Project - Mothers and Infants

Elijah Castro

November 07, 2022

Summary

A cross-classification table of infants by their survival and gestational age, and their mother's age and smoking habits are here analyzed using different statistical methods to see whether or not a mother's age and smoking puts a new-born baby at risk.

Cross-classification of Infants by Their Survival and Gestational Age, and Their Mother's Age and Smoking Habits:

Mother's Age	Smoker	Premature - Died in 1st Year	Premature - Alive at 1st Year	Full Term - Died in 1st Year	Full Term - Alive at 1st Year	Total
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		104	513	45	6189	6851

Introduction

To see whether or not a mother's age and smoking puts a new-born baby at risk, our investigations were conducted with these specific objectives in mind:

a. Whether age of the mother affects their smoking habits
b. Whether smokers/older mothers were more likely to have premature babies
c. Whether smokers/older mothers were more likely to have babies who died in the 1st year
d. How age and smoking habits affected survival and gestational age
e. How age and smoking habits affect survival among premature babies
f. How smoking habits affect survival among premature babies with young/older mothers

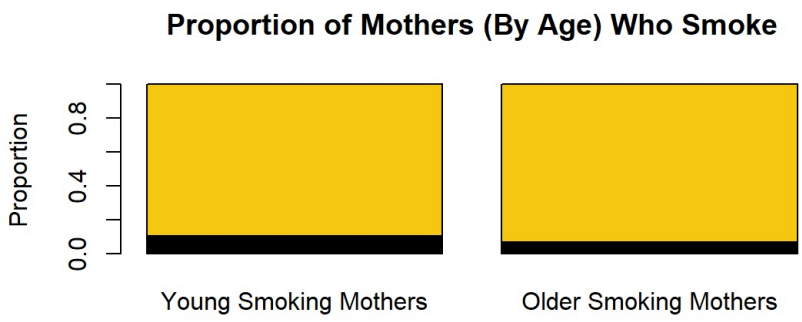
We then conducted the appropriate hypothesis test related to the objective in mind, interpreted the results, and presented our conclusions, in conjunction with supplemental statistical analyses when necessary.

Findings and Conclusions

(a) Here, we are comparing among two groups of mothers:

- Young mothers: n = 4915
- Older mothers: n = 1936

The proportion of mothers who said "yes" to being a smoker in each group is as follows, with its respective bar plot:



Note: For future objectives, bar plots will not be displayed. To view all created bar plots, see the link at the end of this report.

- Group YS - Young mothers: 514 smokers, $p_{YS} = 514/4915 = 0.1046$
- Group OS - Older mothers: 140 smokers, $p_{OS} = 140/1936 = 0.0723$

We want to know whether young mothers or older mothers are more likely to smoke. One way we can interpret this is that we want to know if the proportion of young mothers who smoke is greater than the proportion of older mothers who smoke in the two groups of mothers. Therefore, we will conduct a two-proportion z-test to compare the two observed populations. The null hypothesis is as follows:

$H_0: p_{YS} \leq p_{OS}$ vs. $H_a: p_{YS} > p_{OS}$

From this test, its respective p-value was 0.000026.

Conclusion: At the 5% significance level, we reject H_0. There is sufficient evidence to support the claim that the proportion of young mothers who smoke is greater than the proportion of older mothers who smoke. This means we can say that in terms of this population of mothers, young mothers are **more** likely to smoke. However, we cannot extend this same claim out to **every** single young mother because the samples drawn from any population vary.

(b) For the first part of this objective, we are comparing among two groups of populations:

- Mothers who smoke: n = 654
- Mothers who do not smoke: n = 6197

The number of mothers with premature babies and separated by if they smoke is as follows:

- Group SP - Mothers who smoke: 64 premature babies, $p_{SP} = 64/654 = 0.0979$
- Group NP - Mothers who do not smoke: 553 premature babies, $p_{NP} = 553/6197 = 0.0892$

We want to know if mothers who smoke are more likely to have premature babies. So, our goal is to see if the proportion of smoking mothers with premature babies is greater than the proportion of non-smoking mothers with premature babies. Therefore, we will conduct a two-proportion z-test to compare the two observed populations. The null hypothesis is as follows:

$H_0: p_{SP} \leq p_{NP}$ vs. $H_a: p_{SP} > p_{NP}$

From this test, its respective p-value was 0.2544.

Conclusion: At the 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the proportion of smoking mothers with premature babies is greater than the proportion of non-smoking mothers with premature babies. This does not mean however, that the null hypothesis is true; that is, the proportion of smoking mothers with premature babies is less than or equal to the proportion of non-smoking mothers with premature babies. We have only shown that there is not enough evidence to support the alternative hypothesis, *not* that there *is* enough evidence to support the null hypothesis. Ultimately, we cannot determine that smoking mothers are more likely to have premature babies.

For the second part of this objective, we are comparing among two groups of mothers:

- Older mothers: n = 1936
- Young mothers: n = 4915

The number of mothers who had premature babies in each group is as follows:

- Group OP - Older mothers: 203 premature babies, $p_{OP} = 203/1936 = 0.1049$
- Group YP - Young mothers: 414 premature babies, $p_{YP} = 414/4915 = 0.0842$

Similar to the first question, we want to know if the proportion of older mothers with premature babies is greater than the proportion of younger mothers with premature babies. Therefore, we will conduct a two-proportion z-test to compare the two observed populations. The null hypothesis is as follows:

$H_0: p_{OP} \leq p_{YP}$ vs. $H_a: p_{OP} > p_{YP}$

From this test, its respective p-value was 0.0042.

Conclusion: At the 5% significance level, we reject H_0. There is sufficient evidence to support the claim that the proportion of older mothers with premature babies is greater than the proportion of younger mothers with premature babies. This means we can say that in terms of this population of mothers, older mothers are **more** likely to have premature babies. However, again, we cannot extend this same claim out to **every** single older mother because the samples drawn from any population vary.

(c) For the first part, we are comparing among two groups of populations:

- Mothers who smoke: n = 654
- Mothers who do not smoke: n = 6197

The number of mothers with babies who died in the 1st year and separated by if they smoke is as follows:

- Group SD - Mothers who smoke: 20 babies dead in 1st year, $p_{SD} = 20/654 = 0.0306$
- Group ND - Mothers who do not smoke: 129 babies dead in 1st year, $p_{ND} = 129/6197 = 0.0208$

We want to know if the proportion of smoking mothers with babies dead in the 1st year is greater than the proportion of non-smoking mothers with babies dead in the 1st year. Therefore, we will conduct a two-proportion z-test to compare the two observed populations. The null hypothesis is as follows:

$H_0: p_{SD} \leq p_{ND}$ vs. $H_a: p_{SD} > p_{ND}$

From this test, its respective p-value is 0.068.

Conclusion: At the 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the proportion of smoking mothers with babies who died in the 1st year is greater than the proportion of non-smoking mothers with babies who died in the first year. Similar to our previous objectives and interpretations, this does not mean that the null hypothesis is true. Therefore, we cannot determine that smoking mothers are more likely to have babies who died in the 1st year.

For the second part, we are comparing among two groups of mothers:

- Older mothers: n = 1936
- Young mothers: n = 4915

The number of mothers who had babies who died in the 1st year in each group is as follows:

- Group OD - Older mothers: 60 babies dead in 1st year, $p_{OD} = 60/1936 = 0.0310$
- Group YD - Young mothers: 89 babies dead in 1st year, $p_{YD} = 89/4915 = 0.0181$

We want to know if the proportion of older mothers with babies dead in the 1st year is greater than the proportion of younger mothers with babies dead in the 1st year. Therefore, we will conduct a two-proportion z-test to compare the two observed populations. The null hypothesis is as follows:

$H_0: p_{OD} \leq p_{YD}$ vs. $H_a: p_{OD} > p_{YD}$

From this test, its respective p-value is 0.00069.

Conclusion: At the 5% significance level, we reject H_0. There is sufficient evidence to support the claim that the proportion of older mothers with babies who died in the 1st year is greater than the proportion of younger mothers with babies who died in the 1st year. This means we can say that in terms of this population of mothers, older mothers are **more** likely to have babies dead in the 1st year. However, again, we cannot extend this same claim out to **every** single older mother because the samples drawn from any population vary.

(d) For the first part of the objective, we create tables of the mother's age and their baby's survival conditioned on the smoking habits of the mother.

Note: For future objectives, tables will not be displayed. To view all created tables, see the link at the end of this report.

Conditional on Smoking Mothers:

Age	Died in 1st Year	Alive at 1st Year	Total
Older	5	135	140
Young	15	499	514
Total	20	634	654

Conditional on Non-Smoking Mothers:

Age	Died in 1st Year	Alive at 1st Year	Total
Older	55	1741	1796
Young	74	4327	4401
Total	129	6068	6197

Then, the next step we can do is check if each of the variables in the respective tables are independent. That is, we can conduct a chi-square test for independence. Again, conditional on the smoking habit of the mother, if we find that the variables age and baby's survival are independent, then we can say that age does not have a significant relationship with survival, and thus, probably little effect on survival of the baby. If we find that the variables are dependent on one another, then we can calculate the odds ratio and interpret how the mother's age and smoking habits affect survival. We will now conduct a chi-square test for independence, first conditional on smoking mothers. The null hypothesis is as follows:

H_0: mother's age and baby's survival during the 1st year are independent. vs. H_a: mother's age and baby's survival during the 1st year are dependent.

From this test, its respective p-value is 0.9036.

Conclusion: At the 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the mother's age and the baby's survival during the 1st year are dependent. As previously stated, this does not mean that the null hypothesis is true. Ultimately, we cannot determine how age and smoking habits affect survival **for smoking mothers**.

Now, we will conduct a chi-square test for independence conditional on non-smoking mothers. The null hypothesis is as follows:

H_0: mother's age and baby's survival during the 1st year are independent. vs. H_a: mother's age and baby's survival during the 1st year are dependent.

From this test, its respective p-value is 0.0007898.

Conclusion: At the 5% significance level, we reject H_0. There is sufficient evidence to support the claim that for non-smoking mothers, age and baby's survival during the 1st year are dependent. Now, we will calculate the odds ratio and interpret:

Odds Ratio:

The odds ratio is approximately 1.85. This means that conditional on non-smoking mothers, the odds of an older mother's baby dying in the 1st year is estimated to be about 1.85 times the odds for young mothers. Therefore, in terms of how age and smoking habits affect survival, we now know non-smoking older mothers are at a **higher** risk of losing their baby in the 1st year than non-smoking young mothers.

However, if a mother smokes, we cannot know for sure how their age affects the survival of their baby. It could be possible that there is no effect, i.e, there is no added risk of death in the 1st year based on a smoking mother's age. If we were to calculate its respective odds ratio, this would help provide further evidence of this possibility since the odds ratio is roughly close to 1 (1.23), meaning the odds of a smoking older mother's baby dying in the 1st year is similar to the odds for young mothers. But again, since our chi-square test results were not significant, we can only speculate.

For the second part, we use the created tables of the mother's age and their baby's gestational age conditioned on the smoking habits of the mother. Then, we want to check if the mother's age and their baby's gestational age are independent by conducting a chi-square test for independence. As in the first question of this objective, we will have similar interpretations depending on if the variables are independent or dependent. We will now conduct a chi-square test for independence, first conditional on smoking mothers. The null hypothesis is as follows:

H_0: mother's age and gestational age are independent. vs. H_a: mother's age and gestational age are dependent.

From this test, its respective p-value is 0.7975.

Conclusion: At the 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the mother's age and the baby's gestational age are dependent. This does not mean however, that the null hypothesis is true. Ultimately, we cannot determine how age and smoking habits affect gestational age **for smoking mothers**.

Now, we will conduct a chi-square test for independence conditional on non-smoking mothers. The null hypothesis is as follows:

H_0: mother's age and gestational age are independent. vs. H_a: mother's age and gestational age are dependent.

From this test, its respective p-value is 0.007483.

Conclusion: At the 5% significance level, we reject H_0. There is sufficient evidence to support the claim that for non-smoking mothers, age and baby's gestational age are dependent. Now, we will calculate the odds ratio and interpret:

Odds Ratio:

The odds ratio is approximately 1.29. This means that conditional on non-smoking mothers, the odds of an older mother's baby being premature is estimated to be about 1.29 times the odds for young mothers. Therefore, in terms of how age and smoking habits affect gestational age, we now know non-smoking older mothers are at a **higher** risk of having a premature baby than non-smoking young mothers.

However, if a mother smokes, we cannot know for sure how their age affects the gestational age of their baby. There is a possibility that there is no effect, i.e, there is no added risk of prematurity based on a smoking mother's age. Again, if we were to calculate its respective odds ratio, this would help provide further evidence of this possibility since the odds ratio is roughly close to 1 (1.14), meaning the odds of a smoking older mother's baby being premature are similar to the odds for young mothers. However, since our chi-square test results were not significant, we can only speculate.

(e) Here, we use our created tables of the mother's age and their premature baby's survival conditioned on the smoking habits of the mother. Then, the next step we can do is check if each of the variables in the respective tables are independent by conducting a chi-square test for independence. As in the previous objectives, we will have similar interpretations depending on if the variables age and premature baby's survival are independent or dependent. We will now conduct a chi-square test for independence, first conditional on smoking mothers. The null hypothesis is as follows:

H_0: mother's age and premature baby's survival during the 1st year are independent. vs. H_a: mother's age and premature baby's survival during the 1st year are dependent.

From this test, its respective p-value is 0.7396.

Conclusion: At the 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the mother's age and the premature baby's survival during the 1st year are dependent. Again, this does not mean that the null hypothesis is true. Therefore, we cannot determine how age and smoking habits affect survival among premature babies **for smoking mothers**.

Now, we will conduct a chi-square test for independence conditional on non-smoking mothers. The null hypothesis is as follows:

H_0: mother's age and premature baby's survival during the 1st year are independent. vs. H_a: mother's age and premature baby's survival during the 1st year are dependent.

From this test, its respective p-value is 0.0206.

Conclusion: At the 5% significance level, we reject H_0. There is sufficient evidence to support the claim that for non-smoking mothers, age and premature baby's survival during the 1st year are dependent. Now, we will calculate the odds ratio and interpret:

Odds Ratio:

The odds ratio is approximately 1.76. This means that conditional on non-smoking mothers, the odds of an older mother's premature baby dying in the 1st year is estimated to be about 1.76 times the odds for young mothers. Therefore, in terms of how age and smoking habits affect survival among premature babies, we now know non-smoking older mothers are at a **higher** risk of losing their premature baby in the 1st year than non-smoking young mothers.

Again, if a mother smokes, we cannot know for sure how their age affects the survival of their premature baby. We can do a rough check that there is no effect, i.e, there is no added risk of death in the 1st year based on a smoking mother's age using its respective odds ratio. However, this would not help provide further evidence of this possibility since the odds ratio is roughly far from 1 (1.62). As always, we can only speculate since our chi-square test results were not significant.

(f) For the first part, we use our created table of the mother's smoking habits and their premature baby's survival conditioned on young mothers. Then, the next step we can do is check if each of the variables in the table are independent by conducting a chi-square test for independence. Again, as in previous objectives, we will have similar interpretations depending on if the variables smoking habits and premature baby's survival are independent or dependent. We will now conduct a chi-square test for independence, first conditional on young mothers. The null hypothesis is as follows:

H_0: mother's smoking habit and premature baby's survival during the 1st year are independent. vs. H_a: mother's smoking habit and premature baby's survival during the 1st year are dependent.

From this test, its respective p-value is 0.5091.

Conclusion: At the 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the mother's smoking habit and the premature baby's survival during the 1st year are dependent. As always, this does not mean that the null hypothesis is true. Thus, we cannot determine how smoking habits affect survival among premature babies **for young mothers**.

Now, for the second part, we use our created table of the mother's smoking habits and their premature baby's survival conditioned on older mothers. This is the exact same process, but we will now conduct a chi-square test for independence, conditional on older mothers. The null hypothesis is as follows:

H_0: mother's smoking habit and premature baby's survival during the 1st year are independent. vs. H_a: mother's smoking habit and premature baby's survival during the 1st year are dependent.

From this test, its respective p-value is 0.9101.

Conclusion: At the 5% significance level, we fail to reject H_0. There is not sufficient evidence to support the claim that the mother's smoking habit and the premature baby's survival during the 1st year are dependent. As always, this does not mean that the null hypothesis is true. Ultimately, we cannot determine how smoking habits affect survival among premature babies **for older mothers** as well.

Odds Ratios:

In both of our chi-square tests for independence, we could not reject H_0. This means regardless of a mother's age, young or older, we cannot know for sure how smoking habits affect the survival of their premature baby. There is a possibility that there is no effect, i.e, there is no added risk of death in the 1st year based on a smoking mother, regardless of age. We will calculate their respective odds ratios and check if it is close to 1:

Interpretations:

- Conditional on young mothers, the odds of a smoking mother's premature baby dying in the 1st year is estimated to be about 1.42 times the odds for non-smoking mothers.
- Conditional on older mothers, the odds of a smoking mother's premature baby dying in the 1st year is estimated to be about 1.3 times the odds for non-smoking mothers.

This means the odds of a young or older mother who smokes and have their premature baby die in the 1st year is not similar to the odds for non-smoking young or older mothers. Unfortunately, since our chi-square test results were not significant, we can only speculate and attribute these results to chance.

---

https://github.com/ElijahMC2/PSTAT-220A---Project-1/blob/main/Project_Plots_Tables.pdf (https://github.com/ElijahMC2/PSTAT-220A---Project-1/blob/main/Project_Plots_Tables.pdf)