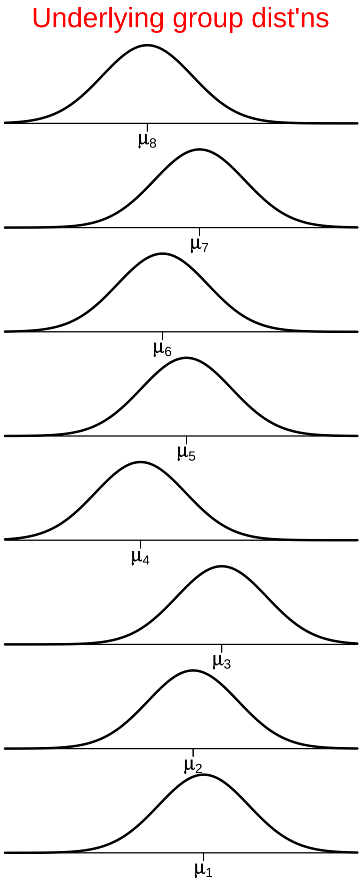
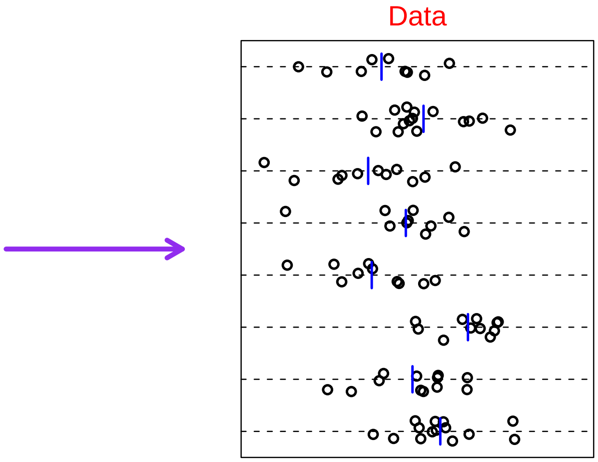
Standard ANOVA model





The ANOVA model

Three different ways to describe the model:

A. $\mathrm{Y_{ti}}$ independent with $\mathrm{Y_{ti}} \sim \mathrm{N}(\mu_{\mathrm{t}}, \sigma^2)$

B. $\mathrm{Y_{ti}} = \mu_{\mathrm{t}} + \epsilon_{\mathrm{ti}}$ where $\epsilon_{\mathrm{ti}} \sim$ iid $\mathrm{N}(0, \sigma^2)$

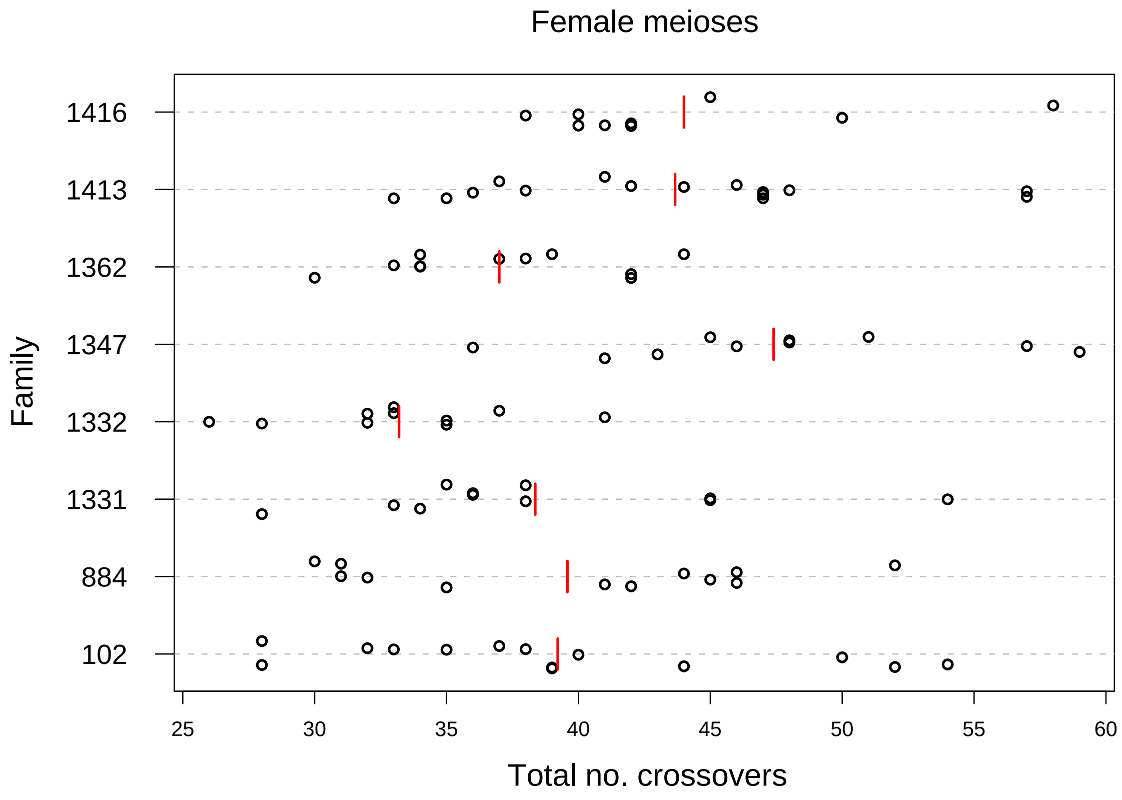
C. $\mathrm{Y_{ti}} = \mu + \tau_{\mathrm{t}} + \epsilon_{\mathrm{ti}}$ where $\epsilon_{\mathrm{ti}} \sim$ iid $\mathrm{N}(0, \sigma^2)$ and $\sum_{\mathrm{t}} \tau_{\mathrm{t}} = 0$

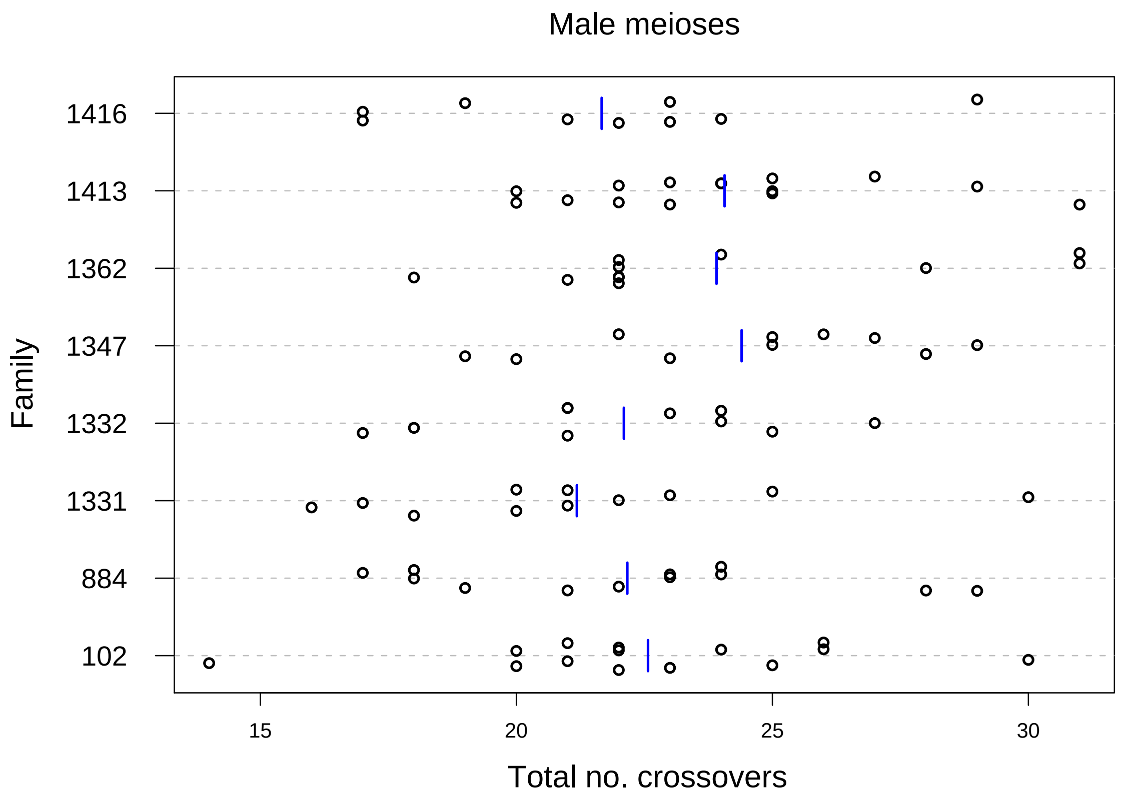
Example

For each of 8 mothers and 8 fathers, we observe (estimates of) the number of crossovers, genome-wide, in a set of independent meiotic products.

Question:

Do the fathers (or mothers) vary in the number of crossovers they deliver?





ANOVA tables

Female meioses:

source	SS	df	MS	F	P-value
between families	1485	7	212.2	4.60	0.0002
within families	3873	84	46.1		
total	5358	91			

Male meioses:

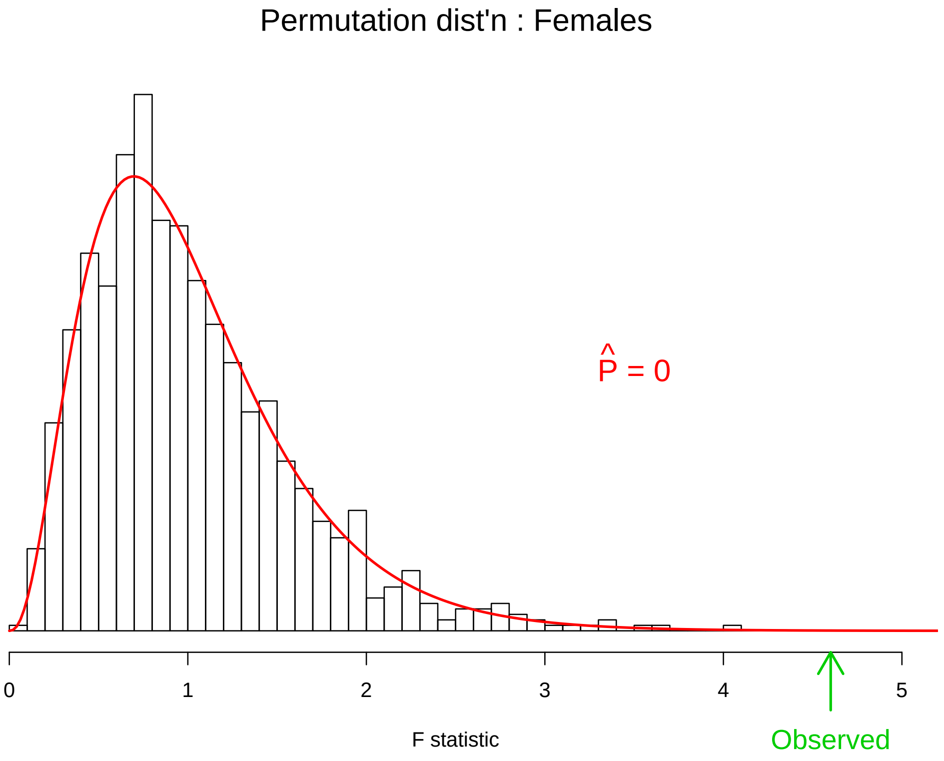
source	SS	df	MS	F	P-value
between families	114	7	16.3	1.23	0.30
within families	1112	84	13.2		
total	1226	91			

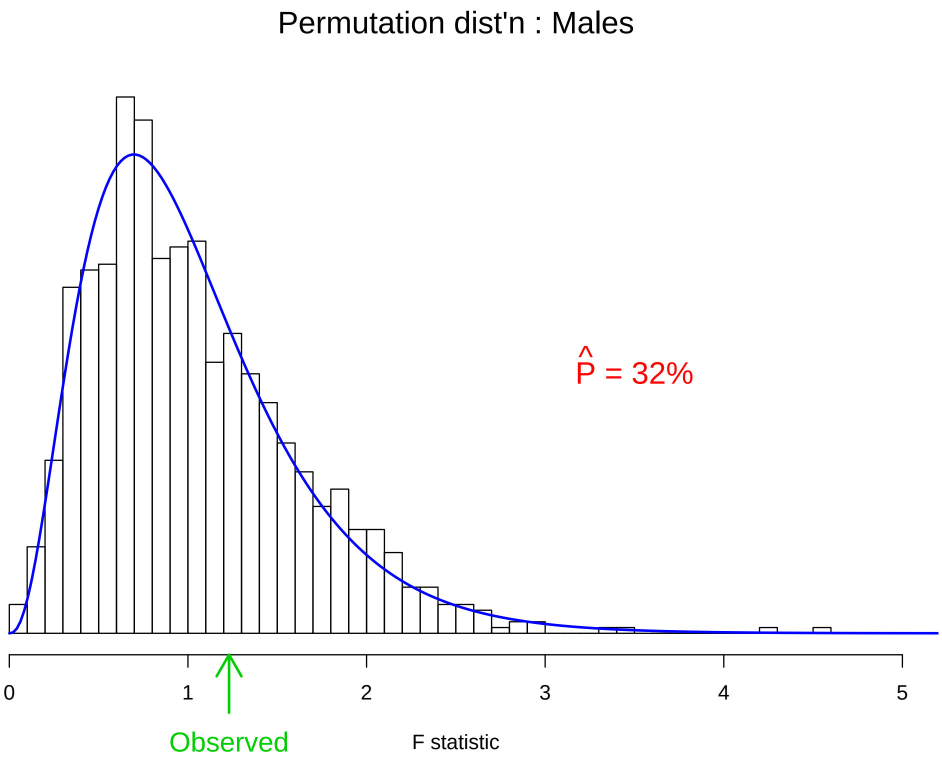
Permutation test

The P-values calculated above rely on the assumption that the measurements in the underlying populations are normally distributed.

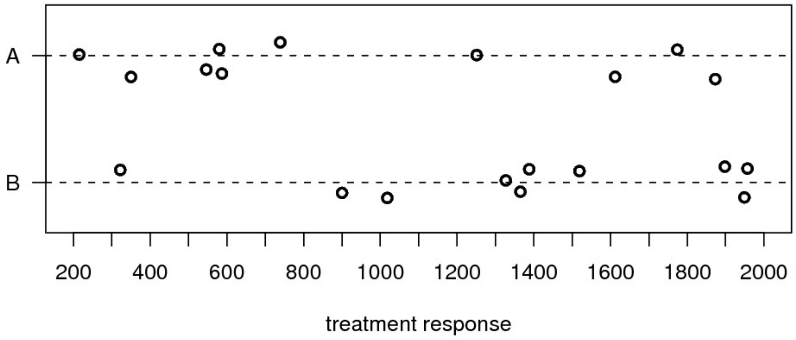
Alternatively, one may use a permutation (aka randomization) test to obtain P-values.

1. Permute (shuffle) the XO counts relative to the family IDs.
2. Re-calculate the F statistic.
3. Repeat (1) and (2) many times (e.g., 1000 or 10,000 times).
4. Estimate the P-value as the proportion of the F statistics from permuted data that are $\geq$ the observed F statistic.





Another example



Are the population means the same?

By now, we know two ways of testing that:

- two-sample t-test
- ANOVA with two treatments

But do they give similar results?

ANOVA table

source	sum of squares	df	mean square
between treatments	$S_B = \sum_t n_t(\bar{Y}_{t\cdot} - \bar{Y}_{\cdot\cdot})^2$	k – 1	$M_B = S_B/(k-1)$
within treatments	$S_W = \sum_t \sum_i (Y_{ti} - \bar{Y}_{t\cdot})^2$	N – k	$M_W = S_W/(N-k)$
total	$S_T = \sum_t \sum_i (Y_{ti} - \bar{Y}_{\cdot\cdot})^2$	(N – 1)	

ANOVA for two groups

The ANOVA test statistic is M_B/M_W, with

$$M_B = n_1(\bar{Y}_1 - \bar{Y}_{\cdot\cdot})^2 + n_2(\bar{Y}_2 - \bar{Y}_{\cdot\cdot})^2$$

and

$$M_W = \frac{\sum_{i=1}^{n_1} (Y_{1i} - \bar{Y}_1)^2 + \sum_{i=1}^{n_2} (Y_{2i} - \bar{Y}_2)^2}{n_1 + n_2 - 2}$$

Two-sample t-test

The test statistic for the two sample t-test is

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{s\sqrt{1/n_1 + 1/n_2}}$$

with

$$s^2 = \frac{\sum_{i=1}^{n_1} (Y_{1i} - \bar{Y}_1)^2 + \sum_{i=1}^{n_2} (Y_{2i} - \bar{Y}_2)^2}{n_1 + n_2 - 2}$$

This also assumes equal variance within the groups!

Result

$$\frac{M_B}{M_W} = t^2$$

Reference distributions

If there was no difference in means, then

$$\frac{M_B}{M_W} \sim F_{1,n_1+n_2-2}$$

$$t \sim t_{n_1+n_2-2}$$

Now does this mean $F_{1,n_1+n_2-2} = (t_{n_1+n_2-2})^2$?

A few facts

$$F_{1,k} = t_k^2$$

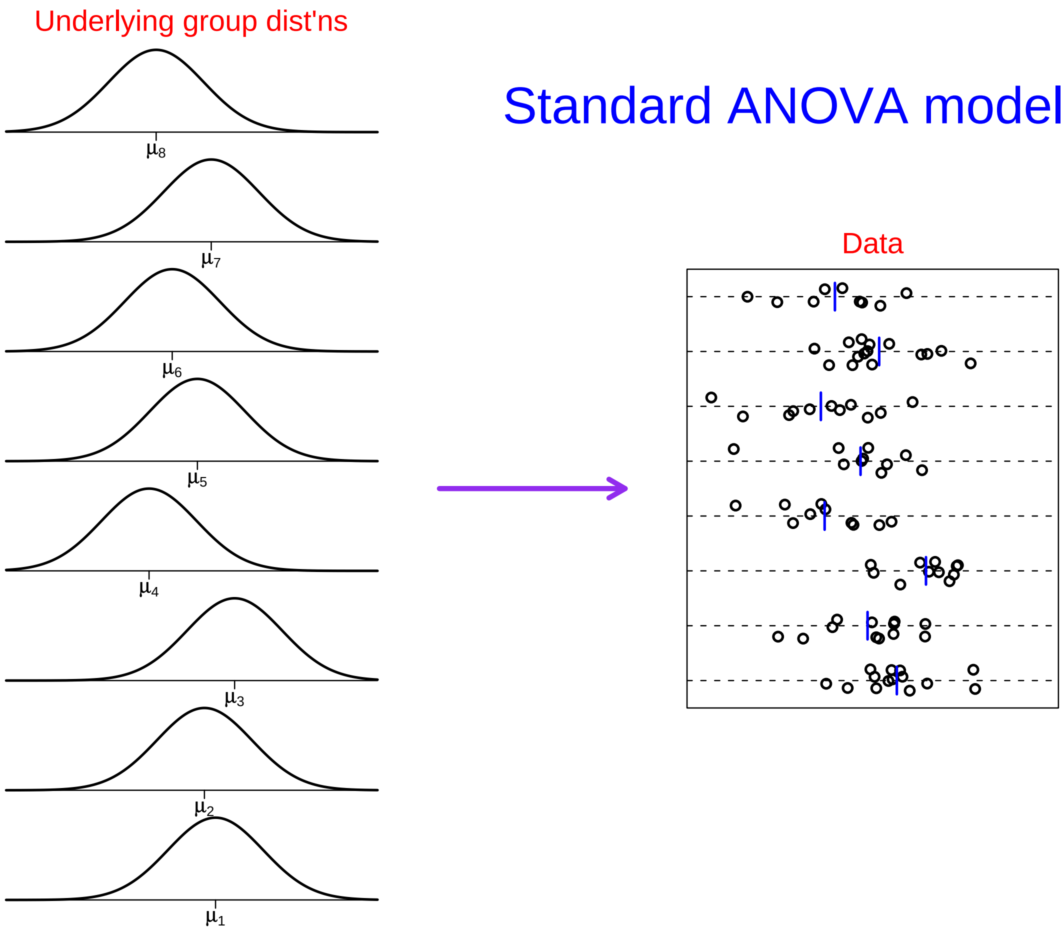
$$F_{k,\infty} = \frac{\chi_k^2}{k}$$

$$N(0,1)^2 = \chi_1^2 = F_{1,\infty} = t_\infty^2$$

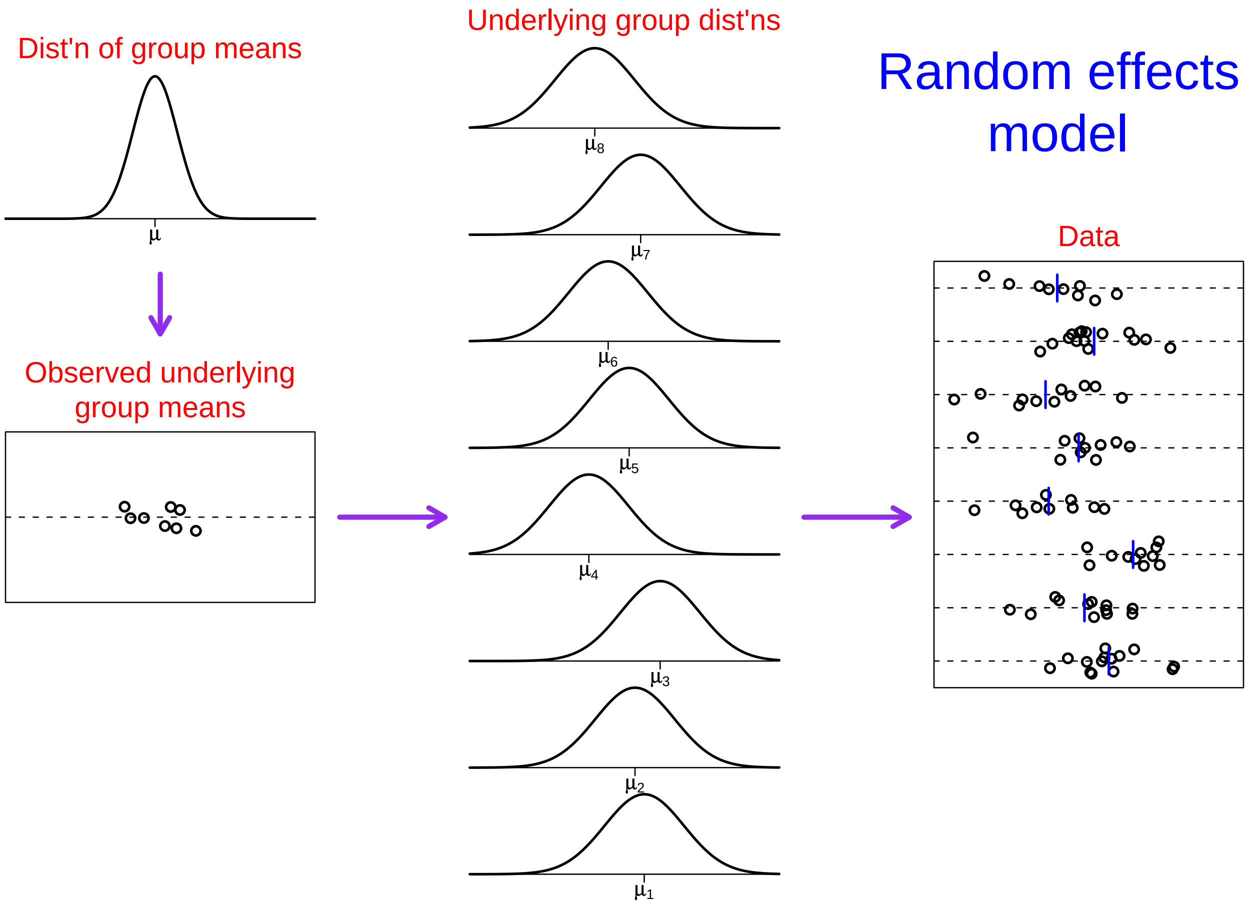
F table

	1	2	·	k_1	·	·	∞
1		·					
2		·					
·		·					
k_2	$t^2_{k_2}$	·		F_{k_1,k_2}			
·		·					
·	·	·	·	·	·	·	·
∞	t^2_{∞}	·		$\dfrac{\chi^2_{k_1}}{k_1}$			

Standard ANOVA model



Random effects model



The random effects model

Two different ways to describe the model:

A. $\mu_t \sim$ iid $N(\mu, \sigma_A^2)$

$$Y_{ti} = \mu_t + \epsilon_{ti} \text{ where } \epsilon_{ti} \sim \text{ iid } N(0, \sigma^2)$$

B. $\tau_t \sim$ iid $N(0, \sigma_A^2)$

$$Y_{ti} = \mu + \tau_t + \epsilon_{ti} \text{ where } \epsilon_{ti} \sim \text{ iid } N(0, \sigma^2)$$

$\longrightarrow$ We add another layer of sampling.

Hypothesis testing

In the standard ANOVA model, we considered the μ_t as fixed but unknown quantities.

We test the hypothesis $H_0 : \mu_1 = \cdots = \mu_k$ using the statistics M_B/M_W from the ANOVA table and the comparing this to an $F(k - 1, N - k)$ distribution.

In the random effects model, we consider the μ_t as random draws from a normal distribution with mean μ and variance σ_A^2.

We seek to test the hypothesis $H_0 : \sigma_A^2 = 0$ versus $H_a : \sigma_A^2 > 0$.

As it turns out, we end up with the same test statistic and same null distribution.

Estimation

For the random effects model it can be shown that

$$E(M_B) = \sigma^2 + n_0 \times \sigma_A^2$$

where

$$n_0 = \frac{1}{k - 1}\left(N - \frac{\sum_t n_t^2}{\sum_t n_t}\right)$$

Recall also that $E(M_W) = \sigma^2$.

Thus, we may estimate σ^2 by $\hat{\sigma}^2 = M_W$.

And we may estimate σ_A^2 by $\hat{\sigma}_A^2 = (M_B - M_W)/n_0$

(provided that this is ≥ 0).

The first example

The samples sizes for the 8 families were (14, 12, 11, 10, 10, 11, 15, 9), for a total sample size of 92.

Thus, $\mathrm{n}_0 \approx 11.45$.

For the female meioses, $\mathrm{M_B}$ = 212 and $\mathrm{M_W}$ = 46. Thus

$\hat\sigma = \sqrt{46} = 6.8$ (Note: overall sample mean = 40.3)

$\hat\sigma_{\mathrm{A}} = \sqrt{(212 - 46)/11.45} = 3.81.$

For the male meioses, $\mathrm{M_B}$ = 16.3 and $\mathrm{M_W}$ = 13.2. Thus

$\hat\sigma = \sqrt{13.2} = 3.6$ (Note: overall sample mean = 22.8)

$\hat\sigma_{\mathrm{A}} = \sqrt{(16.3 - 13.2)/11.45} = 0.52.$