

STAT 22000 2020 Summer HW3 Solutions

Yibi Huang

1. – 9 total points

(1a)

[1pt] Let's first load the NC births data, exclude the case with missing `marital` status, and load the `mosaic` library.

```
nc = read.csv("https://www.openintro.org/stat/data/csv/ncbirths.csv")
nc = subset(nc, is.na(marital) == FALSE)
library(mosaic)
```

The two way table is

```
tally(~ marital + lowbirthweight, data=nc)
```

```
##           lowbirthweight
## marital      low not low
##   married      49    564
##   not married   61    325
```

The R command `'addmargins'` adds the marginal totals to the two-way table.

```
addmargins(tally(~ marital + lowbirthweight, data=nc))
```

```
##           lowbirthweight
## marital      low not low Sum
##   married      49    564 613
##   not married   61    325 386
##   Sum           110    889 999
```

(1b)

[2pts in total, 1pt each]

About $386/999 \approx 0.3864 \approx 38.6\%$ of the babies were born to unmarried mothers.

About $110/999 \approx 0.110 = 11\%$ of the babies had low birth weights.

Alternatively, the answers can be computed in R as follows.

```
prop.table(tally(~ marital, data=nc))
```

```
## marital
##   married not married
## 0.6136136 0.3863864
```

```
prop.table(tally(~ lowbirthweight, data=nc))
```

```
## lowbirthweight  
##      low    not low  
## 0.1101101 0.8898899
```

(1c)

[1pt] $61/386 \approx 0.158 = 15.8\%$. Alternatively, we can compute it using the R command:

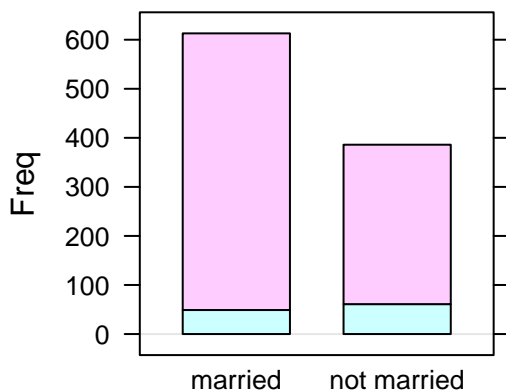
```
prop.table(tally(~ marital + lowbirthweight, data=nc),1)
```

```
##           lowbirthweight  
## marital      low    not low  
## married    0.07993475 0.92006525  
## not married 0.15803109 0.84196891
```

(1d)

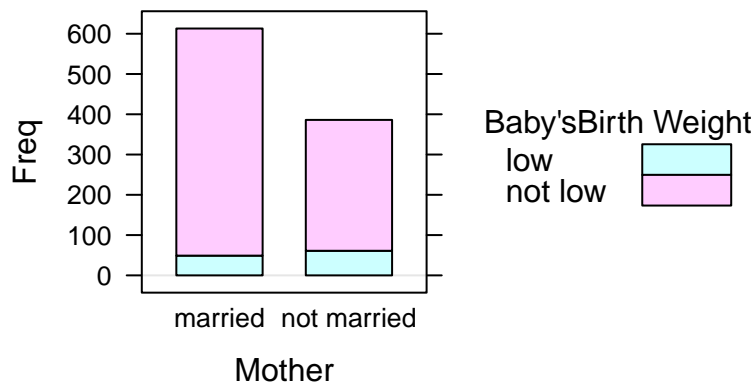
[2pts] The segmented bar chart is as follows.

```
barchart(tally(~ marital + lowbirthweight, data=nc), horizontal=FALSE)
```



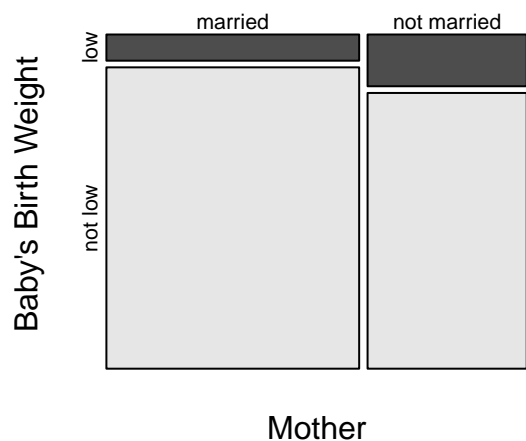
(Not Required) It would be better to add legend and labels to the barplot(s).

```
barchart(tally(~ marital + lowbirthweight, data=nc),  
         horizontal=FALSE,  
         xlab="Mother",  
         auto.key=list(space="right", title="Baby's Birth Weight", cex.title=1))
```



##(1e) [1pt]

```
mosaicplot(tally(~ marital + lowbirthweight, data=nc),
            ylab="Baby's Birth Weight",
            xlab="Mother", color=TRUE, main="")
```



(1f)

[1pt] From the mosaic plot, we see babies born to unmarried women had a higher percentage of having low birth weights than those born to married women.

(1g)

[1pt] NOT independent because the locations of the segment in the two bars do not match or because the unmarried group had a higher percentage of having low birth weight babies.

2. – 10 total points

##(2a) [3pts = 1pt for the table + 2pts for the proportions and the comparisons]

The two way table with marginal total is

```
addmargins(tally(~ marital + habit, data=nc))
```

```
##           habit
## marital      nonsmoker smoker Sum
## married         553     60 613
## not married     320     66 386
## Sum             873    126 999
```

About $60/613 \approx 0.098 \approx 9.8\%$ of the married mothers smoked during pregnancy, and $66/386 \approx 0.171 \approx 17.1\%$ of the unmarried mothers smoked during pregnancy.

Alternatively, we can compute using the R command:

```
prop.table(tally(~ marital + habit, data=nc),1)
```

```
##           habit
## marital      nonsmoker      smoker
## married    0.90212072 0.09787928
## not married 0.82901554 0.17098446
```

So the unmarried mothers were more likely to smoke during pregnancy.

(2b)

```
ftable(tally(~ habit + marital + lowbirthweight, data=nc))
```

```
##           lowbirthweight low not low
## habit      marital
## nonsmoker married          42    511
##           not married          50    270
## smoker      married          7     53
##           not married         11     55
```

```
prop.table(ftable(tally(~ habit + marital + lowbirthweight, data=nc)),1)
```

```
##           lowbirthweight      low      not low
## habit      marital
## nonsmoker married          0.07594937 0.92405063
##           not married          0.15625000 0.84375000
## smoker      married          0.11666667 0.88333333
##           not married          0.16666667 0.83333333
```

(b-i) [2pts] The percentages were $42/(42 + 511) \approx 7.6\%$ for married nonsmoking moms, and $50/(50 + 270) \approx 15.6\%$ for unmarried nonsmoking moms. Married nonsmoking moms were **less likely** to have low-birth-weight babies than unmarried nonsmoking moms.

(b-ii) [2pts] The percentages were $7/(7 + 53) \approx 11.7\%$ for married smoking moms, and $11/(11 + 55) \approx 16.7\%$ for unmarried smoking moms. Married smoking women were **less likely** to have low-birth-weight babies than unmarried smoking women.

(b-iii) [1pt] After adjusting for smoking status, unmarried women still had higher rates of having low-birth-weight babies than married women. Out-of-marriage births are still associated with low-birth-weights, even after accounting for the effect of smoking.

(2c)

[2pts] No, this is an observational study. Even if we tried to rule out one confounding factor of smoking, there can be other confounders like mothers' age, income, or education level that may be associated with mothers' marital status and how likely babies may have low birth weights. We cannot make a causal conclusion here.

3. – 6 total points

(3a)

[2pts] Yes. Since subjects were **randomized** to receive either bupropion or a placebo.

(3b)

[2pts] Yes, the study blocked on whether a subject lived with a smoker.

(3c)

[2pts] No, since the subjects were **not randomized** to live with a smoker or not to live with a smoker. The comparison between those who lived with a smoker and those who didn't is observational and is hence subject to confounding.