Exercise #3

PLSC 501, spring 2024

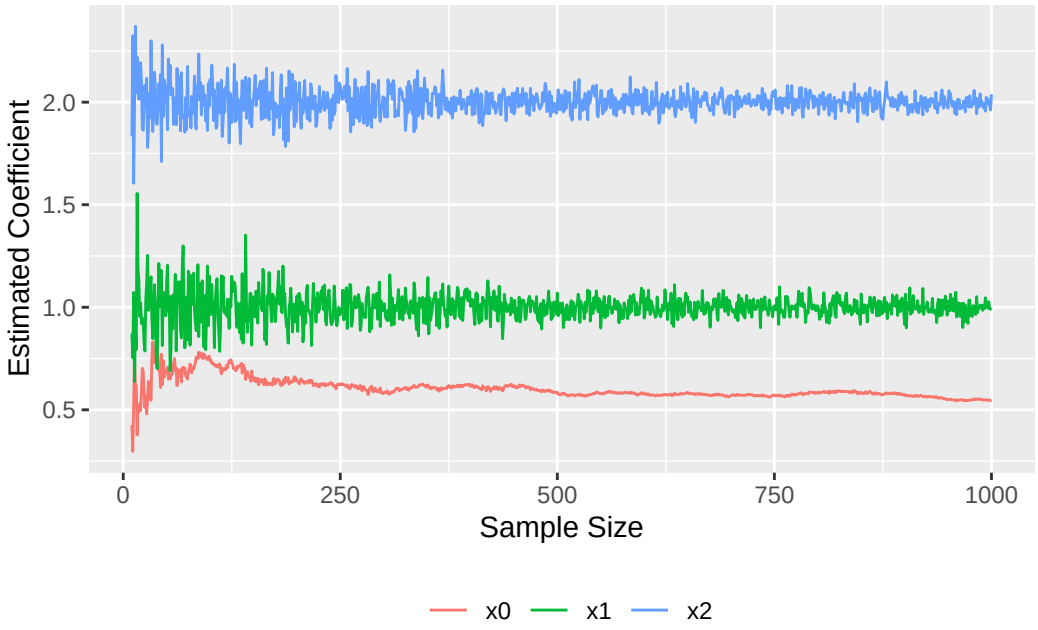
Dave Clark

February 13, 2024

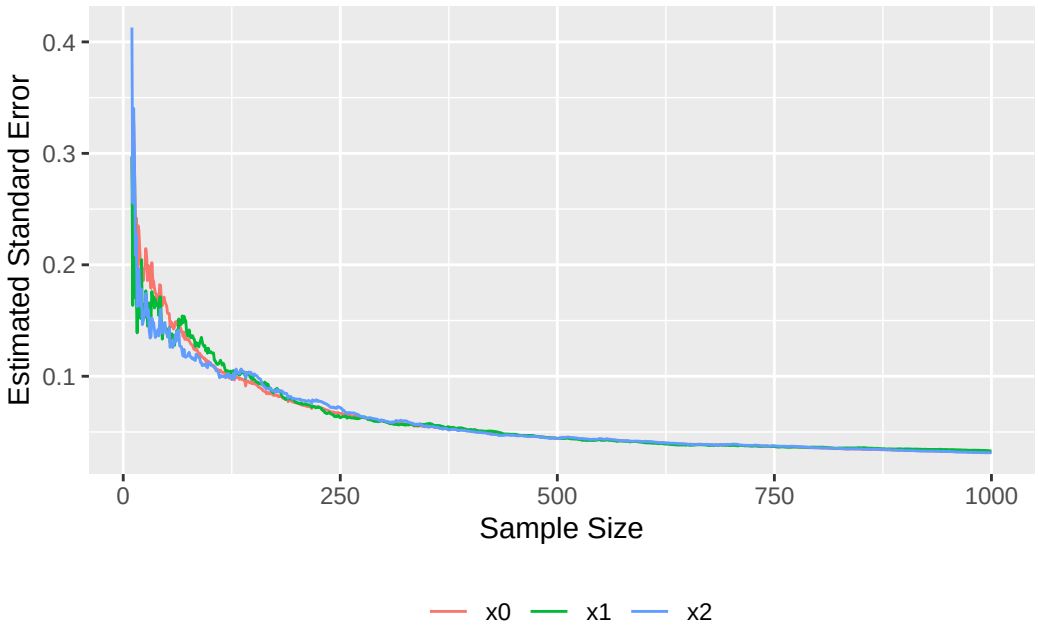
Q1 - sample size simulation

```
results <- data.frame ( var = character (0), coef = numeric(0), se = numeric (0), n = nume

for(i in seq(10,1000,1)) {
  set.seed(12345)
  data <- tibble(
    X <- rnorm_multi(i, 3,
                     mu=c(0, 0, 0),
                     sd=1,
                     r = c(0.0, 0.0, 0.0),
                     varnames=c("x1", "x2", "e"))
  ) %>%
    mutate(y = .5 + 1*x1 + 2*x2 + e)

  mod <- (lm(y ~ x1 + x2, data=data))

  results[((i-9)*3-2):((i-9)*3),1:4] <- data.frame ( var = c("x0","x1","x2"), coef(summary

}

results <-data.frame(results, t=results$coef/results$se)


p1 <- ggplot(results, aes(x=n, y=coef, color=var)) +
  geom_line() +
  labs ( colour = NULL, x = "Sample Size", y =  "Estimated Coefficient" ) +
  theme ( legend.position = "bottom",
```

```
          legend.key = element_blank() )
p1
```




```
p2 <- ggplot(results, aes(x=n, y=se, color=var)) +
  geom_line() +
  labs ( colour = NULL, x = "Sample Size", y =  "Estimated Standard Error" ) +
  theme ( legend.position = "bottom",
          legend.key = element_blank() )
p2
```

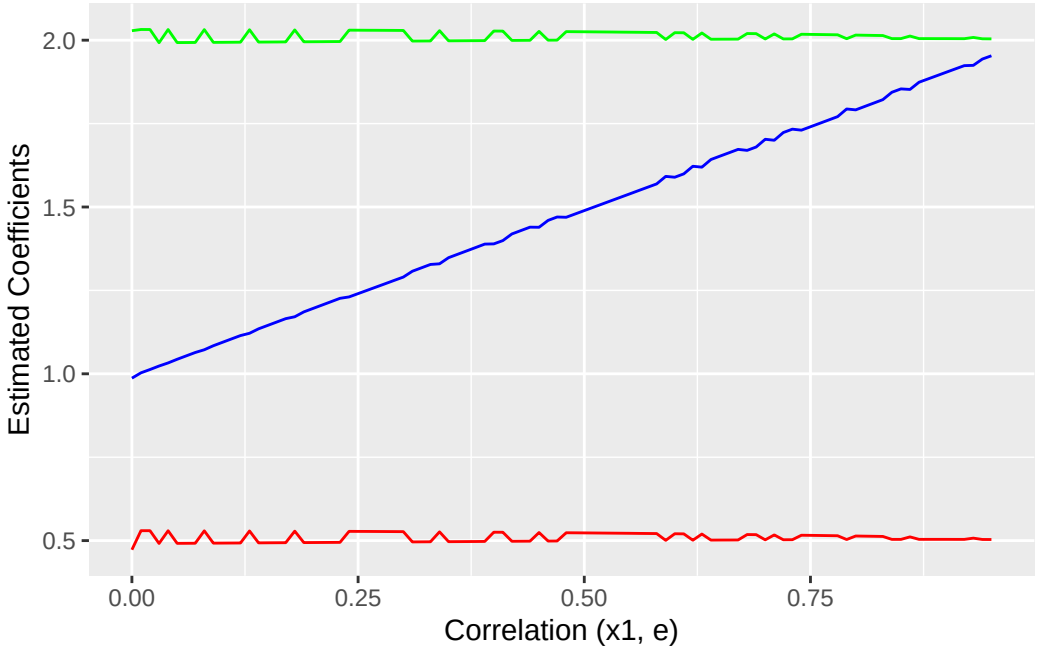


Q2 - correlation simulation

rho(x,e) —-

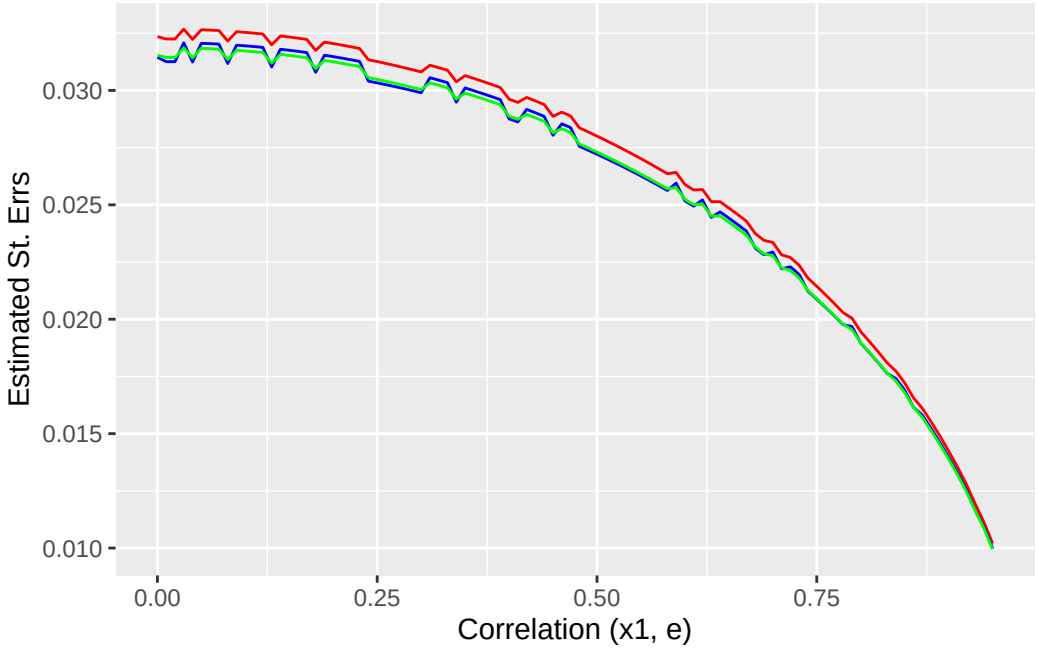
```
results <- data.frame()
i=1
for(r in seq(0, .95, .01)) {
  set.seed(8675309)
  X <- rnorm_multi(1000, 3,
                   mu=c(0, 0, 0),
                   sd=1,
                   r = c(0, r, 0),
                   varnames=c("x1", "x2", "e"))
  y = .5 + X$x1 + 2*X$x2 + X$e
  data <- data.frame(X, y)
  m <- lm(y ~ x1 + x2, data=data)

  b0<-coef(summary(m))[1,1]
  b1<-coef(summary(m))[2,1]
  b2<-coef(summary(m))[3,1]
  se0<-coef(summary(m))[1,2]
  se1<-coef(summary(m))[2,2]
```

```
  se2<-coef(summary(m))[3,2]
  results[i,1:7] <- data.frame(rho = r, b0, b1, b2, se0, se1, se2)

  i = i+1
}

b <- ggplot() +
  geom_line(data=results, aes(x=rho, y=b0), color="red")+
  geom_line(data=results, aes(x=rho, y=b1), color="blue")+
  geom_line(data=results, aes(x=rho, y=b2), color="green")+
  labs ( colour = NULL, x = "Correlation (x1, e)", y =  "Estimated Coefficients" ) +
  theme ( legend.position = "bottom",
          legend.key = element_blank() )
b
```




```
se <- ggplot() +
  geom_line(data=results, aes(x=rho, y=se0), color="red")+
  geom_line(data=results, aes(x=rho, y=se1), color="blue")+
  geom_line(data=results, aes(x=rho, y=se2), color="green") +
  labs ( colour = NULL, x = "Correlation (x1, e)", y =  "Estimated St. Errs" ) +
  theme ( legend.position = "bottom",
          legend.key = element_blank() )
```

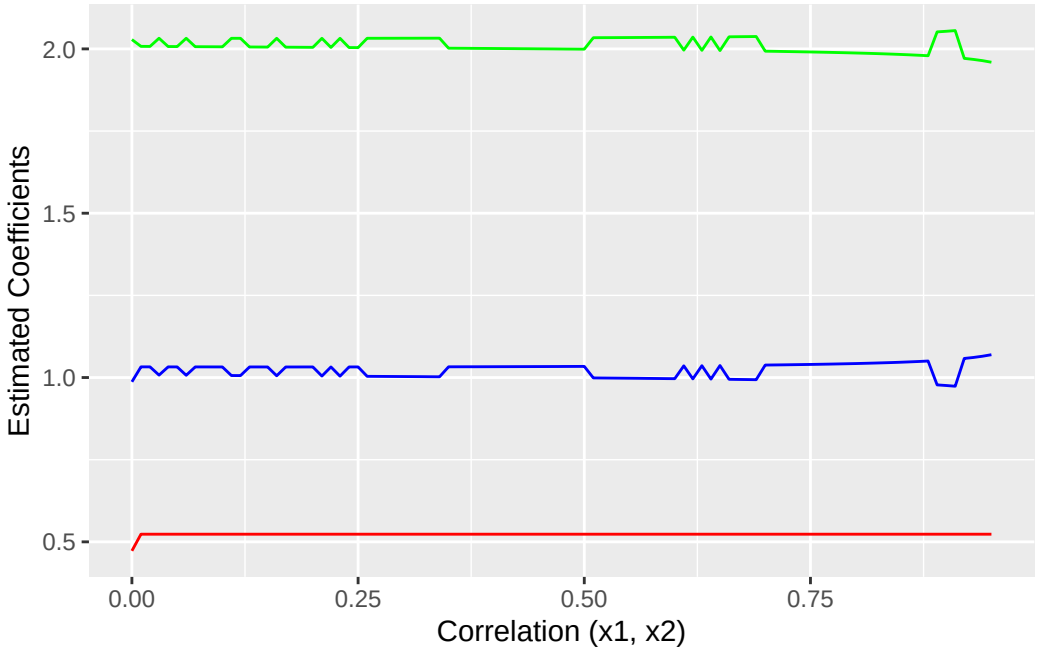
```
se
```



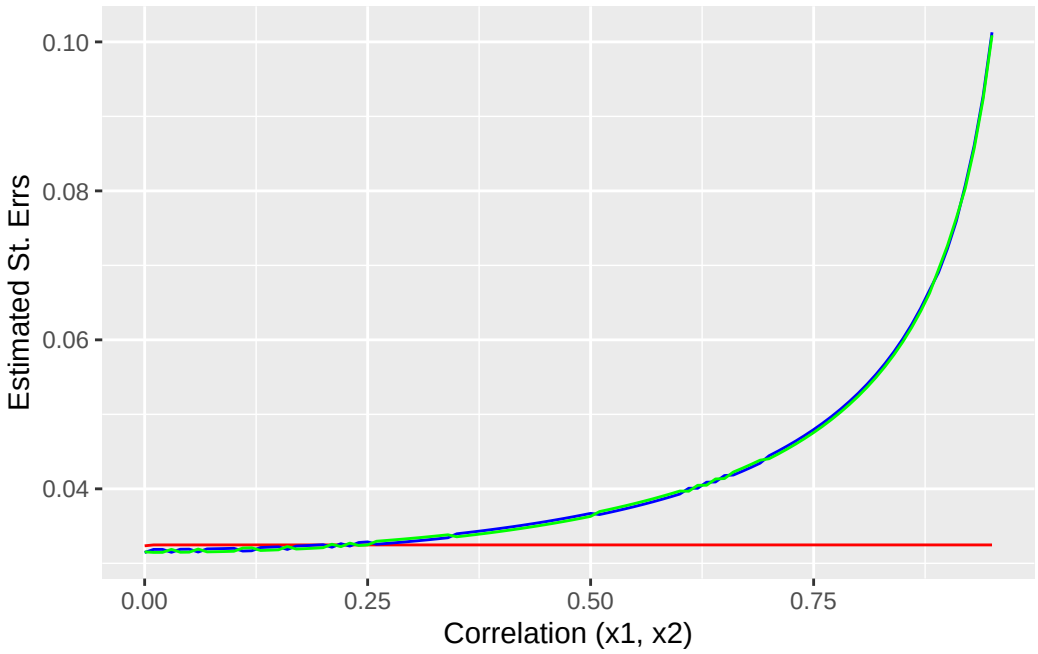
rho(x1,x2) —-

```
results <- data.frame()
i=1
for(r in seq(0, .95, .01)) {
  set.seed(8675309)
  X <- rnorm_multi(1000, 3,
                   mu=c(0, 0, 0),
                   sd=1,
                   r = c(r, 0, 0),
                   varnames=c("x1", "x2", "e"))
  y = .5 + X$x1 + 2*X$x2 + X$e
  data <- data.frame(X, y)
  m <- lm(y ~ x1 + x2, data=data)
  i = i+1
  b0<-coef(summary(m))[1,1]
  b1<-coef(summary(m))[2,1]
  b2<-coef(summary(m))[3,1]
  se0<-coef(summary(m))[1,2]
```

```
  se1<-coef(summary(m))[2,2]
  se2<-coef(summary(m))[3,2]
  results[i,1:7] <- data.frame(rho = r, b0, b1, b2, se0, se1, se2)
}

b <- ggplot() +
  geom_line(data=results, aes(x=rho, y=b0), color="red")+
  geom_line(data=results, aes(x=rho, y=b1), color="blue")+
  geom_line(data=results, aes(x=rho, y=b2), color="green")+
  labs ( colour = NULL, x = "Correlation (x1, x2)", y =  "Estimated Coefficients" ) +
  theme ( legend.position = "bottom",
          legend.key = element_blank() )
b
```




```
se <- ggplot() +
  geom_line(data=results, aes(x=rho, y=se0), color="red")+
  geom_line(data=results, aes(x=rho, y=se1), color="blue")+
  geom_line(data=results, aes(x=rho, y=se2), color="green") +
  labs ( colour = NULL, x = "Correlation (x1, x2)", y =  "Estimated St. Errs" ) +
  theme ( legend.position = "bottom",
          legend.key = element_blank() )
se
```


0.10
0.08
0.06
0.04
Estimated St. Errs
0.00
0.25
0.50
0.75
Correlation (x1, x2)