Statistical Techniques II

EXST7015

Split-plot and Repeated Measures Designs

Split-plot and Repeated Measure Designs

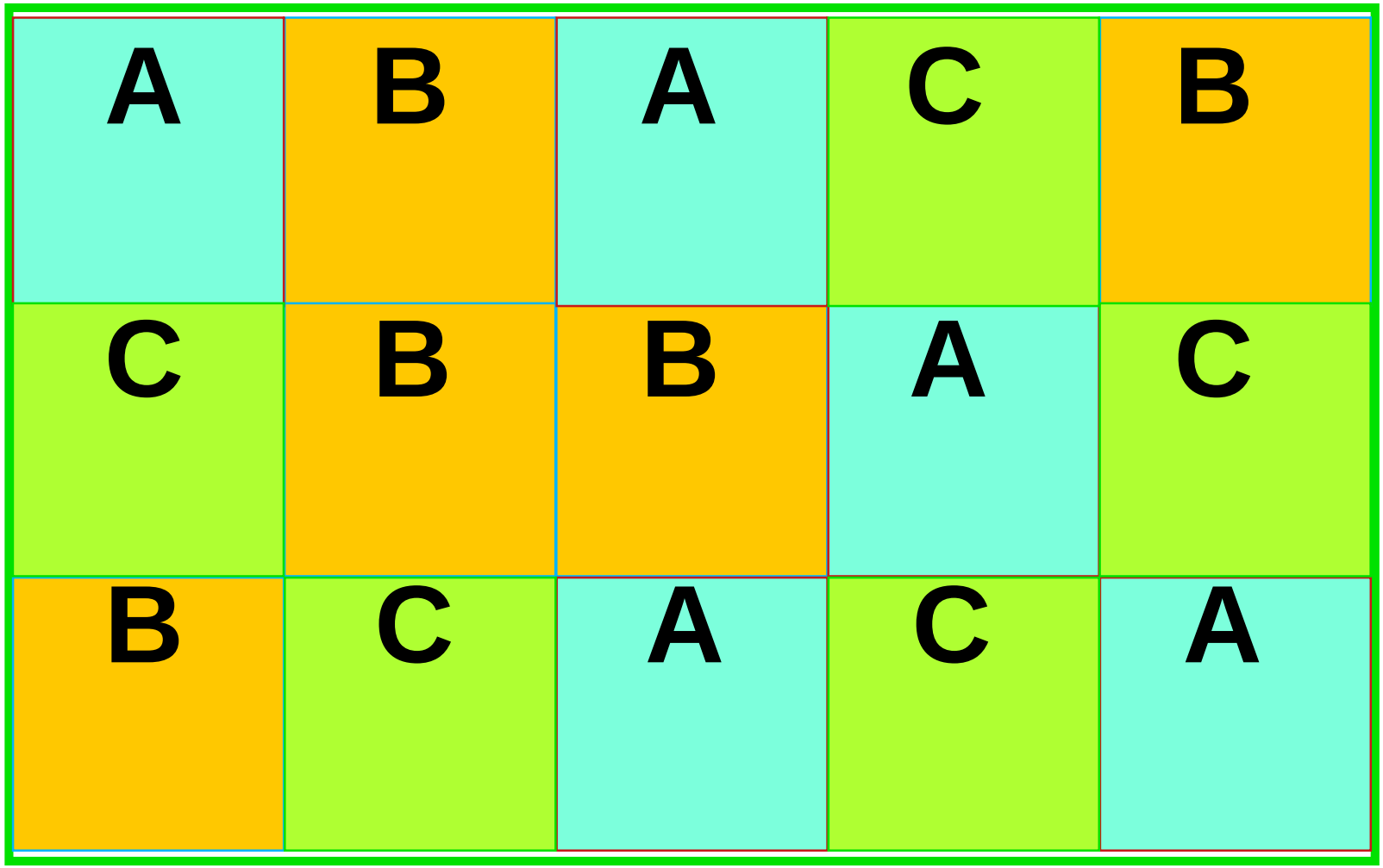
- **The Split-plot and Repeated Measures "Designs" combine elements of design (error structure) and treatment arrangement concepts.**
- **These are designs with two levels, a "Main Plot", with its own treatment and error, and a "Sub-plot", with its own treatment and error.**
- **It is possible to have more than just an "*a priori*" treatment in both levels.**

Split-plot Designs *(continued)*

- **The (minimum of) two treatments (from the main and sub plots) are usually cross classified .**
- **Either Main or Subplot may have nested error structure.**
- **The simplest split plot would have the following model (CRD).**
  - $Y_{ijk} = \mu + \tau_{1i} + \gamma_{ij} + \tau_{2k} + \tau_1\tau_{2ik} + \varepsilon_{ijk}$
-

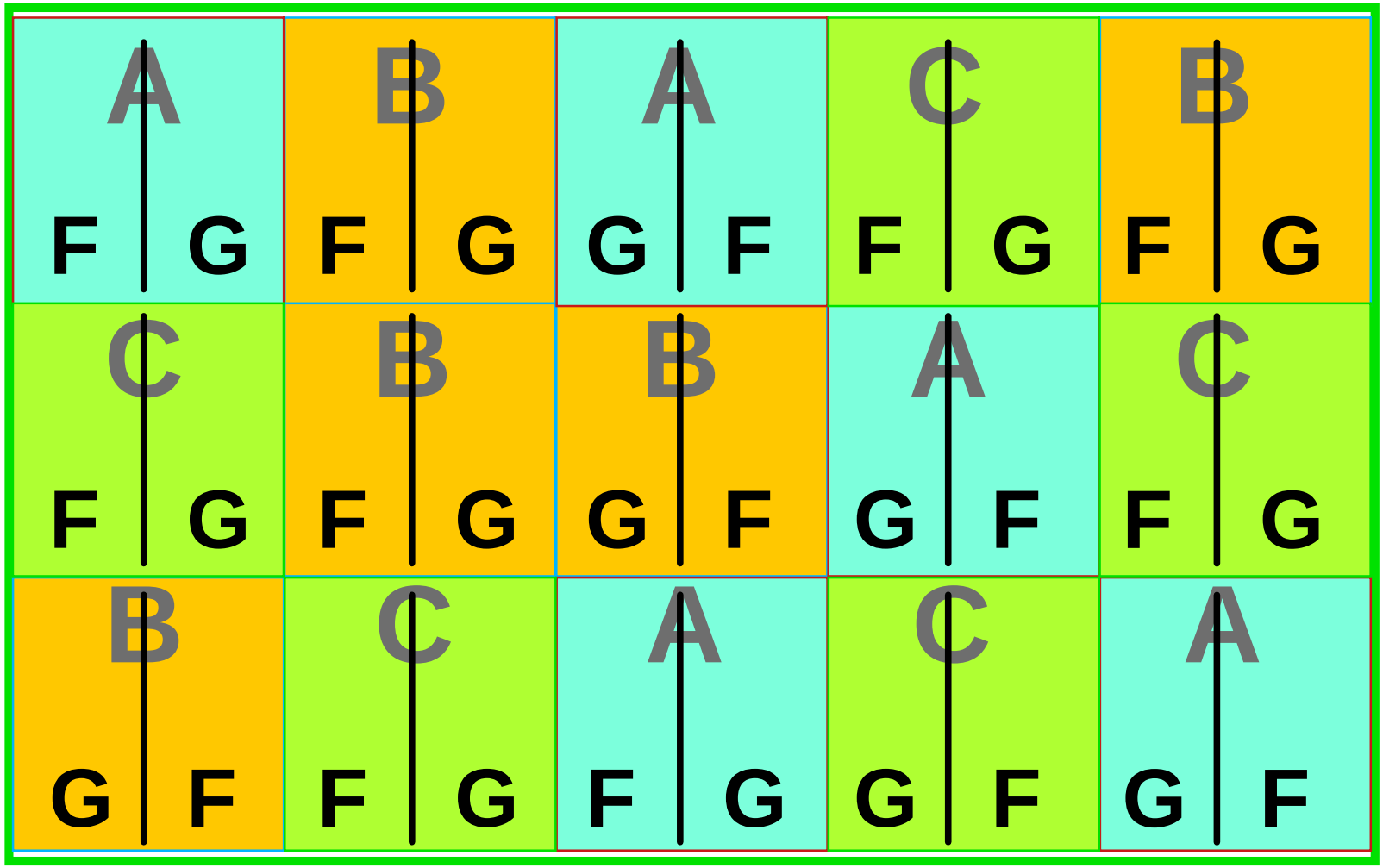
Split-plot Designs *(continued)*

- **Example with CRD main plot.**



Split-plot Designs *(continued)*

- **Each plot SPLIT for a new treatment.**



Split-plot Designs *(continued)*

- **Split-plot design source table. The d.f. for error(b) is the usual $t_1*t_2*(n-1)$ less the d.f. for error(a), $t_1*(n-1)$, giving $t_1*(t_2-1)(n-1)$.**

Source	d.f.
Treatment1	t_1-1 = 2
Error(a)	$t_1(n-1)$ = 12
Treatment 2	t_2-1 = 1
Tmt1*Tmt2	$(t_1-1)(t_2-1)$ = 2
Error(b)	$t_1*(t_2-1)(n-1)$ = 12
Total	$t_1*t_{2*}n-1$ = 29

Split-plot Designs *(continued)*

- **Split-plot design - examples of splits**
  - **We may split a plot to do a new treatment, e.g. an agricultural experiment with fertilizer treatments in plots may have a herbicide applied to half of each plot an not to the other half.**
  - **A soil study of contaminants may measure levels of the chemical of interest at various levels in a soil core (0-5 cm, 6-10 cm, 11-15 cm, etc), so the core is split.**

Split-plot Designs *(continued)*

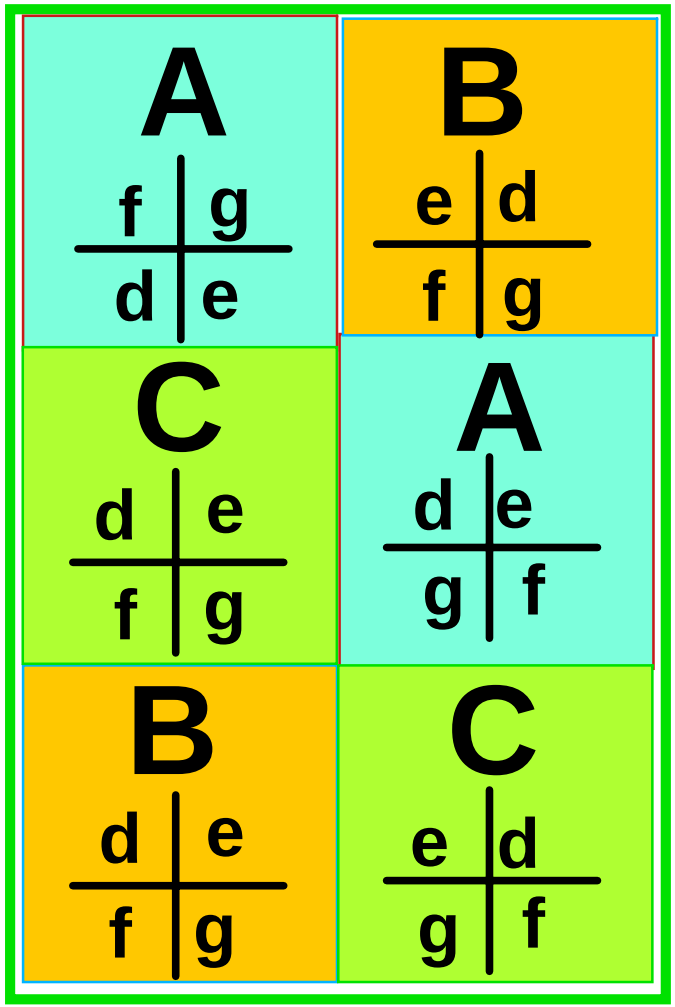
- **Split-plot split examples (continued)**
  - **A study of the growth of plants, e.g. Spartina in a marsh, may split the plant into above ground, root and rhizome biomass.**
  - **Anytime a treatment occurs within an experimental unit, we have a split-plot. If we are studying diets of fish, and put a male and female fish in each aquaria, weight gain of hogs with large and small hogs in each pen, etc.**

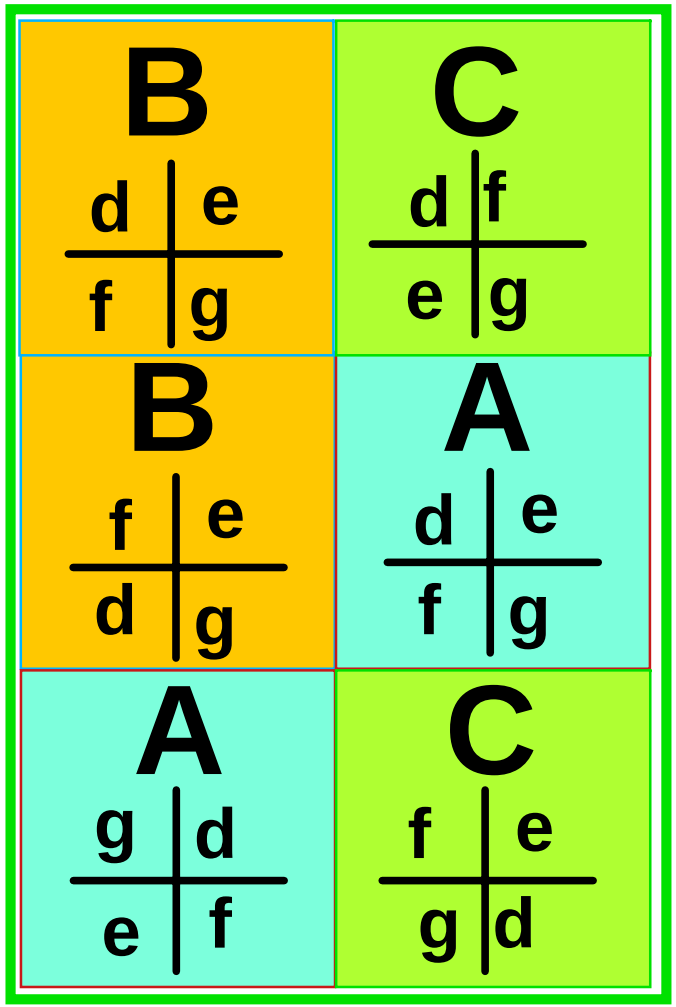
Split-plot Designs *(continued)*

- **More complex designs are possible. The main plot may be an RBD,**
  - **the main plot and/or sub plot treatments may be factorial or nested.**
  - **It is possible to have plots that are split twice,**
  - **or split and measured repeatedly.**
- **These designs are complicated, difficult to analyze and difficult to interpret. So why do you do them?**

Split-plot Designs *(continued)*

- **Split plot design with an RBD main plot.**





Split-plot Designs *(continued)*

- **This design has two blocks, three levels in the main plot treatment and four levels in the subplot treatment.**
- **For the main plot the analysis is the same as any RBD. This one will have treatments, blocks, treatment*block interaction and replicated experimental units in blocks.**
  - $Y_{ijk}=\mu+\beta_i+\tau_{1j}+\tau\beta_{ij}+\gamma_{ijk}+\tau_{2l}+\tau_1\tau_{2il}+\tau_1\tau_2\beta_{ijl}+\varepsilon_{ijkl}$

Split-plot Designs *(continued)*

- **Source table RBD main plot in split-plot.**

Source	d.f.
Block	$b-1 = 1$
Treatment 1	$t_1-1 = 2$
$B*T_1$	$(b-1)(t_1-1) = 2$
Error(a)	$bt_1(n-1) = 6$
Treatment 2	$t_2-1 = 3$
Tmt1*Tmt2	$(t_1-1)(t_2-1) = 6$
$B*T_2+B*T_1*T_2$ (pooled)	$(b-1)(t_1-1)(t_2-1) = 9$
Error(b)	$b*t_1*(t_2-1)(n-1) = 18$
Total	$b*t_1*t_{2*}n-1 = 47$

Repeated measures

- The repeated measures design is similar to a split-plot.
- We have a "main plot", which can be any of the designs we have discussed previously (CRD, RBD, LSD).
- We then take repeated measurements over time within the plots. If these "repeated measures" are independent, then this "time" factor is just cross-classified with the treatment.

Repeated *(continued)*

- **If, however, the measurements are NOT independent, we have a repeated measures design.**
- **Independence? Again? Yep.**
- **What do I mean by independent? For example, if you are sampling sugar content of an ear of corn from a plot, or the height of Spartina in a plot, you ask, "are they independent or not?"**

Repeated *(continued)*

- **If you measure a different ear of corn from a different plant each time, or measure a different Spartina plant, they are probably independent.**
- **However, if you measure a kernel from the same ear of corn, or the same Spartina plant each time, they are NOT independent.**
-

Repeated *(continued)*

- **Some examples of split plot and repeated measures variables.**
  - **Pre-post tests on people, in fact most any experiment where several levels of a treatment(s) are measured on the same subject (= a person).**
  - **Soil samples or water samples at different depths (in the same site).**
  - **Epiphytes on Spartina counted below, at and above the tide line (on the same plant).**

Repeated *(continued)*

- **More examples**
  - **Studies on plants like sugar cane where we measure production in year1, year2 and year3 on the same biological material.**
  - **Ditto for asparagus, trees, artichokes, etc.**
  - **In general, any time your experimental unit has a treatment applied to each experimental unit, this is a split plot. If the experimental unit is measure over time it is repeated measures.**

Repeated *(continued)*

- **Why is this independence important? What can we do about it?**
- **Lets BRIEFLY revisit the X and X'X matrices.**
- **The X matrix for designs consists of columns of 0 values and 1 values, arranged to distinguish between categories.**

Repeated *(continued)*

- **For a simple CRD with 4 treatment levels the X'X matrix may look like the following.**
-

X =	**1**	**0**	**0**	**0**
	1	**0**	**0**	**0**
	0	**1**	**0**	**0**
	0	**1**	**0**	**0**
	0	**0**	**1**	**0**
	0	**0**	**1**	**0**
	0	**0**	**0**	**1**
	0	**0**	**0**	**1**

Repeated *(continued)*

- **For a simple CRD with 4 treatment levels the X'X matrix would look like the following.**

X'X =	n_1	0	0	0
	0	n_2	0	0
	0	0	n_3	0
	0	0	0	n_4

Repeated *(continued)*

- **For a simple CRD with 4 treatment levels the $(X'X)^{-1}$ matrix would look like the following.**

$(X'X)^{-1}$ =	$1/n_1$	0	0	0
	0	$1/n_2$	0	0
	0	0	$1/n_3$	0
	0	0	0	$1/n_4$

Repeated *(continued)*

- **To get the variances and covariances we multiply by the MSE, as you know. This gives MSE/n on the main diagonal, and zeros on the off diagonal.**
- **All those zeros on the off diagonal mean that THERE IS NO COVARIANCE BETWEEN THE TREATMENTS. This is well and good, we do not expect covariances between the independently sampled treatments.**

Repeated *(continued)*

- **But for the split plot and repeated measures, we do actually expect some covariances!!**
  - **Maybe the covariance is simple, perhaps it is a constant. This would be the assumption for split-plot designs, and we can use GLM for our tests (but not for subplot standard errors).**
  - **But much recent scientific investigation has found that often the structure is not simple.**

Repeated *(continued)*

- **See the Handout on covariance structure. A couple that are of particular interest are the variance component structure (split plot) and a favorite repeated measure structure AR(1).**
- **There are many other structures, including some where the structure follows some type of regression line. This is frequent in "spatial statistics" where covariance is modeled as a function of distance between plots.**

Repeated *(continued)*

- **For a simple CRD with 4 treatment levels variance structure is given below. This is the SAS default in the proc mixed repeated statement.**

VarCov =	σ^2	**0**	**0**	**0**
	0	σ^2	**0**	**0**
	0	**0**	σ^2	**0**
	0	**0**	**0**	σ^2

Repeated *(continued)*

- **For a CRD with 4 treatment levels SAS also has a "VC" option for "variance components".**

VarCov =	σ_1^2	**0**	**0**	**0**
	0	σ_2^2	**0**	**0**
	0	**0**	σ_3^2	**0**
	0	**0**	**0**	σ_4^2

Repeated *(continued)*

- **The usual and simplest assumption for variance components for a sub-plot treatment is "Compound Symmetry", SAS option CS. This is the only structure permitted by PROC GLM.**

VarCov =	$\sigma^2+\sigma_1^2$	σ_1^2	σ_1^2	σ_1^2
	σ_1^2	$\sigma^2+\sigma_1^2$	σ_1^2	σ_1^2
	σ_1^2	σ_1^2	$\sigma^2+\sigma_1^2$	σ_1^2
	σ_1^2	σ_1^2	σ_1^2	$\sigma^2+\sigma_1^2$

Repeated *(continued)*

- **Another sub-plot treatment structure is "unstructured", SAS option UN. This takes up the most degrees of freedom to estimate.**

VarCov =	σ_{11}^2	σ_{12}	σ_{13}	σ_{14}
	σ_{12}	σ_{22}^2	σ_{23}	σ_{24}
	σ_{13}	σ_{23}	σ_{33}^2	σ_{34}
	σ_{14}	σ_{24}	σ_{34}^2	σ_{44}^2

Repeated *(continued)*

- **One of the most popular structures for repeated measures designs is the "first order autoregressive", SAS option AR(1).**

VarCov =	σ^2	$\sigma^2\rho^1$	$\sigma^2\rho^2$	$\sigma^2\rho^3$
	$\sigma^2\rho^1$	σ^2	$\sigma^2\rho^1$	$\sigma^2\rho^2$
	$\sigma^2\rho^2$	$\sigma^2\rho^1$	σ^2	$\sigma^2\rho^1$
	$\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho^1$	σ^2

Repeated *(continued)*

- **Other sub-plot treatment structures are possible. Below is a "Toeplitz" structure (SAS option TOEP). The number of bands can be varied.**

VarCov =	σ^2	σ_1	σ_2	σ_3
	σ_1	σ^2	σ_1	σ_2
	σ_2	σ_1	σ^2	σ_1
	σ_3	σ_2	σ_1	σ^2

Repeated *(continued)*

- **A final note on different structures. There is an area of statistics called "spatial statistics" where the covariance structure is a function of distance.**
- **In order to fit these covariance structures and get correct subplot standard error estimates we must use PROC MIXED. These structures are not available in PROC GLM.**

Repeated *(continued)*

- **How are the various structures fitted in SAS? We will look at two examples with different structures.**
- **When fit with PROC GLM we assume compound symmetry, and this is the default in PROC MIXED.**
  - **There are some "adjustments" that can be made by some GLM options, but we will not cover these since all problems are resolved in PROC MIXED.**