title: "36-402 Homework 3"
author: "Jacky Liu"
date: "2/12/21"
output: pdf_document

```
library(graphics)
ucb = as.data.frame(UCBAdmissions)
```

Problem 3

Problem 3 (a)

```
prop.test(x = c(1198, 557), n = c(1198+1493, 557+1278), alternative = "greater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(1198, 557) out of c(1198 + 1493, 557 + 1278)
## X-squared = 91.61, df = 1, p-value < 2.2e-16
## alternative hypothesis: greater
## 95 percent confidence interval:
##  0.1175222 1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.4451877 0.3035422
```

Using the two-proportions z-test as shown above where

$$H0 : P_{Male} = P_{Female}$$

$$H_A : P_{Male} > P_{Female}$$

we get a p-value of 2.2e-16. Since this value is close to 0, we reject our null hypothesis and conclude that the proportion of male applicants admitted is most likely higher than the proportion of female applicants admitted.

Problem 3 (b)

```
# Department A
prop.test(x = c(89, 512), n = c(89+19, 512+313), alternative = "greater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(89, 512) out of c(89 + 19, 512 + 313)
## X-squared = 16.372, df = 1, p-value = 2.603e-05
## alternative hypothesis: greater
## 95 percent confidence interval:
```

```
##  0.1318697 1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.8240741 0.6206061
```

```r
# Department B
prop.test(x = c(17, 353), n = c(17+8, 353+207), alternative = "greater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(17, 353) out of c(17 + 8, 353 + 207)
## X-squared = 0.085098, df = 1, p-value = 0.3853
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.1283321  1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.6800000 0.6303571
```

```r
# Department C
prop.test(x = c(202, 120), n = c(202+391, 120+205), alternative = "greater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(202, 120) out of c(202 + 391, 120 + 205)
## X-squared = 0.63322, df = 1, p-value = 0.7869
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.08541037  1.00000000
## sample estimates:
##    prop 1    prop 2
## 0.3406408 0.3692308
```

```r
# Department D
prop.test(x = c( 131,138), n = c( 131+244,138+279), alternative = "greater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(131, 138) out of c(131 + 244, 138 + 279)
## X-squared = 0.22159, df = 1, p-value = 0.3189
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.03960046  1.00000000
## sample estimates:
##    prop 1    prop 2
## 0.3493333 0.3309353
```

```r
# Department E
prop.test(x = c(94, 53), n = c( 94+299,53+138), alternative = "greater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(94, 53) out of c(94 + 299, 53 + 138)
## X-squared = 0.80805, df = 1, p-value = 0.8157
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.1061656  1.0000000
## sample estimates:
##    prop 1    prop 2
## 0.2391858 0.2774869
```

```r
# Department F
prop.test(x = c(24, 22), n = c(24+317, 22+351), alternative = "greater")
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(24, 22) out of c(24 + 317, 22 + 351)
## X-squared = 0.21824, df = 1, p-value = 0.3202
## alternative hypothesis: greater
## 95 percent confidence interval:
##  -0.02176624  1.00000000
## sample estimates:
##     prop 1     prop 2
## 0.07038123 0.05898123
```

Based on statistical tests, females were only admitted at a significantly higher rate than males in department A. However, factually they were admitted at a higher rate by department A, B, D, and F. This is a near example of Simpson's paradox because the statistics contradict when we look at the groups separately vs when we combine the groups, which shows us that we can't only look at the combined data as there is a third factor involved.

Problem 3 (c)

```r
# x: 1 = Female, 0 = Male
# y: 1 = Admitted, 0 = Rejected
# z: Departments A-F
# vec should be passed in with format: c(admitted male, admitted female, rejected male, rejected female
calc_cond_regression = function(x, vec) {
  if(x == 0){ # male
    p = vec[1] / (vec[1] + vec[3])
  } else { # female
    p = vec[2] / (vec[2] + vec[4])
  }
  return(p)
}
```

```r
# statistics for each department
A = c(512,89,313,19)
B = c(353,17,207,8)
C = c(120,202,205,391)
D = c(138,131,279,244)
E = c(53,94,138,299)
f = c(22,24,351,317)
```

r(0,A):

```r
calc_cond_regression(0, A)
```

```
## [1] 0.6206061
```

r(1,A):

```r
calc_cond_regression(1, A)
```

```
## [1] 0.8240741
```

r(0,B):

```r
calc_cond_regression(0, B)
```

```
## [1] 0.6303571
```

r(1,B):

```r
calc_cond_regression(1, B)
```

```
## [1] 0.68
```

r(0,C):

```r
calc_cond_regression(0, C)
```

```
## [1] 0.3692308
```

r(1,C):

```r
calc_cond_regression(1, C)
```

```
## [1] 0.3406408
```

r(0,D):

```r
calc_cond_regression(0, D)
```

```
## [1] 0.3309353
```

r(1,D):

```r
calc_cond_regression(1, D)
```

```
## [1] 0.3493333
```

r(0,E):

```r
calc_cond_regression(0, E)
```

```
## [1] 0.2774869
```

r(1,E):

```r
calc_cond_regression(1, E)
```

```
## [1] 0.2391858
```

r(0,f):

```r
calc_cond_regression(0, f)
```

```
## [1] 0.05898123
```

r(1,f):

```r
calc_cond_regression(1, f)
```

```
## [1] 0.07038123
```