bootstrap

XC

Create bootstrap samples

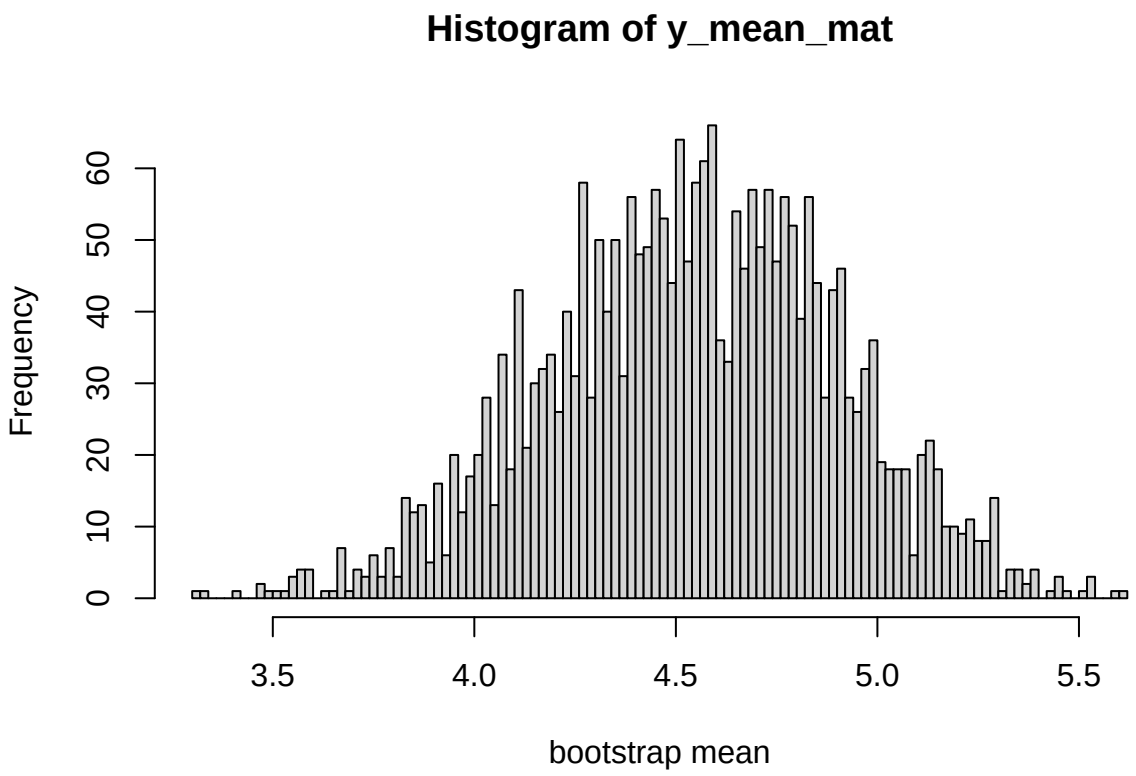
```
set.seed(12-07-2021)

y <- rgamma(8, shape = 4, rate = 1)          # generate samples from Gamma(4, 1)

nBoots <- 2500
y_B <- list()
for(i in 1:nBoots) {
  y_B[[i]] <- sample(y, replace = T)           # bootstrap with replacement
}
```

```
y_mean <- lapply(y_B, function(x) mean(x))
```

```
y_mean_mat <- do.call("rbind", y_mean)
str(y_mean_mat)
```

```
##  num [1:2500, 1] 4.81 5.3 4.38 4.31 4.28 ...
```

```
hist(y_mean_mat, breaks = 100, xlab = "bootstrap mean")
```



Example: boostrap to simple linear regression

Model: $Y_{ij} = \alpha + \beta x_i + \epsilon_{ij}$ * two parameters: α intercept, β slope * $\epsilon_{ij} \sim^{iid} N(0, \sigma^2)$ * we fit the linear model and get $\hat{\alpha}, \hat{\beta}$ * residuals from the fitted model: $\hat{\epsilon_{ij}} = y_{ij} - \hat{\alpha} - \hat{\beta} x_i$ * we bootstrap the residuals to get $\hat{\epsilon_{ij}}*$ * we get new bootstrapped data: $y_{ij}* = y_{ij} + \hat{\epsilon_{ij}}*$

```
x <- seq(-3, 3, length.out = 5)     # num[1:5]
y <- 2 +  4 * x + rnorm(5)          # num[1:5]

fit <- lm(y ~ x)
res <- fit$residuals
fit$coefficients
```

```
## (Intercept)           x
##    1.663603    3.974648
```

```
nBoots <- 2000
coeff_bt <- array(0, dim = c(nBoots, 2))  # 2 parameters alpha, beta
for(i in 1:nBoots) {
  y_bt <- y + sample(res, replace = T)
  fit_bt <- lm(y_bt ~ x)
  coeff_bt[i, ] <- fit_bt$coefficients
}
```

```
par(mfrow = c(1, 2))
hist(coeff_bt[, 1], xlab = "Boot_intercept", breaks = 50)
abline(v = fit$coefficients[1], lwd = 2)
```

```r
abline(v = quantile(coeff_bt[, 1], 0.025), lty = 2)
abline(v = quantile(coeff_bt[, 1], 0.975), lty = 2)


hist(coeff_bt[, 2], xlab = "Boot_slope", breaks = 50)
abline(v = fit$coefficients[2], lwd = 2)
abline(v = quantile(coeff_bt[, 2], 0.025), lty = 2)
abline(v = quantile(coeff_bt[, 2], 0.975), lty = 2)
```