Thank you, Dominic, and here comes to the question B: is there a difference between the birth weight of babies for non-smoking and smoking mothers?

For this question, we made a two-side alternative hypothesis. Our null hypothesis is there is no difference between the birth weights for non-smoking and smoking mothers. And the alternative is that there is a difference.

We made a boxplot of babies’ birthweight from non-smoking and smoking mothers. 115 non-smoking mothers and 74 smoking mothers were observed. We could notice that the mean birthweight of babies from non-smoking mothers are much higher than smoking mothers.

Then, we applied T-test to test our null hypothesis.

Although we assumed that the variances are same, we are not so sure about that, so we run the t-tests in both equal variances assumed and not assumed.

And as we found that the errors are not normally distributed, we also did a non-parametric t-test to compare.

All of the results are quite similar. The p-values we got is much lower than 0.05, then we could reject the null hypothesis. Which means there is a statistical significant difference between the birth weight of babies for non-smoking and smoking mothers.

Then let us welcome Diren to introduce you ours results of question C.

Then we come to the conclusion and outlook parts. From the results of question A, this study helps to decide which babies require additional monitoring due to low birth weight. This also means hospitals or doctors could specialize their criteria for different kinds of babies. And from part B, we found that smoking during pregnancy might result in lower birth weights.

And in the future, it might be helpful to look at some further variables such as the race and size of their fathers.

This also brings us to the end of this presentation. Thank you for listening and please feel free to ask if you have any questions.

细分

Why mean cannot exist?

Tucky? It need same numbers of different samples because is compare in pair.96 26 67

Errors are not normally distributed => non parametric t-test

As data is independent, we used the Mann-Whitney-ilcoxon Test.

p-value (0.0068) is smaller than 0.05 => At 0.05 significance level, we conclude that the data of smokers and non-smokers does not come from the same population.

See t-test.R

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Sum of Squares | df | Mean Square | F | Sig. |
| Between Groups | 5015725.253 | 2 | 2507862.626 | 4.913 | .008 |
| Within Groups | 94953930.557 | 186 | 510505.003 |  |  |
| Total | 99969655.810 | 188 |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| (I) race | (J) race | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
| Lower Bound | Upper Bound |
| 1 | 2 | 383.026 | 157.964 | .055 | -6.76 | 772.82 |
| 3 | 297.435\* | 113.742 | .035 | 16.77 | 578.10 |
| 2 | 3 | -85.591 | 165.089 | .874 | -492.96 | 321.78 |

|  |  |
| --- | --- |
| Mother | N |
| Smoking | 115 |
| Non-smoking | 74 |

|  |  |  |  |
| --- | --- | --- | --- |
| Mother | | Statistic | Std. Error |
| Non-smoking | Mean | 3055.70 | 70.186 |
| Median | 3100.00 |  |
| Variance | 566491.968 |  |
| Std. Deviation | 752.657 |  |
| Range | 3969 |  |
| Smoking | Mean | 2771.92 | 76.681 |
| Median | 2775.50 |  |
| Variance | 435118.158 |  |
| Std. Deviation | 659.635 |  |
| Range | 3529 |  |

|  |  |  |
| --- | --- | --- |
|  | lwt | bwt |
| lwt | 1 | 0.186 |
| bwt | 0.186 | 1 |
| Sig (1-tailed) (p value) | 0.005 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | β (slope) | Std error | T stat | P value | 95% Confidence Interval | |
| α (constant) | 2214.41 | 299.31 | 7.40 | 0.000 | 1623.93 | 2804.90 |
| lwt | 4.18 | 1.74 | 2.40 | 0.02 | 0.74 | 7.62 |
| age | 8.09 | 10.07 | 0.80 | 0.42 | -11.76 | 27.94 |

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Estimate of slope | Standard Error | 95% Confidence Interval |
| Regression | 4.43 | 1.71 | 1.07 - 7.79 |
| Robust Regression | 5.61 | 1.82 | 2.04- 9.18 |
| Bootstrap | 4.52 | 1.46 | 2.03 - 6.8 |

|  |  |
| --- | --- |
| Group | P-value of Shapiro-Wilk test for Residuals |
| White | 0.00058 |
| Black | 0.00346 |
| Other | 0.00065 |
| Smokers | 2.686e-05 |
| Non-smokers | 2.686e-05 |

|  |  |
| --- | --- |
| Comparison between | p-Value of Levene test |
| Ethnic groups | 0.6267 |
| Smokers and non-smokers | 0.229 |

|  |  |
| --- | --- |
| Group | p-value of Shapiro-Wilk test |
| White | 0.4861 |
| Black | 0.8038 |
| Other | 0.2046 |
| Smokers | 0.4195 |
| Non-smokers | 0.3337 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | N | Mean | Std Deviation | Minimum | Maximum |
| Mother’s weight (lwt) | 189 | 129.81 lbs | 30.58 | 80 | 250 |
| Birth weight of the baby (bwt) | 189 | 2945 grams | 729.21 | 709 | 4990 |
| Age | 189 | 23.24 | 5.30 | 14 | 45 |

|  |  |
| --- | --- |
| Method | P-value |
| Independent samples T-test  (Equal variances not assumed) | 0.007 |
| Independent samples T-test  (Equal variances assumed) | 0.008 |
| Non-parametric T-test | 0.0068 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Non-smoking mothers | Smoking mothers |
| low | 0 | 74.78% | 59.46% |
| 1 | 25.22% | 40.54% |

In this part, a two-side alternative hypothesis was made.  Our null hypothesis is there is no difference between the birth weights for non-smoking and smoking mothers. And the alternative is that there is a difference between the birth weights for non-smoking and smoking mothers.

In this study, 115 non-smoking mothers and 74 smoking mothers were observed. The descriptive result was given in Table XXX and visualized by the boxplot in Figure 2. This box plot displayed the distribution of babies' birthweight by non-smoking and smoking mothers. The mean birthweight of babies from non-smoking mothers is 3055.7 grams which is much higher than 2771.92 grams, the mean birthweight of babies from smoking mothers. The range of non-smoking group is larger than the smoking group, that may be caused by the larger sample size.

The birthweight less than 2500 grams was deemed to be low. The Table XXX shows the ratio of low weight babies in all babies of smoking or non-smoking mothers. 40.54% of smoking mothers have a low weight baby while only 25.22% of non-smoking mothers have a low weight baby. Smoking mothers are more likely to have a baby whose birthweight is considered low.

Then, T-test was applied to test the null hypothesis. To exclude the impact of variances, we run the t-tests in both equal variances assumed and not assumed. From Table XXX we could got the p-values of each, 0.008 and 0.007. At the same time, as the errors are not normally distributed, we also did a non-parametric t-test as comparison. The p-value of non-parametric t-test is 0.0068. All of the results are quite similar. The p-values are much lower than 0.05. At 0.05 significance level, we conclude that the data of smokers and non-smokers does not come from the same population. Thus, we reject the null hypothesis and concluded that there is a statistical significant difference between the birth weight of babies for non-smoking and smoking mothers.

There is a statistical significant difference between the birth weight of babies for non-smoking and smoking mothers. The birthweight of babies from non-smoking mothers are higher than the birthweight of babies from smoking mothers. And smoking during pregnancy is more likely to have a baby with low birthweight (under 2500 grams).

From question A, this study helps to decide which baby requires additional monitoring due to low birth weight. As babies from different ethnic groups have different mean birthweight, the original criteria may not suit for all babies. This means hospitals or doctors could specialize their criteria for different kinds of babies. And from part B, we found that smoking during pregnancy might result in lower birth weights. Mothers are advised to pay attention to this result and try to avoid smoking during their pregnancy.

This research only focused on how could mothers' factors impact on their babies' birthweight. But there may be some other factors could also influence the birthweight. In the future, it might be helpful to look at some further variables such as the race and size of their fathers.