Confidence intervals for group effects

Peter Hoff
Duke STA 610

Frequentist confidence intervals for group means

A confidence interval provides a range of plausible values for θ_j.

$$C(\mathbf{y}) \stackrel{?}{=} \bar{y}_j \pm \frac{\hat{\sigma}}{\sqrt{n_j}} t_{1-\alpha/2}$$

- Exact constant coverage:

$$\Pr(\theta_j \in C(\mathbf{y})|\boldsymbol{\theta}) = 1 - \alpha \text{ for all values of } \theta_j.$$

- Narrowest interval among "unbiased" intervals.
- Doesn't use all available information.

Can we do better by sharing information across groups?

Confidence intervals without constant coverage

$$\begin{aligned}\text{Bias}[\hat{\theta}_j|\boldsymbol{\theta}] &= w(\mu - \theta_j)\\ \text{Var}[\hat{\theta}_j|\boldsymbol{\theta}] &= (1-w)^2\sigma^2/n_j\\ \hat{\theta}_j - \theta_j|\boldsymbol{\theta} &\sim N(w(\mu-\theta_j),(1-w)^2\sigma^2/n)\\ w &= (1/\tau^2)/(n_j/\sigma^2 + 1/\tau^2).\end{aligned}$$

If the hierarchical model is correct, then the variation *across groups* is

$\mu - \theta_j \sim N(0,\tau^2)$ (because $\theta_1,\ldots,\theta_p \sim \text{i.i.d.} N(\mu,\tau^2)$)

and so

$\hat{\theta}_j - \theta_j \sim N(0, 1/(n_j/\sigma^2 + 1/\tau^2))$ *marginally, across groups*.

“Prediction” interval:

$$C(\hat{\theta}_j) = \hat{\theta}_j \pm t_{1-\alpha/2}/\sqrt{n_j/\hat{\sigma}^2 + 1/\hat{\tau}^2}$$

- $1-\alpha$ coverage *on average across groups*.
- Could be lower or higher for any given group, and you don’t know which.

Bayes posterior intervals

- "Prior" density: $\theta_j \sim N(\mu, \tau^2)$
- Sampling density: $y_{1,j}, \ldots, y_{n_j,j} | \theta \sim N(\theta_j, \sigma^2)$.

Bayes rule: $\theta_j | y_{1,j}, \ldots, y_{n_j,j}$ is normal, with

$$\mathrm{E}[\theta_j | y_{1,j}, \ldots, y_{n_j,j}] = \frac{\tau^2}{\sigma^2/n_j + \tau^2} \bar{y}_j + \frac{\sigma^2/n_j}{\sigma^2/n_j + \tau^2} \mu$$

$$\mathrm{Var}[\theta_j | y_{1,j}, \ldots, y_{n_j,j}] = 1/(n_j/\sigma^2 + 1/\tau^2)$$

Bayes posterior intervals

This means that

$$\Pr(|\theta_j - \hat{\theta}_j| \times \sqrt{n_j/\sigma^2 + 1/\tau^2} > z_{1-\alpha/2}|\mathbf{y}_j) = 1 - \alpha$$

or equivalently,

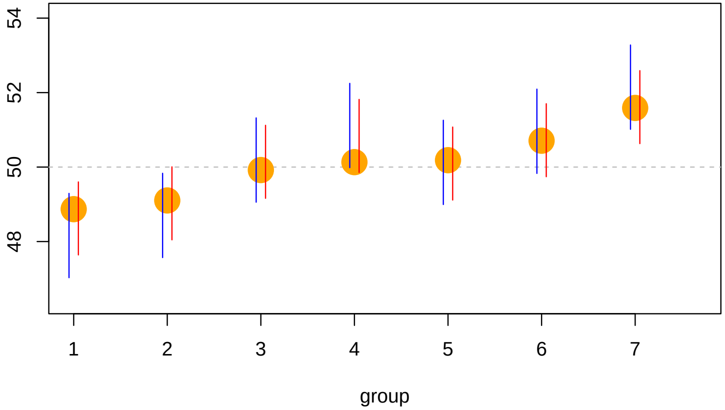
$$\hat{\theta}_j \pm z_{1-\alpha/2}/\sqrt{n_j/\sigma^2 + 1/\tau^2}$$

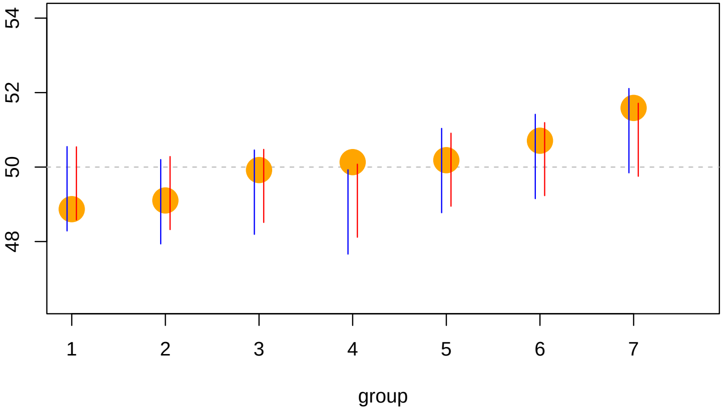
has $1 - \alpha$ *posterior coverage*.

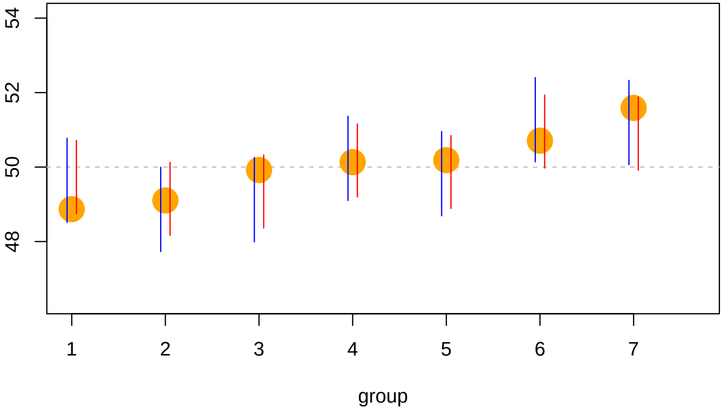
A corresponding Empirical Bayes interval is

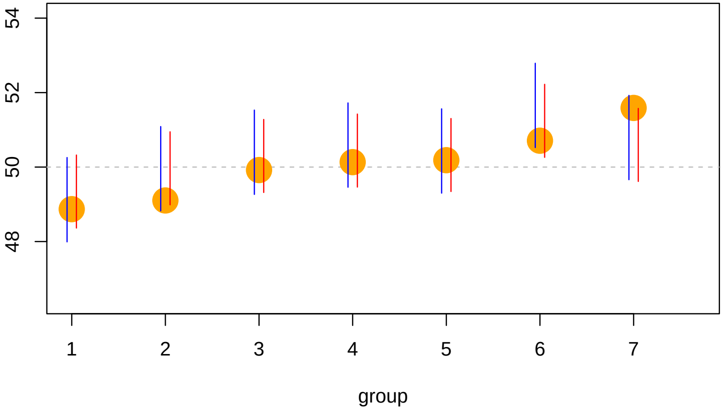
$$C(\hat{\theta}_j) = \hat{\theta}_j \pm t_{1-\alpha/2}/\sqrt{n_j/\hat{\sigma}^2 + 1/\hat{\tau}^2}$$

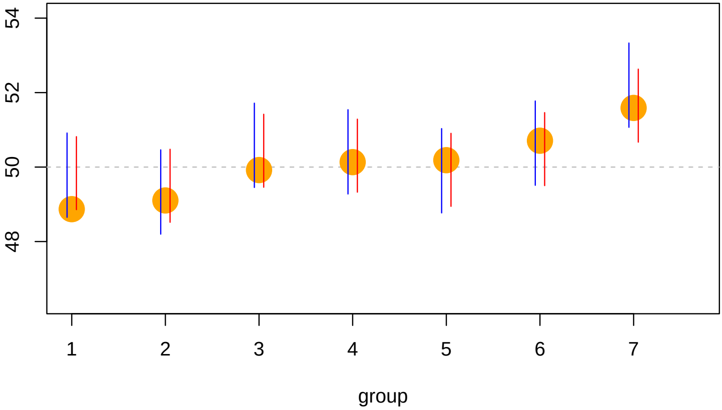
which is the *same* as the prediction interval, but has a different interpretation.

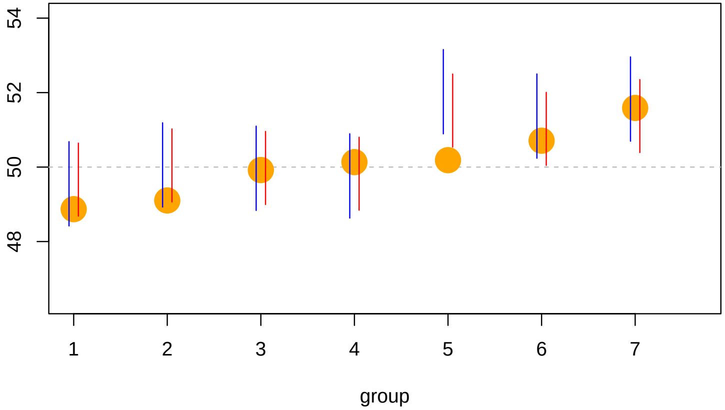

54
52
50
48
1
2
3
4
5
6
7
group

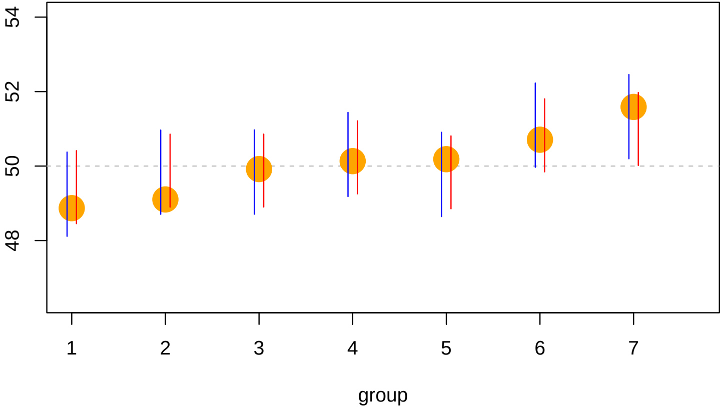

54
52
50
48
1
2
3
4
5
6
7
group

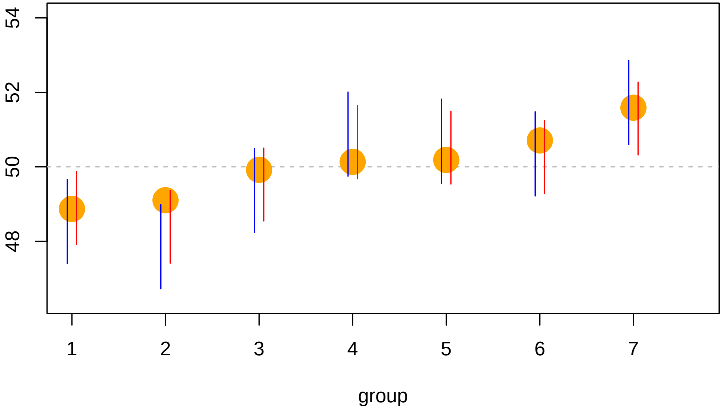

54
52
50
48
1
2
3
4
5
6
7
group

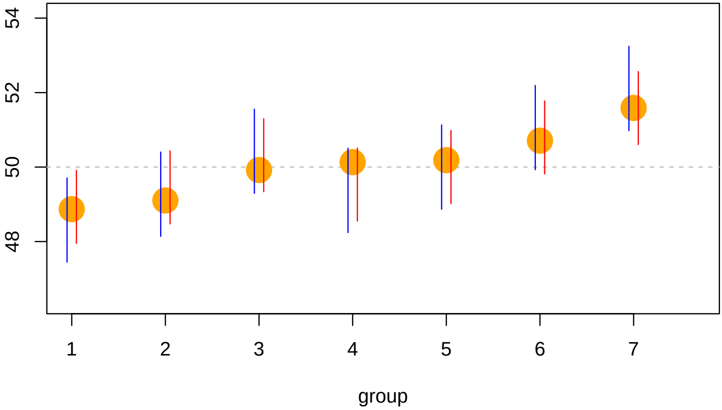

54
52
50
48
1
2
3
4
5
6
7
group

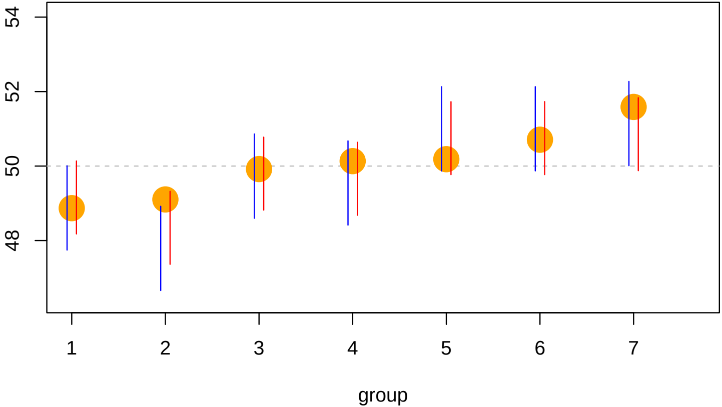

54
52
50
48
1
2
3
4
5
6
7
group


54
52
50
48
1
2
3
4
5
6
7
group

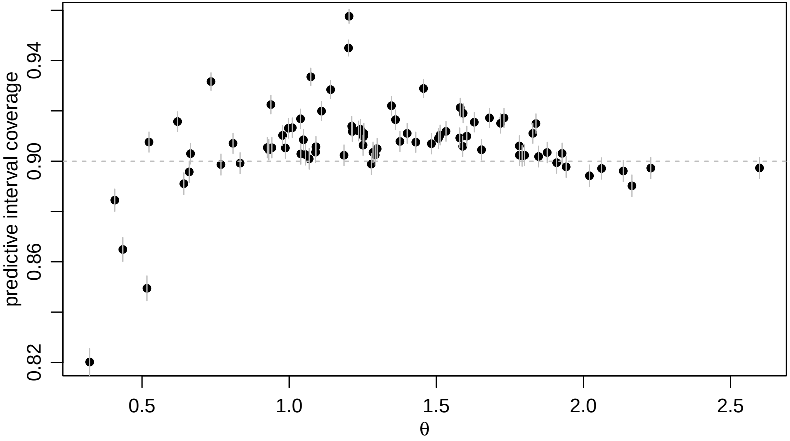

54
52
50
48
1
2
3
4
5
6
7
group


54
52
50
48
1
2
3
4
5
6
7
group


54
52
50
48
1
2
3
4
5
6
7
group


54
52
50
48
1
2
3
4
5
6
7
group

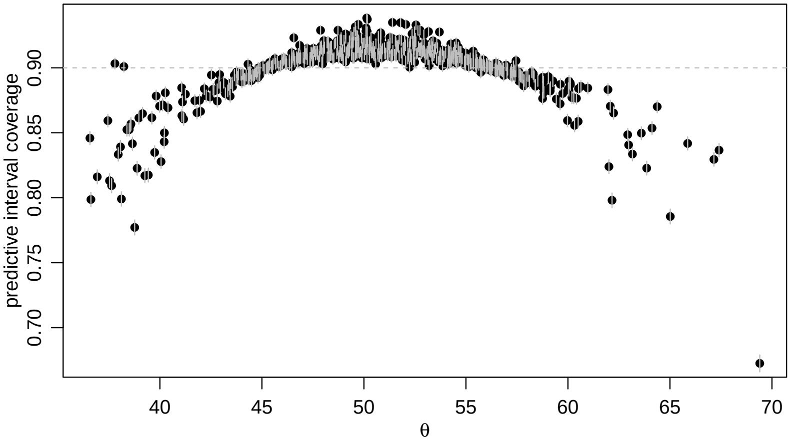
Nonconstant coverage: Radon data



$$\Pr(\theta_j \in C(\hat{\theta}_j)) \approx 1 - \alpha$$

$$\Pr(\theta_j \in C(\hat{\theta}_j)|\boldsymbol{\theta}) \quad \text{depends on } \theta_j.$$

Nonconstant coverage: ELS data



$$\Pr(\theta_j \in C(\hat{\theta}_j)) \approx 1 - \alpha$$

$$\Pr(\theta_j \in C(\hat{\theta}_j)|\boldsymbol{\theta}) \quad \text{depends on } \theta_j.$$

Comparing interval procedures

Interval widths:

- t-interval: $2 \times t_{1-\alpha/2} \times \hat{\sigma}/\sqrt{n}$
- EBayes interval: $2 \times t_{1-\alpha/2}/\sqrt{n/\hat{\sigma}^2 + 1/\tau^2}$

Exercise : Show $\hat{\sigma}/\sqrt{n} > 1/\sqrt{n/\hat{\sigma}^2 + 1/\tau^2}$

EBayes is always narrower, but

- t-interval is centered around high-variance unbiased estimator $\bar{y}_j$;
- EBayes-interval is centered around low-variance biased estimator $\hat{\theta}_j$;

This means coverage of EBayes will be

- higher than $1 - \alpha$ for groups near the center;
- lower than $1 - \alpha$ for groups away from the center.

Valid confidence intervals that share information

Goal: Construct confidence intervals $C^1, \ldots, C^p$ having

- constant coverage: $\Pr(\theta_j \in C^j(\mathbf{y})|\boldsymbol{\theta}) = 1 - \alpha$ for all groups/$\boldsymbol{\theta}$'s.
- improved precision: $\mathrm{E}[|C^j(\boldsymbol{y})|] < 2t_{1-\alpha/2}$ on average across groups/$\boldsymbol{\theta}$'s.

The first criterion is group-specific/frequentist - conditional on θ_j.

The second is study-specific/Bayes - on average across $\theta_1, \ldots, \theta_p$.

All CIPs

Standard procedure:

$$C_{1/2}(y) = \{\theta : y + \sigma z_{\alpha/2} < \theta < y + \sigma z_{1-\alpha/2}\}$$

Any procedure:

$$C_w(y) = \{\theta : y + \sigma z_{\alpha(1-w)} < \theta < y + \sigma z_{1-\alpha w}\}$$

In fact, w may depend on θ: If $w : \mathbb{R} \to [0, 1]$ then

$$C_w(y) = \{\theta : y + \sigma z_{\alpha(1-w(\theta))} < \theta < y + \sigma z_{1-\alpha w(\theta)}\}$$

satisfies $\Pr(\theta \in C_w(y)|\boldsymbol{\theta}) = 1 - \alpha$

- Examples in Bartholomew [1971], Stein [1962].
- Essentially complete class result in Yu and Hoff [2018].

FAB: Bayes-optimal frequentist interval

Simplified model:

- $y|\theta \sim N(\theta, \sigma^2)$, σ^2 known.
- $\pi(\theta)$ is prior information about θ.

Idea: Find the *w*-function that minimizes the prior expected width

$$\int\int |C_w(y)|\, p(dy|\theta)\pi(d\theta) < \int\int |C(y)|\, p(dy|\theta)\pi(d\theta)$$

Such an interval will have

- constant coverage, because C_w has constant coverage for any *w*-function;
- optimal precision on average with respect to π, by construction.

We call it FAB - Frequentist And Bayesian.

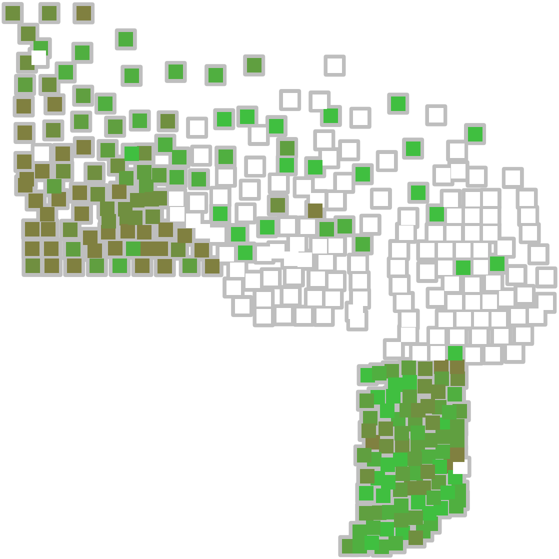
Adaptive FAB for multigroup inference

For each group $j = 1, \dots, p$:

1. Obtain $\hat{\mu}$, $\hat{\tau}^2$, $\hat{\sigma}^2$ using data from groups other than j;
2. Obtain $\hat{w}_j(\theta) = g^{-1}(2\hat{\sigma}(\theta - \hat{\mu})/\hat{\tau}^2)$;
3. Construct $C_{\hat{w}_j}(\bar{y}_j)$.

- Exact $1 - \alpha$ coverage *for each group*, even if hierarchical model is wrong.
- Improved precision *on average across groups*.

Radon data



Small area estimation (Burris and Hoff 2019)

Sampling model: $\bar{y}_j \sim N(\theta_j, \sigma_j^2)$ independently across groups.

Linking Model: $\theta_j = \boldsymbol{\beta}^\top \mathbf{x}_j + e_j$, $\text{Cov}[\boldsymbol{\theta}] = \Sigma$ (spatial FH model).

Direct interval: $\bar{y}_j \pm \hat{\sigma}_j t_{1-\alpha/2}$

AFAB interval: For each area $j = 1, \ldots, p$

1. using areas other than j, obtain estimates of $\boldsymbol{\theta}_{-j}$, $\boldsymbol{\beta}$ and Σ;
2. obtain "prior" distribution for θ_j from estimates and working model;
3. compute optimal w-function and construct FAB interval for θ_j.

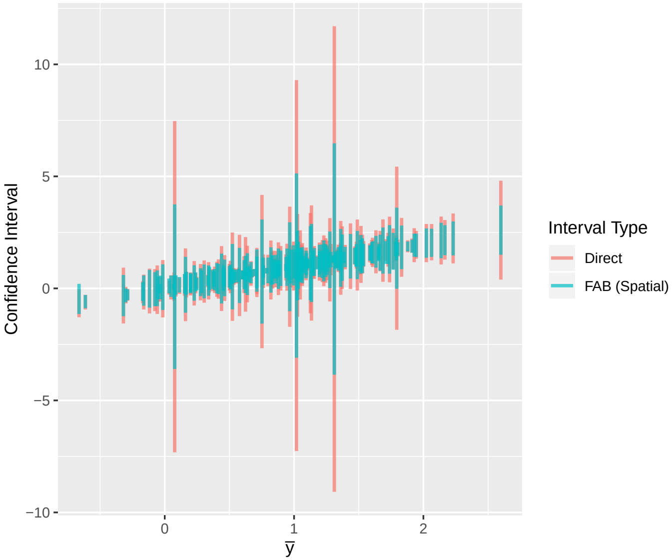
- Both intervals have $1-\alpha$ area-specific coverage, under random sampling within each area. The linking model need not be correct.
- FAB intervals make use of information from neighboring areas and known area-level characteristics (surficial radium).

Interval comparisons

Type	Hierarchical model	relative width	fraction intervals improved
Direct	-	1.0	-
FAB	exchangeable	.77	.898
FAB	covariate	.77	.888
FAB	spatial	.74	.964
FAB	spatial, covariate	.74	.955

By sharing information, hierarchical models can improve across-group performance, even if the hierarchical model is wrong.

Interval comparisons



Computing different intervals

```
y<-log(radon$radon)
g<-radon$county

tapply(y,g,mean)[1:20]

##     AITKIN      ANOKA     BECKER   BELTRAMI     BENTON   BIGSTONE  BLUEEARTH
##   4.293832   4.479973   4.675008   4.793035   4.869503   5.128199   5.522876
##      BROWN    CARLTON     CARVER       CASS   CHIPPEWA    CHISAGO       CLAY
##   5.244160   4.560494   4.971890   5.017782   5.349376   4.670860   5.402667
## CLEARWATER       COOK COTTONWOOD   CROWWING     DAKOTA      DODGE
##   4.609353   4.295244   4.577311   4.571230   4.917210   5.412986

table(g)[1:20]

## g
##     AITKIN      ANOKA     BECKER   BELTRAMI     BENTON   BIGSTONE  BLUEEARTH
##          4         52          3          7          4          3         14
##      BROWN    CARLTON     CARVER       CASS   CHIPPEWA    CHISAGO       CLAY
##          4         10          6          5          4          6         14
## CLEARWATER       COOK COTTONWOOD   CROWWING     DAKOTA      DODGE
##          4          2          4         12         63          3
```

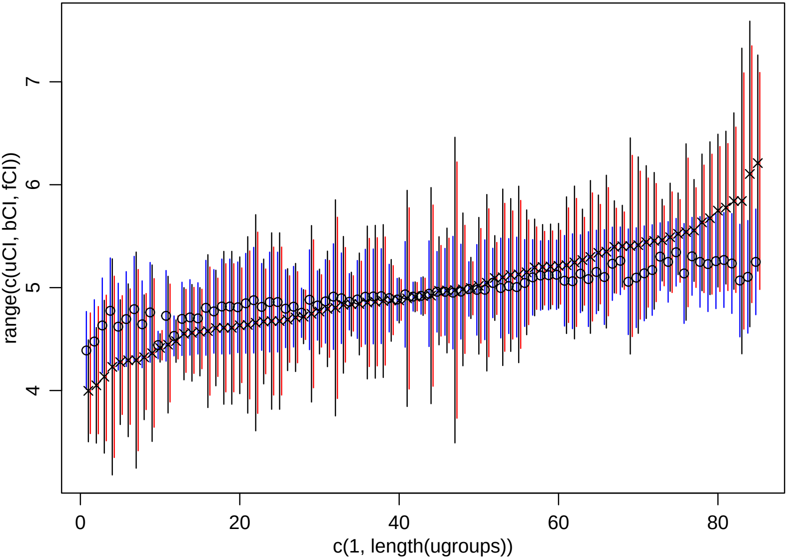
Computing different intervals

```
## unbiased intervals
fitLM<-lm(y ~ -1 + as.factor(g))
uCI<-confint(fitLM)

## EBayes intervals
library(lme4)
fitHM<-lmer(y ~ (1|g))
blupInfo<-as.data.frame(ranef(fitHM,condVar=TRUE))
bEst<-fixef(fitHM) + blupInfo[,4]
bSE<-blupInfo[,5]
bCI<-bEst + qnorm(.975)* outer( bSE ,c(-1,1))


## FAB intervals
library(FABInference)
fit<-lmFAB( y ~ -1,  model.matrix(~ -1+g) )
fCI<-fit$FABci
```

Comparing different intervals



Computing different intervals