How conservative is the Bonferroni correction for deleted Studentized residuals?

```
> rm(list=ls())
>
> source("http://www.utstat.utoronto.ca/~brunner/Rfunctions/rmvn.txt")
> Phi = rbind(c(4,2),
+             c(2,4))
> muX = c(10,10)
> n = 200
> beta0=1; beta1=1; beta2=1; sigmasq = 8
> alpha = 0.05
> tcrit = qt(1-alpha/2,n-1-3); tcrit
[1] 1.972141
> alpha/(2*n) # Bonferroni says use this instead of alpha
[1] 0.000125
> bcrit = qt(1-alpha/(2*n),n-1-3); bcrit
[1] 3.730706
>
> set.seed(9999) # For random number generation
>
```

```
> # Try one
> X = rmvn(n,muX,Phi); x1 = X[,1]; x2 = X[,2]
> epsilon = rnorm(n,0,sqrt(sigmasq))
> Y = beta0 + beta1*x1 + beta2*x2 + epsilon
> cor(cbind(x1,x2,Y))
          x1        x2         Y
x1 1.0000000 0.5233363 0.7102420
x2 0.5233363 1.0000000 0.7146543
Y  0.7102420 0.7146543 1.0000000
> mod0 = lm(Y~x1+x2)
> summary(mod0)

Call:
lm(formula = Y ~ x1 + x2)

Residuals:
    Min      1Q  Median      3Q     Max
-6.2103 -1.6583 -0.0708  1.5276  6.1990

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.27174    0.93950   2.418   0.0165 *
x1           0.92375    0.09633   9.589   <2e-16 ***
x2           0.94239    0.09635   9.781   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.41 on 197 degrees of freedom
Multiple R-squared: 0.6664,   Adjusted R-squared: 0.663
F-statistic: 196.8 on 2 and 197 DF,  p-value: < 2.2e-16

>
> rs = rstudent(mod0)
> length(rs[abs(rs)>bcrit]) # Chance of getting one or more is less than 0.05
[1] 0
> length(rs[abs(rs)>tcrit]) # Expect 10
[1] 10
>
```

```
>
> # Now simulate
> nsim = 10000 # Ten thousand
> nsigt = nsigb = numeric(nsim)
>
> for(sim in 1:nsim)
+     {
+     X = rmvn(n,muX,Phi); x1 = X[,1]; x2 = X[,2]
+     epsilon = rnorm(n,0,sqrt(sigmasq))
+     Y = beta0 + beta1*x1 + beta2*x2 + epsilon
+     mod = lm(Y~x1+x2)
+     rs = rstudent(mod)
+     nsigt[sim] = length(rs[abs(rs)>tcrit]) # Expect 10
+     nsigb[sim] = length(rs[abs(rs)>bcrit])
+     } # End simulation loop
>
```

```
> length(nsigt[nsigt>0])/nsim # Prob of at least one "outlier," uncorrected
[1] 1
> length(nsigb[nsigb>0])/nsim # Prob of at least one "outlier," CORRECTED
[1] 0.0498
>
> mean(nsigt)
[1] 9.9934
> mean(nsigb)
[1] 0.0507
>
> # Is Pr{Type I error} < 0.05 at all?
> Z = sqrt(nsim)*(0.0498-0.05)/sqrt(0.0498*(1-0.0498)); Z
[1] -0.0919407
>
```
