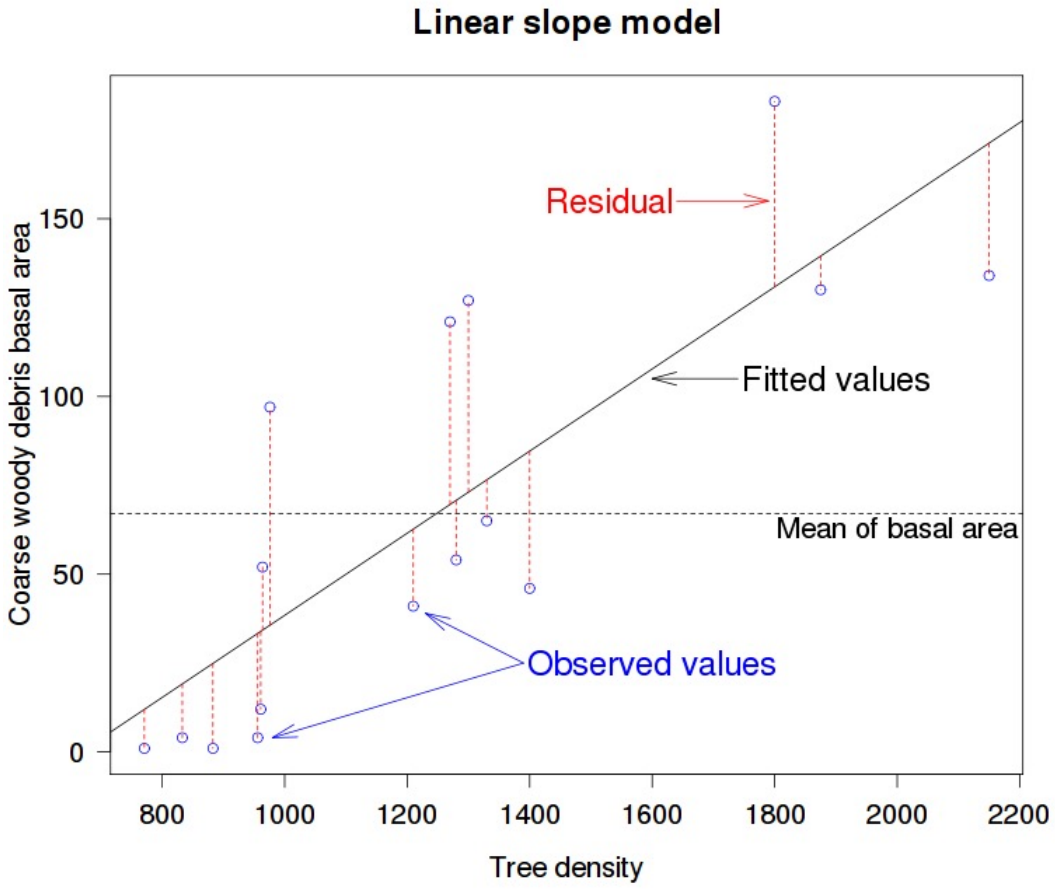
Estimation using least squares

least squares has an arithmetical solution that minimises the sum of squared residual deviations of the observations from the fitted model values



$$\min \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Error (noise; variability)

- classical statistics built around assumption that variability is normally distributed
- but … *normality is in fact rare*
- non-normality is an opportunity to:
    - represent variability in a more realistic way
    - gain insights into the process of interest

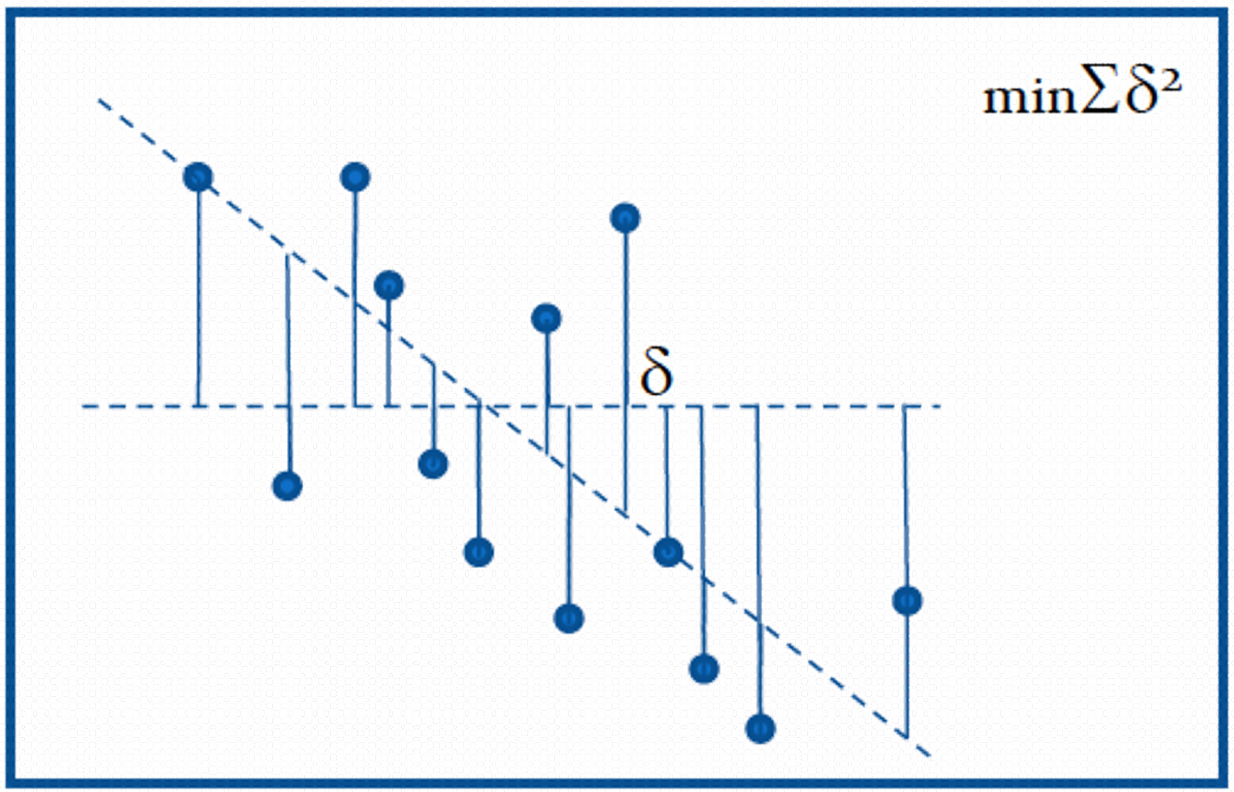
Likelihood

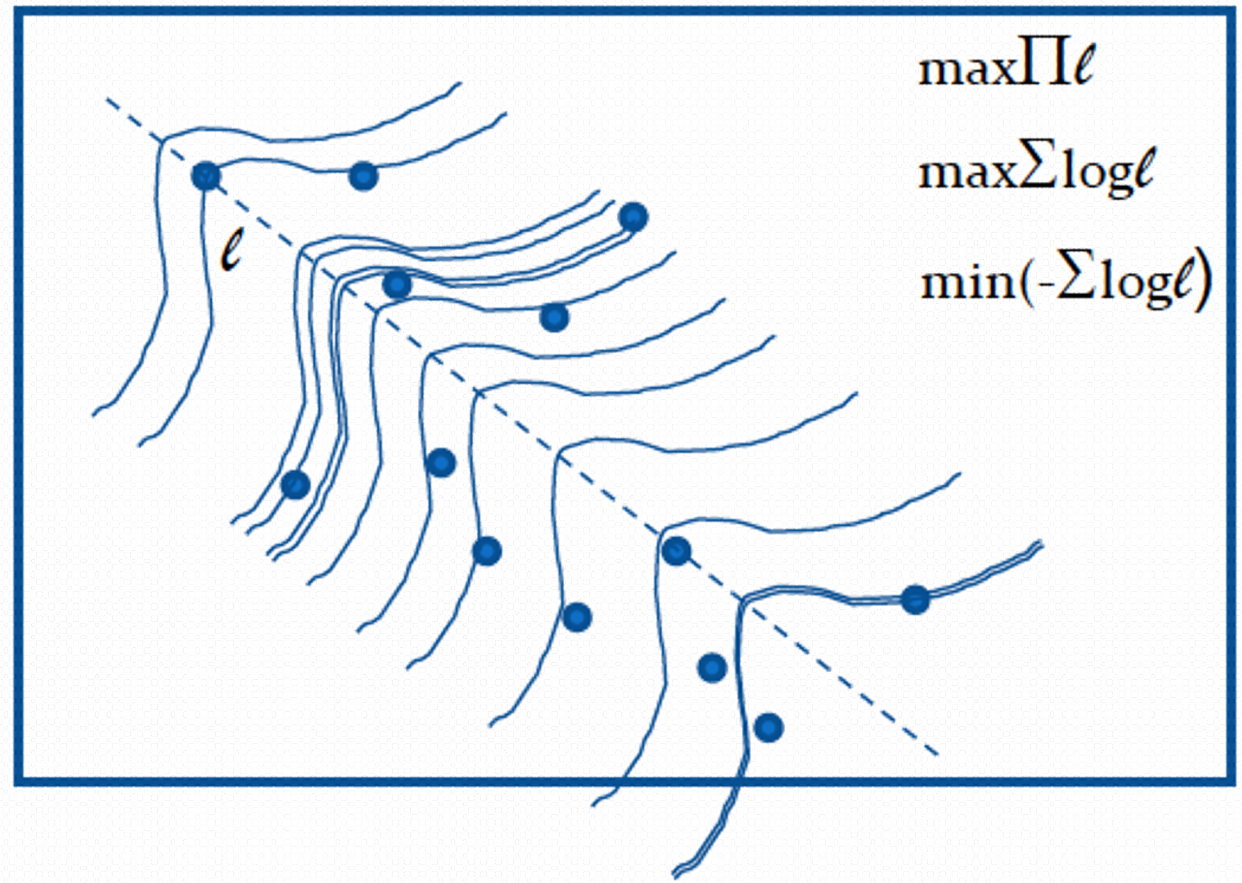
- to define a model:
  - **functions** of systematic effects for deterministic patterns
  - **probability distributions** to describe stochastic patterns
- now need to estimate parameters of models & test models against each other
- estimating parameters: finding values that make model fit data best
- comparing models = which fits best?
- goodness-of-fit metrics based on the *likelihood*

Likelihood

- **likelihood** is a conditional probability
- likelihood = *probability of seeing the data we collected* ***given*** *a particular model*
- likelihood of a population parameter equal to a specific value, given the data, is the probability of obtaining observed data given parameter equals a specific value
- $\mathcal{L}$[parameter ϑ | data] = Pr[data | parameter ϑ]
  - here, a parameter represents a hypothesis, so can write:

Pr[data | H]


$\min\Sigma\delta^2$
δ


$\max\Pi\ell$
$\max\Sigma\log\ell$
$\min(-\Sigma\log\ell)$
ℓ

Neyman-Pearson null-hypothesis testing

Could these observations (or more extreme ones) really have occurred by chance?

Neyman-Pearson null-hypothesis testing

- used primarily for binary detection problem

H_0: $\bar{x}_1 = \bar{x}_2$

H_A: $\bar{x}_1 \neq \bar{x}_2$ or $\bar{x}_1 < \bar{x}_2$ or $\bar{x}_1 > \bar{x}_2$

- fail to reject H_0 OR reject H_0
- probability of false-alarm fixed at some arbitrary value
- probability of rejecting the null hypothesis H_0 when it is true = *Type I error* (false alarm)

Neyman-Pearson null-hypothesis testing

'p' (probability) value:

if H_0 is true, what is probability of observing the metric at least as extreme as the one observed?

'α' (alpha):

threshold probability below which we agree that p is 'significant'

(i.e., H_0 ruled FALSE)

Neyman-Pearson null-hypothesis testing

What's the probability of getting it wrong?

Type II error = probability of making an error when failing to reject H_0 (i.e., H_0 is TRUE)

NHT cannot evaluate this probability!

post hoc methods to determine **POWER** (1 – *Type II error*)

statistical tests

- *t*-test ask the question: given the inherent variability in my samples, does the mean of treatment = mean of control?
- ANOVA (analysis of variance) is similar in principle to a *t*-test, but can test across multiple means by comparing variance within and variance among samples
- regression relates a continuous independent variable (e.g., rainfall) to a dependent one (e.g., growth rate of a plant)
- last 2 are forms of the *general linear model*

Main gripes against NHT

- disconnect between estimates of *Type I* and *II errors*
- conflation of *effect* and *sample* sizes
- ambiguity regarding *how much chance of making an error is acceptable?*
- reduced opportunity for 'control' in biology (high variability)
- considers only 1 dimension of a complex problem

Multiple Working Hypotheses

- multiple causative agents lead to complexity
- single alternatives do not suffice
- ‘ruling hypothesis’ – bound to be subjective
- different ideas (working hypotheses)
- MWH: Thomas C. Chamberlin (1890)

“the dangers of parental affection for a favourite theory can be circumvented”

1. Is this really the full explanation?
2. Are we seeking to establish prematurely the truth of a single factor, when consideration of more than one may be more appropriate?

Multiple Working Hypotheses

- despite attraction, avoids pinning conclusions on a single (simple) explanation
- accounts for cascades (sequences) and emergent properties
- no restriction on number of models to examine (provided *n* large enough)
- models ranked on relative support
- parsimony = (simplest combination of factors providing strongest explanatory power)

Multiple Working Hypotheses

1. define hypotheses to test
2. construct model set from plausible biological mechanisms
3. approach equally suitable to experimental, mensurative or observational studies
4. NOT falsification of all but one model
5. incorporates uncertainty in *model*, not just model parameters

Comparing models

- approximation of **Kullback-Leibler** (K-L) information loss

  states that all models are false because represent incomplete approximations of real (but unreachable) 'truth'

- **Bayesian** (dimension-consistent)

  weight of evidence of one model relative to others; assumes 'true' model included in model set

- **cross-validation**

  jackknife approximation of K-L measures

- **Bayesian inference**

  Bayes theorem – model's posterior distribution based on MCMC simulation (next lecture)

Comparing models

- **Akaike's Information Criterion** (AIC)

  useful for smaller sample sizes (with correction); tends to favour more complex models when tapering effects exist; best for prediction

$$-2\ell + 2k$$

- **Bayesian Information Criterion** (BIC)

  better for larger samples and when main effects are of primary interest; not best for prediction

$$-2\ell + k \log_e n$$

Comparing models

DIC and *w*IC

How to scale and compare models with different IC values?

e.g., 1: ~A+B+C $\ell_1 = -74.625$; $k_1 = 4$

2: ~A+B $\ell_2 = -73.612$; $k_2 = 3$

3: ~A $\ell_3 = -83.010$; $k_2 = 2$

$AIC_1 = 159.248$; $AIC_2 = 153.224$; $AIC_3 = 162.020$

$\Delta AIC_1 = 6.024$; $\Delta AIC_2 = 0.000$; $\Delta AIC_3 = 8.796$

$$\Delta = \mathrm{IC} - \min(\mathrm{IC})$$

$wAIC_1 = 0.046$; $wAIC_2 = 0.942$; $wAIC_3 = 0.012$

$$w = \frac{e^{-\frac{\Delta}{2}}}{\sum e^{-\frac{\Delta}{2}}}$$

Evidence ratio

evidence of one model versus another

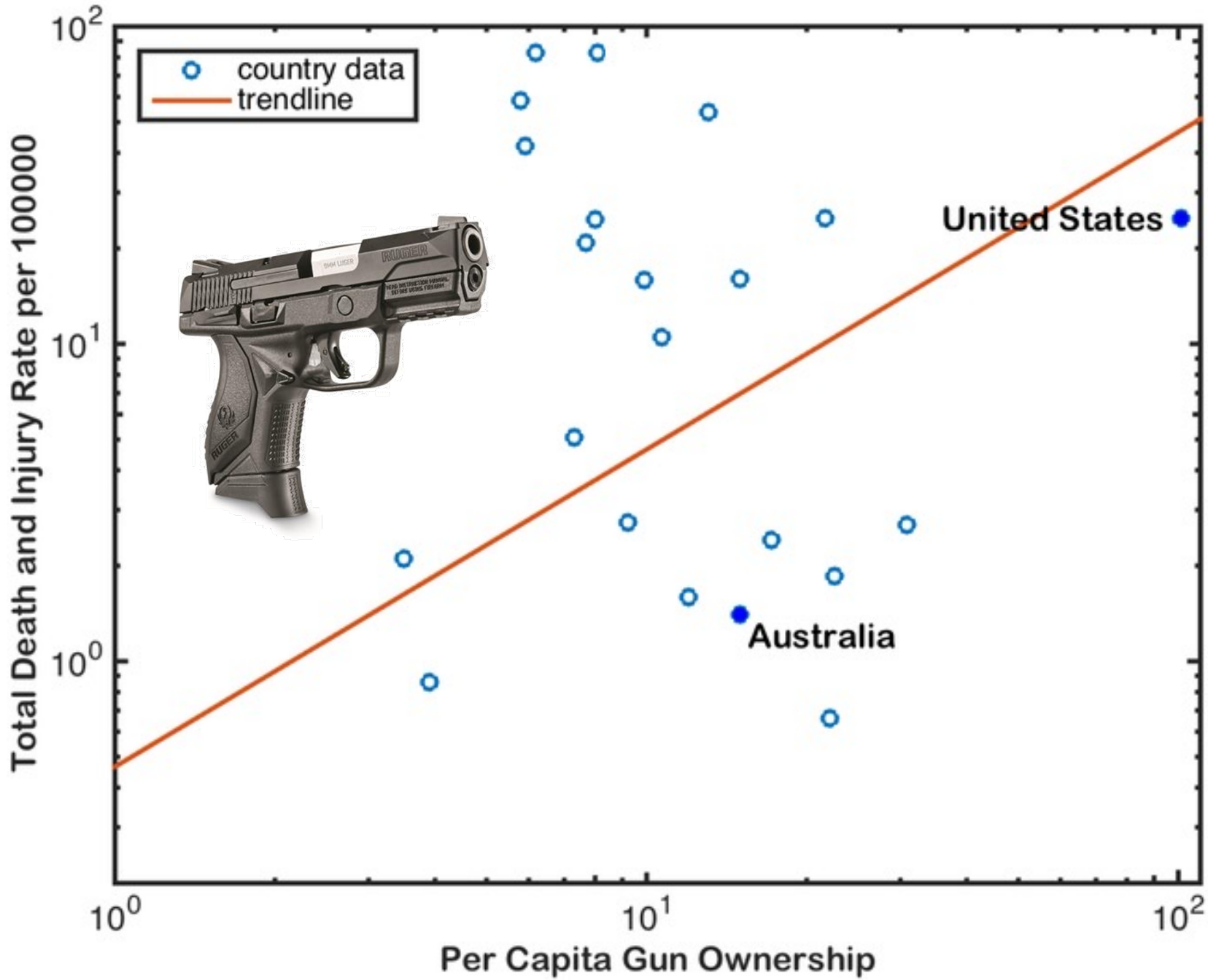
How much more likely, given bias correction, is model 2 than model 1?

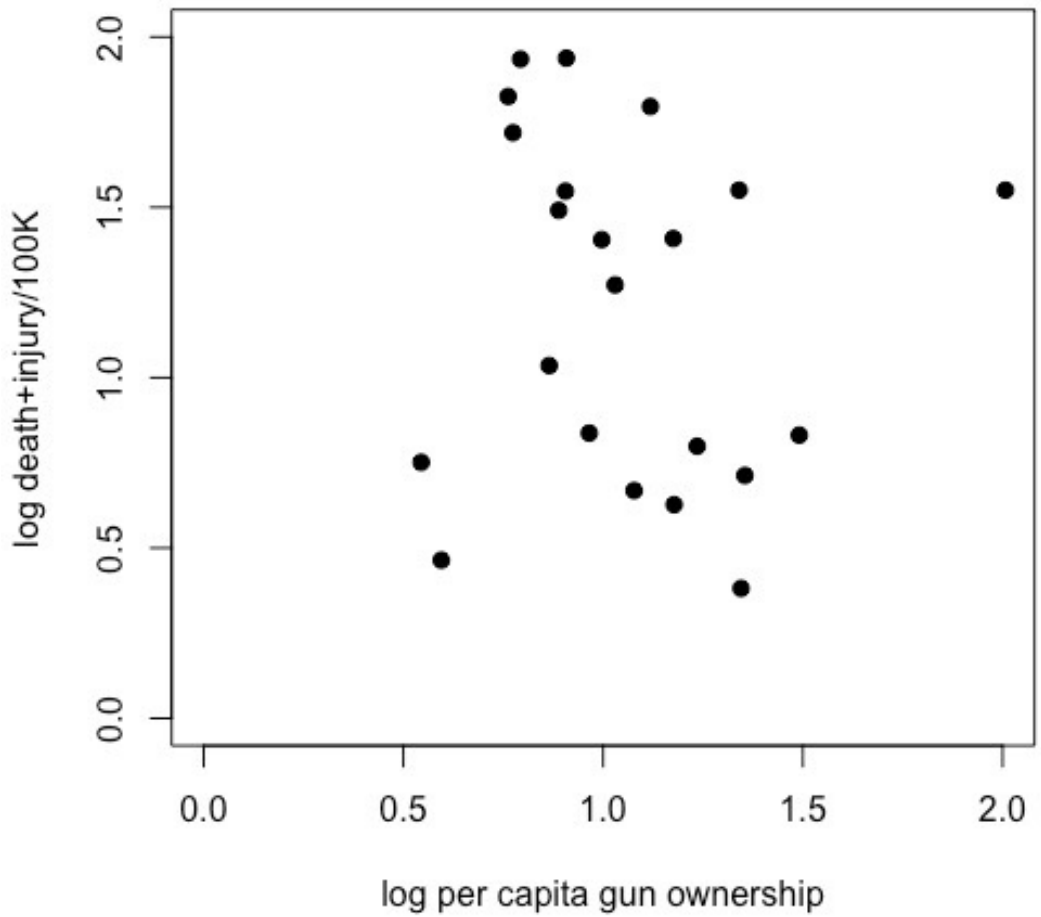
$wAIC_1 = 0.046$; $wAIC_2 = 0.942$; $wAIC_3 = 0.012$

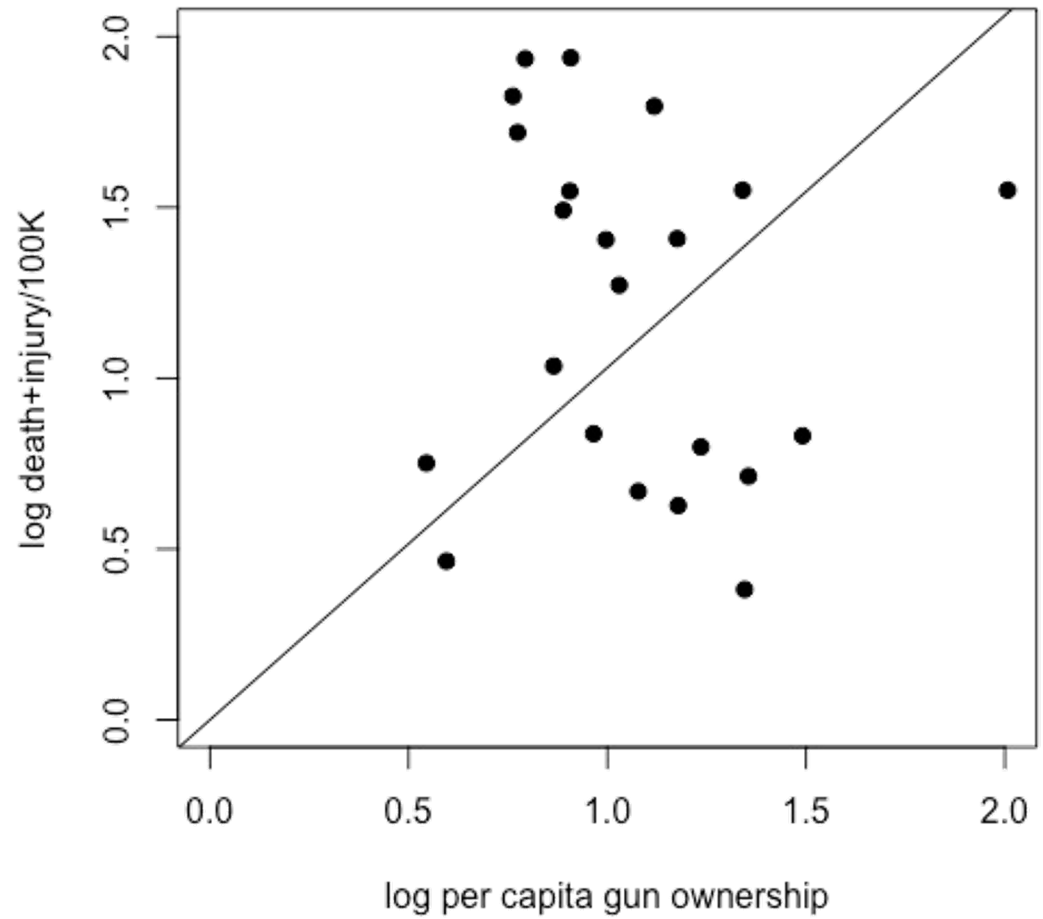
$ER = wAIC_2 \div wAIC_1 = 0.942/0.046 = 20.5$

theconversation.com/gun-control-in-america-by-the-right-and-wrong-numbers-49573

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for
Australian Biodiversity
and Heritage




log death+injury/100K
0.0
0.5
1.0
1.5
2.0
log per capita gun ownership


log death+injury/100K
0.0
0.5
1.0
1.5
2.0
log per capita gun ownership