DATASET VARIABLES:

Variable Label	*Description*
plurality	Number of children born of the pregnancy
sex	Sex of child (1=Male, 2=Female)
mage	Age of mother (years)
weeks	Completed Weeks of Gestation (weeks)
marital	Marital status (1=married, 2=not married)
racemom	Race of Mother (0=Other Non-white, 1=White, 2=Black 3=American Indian, 4=Chinese, 5=Japanese, 6=Hawaiian, 7=Filipino, 8=Other Asian or Pacific Islander)
hispmom	Mother of Hispanic origin (C=Cuban, M=Mexican, N=Non-Hispanic, O=Other and Unknown Hispanic, P=Puerto Rican, S=Central/South American, U=Not Classifiable)
gained	Weight gained during pregnancy (pounds)
smoke	0=mother did not smoke during pregnancy 1=mother did smoke during pregnancy
drink	0=mother did not consume alcohol during pregnancy 1=mother did consume alcohol during pregnancy
tounces	Weight of child (ounces)
tgrams	Weight of child (grams)
low	0=infant was not low birth weight 1=infant was low birth weight
Premie	0=infant was not premature 1=infant was premature premature defined at 36 weeks or sooner

QUESTION 1:

1. Please begin the assignment by deleting the items with missing values ("#NULL!") in it and do analysis only on the remained data set (hint: use the code provided in class). Please provide a frequency table for the percentage of low birth weights and a frequency table for the percentage of smokers. Create a summary table (mean, median, standard deviation, minimum, maximum) for the continuous variables of mage, weeks, and tounces.

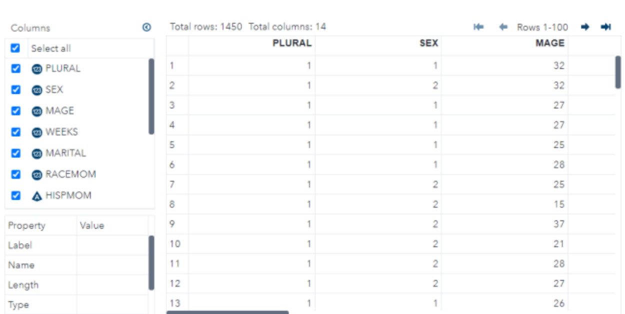


Figure 1. Initial output of loading data.

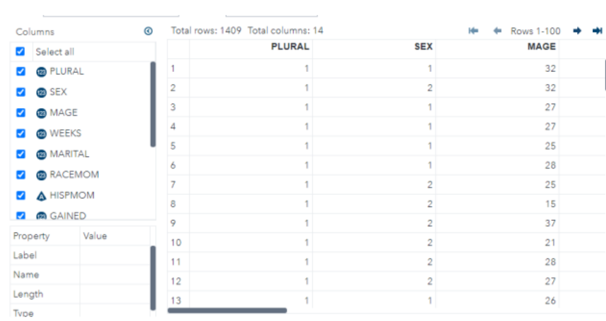


Figure 2. Output data after deleting missing values

The FREQ Procedure

LOW	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1290	91.55	1290	91.55
1	119	8.45	1409	100.00

SMOKE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1203	85.38	1203	85.38
1	206	14.62	1409	100.00

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
MAGE	1409	26.7927608	6.0891510	13.0000000	43.0000000
WEEKS	1409	38.6493967	2.6541672	22.0000000	45.0000000
TOUNCES	1409	116.4407381	22.1255345	12.0000000	181.0000000

QUESTION 2:

[A.] Determine if there is sufficient evidence to conclude the mean age of mothers giving birth in North Carolina is over 25 years of age at the 0.05 level of significance.

NULL & ALTERNATIVE HYPOTHESIS:

$$H_0 : \mu \leq 25$$
$$H_a : \mu > 25$$

The ">" demonstrates that it is a right-tailed test, for a single population mean.

ASSUMPTIONS:

Random Variable: the mean age of mothers giving birth in North Carolina.

The TTEST Procedure

Variable: MAGE

N	Mean	Std Dev	Std Err	Minimum	Maximum
1409	26.7928	6.0892	0.1622	13.0000	43.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
26.7928	26.5258	Infty	6.0892	5.8723	6.3227

DF	t Value	Pr > t
1408	11.05	<.0001

TEST:

Using the P-value approach: p-value = 0.0001. The p-value is the area to the right of the sample mean. Test statistics: The critical value for a right-tailed test is t_c = 1.645. The rejection region is R = {t : t >1.645}

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{26.7928 - 25}{6.0892/\sqrt{1409}} = 11.052$$

Since t =11.05 is greater than t_c = 1.645, it can be concluded that the null hypothesis is rejected.

INTERPRETATION OF THE P-VALUE:

If H_0 is true, there is a 0.0001 probability that the mean age of mothers in North Carolina is 25 years. Because 0.0001 chance is small, the mean age of 25 years is unlikely to have happened randomly, therefore it is a rare event.

Compare α and the p-value: α = 0.05; p-value = 0.0001; α > p-value

DECISION:

Since α > p-value, reject H_0 .
This means that the hypothesis that $\mu = 25$ is rejected.

CONCLUSION:

At the 5% significance level, we conclude that the mean age of mothers in North Carolina is greater than 25 years. The sample data illustrates that there is sufficient evidence that the mean age of mothers in North Carolina is greater than 25 years.

[B.] Determine if there is sufficient evidence to conclude the mean weeks of gestation of mothers giving birth in North Carolina is below 39 weeks.

NULL & ALTERNATIVE HYPOTHESIS

$H_0 : \mu \geq 39 \quad H_a : \mu < 39$

The "<" demonstrates that it is a left-tailed test, for a single population mean.

ASSUMPTIONS

Random Variable: the mean weeks of gestation of mothers giving birth in North Carolina.

The TTEST Procedure

Variable: WEEKS

N	Mean	Std Dev	Std Err	Minimum	Maximum
1409	38.6494	2.6542	0.0707	22.0000	45.0000

Mean	5% CL Mean		Std Dev	5% CL Std Dev	
38.6494	-Infty	38.5330	2.6542	2.6517	2.6579

DF	t Value	Pr < t
1408	-4.96	<.0001

TEST

Using the P-value approach: p-value = 0.0001. The p-value is the area to the right of the sample mean.
Test statistics:
The critical value for a left-tailed test is t_c = 1.645. The rejection region is R = {t : t < -1.645

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{38.6494 - 39}{2.6542/\sqrt{1409}} = -4.9583$$

Obs	tc
1	-1.64594

Since t =-4.95 is less than t_c = -1.645, it can be concluded that the null hypothesis is rejected.

INTERPRETATION OF THE P-VALUE:

If H_0 is true, there is a 0.0001 probability that the mean weeks of gestation in North Carolina greater than or equal to 39 weeks.
Compare α and the p-value: α = 0.95; p-value = 0.0001; α > p-value

DECISION:

Since α > p-value, reject H_0 .
This means that the hypothesis that $\mu \geq 39$ is rejected.

CONCLUSION:

At the 95% confidence interval level, we conclude that the mean weeks of gestation of mothers in North Carolina is less than 39 weeks. The sample data suggest that there is sufficient evidence that the mean weeks of gestation of mothers is less than 39 weeks.

[C]. Determine if there is sufficient evidence to conclude that the mean weight of babies born to mothers in North Carolina is above 7 lbs. (Note that there are 16 ounces in a pound.)

NULL & ALTERNATIVE HYPOTHESIS:

$H_0 : \mu \leq 112 \quad H_a : \mu > 112$

The ">" demonstrates that it is a right-tailed test, for a single population mean.

The TTEST Procedure

Variable: TOUNCES

N	Mean	Std Dev	Std Err	Minimum	Maximum
1409	116.4	22.1255	0.5894	12.0000	181.0

Mean	5% CL Mean		Std Dev	5% CL Std Dev	
116.4	117.4	Infty	22.1255	22.1046	22.1570

DF	t Value	Pr > t
1408	7.53	<.0001

ASSUMPTIONS:

Random Variable: the mean weight of babies born to mothers in North Carolina.

TEST:

Using the P-value approach: p-value = 0.0001. The p-value is the area to the right of the sample mean. Test statistics: The critical value for a right-tailed test is t_c = 1.645. The rejection region is R = {t : t >1.645}

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{116.4 - 112}{22.1255/\sqrt{1409}} = 7.465$$

Obs	tc
1	1.64591

Since t =7.465 is greater than t_c = 1.645, it can be concluded that the null hypothesis is rejected.

INTERPRETATION OF THE P-VALUE:

If H_0 is true, there is a 0.0001 probability that the mean weight of babies born to mothers in North Carolina is less than or equal to 7 pounds / 112 ounces. Because 0.0001 chance is small, the mean weight of 7 pounds / 112 ounces is unlikely to have happened randomly, therefore it is a rare event.
Compare α and the p-value: α = 0.05; p-value = 0.0001; α > p-value

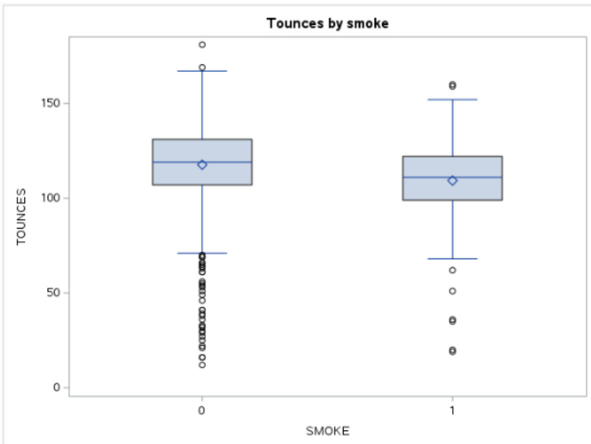
DECISION:

Since α > p-value, reject H_0 .
This means that the hypothesis that $\mu \leq 112$ is rejected.

CONCLUSION:

At the 95% confidence interval level, we conclude that mean weight of babies born to mothers in North Carolina is greater than 7 pounds / 112 ounces. The sample data illustrates that there is sufficient evidence that the mean weight of babies born to mothers in North Carolina is greater than 7 pounds / 112 ounces.

[D]. Construct a side-by-side boxplot for tounces for smokers and non-smokers. Comment on whether you believe you will reject or fail to reject the null hypothesis. Determine if there is sufficient evidence to conclude the mean tounces of smoking mothers is lower than the mean birth weight for non-smoking mothers.



COMMENT: Observing the box plot, it is seen clearly that the mean for when smoke value is 0, which implies that the mothers did not smoke during pregnancy have a mean weight that is higher than those who smoke during pregnancy. It is also seen that the maximum value of the weight of babies from non-smokers mothers is well above 150 in contrast to smoker mothers that have a maximum value around that 150. Therefore, I believe that we will reject the null hypothesis that mean ounces of smoking mothers is greater than or equal to the mean birth weight for non-smoking mothers.

NULL & ALTERNATIVE HYPOTHESIS:

$$H_0: \mu_{smoking} \geq \mu_{non-smoking}$$
$$H_a: \mu_{smoking} < \mu_{non-smoking}$$

The two means demonstrates that it is a two-tailed test, for population mean.

ASSUMPTIONS:

$\mu_{smoking}$ represents the population mean birth weight for smoking mothers.
$\mu_{non-smoking}$ represents the population mean birth weight for non-smoking mothers.
Random Variable: the mean weight of babies born to smoking and non-smoking mothers in North Carolina

Variable: TOUNCES

SMOKE	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		1203	117.7	21.9700	0.6334	12.0000	181.0
1		206	109.3	21.7231	1.5135	19.0000	160.0
Diff (1-2)	Pooled		8.3794	21.9342	1.6539		
Diff (1-2)	Satterthwaite		8.3794		1.6407		

SMOKE	Method	Mean	5% CL Mean		Std Dev	5% CL Std Dev	
0		117.7	117.6	117.7	21.9700	21.9480	22.0043
1		109.3	109.2	109.4	21.7231	21.6912	21.8261
Diff (1-2)	Pooled	8.3794	8.2757	8.4832	21.9342	21.9135	21.9654
Diff (1-2)	Satterthwaite	8.3794	8.2765	8.4824			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	1407	5.07	<.0001
Satterthwaite	Unequal	281.63	5.11	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	1202	205	1.02	0.8528

TEST:
Using the P-value approach: p-value <0.0001. The p-value is the area to both sides of the sample mean.
Test statistics:
The critical value for a two-tailed test is t_c = 1.96. The rejection region is R = {t : t >1.96}

$$t\ value = 5.07$$

Since t =5.07 is greater than t_c = 1.96, it can be concluded that the null hypothesis is rejected.

INTERPRETATION OF THE P-VALUE:
If H_0 is true, there is a 0.0001 probability that the mean weight of babies born to smoking mothers in North Carolina is greater than or equal to the mean weight of babies born to non-smoking mothers. Because 0.0001 chance is small, the equality of the two means is unlikely to have happened randomly, therefore it is a rare event.
Compare α and the p-value: α = 0.05; p-value = 0.0001; α > p-value

DECISION:

Since α > p-value, reject H_0 .
This means that the hypothesis that $\mu_{smoking} \geq \mu_{non-smoking}$ is rejected.

CONCLUSION:

At the 5% significance level, we conclude that the mean weight of babies born to smoking mothers in North Carolina is lesser than the mean weight of babies born to non-smoking mothers. The sample data illustrates that there is sufficient evidence that the mean weight of babies born to smoking mothers in North Carolina is greater than or equal to the mean weight of babies born to non-smoking mothers.

[E]. Determine if there is sufficient evidence to conclude the percentage of low birth weight children in North Carolina is above 6%.

NULL & ALTERNATIVE HYPOTHESIS

$H_0 : p \leq 0.06 \quad H_a : p > 0.06$
The p demonstrates that it is a z-test for proportion.

ASSUMPTIONS:

Random Variable: p represents the population proportion of low birth weight of babies born to mothers in North Carolina.
Assume a significant level of 0.025.

TEST:

Using the P-value approach: p-value = 0.0001.
Test statistics:
The critical value for a right-tailed test is z_c = 1.96. The rejection region is R = {z : z >1.96}

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.0845 - 0.06}{\sqrt{\frac{0.06(1-0.06)}{1409}}} = 3.872$$

Since z =3.872 is greater than z_c = 1.96 @ 0.025, it can be concluded that the null hypothesis is rejected.

Binomial Proportion	
LOW = 1	
Proportion	0.0845
ASE	0.0074
95% Lower Conf Limit	0.0699
95% Upper Conf Limit	0.0990
Exact Conf Limits	
95% Lower Conf Limit	0.0705
95% Upper Conf Limit	0.1002

Test of H0: Proportion = 0.06	
ASE under H0	0.0063
Z	3.8656
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	0.0001

Sample Size = 1409

The FREQ Procedure

LOW	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1290	91.55	1290	91.55
1	119	8.45	1409	100.00

Binomial Proportion	
LOW = 0	
Proportion	0.9155
ASE	0.0074
95% Lower Conf Limit	0.9010
95% Upper Conf Limit	0.9301
Exact Conf Limits	
95% Lower Conf Limit	0.8998
95% Upper Conf Limit	0.9295

Test of H0: Proportion = 0.06	
ASE under H0	0.0063
Z	135.2252
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	<.0001

Sample Size = 1409

Obs	tc
1	1.64591

INTERPRETATION OF THE P-VALUE:

If H_0 is true, there is a 0.0001 probability that the percentage of low birth weight children in North Carolina is lesser than or equal to 6%.
Compare α and the p-value: α = 0.025; p-value = 0.0001; α > p-value

DECISION:

Since α > p-value, reject H_0 .
This means that the hypothesis that $p \leq 0.06$ is rejected.

CONCLUSION:

At the 0.025 level of significance level, we conclude that the percentage of low birth weight children in North Carolina is above 6%. The sample proportion illustrates that there is sufficient evidence that the percentage of low birth weight children in North Carolina is above 6%.

[F]. Determine if there is sufficient evidence to conclude the percentage of mothers who smoke in North Carolina is above 10%.

NULL & ALTERNATIVE HYPOTHESIS:

$H_0 : p \leq 0.10 \quad H_a : p > 0.10$
The p demonstrates that it is a z-test for proportion.

ASSUMPTIONS:

Random Variable: p represents the population proportion of mothers who smoke in North Carolina.

TEST:

Using the P-value approach: p-value = 0.0001.
Test statistics:
The critical value for a right-tailed test is z_c = 1.96. The rejection region is R = {z: z >1.96}

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.1462 - 0.10}{\sqrt{\frac{0.10(1-0.10)}{1409}}} = 5.7806$$

Since z =5.7806 is greater than z_c = 1.96, it can be concluded that the null hypothesis is rejected.

The FREQ Procedure

SMOKE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	1203	85.38	1203	85.38
1	206	14.62	1409	100.00

Binomial Proportion SMOKE = 1	
Proportion	0.1462
ASE	0.0094
95% Lower Conf Limit	0.1278
95% Upper Conf Limit	0.1647
Exact Conf Limits	
95% Lower Conf Limit	0.1282
95% Upper Conf Limit	0.1657

Test of H0: Proportion = 0.1	
ASE under H0	0.0080
Z	5.7810
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	<.0001

Sample Size = 1409

Binomial Proportion SMOKE = 0	
Proportion	0.8538
ASE	0.0094
95% Lower Conf Limit	0.8353
95% Upper Conf Limit	0.8722
Exact Conf Limits	
95% Lower Conf Limit	0.8343
95% Upper Conf Limit	0.8718

Test of H0: Proportion = 0.1	
ASE under H0	0.0080
Z	94.3167
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	<.0001

Sample Size = 1409

INTERPRETATION OF THE P-VALUE:

If H_0 is true, there is a 0.0001 probability that the percentage of mothers who smoke in North Carolina is lesser than or equal to 10%.
Compare α and the p-value: α = 0.05; p-value = 0.0001; α > p-value

DECISION:

Since $\alpha >$ p-value, reject H_0 .
This means that the hypothesis that $p \leq 0.10$ is rejected.

CONCLUSION:

We conclude that the percentage of mothers who smoke in North Carolina is above 10%. The sample proportion illustrates that there is sufficient evidence that the percentage of mothers who smoke in North Carolina is above 10%.

[G]. Determine if there is sufficient evidence to conclude the percentage of low birth weight for smoking mothers is different than the percentage of low birth weight for non-smoking mothers.

Risk Difference Test	
H0: P1 - P2 = 0	Wald Method
Risk Difference	0.0489
ASE (H0)	0.0210
Z	2.3325
One-sided Pr > Z	0.0098
Two-sided Pr > |Z|	0.0197
Column 1 (LOW = 0)	

Risk Difference Test	
H0: P1 - P2 = 0	Wald Method
Risk Difference	-0.0489
ASE (H0)	0.0210
Z	-2.3325
One-sided Pr < Z	0.0098
Two-sided Pr > |Z|	0.0197
Column 2 (LOW = 1)	

NULL & ALTERNATIVE HYPOTHESIS:

$$H_0 : p_{smoking} = p_{non-smoking}$$
$$H_a : p_{smoking} \neq p_{non-smoking}$$

The p demonstrates that it is a z-test for proportion.

ASSUMPTIONS:

Statistics for Table of SMOKE by LOW

Column 1 Risk Estimates

	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.9227	0.0077	0.9076	0.9378	0.9061	0.9372
Row 2	0.8738	0.0231	0.8284	0.9191	0.8206	0.9159
Total	0.9155	0.0074	0.9010	0.9301	0.8998	0.9295
Difference	0.0489	0.0244	0.0011	0.0967		

Difference is (Row 1 - Row 2)

Column 2 Risk Estimates

	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.0773	0.0077	0.0622	0.0924	0.0628	0.0939
Row 2	0.1262	0.0231	0.0809	0.1716	0.0841	0.1794
Total	0.0845	0.0074	0.0699	0.0990	0.0705	0.1002
Difference	-0.0489	0.0244	-0.0967	-0.0011		

Difference is (Row 1 - Row 2)

Random Variable:
$p_{smoking}$ represents the population proportion of low birth weight for smoking mothers.
$p_{non-smoking}$ represents the population proportion of low birth weight for non-smoking mothers.
If the values do not include zero as a likely value of the population proportion difference, the difference would be significant at 0.05 level.

TEST:

Using the P-value approach:
From the Risk Difference test Table, we get these values.
p-value = 0.0197; z-test statistics: -2.3325.

INTERPRETATION OF THE P-VALUE:

If H_0 is true, there is a 0.0197 probability that the percentage of low birth weight for smoking mothers is equal to the percentage of low birth weight for non-smoking mothers.
Compare α and the p-value: α = 0.05; p-value = 0.0197; $\alpha >$ p-value

DECISION:
Since $\alpha >$ p-value, reject H_0 .
This means that the hypothesis that $p_{smoking} = p_{non-smoking}$ is rejected.

CONCLUSION:

At the 5% significance level, we conclude that the percentage of low birth weight for smoking mothers is different than the percentage of low birth weight for non-smoking mothers. The sample proportion illustrates that there is sufficient evidence percentage of low birth weight for smoking mothers is different than the percentage of low birth weight for non-smoking mothers.

[H]. Determine if there is sufficient evidence to conclude the percentage of low birth weight for smoking mothers is lower than the percentage of low birth weight for non-smoking mothers.

NULL & ALTERNATIVE HYPOTHESIS:

$$H_0 : p_{smoking} \geq p_{non-smoking}$$
$$H_a : p_{smoking} < p_{non-smoking}$$

The p demonstrates that it is a z-test for proportion.

Risk Difference Test	
H0: P1 - P2 = 0	Wald Method
Risk Difference	0.0489
ASE (H0)	0.0210
Z	2.3325
One-sided Pr > Z	0.0098
Two-sided Pr > |Z|	0.0197
Column 1 (LOW = 0)	

Statistics for Table of SMOKE by LOW

Column 1 Risk Estimates

	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.9227	0.0077	0.9076	0.9378	0.9061	0.9372
Row 2	0.8738	0.0231	0.8284	0.9191	0.8206	0.9159
Total	0.9155	0.0074	0.9010	0.9301	0.8998	0.9295
Difference	0.0489	0.0244	0.0011	0.0967		

Difference is (Row 1 - Row 2)

Column 2 Risk Estimates

	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.0773	0.0077	0.0622	0.0924	0.0628	0.0939
Row 2	0.1262	0.0231	0.0809	0.1716	0.0841	0.1794
Total	0.0845	0.0074	0.0699	0.0990	0.0705	0.1002
Difference	-0.0489	0.0244	-0.0967	-0.0011		

Statistics for Table of SMOKE by LOW

Statistic	DF	Value	Prob
Chi-Square	1	5.4406	0.0197
Likelihood Ratio Chi-Square	1	4.9015	0.0268
Continuity Adj. Chi-Square	1	4.8265	0.0280
Mantel-Haenszel Chi-Square	1	5.4367	0.0197
Phi Coefficient		0.0621	
Contingency Coefficient		0.0620	
Cramer's V		0.0621	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	1110
Left-sided Pr <= F	0.9909
Right-sided Pr >= F	0.0172
Table Probability (P)	0.0081
Two-sided Pr <= P	0.0289

ASSUMPTIONS:

Random Variable:
$p_{smoking}$ represents the population proportion of low birth weight for smoking mothers.
$p_{non-smoking}$ represents the population proportion of low birth weight for non-smoking mothers.
Assume significance level of 0.025.

TEST:

Using the P-value approach:
From the Risk Difference test Table, we get these values. p-value = 0.0098; z-test statistics: 2.3325
The critical value for a left-tailed test is Z_C= 1.96.

INTERPRETATION OF THE P-VALUE:
If H_0 is true, there is a 0.0098 probability that the percentage of low birth weight for smoking mothers is greater than and equal to the percentage of low birth weight for non-smoking mothers.
Compare α and the p-value: α = 0.05; p-value = 0.0098; α > p-value

DECISION:

Since obtained Z > Z_c, reject H_0 .
This means that the hypothesis $p_{smoking} \geq p_{non-smoking}$ is rejected.

CONCLUSION:

At the 0.025 significance level, we conclude that the percentage of low birth weight for smoking mothers is less than the percentage of low birth weight for non-smoking mothers. The sample proportion illustrates that there is sufficient evidence percentage of low birth weight for smoking mothers is less than the percentage of low birth weight for non-smoking mothers.

[I]. Determine if there is sufficient evidence to conclude the percentage of low birth weight for mothers who did not consume alcohol during pregnancy lower than the percentage of low birth weight for mothers who did consume alcohol during pregnancy.

NULL & ALTERNATIVE HYPOTHESIS:

$$H_0 : p_{non-alcohol} \geq p_{alcohol}$$
$$H_a : p_{non-alcohol} < p_{alcohol}$$

The p demonstrates that it is a z-test for proportion.

ASSUMPTIONS:

Random Variable:
$p_{alcohol}$ represents the population proportion of low birth weight for mothers who did consume alcohol.
$p_{non-alcohol}$ represents the population proportion of low birth weight for mothers who did not consume alcohol.
Assume significance level of 0.025.

TEST:

Using Z-test.
From the Risk Difference test Table, we get these values: p-value = 0.3396; z-test statistics: 0.4136.
The critical value for a left-tailed test is z_c = 1.96.

Risk Difference Test	
H0: P1 - P2 = 0	Wald Method
Risk Difference	0.0408
ASE (H0)	0.0986
Z	0.4136
One-sided Pr > Z	0.3396
Two-sided Pr > |Z|	0.6792
Column 1 (LOW = 0)	

Column 2 Risk Estimates

	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.0842	0.0074	0.0697	0.0988	0.0702	0.1000
Row 2	0.1250	0.1169	0.0000	0.3542	0.0032	0.5265
Total	0.0845	0.0074	0.0699	0.0990	0.0705	0.1002
Difference	-0.0408	0.1172	-0.2704	0.1889		

Difference is (Row 1 - Row 2)

The FREQ Procedure

Frequency / Percent / Row Pct / Col Pct

Table of DRINK by LOW

DRINK	LOW 0	LOW 1	Total
0	1283 91.06 91.58 99.46	118 8.37 8.42 99.16	1401 99.43
1	7 0.50 87.50 0.54	1 0.07 12.50 0.84	8 0.57
Total	1290 91.55	119 8.45	1409 100.00

Statistics for Table of DRINK by LOW

Column 1 Risk Estimates

	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.9158	0.0074	0.9012	0.9303	0.9000	0.9298
Row 2	0.8750	0.1169	0.6458	1.0000	0.4735	0.9968
Total	0.9155	0.0074	0.9010	0.9301	0.8998	0.9295
Difference	0.0408	0.1172	-0.1889	0.2704		

Difference is (Row 1 - Row 2)

INTERPRETATION OF THE P-VALUE:

If H_0 is true, there is a 0.3396 probability that the percentage of low birth weight for mothers who did not consume alcohol is greater than and equal to the percentage of low birth weight for mothers did consume alcohol.
Compare α and the p-value: α = 0.05; p-value = 0.0098; α > p-value

DECISION:

Since obtained Z > Z_c, fail to reject H_0 .
This means that the hypothesis $p_{non-alcohol} \geq p_{alcohol}$ is not rejected.

CONCLUSION:

At the 0.025 significance level, we conclude that the percentage of low birth weight for mothers who did not consume alcohol is less than and equal to the percentage of low birth weight for mothers who did consume alcohol. The sample proportion illustrates that there is sufficient evidence percentage of low birth weight for mothers who did not consume alcohol is less than and equal to the percentage of low birth weight for mothers who did consume alcohol.

[J]**. Determine if there is sufficient evidence to conclude the percentage of low birth weight has any relation with the race of mothers?**

NULL & ALTERNATIVE HYPOTHESIS:

$$H_0 : Percentage\ of\ low\ birth\ weight\ is\ not\ associated\ with\ races\ of\ mothers$$
$$H_a : Percentage\ of\ low\ birth\ weight\ is\ associated\ with\ races\ of\ mothers$$

Since we want to check the independence of two variables, it is a Chi-square test for proportion.

ASSUMPTIONS:

There are 9 races of mothers: Other Non-White, White, Black, American Indian, Chinese, Japanese, Hawaiian, Filipino, Other Asian or Pacific Islander.

TEST:

Using Chi-square test.
From the Chi-square test Table, we get these values: p-value = 0.0432, df = 5, critical statistic at 5% = 11.070, chi-square = 11.45

The FREQ Procedure

Frequency
Percent
Row Pct
Col Pct

Table of LOW by RACEMOM

	RACEMOM						
LOW	1	2	3	4	7	8	Total
0	963 68.35 74.65 92.86	283 20.09 21.94 87.08	20 1.42 1.55 90.91	2 0.14 0.16 100.00	1 0.07 0.08 100.00	21 1.49 1.63 95.45	1290 91.55
1	74 5.25 62.18 7.14	42 2.98 35.29 12.92	2 0.14 1.68 9.09	0 0.00 0.00 0.00	0 0.00 0.00 0.00	1 0.07 0.84 4.55	119 8.45
Total	1037 73.60	325 23.07	22 1.56	2 0.14	1 0.07	22 1.56	1409 100.00

Statistics for Table of LOW by RACEMOM

Statistic	DF	Value	Prob
Chi-Square	5	11.4478	0.0432
Likelihood Ratio Chi-Square	5	10.8225	0.0550
Mantel-Haenszel Chi-Square	1	0.5883	0.4431
Phi Coefficient		0.0901	
Contingency Coefficient		0.0898	
Cramer's V		0.0901	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

Sample Size = 1409

INTERPRETATION OF THE P-VALUE:

If H_0 is true, there is a 0.0432 probability that the percentage of low birth weight is not associated with races of mothers.
Compare α and the p-value: α = 0.05; p-value = 0.0432; α > p-value

DECISION:

Since critical statistic < obtained chi-square statistic, reject H_0 .
This means that the hypothesis that percentage of low birth weight is not associated with races of mothers is rejected.

CONCLUSION:

At the 5% significance level, we conclude that the percentage of low birth weight is associated with races of mothers. The sample proportion illustrates that there is sufficient evidence that the percentage of low birth weight is associated with the races of mothers.

SAS PROGRAM SCRIPT

```
/* Load Data*/

FILENAME REFFILE '/home/u63530461/STATS3316MINIPROJ/ncbirth1450.csv';

/* import data*/
PROC IMPORT DATAFILE=REFFILE
    DBMS=CSV /* data extension*/
    OUT=birth; /* a name for you data, you can pick any name you like*/
    GETNAMES=YES;
RUN;

title 'MATH 3316-003 TERM PROJECT';
title2 'TEMITAYO ADEROUNMU - CHANDNI PATEL';


/*Deleting rows with missing values*/
data new_birthdata;
    set birth;
    if cmiss(of _all_) then delete;
run;

/*Creating frequency tables for the percentage of low birth weights and smokers*/
PROC FREQ DATA=new_birthdata;
    TABLE LOW SMOKE;
RUN;

/*Creating a summary table*/
PROC MEANS DATA=new_birthdata;
    VAR MAGE WEEKS TOUNCES;
RUN;

/*Determine if there is sufficient evidence to conclude the mean age of mothers giving
birth in North Carolina is over 25 years of age at the 0.05 level of significance*/
/* run t-test */
/* alpha: significance level */
/* data: birth */
/* sides: U=right-sided, L=left-sideds, 2=two-sided */
/* h0: mu=25, null hypothesis, h1: mu>25, alternative hypothesis*/
proc ttest alpha=0.05 data=new_birthdata sides=U h0=25;
    VAR MAGE;
run; /* run t-test*/

/*critical value in t-test*/
/* 1450 is the number of observations*/
data critical;
tc=tinv(0.95,1408); /* get the critical value at confidence level C=0.95, and the freedom
of t distribution is 1409*/
proc print data=critical; /* print the critical value*/
run;

/*Determine if there is sufficient evidence to conclude the mean weeks of gestation of
mothers giving birth in North Carolina is below 39 weeks.*/
proc ttest alpha=0.95 data=new_birthdata sides=L h0=39;
    VAR WEEKS;
run;

data criticalb;
tc=tinv(0.05,1408); /* get the critical value at confidence level C=0.95, and the freedom
of t distribution is 1409*/
proc print data=criticalb; /* print the critical value*/
```

```
run;

/*Determine if there is sufficient evidence to conclude that the mean weight of babies
born to mothers in North Carolina is above 7 lbs*/
proc ttest alpha=0.05 data=new_birthdata sides=U h0=112;
    var tounces;
run;

/*Construct a side-by-side boxplot for tounces for smokers and non-smokers.*/
proc sgplot data=new_birthdata;
    vbox tounces / category = smoke;
    title 'Tounces by smoke';
run;

/*Determine if there is sufficient evidence to conclude the mean tounces of smoking mothers
is lower than the mean birth weight for non-smoking mothers.*/
/* h0: mu1>=mu2, null hypothesis, h1: mu1<mu2, alternative hypothesis*/
proc ttest alpha=0.95 data=new_birthdata sides=2 h0=0;
    CLASS SMOKE;
    VAR tounces;
run;

/*Determine if there is sufficient evidence to conclude the percentage of low birth weight
children in North Carolina is above 6%*/
/* np>=10 and n(1-p)>=10* normality test*/
proc freq data=new_birthdata;
    tables LOW/binomial(p=0.06 level=2);
run;

/*Determine if there is sufficient evidence to conclude the percentage of mothers who
smoke in North Carolina is above 10%.*/
proc freq data=new_birthdata;
    tables SMOKE/binomial(p=0.1 level=2);
run;

/*Determine if there is sufficient evidence to conclude the percentage of low birth weight
for smoking mothers is different than the percentage of low birth weight for non-smoking mothers.*/
proc freq data=new_birthdata;
    tables SMOKE * LOW / riskdiff(equal var =null column =2);
run;

/*Determine if there is sufficient evidence to conclude the percentage of low birth weight
for smoking mothers is lower than the percentage of low birth weight for non-smoking
mothers*/
proc freq data=new_birthdata;
    tables SMOKE * LOW / chisq riskdiff(equal var =null column =2 cl = exact);
run;

/* H */
PROC FREQ DATA = NEW_BIRTHDATA;
    TABLES SMOKE * LOW / RISKDIFF( EQUAL VAR = NULL );
RUN;


/* I */
PROC FREQ DATA = NEW_BIRTHDATA;
    TABLES DRINK * LOW / RISKDIFF( EQUAL VAR = NULL );
RUN;


/* J */
PROC FREQ DATA = NEW_BIRTHDATA;
    TABLES LOW * RACEMOM / CHISQ RISKDIFF;
RUN;
```