- Often we wish to make predictions of quantities like $\underline{C\beta} + Du$ for some estimable $C\beta$.

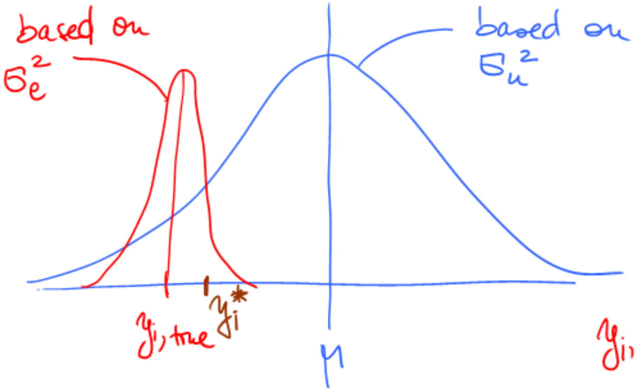
- The BLUP of such a quantity is $C\widehat{\beta}_{\Sigma} + D\widehat{u}$, the BLUE of $C\beta$ plus D times the BLUP of u.

$$\underbrace{\underset{\text{BLUE}}{C\widehat{\beta}_{\Sigma}} + \underset{\text{(E)BLUP}}{D\widehat{u}}}_{\text{(E)BLUP of } C\beta + Du}$$

(E) for empirical whenever we need to estimate our variance components first:

$u_1, \dots, u_n \sim N(0, \sigma_u^2)$ — variability in students reading ability scores

$e_1, \dots, e_n \sim N(0, \sigma_e^2)$ — variability in the actual test scores



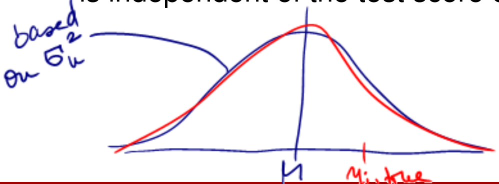
we observe a student's test score: y_i^*

$y_{i,true} = \mu + u_i$ — predict

- Suppose reading ability for a population of students are normally distributed with a mean μ and variance σ_u^2.

- Suppose a reading ability test was given to an i.i.d. sample of such students. sample from $N(\mu, \sigma_u^2)$

- Suppose that, given the true reading ability of a student at that time, the test score for that student is normally distributed with a mean equal to the student's reading ability and a variance σ_e^2 and is independent of the test score of any other student.



given the true μ for the student our observed test score comes from a $N(\mu, \sigma_e^2)$

- Suppose it is known that $\boxed{\sigma_u^2/\sigma_e^2} = 9$.

/ $\bar{y} = 86$

- If the sample mean of the students' test scores was 86, what is the best prediction of the reading ability of a student who scored 96 on the test?

the student's true reading ability

$\boxed{y_{i,\text{true}} = \mu + u_i}$ we observe

however $y_i^* = 96$

u & e are indep.

- Suppose $u_1, \ldots, u_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_u^2)$ independent of $e_1, \ldots, e_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_e^2)$.

- If we let $\mu + u_i$ denote the reading ability of student i $(i = 1, \ldots, n)$, then the reading ability scores of the students are $\mathcal{N}(\mu, \sigma_u^2)$ as in the statement of the problem.

- If we let $y_i = \mu + u_i + e_i$ denote the test score of student i $(i = 1, \ldots, n)$, then $(y_i | \mu + u_i) \sim \mathcal{N}(\mu + u_i, \sigma_e^2)$ as in the problem statement.

We have $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$, where

$$\boldsymbol{X} = \boldsymbol{1}, \quad \boldsymbol{\beta} = \mu, \quad \boldsymbol{Z} = \boldsymbol{I}, \quad \boldsymbol{G} = \sigma_u^2 \boldsymbol{I}, \quad \boldsymbol{R} = \sigma_e^2 \boldsymbol{I}, \quad \text{and}$$

$$\boldsymbol{\Sigma} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^\top + \boldsymbol{R} = \boxed{(\sigma_u^2 + \sigma_e^2)\boldsymbol{I}.}$$

Thus,

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}} = (\boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{X})^- \boldsymbol{X}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y} = (\boldsymbol{1}^\top \boldsymbol{1})^- \boldsymbol{1}^\top \boldsymbol{y} = \bar{y}.$$

in our ex.
observed
mean = 86

and

end lecture
36
4-25-25

$$\boldsymbol{G}\boldsymbol{Z}^\top \boldsymbol{\Sigma}^{-1} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \boldsymbol{I}.$$

Thus, the BLUP for $\boldsymbol{u}$ is

$$\widehat{\boldsymbol{u}} = \boldsymbol{G}\boldsymbol{Z}^{\top}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}}) = \boxed{\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}}(\boldsymbol{y} - \boldsymbol{1}\bar{y}_\cdot).$$

The i^{th} element of this vector is

$$\widehat{u}_i = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}(y_i - \bar{y}_\cdot).$$

Thus, the BLUP for $\mu + u_i$ (the reading ability score of student i) is

$$\widehat{\mu} + \widehat{u}_i = \bar{y}_\cdot + \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}(y_i - \bar{y}_\cdot) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}y_i + \frac{\sigma_e^2}{\sigma_u^2 + \sigma_e^2}\bar{y}_\cdot$$

Note that the BLUP is a convex combination of the individual score and the overall mean score.

$$\hat{\mu} + \hat{u}_i = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} y_i + \frac{\sigma_e^2}{\sigma_u^2 + \sigma_e^2} \bar{y}.$$

empirical BLUP

Because $\frac{\sigma_u^2}{\sigma_e^2}$ is assumed to be 9, the weights are

$$\frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = \frac{\frac{\sigma_u^2}{\sigma_e^2}}{\frac{\sigma_u^2}{\sigma_e^2} + 1} = \frac{9}{9+1} = 0.9$$

and

$$\frac{\sigma_e^2}{\sigma_u^2 + \sigma_e^2} = 0.1.$$

We would predict the IQ of a student who scored 130 on the test to be

$$0.9(96) + 0.1(86) = 95$$

EBLUPs in R for Seedling Dry Weight Example

2 genotype

```
> library(lme4)
>
> #Fit the linear mixed-effects model
> #with fixed genotype effects and
> #random tray effects.
>
> o = lmer(SeedlingWeight~Genotype+(1|Tray),data=d)
```

EBLUPs in R for Seedling Dry Weight Example

```
> #uhat is the vector of the EBLUPs of Tray effects.
>
> ranef(o)
$Tray
  (Intercept)
1   -4.985886
2    2.622632
3   -1.226723
4    3.589977
5    1.200572
6   -1.666317
7    3.104183
8   -2.638439
> uhat = ranef(o)$Tray[[1]]
```

EBLUPs in R for Seedling Dry Weight Example

$\hat{\mu}_i$ $\hat{u}_{ij}$

```
> #Get EBLUPs of genotype mean + tray effects
>
> betahat = fixef(o)
> betahat
(Intercept)   Genotype2
  15.288837   -3.550201
> estGeno1Mean = as.numeric(betahat[1])
> estGeno2Mean = as.numeric(betahat[1] + betahat[2])
> EBLUPs = c(estGeno1Mean + uhat[1:4],
+              estGeno2Mean + uhat[5:8])
```

$\hat{\mu}_1$

$\hat{\mu}_2 - \hat{\mu}_1 = -3.55$

EBLUPs in R for Seedling Dry Weight Example

```
> #Compare EBLUPs with tray averages.
> trayAverages=tapply(d$SeedlingWeight,d$Tray,FUN = mean)
> rbind(trayAverages, EBLUPs)
                1    2    3    4    5    6    7    8
trayAverages 10.0 18.0 14.0 19.0 13.0 10.0 15.0 9.00
EBLUPs       10.3 17.9 14.1 18.9 12.9 10.1 14.8 9.10
> estGeno1Mean
[1] 15.28884
> estGeno2Mean
[1] 11.73864
```

15.28884 − 4.985886

≈ 10.3

end lecture 37

4-28-25