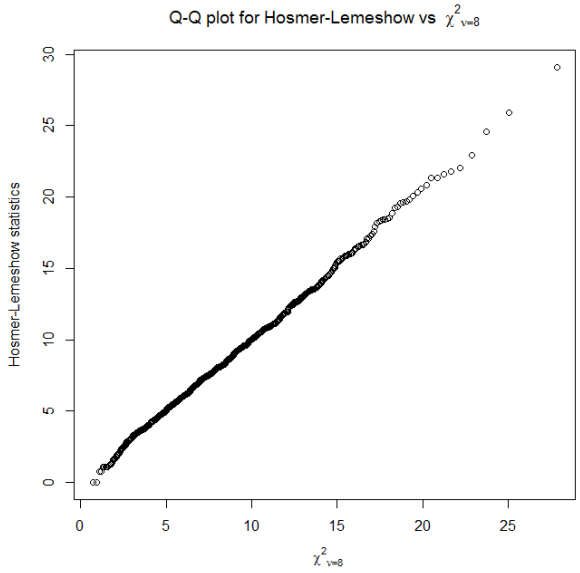
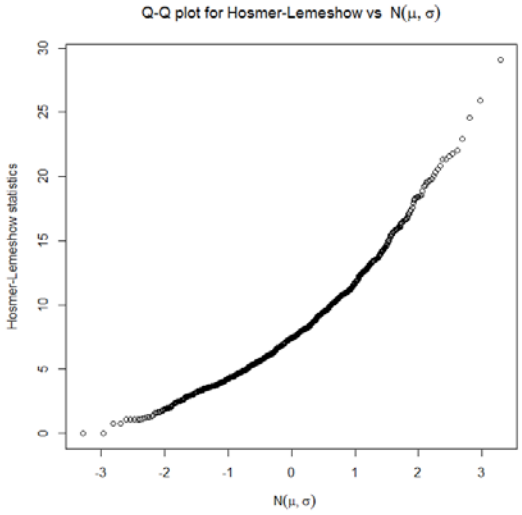
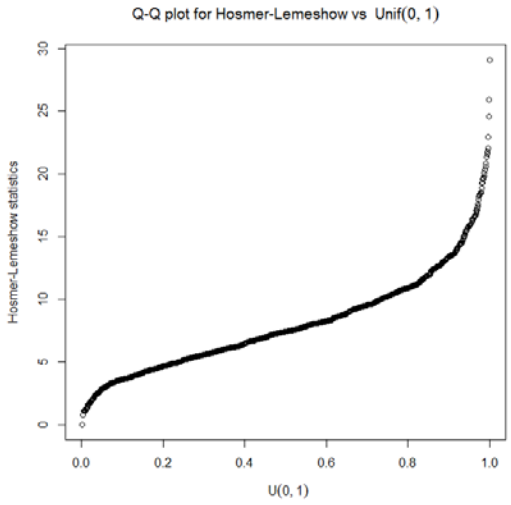
Q4

The HL test statistic is compared to a chi-squared distribution with a certain number of degrees of freedom. The simulation produces 1000 HL statistics for a logit model where the probabilities are divided into 10 groups. The results for Hosmer-Lemeshow vs chi(df=8) are shown in a Q_Q plot .The plot is a straight line. So the distribution for Hosmer-Lemeshow is X^2_{g-2}



We also checked normal and uniform distributions which don't match Hosmer-Lemeshow. (see plots below). To find the degree of freedom for chi-square distribution we can calculate rank of covariance matrix.





Appendix (R code)

```
# install.packages('ResourceSelection')

library("ResourceSelection")


# to store generated HL statistics

hlvals<-c()

# Run the simulation

for (i in 10:1000){

 # Set sample size

 n<-i

 # Making data dependent on each other

 x1<-rnorm(n)

 x2<-0.5*x1+rnorm(n)

 xb<-x1-x2

 # Link function

 pr<- exp(xb)/(1+exp(xb))

 # Response variable

 y <- 1*(runif(n)<pr)

 #' Dataframe for logit model

 dt<- data.frame(y,x1,x2)

 #' The logistic regression model

 model<-glm(y~x1+x2, data=dt, family=binomial(link="logit"))

 # Run the HL test

 hl<-hoslem.test(model$y,fitted(model),g=10)

 # Save the HL stat. Indexing adjusted for
```

```
  hlvals[i-9]<-hl$statistic
}



qqplot(qchisq(ppoints(length(hlvals)), df = 8), hlvals,
       xlab = ~~ {chi^2}[nu == 8], ylab = "Hosmer-Lemeshow statistics",
       main = expression("Q-Q plot for Hosmer-Lemeshow vs" ~~ {chi^2}[nu == 8]))


qqplot(qunif(ppoints(length(hlvals))), hlvals,
       xlab = ~~ {U (0,1)}, ylab = "Hosmer-Lemeshow statistics",
       main = expression("Q-Q plot for Hosmer-Lemeshow vs" ~~ {Unif(0,1)}))


qqplot(qnorm(ppoints(length(hlvals))), hlvals,
       xlab = ~~ {N (mu, sigma)}, ylab = "Hosmer-Lemeshow statistics",
       main = expression("Q-Q plot for Hosmer-Lemeshow vs" ~~ {N (mu, sigma)}))
```