

# GLMs and GLMMs in the Analysis of Randomized Experiments

Steven Gilmour  
University of Southampton, UK

Optimum Design for Mixed Effects in Nonlinear and  
Generalised Linear Models  
August 2011

## Experiments with discrete responses

Myers *et al.* (2010) Response is number of defects out of 20 parts.

Run	A	B	C	D	Y
1	-1	-1	-1	-1	4
2	1	-1	-1	-1	10
3	-1	1	-1	-1	5
4	1	1	-1	-1	7
5	-1	-1	1	-1	8
6	1	-1	1	-1	6
7	-1	1	1	-1	8
8	1	1	1	-1	10
9	-1	-1	-1	1	4
10	1	-1	-1	1	7
11	-1	1	-1	1	5
12	1	1	-1	1	14
13	-1	-1	1	1	7
14	1	-1	1	1	9
15	-1	1	1	1	7
16	1	1	1	1	7

# Experiments with discrete responses

Nelder and Hamada (1996) *JQT* An experiment on car grille opening panels. Response is the number of defects.

Run	A	B	C	D	E	F	G	H	J	Y
1	-1	-1	-1	-1	1	1	1	1	1	56
2	1	-1	-1	-1	1	1	1	-1	-1	17
3	-1	1	-1	-1	1	-1	-1	1	-1	2
4	1	1	-1	-1	1	-1	-1	-1	1	4
5	-1	-1	1	-1	-1	1	-1	-1	1	3
6	1	-1	1	-1	-1	1	-1	1	-1	4
7	-1	1	1	-1	-1	-1	1	-1	-1	50
8	1	1	1	-1	-1	-1	1	1	1	2
9	-1	-1	-1	1	-1	-1	1	1	1	1
10	1	-1	-1	1	-1	-1	1	-1	-1	0
11	-1	1	-1	1	-1	1	-1	1	-1	3
12	1	1	-1	1	-1	1	-1	-1	1	12
13	-1	-1	1	1	1	-1	-1	-1	1	3
14	1	-1	1	1	1	-1	-1	1	-1	4
15	-1	1	1	1	1	1	1	-1	-1	0
16	1	1	1	1	1	1	1	1	1	0

# Outline

- Randomization and continuous responses

# Outline

- Randomization and continuous responses
- Count and binary data in randomization theory

# Outline

- Randomization and continuous responses
- Count and binary data in randomization theory
- Generalized Linear Mixed Models (GLMMs)

# Outline

- Randomization and continuous responses
- Count and binary data in randomization theory
- Generalized Linear Mixed Models (GLMMs)
- Independence - experimental and observational units

# Outline

- Randomization and continuous responses
- Count and binary data in randomization theory
- Generalized Linear Mixed Models (GLMMs)
- Independence - experimental and observational units
- Examples



# Randomization theory

We have  $n$  experimental units (e.g. field plots, human subjects, runs of a process).

# Randomization theory

We have  $n$  experimental units (e.g. field plots, human subjects, runs of a process).

We will compare  $n_t$  treatments (e.g. varieties, drugs, combinations of levels of process inputs).

# Randomization theory

Assumed model for a response variable  $y$  (e.g. log yield, diastolic blood pressure, darkness of output product) is (on unit  $i$  with treatment  $k$  applied)

$$y_{i(k)} = u_i + t_k, \quad (1)$$

where  $u_i$  would have been the response for a treatment with 0 effect,  $t_k$  is the effect of treatment  $k$  and the only assumption needed is additivity of unit and treatment effects.

# Randomization theory

The form of randomization used determines the appropriate analysis, e.g. for completely randomized design with  $\frac{n}{n_t}$  replicates of each treatment, considering all possible outcomes of the randomization, model (1) becomes

$$y_{i(k)} = \sum_{j=1}^n \delta_{ij} u_j + t_k, \quad (2)$$

where  $\delta_{ij} = \begin{cases} 1 & \text{if unit label } i \text{ applies to unit } j; \\ 0 & \text{otherwise.} \end{cases}$

# Randomization theory

This implies that the usual linear models analysis gives:

- minimum variance linear unbiased estimators of  $t_k$  (and linear functions of  $t_k$ s);

# Randomization theory

This implies that the usual linear models analysis gives:

- minimum variance linear unbiased estimators of  $t_k$  (and linear functions of  $t_k$ s);
- minimum variance unbiased estimators of  $\sigma^2$ , the inter-unit variance, and any other variance components;

# Randomization theory

This implies that the usual linear models analysis gives:

- minimum variance linear unbiased estimators of  $t_k$  (and linear functions of  $t_k$ s);
- minimum variance unbiased estimators of  $\sigma^2$ , the inter-unit variance, and any other variance components;
- unbiased estimators of  $s.e.(\hat{t}_k)$  (and linear functions);

# Randomization theory

This implies that the usual linear models analysis gives:

- minimum variance linear unbiased estimators of  $t_k$  (and linear functions of  $t_k$ s);
- minimum variance unbiased estimators of  $\sigma^2$ , the inter-unit variance, and any other variance components;
- unbiased estimators of  $s.e.(\hat{t}_k)$  (and linear functions);
- $E(MS_{Treat}) = \sigma^2$  if  $t_k = 0 \forall k$ .



# Randomization theory

However, exact inference requires the use of permutation tests since the sampling distribution of  $\hat{t}_k$  is not of a regular form.

# Randomization theory

However, exact inference requires the use of permutation tests since the sampling distribution of  $\hat{t}_k$  is not of a regular form.

We usually (for continuous  $y$ ) use a Normal approximation, i.e. model (2) becomes

$$y_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$$

# Randomization theory

However, exact inference requires the use of permutation tests since the sampling distribution of  $\hat{t}_k$  is not of a regular form.

We usually (for continuous  $y$ ) use a Normal approximation, i.e. model (2) becomes

$$y_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$$
$$\text{i.e. } y_{i(k)} = \mu + t_k + \epsilon_i; \epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$

# Randomization theory

However, exact inference requires the use of permutation tests since the sampling distribution of  $\hat{t}_k$  is not of a regular form.

We usually (for continuous  $y$ ) use a Normal approximation, i.e. model (2) becomes

$$\begin{aligned} y_{i(k)} &= \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2) \\ \text{i.e. } y_{i(k)} &= \mu + t_k + \epsilon_i; \epsilon_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2) \\ \text{i.e. } y_{i(k)} &\stackrel{\text{ind}}{\sim} N(\mu_{ik}, \sigma^2); \mu_{ik} = \mu + t_k = \eta_{ik} \end{aligned} \quad (3)$$

# Randomization theory

Extensions of this normal approximation often used almost automatically.

# Randomization theory

Extensions of this normal approximation often used almost automatically.

They are the basis for mixed models analysis of nonorthogonal split-plot and multi-stratum designs, etc.:

- “sum of errors” argument;

# Randomization theory

Extensions of this normal approximation often used almost automatically.

They are the basis for mixed models analysis of nonorthogonal split-plot and multi-stratum designs, etc.:

- “sum of errors” argument;
- model checking;

# Randomization theory

Extensions of this normal approximation often used almost automatically.

They are the basis for mixed models analysis of nonorthogonal split-plot and multi-stratum designs, etc.:

- “sum of errors” argument;
- model checking;
- robustness of Normal inference.



# Randomization theory for discrete data

If the response variable is discrete, a direct Normal approximation is inappropriate. In practice GLMs are often used, e.g. logistic regression or log-linear models.

# Randomization theory for discrete data

If the response variable is discrete, a direct Normal approximation is inappropriate. In practice GLMs are often used, e.g. logistic regression or log-linear models.

However, if we can assume additivity on some scale, the exact randomization theory should still apply.

# Randomization theory for discrete data

If the response variable is discrete, a direct Normal approximation is inappropriate. In practice GLMs are often used, e.g. logistic regression or log-linear models.

However, if we can assume additivity on some scale, the exact randomization theory should still apply.

Problem: there is no way to write a GLM in the form of (1).

# Generalized Linear Mixed Models

Apply randomization theory by assuming additivity in the (unobservable) linear predictor  $\eta$ , i.e.

$$\eta_{i(k)} = u_i + t_k.$$

# Generalized Linear Mixed Models

Apply randomization theory by assuming additivity in the (unobservable) linear predictor  $\eta$ , i.e.

$$\eta_{i(k)} = u_i + t_k.$$

Under randomization, we get

$$\eta_{i(k)} = \sum_{j=1}^n \delta_{ij} u_j + t_k.$$

# Generalized Linear Mixed Models

The usual Normal approximation gives

$$\eta_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).$$

# Generalized Linear Mixed Models

The usual Normal approximation gives

$$\eta_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).$$

Now the appropriate distributional assumption for observational units gives a GLMM, e.g.

$$c_{i(k)} \stackrel{\text{ind}}{\sim} \text{Poisson}(\lambda_{i(k)});$$
$$\ln(\lambda_{i(k)}) = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).$$

# Generalized Linear Mixed Models

The usual Normal approximation gives

$$\eta_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).$$

Now the appropriate distributional assumption for observational units gives a GLMM, e.g.

$$c_{i(k)} \stackrel{\text{ind}}{\sim} \text{Poisson}(\lambda_{i(k)});$$

$$\ln(\lambda_{i(k)}) = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).$$

$$r_{i(k)} \stackrel{\text{ind}}{\sim} \text{Binom}(m, \pi_{i(k)});$$

$$\ln\left(\frac{\pi_{i(k)}}{1 - \pi_{i(k)}}\right) = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).$$





# Generalized Linear Models

Simple GLMs are, e.g.

$$c_{i(k)} \stackrel{\text{ind}}{\sim} \text{Poisson}(\lambda_{ik}); \ln(\lambda_{ik}) = \mu + t_k = \eta_{ik} \quad (4)$$

$$r_{i(k)} \stackrel{\text{ind}}{\sim} \text{Binom}(m, \pi_{ik}); \ln\left(\frac{\pi_{ik}}{1 - \pi_{ik}}\right) = \mu + t_k = \eta_{ik} \quad (5)$$

# Generalized Linear Models

Simple GLMs are, e.g.

$$c_{i(k)} \stackrel{\text{ind}}{\sim} \text{Poisson}(\lambda_{ik}); \ln(\lambda_{ik}) = \mu + t_k = \eta_{ik} \quad (4)$$

$$r_{i(k)} \stackrel{\text{ind}}{\sim} \text{Binom}(m, \pi_{ik}); \ln\left(\frac{\pi_{ik}}{1 - \pi_{ik}}\right) = \mu + t_k = \eta_{ik} \quad (5)$$

They have no justification based on randomization.

## Independence assumptions (or constant probability)

The problem arises because experimental units and observational units are fundamentally different concepts, *even when they have the same physical meaning*, e.g. a field plot might be both the experimental unit and the observational unit.

## Independence assumptions (or constant probability)

The problem arises because experimental units and observational units are fundamentally different concepts, *even when they have the same physical meaning*, e.g. a field plot might be both the experimental unit and the observational unit.

Assuming independence between sub-units of the same observational unit (and constant probability over observational units) is reasonable (as a working assumption) and leads to Poisson and Binomial distributions.

## Independence assumptions (or constant probability)

Assuming independence between sub-units of the same experimental unit (and constant probability over experimental units) is *clearly not a reasonable working assumption*.

# Independence assumptions (or constant probability)

Assuming independence between sub-units of the same experimental unit (and constant probability over experimental units) is *clearly not a reasonable working assumption*.

Binomial distribution assumes:

- number of successes in *independent* trials;

# Independence assumptions (or constant probability)

Assuming independence between sub-units of the same experimental unit (and constant probability over experimental units) is *clearly not a reasonable working assumption*.

Binomial distribution assumes:

- number of successes in *independent* trials;
- *constant probability* of success.

# Independence assumptions (or constant probability)

Assuming independence between sub-units of the same experimental unit (and constant probability over experimental units) is *clearly not a reasonable working assumption*.

Binomial distribution assumes:

- number of successes in *independent* trials;
- *constant probability* of success.



# Independence assumptions (or constant probability)

Assuming independence between sub-units of the same experimental unit (and constant probability over experimental units) is *clearly not a reasonable working assumption*.

Binomial distribution assumes:

- number of successes in *independent* trials;
- *constant probability* of success.

Modelling observations (one or many) from a single experimental unit as samples from a Binomial population seems reasonable.

## Independence assumptions (or constant probability)

Consider modelling observations from different experimental units, both with treatment  $k$ , as samples from a Binomial population (i.e. model (5)).

## Independence assumptions (or constant probability)

Consider modelling observations from different experimental units, both with treatment  $k$ , as samples from a Binomial population (i.e. model (5)).

Is the probability of success constant over different experimental units? Are observations in the same experimental unit independent?

## Independence assumptions (or constant probability)

Consider modelling observations from different experimental units, both with treatment  $k$ , as samples from a Binomial population (i.e. model (5)).

Is the probability of success constant over different experimental units? Are observations in the same experimental unit independent?

Of course not! The physical process of “treating” them ensures not.

# Independence assumptions (or constant probability)

Poisson distribution assumes:

- number of events which occur *independently*;

# Independence assumptions (or constant probability)

Poisson distribution assumes:

- number of events which occur *independently*;
- constant rate (in time, space, etc.).

# Independence assumptions (or constant probability)

Poisson distribution assumes:

- number of events which occur *independently*;
- constant rate (in time, space, etc.).

# Independence assumptions (or constant probability)

Poisson distribution assumes:

- number of events which occur *independently*;
- constant rate (in time, space, etc.).

Again, modelling observations from a single experimental unit as samples from a Poisson population seems reasonable.



# Independence assumptions (or constant probability)

Poisson distribution assumes:

- number of events which occur *independently*;
- constant rate (in time, space, etc.).

Again, modelling observations from a single experimental unit as samples from a Poisson population seems reasonable.

Modelling observations from replicate experimental units as samples from a Poisson distribution (i.e. model (4)) is clearly wrong.

# Independence assumptions (or constant probability)

Poisson distribution assumes:

- number of events which occur *independently*;
- constant rate (in time, space, etc.).

Again, modelling observations from a single experimental unit as samples from a Poisson population seems reasonable.

Modelling observations from replicate experimental units as samples from a Poisson distribution (i.e. model (4)) is clearly wrong.

“Treating” units ensures the rate of events will be different in different experimental units.

# Binomial Example

Myers *et al.* fit a GLM (binomial distribution, logit link) with main effects and two-factor interactions in the linear predictor, then perform model simplification.

# Binomial Example

Myers *et al.* fit a GLM (binomial distribution, logit link) with main effects and two-factor interactions in the linear predictor, then perform model simplification.

This suggests strong evidence that  $A$  has a large main effect and clear evidence of an  $A \times C$  interaction, with absolutely no evidence of overdispersion.

# Binomial Example

Myers *et al.* fit a GLM (binomial distribution, logit link) with main effects and two-factor interactions in the linear predictor, then perform model simplification.

This suggests strong evidence that  $A$  has a large main effect and clear evidence of an  $A \times C$  interaction, with absolutely no evidence of overdispersion.

A similar analysis using GLMMs with the runs random effect suggests only moderate evidence of a main effect of  $A$  and a small, but non-zero random runs variance component.

# Binomial Example

Myers *et al.* fit a GLM (binomial distribution, logit link) with main effects and two-factor interactions in the linear predictor, then perform model simplification.

This suggests strong evidence that  $A$  has a large main effect and clear evidence of an  $A \times C$  interaction, with absolutely no evidence of overdispersion.

A similar analysis using GLMMs with the runs random effect suggests only moderate evidence of a main effect of  $A$  and a small, but non-zero random runs variance component.

The GLM analysis wrongly attributes the run to run variation to fixed effects.

# Car Grilles Experiment

Nelder and Hamada fit a GLM (Poisson distribution, log link) with main effects in the linear predictor, then drop main effects and add interactions to get a plausible model.

# Car Grilles Experiment

Nelder and Hamada fit a GLM (Poisson distribution, log link) with main effects in the linear predictor, then drop main effects and add interactions to get a plausible model.

This suggests a model with main effects of  $A$ ,  $C$ ,  $D$ ,  $G$  and  $H$  and interactions  $A \times D$  and  $D \times G$ , the last two significant at the 0.01% level. There is slight evidence of overdispersion.



# Car Grilles Experiment

Nelder and Hamada fit a GLM (Poisson distribution, log link) with main effects in the linear predictor, then drop main effects and add interactions to get a plausible model.

This suggests a model with main effects of  $A$ ,  $C$ ,  $D$ ,  $G$  and  $H$  and interactions  $A \times D$  and  $D \times G$ , the last two significant at the 0.01% level. There is slight evidence of overdispersion.

The GLMM analysis suggests only moderate evidence of a  $D$  main effect and a fairly large runs variance component.

# Car Grilles Experiment

Nelder and Hamada fit a GLM (Poisson distribution, log link) with main effects in the linear predictor, then drop main effects and add interactions to get a plausible model.

This suggests a model with main effects of  $A$ ,  $C$ ,  $D$ ,  $G$  and  $H$  and interactions  $A \times D$  and  $D \times G$ , the last two significant at the 0.01% level. There is slight evidence of overdispersion.

The GLMM analysis suggests only moderate evidence of a  $D$  main effect and a fairly large runs variance component.

The GLM analysis suggests a very complex system, when the data support nothing other than large run to run variation.

# Back to GLMMs

What about Normal distribution?

# Back to GLMMs

What about Normal distribution?

$$\begin{aligned}\mu_{i(k)} &= \eta_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2); \\ y_{i(k)} &\stackrel{\text{ind}}{\sim} N(\mu_{i(k)}, \sigma_e^2); \mu_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).\end{aligned}$$

## Back to GLMMs

What about Normal distribution?

$$\begin{aligned}\mu_{i(k)} &= \eta_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2); \\ y_{i(k)} &\stackrel{\text{ind}}{\sim} N(\mu_{i(k)}, \sigma_e^2); \mu_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).\end{aligned}$$

reduces to

$$y_{i(k)} \stackrel{\text{ind}}{\sim} N(\mu + t_k, \sigma_u^2 + \sigma_e^2).$$

# Back to GLMMs

What about Normal distribution?

$$\begin{aligned}\mu_{i(k)} &= \eta_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2); \\ y_{i(k)} &\stackrel{\text{ind}}{\sim} N(\mu_{i(k)}, \sigma_e^2); \mu_{i(k)} = \mu_i + t_k; \mu_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma_u^2).\end{aligned}$$

reduces to

$$y_{i(k)} \stackrel{\text{ind}}{\sim} N(\mu + t_k, \sigma_u^2 + \sigma_e^2).$$

Write this as

$$y_{i(k)} \stackrel{\text{ind}}{\sim} N(\mu_{ik}, \sigma^2),$$

where  $\mu_{ik} = \mu + t_k$ ;  $\sigma^2 = \sigma_u^2 + \sigma_e^2$  and we have equation (3) again.

# Conclusions

- For count and proportion data in experiments we should *always* use GLMMs rather than GLMs.

# Conclusions

- For count and proportion data in experiments we should *always* use GLMMs rather than GLMs.
- GLMs will often show overdispersion and appropriate modelling of this *might* give the same conclusions as GLMMs. However, there are several different ways to model overdispersion, e.g including more terms in the linear predictor.



# Conclusions

- Designs with restricted randomization (e.g. blocks, rows and columns, etc.) are easily dealt with in the usual way, by introducing a random effect for each level of randomization and the product of crossed blocking factors, e.g.

$$\eta_{ij(k)} = \mu + \rho_i + \gamma_j + \epsilon_{ij} + t_k;$$

$$\rho_i \stackrel{\text{iid}}{\sim} N(0, \sigma_r^2), \quad \gamma_j \stackrel{\text{iid}}{\sim} N(0, \sigma_c^2), \quad \epsilon_{ij} \stackrel{\text{iid}}{\sim} N(0, \sigma_u^2).$$

# Ongoing Work

Heiko Großmann and I have been trying to do an exact analysis based on the proper randomization distribution, rather than the normal approximation.

# Ongoing Work

Heiko Großmann and I have been trying to do an exact analysis based on the proper randomization distribution, rather than the normal approximation.

This is hard, perhaps impossible.

# Ongoing Work

Heiko Großmann and I have been trying to do an exact analysis based on the proper randomization distribution, rather than the normal approximation.

This is hard, perhaps impossible.

Practical questions:

- In practice, is a normal approximation good enough?

# Ongoing Work

Heiko Großmann and I have been trying to do an exact analysis based on the proper randomization distribution, rather than the normal approximation.

This is hard, perhaps impossible.

Practical questions:

- In practice, is a normal approximation good enough? When is it not?

# Ongoing Work

Heiko Großmann and I have been trying to do an exact analysis based on the proper randomization distribution, rather than the normal approximation.

This is hard, perhaps impossible.

Practical questions:

- In practice, is a normal approximation good enough? When is it not?
- Are diagnostics useful?

# Ongoing Work

Heiko Großmann and I have been trying to do an exact analysis based on the proper randomization distribution, rather than the normal approximation.

This is hard, perhaps impossible.

Practical questions:

- In practice, is a normal approximation good enough? When is it not?
- Are diagnostics useful? When do they work?

# Ongoing Work

Heiko Großmann and I have been trying to do an exact analysis based on the proper randomization distribution, rather than the normal approximation.

This is hard, perhaps impossible.

Practical questions:

- In practice, is a normal approximation good enough? When is it not?
- Are diagnostics useful? When do they work?
- If the normal approximation seems to be failing, what should we try next?



## Back to the conference topic

Obviously this has implications for optimal design of experiments with discrete responses.

## Back to the conference topic

Obviously this has implications for optimal design of experiments with discrete responses.

Thank you.

Questions or comments welcome!