y_i, r_i | X_i, \theta, \psi) \]

Selection Models: \( f(y_i | X_i, \theta)f(r_i | X_i, y_i^0, y_i^m, \psi) \)

MCAR/simple \( \rightarrow \) MAR \( \rightarrow \) MNAR

- joint model?
- sensitivity analysis!
- PMM
- MI (MGK, J&J)
- local influence
- interval ignorance
- IPW based
Theme III

Research Integrity
Statistics in the Middle of a Crisis: Another Infamous Triangle

• Smoking and lung cancer:
  ▶ The tobacco industry versus the states of the United States
  ▶ Donald B. Rubin versus Scott L. Zeger

• Waste incineration (dioxin exhaust) and congenital malformation

• Cadmium contamination (soil) and lung cancer
The Toenail Data

- **Toenail Dermatophyte Onychomycosis**: Common toenail infection, difficult to treat, affecting more than 2% of population.

- Classical treatments with antifungal compounds need to be administered until the whole nail has grown out healthy.

- New compounds have been developed which reduce treatment to 3 months.

- Randomized, double-blind, parallel group, multicenter study for the comparison of two such new compounds (A and B) for oral treatment.

- Research question:

  Severity relative to treatment of TDO?

---

75 Year Anniversary NC State University Department of Statistics
• 2 × 189 patients randomized, 36 centers

• 48 weeks of total follow up (12 months)

• 12 weeks of treatment (3 months)

• measurements at months 0, 1, 2, 3, 6, 9, 12.
- Frequencies at each visit (both treatments):

![Toenail data graph]

- Time (months)
- % severe infections
- Treatment A
- Treatment B

Toenail data
Application to the Toenail Data

- Consider the model:

\[
Y_{ij} \sim \text{Bernoulli}(\mu_{ij})
\]

\[
\log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}
\]

\[
\text{Corr}(Y_{ij}, Y_{ij'}) = \alpha \quad \text{(working correlation)}
\]

- \(Y_{ij}\): severe infection (yes/no) at occasion \(j\) for patient \(i\)

- \(t_{ij}\): measurement time for occasion \(j\)

- \(T_i\): treatment group
Standard GEE

• SAS Code:

```sas
proc genmod/gee data=test descending;
class idnum timeclss;
model onyresp = treatn time treatn*time
   / dist=binomial;
repeated subject=idnum / withinsubject=timeclss
   type=exch covb corrw modelse;
run;
```

• Selected output:

▶ Regression parameters:

```
Analysis Of Initial Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Wald 95% Confidence Limits</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.5571</td>
<td>0.1090</td>
<td>-0.7708 -0.3433</td>
<td>26.10</td>
</tr>
<tr>
<td>treatn</td>
<td>1</td>
<td>0.0240</td>
<td>0.1565</td>
<td>-0.2827 0.3307</td>
<td>0.02</td>
</tr>
<tr>
<td>time</td>
<td>1</td>
<td>-0.1769</td>
<td>0.0246</td>
<td>-0.2251 -0.1288</td>
<td>51.91</td>
</tr>
<tr>
<td>treatn*time</td>
<td>1</td>
<td>-0.0783</td>
<td>0.0394</td>
<td>-0.1556 -0.0010</td>
<td>3.95</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td>3.95</td>
</tr>
</tbody>
</table>
```
### Analysis Of GEE Parameter Estimates

#### Empirical Standard Error Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.5840</td>
<td>0.1734</td>
<td>-0.2441</td>
<td>-3.37</td>
<td>0.0008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn</td>
<td>0.0120</td>
<td>0.2613</td>
<td>-0.5001</td>
<td>0.5241</td>
<td>0.05</td>
<td>0.9633</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>-0.1770</td>
<td>0.0311</td>
<td>-0.1161</td>
<td>-0.2380</td>
<td>5.69</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>treatn*time</td>
<td>-0.0886</td>
<td>0.0571</td>
<td>0.0233</td>
<td>-1.55</td>
<td>0.1208</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Analysis Of GEE Parameter Estimates

#### Model-Based Standard Error Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.5840</td>
<td>0.1344</td>
<td>-0.3204</td>
<td>-4.34</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn</td>
<td>0.0120</td>
<td>0.1866</td>
<td>0.3777</td>
<td>0.06</td>
<td>0.9486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>-0.1770</td>
<td>0.0209</td>
<td>-0.1361</td>
<td>-8.47</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn*time</td>
<td>-0.0886</td>
<td>0.0362</td>
<td>-0.0177</td>
<td>-2.45</td>
<td>0.0143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

▷ The working correlation:

Exchangeable Working Correlation

Correlation 0.420259237
Inference on Key Parameter: $\beta_3$

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate (s.e.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td>-0.0783 (0.0394)</td>
<td>0.0469</td>
</tr>
<tr>
<td>Model-based (naive)</td>
<td>-0.0886 (0.0362)</td>
<td>0.0143</td>
</tr>
<tr>
<td>Empirically corrected (robust)</td>
<td>-0.0886 (0.0571)</td>
<td>0.1208</td>
</tr>
</tbody>
</table>
### Inference on Key Parameter: $\beta_3$. Story 1.

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate (s.e.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td>-0.0783 (0.0394)</td>
<td>0.0469</td>
</tr>
<tr>
<td>Model-based (naive)</td>
<td>-0.0886 (0.0362)</td>
<td>0.0143</td>
</tr>
<tr>
<td>Empirically corrected (robust)</td>
<td>-0.0886 (0.0571)</td>
<td>0.1208</td>
</tr>
</tbody>
</table>

“The initial model is the most efficient estimator, because it assumes that each data point provides an independent piece of information. Based on this model, the treatment effect is marginally significant.”
Inference on Key Parameter: $\beta_3$. Story 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate (s.e.)</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td>-0.0783 (0.0394)</td>
<td>0.0469</td>
</tr>
<tr>
<td>Model-based (naive)</td>
<td>-0.0886 (0.0362)</td>
<td>0.0143</td>
</tr>
<tr>
<td>Empirically corrected (robust)</td>
<td>-0.0886 (0.0571)</td>
<td>0.1208</td>
</tr>
</tbody>
</table>

“The model-based estimator assumes that the various pairs of measurements per patient exhibit a common correlation. This is estimated to be $\hat{\alpha} = 0.42$, considered to be a plausible value. Therefore, inferences are based on the model-based estimator; this leads to a significant effect of treatment, with $p = 0.0143$.”
Inference on Key Parameter: $\beta_3$. Story 3.

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimate (s.e.)</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td>-0.0783 (0.0394)</td>
<td>0.0469</td>
</tr>
<tr>
<td>Model-based (naive)</td>
<td>-0.0886 (0.0362)</td>
<td>0.0143</td>
</tr>
<tr>
<td>Empirically corrected (robust)</td>
<td>-0.0886 (0.0571)</td>
<td>0.1208</td>
</tr>
</tbody>
</table>

“The empirically-corrected estimator assumes that the various pairs of measurements per patient exhibit a common correlation, but that, at the same time, this correlation assumption may be incorrect. In other words, it protects against misspecification. Inferences are based on this estimator. We conclude that there is no significant effect of treatment, with $p = 0.1208$.”
### Inference on Key Parameter: $\beta_3$. Story 4.

<table>
<thead>
<tr>
<th>Model</th>
<th>Working corr. $\alpha$</th>
<th>Estimate (s.e.)</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial model</td>
<td></td>
<td>-0.078 (0.039)</td>
<td>0.0469</td>
</tr>
<tr>
<td>Model-based (naive)</td>
<td>exchangeable</td>
<td>-0.089 (0.036)</td>
<td>0.0143</td>
</tr>
<tr>
<td>Emp. corr. (robust)</td>
<td>independence</td>
<td>-0.078 (0.055)</td>
<td>0.1515</td>
</tr>
<tr>
<td>Emp. corr. (robust)</td>
<td>exchangeable</td>
<td>-0.089 (0.057)</td>
<td>0.1208</td>
</tr>
<tr>
<td>Emp. corr. (robust)</td>
<td>unstructured</td>
<td>-0.114 (0.052)</td>
<td>0.0275</td>
</tr>
</tbody>
</table>

“The empirically-corrected estimator assumes that the various pairs of measurements per patient exhibit a certain structure, but that, at the same time, this correlation assumption may be incorrect. The working correlation that is closest to the true structure is generally most efficient. Inferences are based on this estimator, with unstructured working correlation. We conclude that there is a significant effect of treatment, with $p = 0.0275$.”
Here is to the NC State Statistics Department!