Task 4 - Permutation

Permutation

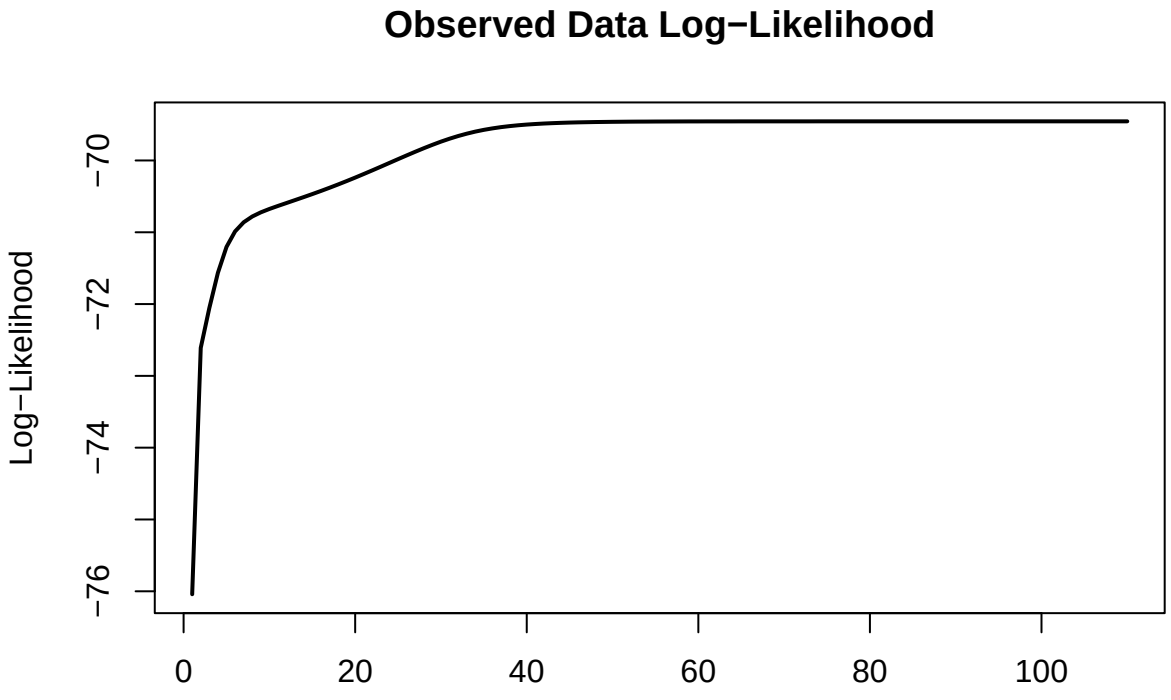
Histogram of Years of Experience

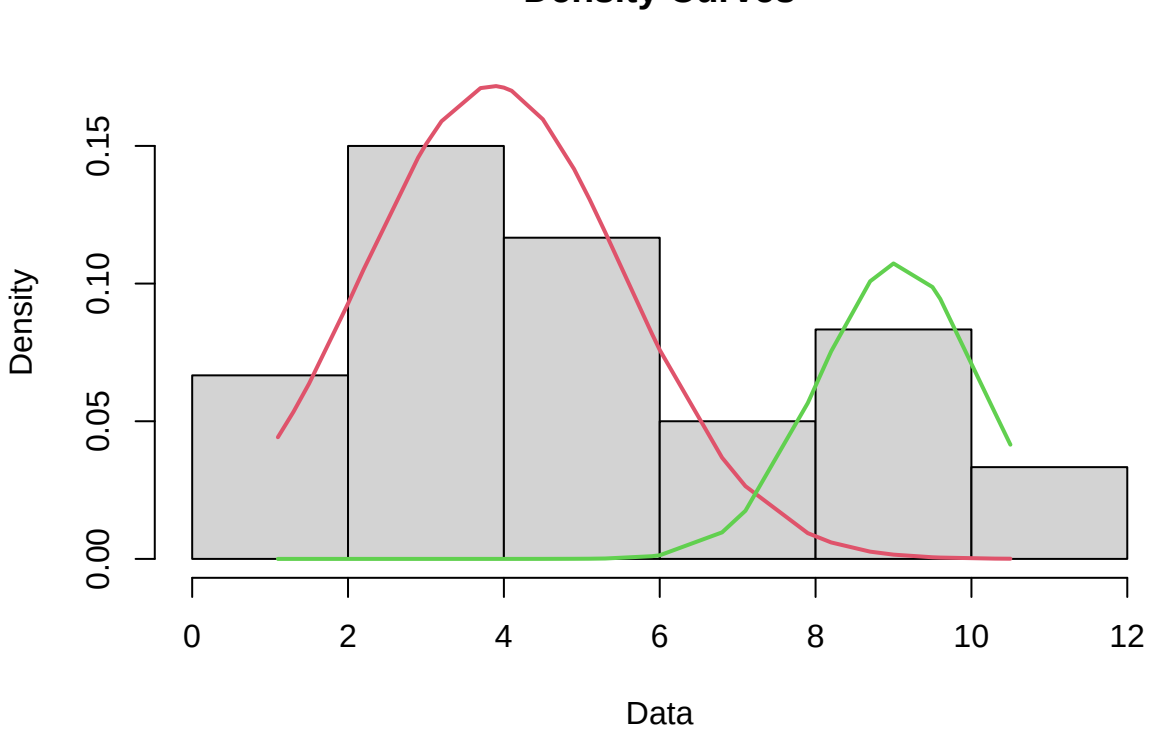
Fitting mixture distribution

Seeing as there are two peaks, the distribution might be a mixture. Let's fit two normal distributions.

```
## number of iterations= 109
```

```
plot(exp, density = TRUE)
```






```
cat("Mixture Log Likelihood:", exp$loglik,"\n")
```

```
## Mixture Log Likelihood: -69.4553
```

```
cat("First Normal Distribution proportion:", exp$lambda[1], ", Mean:",exp$mu[1], ", and Standard Deviati
```

```
## First Normal Distribution proportion: 0.7210976 , Mean: 3.859548 , and Standard Deviation: 1.674672
```

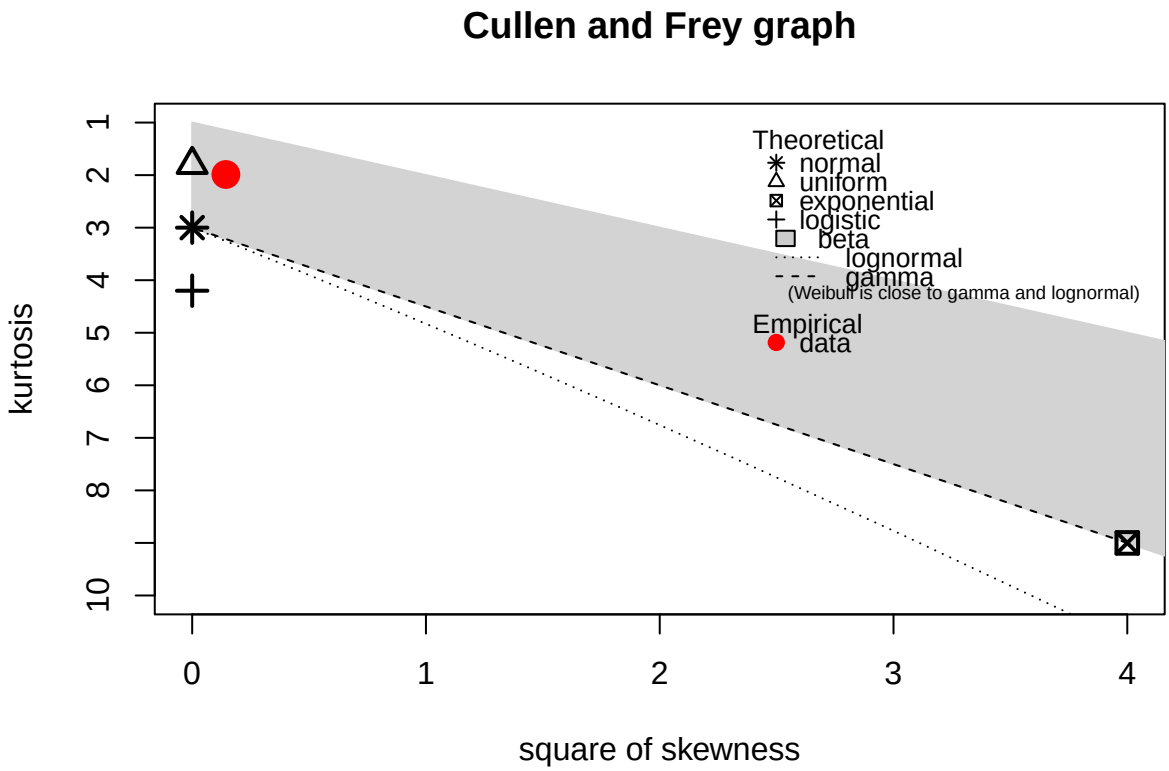
```
cat("\nSecond Normal Distribution proportion:", exp$lambda[2], ", Mean:",exp$mu[2], ", and Standard Devi
```

```
##
## Second Normal Distribution proportion: 0.2789024 , Mean: 9.07207 , and Standard Deviation: 1.034407
```

Fitting one distribution

Now, let's fit try fitting just one theoretical distribution

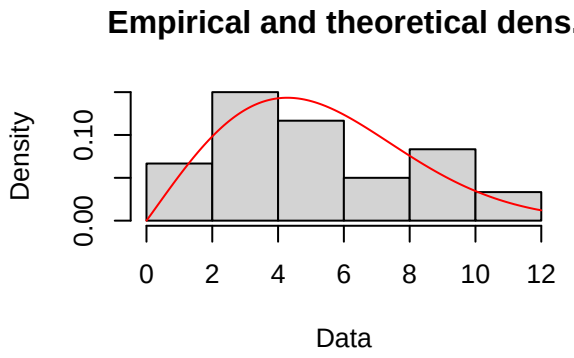
```
descdist(salary_df$YearsExperience, discrete = FALSE)
```

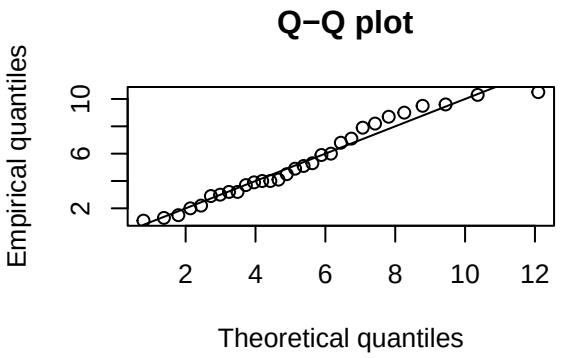


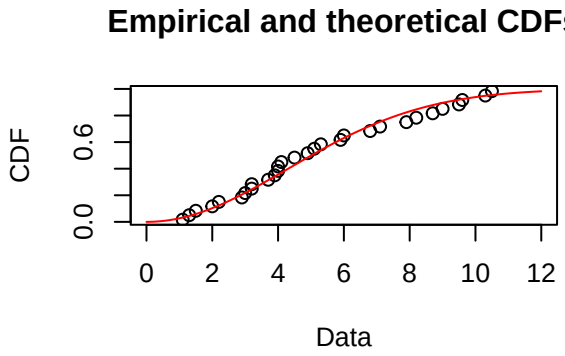

```
## summary statistics
## ------
## min:  1.1   max:  10.5
## median:  4.7
## mean:  5.313333
## estimated sd:  2.837888
## estimated skewness:  0.3795602
## estimated kurtosis:  1.987788
```

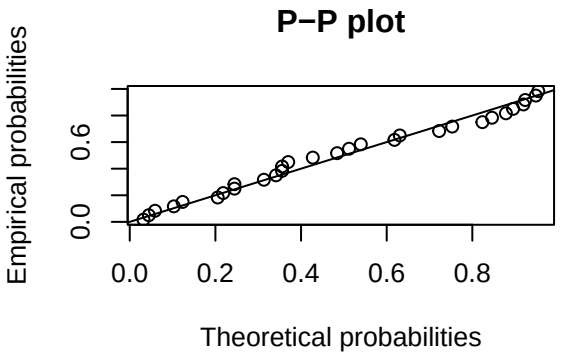
```
fit.weibull <- fitdist(salary_df$YearsExperience, "weibull")
fit.norm <- fitdist(salary_df$YearsExperience, "norm")
```

```
plot(fit.weibull)
```

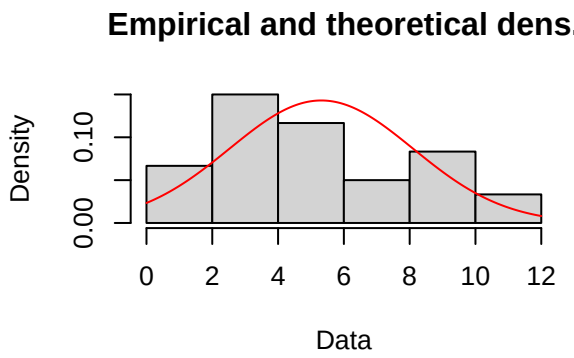


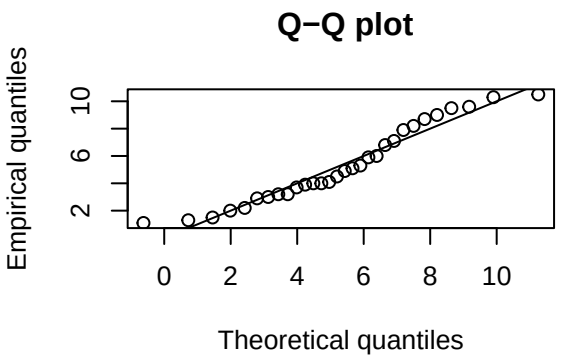


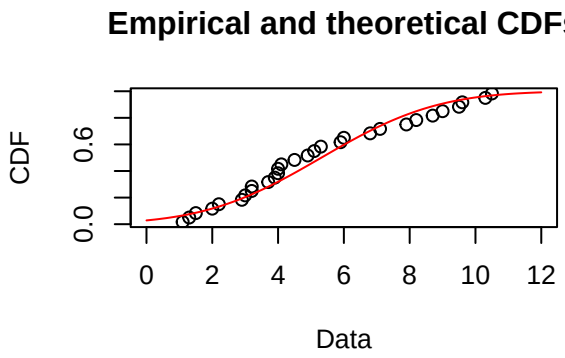


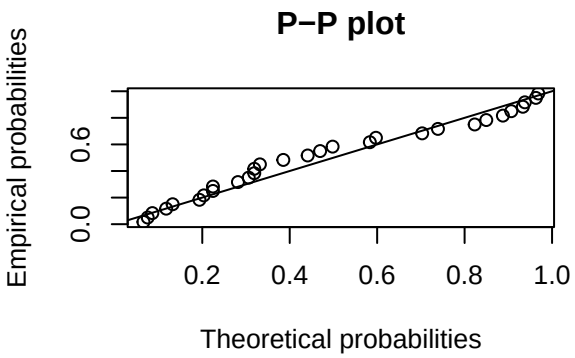



```
plot(fit.norm)
```

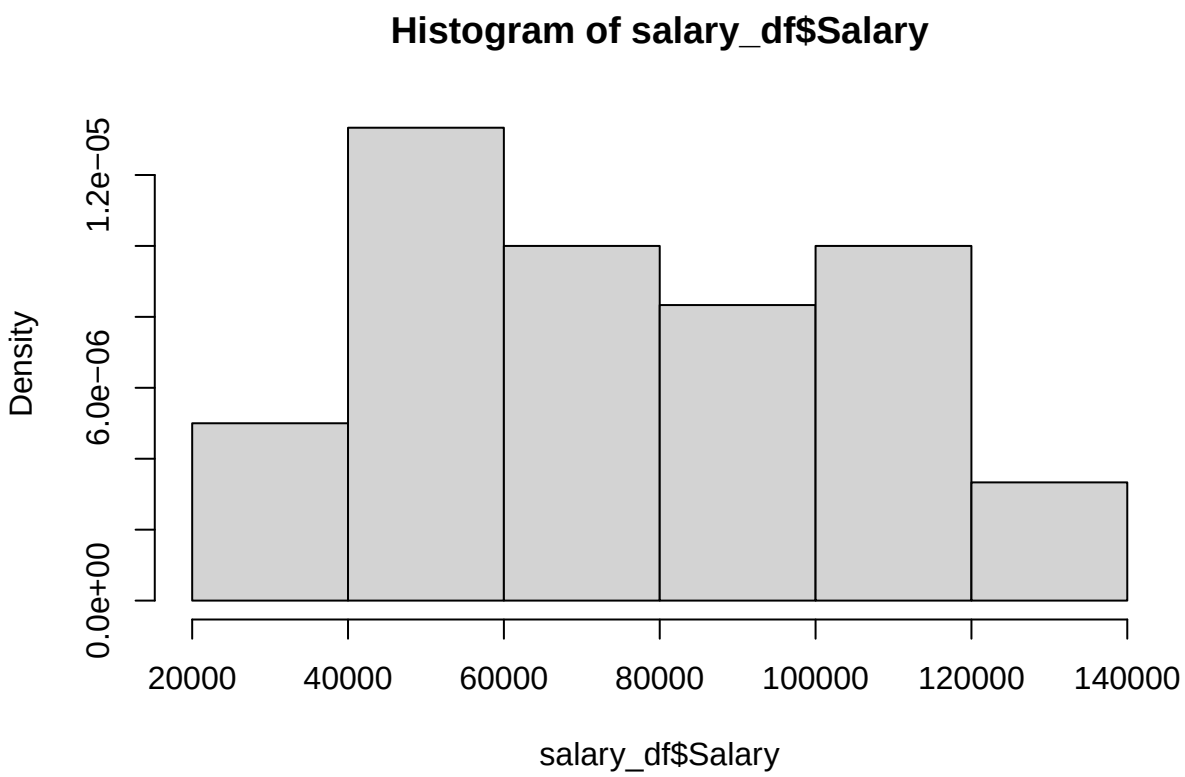








```r
hist(salary_df$Salary, breaks = "Sturges", freq = FALSE)
```




```r
cat("Fitted Weibull AIC:",fit.weibull$aic )
```

```
## Fitted Weibull AIC: 147.1482
```

```r
cat("Fitted Normal AIC:", fit.norm$aic )
```

```
## Fitted Normal AIC: 150.7029
```

The Weibull distribution seems to explain our sample better, in the case of fitting one distribution.

The estimated parameters are:

```r
fit.weibull$estimate
```

```
##    shape    scale
## 2.016509 6.012105
```

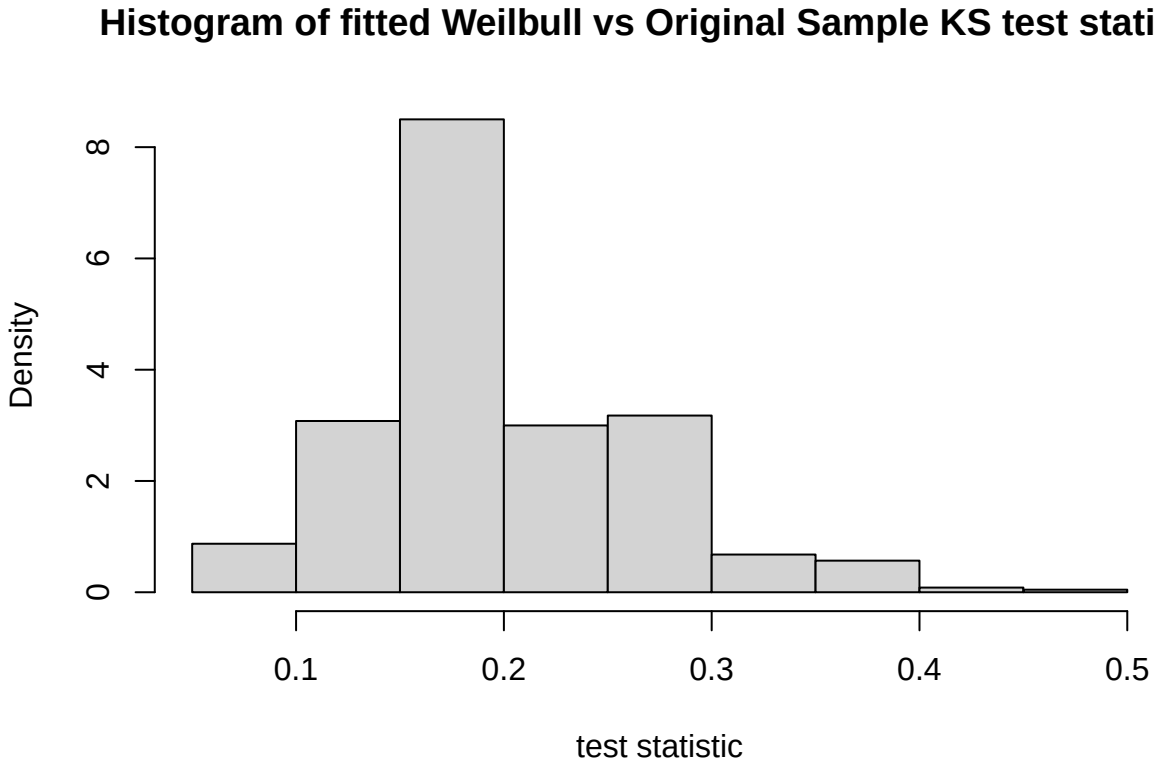
Weibull Permutation Testing

```r
weibull.samp <- rweibull(30, shape = fit.weibull$estimate["shape"], scale = fit.weibull$estimate["scale
```

```
## The p-value of the initial Kolmogorov-Smirnov test is: 0.8039413
```

```
##
## The Test Statistic is: 0.1666667
```

```
## The obtained ASL is: 0.8025197
```



Mixed Normal Distribution Permutation Testing

```
N <- 30
sample_idx <- sample(1:2, prob = exp$lambda, size = N, replace = TRUE)
mixnorm.samp <- rnorm(N, mean = exp$mu[sample_idx], sd = exp$sigma[sample_idx])
```

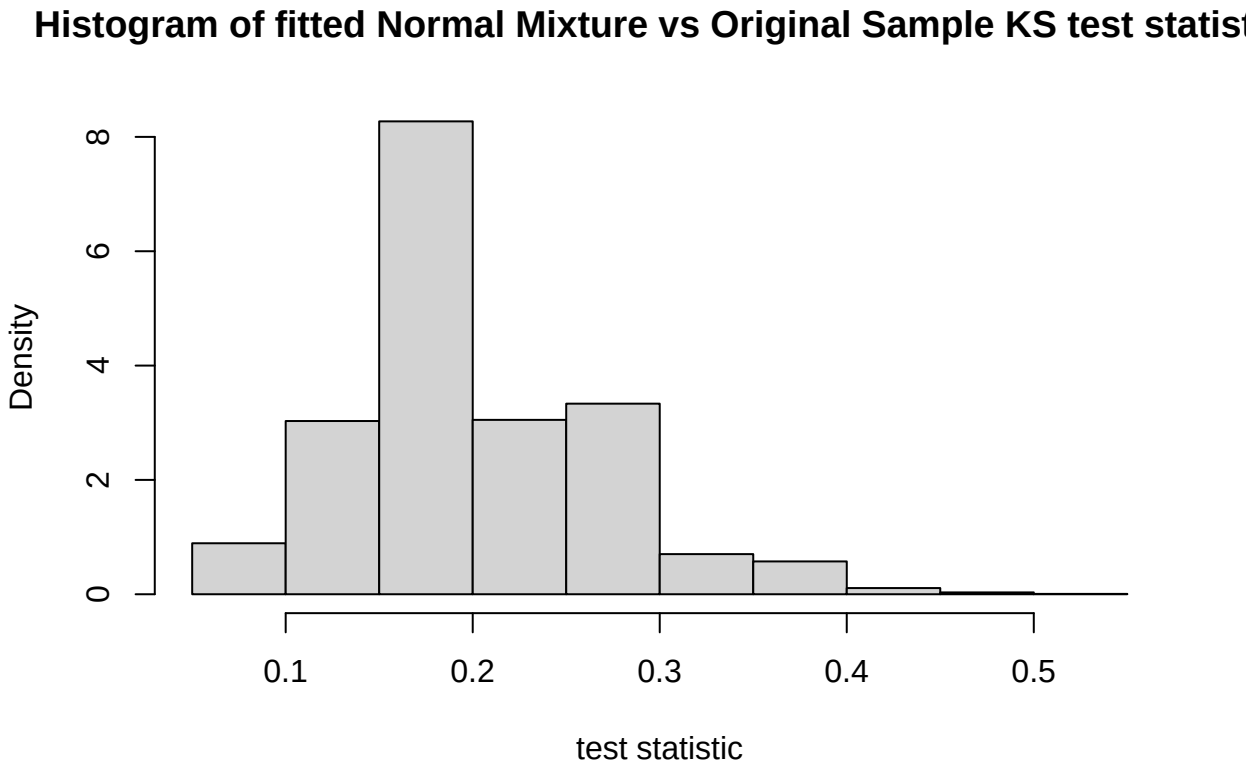
```
mixnorm.ks.test.initial <- ks.test(salary_df$YearsExperience, mixnorm.samp)

cat("The p-value of the initial Kolmogorov-Smirnov test is:", mixnorm.ks.test.initial$p.value)
```

```
## The p-value of the initial Kolmogorov-Smirnov test is: 0.9560382
```

```
mixnorm.ks.test.D0 <- mixnorm.ks.test.initial$statistic
cat("\nThe Test Statistic is:", mixnorm.ks.test.D0)
```

```
##
## The Test Statistic is: 0.1333333
```

```
## The obtained ASL is: 0.9554045
```


Histogram of fitted Normal Mixture vs Original Sample KS test statist
Density
0
2
4
6
8
0.1
0.2
0.3
0.4
0.5
test statistic