MAMAT JASSEH

STUDENT ID 141305227

STATISTICS FOR ANALYTICS ASSIGNMENT

BAN100

SAS CODINGS

```
proc format;
Value blackB 1= 'black'
          0= 'non_black';
Value marriedB 1= 'married'
           0= 'non_married';
value boyB     1= 'baby_boy'
           0= 'baby_girl';
value momsmokeB 1= 'smoker'
           0='nonsmoker';
run;


data birth1;
set work.birth;
label weight = 'infantb_birth'
     married = 'married_mother'
     boy = 'baby_boy'
     momsmoke = 'smoking_mother';
format black blackB.
      married marriedB.
      boy boyB.
      momsmoke momsmokeB.;
run;
```

1. **Using ttest to investigate the weight difference between Black individuals and nonblack individuals.**

SAS CODE

```
proc ttest data=birth;
class black;
```

var weight;

run;

The TTEST Procedure

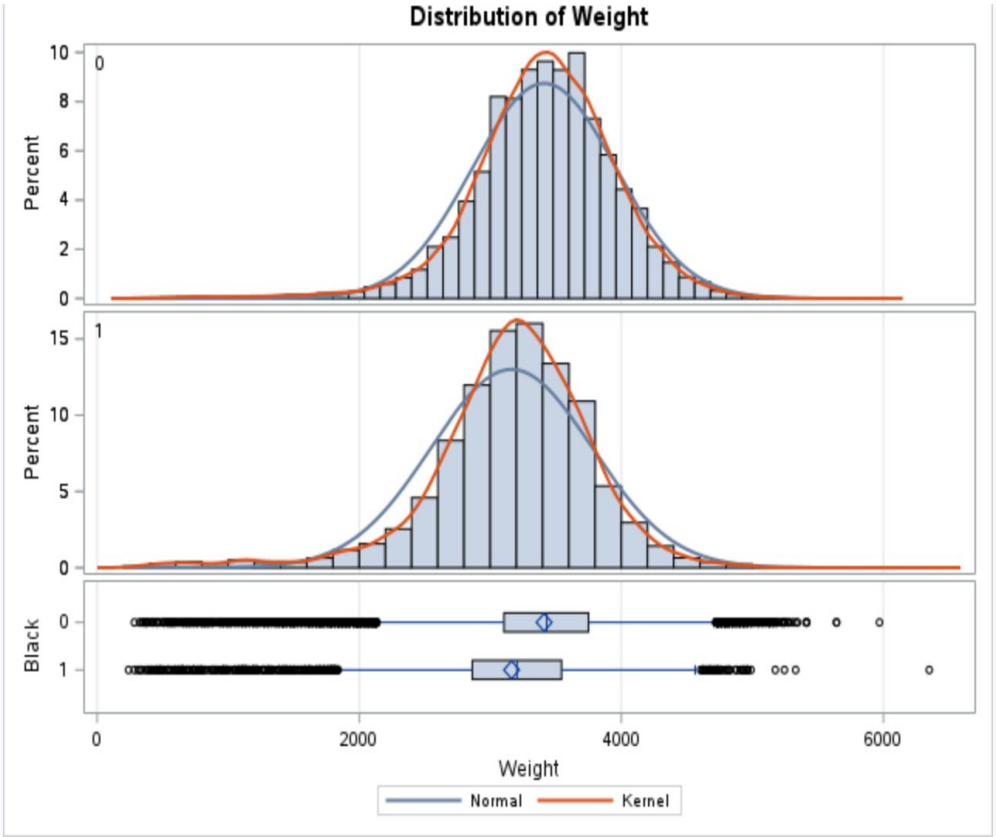
Variable: Weight (Weight)

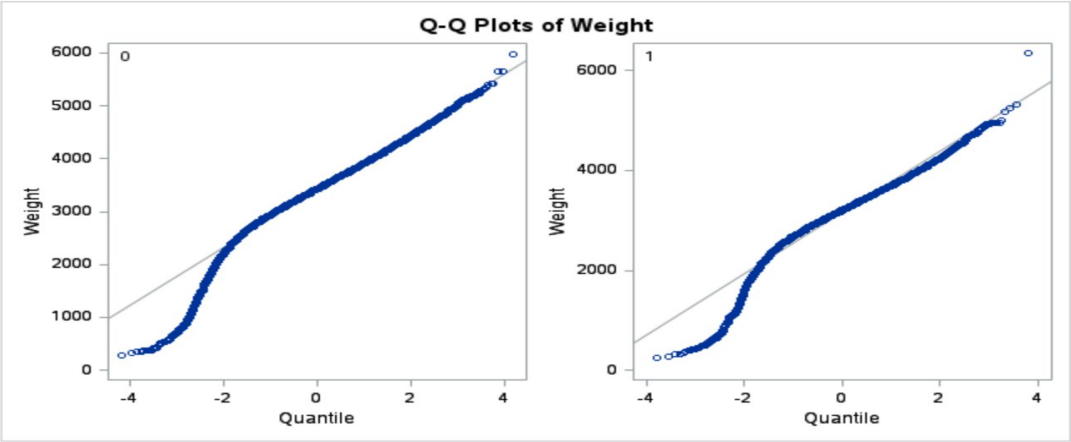
Black	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		41858	3411.2	547.6	2.6766	284.0	5970.0
1		8142	3162.7	613.7	6.8011	240.0	6350.0
Diff (1-2)	Pooled		248.6	558.9	6.7697		
Diff (1-2)	Satterthwaite		248.6		7.3088		

Black	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		3411.2	3406.0	3416.5	547.6	543.9	551.4
1		3162.7	3149.3	3176.0	613.7	604.4	623.3
Diff (1-2)	Pooled	248.6	235.3	261.8	558.9	555.5	562.4
Diff (1-2)	Satterthwaite	248.6	234.2	262.9			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	49998	36.72	<.0001
Satterthwaite	Unequal	10808	34.01	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	8141	41857	1.26	<.0001





From the statistical analysis provided, we gathered some key information regarding the "Weight" variable, specifically comparing the two groups, Blacks and non-blacks.

On average, black babies weigh about 3,162.7, while non-black babies weigh on average about 3,411.2. This means that on average, non-black babies weigh about 248.6 more than black babies.
Both groups' weight distributions are not equal. The weight standard deviation for Black babies is roughly 613.7, while it is roughly 547.6 for Non-Black babies. This suggests that Black babies experience greater weight variance.

We have 95% confidence level that the true mean weight for Blacks is between 3,149.3 and 3,176.0 and between 3,406.0 and 3,416.5 for non-Blacks. The range from 235.3 to 261.8 is the 95% confidence interval for the mean difference between the two groups.

According to the analysis, there are no equal weight variances between the two groups. Both the Pooled and Satterthwaite variance comparison approaches show a substantial difference. Additionally, the Q-Q plots can be examined to visualize the distribution of weights in the two groups and evaluate their normality.

The study shows that there is a statistically significant weight difference between Black and Non-Black babies, with Non-Black babies weighing more on average. However, it's crucial to remember that weight varies significantly within each group, and there is a significant difference in variances between the groups. We conclude that we reject the null hypothesis Ho because the proc ttest reveals that the P-value <.00001 is < our selected alpha (α) level which is 0.005.

2. **Using ttest to investigate the weight difference between married individuals and nonmarried individuals.**

SAS CODE

```
proc ttest data=birth;

class married;

var weight;

run;
```

The TTEST Procedure

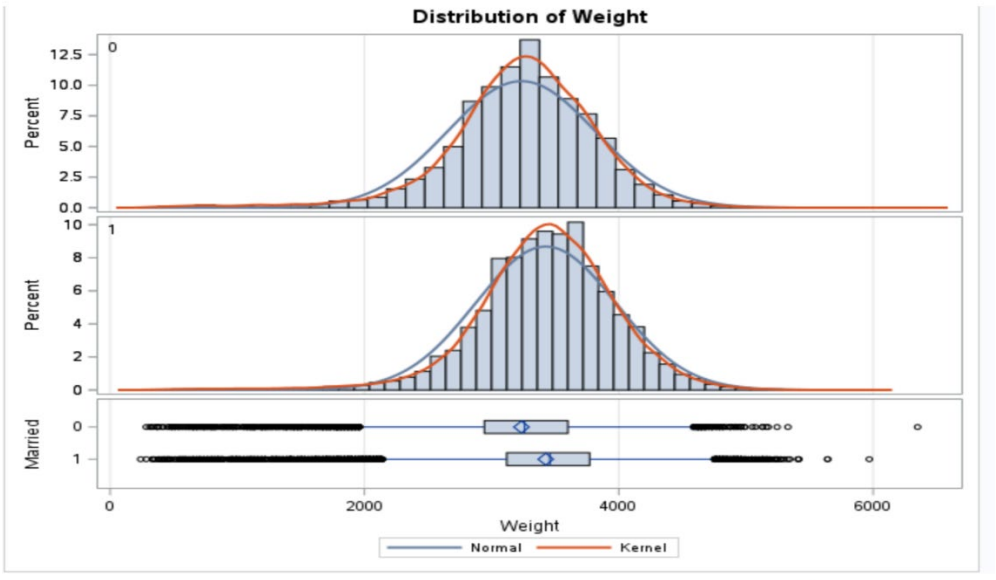
Variable: Weight (Weight)

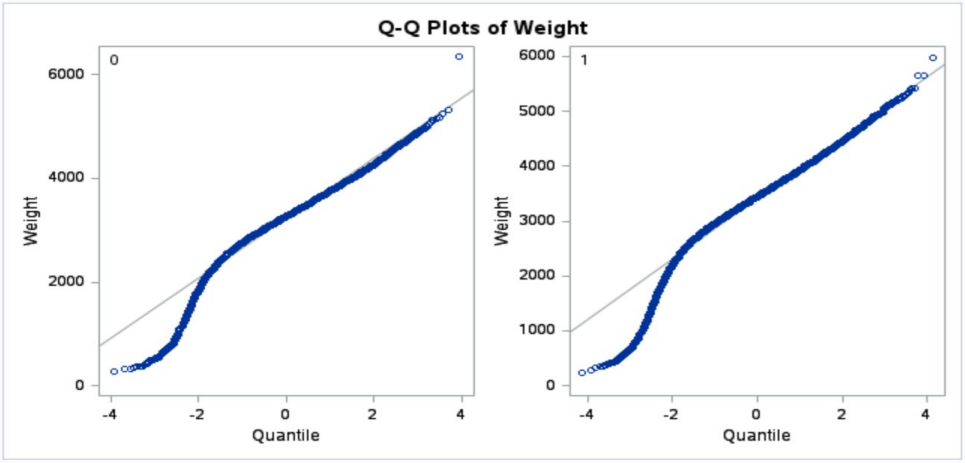
Married	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		14369	3234.4	579.0	4.8302	284.0	6350.0
1		35631	3425.7	551.8	2.9231	240.0	5970.0
Diff (1-2)	Pooled		-191.3	559.7	5.5315		
Diff (1-2)	Satterthwaite		-191.3		5.6459		

Married	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		3234.4	3225.0	3243.9	579.0	572.4	585.8
1		3425.7	3420.0	3431.5	551.8	547.8	555.9
Diff (1-2)	Pooled	-191.3	-202.1	-180.5	559.7	556.3	563.2
Diff (1-2)	Satterthwaite	-191.3	-202.4	-180.2			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	49998	-34.58	<.0001
Satterthwaite	Unequal	25443	-33.88	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	14368	35630	1.10	<.0001





In our investigation, we looked at the connection between marital status and weight. Individuals who are married and those who are not married are separated from the data into two groups.

The t-test results show that there is a statistically significant difference in average weight between married and unmarried mothers. Married mothers have babies with a higher average weight (3425.7) than unmarried mothers, who have an average weight of 3234.4. meaning the average weight of married is 191.3 higher than that of unmarried.

The investigation also indicated that the difference in weight between the two groups was statistically significant. The p-value of the Pooled and Satterthwaite methods is less than 0.0001. The low p-value for the equality of variances test implies that the variances in the two groups are significantly different.

Since our alpha value is 0.005 and the p-value is much smaller than the alpha value, we will reject the null hypothesis. When testing a hypothesis, a small p-value less than alpha indicates strong evidence against the null hypothesis, meaning there is a statistically significant difference between the two groups being compared. Therefore, based on the provided alpha value of 0.005, we would reject the null hypothesis in favor of the alternative hypothesis, which suggests that there is a significant difference in the mean between the two groups.

3. **Using ttest to investigate the weight difference between baby boy and baby girl.**

SAS CODE

```
proc ttest data=birth;

class boy;

var weight;

run;
```

The TTEST Procedure

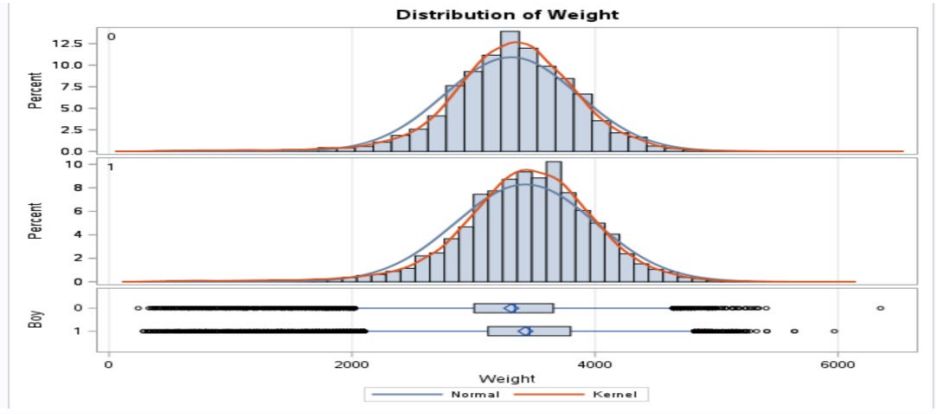
Variable: Weight (Weight)

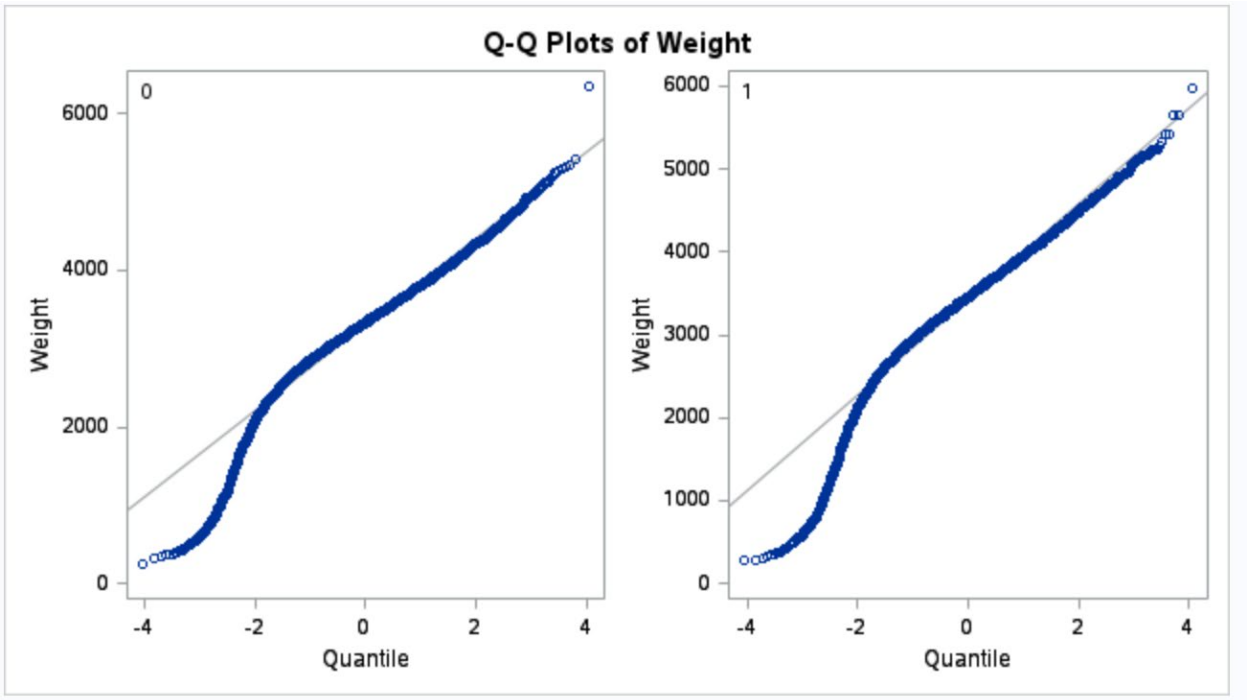
Boy	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		24208	3310.6	547.7	3.5204	240.0	6350.0
1		25792	3427.3	577.7	3.5970	284.0	5970.0
Diff (1-2)	Pooled		-116.7	563.4	5.0416		
Diff (1-2)	Satterthwaite		-116.7		5.0331		

Boy	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		3310.6	3303.7	3317.5	547.7	542.9	552.7
1		3427.3	3420.2	3434.3	577.7	572.7	582.7
Diff (1-2)	Pooled	-116.7	-126.6	-106.8	563.4	559.9	566.9
Diff (1-2)	Satterthwaite	-116.7	-126.6	-106.8			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	49998	-23.15	<.0001
Satterthwaite	Unequal	49993	-23.18	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	25791	24207	1.11	<.0001





The t-test's findings are shown for the variable "Weight". Labeled boys and girls are contrasted in the data set. Based on the sample, girls' average weight is 3,310.6, with a standard deviation of 547.7 boys' average weight is 3427.3, while their standard deviation is 577.7, which is slightly greater. Boys are heavier on average than girls, according to the observed difference in mean weight between the two groups, which is -116.7.

The 95% confidence interval for this difference in means ranges from -126.6 to -106.8. Since this range does not include zero, it implies statistically significant differences between groups. This interpretation is further supported by the t values, which are highly negative for the equal and unequal variance assumptions (−23.15 and −23.18, respectively). The corresponding p-value is less than 0.0001, indicating a very significant difference between the two groups. Additionally, a separate test for equality of variances yielded a p-value less than 0.0001, indicating a significant difference in variance between the two methods.

The p-values for the difference in mean values between the two groups were all less than 0.0001. This value is much lower than the defined alpha level of 0.005. Based on the given alpha level of 0.005, we reject the null hypothesis. This means there is a statistically significant difference in weight between the two groups.

4. **Using ttest to investigate the weight difference between momsmoke and non-smoking individuals.**

SAS CODE

```
proc ttest data=birth;
class momsmoke;
var weight;
run;
```

The TTEST Procedure

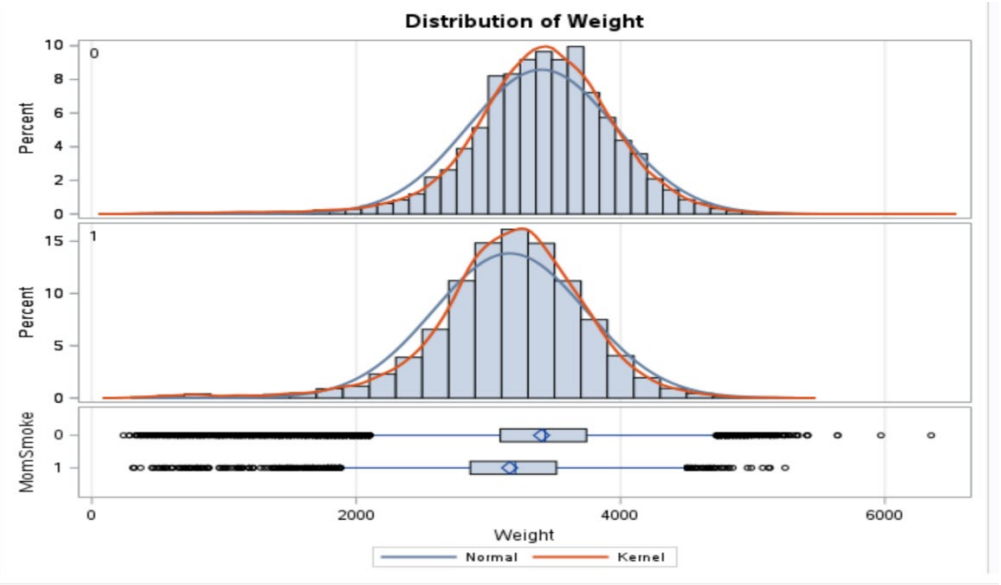
Variable: Weight (Weight)

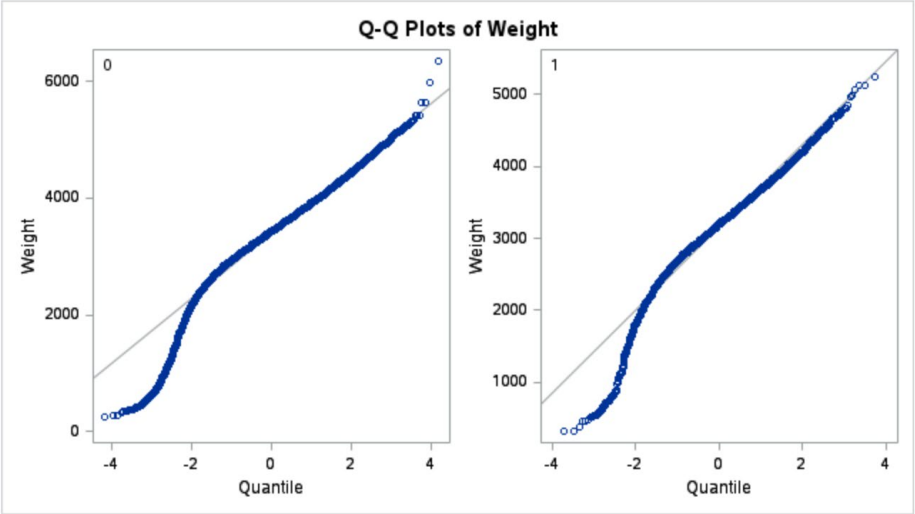
MomSmoke	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		43467	3402.3	558.0	2.6766	240.0	6350.0
1		6533	3160.9	576.8	7.1358	312.0	5245.0
Diff (1-2)	Pooled		241.5	560.5	7.4376		
Diff (1-2)	Satterthwaite		241.5		7.6213		

MomSmoke	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		3402.3	3397.1	3407.6	558.0	554.3	561.8
1		3160.9	3146.9	3174.8	576.8	567.0	586.8
Diff (1-2)	Pooled	241.5	226.9	256.0	560.5	557.1	564.0
Diff (1-2)	Satterthwaite	241.5	226.5	256.4			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	49998	32.46	<.0001
Satterthwaite	Unequal	8474.1	31.68	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	6532	43466	1.07	0.0004





The main objective of this study is to examine differences in weight between mothers who smoke and those who do not smoke. The sample size for the proportion of mothers who did not smoke is significantly higher 43,467 than that of mothers who smoked 6,533. This imbalance can affect the results of the t-test. On average, non-smoking mothers had babies with a higher average weight 3,402.3 than smoking mothers 3,160.9. The difference in means is 241.5.

The t-test findings show that the weights of the two groups differ statistically significantly. The p-values for the methods Pooled and Satterthwaite are much below the 0.005 threshold.

It's important to note that although both procedures have the same mean difference, their computed t-values differ somewhat. This is because of how each approach handles variance and sample size.

An important assumption for the t-test is equality of variances. Since the p-value of the fold F test is less than 0.005, this assumption is likely incorrect. Pooled t-test results may no longer be valid due to this violation. The Satterthwaite method should be used instead because the assumption of equal variances is broken.

Given that both p-values obtained are below the alpha threshold of 0.005, there is a statistically significant difference in weight between the two groups of mothers who smoke and those who don't. The null hypothesis, according to which there is no weight difference between the two groups, is rejected.