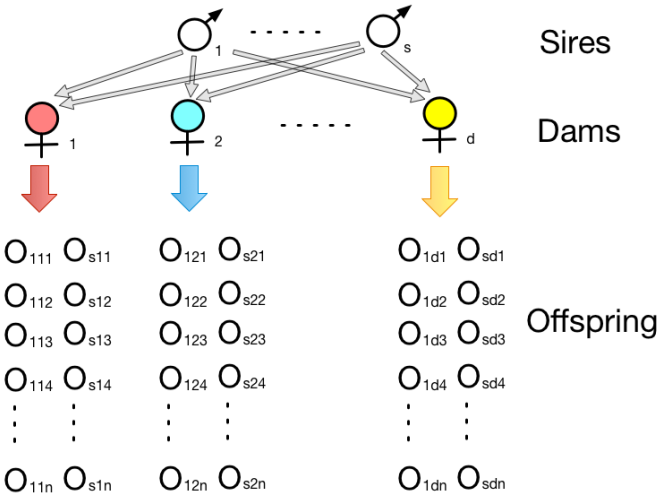
Heritability: precision of estimation

Jinliang Yang
Oct. 31, 2018

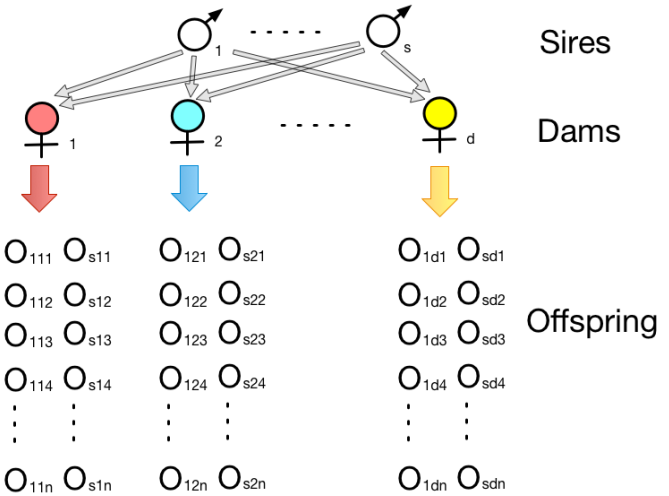
A sib design

Summary of a (balanced) nested analysis of variance involving s sires, d dams per sire and n offspring per dam.



A sib design

Summary of a (balanced) nested analysis of variance involving s sires, d dams per sire and n offspring per dam.



$$p_{ijk} = \mu + s_i + d_{ij} + w_{ijk}$$

- p_{ijk} is the phenotypic value of the kth offspring from the family of the i th sire and jth dam.
- s_i is the effect of the ith sire, d_{ij} is the effect of jth dam mated to the ith sire, and w_{ijk} is the within dam residual deviation.

A sib design

$$p_{ijk} = \mu + s_i + d_{ij} + w_{ijk}$$

Phenotypic variance

- Under the assumption that individuals are random members of the same population
- s_i, d_{ij}, and w_{ijk} are independent

Therefore,

$$\sigma_P^2 = \sigma_S^2 + \sigma_D^2 + \sigma_W^2$$

A sib design

$$p_{ijk} = \mu + s_i + d_{ij} + w_{ijk}$$

A sib design

$$p_{ijk} = \mu + s_i + d_{ij} + w_{ijk}$$

ANOVA Table:

Source	df	Sums of Squares	MS	E(MS)
Sires	s-1	?	MS_s	$= \sigma^2_W + n\sigma^2_D + dn\sigma^2_S$
Dams (Sires)	s(d-1)	?	MS_d	$= \sigma^2_W + n\sigma^2_D$
Sibs (Dams)	sd(n-1)	?	MS_w	$= \sigma^2_W$

- $\bar{p}_{ij}$: the mean value for the **full-sib family** of the ith sire and j dam.
- $\bar{p}_i$: the mean value for the **half-sib family** of the ith sire.
- $\bar{p}$: the **overall** mean.

A sib design

$$p_{ijk} = \mu + s_i + d_{ij} + w_{ijk}$$

ANOVA Table:

Source	df	Sums of Squares	MS	E(MS)
Sires	s-1	$dn\sum_{i=1}^{s}(\bar{p}_i - \bar{p})^2$	MS_s	$= \sigma_W^2 + n\sigma_D^2 + dn\sigma_S^2$
Dams (Sires)	s(d-1)	$n\sum_{i=1}^{s}\sum_{j=1}^{d}(\bar{p}_{ij} - \bar{p}_i)^2$	MS_d	$= \sigma_W^2 + n\sigma_D^2$
Sibs (Dams)	sd(n-1)	$\sum_{i=1}^{s}\sum_{j=1}^{d}\sum_{k=1}^{n}(p_{ijk} - \bar{p}_{ij})^2$	MS_w	$= \sigma_W^2$

- $\bar{p}_{ij}$: the mean value for the **full-sib family** of the ith sire and j dam.
- $\bar{p}_i$: the mean value for the **half-sib family** of the ith sire.
- $\bar{p}$: the **overall** mean.

A sib design

What is the E(SS) for sibs?

$$E\left(\sum_{i=1}^{s}\sum_{j=1}^{d}\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2\right)$$

A sib design

What is the E(SS) for sibs?

$$E\left(\sum_{i=1}^{s}\sum_{j=1}^{d}\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2\right)$$
$$=\sum_{i=1}^{s}\sum_{j=1}^{d}E\left(\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2\right)$$

A sib design

What is the E(SS) for sibs?

$$E\left(\sum_{i=1}^{s}\sum_{j=1}^{d}\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2\right)$$
$$=\sum_{i=1}^{s}\sum_{j=1}^{d}E\left(\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2\right)$$

Where, by definition

$$\frac{\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2}{n-1}=\sigma_W^2$$

is an **unbiased estimate of the variance among full-sibs** and from our assumption it equals to σ_W^2.

A sib design

What is the E(SS) for sibs?

$$E\left(\sum_{i=1}^{s}\sum_{j=1}^{d}\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2\right)$$
$$=\sum_{i=1}^{s}\sum_{j=1}^{d}E\left(\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2\right)$$

Where, by definition

$$\frac{\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2}{n-1}=\sigma_W^2$$

is an **unbiased estimate of the variance among full-sibs** and from our assumption it equals to σ_W^2.

E(SS) $= sd(n-1)\sigma_W^2$

A sib design

ANOVA Table:

Source	df	Sums of Squares	MS	E(MS)
Sires	s-1	$dn\sum_{i=1}^{s}(\bar{p}_i-\bar{p})^2$	MS_s	$=\sigma_W^2+n\sigma_D^2+dn\sigma_S^2$
Dams (Sires)	s(d-1)	$n\sum_{i=1}^{s}\sum_{j=1}^{d}(\bar{p}_{ij}-\bar{p}_i)^2$	MS_d	$=\sigma_W^2+n\sigma_D^2$
Sibs (Dams)	sd(n-1)	$\sum_{i=1}^{s}\sum_{j=1}^{d}\sum_{k=1}^{n}(p_{ijk}-\bar{p}_{ij})^2$	MS_w	$=\sigma_W^2$

- $\bar{p}_{ij}$: the mean value for the **full-sib family** of the ith sire and j dam.
- $\bar{p}_i$: the mean value for the **half-sib family** of the ith sire.
- $\bar{p}$: the **overall** mean.

Estimation of heritability

The heritability is defined as the ratio of additive genetic variance to phenotypic variance:

$$h^2 = \frac{V_A}{V_P} = \frac{\sigma_A^2}{\sigma_P^2}$$

Estimation of heritability

The heritability is defined as the ratio of additive genetic variance to phenotypic variance:

$$h^2 = \frac{V_A}{V_P} = \frac{\sigma_A^2}{\sigma_P^2}$$

From the nested sib design:

$$\sigma_P^2 = \sigma_S^2 + \sigma_D^2 + \sigma_W^2$$
$$\sigma_A^2 = 4 \times \sigma_S^2$$

Interpretation of heritability

- h^2 influenced by allele frequencies, and therefore differ from one **population** to another
- Depends on **environments** and **number of measurements**
- Varies from **traits** to traits
- Varies from **species** to species

Interpretation of heritability

- h^2 influenced by allele frequencies, and therefore differ from one **population** to another
- Depends on **environments** and **number of measurements**
- Varies from **traits** to traits
- Varies from **species** to species

> "The choice of method", (hence design), "is usually determined more by the practical consideration and be freedom from bias, than by precision". - F & M.

In plant breeding

We are normally using the **average of a family in a plot**, e.g. ear numbers in a row, and thus we will need to consider heritability of a family average.

In plant breeding

We are normally using the **average of a family in a plot**, e.g. ear numbers in a row, and thus we will need to consider heritability of a family average.

Or, usually using **inbreed lines**, therefore, estimate broad sense heritability (H^2).

Class	Trait (abbreviation)	H^2	Joint linkage QTL	GWAS SNPs (RMIP>4)
Tassel	Tassel length (TL)	0.93	37	241
Tassel	Spike length (SL)	0.92	33	286
Tassel	Branch zone (BZ)	0.92	26	303
Tassel	Branch number (BN)	0.94	39	325
Ear	Cob length (CL)	0.87	26	233
Ear	Cob diameter (CD)	0.90	39	317
Ear	Ear row number (ERN)	0.89	36	261

All traits were measured in 8 environments, and best linear unbiased predictors (BLUPs) were used to detect joint linkage QTL and GWAS SNPs.
doi:10.1371/journal.pgen.1002383.t001

> Brown et al., 2011

In animal breeding

We are mostly working with **individuals** when using the concept of heritability.

In animal breeding

We are mostly working with **individuals** when using the concept of heritability.

Trait	Paternal half-sibs ($h^2 \pm se$)	Maternal half-sibs ($h^2 \pm se$)
ACP1	0.47 ± 0.26	0.49 ± 0.35
GRP1	0.36 ± 0.25	0.58 ± 0.34
AFA1	0.49 ± 0.26	0.49 ± 0.35
ACP2	$\mathbf{0.09 \pm 0.32}$	$\mathbf{0.91 \pm 0.43}$
GPR2	0.37 ± 0.35	NA
AFA2	0.13 ± 0.32	0.87 ± 0.43
WNCP	0.24 ± 0.34	0.03 ± 0.53
FAIN	0.27 ± 0.34	0.24 ± 0.50

> Irgang and Robinson, 1984.
>
> - ACP1 and ACP2: ages at first and second conception
> - AFA1 and AFA2: ages at first and second farrowing
> - GPR1 and GPR2: ages at first and second gestation periods
> - WNCP: weaning to conception interval
> - FAIN: farrowing interval

Narrow sense heritability

- Parents transmit only one allele to offspring
- Most relatives share only one or zero alleles that are IBD, therefore, only share the average effect of one allele.

Therefore, the **narrow sense heritability (h^2)** is the most important component.

Narrow sense heritability

- Parents transmit only one allele to offspring
- Most relatives share only one or zero alleles that are IBD, therefore, only share the average effect of one allele.

Therefore, the **narrow sense heritability (h^2)** is the most important component.

Important for breeding

- is a fundamental statistics we use in **predicting response** to selection
- and is very informative for **designing breeding schemes**
- enters to **almost every formula** connected with breeding methods.

Regression

From Math foundation:

$$b_{XY} = \frac{Cov(X, Y)}{Var(Y)}$$

Regression

From Math foundation:

$$b_{XY} = \frac{Cov(X, Y)}{Var(Y)}$$

Breeding value and phenotypic value

$$\begin{aligned} b_{AP} &= \frac{Cov(A, P)}{\sigma_P^2} \\ &= \frac{Cov(A, A + D + I + E)}{\sigma_P^2} \\ &= \frac{Cov(A, A)}{\sigma_P^2} = \frac{\sigma_A^2}{\sigma_P^2} = h^2 \end{aligned}$$

Regression

From Math foundation:

$$b_{XY} = \frac{Cov(X, Y)}{Var(Y)}$$

Breeding value and phenotypic value

$$\begin{aligned} b_{AP} &= \frac{Cov(A, P)}{\sigma_P^2} \\ &= \frac{Cov(A, A + D + I + E)}{\sigma_P^2} \\ &= \frac{Cov(A, A)}{\sigma_P^2} = \frac{\sigma_A^2}{\sigma_P^2} = h^2 \end{aligned}$$

Therefore, $h^2 = b_{AP}$ is the **regression coefficient** of breeding value on phenotypic value: $A = h^2 P$.

Bristle number in Drosophila

In total, 38 families with mother's score and mean score of her offspring.

> Data from Bob Sheehy, 1996.

```
d <- read.csv('https://jyanglab.com/AGRO-931-2018/data/drosophila.csv')
dim(d)
```

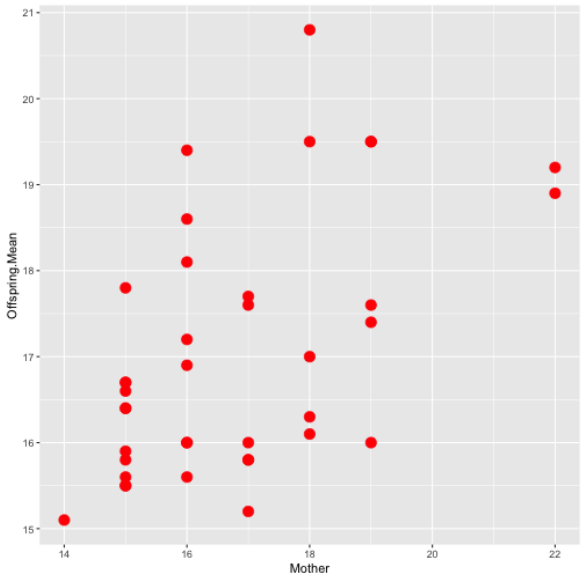
```
## [1] 38  2
```

```
head(d)
```

```
##   Mother Offspring.Mean
## 1     17           17.7
## 2     16           16.0
## 3     22           19.2
## 4     17           15.2
## 5     15           17.8
## 6     15           16.7
```

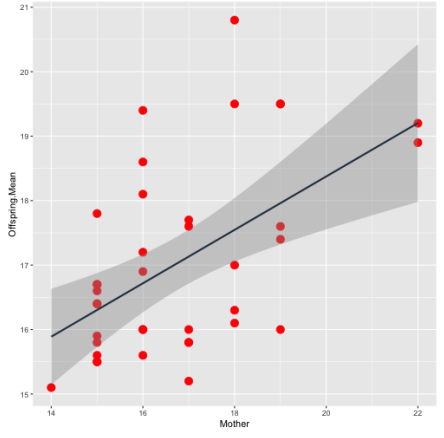
Bristle number in Drosophila

```
library(ggplot2)
ggplot(d, aes(x=Mother, y=Offspring.Mean)) +
  geom_point(color='red', size = 4)
```



Bristle number in Drosophila

```
ggplot(d, aes(x=Mother, y=Offspring.Mean)) +
  geom_point(color='red', size = 4) +
  geom_smooth(method=lm, color='#2C3E50')
```




```
lm(Offspring.Mean ~ Mother, data=d)
```

```
##
## Call:
## lm(formula = Offspring.Mean ~ Mother, data = d)
##
## Coefficients:
## (Intercept)       Mother
##     10.0875       0.4144
```

Parent-Offspring correlation

$$
\begin{aligned}
b_{OP} &= \frac{Cov(O,P)}{\sigma^2_P} \\
&= \frac{Cov(1/2A, A+D+I+E)}{\sigma^2_P} \\
&= \frac{Cov(A,1/2A)}{\sigma^2_P} = \frac{1\sigma^2_A}{2\sigma^2_P} = 1/2h^2
\end{aligned}
$$

Parent-Offspring correlation

$$
\begin{aligned}
b_{OP} &= \frac{Cov(O,P)}{\sigma_P^2} \\
&= \frac{Cov(1/2A, A+D+I+E)}{\sigma_P^2} \\
&= \frac{Cov(A,1/2A)}{\sigma_P^2} = \frac{1\sigma_A^2}{2\sigma_P^2} = 1/2h^2
\end{aligned}
$$

In this case,

The estimated heritability is $2 \times b_{OP} = 2 \times 0.4144 = 0.8288$.

Parent-Offspring correlation

$$
\begin{aligned}
b_{OP} &= \frac{Cov(O,P)}{\sigma_P^2} \\
&= \frac{Cov(1/2A, A+D+I+E)}{\sigma_P^2} \\
&= \frac{Cov(A,1/2A)}{\sigma_P^2} = \frac{1\sigma_A^2}{2\sigma_P^2} = 1/2h^2
\end{aligned}
$$

In this case,

The estimated heritability is $2 \times b_{OP} = 2 \times 0.4144 = 0.8288$.

To predict a mother with score=20?

Parent-Offspring correlation

$$
\begin{aligned}
b_{OP} &= \frac{Cov(O,P)}{\sigma_P^2} \\
&= \frac{Cov(1/2A, A+D+I+E)}{\sigma_P^2} \\
&= \frac{Cov(A,1/2A)}{\sigma_P^2} = \frac{1\sigma_A^2}{2\sigma_P^2} = 1/2h^2
\end{aligned}
$$

In this case,

The estimated heritability is $2 \times b_{OP} = 2 \times 0.4144 = 0.8288$.

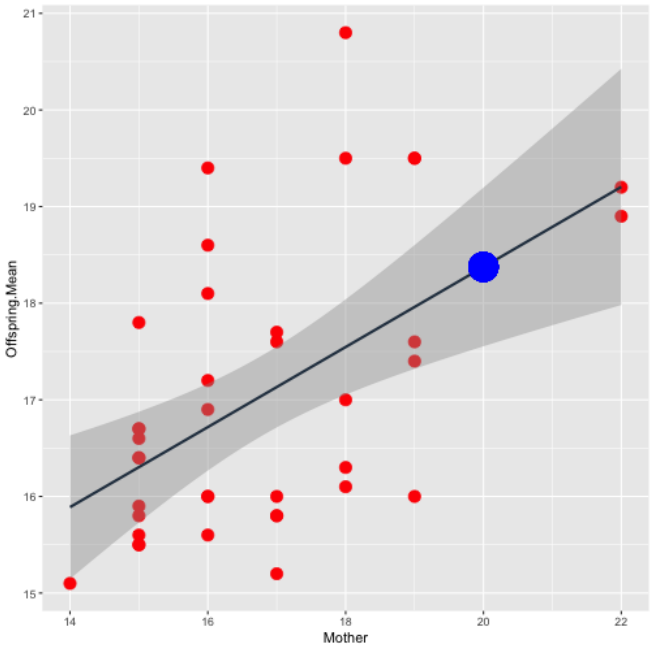
To predict a mother with score=20?

$$E(O) = b_{OP}P = 0.4144 \times 20 + \mu$$

Predicted bristle number

```
fit <- lm(Offspring.Mean ~ Mother, data=d)
o <- predict(fit, data.frame(Mother=20, Offspring.Mea=NA))

ggplot(d, aes(x=Mother, y=Offspring.Mean)) +
  geom_point(color='red', size = 4) +
  geom_smooth(method=lm, color='#2C3E50') +
  geom_point(aes(x=20, y=o), colour="blue", size=10)
```



Precision and design

Data collection must be practical

- time/cost/etc is usually the important determinant

Precision and design

Data collection must be practical

- time/cost/etc is usually the important determinant

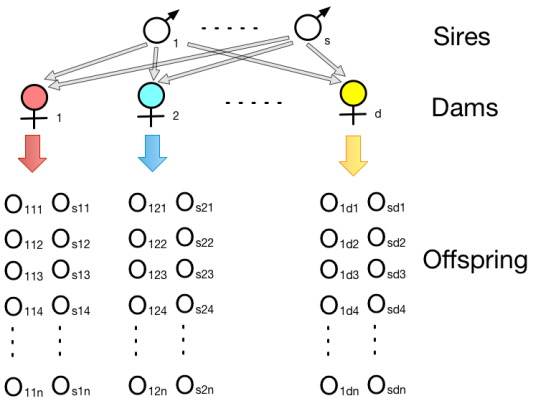
Be freedom from bias

- Random mating
- Absence of common environmental effects, e.g. maternal effects.

Precision and design

If we want to design a balanced experiment:

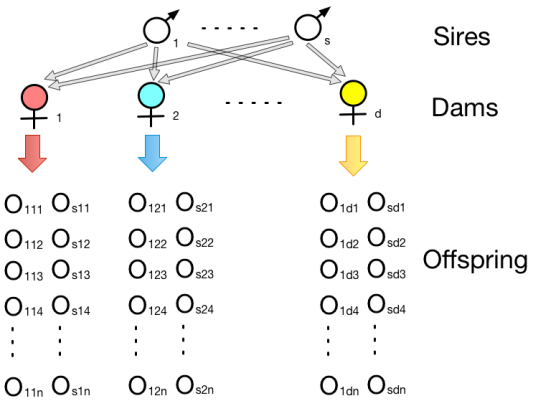
- s sires each mated to d dams
- Each dam has n progenies.



Precision and design

If we want to design a balanced experiment:

- s sires each mated to d dams
- Each dam has n progenies.



Questions before experimental design?

1. Parent-offspring, half-sib, full-sib, or others?
2. How many families?
3. Numbers of progeny?
4. What if it is unbalanced?

Sampling variance of b

F & M page 178, the sampling variance of the parent-offspring regression is approximately:

$$SV_b = \frac{k[1 + (n-1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring per family
- t the intra-class correlation between offspring in a family

Sampling variance of b

F & M page 178, the sampling variance of the parent-offspring regression is approximately:

$$SV_b = \frac{k[1 + (n-1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring per family
- t the intra-class correlation between offspring in a family

One parent

Sampling variance is minimal when $n = 1$, i.e. $(n-1)t = 0$.

Sampling variance of b

F & M page 178, the sampling variance of the parent-offspring regression is approximately:

$$SV_b = \frac{k[1 + (n-1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring per family
- t the intra-class correlation between offspring in a family

One parent

Sampling variance is minimal when $n = 1$, i.e. $(n-1)t = 0$.

The most efficient design:

1. as many as families as possible
2. measure only one offspring per family

One parent

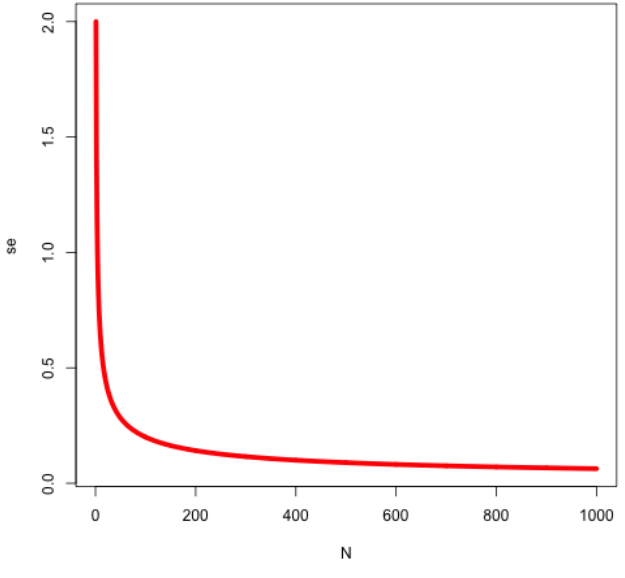
$$SV(h^2 = 2b) = 4/N$$
$$s.e.(h^2) = 2/\sqrt{N}$$

One parent

$$SV(h^2 = 2b) = 4/N$$
$$s.e.(h^2) = 2/\sqrt{N}$$

```
N=1:1000
se <- 2/sqrt(N)

plot(N, se, type="l", lwd=5, col="red")
```



Sampling variance of b

F & M page 178, the sampling variance of the regression is approximately:

$$SV_b = \frac{k[1 + (n-1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring
- t the intra-class correlation between offspring in a family

Sampling variance of b

F & M page 178, the sampling variance of the regression is approximately:

$$SV_b = \frac{k[1 + (n-1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring
- t the intra-class correlation between offspring in a family

Both parent

When $k = 2$ (use mid-parent values), the $SV(h^2 = b) = 2/N$. So the standard error, $se(h^2) = \sqrt{2/N}$

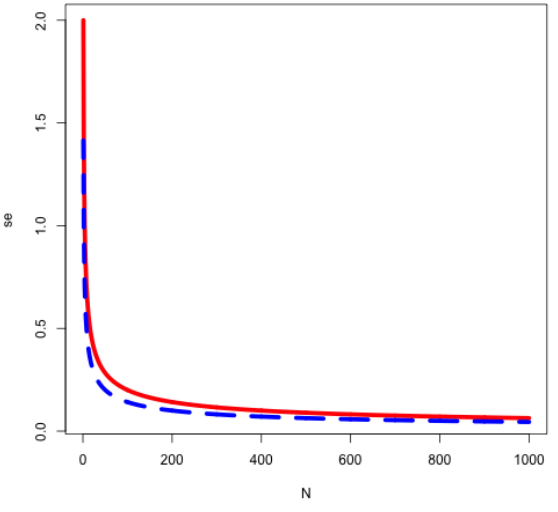
Both-parents

- One parent: $s.e.(h^2) = 2/\sqrt{N}$
- Both-parent: $s.e.(h^2) = \sqrt{2/N}$

Both-parents

- One parent: $s.e.(h^2) = 2/\sqrt{N}$
- Both-parent: $s.e.(h^2) = \sqrt{2/N}$

```
N=1:1000
se <- 2/sqrt(N)
se2 <- sqrt(2/N)
plot(N, se, type="l", lwd=5, col="red")
lines(N, se2, type="l", lwd=5, col="blue", lty=2)
```



Sib analyses

The correlation between A and P, r_{AP}:

$$
\begin{aligned}
r_{AP} &= \frac{Cov(A,P)}{\sqrt{\sigma_A^2 \sigma_P^2}} \\
&= \sqrt{\frac{\sigma_A^2}{\sigma_P^2}} \\
&= h
\end{aligned}
$$

Sib analyses

The correlation between A and P, r_{AP}:

$$
\begin{aligned}
r_{AP} &= \frac{Cov(A,P)}{\sqrt{\sigma_A^2 \sigma_P^2}} \\
&= \sqrt{\frac{\sigma_A^2}{\sigma_P^2}} \\
&= h
\end{aligned}
$$

Therefore, the **intraclass correlation (t)**, is also a function of h.

Sib analyses

Now we look at the sampling variance of the **intraclass correlation (t)**. According to F & M page 180, the sampling variance of **t** is:

$$SV_t = \frac{2[1 + (n-1)t]^2 \times (1-t)^2}{n(n-1)(N-1)}$$

- N families (offspring and parents)
- n offspring per family
- t the intra-class correlation between offspring in a family

Sib analyses

Now we look at the sampling variance of the **intraclass correlation (t)**. According to F & M page 180, the sampling variance of **t** is:

$$SV_t = \frac{2[1 + (n-1)t]^2 \times (1-t)^2}{n(n-1)(N-1)}$$

- N families (offspring and parents)
- n offspring per family
- t the intra-class correlation between offspring in a family

If we let $T = nN$, the total number measured in a generation, the sampling variance is minimized when **n (number of offspring per family) = 1/t**.

In the simplest cases with no common environmental effect in families, then

- $t_{HS} = h^2/4$
- $t_{FS} = h^2/2$

In the simplest cases with no common environmental effect in families, then

- $t_{HS} = h^2/4$
- $t_{FS} = h^2/2$

With $n = 1/t$, approximately

$$SV_t = 8t/(nN)$$

Half-sib families

$$SV(h^2) = 32h^2/(nN)$$

Full-sib families

$$SV(h^2) = 16h^2/(nN)$$

In the simplest cases with no common environmental effect in families, then

- $t_{HS} = h^2/4$
- $t_{FS} = h^2/2$

With $n = 1/t$, approximately

$$SV_t = 8t/(nN)$$

Half-sib families

$$SV(h^2) = 32h^2/(nN)$$

Full-sib families

$$SV(h^2) = 16h^2/(nN)$$

Thus, some assumption of h^2 ahead of time is essential in planning the data collection!