Inference for Loglinear Models

Grinnell College

November 6, 2026

Review: Main Effects

For a two-way independence model,

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y.$$

With the first row and column as reference categories:

- e^{λ} is the expected reference-cell count.
- $e^{\lambda_i^X}$ compares row i with the reference row within a column.
- $e^{\lambda_j^Y}$ compares column j with the reference column within a row.

With interactions, a main effect describes the comparison at the interacting variables' reference levels.

Review: Interactions and Odds Ratios

In a 2×2 table,

$$\log \theta = \log \mu_{11} + \log \mu_{22} - \log \mu_{12} - \log \mu_{21}.$$

The intercept and main effects cancel, leaving

$$\log \theta = \lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY}.$$

With reference-category coding, only λ_{22}^{XY} remains. Exponentiating gives the odds ratio, rather than the odds within a single row.

Review: Types of Association

Highest-order terms	Model
(X, Y, Z)	Mutual independence
(XZ, YZ)	$X \perp Y \mid Z$
(XY, XZ, YZ)	Homogeneous association
(XYZ)	Saturated

- Conditional independence: $OR(XY \mid Z) = 1$ at every level of Z.
- Homogeneous association: each pair's conditional OR is constant across levels of the third variable, but need not be 1.

Conditional and marginal associations need not agree.

Student Substance Use

Dayton, Ohio high-school students reported whether they had ever used alcohol (A), cigarettes (C), or marijuana (M).

		Marijuana Use	
Alcohol Use	Cigarette Use	No	Yes
No	No	279	2
	Yes	43	3
Yes	No	456	44
	Yes	538	911

Last time, we fitted several models to these eight counts.

Models for Substance Use

```
> mutual <- glm(count ~ A + C + M, family = poisson, data = Drugs)
> conditional <- glm(count ~ A * M + C * M,
+                    family = poisson, data = Drugs)
> homogeneous <- glm(count ~ (A + C + M)^2,
+                    family = poisson, data = Drugs)
> saturated <- glm(count ~ A * C * M,
+                  family = poisson, data = Drugs)
```

What restrictions does each model impose on association?

Fitted Counts under Different Models

Observed counts vs fitted counts across the models

Observed	Mutual	Conditional	Homogeneous	Saturated
911	539.98	909.24	910.38	911
538	740.23	438.84	538.62	538
44	282.09	45.76	44.62	44
456	386.70	555.16	455.38	456
3	90.60	4.76	3.62	3
43	124.19	142.16	42.38	43
2	47.33	0.24	1.38	2
279	64.88	179.84	279.62	279

Rows follow the original data-file order. How close are the homogeneous fitted counts to the observations?

Conditional Odds Ratios under Homogeneous Association

```
> pair_terms <- c("Ayes:Cyes", "Ayes:Myes", "Cyes:Myes")
> round(coef(summary(homogeneous))[pair_terms, ], 4)
           Estimate Std. Error z value Pr(>|z|)
Ayes:Cyes    2.0545     0.1741  11.803        0
Ayes:Myes    2.9860     0.4647   6.426        0
Cyes:Myes    2.8479     0.1638  17.382        0
```

- AC: about 7.8 at either level of marijuana use.
- AM: about 19.8 at either level of cigarette use.
- CM: about 17.3 at either level of alcohol use.

How uncertain are these estimated associations?

Profile Likelihood Confidence Intervals

For a candidate log odds ratio b, hold that coefficient fixed and refit all other coefficients:

$$\ell_p(b) = \max_{\text{other coefficients}} \ell(b, \text{other coefficients}).$$

- Compare the best fit at b with the unrestricted best fit.
- Keep values satisfying

$$2\{\ell(\hat{\beta}) - \ell_p(b)\} \leq \chi^2_{1,0.95} \approx 3.84.$$

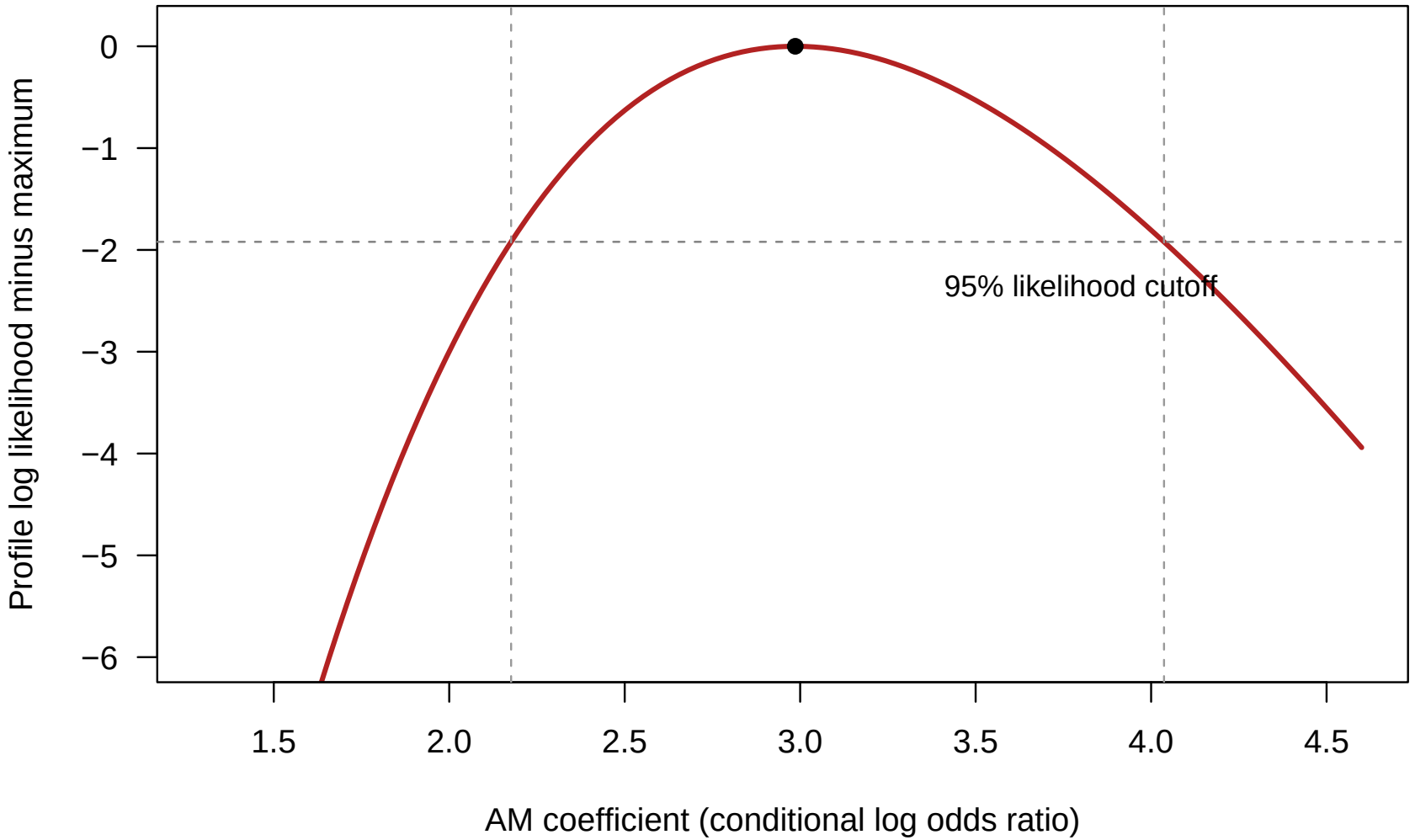
- Equivalently, keep all values where

$$\ell_p(b) - \ell(\hat{\beta}) \leq 3.84/2 = -1.92.$$

- Exponentiate the endpoints to obtain an OR interval.

Profiling the Alcohol–Marijuana Association

Unlike the Wald CI, profile CI doesn't require that our interval be symmetric



Profile Intervals for the Conditional Odds Ratios

```
> profile_ci <- confint(homogeneous, parm = pair_terms)
> round(cbind(OR = exp(coef(homogeneous)[pair_terms]),
+             exp(profile_ci)), 2)
             OR 2.5 % 97.5 %
Ayes:Cyes  7.80  5.60  11.10
Ayes:Myes 19.81  8.81  56.64
Cyes:Myes 17.25 12.65  24.07
```

For AM, the conditional OR is estimated as 19.8, with 95% interval $(8.8, 56.6)$, at either level of cigarette use. All three intervals exclude 1: strong positive associations under the homogeneous association model.

Coefficient Uncertainty and Model Fit

An OR confidence interval describes uncertainty **under the assumed model**.

- Even precise conditional OR estimates do not establish that homogeneous association fits well.
- The intervals do not test every restriction imposed by the model.
- We still need to compare the observed and fitted counts.

Where does each model miss the data, and by how much?

Standardized Residuals

For Poisson cell counts, the Pearson residual is

$$r_i = \frac{n_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}.$$

A standardized Pearson residual also adjusts for fitting:

$$r_i^* = \frac{n_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i(1 - h_{ii})}}.$$

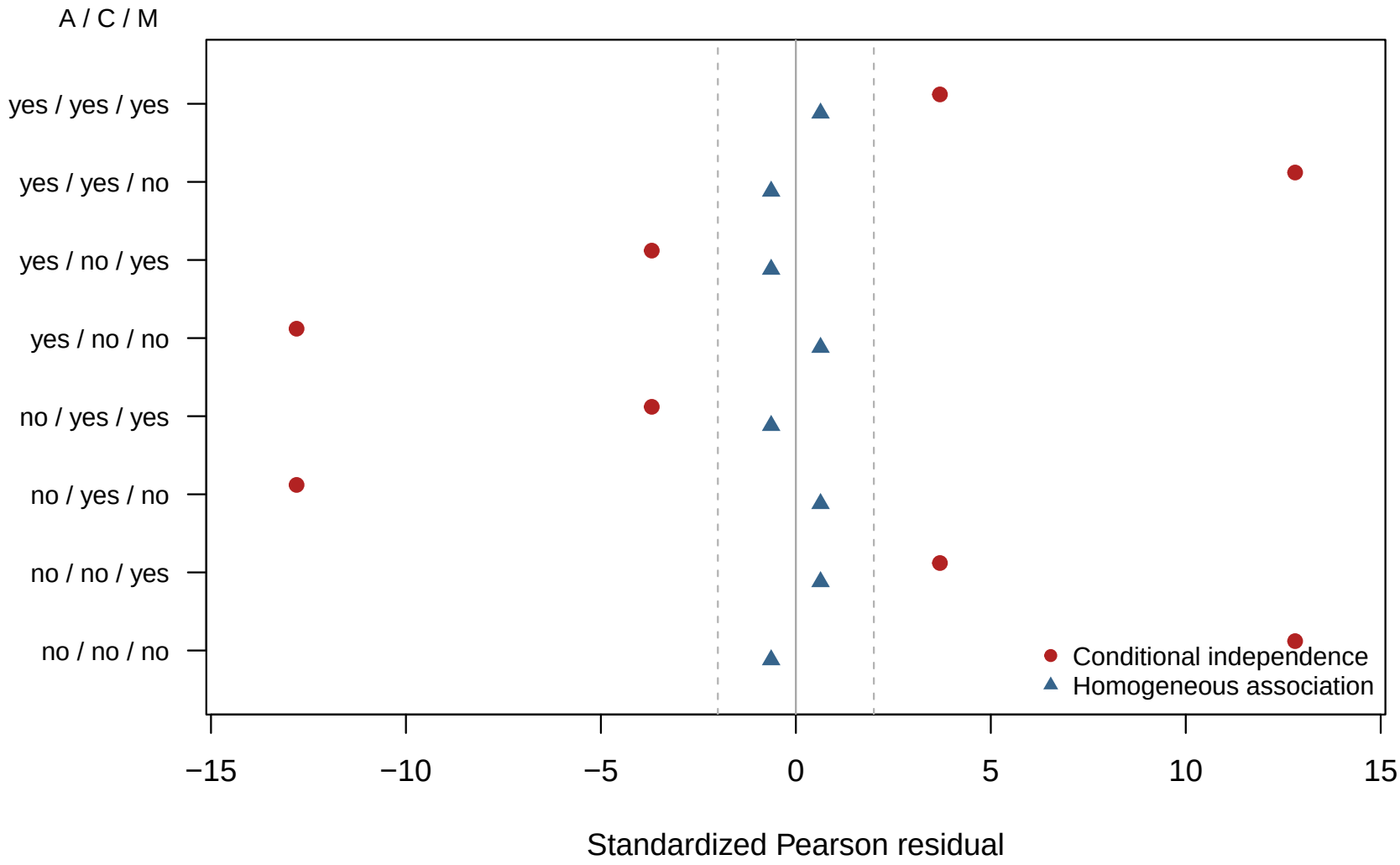
- Positive: more observations than predicted.
- Negative: fewer observations than predicted.
- Magnitudes around 2 or 3 suggest discrepancies to investigate.

Comparing Standardized Residuals

A	C	M	(AM, CM)	(AC, AM, CM)
yes	yes	yes	3.70	0.63
yes	yes	no	12.80	-0.63
yes	no	yes	-3.70	-0.63
yes	no	no	-12.80	0.63
no	yes	yes	-3.70	-0.63
no	yes	no	-12.80	0.63
no	no	yes	3.70	0.63
no	no	no	12.80	-0.63

The conditional-independence model has large discrepancies. The homogeneous model's residuals are all small.

Residual Patterns



Under conditional independence, both-use and neither-use cells have positive residuals; mixed-use cells have negative residuals.

What the Residuals Tell Us

- Within each marijuana category, more students used both alcohol and cigarettes, or neither, than conditional independence predicts.
- Adding AC allows this association within marijuana categories.
- For (AM, CM), the two distinct residual magnitudes correspond to the two strata.
- The homogeneous model has one residual df; its eight residuals have a common magnitude of about 0.63.

Likelihood-Ratio Comparisons

```
> anova(conditional, homogeneous, saturated, test = "LRT")
Analysis of Deviance Table
Model 1: count ~ A * M + C * M
Model 2: count ~ (A + C + M)^2
Model 3: count ~ A * C * M
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         2      187.8
2         1        0.4  1    187.4   <2e-16 ***
3         0        0.0  1      0.4     0.54
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Adding AC improves fit: $187.75 - 0.37 = 187.38$ on 1 df. Adding ACM gives little improvement ($p \approx 0.54$).

How Much Lack of Fit Matters?

Residuals and deviance describe discrepancies relative to sampling variation.

- A small discrepancy can be statistically significant in a large sample.
- A nonsignificant test does not establish exact fit.
- We can also describe the discrepancy on the scale of the table's proportions.

How much of the observed table would have to be redistributed to match the fitted table?

The Dissimilarity Index

For observed counts n_i, fitted counts $\hat{\mu}_i$, and common total n,

$$D = \frac{\sum_i |n_i - \hat{\mu}_i|}{2n} = \frac{1}{2}\sum_i |p_i - \hat{\pi}_i|.$$

- $0 \leq D \leq 1$; smaller values indicate closer proportions.
- Interpretable as the fraction of the table's mass that must move between cells to match the fit.
- Divide by 2 because moving mass removes it from one cell and adds it to another.

Dissimilarity in R

```
> D <- function(fit) {
+   sum(abs(fit$y - fitted(fit))) / (2 * sum(fit$y))
+ }
> round(c(conditional = D(conditional),
+         homogeneous = D(homogeneous)), 5)
conditional homogeneous
    0.08868     0.00108
```

Conditional independence needs about 8.87% of the mass redistributed; homogeneous association needs about 0.108%.

Statistical and Practical Fit for Substance Use

Model	G^2	df	AIC	D
Mutual independence	1286.020	4	1343.06	0.28754
Conditional independence	187.754	2	248.80	0.08868
Homogeneous association	0.374	1	63.42	0.00108
Saturated	0.000	0	65.04	0.00000

Both the inferential comparisons and the proportional discrepancies favor homogeneous association over conditional independence.

A small D can conceal a large relative discrepancy in a rare cell. Use it alongside cell diagnostics and substantive comparisons.

Automobile Accidents and Seat Belts

G = gender, L = location, S = seat-belt use, I = injury. The four binary classifications give 16 cell counts.

			Injury	
Gender	Location	Seat belt	No	Yes
Female	Rural	No	3,246	973
		Yes	6,134	757
	Urban	No	7,287	996
		Yes	11,587	759
Male	Rural	No	6,123	1,084
		Yes	6,693	513
	Urban	No	10,381	812
		Yes	10,969	380

We will focus on model comparisons rather than individual cells.

The Book's Candidate Models

Model	G^2	df	AIC
(G, L, S, I)	2792.77	11	2956.23
All pairwise	23.35	5	198.81
All three-way	1.33	1	184.78
(GLS, GI, LI, SI)	7.46	4	184.92
(GLI, GS, LS, SI)	18.57	4	196.03
(GSI, GL, LS, LI)	22.85	4	200.31
(LSI, GL, GS, GI)	20.63	4	198.09

These candidates are not all nested. The next comparison follows a nested sequence: all pairwise terms, then add GLS, then add the other three-way terms.

Fitting the Accident Models

```
acc_pairs <- glm(count ~ (G + L + S + I)^2,
                 family = poisson, data = Accidents)
acc_GLS <- glm(count ~ G*L*S + G*I + L*I + S*I,
               family = poisson, data = Accidents)
acc_triples <- glm(count ~ (G + L + S + I)^3,
                   family = poisson, data = Accidents)
```

Adding *GLS* allows associations among gender, location, and seat-belt use to vary across the third variable. Injury's pairwise associations remain homogeneous across the other classifications.

Nested Accident-Model Comparisons

```
> anova(acc_pairs, acc_GLS, acc_triples, test = "LRT")
Analysis of Deviance Table
Model 1: count ~ (G + L + S + I)^2
Model 2: count ~ G * L * S + G * I + L * I + S * I
Model 3: count ~ (G + L + S + I)^3
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         5      23.35
2         4       7.46  1    15.89  6.7e-05 ***
3         1       1.33  3     6.14     0.11
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Adding *GLS* reduces deviance by 15.89 on 1 df ($p \approx 0.000067$). Adding the other three-way terms is not significant at 5%.

AIC, Deviance, and Dissimilarity

```
> round(rbind(pairwise = c(G2 = deviance(acc_pairs),
+                          AIC = AIC(acc_pairs), D = D(acc_pairs)),
+             add_GLS = c(G2 = deviance(acc_GLS),
+                          AIC = AIC(acc_GLS), D = D(acc_GLS))), 5)
               G2    AIC       D
pairwise 23.3510 198.81 0.00822
add_GLS   7.4645 184.92 0.00251
```

- AIC decreases by about 13.9 when *GLS* is added.
- Yet D is only 0.822% under the pairwise model and 0.251% after adding *GLS*.

There is a detectable departure from homogeneous association, but both fitted tables are close in overall proportions.

How Much Do the Associations Change?

Consider the gender–location odds ratio:

	Seat belt	
Model	No	Yes
All pairwise	0.81	0.81
Add *GLS*	0.75	0.86

The data detect a difference. Whether it matters requires judging the size of these changes, not just their statistical significance.

A Sample-Size Experiment

To isolate sample size, use an illustrative table whose counts divide evenly by ten:

	$Y =$ No	$Y =$ Yes
$X =$ No	120	180
$X =$ Yes	180	520

Fit independence and saturation after multiplying every count by 0.1, 1, and 10.

Predict what happens to the OR, dissimilarity, deviance, and preference by AIC.

Same Proportions, Different Evidence

Scale	Total	G^2	p	ΔAIC	D
$n/10$	100	1.99	0.159	-0.01	0.060
n	1000	19.86	8.35×10^{-6}	17.86	0.060
$10n$	10000	198.56	$< 10^{-12}$	196.56	0.060

- The observed OR stays at 1.926 and D stays at 0.06.
- The residual df stays at 1, but deviance scales with the total.
- ΔAIC = AIC(independence) $-$ AIC(saturated); positive values favor association.

Compare AIC values within the same dataset, not across sample sizes.

Why the Measures Behave Differently

Scaling every count by k also scales the fitted counts by k:

$$G^2_{\text{new}} = 2\sum_i kn_i \log \frac{kn_i}{k\hat{\mu}_i} = kG^2_{\text{old}},$$

while

$$D_{\text{new}} = \frac{\sum_i |kn_i - k\hat{\mu}_i|}{2kn} = D_{\text{old}}.$$

The number of cells and parameters has not changed. Degrees of freedom do not remove sample-size sensitivity.

What We Learn about the Accident Models

- The extra interaction captures a real pattern in these observed proportions.
- The large sample makes a modest departure statistically detectable.
- AIC supports the extra parameter, while dissimilarity shows how small the overall redistribution is.
- Practical importance depends on the scientific question and changes in the associations of interest.

Review

1. Why do OR confidence intervals not establish model adequacy?
2. What did residuals reveal about (AM, CM)?
3. What does the LRT compare when AC is added?
4. What does dividing by $2n$ accomplish in the dissimilarity index?
5. Why can a model have a significant lack-of-fit test and a small D?