Multiple Comparisons: Homework - 2

Submitted by:

Eyal Bar-Natan - 207630658

Omer Shubi - 312236219

Question 1

Part A

We perform $iterations = 5,000$ simulations, in which each time we randomly pick $m = 5$ values out a Uniform Distribution ($U \sim Uni[0, 1]$). In each simulation, out of the picked values we select the one with the minimum value. The result of these simulations is stored in $vecmin\_m$.

```
iterations <- 5000
```

```
m <- 5
mat_5 <- replicate(iterations, runif(m,0, 1))
vecmin_5 <- apply(mat_5, 2, min)
```

and similarly with $m = 20$:

```
m <- 20
mat_20 <- replicate(iterations, runif(m,0, 1))
vecmin_20 <- apply(mat_20, 2, min)
```

and $m = 100$:

```
m <- 100
mat_100 <- replicate(iterations, runif(m,0, 1))
vecmin_100 <- apply(mat_100, 2, min)
```

1.

The proportion of $U_{(1)} < 0.05$ when $m = 5$:

```
length(vecmin_5[vecmin_5<0.05])/iterations
```

```
## [1] 0.22
```

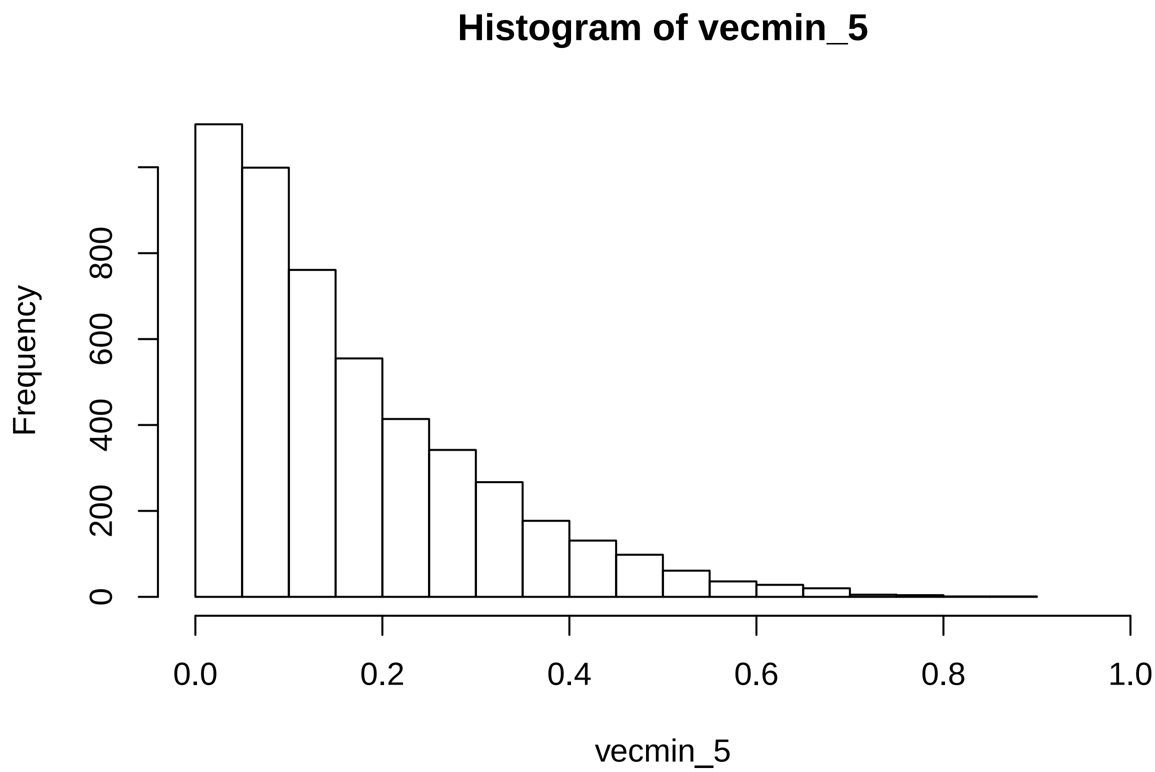
Similarly to section $\aleph$ in H.W. 1 - in which $p_{value} \sim Uni[0, 1]$, and we were requested to retrieve the minimal p_{value}.

This is exactly what we are asked to find in this question as well, by defining $U_i = p_{value\,i} \sim Uni[0, 1]$ and $U_{(1)} = min(U_1, U_2, \ldots, U_m)$ we are dealing with the same thing.

2.

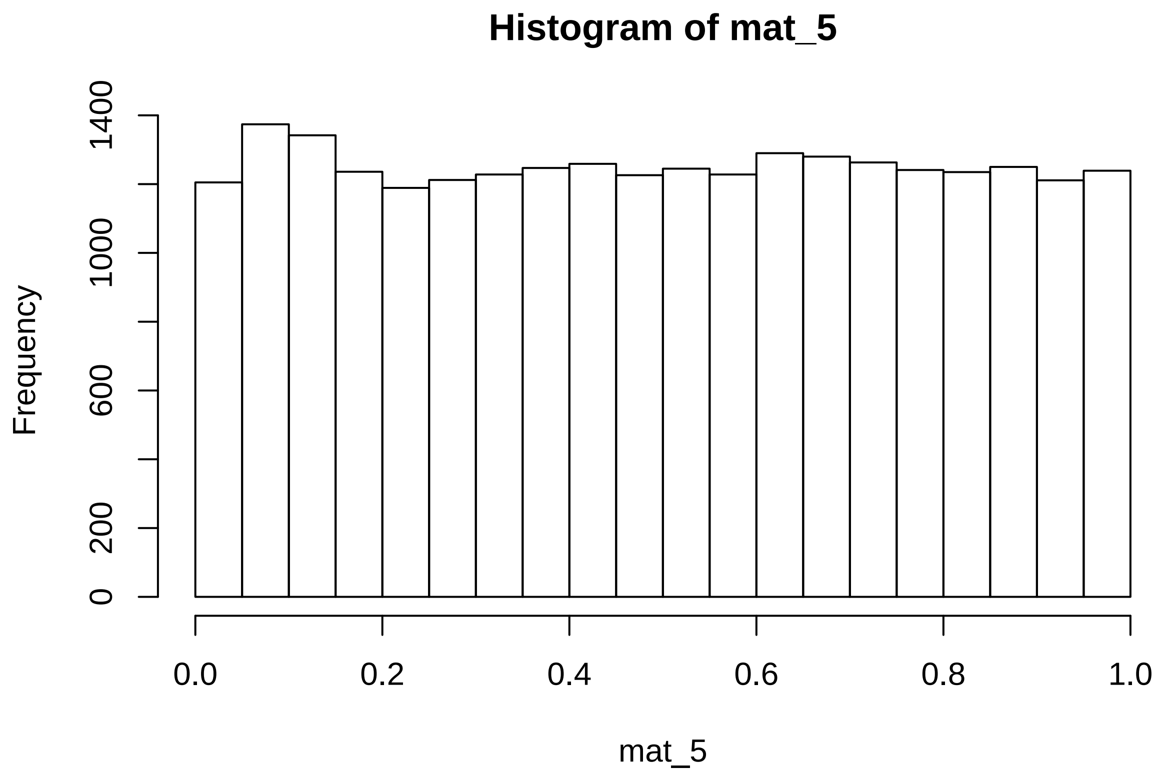
Histogram of $U_{(1)}$ with $m = 5$:

```
hist(vecmin_5, xlim=c(0,1))
```



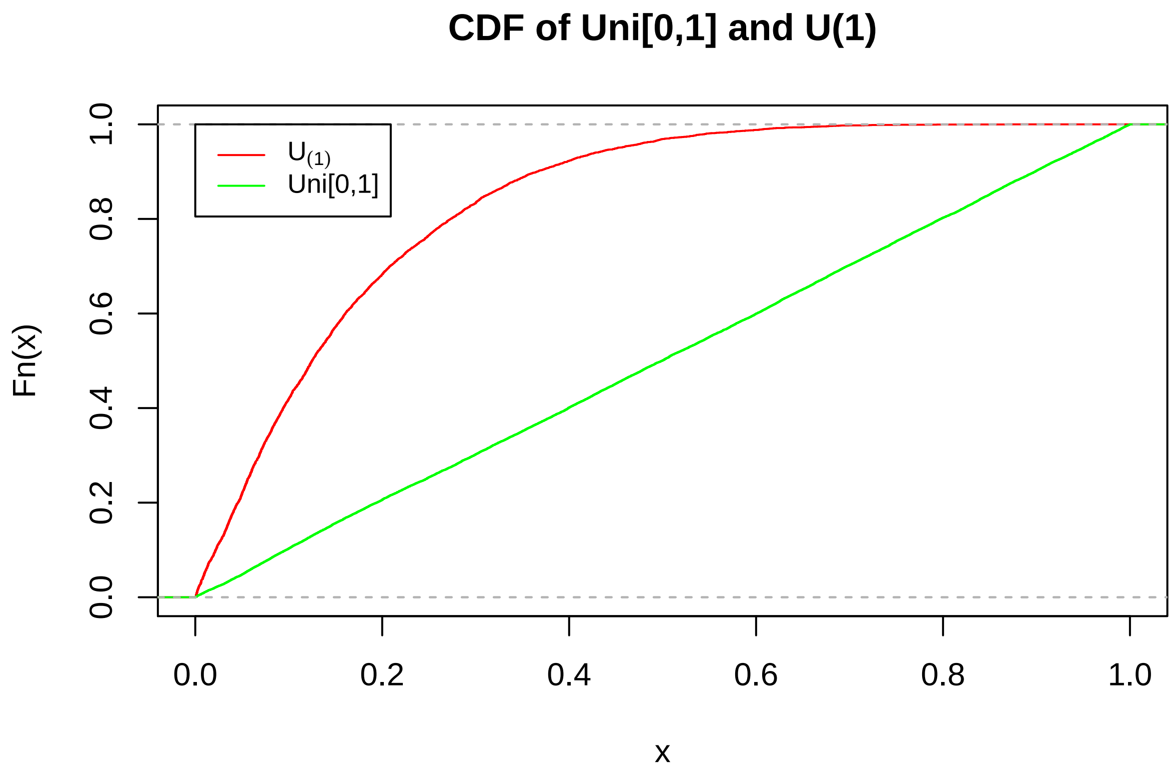
Histogram of U_i with $m = 5$:

```
hist(mat_5, xlim=c(0,1))
```



By running the ecdf function we can visual the cdf of both functions clearly.

```
plot(ecdf(vecmin_5),col='red', main='CDF of Uni[0,1] and U(1)', xlim=c(0,1))
lines(ecdf(mat_5),col='green')
legend(0, 1, legend=c(expression(U[(1)]), 'Uni[0,1]'), col=c("red", "green"), lty=1, cex=0.8)
```



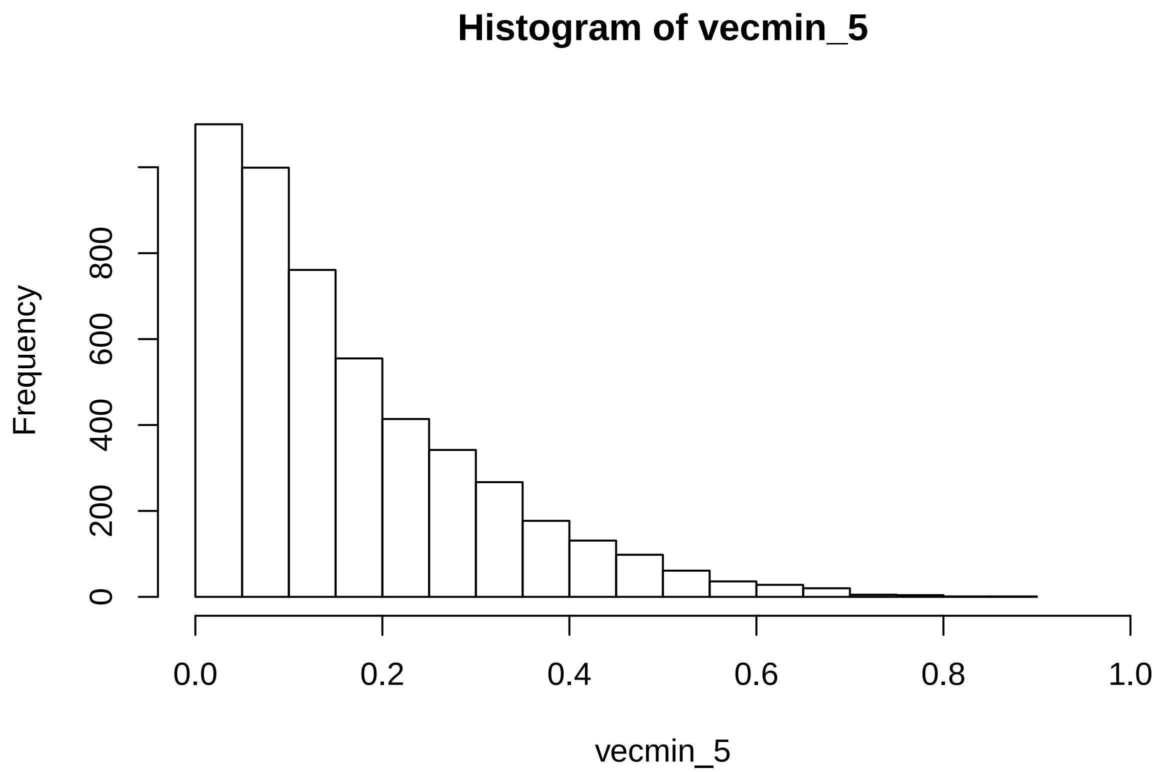
As can be seen from both the histograms and the cdf functions,

it is clear that $U_{(1)}$ is stochastically smaller $(\prec)$ than $Uni[0,1]$.

3.

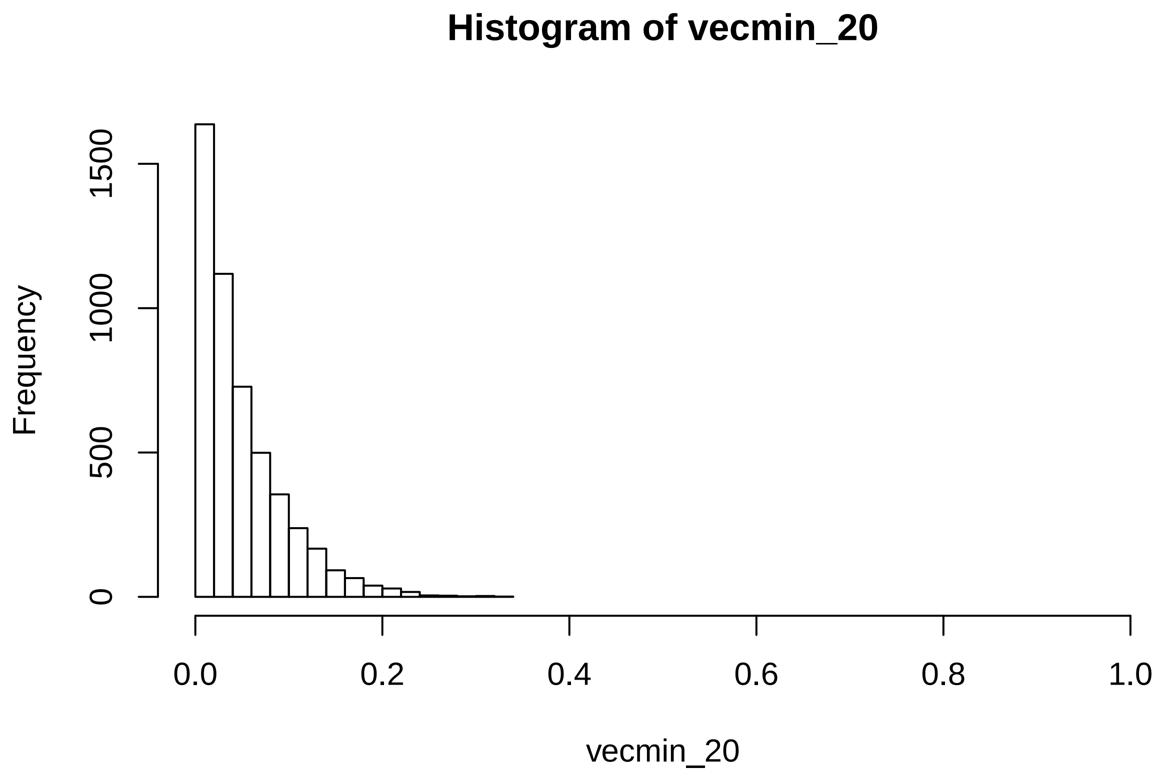
Histogram of $U_{(1)}$ with $m = 5$:

```
hist(vecmin_5, xlim=c(0,1))
```



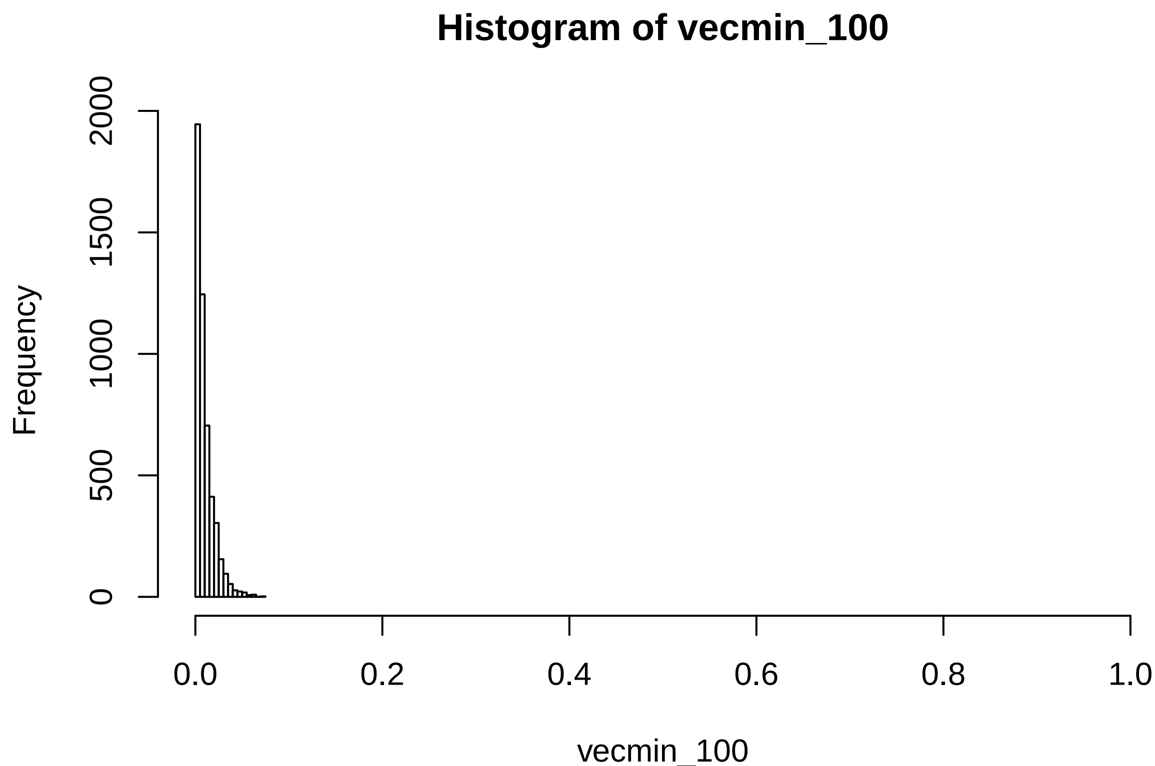
Histogram of $U_{(1)}$ with $m = 20$:

```
hist(vecmin_20, breaks = 20, xlim=c(0,1))
```



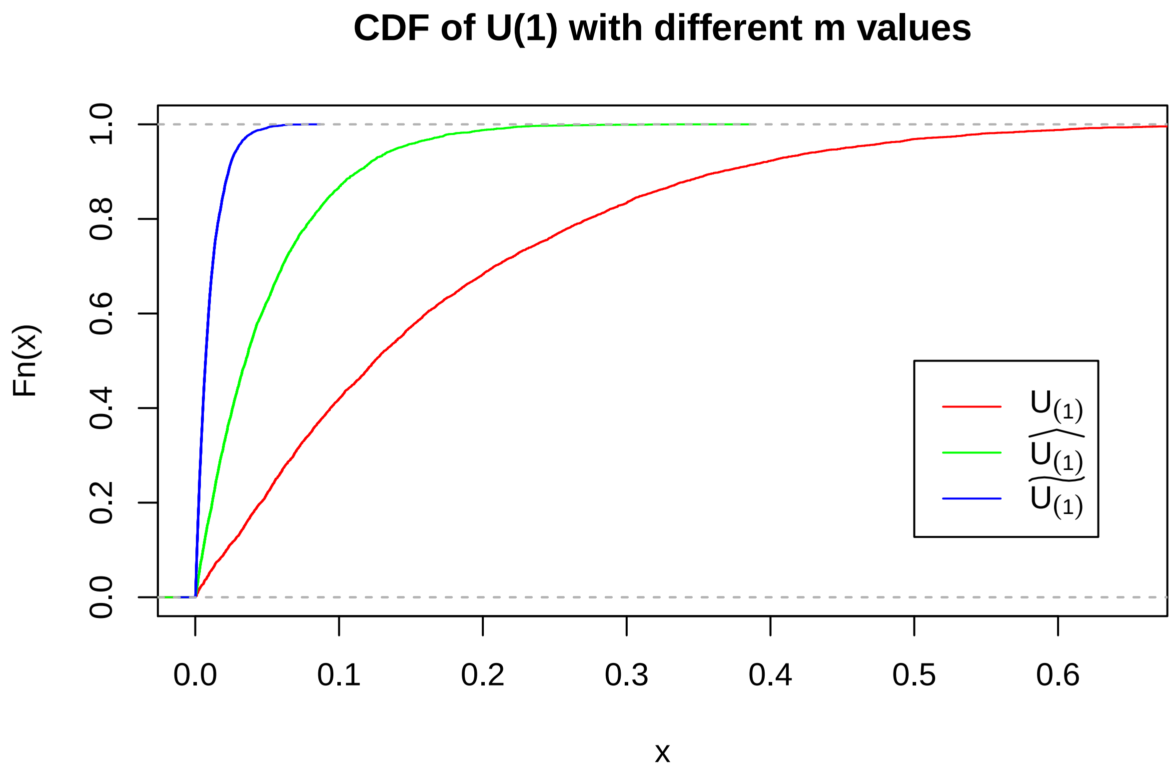
Histogram of $U_{(1)}$ with $m = 100$:

```
hist(vecmin_100, xlim=c(0,1))
```




```r
plot(ecdf(vecmin_5),col='red', main='CDF of U(1) with different m values', xlim=c(0,0.65))
lines(ecdf(vecmin_20),col='green')
lines(ecdf(vecmin_100),col='blue')

legend(0.5, 0.5, legend=c(expression(U[(1)],widehat(U[(1)]),widetilde(U[(1)]))), col=c("red", "green",
```



$U_{(1)}$ is with $m = 5$, $\widehat{U}_{(1)}$ is with $m = 20$ and $\widetilde{U}_{(1)}$ is with $m = 100$.

It can be clearly observed that:

if $m > m'$ then $U_{(1)} \prec U'_{(1)}$

if $m < m'$ then $U_{(1)} \succ U'_{(1)}$

Part B.

$F_{U_{(1)}}(x) = P(U_{(1)} < x) = 1 - P(U_{(1)} \geq x) = 1 - P(U_1 \geq x, \ldots, U_m \geq x) = 1 - P(U_1 \geq x) * \ldots * P(U_m \geq x) = 1 - (1-x)^m$, $\forall x : 0 \leq x \leq 1$

This comes from independence of U_i, definition of $U_{(1)}$ and uniform distribution.

Proof of part 2:

$U_{(1)} \prec U_i \iff P(U_{(1)} \leq a) \geq P(U_i \leq a) \iff F_{U_{(1)}}(a) \geq F_{U_i}(a) \iff 1 - (1 - F_{U_i}(a))^m \geq F_{U_i}(a) \iff 1 - F_{U_i}(a) \geq (1 - F_{U_i}(a))^m \iff 1 \geq (1 - F_{U_i}(a))^{m-1} \iff 1 \geq 1 - F_{U_i}(a) \iff F_{U_i}(a) \geq 0$ ■

Proof of part 3:

$U'_{(1)} \prec U_{(1)} \iff F_{U'_{(1)}}(a) \geq F_{U_{(1)}}(a) \iff 1 - (1 - F_{U'_i}(a))^{m'} \geq 1 - (1 - F_{U_i}(a))^m \iff (1 - F_{U'_i}(a))^{m'} \leq (1 - F_{U_i}(a))^m \iff (1 - F_{U_i}(a))^{m'} \leq (1 - F_{U_i}(a))^m \iff 1 \geq (1 - F_{U_i}(a))^{m'-m}, m' \geq m \implies 0 \leq 1 - F_{U_i}(a) \leq 1$, which is always true. ■

Part C.

$F_{U_{(1)}}(x) = 1 - (1 - F_{U_i}(x))^m = 1 - (1-x)^m$

Therefore,

$P(U_{(1)} \leq g(t,m)) = F_{U_{(1)}}(g(t,m)) = 1-(1-g(t,m))^m = t \iff 1-t = (1-g(t,m))^m \iff (1-t)^{1/m} = 1-g(t,m) \iff g(t,m) = 1-(1-t)^{1/m}$

Therefore, the function is : $g(t,m) = 1-(1-t)^{1/m}$.

Part D.

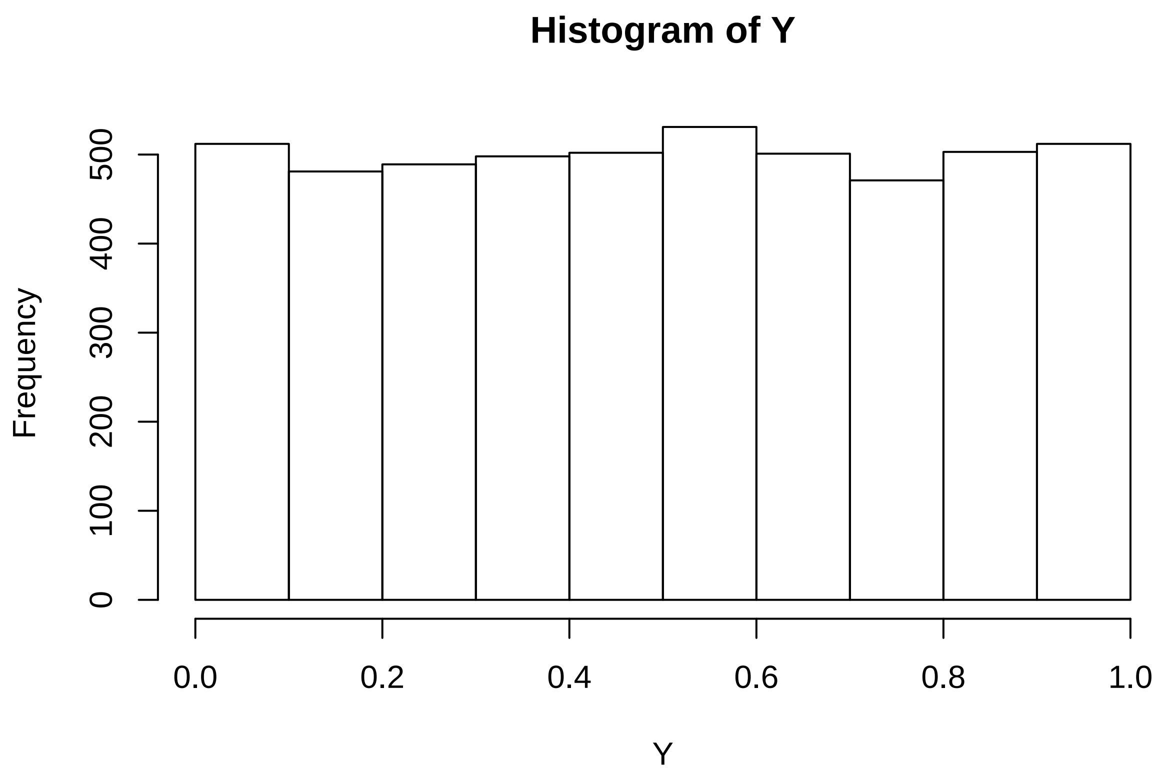
$P(U_{(1)} \leq 1-(1-t)^{1/m}) = t \iff P(1-U_{(1)} \geq (1-t)^{1/m}) = t \iff$

$P((1-U_{(1)})^m \geq 1-t) = t \iff P(1-(1-U_{(1)})^m \leq t) = t$

Therefore the random variable Y is given by: $Y = h_m(U_{(1)}) = 1-(1-U_{(1)})^m$

Part E.

```
Y <- 1-(1-vecmin_100)^100
hist(Y)
```



For $Y = 1-(1-U_{(1)})^m$,

as can be seen, the random variable Y that we found , holds $Y \sim Uni[0,1]$ as expected.

Part F.

1.

This corresponds exactly to the function we found in part C:

$p_{value}^{adj} = 1-(1-P_{value})^{\frac{1}{m}}$

where m is the total number of courses (because $m = m_0$ according to the question assumptions).

So for $p_{value} = 0.05$ we get:

$p^{adj}_{value} = 1 - (1 - 0.05)^{\frac{1}{m}} = 1 - (0.95)^{\frac{1}{m}}$

As we showed and proved in previous sections, this equation is provides the adjusted pvalue, as desired.

Question 2

Part 1

$P_i \leq 1 - (1-\alpha)^{\frac{1}{m}} = \alpha_{sid} \iff P_i + (1-\alpha)^{\frac{1}{m}} \leq 1 \iff (1-\alpha)^{\frac{1}{m}} \leq 1 - P_i \iff 1 - \alpha \leq (1 - P_i)^m \iff \alpha \geq 1 - (1 - P_i)^m$

Therefore,

$q_{i_{sid}} = 1 - (1 - P_i)^m$

Part 2

$P(V > 0) \underbrace{=}_{(1)} P(\exists i : \mu_i \notin I'_i(X)) \underbrace{\leq}_{(2)} \alpha$

(1) By definition of the test.
(2) By definition of simultaneous confidence intervals.

Therefore, the given test is a FWER-controlling procedure at level α.

Part 3

From definition and independence,

$P(\forall i, \mu_i \in I'_i(X)) = P(\mu_1 \in I'_1(X)) \cdot \ldots \cdot P(\mu_m \in I'_m(X)) = (1 - \frac{\alpha}{m})^m$

To prove that Bonferroni CI are simultaneous CI at confidence level of at least $1 - \alpha$ we need to show that $(1 - \frac{\alpha}{m})^m \geq 1 - \alpha$.

By applying the *log* function (allowed because both sides are non negative for $m \geq 0$, $0 \leq \alpha \leq 1$) on both sides we get:

$(1 - \frac{\alpha}{m})^m \geq 1 - \alpha \iff (log(1 - \frac{\alpha}{m}))^m \geq log(1 - \alpha) \iff mlog(1 - \frac{\alpha}{m}) \geq log(1 - \alpha)$

Assuming $m \geq 1$ we show that

$log(1 - \frac{\alpha}{m}) \geq log(1 - \alpha) \iff 1 - \frac{\alpha}{m} \geq 1 - \alpha \iff m \geq 1.$

From the assumption this is always true. ■

Part 4