

How to calculate the mean

KE Lotterhos

3/9/2021

```
require(lme4)

## Loading required package: lme4
## Loading required package: Matrix
require(emmeans)

## Loading required package: emmeans
## Welcome to emmeans.
## NOTE -- Important change from versions <= 1.41:
##     Indicator predictors are now treated as 2-level factors by default.
##     To revert to old behavior, use emm_options(cov.keep = character(0))

# Different ways of calculating means ####
getmeans <- function(ind=1:length(a), true){

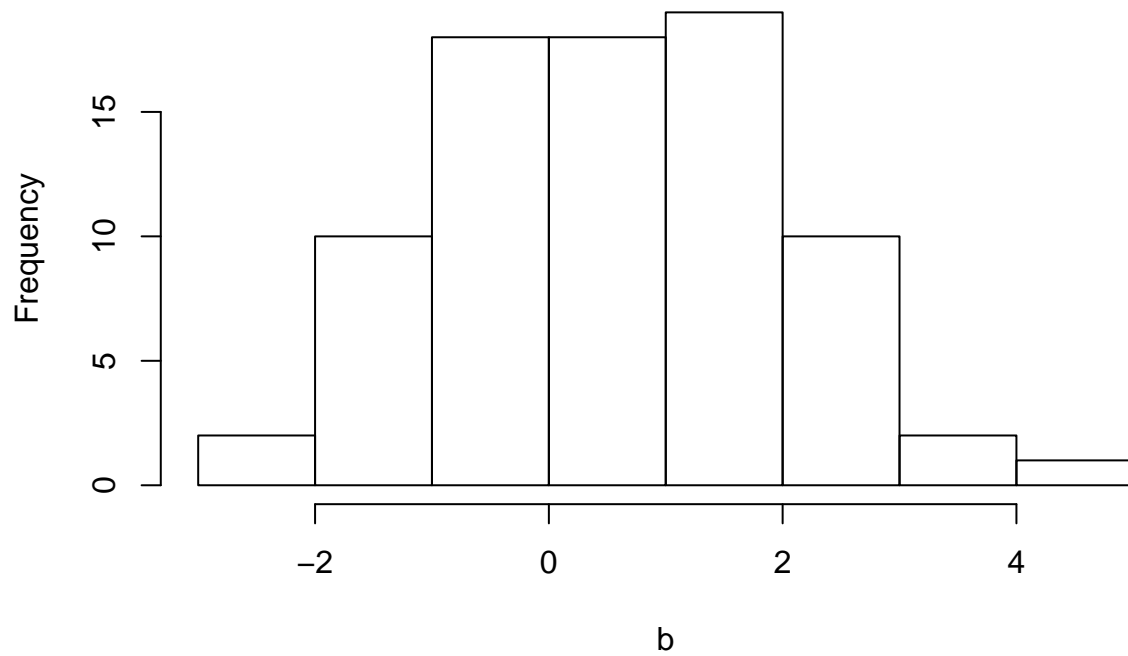
  # model fit
  G <- as.factor(G_lev[ind])
  E <- as.factor(E_lev[ind])
  m1 <- lm(b[ind] ~ G * E)
  emms <- as.data.frame(emmeans(m1, ~ G * E))
  emmG <- as.data.frame(emmeans(m1, "G"))
  emmE <- as.data.frame(emmeans(m1, "E"))
  obs_all <- mean(b[ind]) # This may or may not be what your eMM does
  obs_mean_GE <- mean(tapply(b[ind],a[ind], mean))
  obs_mean_G <- mean(tapply(b[ind],G_lev[ind], mean))
  obs_mean_E <- mean(tapply(b[ind],E_lev[ind], mean))
  obs_mean_cGE <- mean(c(obs_mean_G, obs_mean_E)) # Note they count unequally in the common garden N x
  EMM_mean_GE <- mean(emms$emmean)
  EMM_mean_G <- mean(emmG$emmean)
  EMM_mean_E <- mean(emmE$emmean)
  EMM_mean_cGE <- mean(c(EMM_mean_E,EMM_mean_G))
  out <- cbind(obs_all, obs_mean_GE, obs_mean_G,obs_mean_E, obs_mean_cGE, EMM_mean_GE, EMM_mean_G, EMM_mean_E, EMM_mean_cGE)
  round(out - true, 4)
}
```

consider 2x2 case

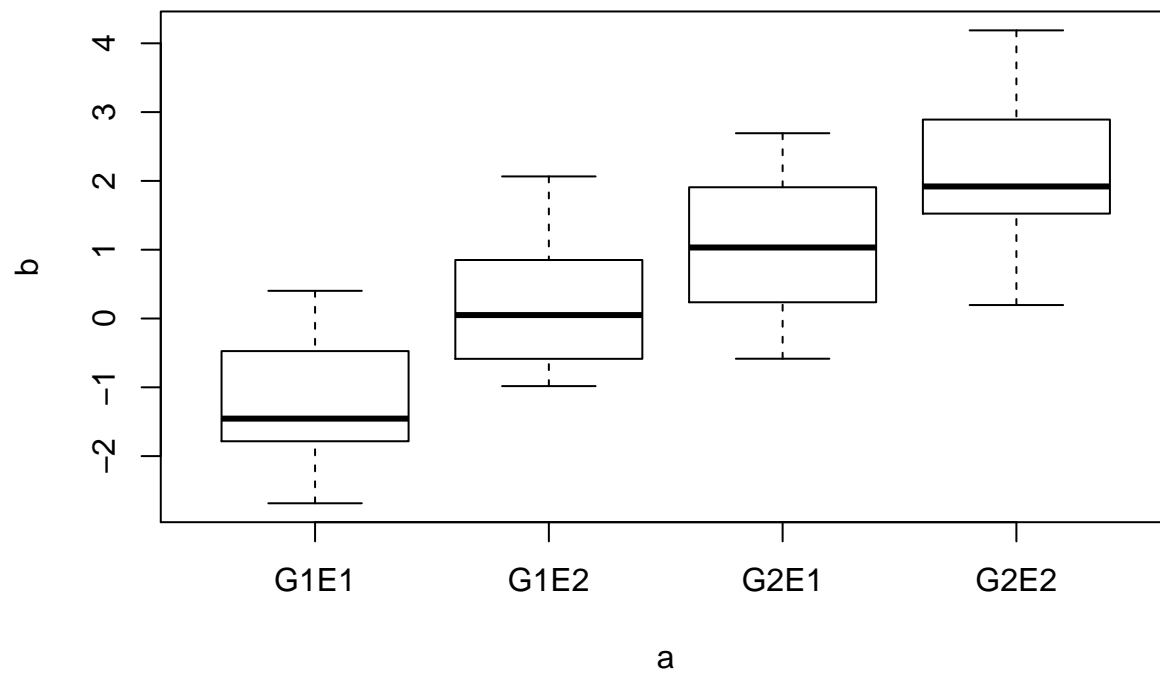
First, consider the 2x2 case, with different GE means for each level. Compare results for imbalance in the level with the lowest mean, and imbalance in the level with the highest mean.

```
a <- rep(c("G1E1","G1E2", "G2E1", "G2E2"), each=20)
G_lev <- substr(a, start=1, stop=2)
E_lev <- substr(a, start=3, stop=4)
b <- c(rnorm(length(a)/4, -1), rnorm(length(a)/4, 0), rnorm(length(a)/4, 1), rnorm(length(a)/4, 2))
hist(b)
```

Histogram of b



```
boxplot(b~a)
```



```
true1 <- mean(c(-1,0,1,2)) # true mean
getmeans(1:length(a), true1) # no imbalance
```

```
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
##      obs_all obs_mean_GE obs_mean_G obs_mean_E obs_mean_cGE EMM_mean_GE
```

```
getmeans(15:length(a), true1) # imbalance in first category
```

```
getmeans(1:(length(a)-15), true1) # imbalance in last category
```

```
par(mar=c(10,3,1,1))
```

3

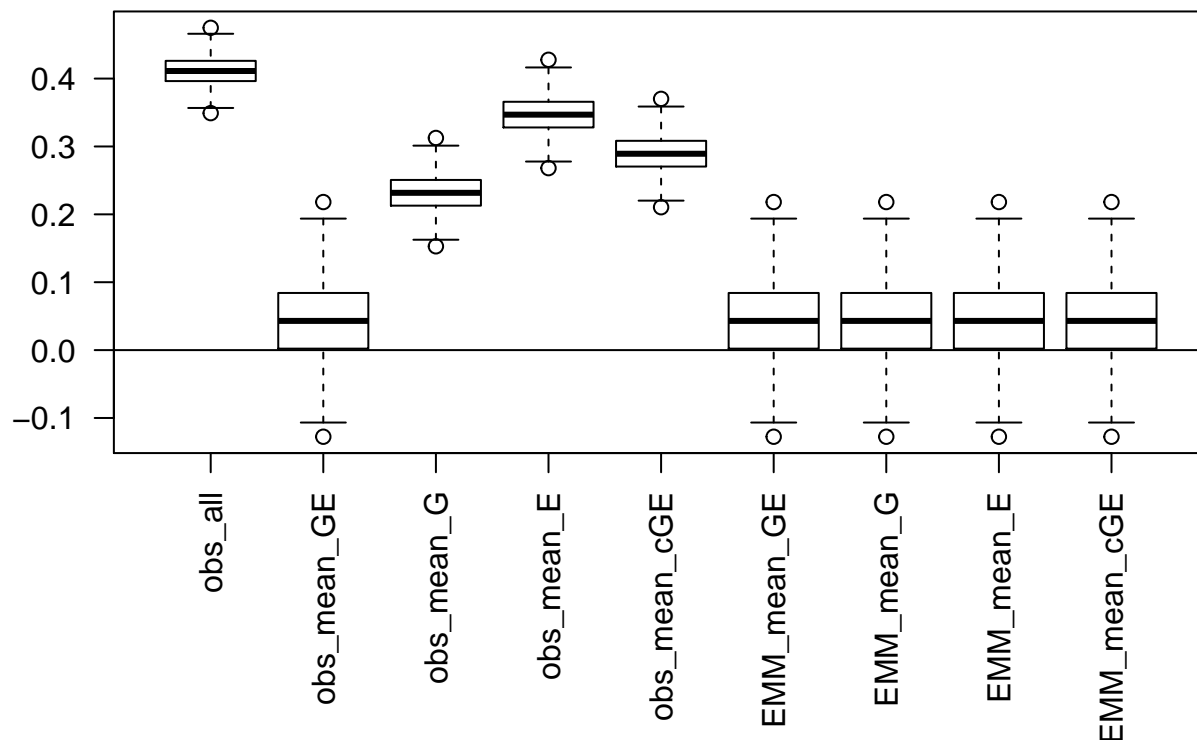
[illegible]

[illegible]

[illegible]

```
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
```

```
rownames(bob) <- c("obs_all", "obs_mean_GE", "obs_mean_G", "obs_mean_E", "obs_mean_cGE", "EMM_mean_GE",
#head(bob)
boxplot(t(bob), las=2)
abline(h=0)
```



```
bob <- replicate(100, getmeans(c(1:(length(a)-20), sample((length(a)-20):length(a), 5)), true1), simplify = FALSE)
```

```
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
```


[illegible]

[illegible]

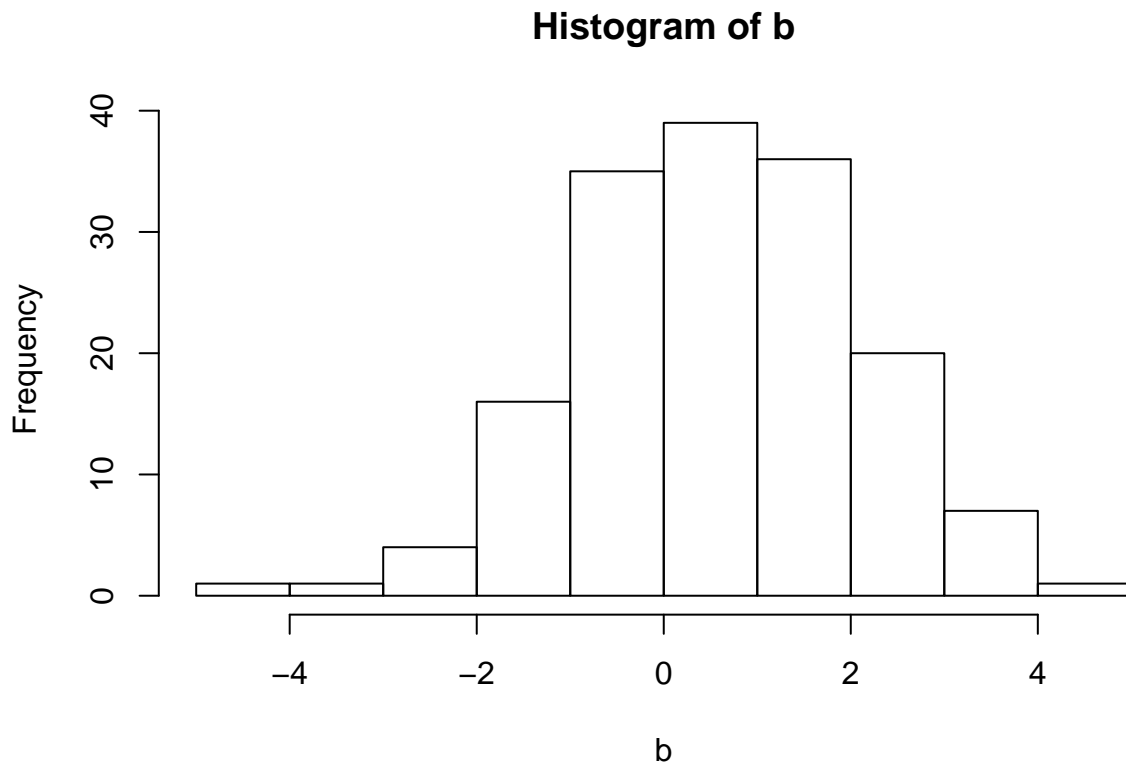
```
#head(bob)
rownames(bob) <- c("obs_all", "obs_mean_GE", "obs_mean_G", "obs_mean_E", "obs_mean_cGE", "EMM_mean_GE",
#head(bob)
boxplot(t(bob), las=2)
abline(h=0)
```



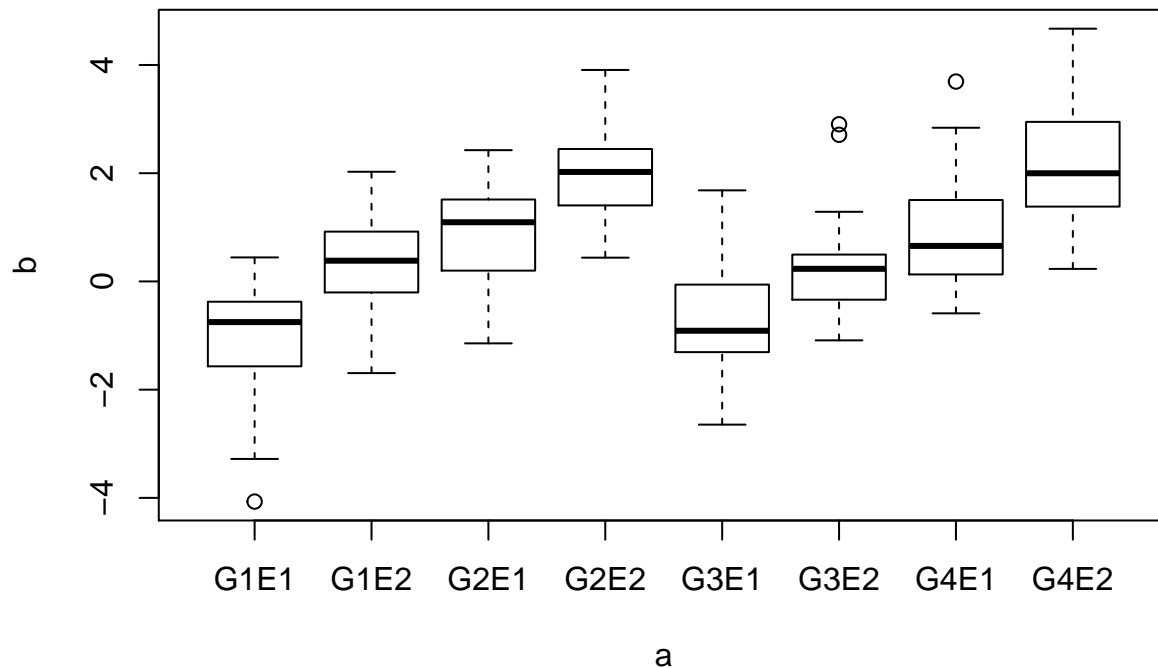
consider 2x4 common garden case

First, consider the 2x4 case, with different GE means for each level. Compare results for imbalance in the level with the lowest mean, and imbalance in the level with the highest mean.

```
a <- rep(c("G1E1", "G1E2", "G2E1", "G2E2", "G3E1", "G3E2", "G4E1", "G4E2"), each=20)
G_lev <- substr(a, start=1, stop=2)
E_lev <- substr(a, start=3, stop=4)
b <- c(rnorm(length(a)/8, -1),
      rnorm(length(a)/8, 0),
      rnorm(length(a)/8, 1),
      rnorm(length(a)/8, 2),
      rnorm(length(a)/8, -1),
      rnorm(length(a)/8, 0),
      rnorm(length(a)/8, 1),
      rnorm(length(a)/8, 2))
hist(b)
```



```
boxplot(b~a)
```



```
par(mar=c(10,3,1,1))
true2 <- mean(c(-1,0,1,2, -1, 0, 1,2,-1,0,1,2)) # true mean
bob <- replicate(100,getmeans(c(sample(1:20,5),20:length(a)), true2),
                  simplify = TRUE)
```

[illegible]

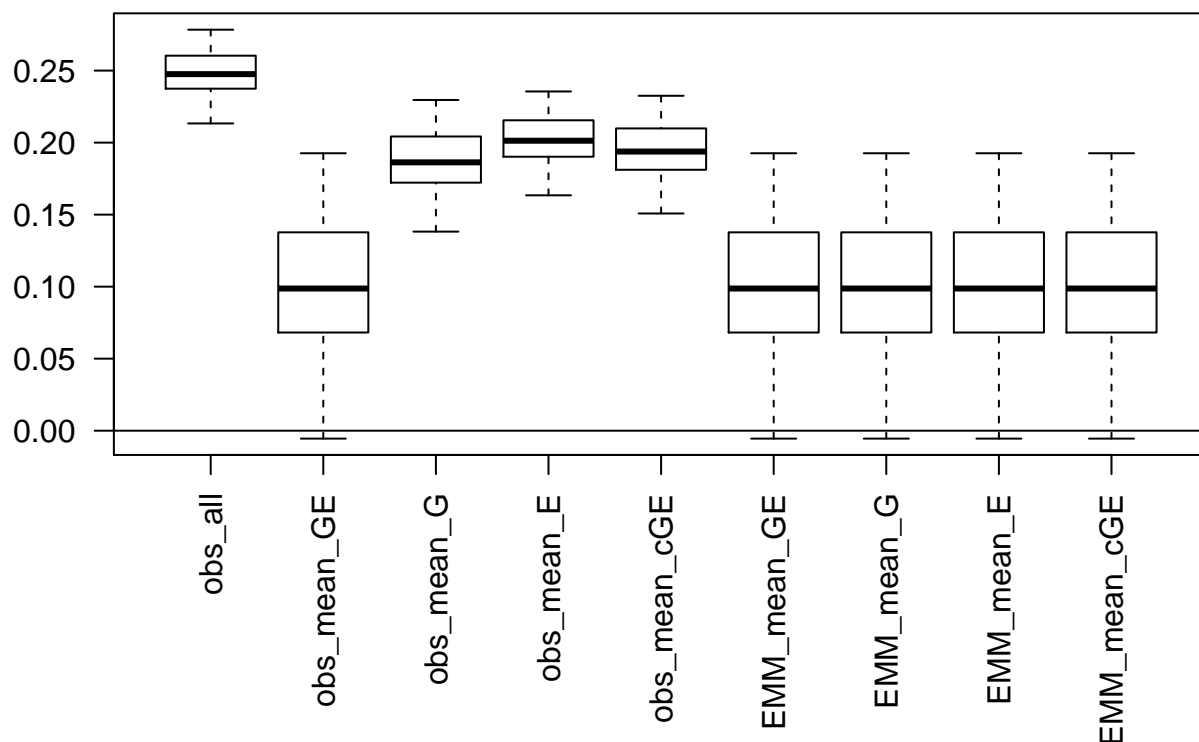
[illegible]

[illegible]

[illegible]


```
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
```

```
rownames(bob) <- c("obs_all", "obs_mean_GE", "obs_mean_G", "obs_mean_E", "obs_mean_cGE", "EMM_mean_GE",
#head(bob)
boxplot(t(bob), las=2)
abline(h=0)
```



```
bob <- replicate(100, getmeans(c(1:(length(a)-20),
                                sample((length(a)-20):length(a), 5)), true2),
                simplify = TRUE)
```

```
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
## NOTE: Results may be misleading due to involvement in interactions
```

[illegible]

[illegible]

[illegible]

```
rownames(bob) <- c("obs_all", "obs_mean_GE", "obs_mean_G", "obs_mean_E", "obs_mean_cGE", "EMM_mean_GE",
#head(bob)
boxplot(t(bob), las=2)
abline(h=0)
```

