Statistical Analysis of the Comet Assay using a Mixture of Gamma Distributions

Bryan Shepherd, Bruce Schaalje, and Kim O'Neill

The Comet Assay

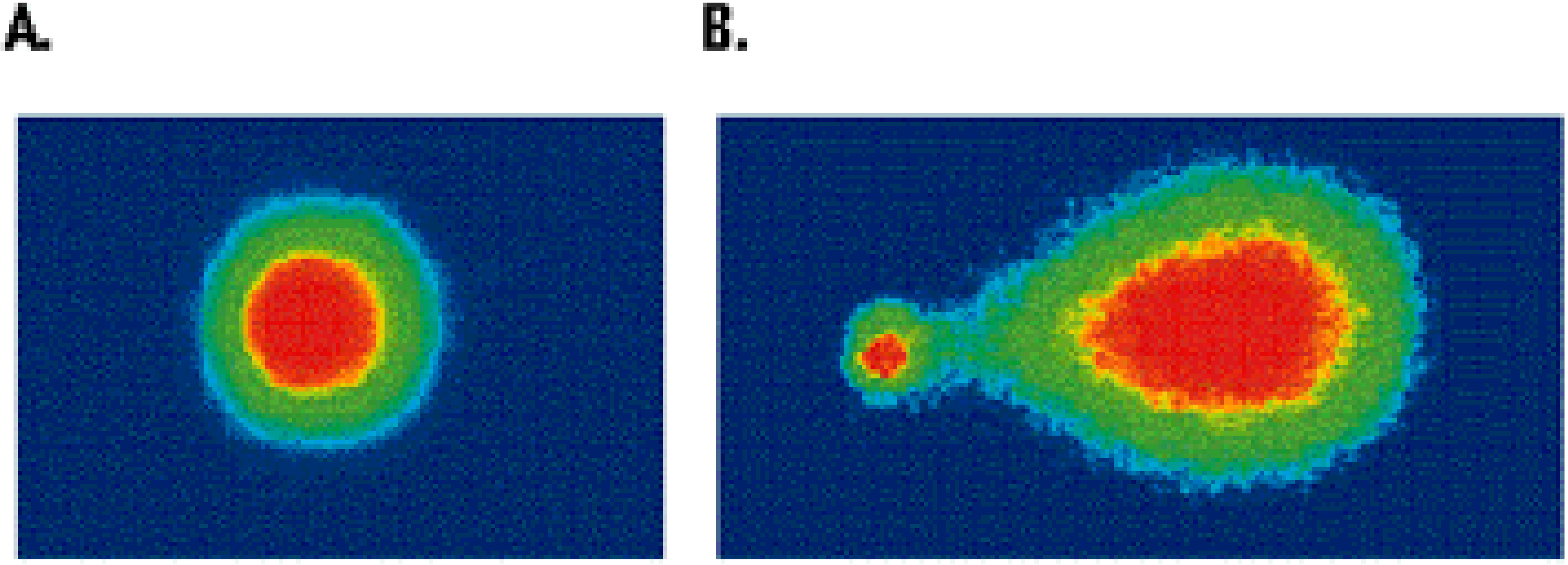
- Sensitive Method of detecting DNA damage
  – Cancer research
    - Identify carcinogens and tumor suppressors
  – Cardiovascular research

  (Fabiani R, *et al*, 2001; Sierens J, *et al*, 2001; Anderson D, *et al*, 2000)

Electrophoresis separates DNA strands from a single cell, showing amount of cellular DNA damage

Undamaged Cell

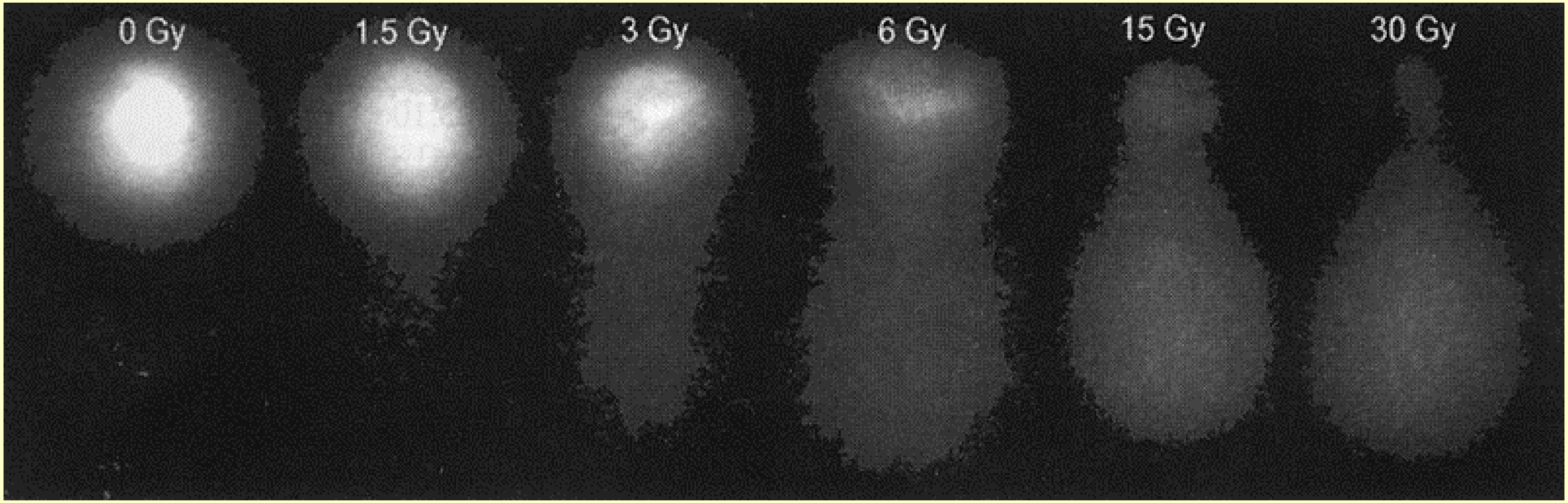
Damaged Cell



Tail Moments

- Numerical Measurement of the DNA damage
- Calculated as the product of the length of the “tail” of DNA trailing the nucleus and the percent of total DNA in the tail
- Higher tail moments indicate greater DNA damage

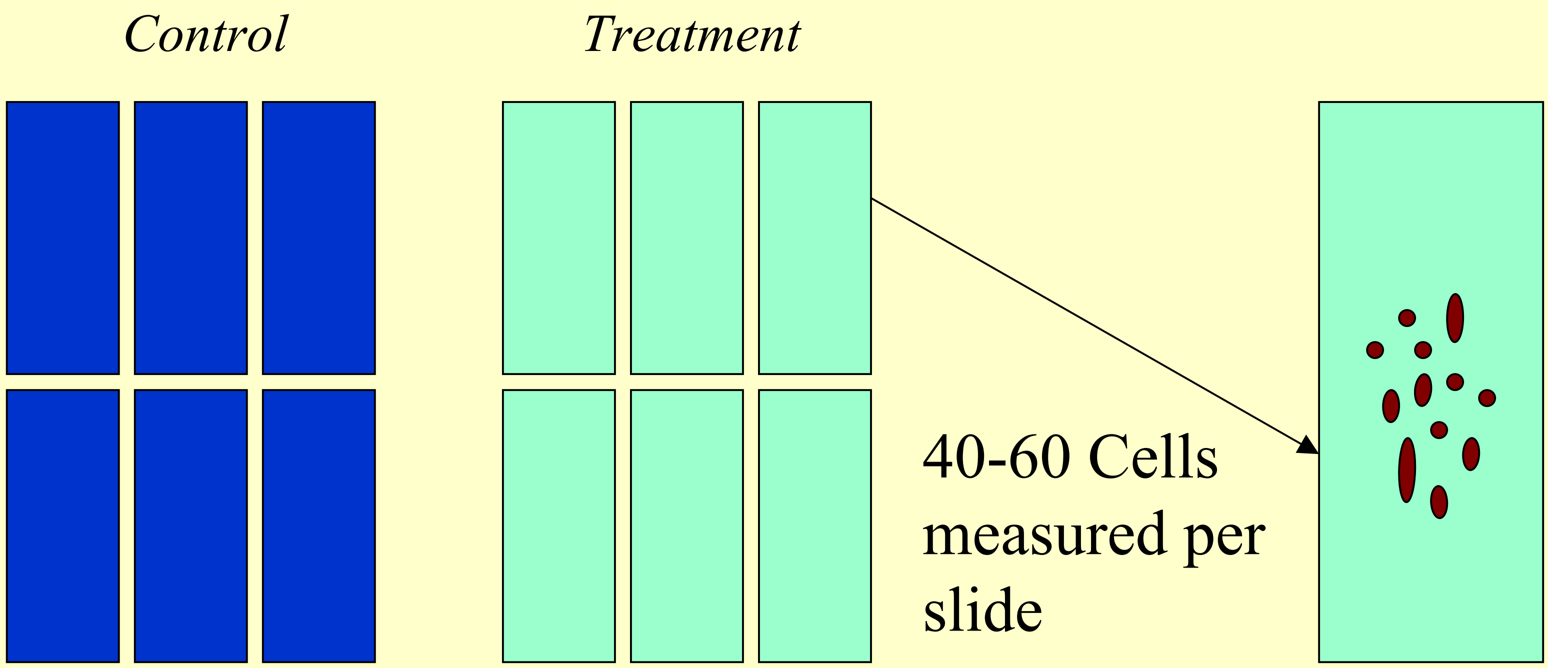
Tail Moments



A Common Experiment

Hypothesis:

Treated cells are less damaged than controls.

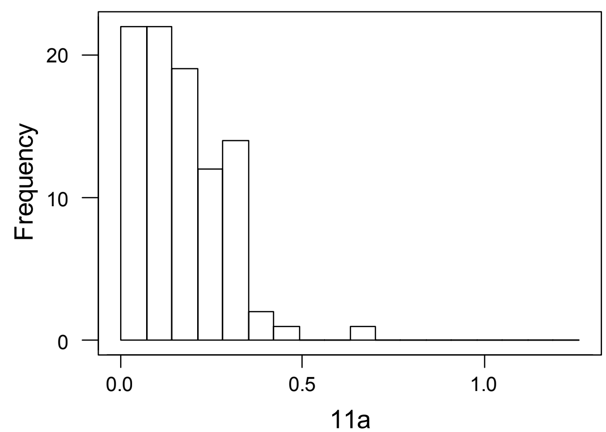


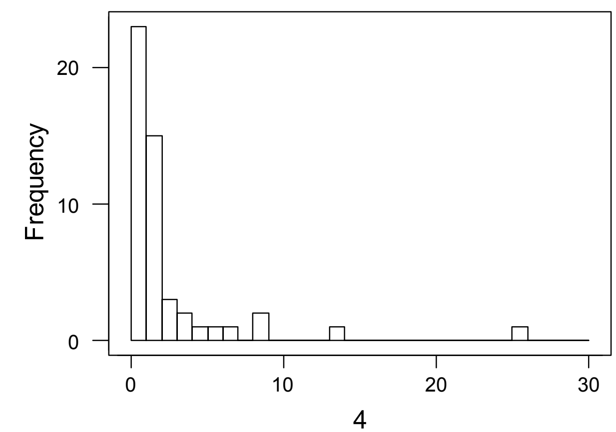
Analysis of Variance Model

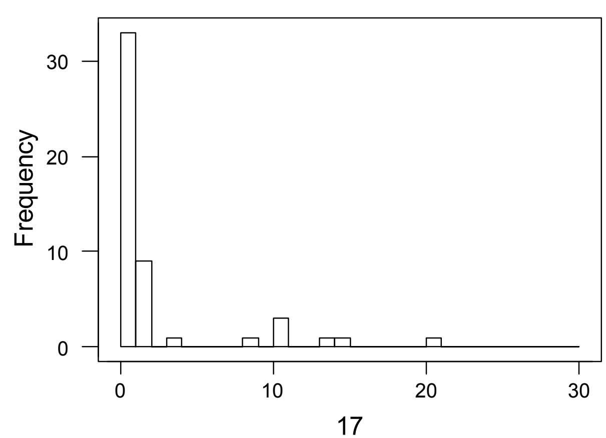
$$Y_{ijk} = \mu + \gamma_i + \varepsilon_{ij} + \delta_{ijk}$$

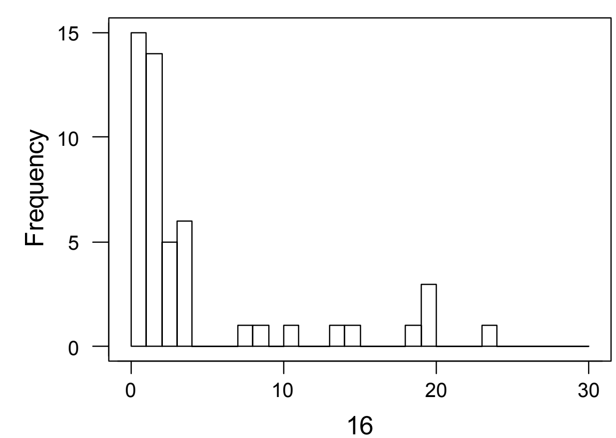
Y_{ijk} = log-transformed tail moment
μ = overall mean
γ_i = main effect of treatment i
ε_{ij} = random error associated with slide j of the i treatment
δ_{ijk} = random error associated with cell k on slide j.

Distribution of Tail Moments on a Slide

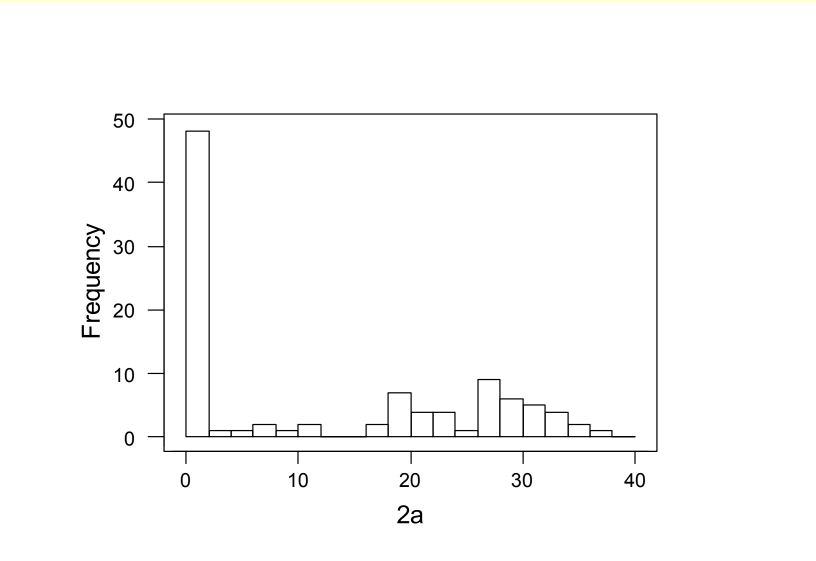


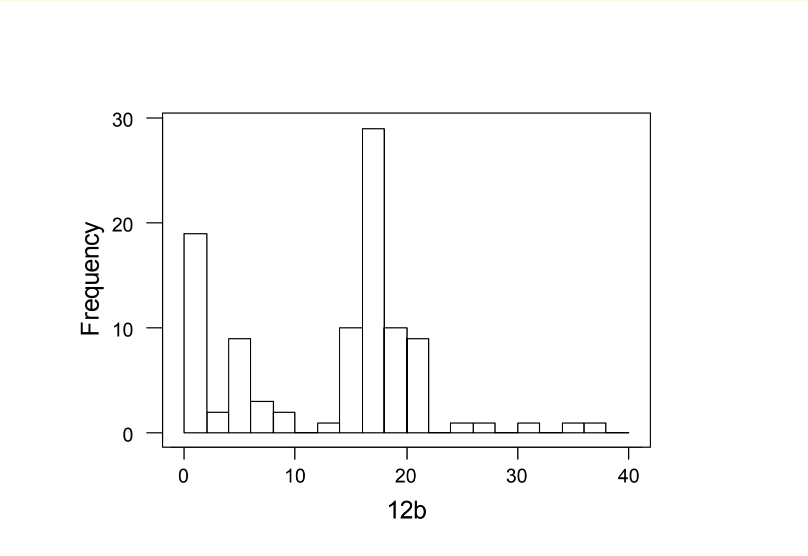


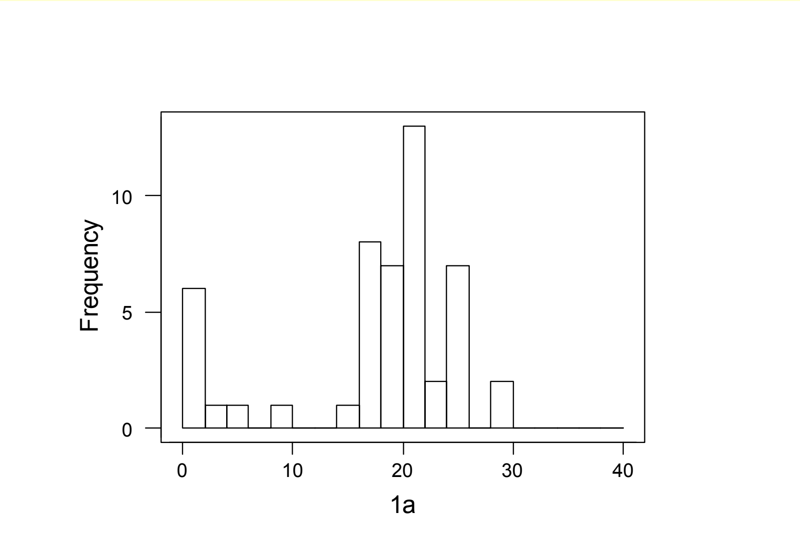




More Tail Moment Distributions – Bimodality?







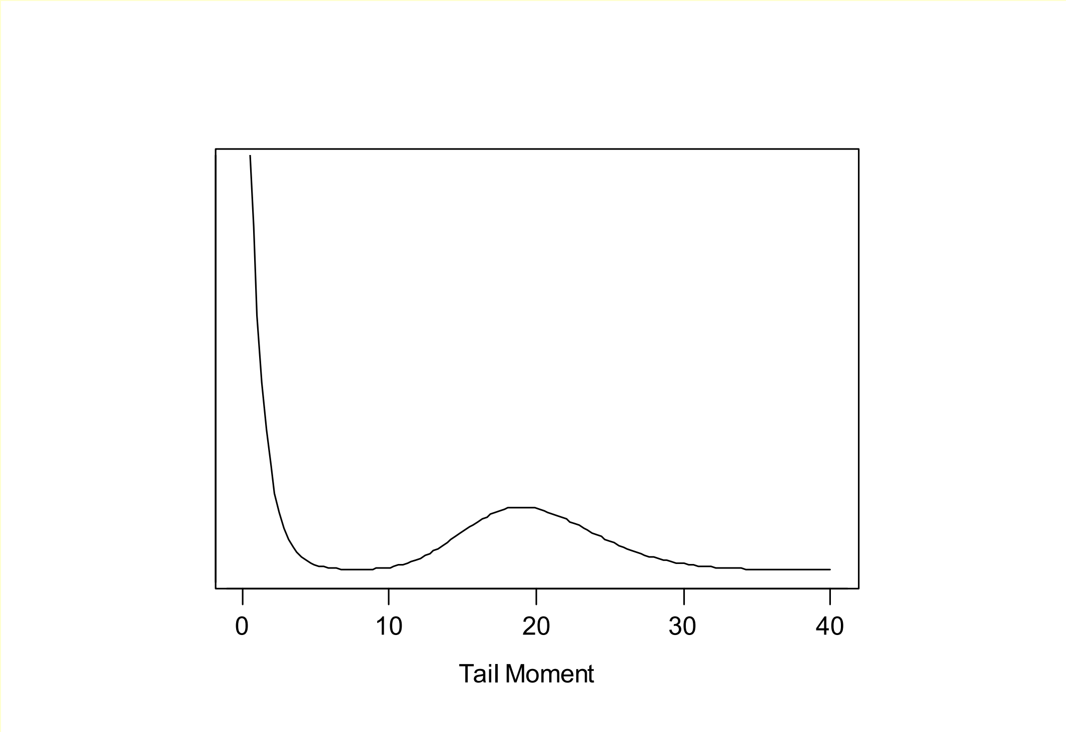
Mixture of Gamma Distributions

$$f(x; p, \kappa_1, \kappa_2, \theta_1, \theta_2) = pf_1 + (1-p)f_2$$

where
$$f_1 = \frac{1}{\theta_1^{\kappa_1}\Gamma(\kappa_1)} x^{\kappa_1 - 1} e^{(-x/\theta_1)}$$

$$f_2 = \frac{1}{\theta_2^{\kappa_2}\Gamma(\kappa_2)} x^{\kappa_2 - 1} e^{(-x/\theta_2)}$$

An example of a density function for a gamma mixture with μ_1=1, σ_1^2=1, μ_2=20, σ_2^2=20, and p=0.5. (where $\mu = \kappa\theta$, $\sigma^2 = \kappa\theta^2$)



One simple method of analysis

- Fit gamma mixture model to every slide in the experiment
  - Take into account the left censored nature of the data
  - Obtain estimated parameters and standard errors.
  - Easily done in SAS PROC NLP
- Perform a weighted analysis of variance on the parameter estimates

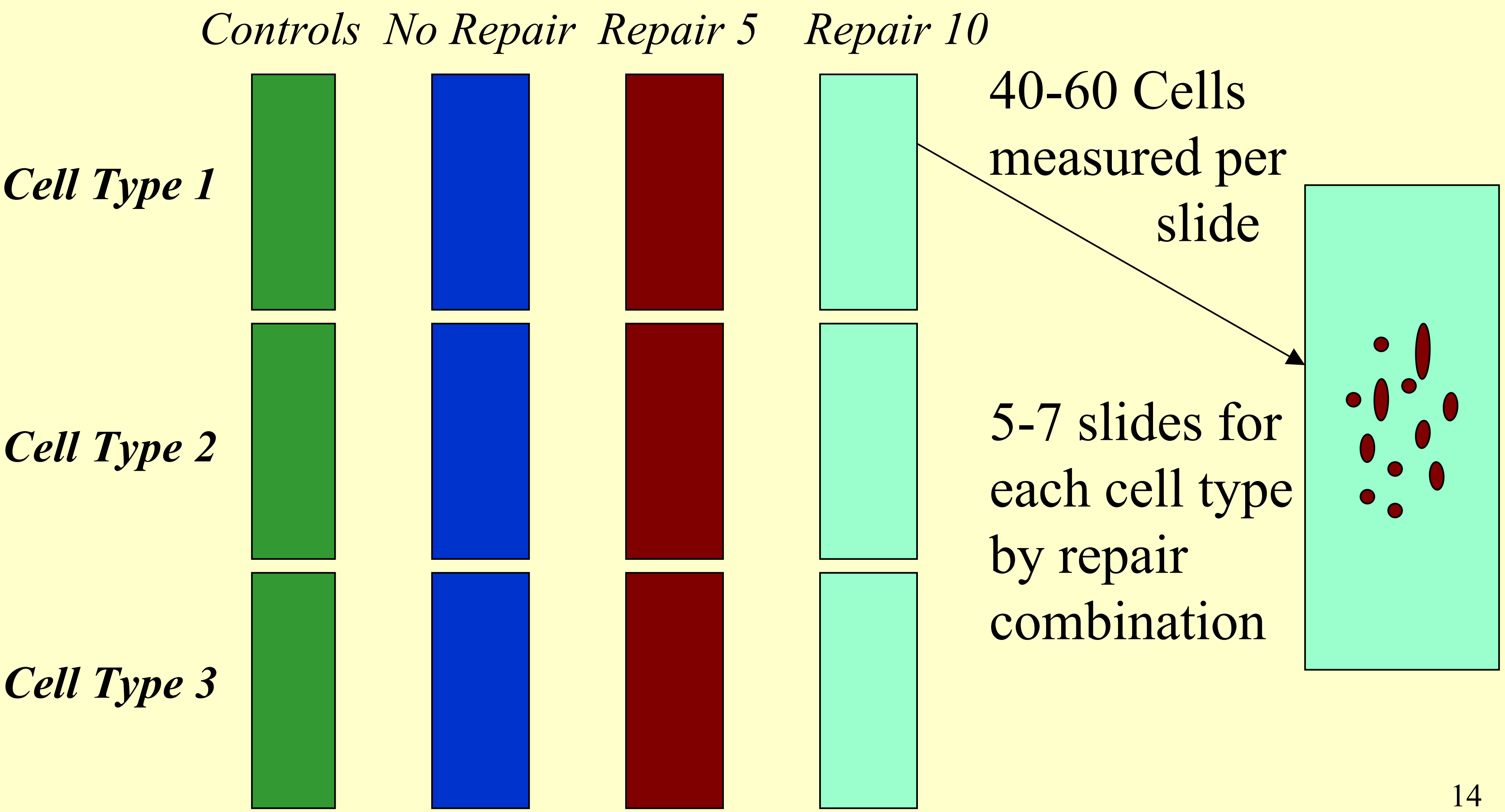
Analysis of Variance Model

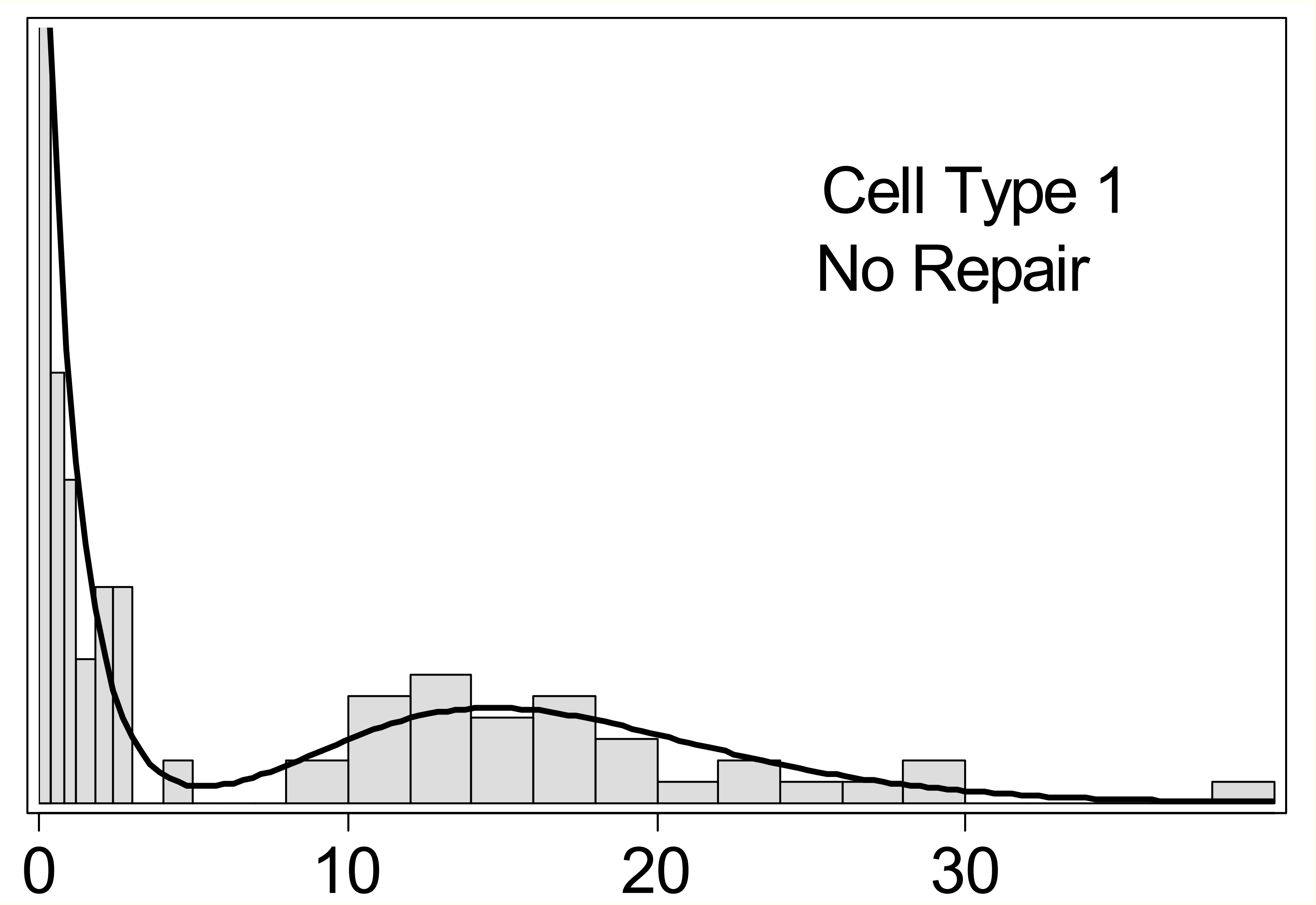
$$Y_{ij} = \mu + \gamma_i + \varepsilon_{ij}$$

Y_{ij} = parameter estimate
μ = overall mean
γ_i = main effect of treatment i
ε_{ij} = random error associated with slide j of the i treatment.

An Example

The TK Experiment




Cell Type 1
No Repair
0
10
20
30

Goodness of Fit Tests
(Number of Slides with Significant Lack of Fit)

Mixture of Gammas	Log-normal	Single Gamma
0	17	32

(Total of 80 slides)

Analysis of Variance Model

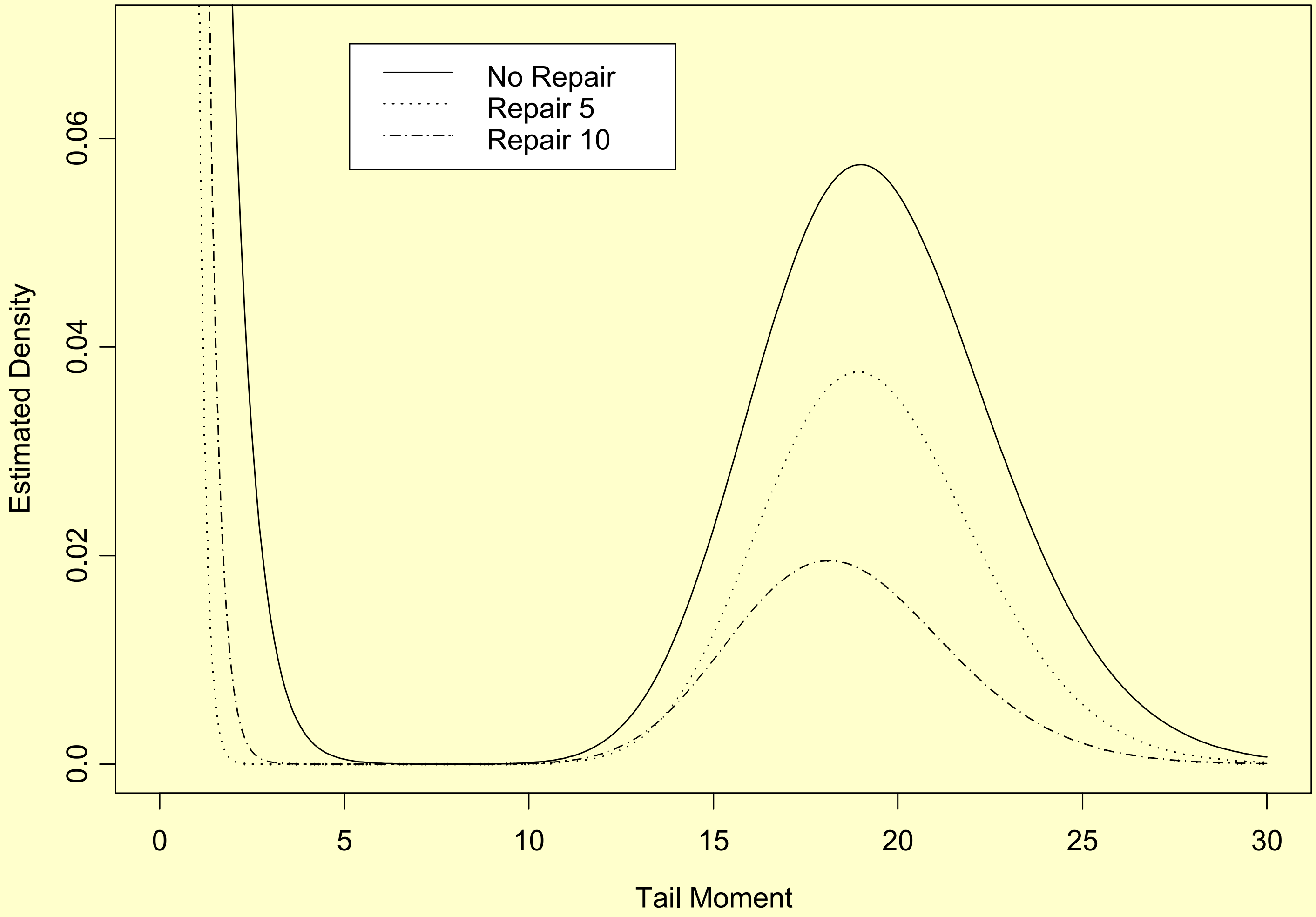
$$Y_{ijk} = \mu + \gamma_i + \rho_j + (\gamma\rho)_{ij} + \varepsilon_{ijk}$$

Y_{ijk} = parameter estimate
μ = overall mean
γ_i = main effect of cell type i
ρ_j = main effect of damage-repair treatment j,
$(\gamma\rho)_{ij}$ = interaction effect of cell type i combined with damage-repair treatment j
ε_{ijk} = random error associated with slide k of the ij treatment.

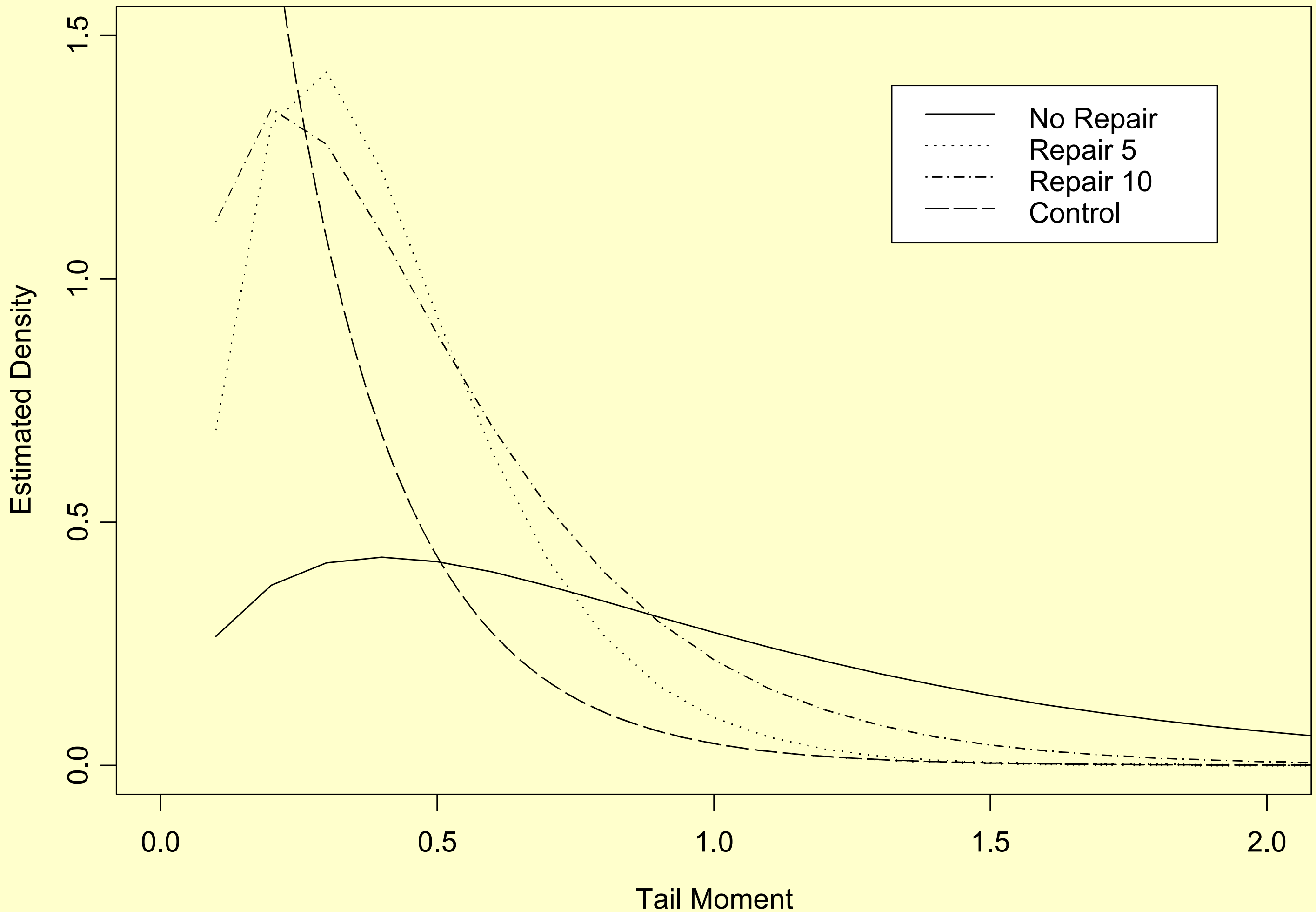
Least squares means (standard errors) of estimates of the parameters p, μ_1, σ_1^2, μ_2, and σ_2^2 at different levels of the damage-repair factor and p-values from tests for equality of parameters.

Damage-repair treatment	Parameter p	μ_1	σ_1^2	μ_2	σ_2^2
Control	0	0.207 (0.028)	0.047 (0.009)	—	—
No repair	0.455 (0.048)	0.924 (0.185)	0.486 (0.165)	19.51 (0.648)	10.19 (2.03)
Repair 5	0.269 (0.042)	0.412 (0.087)	0.058 (0.027)	19.35 (0.759)	8.31 (2.04)
Repair 10	0.141 (0.035)	0.463 (0.090)	0.116 (0.041)	18.59 (0.913)	8.49 (2.70)
P-values					
Controls vs. others	—	0.0001	0.0042	—	—
Among others	0.0001	0.0459	0.0283	0.6905	0.7495

Density Estimates



Density Estimates



Advantages to analysis based on Mixture of Gamma distributions

- More information can be obtained about the nature of the DNA damage and repair at different treatment levels
- Analysis is more sensitive to detect differences between treatment levels
- Analysis complies better to ANOVA assumptions

Disadvantages

- More complicated
- Forcing data into particular fit

Summary

- A mixture model of two gamma distributions has a good fit to the tail moment data
- Analysis can be performed by weighted ANOVA on the parameter estimates