PSTAT200A Project 1

Samantha Sambado

11/2/2021

Title

My body is your body: a mother's impact on their babies health

Executive Summary

Young mothers that do not smoke are more likely to have a baby that is brought to full term and survives to year 1. Although smoking in general decreases the probability of a baby being brought to full term and surviving, based on the given data, a mother's age may be a more important factor in gestation and survival outcomes.

Introduction

Mothers play a fundamental role in the health of their children, especially during pregnancy when babies rely solely on their mother for nutrients. Oxygen, food, and chemicals from the mother's blood can be shared with the baby via the umbilical cord (**Dawe et al. 2007**). This kind of physiological connection can be beneficial in most cases or negative if the mother is a smoker. Smoking during pregnancy increases the risk of health problems for babies and moms (**Banderali et al. 2015**).

Current literature has shown that smoking causes lung cancer and can cause serious problems for babies while in utero. Some potential negative effects for babies could be babies born premature and birth defects which can create long term health problems or sudden infant death syndrome (SIDS) (**CDC 2017**). The literature also suggests that older women, particularly over the age of 40, have an increased rate of complications during pregnancy that can result in negative health effects for both baby and mom, but this literature is not as clear as the effects of smoking habits on baby's health.

A survey of **6,851** responses was collected regarding mother and baby demographic information. For the mother, their age (young or older) and smoking habits (no or yes) were collected. For the baby, the gestation period (premature or full term) and survival outcome within its 1st year (died or alive) was recorded. Majority of the respondents were young mother's (**n = 4,915; 72%**) or non smokers (**n = 6,197; 90%**). Majority of babies were brought to full term and were alive at their 1st year (**n = 6,189; 90%**). A potential outlier, or inconsistency, within this dataset would be an older mother who smoked and gave birth to 1) a full term baby that died at year 1 (**n = 1**), or 2) a premature baby that died at year 1 (**n = 4**).

Methods

For each part of the report, I manually summed counts for the particular groupings of interest to address the individual questions. Summary tables of counts for each questions and visualized with mosaic plots can be found in the Appendix (**Table 2 - 8**). All analyses in R were completed with version 1.4.1717.

Table 1: Full data table

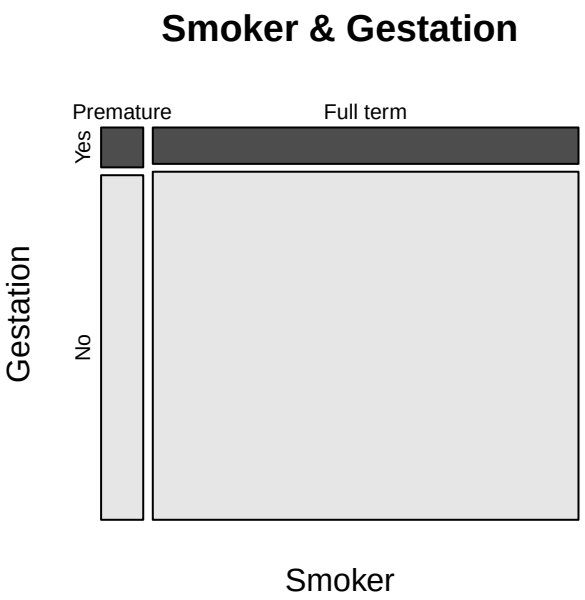
	Age	Smoker	Premature (died)	Premature (alive)	Full term (dead)	Full term (alive)	Total_row
	Young	No	50	315	24	4012	4401
	Young	Yes	9	40	6	459	514
	Older	No	41	147	14	1594	1796
	Older	Yes	4	11	1	124	140
Total_column			104	513	45	6189	6851

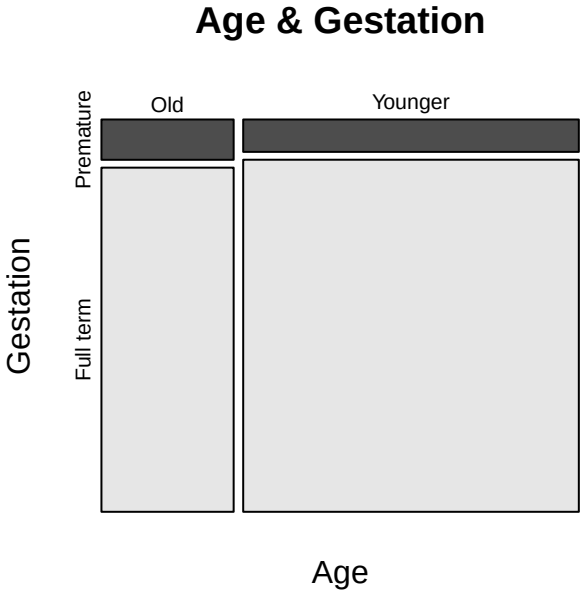
To compare the age of mother's to their smoking habit (**part A**), I performed a proportion comparison using the `prop.test()` function in R. I compared the counts of young and older mother's that reported yes for smoking. To test the hypothesis: $H_0 : p_{young} = p_{old}$; $H_1 : p_{young} \neq p_{old}$

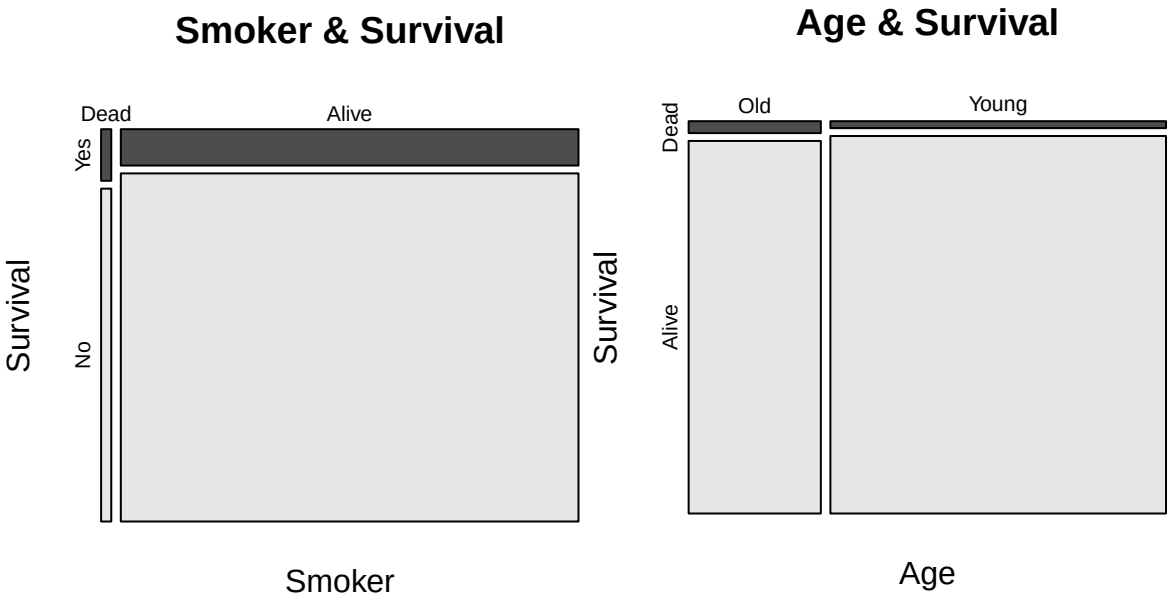
To test the independence of two variables (**part B & C**), particularly the health of the baby with characteristics of the mother, a Pearson's Chi-squared test with Yates' continuity correction was conducted with the `chi.square()` function in R. We tested the following hypotheses:

- **Part B1** $H_0 : p_{smokerhabit} = p_{gestation}$; $H_1 : p_{smokerhabit} \neq p_{gestation}$.
- **Part B2** $H_0 : p_{age} = p_{gestation}$; $H_1 : p_{age} \neq p_{gestation}$.
- **Part C1** $H_0 : p_{smokerhabit} = p_{survival}$; $p_{smokerhabit} \neq p_{survival}$.
- **Part C2** $H_0 : p_{age} = p_{survival}$; $p_{age} \neq p_{survival}$.

To look at how mother's characteristics affect baby survival (**part D, E, F**), a binomial logistic model was performed with `glm()` function. A logistic regression model allows the outcome variable, y, to be binary for 1 = success and 0 = failure. The binomial outcome was deemed **success** if the *baby survived (alive)* or if gestation was carried to *full term* and **failure** if the baby did *not survive (dead)* or was a *premature* birth. Tables for data used in each binomial model(**Table 9 - 11**) and model outputs (**Fig 1 - 5**)are in the Appendix (Table).







Results

Part A: Younger mothers are more likely to smoke. Based on the proportion test, the proportion of young mothers and the proportion of older mothers that are likely to smoke are significantly not equal (**X-squared = 16.372, df = 1, p-value = 5.204e-05**). The proportion of young and older mothers that smoke is 0.10 and 0.07, respectively.

Part B: Smokers are more likely to have premature babies than non-smokers, 78.6% and 8.9%, respectively. Based on the chi-square test, smoking habit and gestation outcome preterm are not independent of each other (**X-squared = 2178.5, df = 1, p-value < 2.2e-16**). Although not as statistically significant as smoking habit, older mothers are more likely to have premature babies compared to young mothers, 10.4% and 9.2%, respectively. Based on the chi-square test, mother age is not independent of gestation outcome preterm (**X-squared = 6.9592, df = 1, p-value = 0.008339**).

Part C: Smokers are not more likely than non smokers to have babies that died in the 1st year, 3.0% and 2.1%, respectively. Based on the chi square test, smoking habit is independent of baby survival in the 1st year (**X-squared = 2.212, df = 1, p-value = 0.1369**). Older mothers are more likely to have babies who die in the 1st year compared to young mothers, 3.1% and 1.8%, respectively.Based on the chi-quare test, age and survival outcome died are not independent of each other (**X-squared = 10.239, df = 1, p-value = 0.001375**).

Part D: For part 1, *age* significantly affects *survival* whereas smoking does not. Specially, young mothers are positive coefficient in the model (Appendix **Fig 1**: **estimate = 0.5679, std. error = 0.1697, z-value = 3.347, p-value = 0.000816**) with the probability of baby survival (success = 1) increasing by 0.56, or odds ratio of 1.76, with a young mother. For part 2, *age* significantly affects *gestational* outcome whereas smoking does not. Specially, young mothers are a postivie coefficient in the model (Appendix **Fig 2**: **estimate = 0.24589, std. error = 0.09037, z-value = 2.721, p-value = 0.00651**) with the probability of gestational outcome full term (success = 1) increasing by 0.25, or odds ratio of 1.74, with a young mother.

Part E : *Age* significantly affects *survival* among *premature* babies whereas smoking does not. Specially, young mothers are positive coefficient in the model (Appendix **Fig 3**: **estimate = 0.5551, std. error = 0.2208, z-value = 2.514, p-value = 0.0119**) with the probability of premature baby survival (success = 1) increasing by 0.56, or odds ratio of 1.74, with a young mother.

Part F: Smoking habits does significantly affect survival among premature babies with young mothers (Appendix **Fig 4**: **estimate = 1.8405, std. error = 0.1522, z-value = 12.090, p-value <2e-16**) and

with older mothers - although older mothers have less of an affect (Appendix **Fig 5**: **estimate = 1.2769, std. error = 0.1766, z-value = 7.230, p-value = 4.84e-13**).

Conclusions & Recommendations

Contemporary trends highly encourage women to not smoke (or drink) during pregnancy, although the data collected for this survey suggests that a woman's smoking habit may be less of an effect than a mother's age on gestation and survival. However, I am cautious to recommend smoking for pregnant women and would like to explore these questions with a fuller set of explanatory variables such as a woman's socioeconomic status, access to health care, and adherence to pregnancy vitamins and a healthy diet. It is possible that when this data set was collected (pre-1980?), smoking was not as taboo as it is now in 2000s. Before smoking became a taboo, smoking was a sign of wealth because cigarettes are expensive. It may be that when this data was collected, smoking was partaken by wealthy women who had the means to achieve a healthy and successful pregnancy due to their access to health care and support system.

However, a trend that has remained in place is the negative association and outcomes for older women who decide to give birth. This current trend of delayed childbearing is exacerbated in cities with highly educated women who delay childbearing in order to pursue a career (**Cooke et al. 2012**). With an advancement of a career, older woman may have increased access to similar wealth metrics that may have skewed the smoking data trend in the previous paragraphs. However, despite these economic advantages, physiology dictates that it is biologically easier to have a successful pregnancy at an earlier age no matter the economic status.

Another factor to keep in mind which may skew results is the country where this study took place. The prevalence of smoking during pregnancy varies globally based on cultural norms. In industrialized countries, prevalence rates have become to decline whereas in other countries smoking is becoming increasingly common among young women (**Cnattingius et al. 2004**). More epidemiological information regarding survey respondents place of residence would be beneficial to make sure all respondents are either coming from the same region or country or are randomly sampled from a global population.

I do not make recommendations for when a woman decides to have a child or not and believe such recommendations should be personal with access to the most current available data and advice from a physician. However, I will generally recommend that smoking is not the best chemical to share with your baby during pregnancy if given the option for a healthier lifestyle. It is never to late to stop smoking to protect the mother's and baby's health.

Additional links to support for quitting during pregnancy

> Pregnancy, Motherhood, and Smoking
>
> Feelings, Partners, and Friends

Appendix

References

Scientific literature

- **Banderali**, G., A. Martelli, M. Landi, F. Moretti, F. Betti, G. Radaelli, C. Lassandro, and E. Verduci. 2015. Short and long term health effects of parental tobacco smoking during pregnancy and lactation: a descriptive review. Journal of Translational Medicine 13:327.
- **CDC**. 2017. Tobacco use and pregnancy.
- **Cnattingius**, S. 2004. The epidemiology of smoking during pregnancy: Smoking prevalence, maternal characteristics, and pregnancy outcomes. Nicotine & Tobacco Research 6:125–140.

- **Cooke**, A., T. A. Mills, and T. Lavender. 2012. Advanced maternal age: delayed childbearing is rarely a conscious choice a qualitative study of women's views and experiences. International journal of nursing studies 49:30–9. Dawe, G. S., X. W. Tan, and Z.-C. Xiao. 2007. Cell migration from baby to mother. Cell adhesion & migration 1:19–27.

R packages

```
## @Manual{R-ggplot2,
##   title = {ggplot2: Create Elegant Data Visualisations Using
##     the Grammar of Graphics},
##   author = {Hadley Wickham and Winston Chang and Lionel
##     Henry and Thomas Lin Pedersen and Kohske Takahashi and
##     Claus Wilke and Kara Woo and Hiroaki Yutani and Dewey
##     Dunnington},
##   year = {2021},
##   note = {R package version 3.3.5},
##   url = {https://CRAN.R-project.org/package=ggplot2},
## }
##
## @Manual{R-gtsummary,
##   title = {gtsummary: Presentation-Ready Data Summary and
##     Analytic Result Tables},
##   author = {Daniel D. Sjoberg and Michael Curry and Margie
##     Hannum and Joseph Larmarange and Karissa Whiting and
##     Emily C. Zabor},
##   year = {2021},
##   note = {R package version 1.4.2},
##   url = {https://CRAN.R-project.org/package=gtsummary},
## }
##
## @Manual{R-kableExtra,
##   title = {kableExtra: Construct Complex Table with kable
##     and Pipe Syntax},
##   author = {Hao Zhu},
##   year = {2021},
##   note = {R package version 1.3.4},
##   url = {https://CRAN.R-project.org/package=kableExtra},
## }
##
## @Manual{R-knitr,
##   title = {knitr: A General-Purpose Package for Dynamic
##     Report Generation in R},
##   author = {Yihui Xie},
##   year = {2021},
##   note = {R package version 1.36},
##   url = {https://yihui.org/knitr/},
## }
##
## @Manual{R-tidyverse,
##   title = {tidyverse: Easily Install and Load the
##     Tidyverse},
##   author = {Hadley Wickham},
##   year = {2021},
```

```
##   note = {R package version 1.3.1},
##   url = {https://CRAN.R-project.org/package=tidyverse},
## }
##
## @Book{ggplot22016,
##   author = {Hadley Wickham},
##   title = {ggplot2: Elegant Graphics for Data Analysis},
##   publisher = {Springer-Verlag New York},
##   year = {2016},
##   isbn = {978-3-319-24277-4},
##   url = {https://ggplot2.tidyverse.org},
## }
##
## @Book{knitr2015,
##   title = {Dynamic Documents with {R} and knitr},
##   author = {Yihui Xie},
##   publisher = {Chapman and Hall/CRC},
##   address = {Boca Raton, Florida},
##   year = {2015},
##   edition = {2nd},
##   note = {ISBN 978-1498716963},
##   url = {https://yihui.org/knitr/},
## }
##
## @InCollection{knitr2014,
##   booktitle = {Implementing Reproducible Computational
##     Research},
##   editor = {Victoria Stodden and Friedrich Leisch and Roger
##     D. Peng},
##   title = {knitr: A Comprehensive Tool for Reproducible
##     Research in {R}},
##   author = {Yihui Xie},
##   publisher = {Chapman and Hall/CRC},
##   year = {2014},
##   note = {ISBN 978-1466561595},
##   url = {http://www.crcpress.com/product/isbn/
##     9781466561595},
## }
##
## @Article{tidyverse2019,
##   title = {Welcome to the {tidyverse}},
##   author = {Hadley Wickham and Mara Averick and Jennifer
##     Bryan and Winston Chang and Lucy D'Agostino McGowan and
##     Romain François and Garrett Grolemund and Alex Hayes
##     and Lionel Henry and Jim Hester and Max Kuhn and Thomas
##     Lin Pedersen and Evan Miller and Stephan Milton Bache
##     and Kirill Müller and Jeroen Ooms and David Robinson and
##     Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi
##     and Davis Vaughan and Claus Wilke and Kara Woo and
##     Hiroaki Yutani},
##   year = {2019},
##   journal = {Journal of Open Source Software},
##   volume = {4},
##   number = {43},
```

```
##   pages = {1686},
##   doi = {10.21105/joss.01686},
## }
```

Addition Tables and Figures

Count tables for analyses

Table 2: Age and smoking habits

	Smoker	
	Yes	No
Young	514	4,401
Older	140	1,796

Part A

Table 3: Gestation and smoking habits

	Smoker	
	Yes	No
Premature	64	553
Full term	590	5,644

Table 4: Gestation and age

	Gestation	
	Premature	Full term
Old	203	1,733
Younger	414	4,501

Part B

Table 5: Baby survival and smoking habits

	Smoker	
	Yes	No
Dead	20	129
Alive	634	6,068

Table 6: Baby survival and age

	Baby Survival	
	Dead	Alive
Old	60	1,876
Young	89	4,826

Part C

Table 7: Gestation (died = 1, alive = 0)

Survival	Age	Smoker	Count	Percentage
0	young	no	74	1.1
1			4,327	63.2
0		yes	15	0.2
1			499	7.3
0	old	no	55	0.8
1			1,741	25.4
0		yes	5	0.1
1			135	2.0

Table 8: Gestation (full term = 1, premature = 0)

Gestation	Age	Smoker	Count	Percentage
0	young	no	365	5.3
1			4,036	58.9
0		yes	49	0.7
1			465	6.8
0	old	no	188	2.7
1			1,608	23.5
0		yes	15	0.2
1			125	1.8

Part D

Table 9: Survival of premature babbies (alive = 1, dead = 0)

Premature survival	Age	Smoker	Count	Percentage
0	young	no	50	8.1
1			315	51.1
0		yes	9	1.5
1			40	6.5
0	old	no	41	6.6
1			147	23.8
0		yes	4	0.6
1			11	1.8

Part E

Table 10: Survival of premature babbies with young mothers (alive = 1, dead = 0)

Premature survival	Smoker	Count	Percentage
0	no	50	12.1
1	no	315	76.1
0	yes	9	2.2
1	yes	40	9.7

Table 11: Survival of premature babbies with older mothers (alive = 1, dead = 0)

Premature survival	Smoker	Count	Percentage
0	no	41	20.2
1	no	147	72.4
0	yes	4	2.0
1	yes	11	5.4

Part F

Model outputs

Model outputs for D

```
##
## Call:
## glm(formula = Survival ~ Age + Smoker, family = binomial(link = "logit"),
##     data = d1_uncount)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.8519   0.1859   0.1859   0.2462   0.3060
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   3.4816     0.1337  26.043  < 2e-16 ***
## Ageyoung      0.5679     0.1697   3.347 0.000816 ***
## Smokeryes    -0.4438     0.2447  -1.813 0.069764 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1435.5  on 6850  degrees of freedom
## Residual deviance: 1422.4  on 6848  degrees of freedom
## AIC: 1428.4
##
## Number of Fisher Scoring iterations: 6


##
## Call:
```

Characteristic	OR[1]	95% CI[1]	p-value
Age			
old	—	—	
young	1.76	1.26, 2.45	**<0.001**
Smoker			
no	—	—	
yes	0.64	0.41, 1.07	0.070

[1] OR = Odds Ratio, CI = Confidence Interval

Figure 1: Survival (0/1) ~ Age + Smoker (Part D1)

```
## glm(formula = Gestation ~ Age + Smoker, family = binomial(link = "logit"),
##     data = d2_uncount)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.2300   0.4168   0.4168   0.4686   0.4963
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.15358    0.07500  28.716  < 2e-16 ***
## Ageyoung     0.24589    0.09037   2.721  0.00651 **
## Smokeryes   -0.12139    0.13922  -0.872  0.38323
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4147.3  on 6850  degrees of freedom
## Residual deviance: 4139.5  on 6848  degrees of freedom
## AIC: 4145.5
##
## Number of Fisher Scoring iterations: 5
```

Model outputs for E

```
##
## Call:
## glm(formula = Survival_premature ~ Age + Smoker, family = binomial(link = "logit"),
##     data = e_uncount)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

Characteristic	OR[1]	95% CI[1]	p-value
Age			
old	—	—	
young	1.28	1.07, 1.52	**0.007**
Smoker			
no	—	—	
yes	0.89	0.68, 1.17	0.4

[1] OR = Odds Ratio, CI = Confidence Interval

Figure 2: Gestation (0/1) ~ Age + Smoker (Part D2)

```
## -1.9923   0.5437   0.5437   0.6999   0.8058
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.2818     0.1715   7.474 7.81e-14 ***
## Ageyoung      0.5551     0.2208   2.514   0.0119 *
## Smokeryes    -0.3235     0.3342  -0.968   0.3330
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 559.73  on 616  degrees of freedom
## Residual deviance: 552.95  on 614  degrees of freedom
## AIC: 558.95
##
## Number of Fisher Scoring iterations: 4
```

Model outputs for F

```
##
## Call:
## glm(formula = Survival_premature_young ~ Smoker, family = binomial(link = "logit"),
##     data = f1_uncount)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9939   0.5428   0.5428   0.5428   0.6371
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
```

Characteristic	OR[1]	95% CI[1]	p-value
Age			
old	—	—	
young	1.74	1.13, 2.68	**0.012**
Smoker			
no	—	—	
yes	0.72	0.39, 1.44	0.3

[1] OR = Odds Ratio, CI = Confidence Interval

Figure 3: Premature Baby Survival (0/1) ~ Age + Smoker

```
## (Intercept)   1.8405     0.1522  12.090   <2e-16 ***
## Smokeryes    -0.3489     0.3991  -0.874    0.382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 339.06  on 413  degrees of freedom
## Residual deviance: 338.34  on 412  degrees of freedom
## AIC: 342.34
##
## Number of Fisher Scoring iterations: 4
```

Characteristic	OR[1]	95% CI[1]	p-value
Smoker			
no	—	—	
yes	0.71	0.34, 1.63	0.4

[1] OR = Odds Ratio, CI = Confidence Interval

Figure 4: Young Mothers: Premature Baby Survival (0/1) ~ Smoker

```
##
## Call:
## glm(formula = Survival_premature_old ~ Smoker, family = binomial(link = "logit"),
```

```
##     data = f2_uncount)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7452   0.7014   0.7014   0.7014   0.7876
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.2769     0.1766   7.230 4.84e-13 ***
## Smokeryes    -0.2653     0.6100  -0.435    0.664
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 214.78  on 202  degrees of freedom
## Residual deviance: 214.60  on 201  degrees of freedom
## AIC: 218.6
##
## Number of Fisher Scoring iterations: 4
```

Characteristic	OR[1]	95% CI[1]	p-value
Smoker			
no	—	—	
yes	0.77	0.25, 2.88	0.7

[1] OR = Odds Ratio, CI = Confidence Interval

Figure 5: Older Mothers: Premature Baby Survival (0/1) ~ Smoker

Mosaic plots for Part A