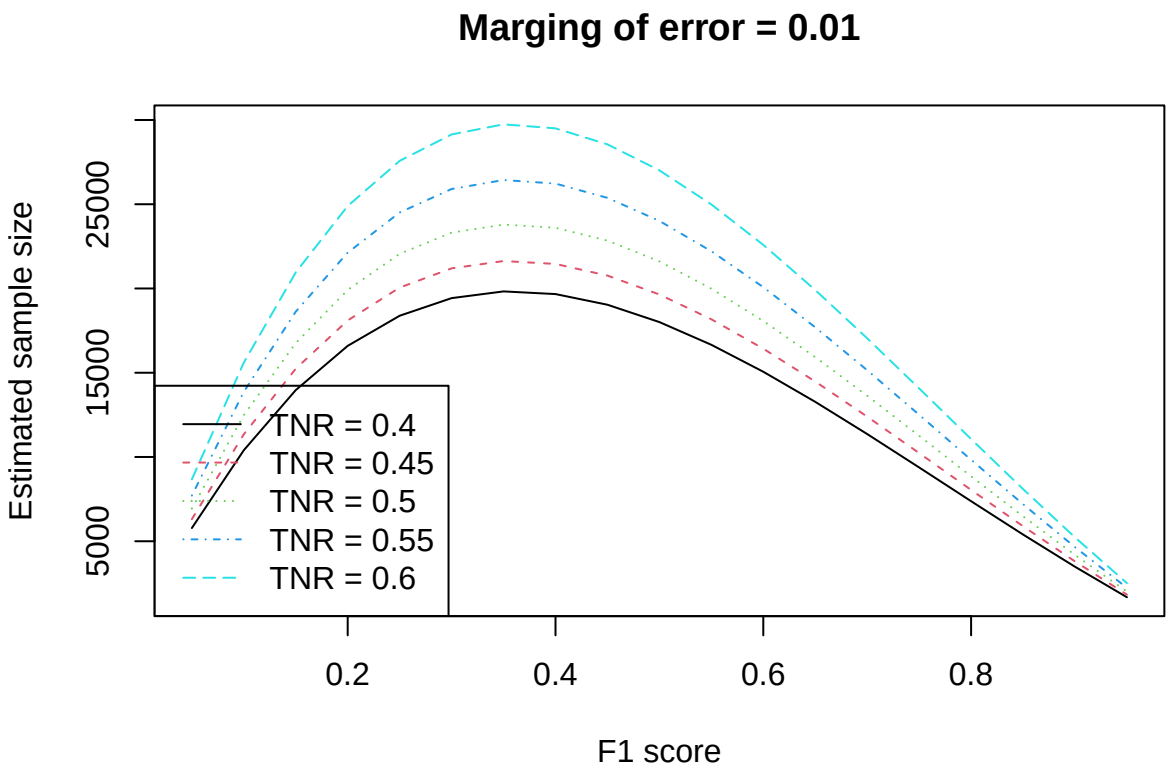
Sample size F1 score confidence intereval

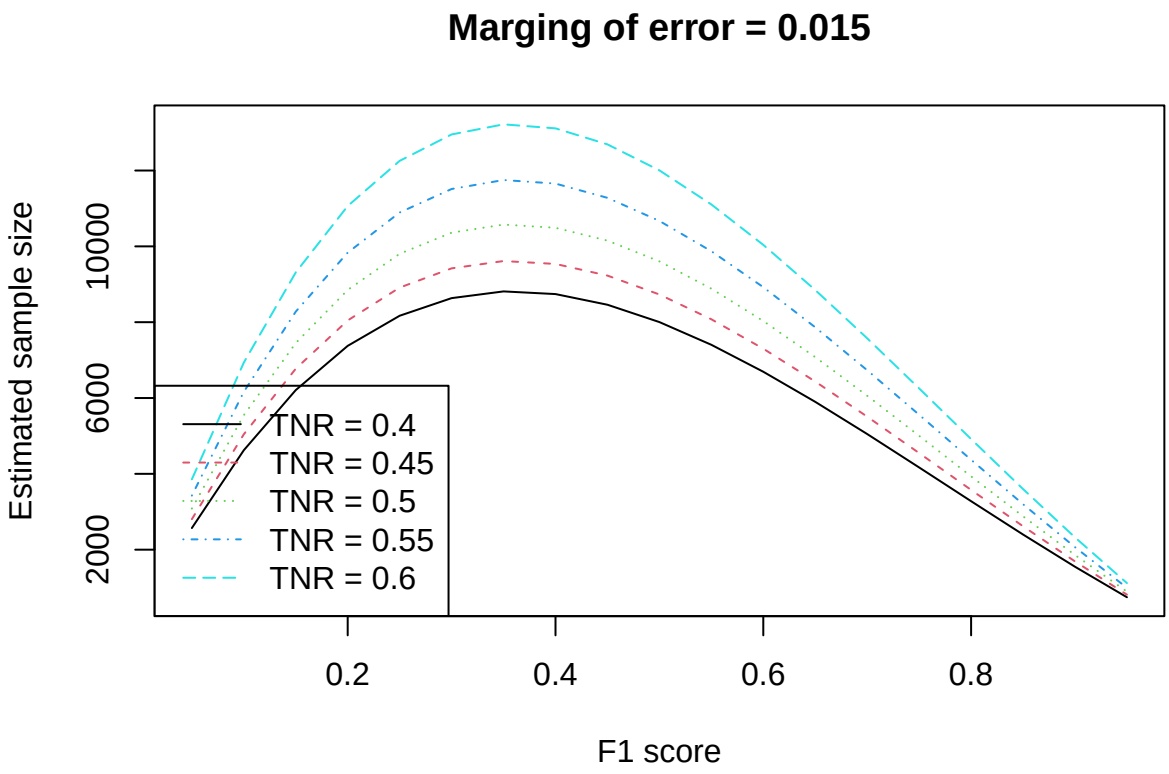
Annika Tillander

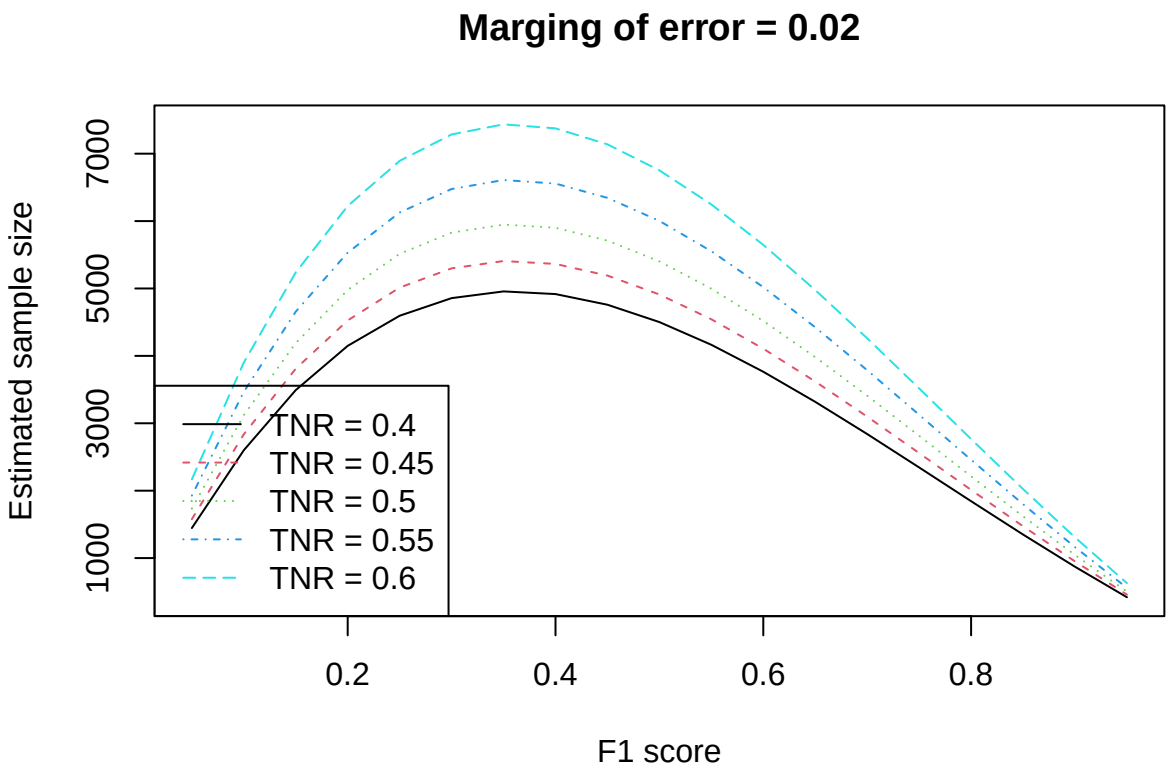
2024-05-20

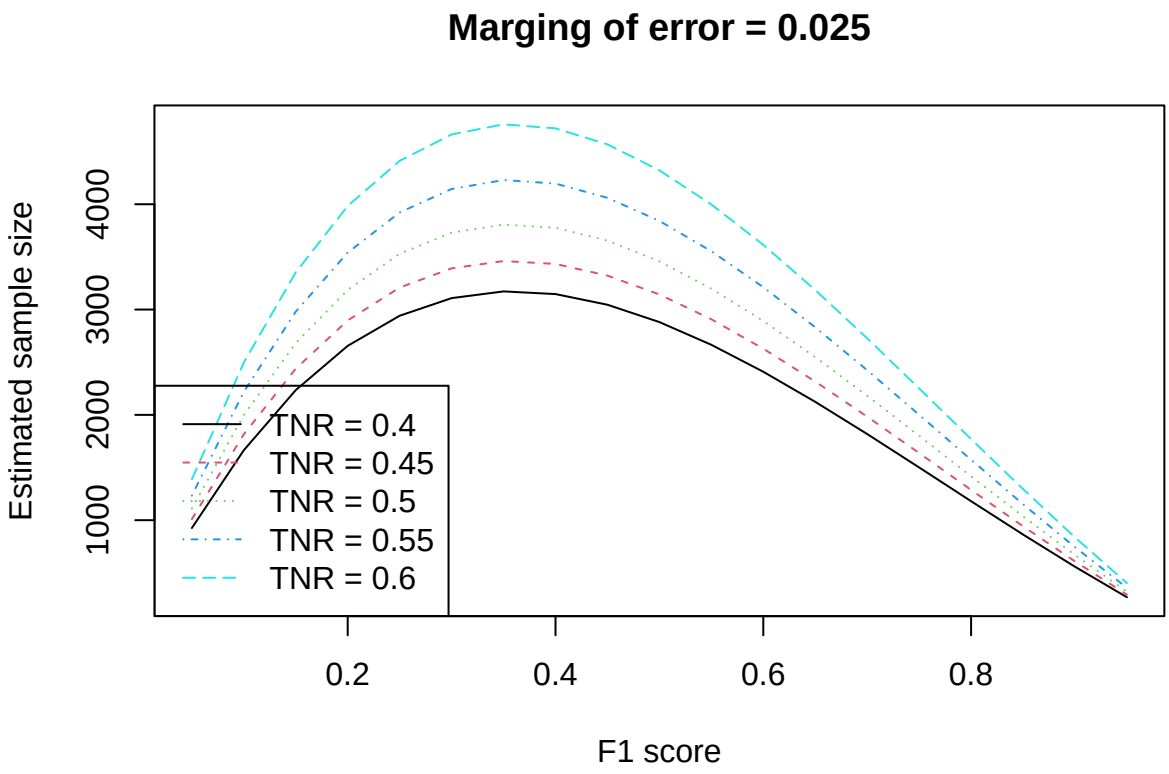
Wald method confidence interval for F1 score Based on arXiv paper by Kevin Fu Yuan Lam

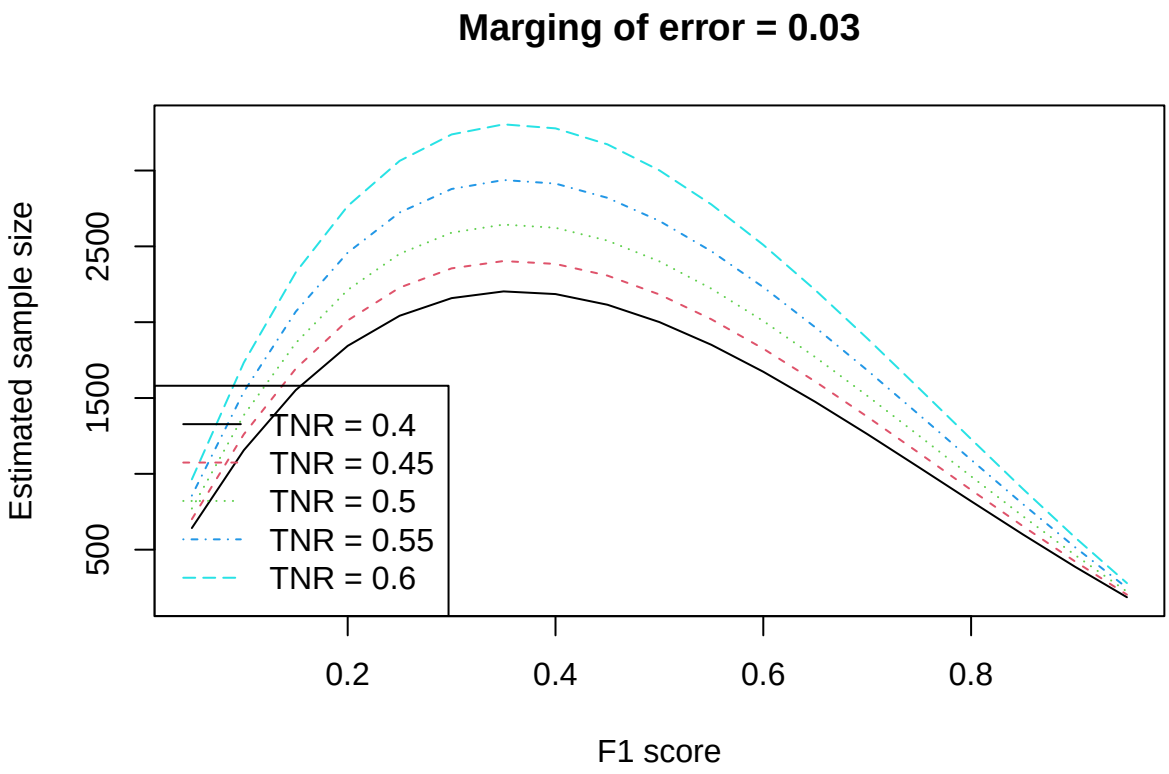
Sample size: n
True positive: n_{11}
False positive: n_{01}
False negative: n_{10}
True negative: n_{00}
TPR true positive rate: p_{11}
FPR false positive rate: p_{01}
FNR false negative rate: p_{10}
TNR true negative rate: p_{00}
$\nu = n_{11} + n_{01} + n_{10} = n - n_{00} \rightarrow n = \frac{\nu}{1-p_{00}}$
F1 score: $F_1 = \frac{2 \times n_{11}}{2 \times n_{11} + n_{01} + n_{10}} = \frac{2 \times p_{11}}{2 \times p_{11} + p_{01} + p_{10}}$
Variance for F1 score: $Var[F_1] \approx \frac{F_1 \times (1-F_1) \times (2-F_1)^2}{2 \times \nu}$
Confidence interval lower: $CI_{low} = F_1 - z_{\alpha/2} \times \sqrt{Var[F_1]}$
Confidence interval upper: $CI_{upp} = F_1 + z_{\alpha/2} \times \sqrt{Var[F_1]}$
$2 \times ME = CI_{upp} - CI_{low} = F_1 + z_{\alpha/2} \times \sqrt{Var[F_1]} - (F_1 + z_{\alpha/2} \times \sqrt{Var[F_1]}) = 2 \times (z_{\alpha/2} \times \sqrt{Var[F_1]})$
Margin of error: $ME = z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \rightarrow ME^2 = z^2_{\alpha/2} \times \frac{F_1 \times (1-F_1) \times (2-F_1)^2}{2 \times \nu}$
General form for sample size estimatation: $n_0 = \frac{z^2_{\alpha/2} \sigma^2}{ME^2}$
For F1 score?: $\nu_0 = \frac{z^2_{\alpha/2} \times [F_1 \times (1-F_1) \times (2-F_1)^2]}{2 \times ME^2} \rightarrow n_0 = \frac{z^2_{\alpha/2} \times [F_1 \times (1-F_1) \times (2-F_1)^2]}{2 \times ME^2} / (1 - p_{00})$

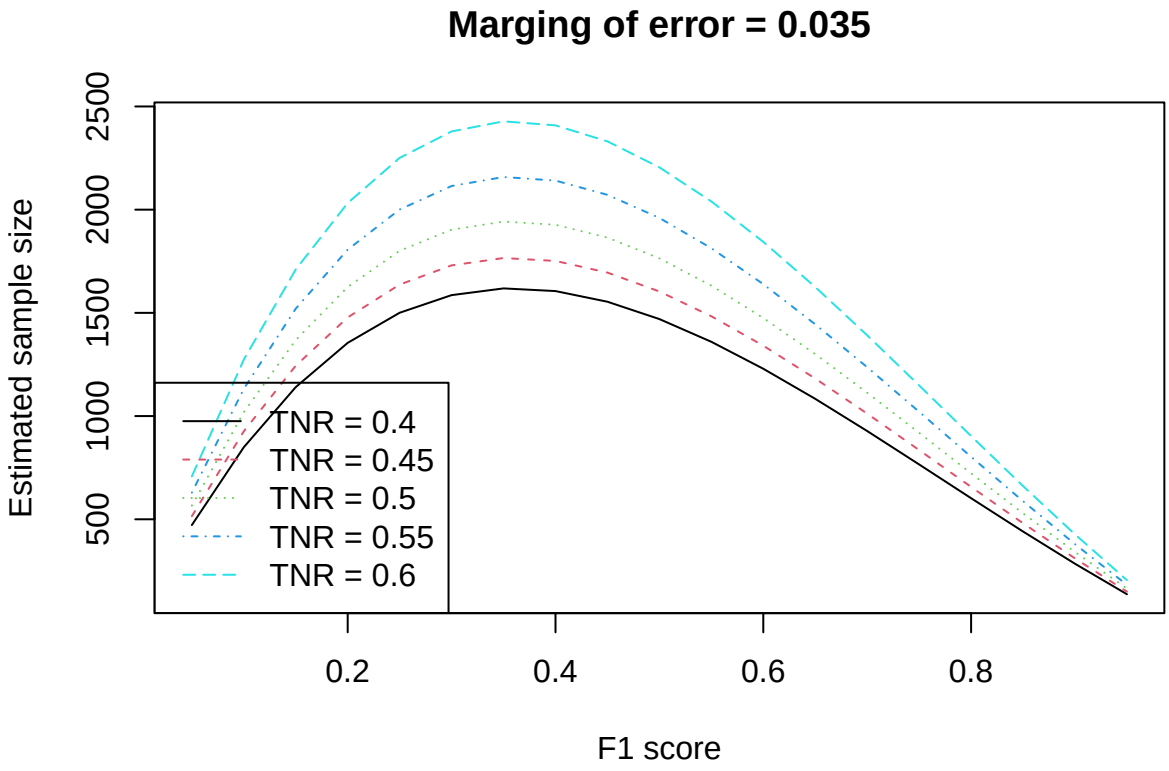

Marging of error = 0.01
Estimated sample size
5000
15000
25000
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6

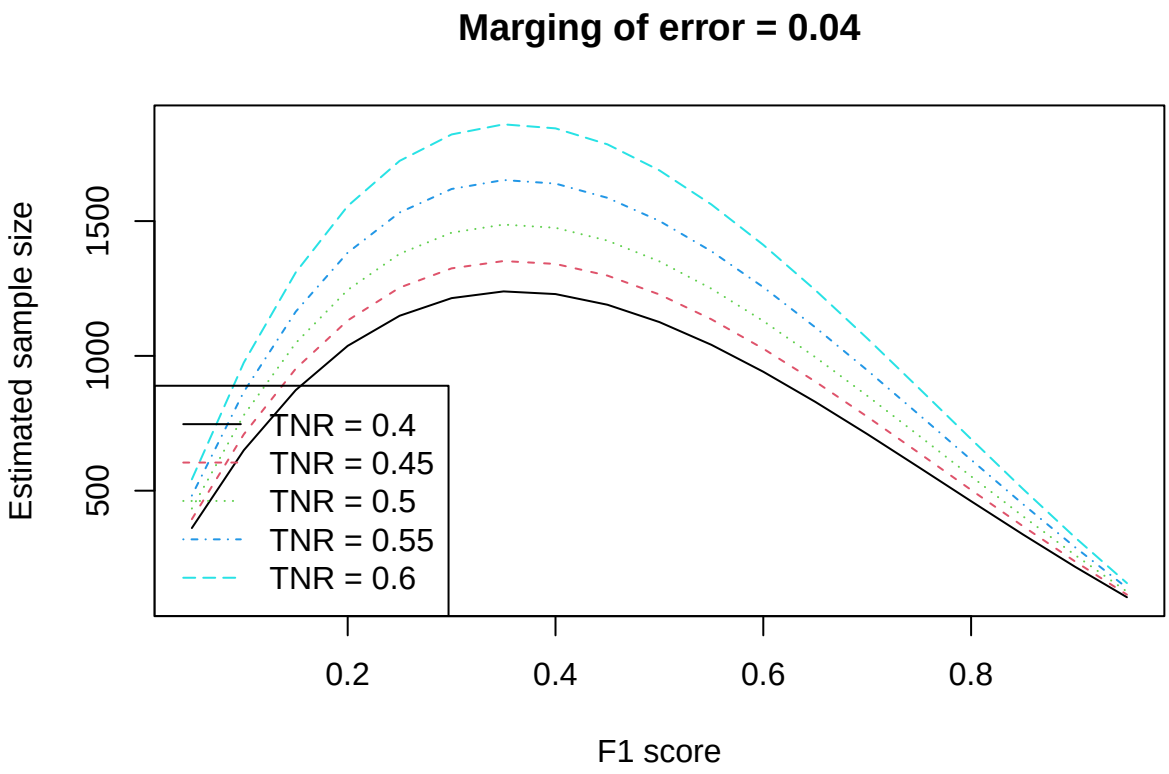

Marging of error = 0.015
Estimated sample size
2000
6000
10000
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6

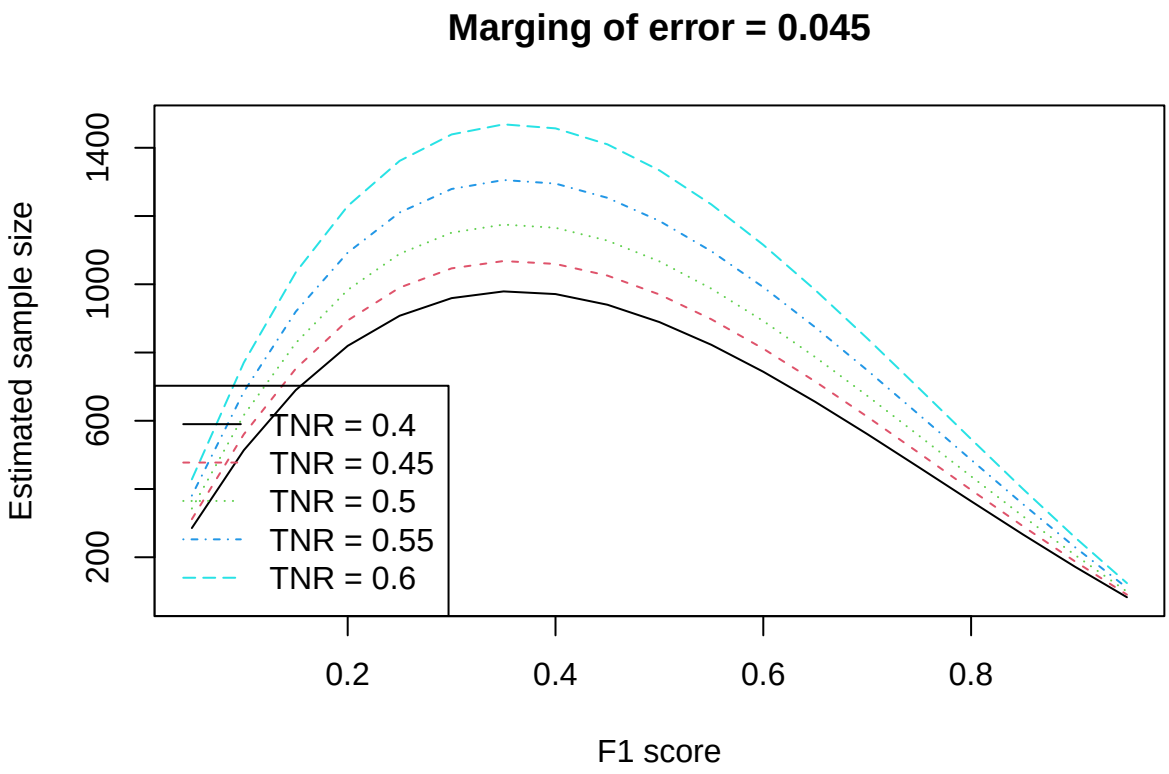

Marging of error = 0.02
Estimated sample size
1000
3000
5000
7000
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6

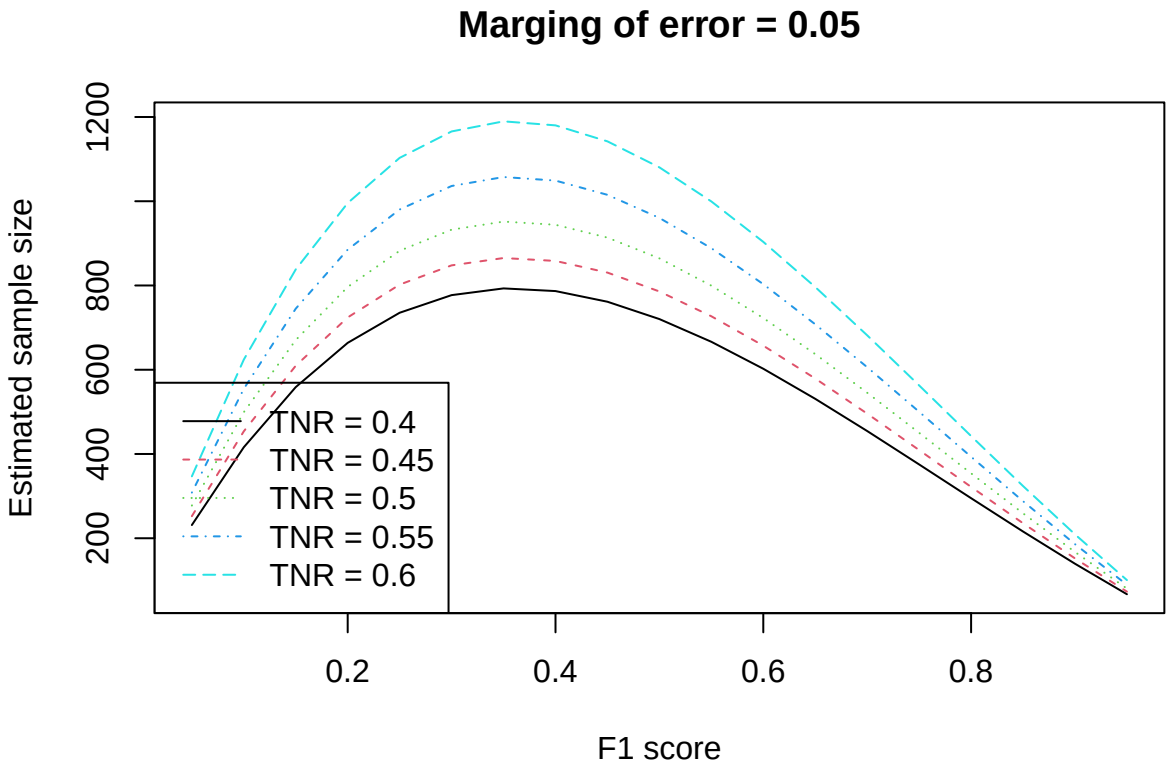

Marging of error = 0.025
Estimated sample size
1000
2000
3000
4000
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6

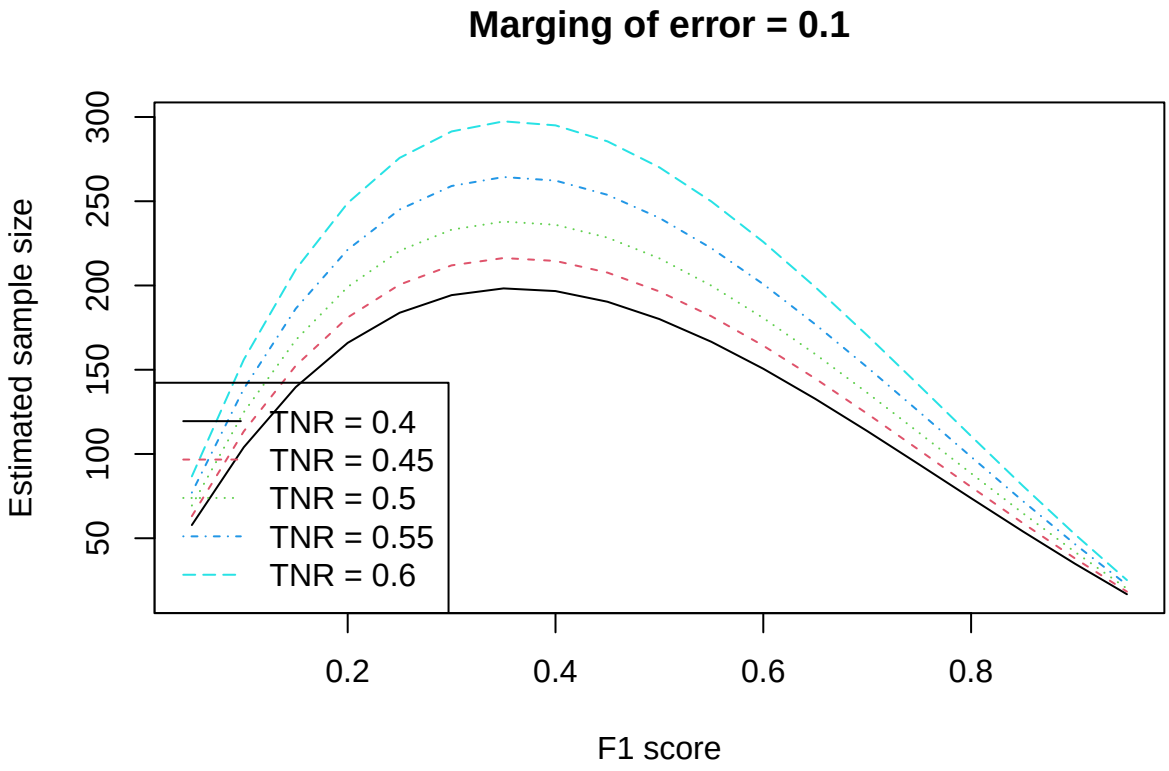

Marging of error = 0.03
Estimated sample size
500
1500
2500
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6


Marging of error = 0.035
Estimated sample size
500
1000
1500
2000
2500
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6


Marging of error = 0.04
Estimated sample size
500
1000
1500
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6


Marging of error = 0.045
Estimated sample size
200
600
1000
1400
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6


Marging of error = 0.05
Estimated sample size
200
400
600
800
1200
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6


Marging of error = 0.1
Estimated sample size
50
100
150
200
250
300
0.2
0.4
0.6
0.8
F1 score
TNR = 0.4
TNR = 0.45
TNR = 0.5
TNR = 0.55
TNR = 0.6