Statistical Analysis of Time-Repeated Measurements on each Experimental Subject

Multiple Measurements on Same Subject

When Measurements are Separated by enough Time to be

<u>Uncorrelated</u>

$$\Sigma_{nxn} = \sigma_\varepsilon^2 \cdot I_{nxn} = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\varepsilon^2 \end{pmatrix}_{nxn}$$

n measurements on the same subject

Multiple Measurements on Same Subject

When Measurements are near enough in Time to be ***correlated***

$$\Sigma_{nxn} = \begin{pmatrix} \sigma_{\varepsilon}^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{\varepsilon}^2 & \cdots & \sigma_{21} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{\varepsilon}^2 \end{pmatrix}_{nxn}$$

n measurements on the same subject

Multiple Measurements on Same Subject

When Measurements are near enough in Time to be

Correlated [Simplified Notation: ρ indicates correlation]

$$\mathbf{\Sigma}_{nxn} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}_{nxn}$$

n measurements on the same subject

Correlation Influences Hypothesis Tests

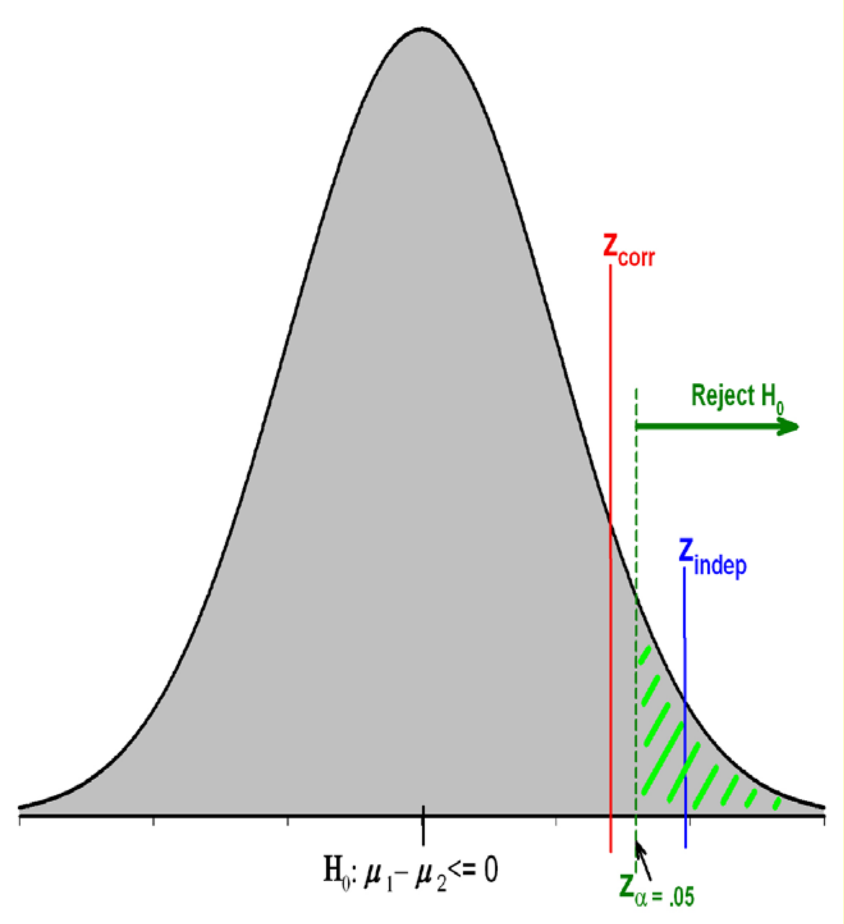
$$z_{indep} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \cdot \sqrt{\frac{2}{n}}}$$

$$Z_{corr} = \frac{\bar{x}_1 - \bar{x}_2}{\sigma \cdot \sqrt{\frac{2\{1 + (n-1) \cdot \rho\}}{n}}}$$

If positive correlation is present and ignored, a treatment effect can be incorrectly declared significant.

Divisor: n for z_{indep}

$n_{effective}$ for Z_{corr}



In Presence of Correlation
Need Larger Sample Size to be as "Effective"

"…*positive autocorrelation* results in *'loss of information'.*

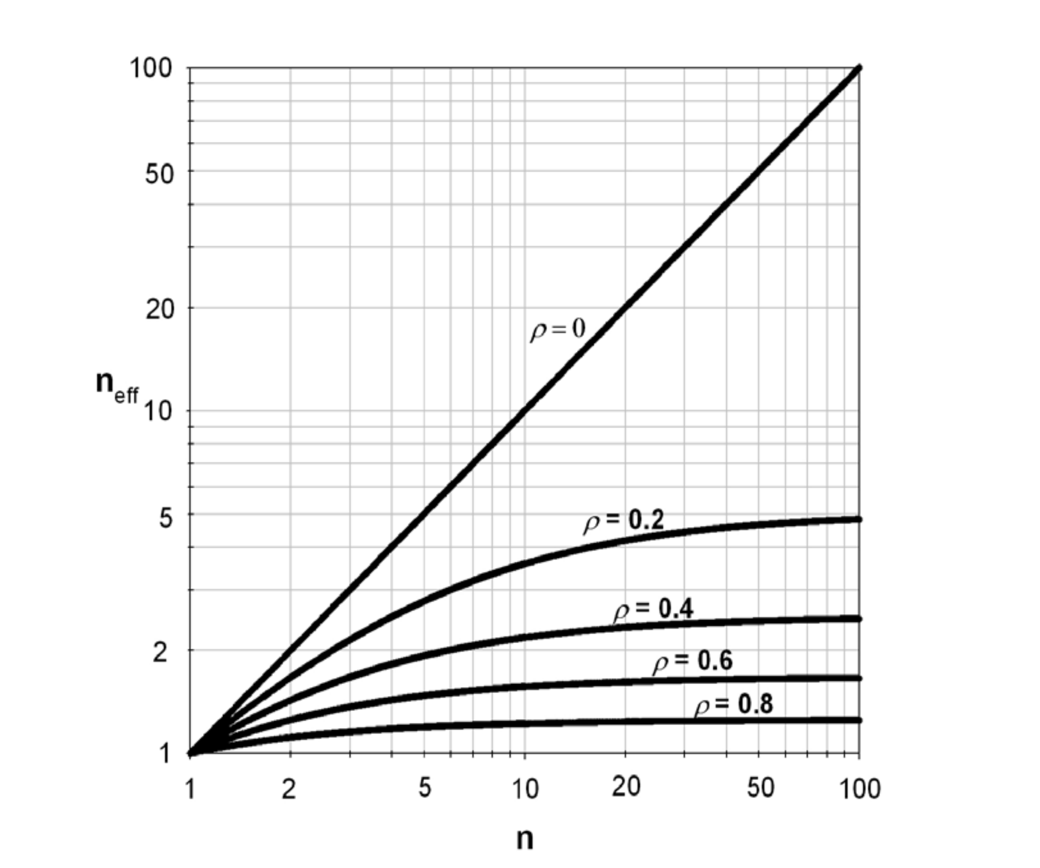
$$n_{effective} = \frac{n_{corr}}{[1+(n_{corr}-1)\rho]}$$

$n_{effective}$ = uncorrelated (independent) samples

n_{corr} = correlated (dependent) samples

where ρ is autocorrelation

with $0 \le \rho \le 1$.



Multiple Measurements on Same Subject

Focus is on modeling

Small-Scale Variability

when there is dependence or correlation among observed data values.

Correlation

implies

$\mathbf{\Sigma}_{nxn}$ is not diagonal

$$\mathbf{\Sigma}_{nxn} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{21} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}_{nxn}$$

Primary Goal of Applied Statistics

Use observed Y values

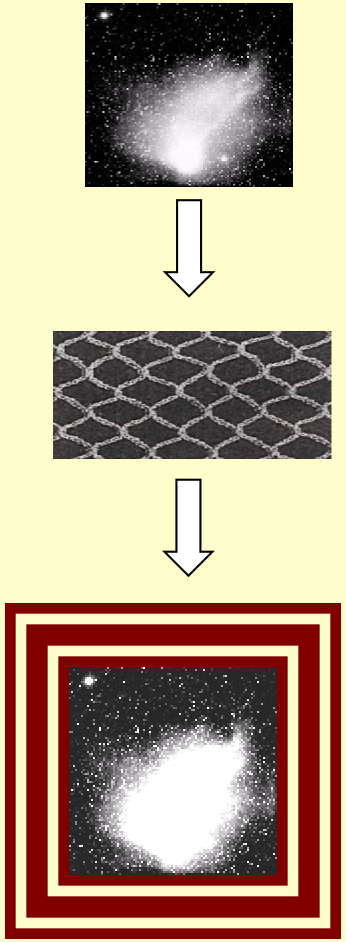
together with scientific knowledge

to create a statistical model

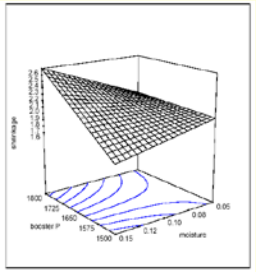
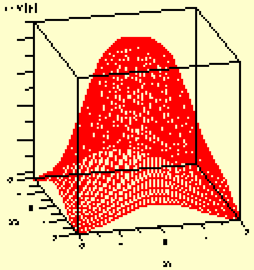
and obtain accurate predictions ($\hat{Y}$)

of unobserved Y values

$\hat{Y}$ =



To Predict Y: First Model “Large-Scale” Trends

$\mathbf{Y}$ =   + ε^*

$\hat{\mathbf{Y}}$

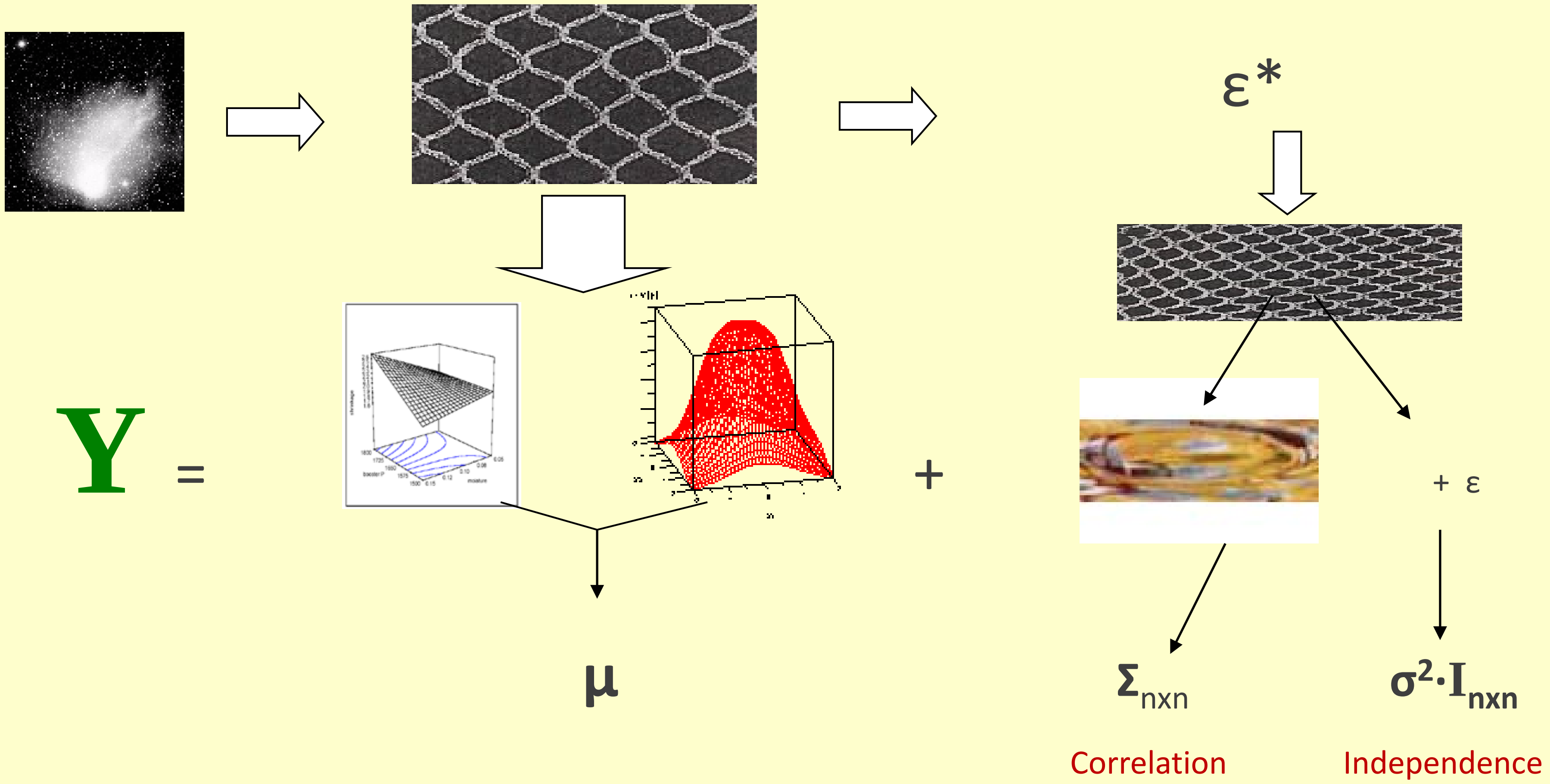
where

$\mathbf{Y}$ is predicted by fitting a ‘large-scale’ trend to the observed data.

ε^* is data variability remaining after the model is fit.

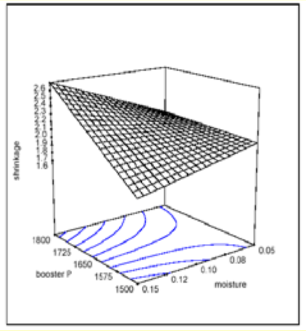
Refine the Model to Predict Y

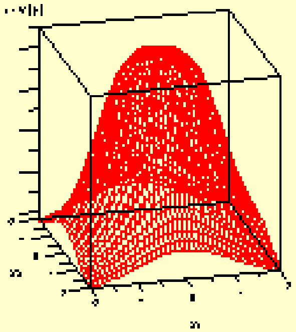
Model ‘Small-Scale’ Variability



Decomposing the Data Variability
ANOVA Terminology

Large-Scale





Small-Scale

All ‘residual’ variation

(eg., a raindrop on water surface)

Fixed Effects

Means
- Deterministic Functions
- Regressors(Covariates)
- Treatments

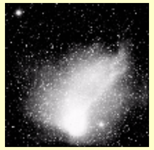
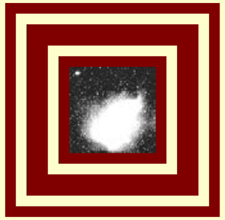
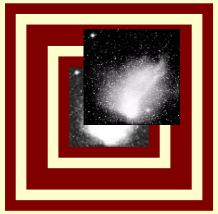
Random Effects

Variance Components
- Variances
- Covariances/**Correlations**

Various Ways to Write
Components of the General Linear Model (GLM)

Y	=	Large-Scale Variation	+	Small-Scale Variation
Y	=	Fixed Effects	+	Random Effects
Y	=	Mean &/or Covariates	+	Variances & Covariances
Y_{nx1}	=	$X_{nxp} \cdot \beta_{px1}$	+	ε_{nx1}
Y_{nx1}	=	μ_{nx1}	+	ε_{nx1}
Y_{nx1}	=	$\hat{y}_{nx1}$	+	ε_{nx1}

The General Linear Model (GLM)
Assumptions – the i.i.d. Mantra

Observed Data	-	Model Prediction	=	Model Error
$\mathbf{y}_{nx1}$	-	$\hat{\mathbf{y}}_{nx1}$	=	$\boldsymbol{\varepsilon}_{nx1}$

Classical GLM assumptions: ε_i are ***i.i.d.***

$\boldsymbol{\varepsilon}_i \sim$ Normal (0, σ^2_ε)

- **i**ndependent(No correlation among the n data values)
- **i**dentically **d**istributed

What is the "**Covariance Structure**" for your Specific Model?

It is seldom true that each different pair of repeated measurements has a different covariance than any other pair.

Typically, there is a simpler "pattern" of covariance.

$$\boldsymbol{\Sigma}_{nxn} = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_\varepsilon^2 & \cdots & \sigma_{21} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_\varepsilon^2 \end{pmatrix}_{nxn}$$

n measurements on the same subject

What is the "**Covariance Structure**" for your Specific Model?

Repeated measurements on the same subject typically share some covariance

These will not be zero

Modeling this diagonal (independence) "covariance structure" will produce incorrect hypothesis test results

$$\Sigma_{nxn} = \sigma_{\varepsilon}^{2} \cdot I_{nxn} = \begin{pmatrix} \sigma_{\varepsilon}^{2} & 0 & \cdots & 0 \\ 0 & \sigma_{\varepsilon}^{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{\varepsilon}^{2} \end{pmatrix}_{nxn}$$

n measurements on the same subject

Typical Covariance Structures
for Multiple Measurements on Same Subject

All pairs of measurements are equally correlated

Compound Symmetry

In SAS: Type= ***CS***

$$\Sigma_{4x4} = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_\varepsilon^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_\varepsilon^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma_\varepsilon^2 + \sigma_1 \end{pmatrix}_{4x4}$$

4 measurements on the same subject

Typical Covariance Structures
for Multiple Measurements on Same Subject

Covariance between pairs of measurements

is a function of their distance (in time)

<u>Equi-Distant</u> Times: 1 2 3 4

<u>Toeplitz</u>

In SAS: Type= **TOEP**

$$\Sigma_{4x4} = \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_1 & \sigma_2 & \sigma_3 \\ \sigma_1 & \sigma_\varepsilon^2 & \sigma_1 & \sigma_2 \\ \sigma_2 & \sigma_1 & \sigma_\varepsilon^2 & \sigma_1 \\ \sigma_3 & \sigma_2 & \sigma_1 & \sigma_\varepsilon^2 \end{pmatrix}_{4x4}$$

4 measurements on the same subject

Typical Covariance Structures
for Multiple Measurements on Same Subject

Covariance between pairs of measurements

is a ***specific [ρ] function*** of their distance (in time)

Equi-Distant Times: 1 2 3 4

1st-Order Auto-Regressive

In SAS: Type= ***AR(1)***

$$\Sigma_{4x4} = \begin{pmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon^2 & \rho^2\sigma_\varepsilon^2 & \rho^3\sigma_\varepsilon^2 \\ \rho\sigma_\varepsilon^2 & \sigma_\varepsilon^2 & \rho\sigma_\varepsilon^2 & \rho^2\sigma_\varepsilon^2 \\ \rho^2\sigma_\varepsilon^2 & \rho\sigma_\varepsilon^2 & \sigma_\varepsilon^2 & \rho\sigma_\varepsilon^2 \\ \rho^3\sigma_\varepsilon^2 & \rho^2\sigma_\varepsilon^2 & \rho\sigma_\varepsilon^2 & \sigma_\varepsilon^2 \end{pmatrix}_{4x4}$$

4 measurements on the same subject

Typical Covariance Structures

for Multiple Measurements on Same Subject

When time measurements

NOT EQUALLY SPACED use

1st-Order Ante-Dependence Type=ANTE(1)

or

Spatial Exponential Type=SP(EXP)

Typical Covariance Structures

for Multiple Measurements on Same Subject

When variances are *heterogeneous*

(*i.e., different magnitudes*) across times

SAS Proc MIXED

Covariance Structures

include:

Type= CSH, TOEPH, ARH(1)

$$\Sigma_{4x4} = \begin{pmatrix} \sigma_1^2 & ? & ? & ? \\ ? & \sigma_2^2 & ? & ? \\ ? & ? & \sigma_3^2 & ? \\ ? & ? & ? & \sigma_4^2 \end{pmatrix}_{4x4}$$

How Do I Check that the Chosen Covariance Structure Fits my Data Well?

1) **Likelihood Ratio Test** – significance indicates chosen Covariance Structure fits data better than "independence" (i.e., diagonal)

2) **Smallest value for AICC fit statistic**

   –indicates "best" fit

Each *Cow* receives 1 of 4 Treatments and is Measured on Days 3, 6, 9 and 21

Experiment Layout:

		(IA=0, SC=20)				(IA=20, SC=20)				(IA=20, SC=0)				(IA=0, SC=0)			
Whole-Plot Factor(**Cow**):		**1**	**8**	**10**	**15**	**2**	**7**	**9**	**16**	**3**	**5**	**12**	**14**	**4**	**6**	**11**	**13**
Subplot Factor (**Days**)	**3**	258	192	234	256	233	152	186	221	260	228	224	197	269	196	202	212
	6	251	185	233	249	236	148	186	219	255	221	217	196	258	191	202	212
	9	245	183	228	237	232	144	185	225	245	221	209	190	249	181	202	212
	21	242	181	219	247	219	139	167	201	249	214	201	281	253	195	189	219

Covariance Structure [General Notation]

for 4 Repeated Measurements (Days 3, 6, 9, 21) on each Cow (k)

$$R_k = \begin{bmatrix} \sigma^2_{cow} + \sigma^2_{day3(cow)} & \sigma^2_{cow} + \sigma_{(day3,day6)(cow)} & \sigma^2_{cow} + \sigma_{(day3,day9)(cow)} & \sigma^2_{cow} + \sigma_{(day3,day21)(cow)} \\ \sigma^2_{cow} + \sigma_{(day6,day3)(cow)} & \sigma^2_{cow} + \sigma^2_{day6(cow)} & \sigma^2_{cow} + \sigma_{(day6,day9)(cow)} & \sigma^2_{cow} + \sigma_{(day6,day21)(cow)} \\ \sigma^2_{cow} + \sigma_{(day9,day3)(cow)} & \sigma^2_{cow} + \sigma_{(day9,day6)(cow)} & \sigma^2_{cow} + \sigma^2_{day9(cow)} & \sigma^2_{cow} + \sigma_{(day9,day21)(cow)} \\ \sigma^2_{cow} + \sigma_{(day21,day3)(cow)} & \sigma^2_{cow} + \sigma_{(day21,day6)(cow)} & \sigma^2_{cow} + \sigma_{(day21,day9)(cow)} & \sigma^2_{cow} + \sigma^2_{day21(cow)} \end{bmatrix}$$

Goodness-of-Fit Summary
Which Covariance Structure Fit the Data Best?

Chosen	Type= Proc Mixed Repeated Statement	Covariance Structure	AICC Fit Statistic	# of Covariance Parameters Estimated
	un	Unstructured	407.2	10
	vc or simple	Variance Components (Independence)	492.2	1
	cs	Compound Symmetry	444.4	2
	csh	Heterogeneous Compound Symmetry	449.6	5
	ante(1)	1^{st}-Order Ante-Dependence	401.5	7
√	sp(exp)	Spatial Exponential	411.5	2

Two Covariance Parameters Estimated by SP(exp)

Cov Parm	Subject	Estimate
$\sigma^2_{sp(exp)(Day)}$	**cow(ia*sc)**	**92.88**
σ^2_{Resid}		**1033.64**

The values in each element of a cow's covariance matrix is calculated by plugging the above 2 estimates into the below formula.

$$\mathbf{Cov}\,(\,\mathbf{Day}_i\,,\,\mathbf{Day}_j\,) = \sigma^2_{\mathbf{Resid}} \cdot (\,\mathbf{exp}[-\mathbf{D}(i\,,\,j)\,/\sigma^2_{\mathbf{sp(Exp)(Day)}}\,]\,)$$

where $D(\,i, j\,) = |\,i - j\,|$ and $i\,, j = 3, 6, 9$, or 21

Estimated SP(exp) Covariance Structure

for each Cow

$$\mathbf{R}_{k\ 4x4} = \mathbf{C}_{4x4} = \begin{array}{c|cccc} & \text{Day 3} & \text{Day 6} & \text{Day 9} & \text{Day 21} \\ \begin{matrix}\text{Day 3} \\ \text{Day 6} \\ \text{Day 9} \\ \text{Day 21}\end{matrix} & 1033.64 \cdot \left[\begin{matrix} 1 & e^{-3/92.88} & e^{-6/92.88} & e^{-18/92.88} \\ e^{-3/92.88} & 1 & e^{-3/92.88} & e^{-15/92.88} \\ e^{-6/92.88} & e^{-3/92.88} & 1 & e^{-12/92.88} \\ e^{-18/92.88} & e^{-15/92.88} & e^{-12/92.88} & 1 \end{matrix}\right]_{4x4} \end{array}$$

Note: $\sigma^2_{residual}$= 1033.64 has been factored out.

Exponent: Numerator = Day difference. Denominator = $\sigma^2_{sp(Exp)(Day)}$= 92.88

Estimated SP(exp) Covariance Structure

for each Cow

$$
\begin{array}{ccccc}
\text{Days:} & 3 & 6 & 9 & 21
\end{array}
$$

$$
1033.64 \cdot \begin{bmatrix} 1 & .9682 & .9374 & .8238 \\ .9682 & 1 & .9682 & .8509 \\ .9374 & .9682 & 1 & .8788 \\ .8238 & .8509 & .8788 & 1 \end{bmatrix}_{4x4}
$$

The numeric estimate illustrate how measurements at two times share less covariance when there is greater time between measurements.

The Model's Covariance Matrix has an 4x4 SP(exp) Covariance on the Diagonal for each Cow

$$Var(Y_{64x1})_{64x64} = R_{64x64} = \begin{pmatrix} C_{4x4} & 0_{4x4} & \cdots & 0_{4x4} \\ 0_{4x4} & C_{4x4} & \cdots & 0_{4x4} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{4x4} & 0_{4x4} & \cdots & C_{4x4} \end{pmatrix}_{64x64}$$

Measurements on different cows are independent; indicated by zero covariance.