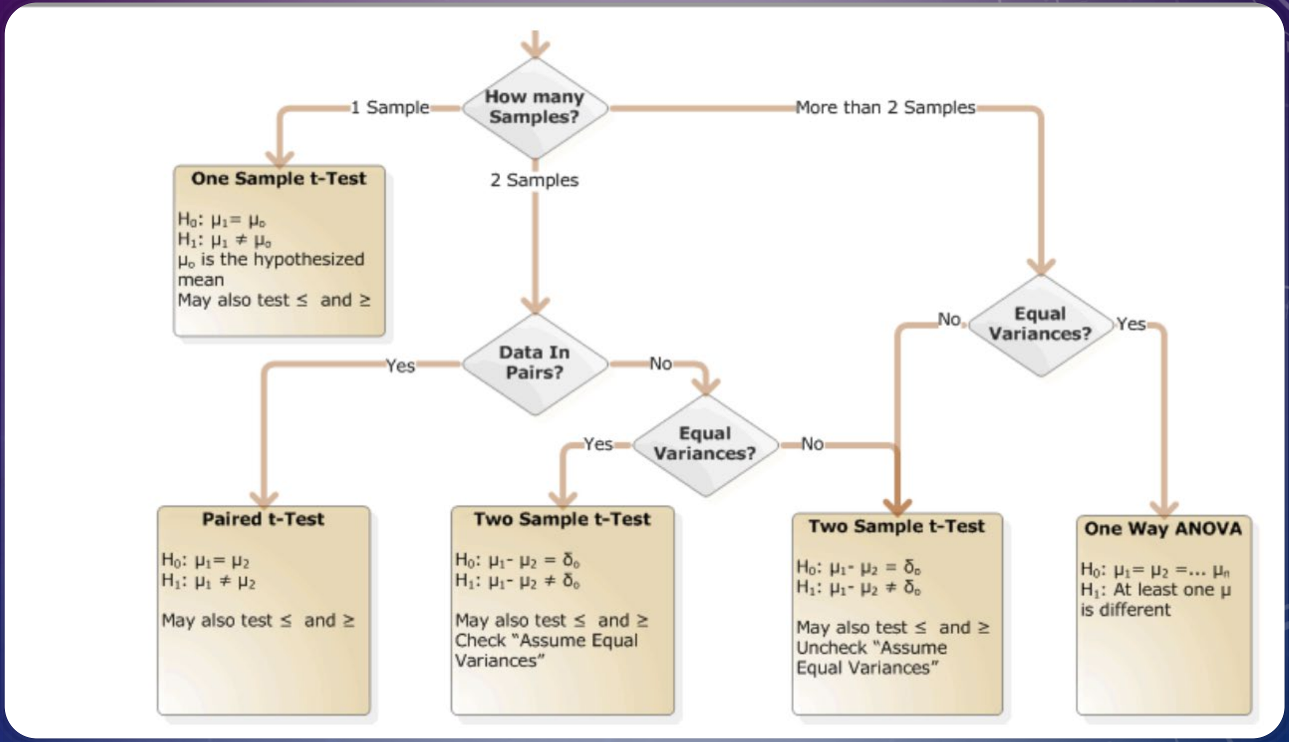
ASSIGNMENT 2 (HYPOTHESIS TESTING) FILE: BIRTH *TECHNICAL DECK*

- Sukanya Mukherjee
- 128041217
- BAN100ZBB
- Statistics for Analytics
- Samaneh Gholami

VARIABLES WITH NO NEGATIVE VALUES AND CLASS VARIABLE HAS ONE OR TWO LEVELS

DATA AUDIT AND DETAILS OF T-TEST

- If we assume all variables as samples of datasets, we must Investigate relationship between the variable "weight" and other variables (does weight vary with other variables) using T-test where feasible.
- Firstly, we will check if all the variables are paired or not.
- Secondly, we will check if all variables have variance equal to weight or not.



PAIRED T-TEST OR TWO SAMPLE T-TEST?

- **To check if all the variables are paired or not:**
- Data is described as unpaired or independent when the sets of data arise from separate individuals.
- Data is described as paired when it arises from the same individual at different points in time.
- So here, we conclude that there are two separate individuals' - mother and baby.
- We go ahead with two-sample t-test and check if the data has equal variances or not.

PROC MEANS TABLE FOR ALL VARIABLES

Variable	Label	Mean	Std Dev	Minimum	Maximum	Median	N	N Miss	Std Error	Variance	Mode	Range	Sum	Lower 95% CL for Mean	Upper 95% CL for Mean	Coeff of Variation	Skewness	Kurtosis	Lower Quartile	Upper Quartile	90th Pctl	99th Pctl	Quartile Range
Weight	Weight	3370.76	566.3850556	240.0000000	6350.00	3402.00	50000	0	2.5329510	320792.03	3402.00	6110.00	168537832	3365.79	3375.72	16.8029056	-0.7554476	2.6900591	3062.00	3720.00	4026.00	4605.00	658.0000000
Black	Black	0.1628400	0.3692233	0	1.0000000	0	50000	0	0.0016512	0.1363259	0	1.0000000	8142.00	0.1596036	0.1660764	226.7399418	1.8263932	1.3357657	0	0	1.0000000	1.0000000	0
Married	Married	0.7126200	0.4525448	0	1.0000000	1.0000000	50000	0	0.0020238	0.2047968	1.0000000	1.0000000	35631.00	0.7086532	0.7165868	63.5043697	-0.9397014	-1.1170060	0	1.0000000	1.0000000	1.0000000	1.0000000
Boy	Boy	0.5158400	0.4997540	0	1.0000000	1.0000000	50000	0	0.0022350	0.2497541	1.0000000	1.0000000	25792.00	0.5114594	0.5202206	96.8815968	-0.0633937	-1.9960611	0	1.0000000	1.0000000	1.0000000	1.0000000
MomAge	MomAge	0.4161400	5.7284539	-9.0000000	18.0000000	0	50000	0	0.0256184	32.8151838	-1.0000000	27.0000000	20807.00	0.3659276	0.4663524	1376.57	0.2713586	-0.6666093	-4.0000000	5.0000000	8.0000000	14.0000000	9.0000000
MomSmoke	MomSmoke	0.1306600	0.3370315	0	1.0000000	0	50000	0	0.0015073	0.1135902	0	1.0000000	6533.00	0.1277058	0.1336142	257.9454357	2.1918116	2.8041500	0	0	1.0000000	1.0000000	0
CigsPerDay	CigsPerDay	1.4766200	4.6540656	0	60.0000000	0	50000	0	0.0208136	21.6603266	0	60.0000000	73831.00	1.4358251	1.5174149	315.1837031	3.8870572	17.7074723	0	0	5.0000000	20.0000000	0
MomWtGain	MomWtGain	0.7092200	12.8761168	-30.0000000	68.0000000	0	50000	0	0.0575837	165.7943829	0	98.0000000	35461.00	0.5963552	0.8220848	1815.53	0.4620977	1.1342923	-8.0000000	9.0000000	17.0000000	36.0000000	17.0000000
Visit	Visit	2.6999800	0.7154986	0	3.0000000	3.0000000	50000	0	0.0031998	0.5119382	3.0000000	3.0000000	134999.00	2.6937083	2.7062517	26.5001442	-2.0923650	2.7215145	3.0000000	3.0000000	3.0000000	3.0000000	0
MomEdLevel	MomEdLevel	1.2189200	1.0899072	0	3.0000000	1.0000000	50000	0	0.0048742	1.1878978	0	3.0000000	60946.00	1.2093665	1.2284735	89.4158140	0.2974152	-1.2507541	0	2.0000000	3.0000000	3.0000000	2.0000000

VARIANCE TABLE FOR ALL VARIABLES

S.No	Variable	Variance
1	Weight	320792.03
2	Black	0.1363259
3	Married	0.2047968
4	Boy	0.2497541
5	MomAge	32.8151838
6	MomSmoke	0.1135902
7	CigsPerDay	21.6603266
8	MomWtGain	165.7943829
9	Visit	0.5119382
10	MomEdLevel	1.1878978

- If we see and compare variances of "Weight" with all other variables, variances are NOT EQUAL.
- So, we go ahead with two-sample t-test.

VARIABLES: WEIGHT AND ETHNICITY

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H*0: variance of weight = variance of ethnicity.

<u>*HA*: The variances are unequal.</u>

<u>Variance of variable weight (320792.03) > Variance of variable black (0.1363259)</u>

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: t-statistics is 34.01 with degree of freedom 10808, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: µ1 = µ2

<u>*HA*: The means are unequal.</u>

<u>Mean of variable weight (3370.76) > Mean of variable black (0.1628400).</u>

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

ONE-TAIL TEST WITH UNEQUAL VARIANCES

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

Ztest:

Z= $(3370.76 \; - 0.162840) - (0) \div \sqrt{(566.3850556)^2 + (0.3692233)}$ ^{2}/50000

Z=3370.59716/2.53295151

Z= 1330.69944

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

Confidence Interval Range:

3365.63258 to 3375.56174

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of “no effect” or “no difference.”
- Here, weight of the baby “makes no difference and is independent” of the ethnicity of the mother.

H0: µ1 – µ2 = 0, which is the same as *H0*: µ1 = µ2

<u>Since p-value < 0.05 so we reject *H0*: µ1 = µ2</u>

- The alternative hypothesis, *Ha*,

Ha: µ1 – µ2 > 0, which is the same as *Ha*: µ1 > µ2

<u>This is TRUE and means that Weight of the baby is “DEPENDENT” on the Ethnicity of the mother.</u>

µ1= mean of variable weight, µ2= mean of variable black.

VARIABLES: WEIGHT AND MARITAL STATUS

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H*0: variance of weight = variance of marital status.

HA: The variances are unequal.

Variance of variable weight (320792.03) > Variance of variable married (0.2047968)

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: t-statistics is -33.88 with degree of freedom 25443, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: µ1 = µ2

HA: The means are unequal.

Mean of variable weight (3370.76) > Mean of variable married (0.7126200).

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

ONE-TAIL TEST WITH UNEQUAL VARIANCES

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

<u>Ztest:</u>

Z= $(3370.76 - 0.7126200) - (0) \div$
$\sqrt{(566.3850556)^2 + (0.4525448)} \; {}^2/50000$

Z=3370.047388/2.53295178

Z= 1330.48225

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

<u>Confidence Interval Range:</u>

3365.0828 to 3375.01197

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of “no effect” or “no difference.”
- Here, weight of the baby “makes no difference and is independent” of the marital status of the mother.

H0: $\mu 1 - \mu 2 = 0$, which is the same as *H0*: $\mu 1 = \mu 2$

<u>Since p-value < 0.05 so we reject *H0*: $\mu 1 = \mu 2$</u>

- The alternative hypothesis, *Ha*,

Ha: $\mu 1 - \mu 2 > 0$, which is the same as *Ha*: $\mu 1 > \mu 2$

<u>This is TRUE and means that Weight of the baby is “DEPENDENT” on the Marital status of the mother.</u>

$\mu 1$= mean of variable weight, $\mu 2$= mean of variable married.

VARIABLES: WEIGHT AND GENDER

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H*0: variance of weight = variance of gender.

<u>*HA*: The variances are unequal.</u>

<u>Variance of variable weight (320792.03) > Variance of variable boy (0.2497541)</u>

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: t-statistics is -23.18 with degree of freedom 49993, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: μ1 = μ2

<u>*HA*: The means are unequal.</u>

<u>Mean of variable weight (3370.76) > Mean of variable boy (0.5158400).</u>

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

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

ONE-TAIL TEST WITH UNEQUAL VARIANCES

Ztest:

Z= $(3370.76 \ - 0.5158400) - (0) \div \sqrt{(566.3850556)^2 + (0.4997540)} \ {}^2$/50000

Z=3370.24416/2.53295196

Z= 1330.55984

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

Confidence Interval Range:

3365.27957 to 3375.20875

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of “no effect” or “no difference.”
- Here, weight of the baby “makes no difference and is independent” of the gender of the baby.

H0: $\mu1 – \mu2 = 0$, which is the same as *H0*: $\mu1 = \mu2$

<u>Since p-value < 0.05 so we reject *H0*: $\mu1 = \mu2$</u>

- The alternative hypothesis, *Ha*,

Ha: $\mu1 – \mu2 > 0$, which is the same as *Ha*: $\mu1 > \mu2$

<u>This is TRUE and means that Weight of the baby is “DEPENDENT” on the Gender of the baby.</u>

$\mu1$= mean of variable weight, $\mu2$= mean of variable boy.

VARIABLES: WEIGHT AND MOM SMOKES

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H*0: variance of weight = variance of mom smokes or not.

HA: The variances are unequal.

Variance of variable weight (320792.03) > Variance of variable momsmoke (0.1135902)

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: t-statistics is 31.68 with degree of freedom 8474.1, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: $\mu 1 = \mu 2$

HA: The means are unequal.

Mean of variable weight (3370.76) > Mean of variable momsmoke (0.1306600).

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

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

ONE-TAIL TEST WITH UNEQUAL VARIANCES

Ztest:

Z= $(3370.76 - 0.1306600) - (0) \div \sqrt{(566.3850556)^2 + (0.3370315)}$ ^{2}/50000

Z=3370.62934/2.53295142

Z= 1330.71219

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

Confidence Interval Range:

3365.66476 to 3375.59392

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of "no effect" or "no difference."
- Here, weight of the baby "makes no difference and is independent" of the smoking mother.

H0: $\mu 1 - \mu 2 = 0$, which is the same as *H0*: $\mu 1 = \mu 2$

<u>**Since p-value < 0.05 so we reject *H0*: $\mu 1 = \mu 2$**</u>

- The alternative hypothesis, *Ha*,

Ha: $\mu 1 - \mu 2 > 0$, which is the same as *Ha*: $\mu 1 > \mu 2$

<u>**This is TRUE and means that Weight of the baby is "DEPENDENT" on the Smoking mother.**</u>

$\mu 1$= mean of variable weight, $\mu 2$= mean of variable momsmoke.

VARIABLES WITH NO NEGATIVE VALUES AND CLASS VARIABLE HAS ONE OR TWO LEVELS

NEGATIVE VALUES :

NEWMOMAGE = MOMAGE + 25 (TO REMOVE ALL NEGATIVE VALUES)

NEWMOMWTGAIN = MOMWTGAIN +30 (TO REMOVE ALL NEGATIVE VALUES)

CLASS VARIABLE MORE THAN TWO LEVELS :

```
PROC GLM DATA=SASDATASETNAME;
CLASS COMPARISONVARIABLE;
MODEL WEIGHT = COMPARISONVARIABLE;
LSMEANS COMPARISONVARIABLE / PDIFF;
```

MEANS AND VARIANCES

Variable	Label	Mean	Std Dev	N	Variance	Mode
newmomage	newmomage	25.4161400	5.7284539	50000	32.8151838	24.0000000
newmomwtgain	newmomwtgain	30.7092200	12.8761168	50000	165.7943829	30.0000000
CigsPerDay	CigsPerDay	1.4766200	4.6540656	50000	21.6603266	0
Visit	Visit	2.6999800	0.7154986	50000	0.5119382	3.0000000
MomEdLevel	MomEdLevel	1.2189200	1.0899072	50000	1.1878978	0

VARIABLES: WEIGHT AND MOM AGE

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H*0: variance of weight = variance of mom age.

<u>*HA*: The variances are unequal.</u>

<u>Variance of variable weight (320792.03) > Variance of variable newmomage (32.8151838)</u>

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: degree of freedom 27, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: µ1 = µ2

<u>*HA*: The means are unequal.</u>

<u>Mean of variable weight (3370.76) > Mean of variable newmomage (25.4161400).</u>

Dependent Variable: Weight Weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	27	218477479	8091758	25.56	<.0001
Error	49972	15820803288	316593		
Corrected Total	49999	16039280767			

R-Square	Coeff Var	Root MSE	Weight Mean
0.013621	16.69258	562.6663	3370.757

Source	DF	Type I SS	Mean Square	F Value	Pr > F
newmomage	27	218477478.9	8091758.5	25.56	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
newmomage	27	218477478.9	8091758.5	25.56	<.0001

ONE-TAIL TEST WITH UNEQUAL VARIANCES

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

Ztest:

Z= $(3370.76 \ - 25.41614) - (0) \div$

$\sqrt{(320792.031)}$ + (32.8151838)/50000

Z=3345.34386/ $\sqrt{(6.416496924)}$

Z= 1320.662266

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

Confidence Interval Range:

3340.379022 to 3350.308698

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of “no effect” or “no difference.”
- Here, weight of the baby “makes no difference and is independent” of the age of mother.

H0: $\mu 1 - \mu 2 = 0$, which is the same as *H0*: $\mu 1 = \mu 2$

Since p-value < 0.05 so we reject *H0*: $\mu 1 = \mu 2$

- The alternative hypothesis, *Ha*,

Ha: $\mu 1 - \mu 2 > 0$, which is the same as *Ha*: $\mu 1 > \mu 2$

This is TRUE and means that Weight of the baby is “DEPENDENT” on the Age of mother.

$\mu 1$= mean of variable weight, $\mu 2$= mean of variable newmomage.

VARIABLES: WEIGHT AND MOM WEIGHT GAIN

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H*0: variance of weight = variance of mom gained weight or not.

HA: The variances are unequal.

Variance of variable weight (320792.03) > Variance of variable newmomwtgain (165.7943829)

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: degree of freedom 95, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: µ1 = µ2

HA: The means are unequal.

Mean of variable weight (3370.76) > Mean of variable newmomwtgain (30.7092200).

Dependent Variable: Weight Weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	95	781626446	8227647	26.91	<.0001
Error	49904	15257654321	305740		
Corrected Total	49999	16039280767			

R-Square	Coeff Var	Root MSE	Weight Mean
0.048732	16.40396	552.9377	3370.757

Source	DF	Type I SS	Mean Square	F Value	Pr > F
newmomwtgain	95	781626445.9	8227646.8	26.91	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
newmomwtgain	95	781626445.9	8227646.8	26.91	<.0001

ONE-TAIL TEST WITH UNEQUAL VARIANCES

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

Ztest:

Z= $(3370.76 - 30.70922) - (0) \div \sqrt{(320792.031)} + (165.7943829)/50000$

Z=3370.62934/ $\sqrt{(6.419156508)}$

Z= 1318.2995

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

Confidence Interval Range:

3335.084913 to 3345.016647

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of “no effect” or “no difference.”
- Here, weight of the baby “makes no difference and is independent” of the weight gained by mother.

H0: $\mu1 - \mu2 = 0$, which is the same as *H0*: $\mu1 = \mu2$

<u>Since p-value < 0.05 so we reject *H0*: µ1 = µ2</u>

- The alternative hypothesis, *Ha*,

Ha: $\mu1 - \mu2 > 0$, which is the same as *Ha*: $\mu1 > \mu2$

<u>This is TRUE and means that Weight of the baby is “DEPENDENT” on the Weight gained by mother.</u>

$\mu1$= mean of variable weight, $\mu2$= mean of variable newmomwtgain.

VARIABLES: WEIGHT AND CIGARETTES PER DAY

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H*0: variance of weight = variance of how many cigs mom had per day.

HA: The variances are unequal.

Variance of variable weight (320792.03) > Variance of variable cigsperday (21.6603266)

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: degree of freedom 31, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: μ1 = μ2

HA: The means are unequal.

Mean of variable weight (3370.76) > Mean of variable cigsperday (1.4766200).

Dependent Variable: Weight Weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	31	346199537	11167727	35.56	<.0001
Error	49968	15693081230	314063		
Corrected Total	49999	16039280767			

R-Square	Coeff Var	Root MSE	Weight Mean
0.021584	16.62573	560.4129	3370.757

Source	DF	Type I SS	Mean Square	F Value	Pr > F
CigsPerDay	31	346199537.0	11167727.0	35.56	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
CigsPerDay	31	346199537.0	11167727.0	35.56	<.0001

ONE-TAIL TEST WITH UNEQUAL VARIANCES

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

Ztest:

Z= $(3370.76 \; - 1.47662) - (0) \div$

$\sqrt{(320792.031} + (21.6603266)$ /50000

Z=3370.62934/ $\sqrt{(6.416273827}$

Z= 1330.136144

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

Confidence Interval Range:

3364.318628 to 3374.248132

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of "no effect" or "no difference."
- Here, weight of the baby "makes no difference and is independent" of the amount of cigarettes mother takes per day.

H0: $\mu1 - \mu2 = 0$, which is the same as *H0*: $\mu1 = \mu2$

<u>Since p-value < 0.05 so we reject *H0*: $\mu1 = \mu2$</u>

- The alternative hypothesis, *Ha*,

Ha: $\mu1 - \mu2 > 0$, which is the same as *Ha*: $\mu1 > \mu2$

<u>This is TRUE and means that Weight of the baby is "DEPENDENT" on the Amount of cigarettes mother takes per day.</u>

$\mu1$= mean of variable weight, $\mu2$= mean of variable cigsperday.

VARIABLES: WEIGHT AND MOM'S PRENATAL VISITS

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H0*: variance of weight = variance of mom's prenatal visits.

<u>*HA*: The variances are unequal.</u>

<u>Variance of variable weight (320792.03) > Variance of variable visit (0.5119382)</u>

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: degree of freedom 3, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: µ1 = µ2

<u>*HA*: The means are unequal.</u>

<u>Mean of variable weight (3370.76) > Mean of variable visit (2.6999800).</u>

Dependent Variable: Weight Weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	90668628	30222876	94.74	<.0001
Error	49996	15948612139	318998		
Corrected Total	49999	16039280767			

R-Square	Coeff Var	Root MSE	Weight Mean
0.005653	16.75585	564.7989	3370.757

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Visit	3	90668627.80	30222875.93	94.74	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Visit	3	90668627.80	30222875.93	94.74	<.0001

ONE-TAIL TEST WITH UNEQUAL VARIANCES

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

Ztest:

Z= $(3370.76\ - 2.69998) - (0) \div$

$\sqrt{(320792.031} + (0.5119382)$ /50000

Z=3370.62934/ $\sqrt{(6.415850859}$

Z= 1329.69701

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

Confidence Interval Range:

3363.095432 to 3373.024608

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of "no effect" or "no difference."
- Here, weight of the baby "makes no difference and is independent" of the prenatal visit of the mother.

H0: $\mu1 - \mu2 = 0$, which is the same as *H0*: $\mu1 = \mu2$

Since p-value < 0.05 so we reject *H0*: $\mu1 = \mu2$

- The alternative hypothesis, *Ha*,

Ha: $\mu1 - \mu2 > 0$, which is the same as *Ha*: $\mu1 > \mu2$

This is TRUE and means that Weight of the baby is "DEPENDENT" on the Prenatal visit of the mother.

$\mu1$= mean of variable weight, $\mu2$= mean of variable visit.

VARIABLES: WEIGHT AND MOM'S EDUCATION LEVEL

- Step 1: Check equal variance:

4th table: The p-value for the F test using SAS is not significant at α=0.05 (p = < 0.0001),

so, we **reject** *H*0: variance of weight = variance of mom's education level.

HA: The variances are unequal.

Variance of variable weight (320792.03) > Variance of variable momedlevel (1.1878978)

- Step 2: Test the null hypothesis of unequal means using the t-test assuming not equal variances:

3rd table: degree of freedom 3, with p-value= < 0.0001.

Since p-value < 0.05 so we **reject** *H0*: µ1 = µ2

HA: The means are unequal.

Mean of variable weight (3370.76) > Mean of variable momedlevel (1.2189200).

Dependent Variable: Weight Weight

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	240093609	80031203	253.26	<.0001
Error	49996	15799187158	316009		
Corrected Total	49999	16039280767			

R-Square	Coeff Var	Root MSE	Weight Mean
0.014969	16.67717	562.1468	3370.757

Source	DF	Type I SS	Mean Square	F Value	Pr > F
MomEdLevel	3	240093608.9	80031203.0	253.26	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
MomEdLevel	3	240093608.9	80031203.0	253.26	<.0001

$$Z = \frac{\text{(Observed difference in sample means)} - \text{(Hypothesized difference in population means)}}{\text{Standard error}}$$

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the average of two samples.
- s_1 and s_2 are the standard deviation of the two samples
- Since the null hypothesis assumes there is no difference in the population means, the expression $(\mu_1 - \mu_2)$ is always zero.

ONE-TAIL TEST WITH UNEQUAL VARIANCES

Ztest:

Z= $(3370.76 \; - 1.21892) - (0) \div$

$\sqrt{(320792.031} + (1.1878978)$ /50000

Z=3370.62934/ $\sqrt{(6.415864378}$

Z= 1330.280325

95% confidence interval= 1-0.95= 0.05 (one tail test)

Z critical: 1.96

Confidence Interval Range:

3364.576487 to 3374.505673

DETERMINE AND CONCLUDE THE HYPOTHESIS

- The null hypothesis, *H0*, is a statement of “no effect” or “no difference.”
- Here, weight of the baby “makes no difference and is independent” of the education level of the mother.

H0: $\mu1 - \mu2 = 0$, which is the same as *H0*: $\mu1 = \mu2$

Since p-value < 0.05 so we reject *H0*: $\mu1 = \mu2$

- The alternative hypothesis, *Ha*,

Ha: $\mu1 - \mu2 > 0$, which is the same as *Ha*: $\mu1 > \mu2$

This is TRUE and means that Weight of the baby is “DEPENDENT” on the Education level of the mother.

$\mu1$= mean of variable weight, $\mu2$= mean of variable momedlevel.