STOR 655 HW 10

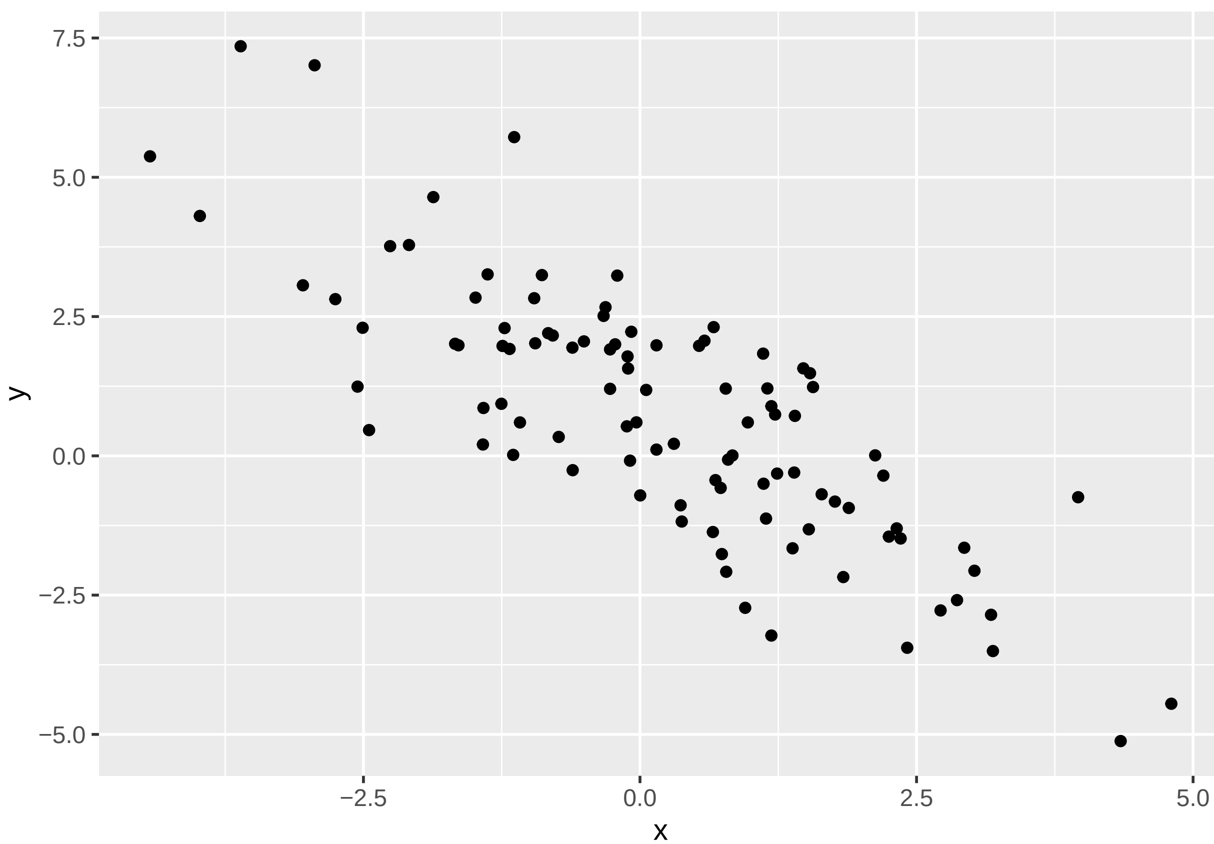
Brian N. White

4/21/2021

Problem 2 Part B (Bootstrap)

The model in question is $Y_i = \alpha + \beta X_i + e_i$ where I assume $e \sim N(0, \sigma^2)$ iid. Let $\theta = (\alpha, \beta)^T$

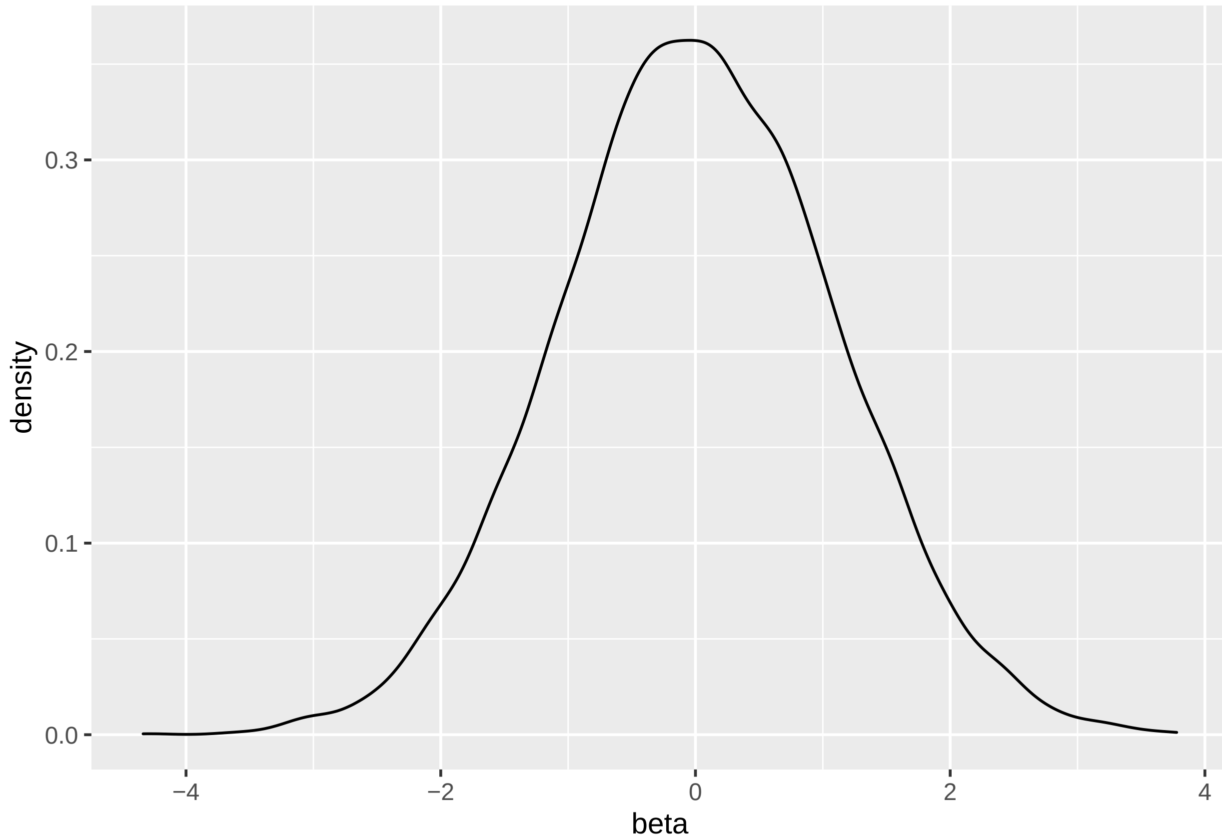
```
#import data and examine relationship between x & y
data <- read.csv("HW10.csv", row.names=1)
data %>%
  ggplot(aes(x=x, y=y)) +
  geom_point()
```




```
#for reference
lm_data <- lm(y~x, data)
```

```r
#sample size
n <- length(data$x)

#theta_hat computation
X <- matrix(c(rep(1, 100), data$x), nrow=100, ncol=2)
theta_hat <- solve(t(X)%*%X)%*%t(X)%*%data$y

#residuals
resid <- data$y-X%*%theta_hat

#unbiased estimator of variance of the random errors
sigma_hat <- sum(resid^2)/(n-2)

#estimate of variance of the random errors
var_theta_hat <- sigma_hat*solve(t(X)%*%X)
```

```r
#dummy vector to store output of bootstrap
boot <- vector()

#sample with replacement from original data
nboot <- 10000
for(k in 1:nboot){
 data.sample <- data %>%
    sample_n(100, replace = T)

 x_star <- data.sample$x
 y_star <- data.sample$y

  X_star <- matrix(c(rep(1, 100), x_star), nrow=100, ncol=2)
  theta_hat_star <- solve(t(X_star)%*%X_star)%*%t(X_star)%*%y_star
  resid_star <- y_star-X_star%*%theta_hat_star
  sigma_hat_star <- sum(resid_star^2)/(n-2)
  var_theta_hat_star <- sigma_hat_star*solve(t(X_star)%*%X_star)

  #z_k (i.e. normalized beta estimate for kth bootstrap sample)
  boot[k] <- (theta_hat_star[2]-theta_hat[2])/sqrt(var_theta_hat_star[2, 2])
}

#plot of bootstrap sampling distribution of normalized beta
data.frame(beta=boot) %>%
  ggplot(aes(beta)) +
  geom_density()
```




```
theta_hat[2]-sqrt(var_theta_hat)[2,2]*quantile(boot, c(.975, .025))
```

```
## Warning in sqrt(var_theta_hat): NaNs produced
```

```
##      97.5%       2.5%
## -1.1874239 -0.8582049
```

```
confint(lm_data)[2,]
```

```
##      2.5 %     97.5 %
## -1.1729667 -0.8676194
```