3. (a) Let us assume $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. It is shonw that the maximum likelihood estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Consider then the estimator $\mathbf{X}\hat{\boldsymbol{\beta}}$. Calculate the expected value $\mathrm{E}(\mathbf{X}\hat{\boldsymbol{\beta}})$ and the covariance matrix $\mathrm{Cov}(\mathbf{X}\hat{\boldsymbol{\beta}})$. What distribution the fitted values $\mathbf{X}\hat{\boldsymbol{\beta}}$ are following?

(2 points)

Solution:

Since $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ the expected value $\mathrm{E}(\mathbf{X}\hat{\boldsymbol{\beta}})$ is

$$\begin{aligned}\mathrm{E}(\mathbf{X}\hat{\boldsymbol{\beta}}) &= \mathrm{E}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{E}(\mathbf{y}) \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\cdot\mathbf{I}\cdot\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}.\end{aligned}$$

The covariance matrix $\mathrm{Cov}(\mathbf{X}\hat{\boldsymbol{\beta}})$ is

$$\begin{aligned}\mathrm{Cov}(\mathbf{X}\hat{\boldsymbol{\beta}}) &= \mathrm{Cov}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\mathrm{Cov}(\mathbf{y})(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')' \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\cdot(\sigma^2\mathbf{I})\cdot\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ &= \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\cdot\mathbf{I}\cdot\mathbf{X}' = \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.\end{aligned}$$

Furthermore, since $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{H}\mathbf{y}$, where $\mathbf{G} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, is linear transformation of the normally distributed random vector $\mathbf{y}$, it also follows normal distribution $\mathbf{X}\hat{\boldsymbol{\beta}} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$.

(b) Consider the linear model

$$\begin{aligned}&\mathbf{y} \sim N(\boldsymbol{\mu}, \sigma^2\mathbf{I}),\\ &\boldsymbol{\mu} = \mathbf{1}\beta_0,\end{aligned}$$

where $\mathbf{1}$ is a vector of ones $\mathbf{1} = (1, 1, \ldots, 1)'$. Use the fundamental equation of the BLUE to show that the sample mean

$$\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{1}{n}\mathbf{1}'\mathbf{y}$$

is the best linear unbiased estimator for the parameter β_0, i.e., $\hat{\beta}_0 = \bar{y}$.

(2 points)

Solution:

In the model $\mathbf{y} = \mathbf{1}\beta_0 + \boldsymbol{\varepsilon}$, the model matrix is $\mathbf{X} = \mathbf{1}$ and the matrix M is

$$\mathbf{M} = \mathbf{I} - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}' = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'.$$

With these $\mathbf{X}$ and $\mathbf{M}$ matrices, the sample mean $\frac{1}{n}\mathbf{1}'\mathbf{y}$ is the BLUE of β_0 if and only if the following BLUE equations are holding:

$$\frac{1}{n}\mathbf{1}'(\mathbf{1} : \mathbf{M}) = (1 : \mathbf{0}).$$

Now $\frac{1}{n}\mathbf{1}'\mathbf{1} = \frac{1}{n} \cdot n = 1$ and

$$\begin{aligned}
\frac{1}{n}\mathbf{1}'\mathbf{M} &= \frac{1}{n}\mathbf{1}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right) = \frac{1}{n}\mathbf{1}' - \frac{1}{n}\mathbf{1}' \cdot \frac{1}{n}\mathbf{1}\mathbf{1}' \\
&= \frac{1}{n}\mathbf{1}' - \frac{1}{n^2}\mathbf{1}'\mathbf{1}\mathbf{1}' = \frac{1}{n}\mathbf{1}' - \frac{n}{n^2}\mathbf{1}' \\
&= \frac{1}{n}\mathbf{1}' - \frac{1}{n}\mathbf{1}' = \mathbf{0}.
\end{aligned}$$

Thus the fundamental BLUE equations $\frac{1}{n}\mathbf{1}'(\mathbf{1} : \mathbf{M}) = (1 : \mathbf{0})$ hold and hence the sample mean $\bar{y} = \frac{1}{n}\mathbf{1}'\mathbf{y}$ is the BLUE of β_0.

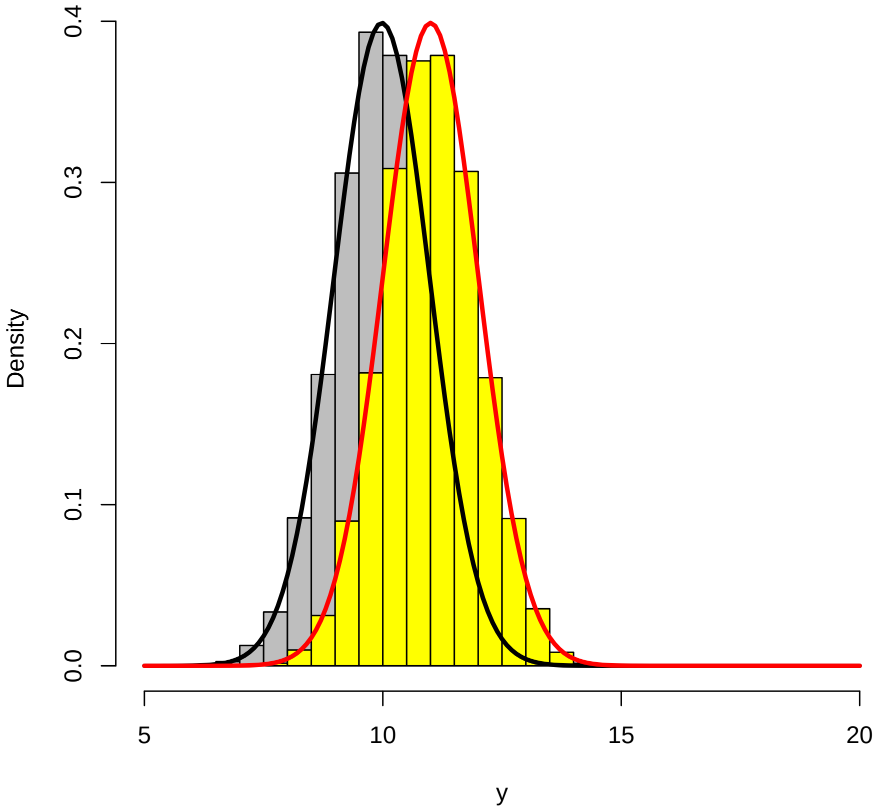
(c) Let us consider the large sample situation where we have simulated $n_j = 10000$ observations from the normal distributions $Y_{iA} \sim N(\mu, \sigma^2)$ and $Y_{iB} \sim N(\mu + \delta, \sigma^2)$. The aim is investigate in which values $\delta > 0$ the predictive effect size between Y_{iA} and Y_{iB} is large enough that you would be ready to declare that there is a real effect size between the random variables Y_{iA} and Y_{iB}. Copy the R-code given below. Start changing the value of δ in code and based on the histograms and estimated density curves, decide yourself which values of δ are such that the values of the random variables Y_{iA} and Y_{iB} are "mostly" different. What are your corresponding p-value and d-value when you are feeling that there is a effect size difference?

```
muA<-10
delta<-1         # You should change this value
muB<-muA+delta
x<-rep(c("A","B"),each=10000)
yA<-rnorm(10000, mean=muA, sd=1)
yB<-rnorm(10000, mean=muB, sd=1)
y<-c(yA,yB)
model<-lm(y~factor(x))
betahat<-coef(model)

k1<-c(0,1)
K<-cbind(k1)
q<-1
Wald<-(t(t(K)%*%betahat)%*%solve(t(K)%*%vcov(model)%*%K)%*%t(K)%*%betahat)/q
Wald
p.value<-pf(Wald, 1, 19998, lower.tail = FALSE)
p.value

x1<-cbind(c(1,0))
x2<-cbind(c(1,1))
pred<-(t(x2)-t(x1))%*%betahat
sigma2<-sigma(model)^2
X<-model.matrix(model)
T<-pred/sqrt(sigma2*(2+(t(x2)-t(x1))%*%solve(t(X)%*%X)%*%(x2-x1)))
d.value<-2*pt(abs(T),df=19998, lower.tail = FALSE)
d.value

hist(yA, xlim=c(5,20), col="grey", main="", freq=FALSE, xlab="y")
hist(yB, xlim=c(5,20),add=TRUE, col="yellow", freq=FALSE)
lines(seq(5,20,0.1),dnorm(seq(5,20,0.1), mean=betahat[1],sd=sigma(model)),
col="black",lwd=3)
lines(seq(5,20,0.1),dnorm(seq(5,20,0.1), mean=betahat[1]+betahat[2],sd=sigma(model)),
col="red",lwd=3)
```



(2 points)