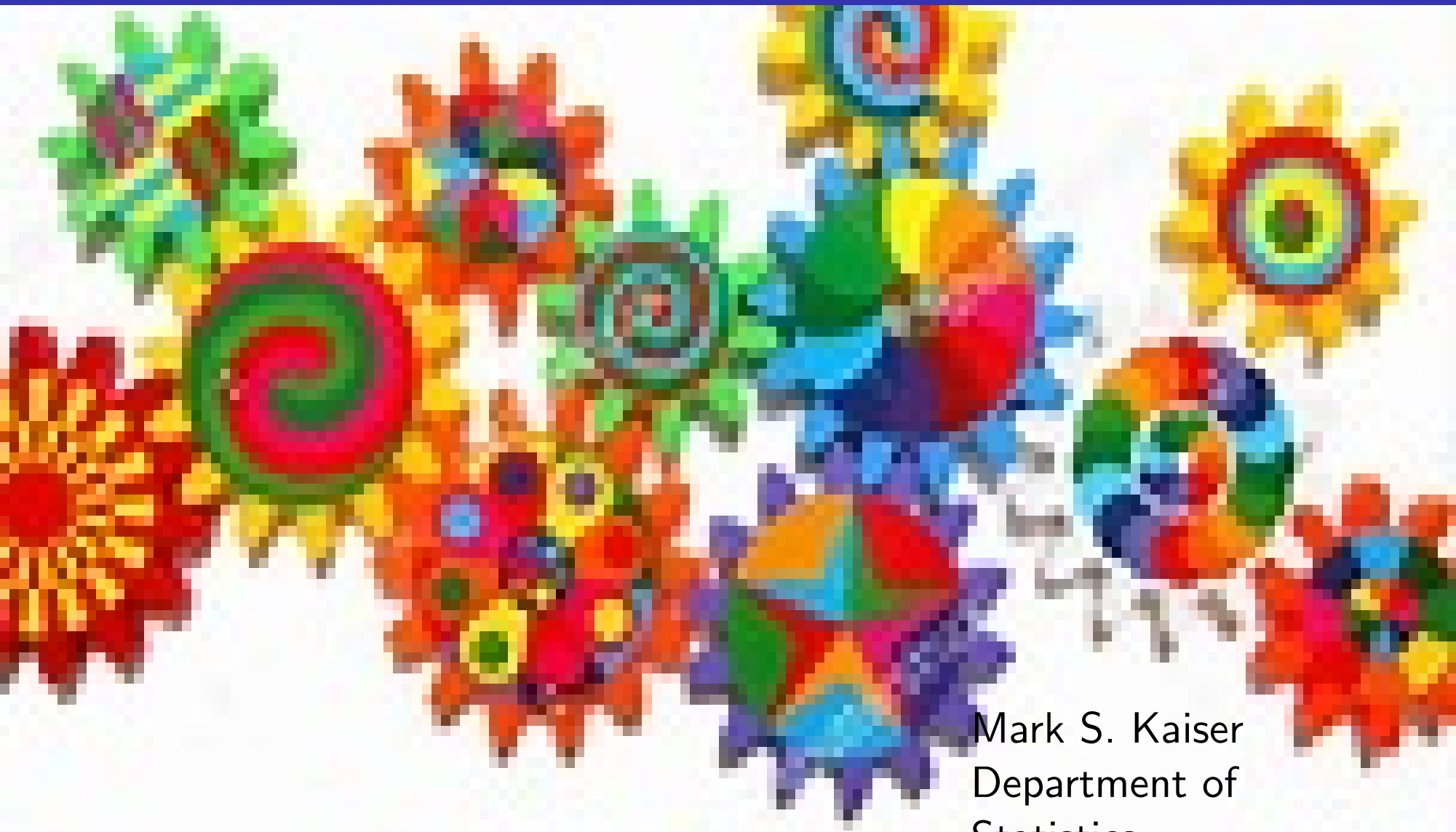
How Do the Pieces Fit Together: A Brief Overview of Meta-Analysis



Mark S. Kaiser
Department of Statistics
Iowa State University

The Broad Context

- Term Meta-analysis coined in 1976 by G.V. Glass – "Meta-analysis refers to the analysis of analyses"
- Overall Goal: systematic summary of evidence on effectiveness of treatment for some condition
- Statistical Aspect: combine estimates from multiple studies with a common "conceptual basis"
- Early Application: Karl Pearson considered typhoid in 1904 (British Medical Journal)
- Today: Appears in a large number of areas in the health sciences – emphasis on multiple studies rather than single studies, and effect size rather than significance

Steps in a Meta-Analysis

The Standard

Cochrane Handbook for Systematic Reviews of Interventions

Steps in the Process – Less Statistical

1. Formulate Question
   - Type of intervention of concern (treatment)
   - Type of population (participants)
   - Definition of response variable(s) (outcomes)
2. Search of Literature
   - Broad: include gray literature, as many journals as possible
   - Focused: articles published a small number of key journals
3. Selection of Studies to Include in Analysis
   - Acceptable study protocols, size, design
   - Consider target population of studies – often not exactly identical
4. Consider Possible Publication Bias
5. Determine Response or Outcome to Use

Steps in a Meta-Analysis

Steps in the Process – More Statistical

5. Determine Response or Outcome to Use
    - Defininiton of response random variables
    - Allowable estimators
6. Consider Inter-Study Heterogeneity
    - Clinical Heterogeneity: Are definitnions of study groups matched or are sub-sets available
    - Methodological Heterogeneity: Design variability, reporting bias (lack of blinding), selection bias, non-ignorable missingness
    - Statistical Heterogeneity: Sometimes attributed to the other types of heterogeneity listed, but could be due to unknown or uncontrolled factors
7. Produce Aggregate (or Meta-Analytic) Estimate of Effect Size

Examination of Heterogeneity

Cochran's Q

- This is Cochran the statistican, not Cochrane the Handbook
- Suppose k studies having individual effect estimates $T_1, T_2, \ldots, T_k$
- Suppose there are weights for the studies, $w_1, w_2, \ldots, w_k$
- Cochran's Q statistic is

$$\begin{aligned} Q &= \sum_{i=1}^{k} w_i (T_i - \bar{T})^2 \\ &= \sum_{i=1}^{k} w_i \, T_i^2 - \frac{\left(\sum_{i=1}^{k} w_i \, T_i \right)^2}{\sum_{i=1}^{k} w_i} \end{aligned}$$

- Compare observed Q to $\chi^2(k-1)$

Problems with Q

1. Small k, low power to detect heterogeneity
2. A "backwards" hypothesis test
3. Huge k, excessive power to detect minor (meaningless) heterogeneity

For examples see:

Higgins, J.P.T, Thompson, S.G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539-1558.

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. and Altman, D.G. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal* **327**, 557-560.

Examination of Heterogeneity

Inconsistency Index

- Now recommended by the Cochrane Handbook
- Inconsistency Index

$$I^2 = 100 \left(\frac{Q - df}{Q} \right),$$

  Q is Cochran's Q and df are usually $k - 1$
- I^2 lies between 0 (by definition) and 100 (by computation)
- Larger values indicating greater heterogeneity
- The claim is that I^2 measures the percentage of variability in effect estimates due to heterogeneity rather than sampling error
  This seems to come from the fact that if the true treatment effect in all studies is the same, $E(Q) = df = k - 1$

Reference Scale for Inconsistency Index

The Cochrane Handbook suggests a scale with overlapping endpoints,

- 0 to 40 might not be important
- 30 to 60 may represent moderate heterogeneity
- 50 to 90 may represent substantial heterogeneity
- 75 to 100 represents considerable heterogeneity

The Handbook also suggests that both Q and I be used.

Meta-Analytic Estimation: The Concepts

Fundamental Concept

There is a TRUE treatment effect that is constant across studies

Error and Variability

- All studies estimate the treatment effect with error
- There is variability among estimated treatment effects reported across studies
- If error is the primary source of variability then heterogeneity among studies will be small
- Uncontrolled sources of heterogeneity result in greater variability than would be seen with only error in operation

Traditional Models

- Fixed Effect
- Random Effects

Meta-Analytic Estimation: The Conceptual Problem

Claim: The fundamental concept presented previously is overly simplistic

1. What is a TRUE treatment effect?
   Q: Suppose you could give the treatment to the entire target population in the world. Would that allow estimation of the true treatment effect?

2. What actually is ERROR?
   Q: Is the reason observations of *iid* random variables are not all identical due to *error*?

Meta-Analytic Estimation: The Conceptual Problem

Claim: The fundamental concept presented previously is overly simplistic

1. What is a TRUE treatment effect?
   Q: Suppose you could give the treatment to the entire target population in the world. Would that allow estimation of the true treatment effect?

   A: Maybe for this year

2. What actually is ERROR?
   Q: Is the reason observations of *iid* random variables are not all identical due to *error*?

   A: Only if you believe error terms somehow represent actual error

Meta-Analytic Estimation: A Better Concept

- The effect of a treatment is specific to an individual to whom it is given
- Assuming independence and the lack of obvious factors that might introduce bias, these effects can be modeled according to some distribution
- There may, however, be additional structure in data that result in groupings – and these may or may not correspond to identifiable factors

The Concept

- All meta-analyses involve intra-study distributions of treatment effect as well as other random effects due to heterogeneity
- The question is whether such heterogeneity can be paired with study labels or identiities

Meta-Analytic Estimation: The Fixed Effect Model

Notation

- Let T_s denote the estimate of effect from study $s = 1, \ldots, S$
- Let V_s denote the estimate of the variance of T_s from study s – note that this is not the variance of individual responses in study s, nor is it the standard error of T_s, it is the variance of T_s

The Model

The fixed effects model is, for $s = 1, \ldots, S$

$$T_s \sim \text{indep } \mathrm{N}(\mu,\ V_s)$$

Note: The V_s are taken as known

Traditional Analysis – Fixed Effect

- For a set of observed t_s; $s = 1, \ldots, S$, the log likelihood is

$$\ell(\mu) = \sum_{s=1}^{S} \left[\frac{-1}{2V_s}(t_s - \mu)^2 - \frac{1}{2}\log(2\pi V_s) \right]$$

- For maximum likelihood estimation of μ,

$$\frac{d}{d\mu}\ell(\mu) = \sum_{s=1}^{S} \left(\frac{t_s}{V_s} \right) - \mu \sum_{s=1}^{S} \frac{1}{V_s}$$

Setting this derivative equal to zero and solving for μ gives

$$\hat{\mu} = \frac{\sum_{s=1}^{S} w_s\, t_s}{\sum_{s=1}^{S} w_s},$$

where $w_s = 1/V_s$

Traditional Analysis – Fixed Effect

The random version of $\hat{\mu}$ is in the form of a weighted average of the study specific estimated effects $\{T_s : s = 1, \ldots, S\}$

$$\begin{aligned} \hat{\mu} &= \frac{\sum_{s=1}^{S} w_s T_s}{\sum_{s=1}^{S} w_s} \\ &= \sum_{s=1}^{S} a_s T_s \end{aligned}$$

where

$$\sum_{s=1}^{S} a_s = 1$$

Traditional Analysis – Fixed Effect

This implies that

- $\hat{\mu}$ is unbiased for μ in $T_s \sim$ indep$N(\mu, V_s)$
- The variance of $\hat{\mu}$ is,

$$\begin{aligned} \text{var}(\hat{\mu}) &= \sum_{s=1}^{S} a_s^2 \text{var}(T_s) \\ &= \frac{1}{\sum_{s=1}^{S} w_s} \end{aligned}$$

  To get this, use $w_s = 1/V_s$ and $\text{var}(T_s) = V_s$.
- $\hat{\mu}$ is a linear combination of the independent and normally distributed T_s; $s = 1, \ldots, S$ and hence has a normal distribution
- An exact (or small sample theory) confidence interval for μ is then

$$\hat{\mu} \pm z_{1-\alpha/2} \left[\text{var}(\hat{\mu})\right]^{1/2}$$

Meta-Analytic Estimation: The Random Effects Model

The random effects model may be written as

$$\begin{aligned} T_s &\sim \text{indep } \mathsf{N}(\theta_s,\, V_s) \\ \theta_s &\sim \text{iid } \mathsf{N}(\mu,\, \tau^2) \end{aligned}$$

This model may also be written as

$$T_s = \theta_s + \epsilon_s,$$

where $\theta_s \sim$ iid $\mathsf{N}(\mu,\, \tau^2)$ and $\epsilon_s \sim$ iid $\mathsf{N}(0,\, V_s)$
Or, it may also be written as

$$T_s = \mu + \delta_s + \epsilon_s,$$

where $\delta_s \sim$ iid $\mathsf{N}(0,\, \tau^2)$ and $\epsilon_s \sim$ iid $\mathsf{N}(0,\, V_s)$

Meta-Analytic Estimation: The Random Effects Model

For any of these equivalent forms,

$$T_s \sim \text{indep } \mathsf{N}(\mu, \tau^2 + V_s)$$

According to the Better Concept for Meta-Analysis presented previously,

- The manifestation of the mechanism in a given situation indexed by s is reflected by θ_s or $\mu + \delta_s$, depending on how the model was written
- Also note that under any formulation of the model the parameter μ now represents the expected effect resulting from a random draw of the mechanism from its distribution. So, while the distribution of effects is of interest, so is the particular value μ
- But note that the random effect or hierarchical grouping is in terms of studies – if heterogeneity arises due to differences in *individuals* this implies that there might be discernible differences among subjects used in different studies

Traditional Analysis – Random Effects

Estimation Problem

- There are two parameters to be estimated, μ and τ^2
- No single agreed upon approach (frequentist)
- The Original: moment-based estimation of τ^2 combined with weighted least squares for μ
- Several moment-based estimators of τ^2 have been proposed
- Why not simply use maximum likelihood

Log Likelihood for Random Effects

$$\ell(\mu, \tau^2) = -\frac{1}{2}\sum_{s=1}^{S} \log[2\pi(\tau^2 + V_s)] - \frac{1}{2}\sum_{s=1}^{S}\left(\frac{(t_s - \mu)^2}{\tau^2 + V_s}\right)$$

Note: the V_s are still being treated as known

Traditional Analysis – Random Effects

Taking the derivatives with respect to μ and τ^2 and setting equal to zero gives the equations

$$\hat{\mu} = \frac{\sum_{s=1}^{S} \left(\frac{t_s}{\tau^2 + V_s} \right)}{\sum_{s=1}^{S} \left(\frac{1}{\tau^2 + V_s} \right)} \tag{1}$$

$$\sum_{s=1}^{S} \frac{1}{\tau^2 + V_s} = \sum_{s=1}^{S} \left(\frac{t_s - \mu}{\tau^2 + V_s} \right)^2 \tag{2}$$

To locate maximum likelihood estimtaes numerically, use profile likelihood approach.

Substitute (1) into (2) and solve the resulting one-dimensional optimization problem (in τ^2) using something like equal interval search – then substitute solution into (1)

Traditional Analysis – Random Effects

Variances result from taking second derivatives. Here

$$\frac{\partial^2}{\partial \mu^2}\ell(\mu, \tau^2) = -\sum_{s=1}^{S} \frac{1}{\tau^2 + V_s}$$

$$\frac{\partial^2}{\partial \mu\, \partial \tau^2}\ell(\mu, \tau^2) = -\sum_{s=1}^{S} \frac{(t_s - \mu)}{(\tau^2 + V_s)^2}$$

$$\frac{\partial^2}{\partial (\tau^2)^2}\ell(\mu, \tau^2) = \frac{1}{2}\sum_{s=1}^{S} \frac{1}{(\tau^2 + V_s)^2} - \sum_{s=1}^{S} \frac{(t_s - \mu)^2}{(\tau^2 + V_s)^3}$$

Note that taking negative expected values simplifies these expressions. If we define

$$w_s^* = \frac{1}{\tau^2 + V_s}$$

the expected (or Fisher) information matrix becomes

$$I(\mu, \tau^2) = \begin{pmatrix} \sum w_s^* & 0 \\ 0 & \frac{1}{2}\sum (w_s^*)^2 \end{pmatrix}$$

Traditional Analysis – Random Effects

Finally, an approximate $(1-\alpha)100\%$ Wald theory interval is then,

$$\hat{\text{var}}(\hat{\mu}) = \frac{1}{\sum_{s=1}^{S} w_s^*}$$

where the weights w_s^* are evaluated at the maximum likelihood estimates $\hat{\tau}^2$.

A benefit of full likelihood estimation is that we also get an interval for τ^2. The appropriate variance here is

$$\hat{\text{var}}(\hat{\tau}^2) = \frac{2}{\sum_{s=1}^{S} (w_s^*)^2}$$

again with the w_s^* evaluated at $\hat{\tau}^2$. Then an approximate interval for τ^2,

$$\hat{\tau}^2 \pm z_{1-\alpha/2} \left[\hat{\text{var}}(\hat{\tau}^2)\right]^{1/2}$$

Bayesian Analyses – Fixed Effect Model

- Bayesian analysis of the fixed effect model is straightforward as there is only one unknown parameter μ. The posterior of μ will be available in closed form.
- Recall the model that, for $s = 1, \ldots, S$ and known variances V_s (estimated in the individual studies)

$$T_s \sim \text{indep } \mathrm{N}(\mu,\, V_s)$$

- Let $\mu \sim N(M_0, V_0)$ be the prior.
- Then,

$$\begin{aligned} p(\mu|\boldsymbol{t}) &\propto \pi(\mu)\, f(\boldsymbol{t}|\mu) \\ &\propto \exp\left[-\frac{1}{2}\sum_{s=1}^{S}\frac{(t_s-\mu)^2}{V_s} - \frac{1}{2}\frac{(\mu-M_0)^2}{V_0}\right] \\ &\propto \exp\left[-\frac{1}{2}\left\{\left(\sum_{s}^{S}\frac{1}{V_s}+\frac{1}{V_0}\right)\mu^2 - 2\left(\sum_{s}^{S}\frac{t_s}{V_s}+\frac{M_0}{V_0}\right)\mu\right\}\right] \end{aligned}$$

Bayesian Analyses – Fixed Effect Model

- Upon completing the square, the posterior $p(\mu|\boldsymbol{t})$ is normal with mean

$$M_\mu = \frac{\sum_{s=1}^{S} w_s\, t_s + \dfrac{M_0}{V_0}}{\sum_{s=1}^{S} w_s + \dfrac{1}{V_0}}$$

and variance

$$V_\mu = \frac{1}{\sum_{s=1}^{S} w_s + \dfrac{1}{V_0}},$$

where $w_s = 1/V_s$ as in the previous material on the fixed effect model

Bayesian Analyses – Random Effects Model

Bayesian analysis of the random effects model becomes more complex than the fixed effect model, requiring the use of MCMC methods to simulate from the joint posterior,

$$p(\mu, \tau^2, \{\theta_s : s = 1, \ldots, S\}|\boldsymbol{t})$$

which means we have also simulated from

$$p(\mu, \tau^2|\boldsymbol{t}) \qquad p(\mu|\boldsymbol{t}) \quad \text{and} \quad p(\tau^2|\boldsymbol{t})$$

and inference involves only making probability statements on the basis of these posteriors

What About the V_s?

There are two major issues that arise if we begin to consider the role of the variances V_s in a Meta-Analytic estimation problem:

1. Throughout, these estimated variances have been taken as true values for var(T_s)
2. All estimates of overall effect (μ), regardless of model or estimation approach end up being functions of the T_s that are weighted by something that is smaller for larger vlaues of V_s and larger for smaller values of V_s

Models for Random V_s

Consider here the simple case in which the T_s are sample means $V_s = \hat{v}(T_s)$

Joint Normality: Approximate Sampling Distributions

$$\begin{aligned} T_s &\sim \text{indep } N(\theta_s, \sigma^2) \\ V_s &\sim \text{iid } N(\eta, \psi^2) \\ \theta_s &\sim \text{iid } N(\mu, \tau^2) \end{aligned}$$

- Straightforward extension of previous models
- Does not maintain weighting of individual effect estimates T_s by estimated variances V_s

Models for Random V_s

Now modify V_s to be the variance of responses (for which T_s is an average) so that $\hat{v}(T_j) = V_j/n_j$

Relating V_s and var(T_s) Through n_s

$$\begin{aligned} T_s &\sim \text{indep } N(\theta_s, \sigma^2/n_s) \\ V_s &\sim \text{indep } N(\sigma^2, 2(\sigma^2)^2/n_s) \\ \theta_s &\sim \text{iid } N(\mu, \tau^2) \end{aligned}$$

- Now differences in var(T_s) rely on differences in n_s, but not differences in V_s
- Distribution for V_s motivated by inverse information in asymptotic normality of mle for variance in a one-sample normal problem

Models for Random V_s

Now relax assumption that (given θ_s) response variables on which T_s is based all have the same variance σ^2

Unequal True Variances

$$
\begin{aligned}
T_s &\sim \text{indep } N(\theta_s, \sigma_s^2/n_s) \\
V_s &\sim \text{indep } N(\sigma_s^2, 2(\sigma_s^2)^2/n_s) \\
\theta_s &\sim \text{iid } N(\mu, \tau^2) \\
\sigma_s^2 &\sim \text{iid IG}(\alpha, \beta),
\end{aligned}
$$

where IG(α, β) denotes an inverse gamma distribution with parameters α and β.

- Variances var(T_s) may now differ due to more than only differences in sample sizes n_s
- Distributions of T_s and V_s still motivated by asymptotic normality of mles in one-sample normal problem

The Precision Fallacy

Go back to the weights using in computation of the mle for $\hat{\mu}$ in the fixed effect model:

$$w_s = \frac{1}{V_s}$$

- Under the model, these are measures of the precisions with which the (unbiased) T_s estimate μ
- But, since the V_s are, in truth, estimates, the weights are no longer measures of this precision

Simple Illustration

One random sample of $n = 100$ from $N(12, 3)$, divided into two equal groups with $n_1 = 50$ and $n_2 = 50$.

$\bar{y}_1 = 12.18$, $\bar{y}_2 1 = 11.80$, $s_1^2 = 3.89$, $s_2^2 = 2.48$, so

$V_1 = 0.0778$ and $V_2 = 0.0496$

Q: Here, V_2 is only 64% of the value of V_1. Does this mean $\bar{y}_2$ is a more precise estimate of μ than is $\bar{y}_1$?

Avoiding the Fallacy

Aggregate Estimates May Be Weighted By

1. Variances
2. Sample Sizes

But Should Not Be Weighted By

1. Estimated Variances

Q: How would results of a Meta-Analysis compare if one weighted by either true variances or sample sizes versus the traditional approach?