1) Let $S = (R_t - \hat{R}_t)^2$

we want to minimize this i.e $\frac{\partial S}{\partial \beta_i} = 0$

$S = (R_t - (\beta_0 + \beta_1 F_{1,t} + \dots + \beta_p F_{p,t}))^2$

$\beta_0$ are the constant terms

OLS estimators satisfy $\frac{\partial S}{\partial \beta_i}(\hat{\beta_i}) = 0$

Substitute in to $\beta_0$ $\quad i = 0, \dots, p$

$$-2(R_t - \hat{R}_t) = 0$$

so, $-2 \sum_{t=1}^{n} (R_t - \hat{R}_t) = 0$

$$\sum_{t=1}^{n} R_t = \sum_{t=1}^{n} \hat{R}_t$$

$$\frac{\sum_{t=1}^{n} R_t}{n} = \frac{\sum_{t=1}^{n} \hat{R}_t}{n}$$

so $\bar{R}_t = \bar{R} = \hat{\mu}_R$

2) $\hat{ES}^{norm}(\alpha) = \frac{1}{\alpha}\int_0^\alpha VaR(y)\,dy$

$VaR^{norm}(y) = -S_t\{\hat{\mu} + \hat{\sigma}\,\Phi^{-1}(y)\}$

$\hookrightarrow$ cdf of $N(0,1)$

$\frac{1}{\alpha}\int_0^\alpha \left[-S_t(\hat{\mu} + \hat{\sigma}\,\Phi^{-1}(y))\right]dy$

$= \frac{1}{\alpha}\int_0^\alpha -S_t\hat{\mu}\,dy + \frac{1}{\alpha}\int_0^\alpha -S_t\hat{\sigma}\,\Phi^{-1}(y)\,dy$

$= -S_t\hat{\mu} + \frac{-S_t\hat{\sigma}}{\alpha}\int_0^\alpha \Phi^{-1}(y)\,dy = Ans$

Now we do change of variable $x = \Phi^{-1}(y)$ $\Phi(x) = y$

$\frac{d\Phi(x)}{dx} = \frac{dy}{dx} \rightarrow dy = \Phi'(x)dx$ $y \in (0,\alpha)$, pdf $\phi$

$\Phi^{-1}(y) \in (-\infty, \Phi^{-1}(\alpha))$

Please note: $x\Phi'(x) : x\phi(x) = -\phi'(x)$

$Ans = -S_t\hat{\mu} + \frac{-S_t\hat{\sigma}}{\alpha}\int_{-\infty}^{\Phi^{-1}(\alpha)} x\,\Phi'(x)\,dx$

$= -S_t\hat{\mu} + \frac{-S_t\hat{\sigma}}{\alpha}\int_{-\infty}^{\Phi^{-1}(\alpha)} -\phi'(x)\,dx$

$= -S_t\hat{\mu} + \frac{S_t\hat{\sigma}}{\alpha}\left[\phi(x)\right]_{-\infty}^{\Phi^{-1}(\alpha)} = -S_t\hat{\mu} + \frac{S_t\hat{\sigma}}{\alpha}\phi(\Phi^{-1}(\alpha))$

$= S_t\left\{-\hat{\mu} + \hat{\sigma}\left(\frac{\phi(\Phi^{-1}(\alpha))}{\alpha}\right)\right\}$

```{r}
library(fGarch)
rm(list=ls())


var_diff = function(a) {
  (-qnorm(a,mean=0.036,sd=1.52)) - (-qstd(a,mean=0.036,sd=1.52,nu=2.8))
}
uniroot(var_diff,c(0.001,0.05))

```

```
$root
[1] 0.01560864

$f.root
[1] -0.0002073335

$iter
[1] 7

$init.it
[1] NA

$estim.prec
[1] 6.103516e-05
```

````
```{r}
normal_var = function(a) {-qnorm(a,mean=0.036,sd=1.52)}
std_var = function(a) {-qstd(a,mean=0.036,sd=1.52,nu=2.8)}
plot(normal_var, from=0, to=0.05, xlab="alpha",ylab="VaR",col="blue")
plot(std_var, from=0, to=0.05,xlab="alpha",ylab="VaR",add=TRUE)
legend("topright",inset=0.08, legend=c("normal","std"),col=c("blue","black"),lty=1:1,lwd=3)

#When alpha is 0.01560864 the difference is 0 for var normal vs var std because it is the root
#Alpha needs to be bigger than the root i.e 0.01560864 for VaR assuming normally-distributed returns to be larger
#than VaR assuming std(v=2.8) distributed returns.
```
````