In Review: Inference for Linear Models

Problem Set Debrief

Math 392

Inference in MLR

1. $Var(\hat{\beta})$

$$\begin{aligned}
Var(\hat{\beta}|X) &= Var((X'X)^{-1}X'Y|X) \\
&= (X'X)^{-1}X'Var(Y|X)X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}
\end{aligned}$$

Inference in MLR, cont.

2. $Var(\hat{E}(Y|X))$

$$\begin{aligned} Var(\hat{E}(Y|X = x_s)) &= Var(x_s\hat{\beta}) \\ &= x_s Var(\hat{\beta}) x_s' \\ &= x_s \sigma^2 (X'X)^{-1} x_s' \end{aligned}$$

Inference in MLR, cont.

3. $Var(Y|X = x_s))$

$$\begin{aligned} Var(Y|X = x_s) &= Var(x_s\hat{\beta} + \epsilon) \\ &= Var(x_s\hat{\beta}) + Var(\epsilon) \\ &= x_s\sigma^2(X'X)^{-1}x_s' + \sigma^2 \end{aligned}$$

Marginal distribution of the Error

$$\epsilon \sim N(0, \sigma^2 I)$$

```r
epsilon <- rnorm(n, mean = 0, sd = sigma_sq)
```

Change the marginal distribution of the ϵ (though it still should be centered at 0).

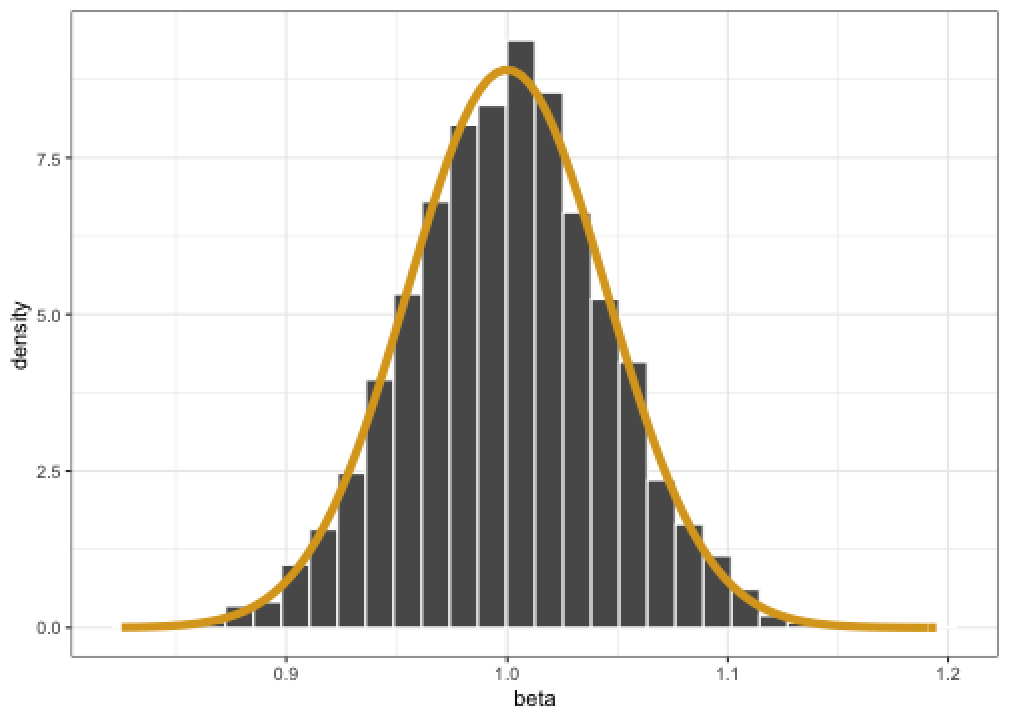
$$\epsilon \sim Unif(-1, 1)$$

```r
epsilon <- runif(n, -1, 1)
```

$$\epsilon \sim Lap(0, b)$$

```r
library(rmutil)
epsilon <- rlaplace(n, m = 0, s = sqrt((1/2) * sigma_sq))
```

$\hat{\beta}$

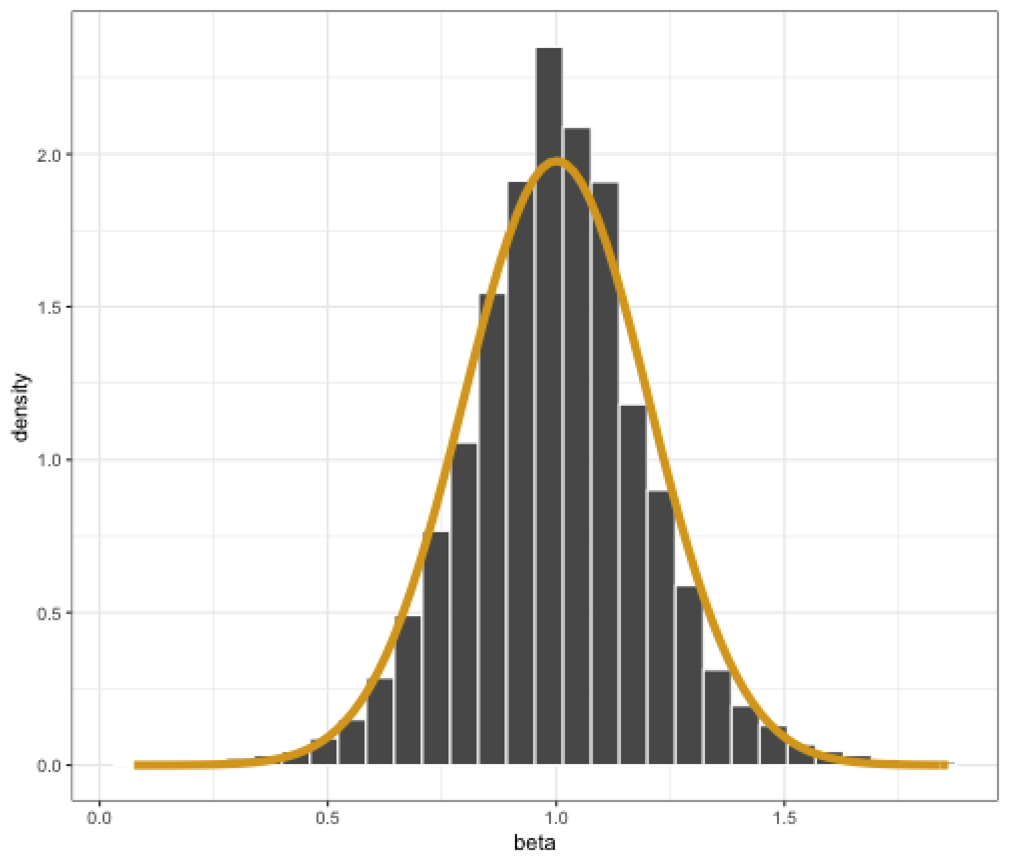


1. Is the variance still $\sigma^2(X'X)^{-1}$?
2. Is the distribution of $\hat{\beta}$ still normal?

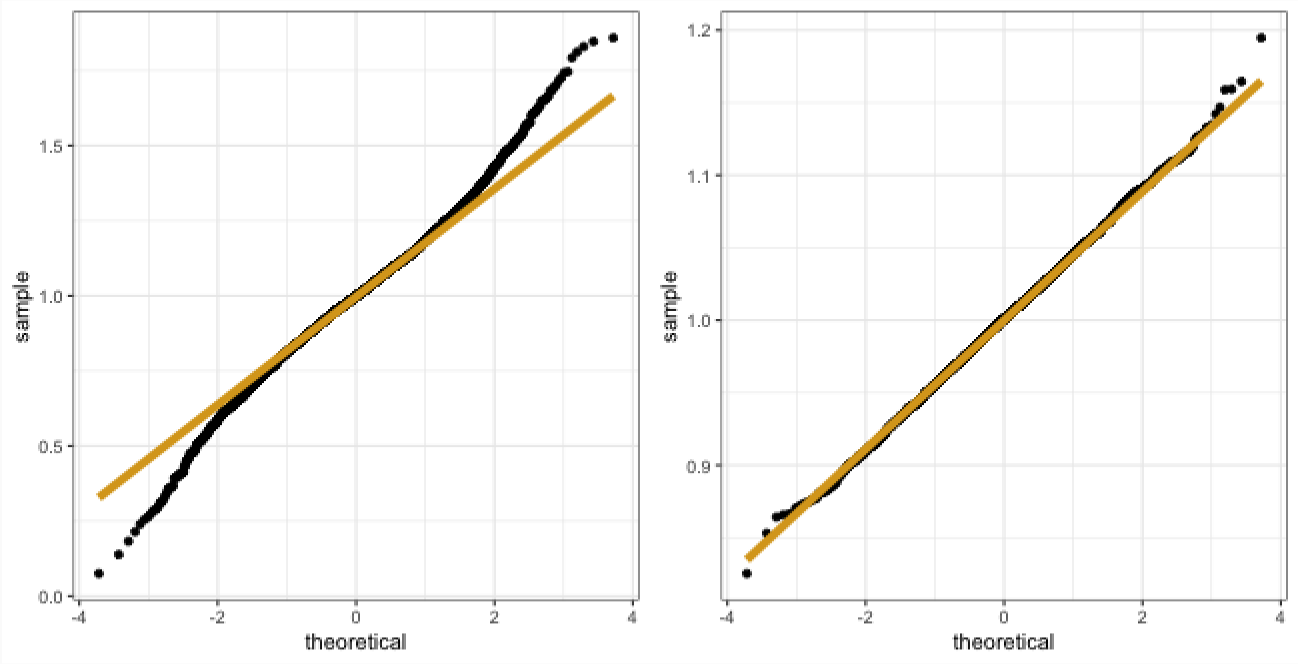
Is the variance still $\sigma^2(X'X)^{-1}$?

Is the distribution of $\hat{\beta}$ still normal?

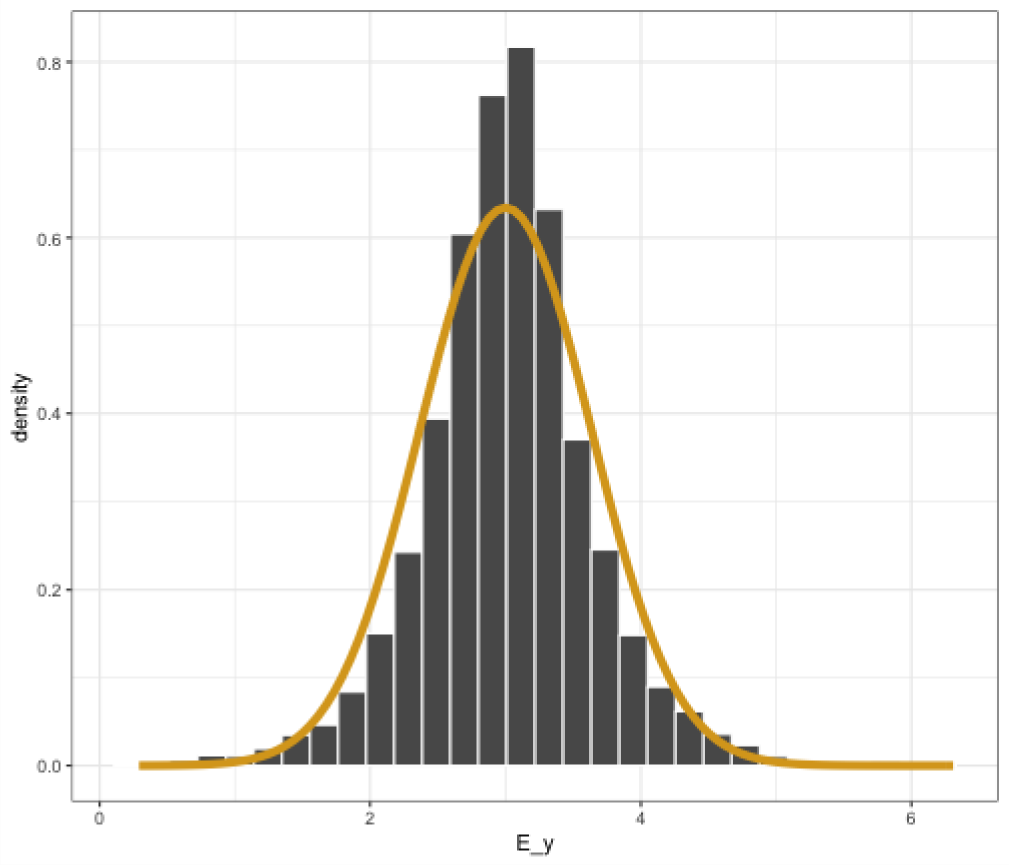
$\hat{\beta}, n = 4$



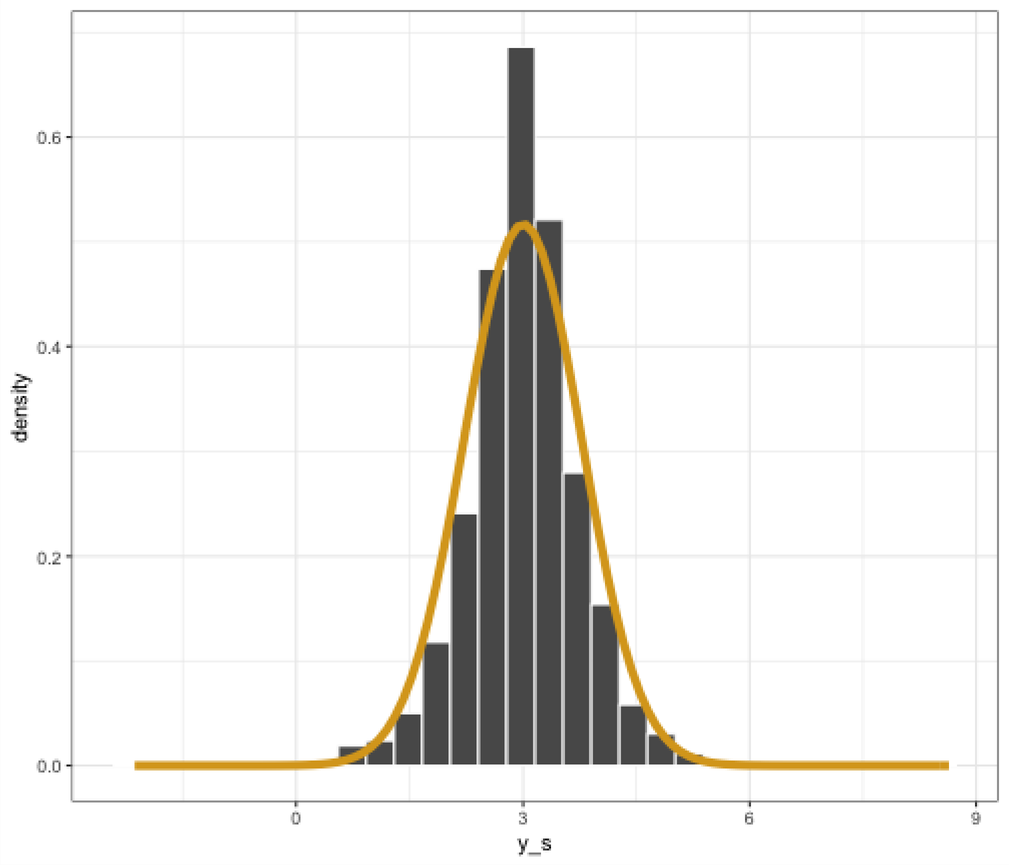
$\hat{\beta}, n = 4$, cont.



$\hat{E}(Y|X = x_s), n = 4$



$Y|X = x_s, n = 4$



Distribution of the X

$$X \sim \quad ?$$

Introduce non-zero covariance into the joint distribution of the X (`rvmnorm()` is helpful here).

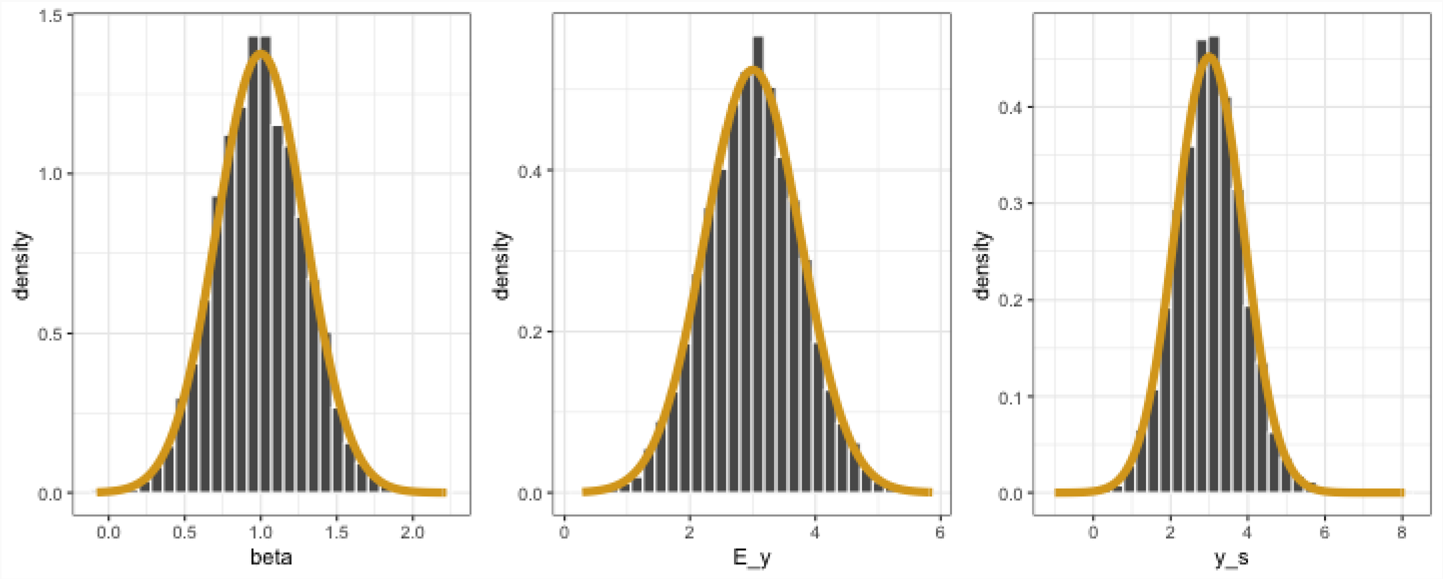
$$X \sim N\left(\mu, \sum\right)$$

$$\mu = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \sum = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

```r
Sigma <- matrix(c(1, 0.5, 0.5, 1),
                byrow = TRUE,
                ncol = 2)
X <- cbind(1, rmvnorm(n, mu, Sigma))
```

Will this mess up the variances? The distributions?

$\hat{\beta}, \hat{E}(Y|X = x_s), Y|X = x_s$



Covariance of the ϵ

$$\epsilon \sim N(0, \sigma^2 I_{n\times n})$$

Introduce non-zero covariance into the joint distribution of the ϵ.

$$\epsilon \sim N\left(\mu, \sum_{n\times n}\right)$$

$$\mu = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \sum = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

```r
sigma_sq <- .2
cov <- .19
Sigma <- matrix(rep(cov, n^2), ncol = n)
diag(Sigma) <- sigma_sq
rmvnorm(1, mean = rep(0, n), sigma = Sigma)
```

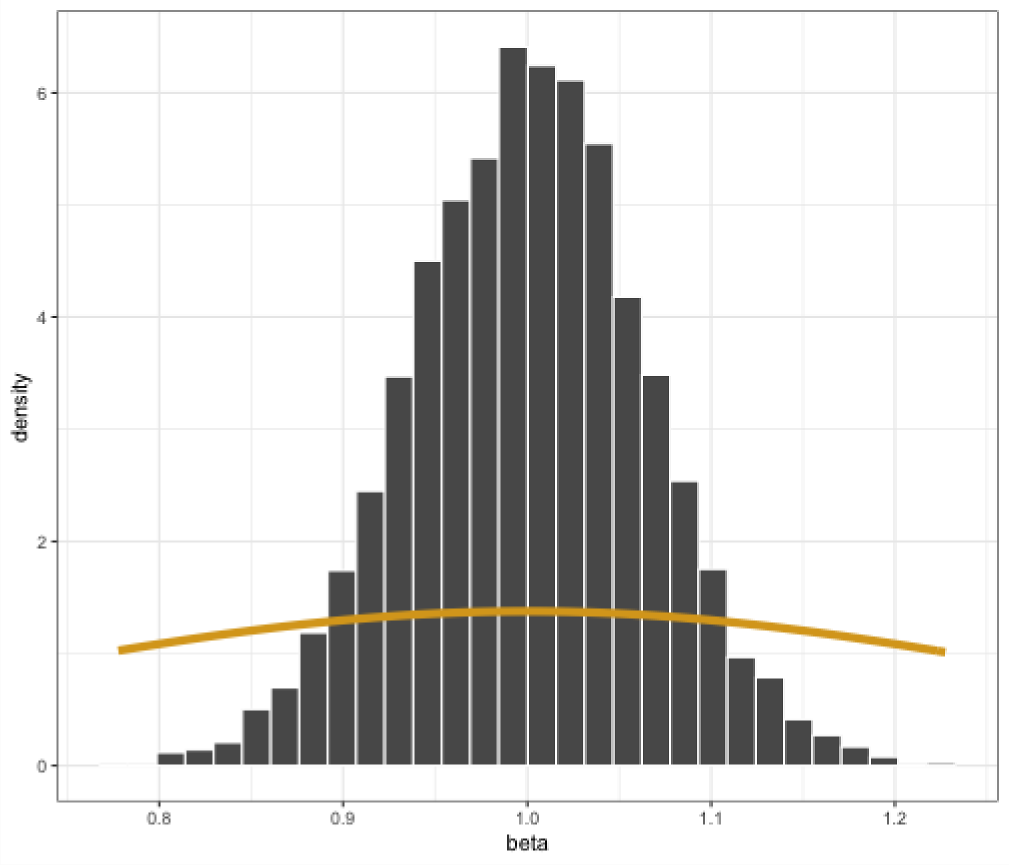
```
##              [,1]      [,2]      [,3]      [,4]
## [1,] 0.05937835 0.1885743 0.1589048 0.1862184
```

Covariance of the ϵ

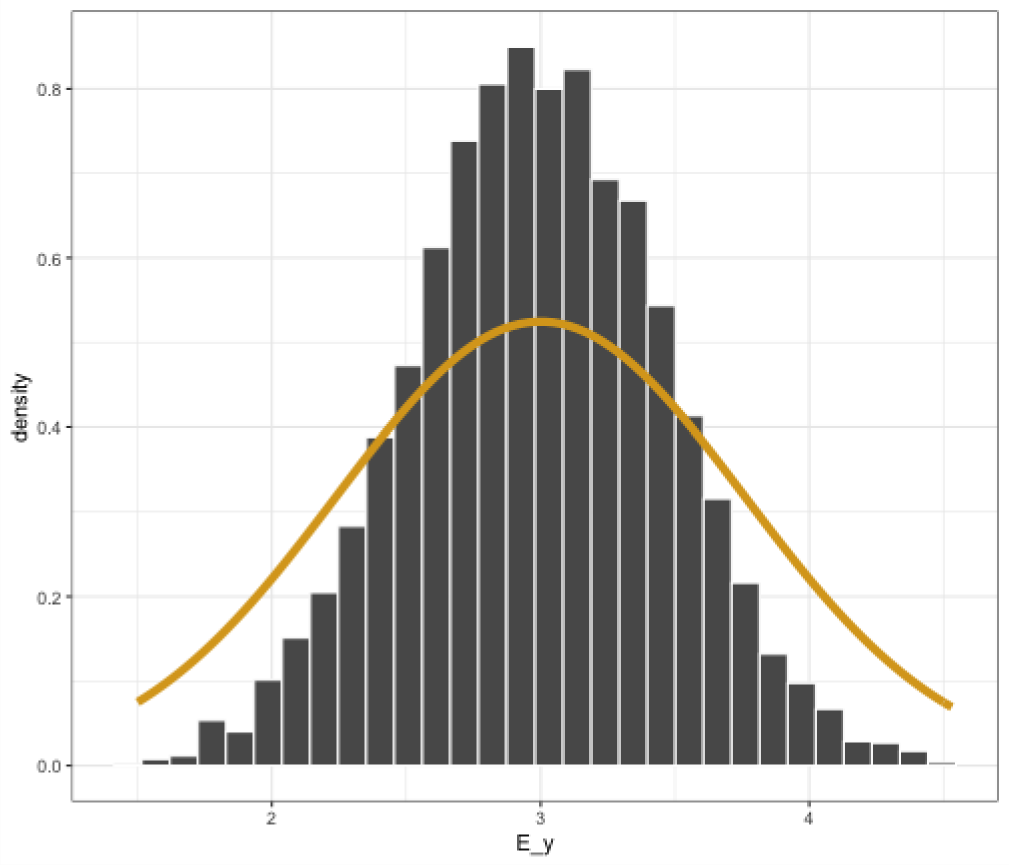
1. What distributions do you expect for the various statistics?

1. Do you expect the variances to be accurate? Underestimate? Overestimate?

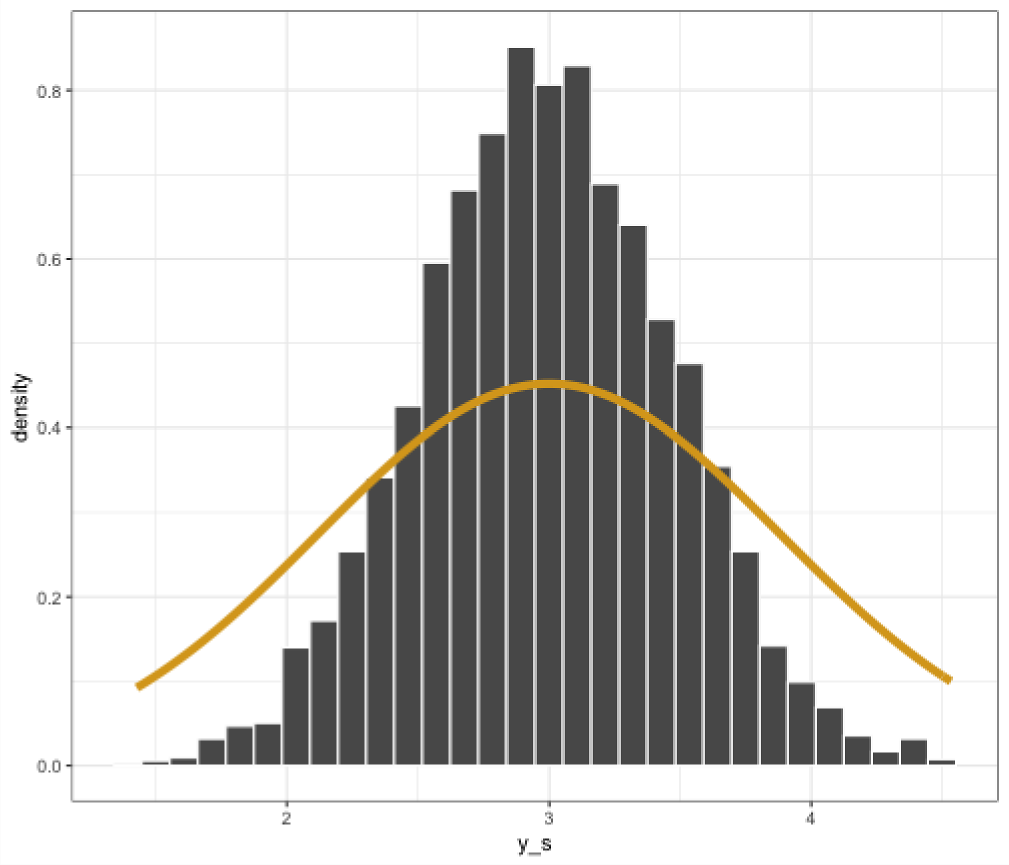
$\hat{\beta}$



$\hat{E}(Y|X = x_s)$

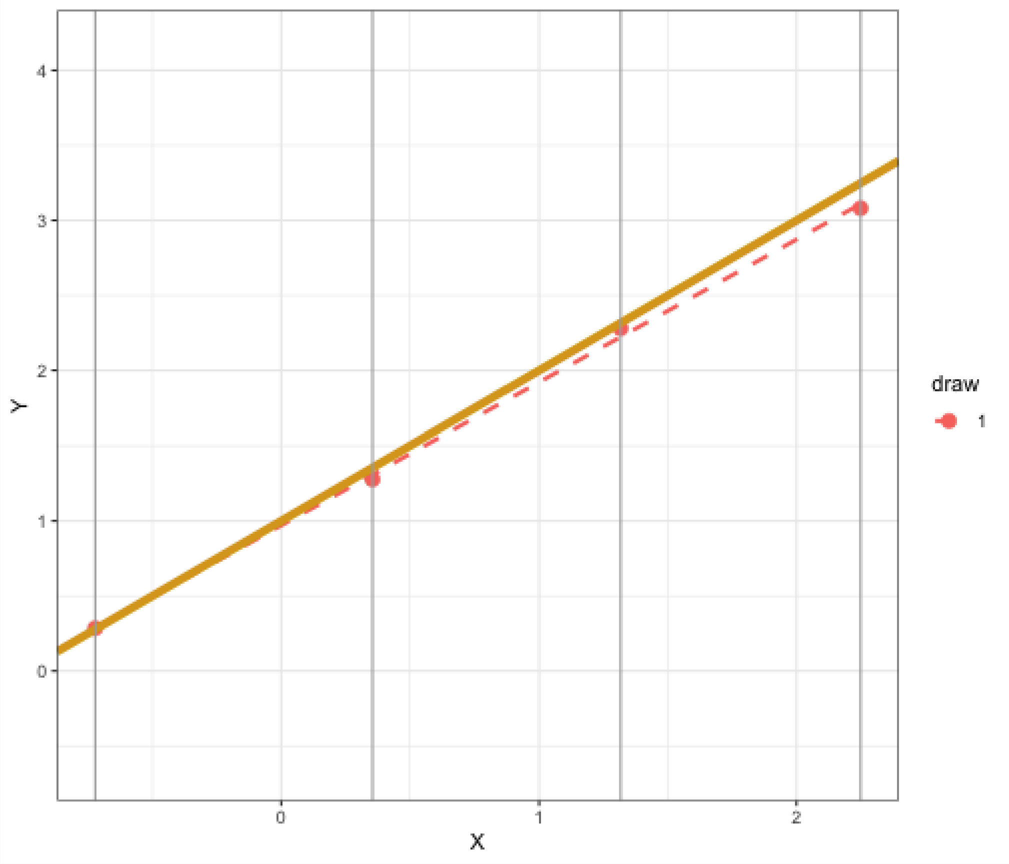


$Y|X = x_s$



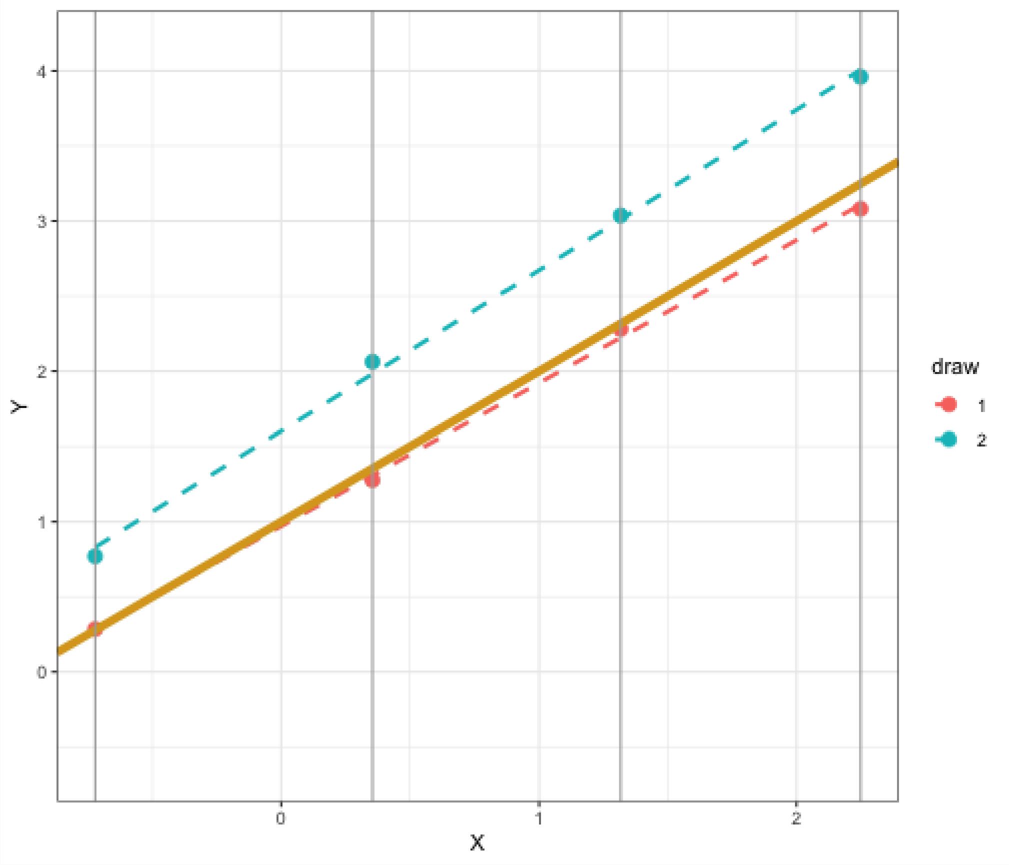
Visualizing correlated errors

One draw.



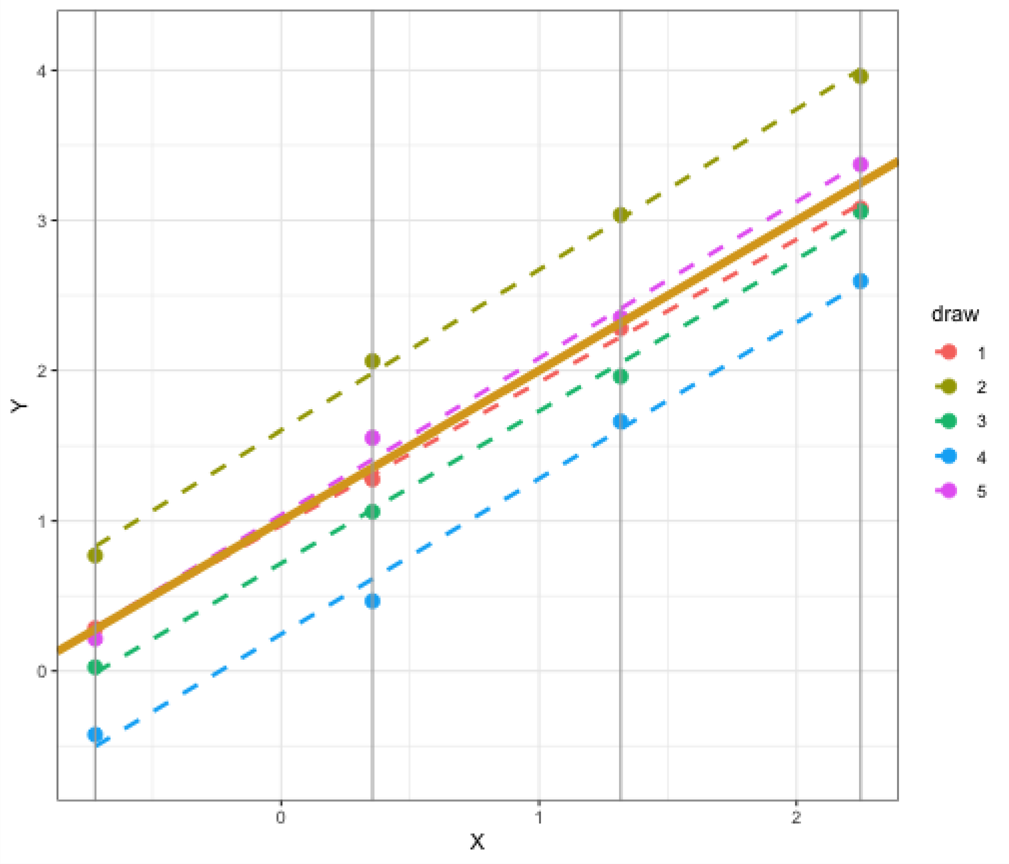
Visualizing correlated errors, cont.

Two draws.

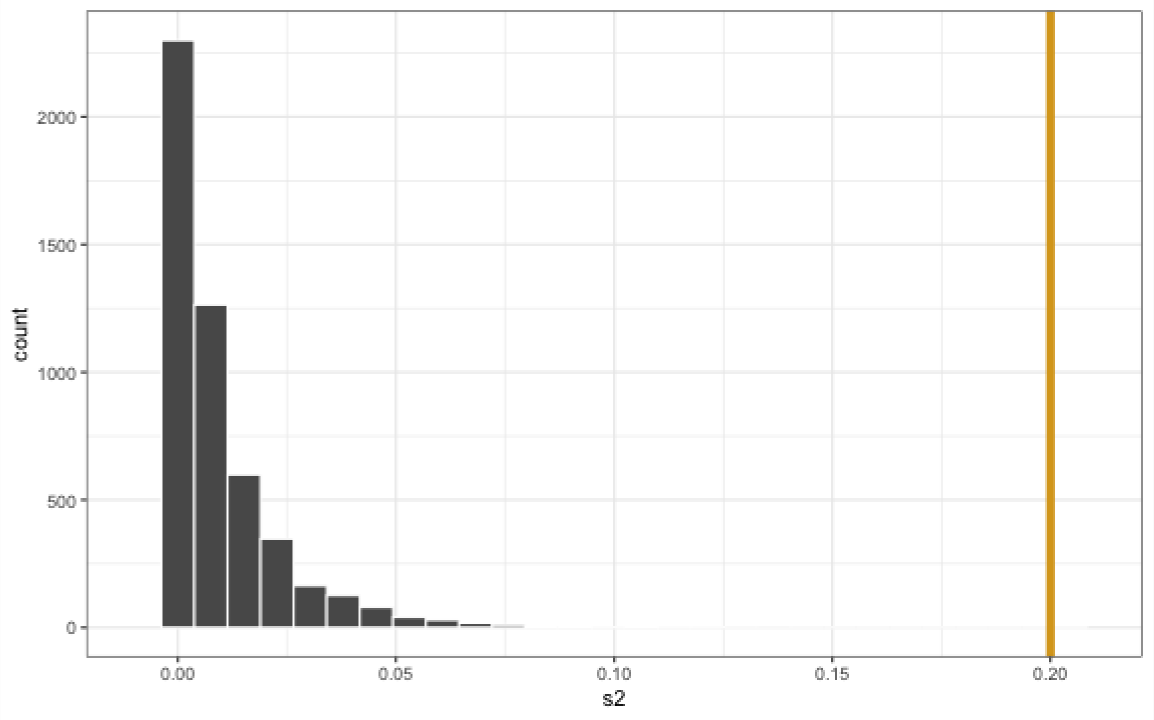


Visualizing correlated errors, cont.

Five draws.



Estimating $\sigma^2



We will dramatically underestimate σ^2, which goes into the SE calculations of all of our statistics.

In Review: the Asymptotical Normality of the MLE

Any MLE, $\hat{\theta}^{MLE}$ *will be normally distributed as* $n \to \infty$ *with expected value* θ *and standard deviation* $\frac{1}{\sqrt(nI(\theta))}$.

Example: $\hat{\beta}^{OLS}$

```r
m1 <- lm(mpg ~ disp + hp + wt, data = mtcars)
summary(m1)
```

```
##
## Call:
## lm(formula = mpg ~ disp + hp + wt, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.891 -1.640 -0.172  1.061  5.861
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.105505   2.110815  17.579  < 2e-16 ***
## disp        -0.000937   0.010350  -0.091  0.92851
## hp          -0.031157   0.011436  -2.724  0.01097 *
## wt          -3.800891   1.066191  -3.565  0.00133 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.639 on 28 degrees of freedom
```

Example: $\hat{\beta}^{OLS}$, cont.

Example: $\hat{\beta}^{Log}$

```
m2 <- glm(factor(am) ~ disp + hp + wt, data = mtcars, family
summary(m2)
```

```
##
## Call:
## glm(formula = factor(am) ~ disp + hp + wt, family = "binomial",
##     data = mtcars)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2074  -0.1285  -0.0092   0.1346   1.3480
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 14.37948    7.65348   1.879   0.0603 .
## disp        -0.02731    0.03922  -0.696   0.4863
## hp           0.06105    0.05219   1.170   0.2421
## wt          -5.95398    3.23118  -1.843   0.0654 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Example: $\hat{\beta}^{Log}$, cont.