Details of T-test and Data Audit

T-test Taking Class as MomSmoke and Variable as Weight.

- P-value for Equality of variance is 0.0004
- Since P-value is below the significance value of 0.05. this means that a Satterthwaite t-test is the best option to use to compare birth weights.

Two Sample T-test:
Weight by MomSmoke

P-Value is <.0001
Alternative Hypothesis: True difference in means is not equal to 0.

95 % Confidence Interval:
226.5   256.4

Mean in Group 0 is 3402.3 and mean in group 1 is 3160.9.

Conclusion:
Based on Our Satterthwaite t-test. We can see that the P-value falls under the significance level of 0.05. This means that we can reject the null hypothesis and conclude that the mean Weight of the babies of smoking mother and non smoking mother are different. We Support:
Ha: Mean of mothers who don't smoke (0) ≠ Mean of mothers who smoke (1)
That means weight of the infants of smoking mom differs from non smoking mom.

The TTEST Procedure

Variable: Weight (Weight)

Mom Smoke	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		

Mom Smoke	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

T-test Taking Class as Boy and Variable as Weight.

- P-value for Equality of variance is <.0001
- Since P-value is below the significance value of 0.05. this means that a Satterthwaite t-test is the best option to use to compare birth weights.

Two Sample T-test:
Weight by Boy

P-Value is <.0001
Alternative Hypothesis: True difference in means is not equal to 0.

95 % Confidence Interval:
-126.6   -106.8

Mean in Group 0 is 3303.7 and mean in group 1 is 3427.3.

Conclusion:
Based on Our Satterthwaite t-test. We can see that the P-value falls under the significance level of 0.05. This means that we can reject the null hypothesis and conclude that the mean Weight of the baby Boy and baby Girl is not same. We support:
Ha: Mean weight of girl (0) ≠ Mean weight of boy(1)
That means weight of baby boy and baby girl are different.

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

T-test Taking Class as Black and Variable as Weight.

- P-value for Equality of variance is <.0001
- Since P-value is below the significance value of 0.05. this means that a Satterthwaite t-test is the best option to use to compare birth weights.

Two Sample T-test:
Weight by Black

P-Value is <.0001
Alternative Hypothesis: True difference in means is not equal to 0.

95 % Confidence Interval:
234.2   262.9

Mean in Group 0 is 3411.2 and mean in group 1 is 3162.7.

Conclusion:

Based on Our Satterthwaite t-test. We can see that the P-value falls under the significance level of 0.05. This means that we can reject the null hypothesis and conclude that the mean Weight of the baby with black mother and non black mother is not same. We support:

Ha: Mean weight of babies with black mother(1) ≠ Mean weight of babies with non black mother(0)

That means weight of babies with black mother is different from non black mother.

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	10806	34.01	<.0001

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

T-test Taking Class as Married and Variable as Weight.

- P-value for Equality of variance is <.0001
- Since P-value is below the significance value of 0.05. this means that a Satterthwaite t-test is the best option to use to compare birth weights.

Two Sample T-test:
Weight by Married

P-Value is <.0001
Alternative Hypothesis: True difference in means is not equal to 0.

95 % Confidence Interval:
-202.4   -180.2

Mean in Group 0 is 3234.4 and mean in group 1 is 3425.7.

Conclusion:
Based on Our Satterthwaite t-test. We can see that the P-value falls under the significance level of 0.05. This means that we can reject the null hypothesis and conclude that the mean Weight of the baby of married mother and mean Weight of the baby of single mother is not same. We support:
Ha: Mean weight of the babies of single moms(0) ≠ Mean weight of babies of married moms(1)
That means weight of the babies of single mom and married mom varies.

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

Metadata description

Variable	Description
Weight	Weight of the newly born babies.
Black	Black mother (1 = Yes, 0 = No)
Married	Mother is married or not (1=Yes, 0=No)
Boy	Infant is boy or girl (1=Boy, 0=Girl)
MomAge	Age is centered to 27. Thus MomAge=-7 means the mother was 20 years old whereas MomAge=5 means that the mother was 32 years old.
MomSmoke	Mother giving birth is a smoker or non smoker(1=smokes,0=Doesn't smoke)

Variables in Creation Order					
#	Variable	Type	Len	Format	Label
1	Weight	Num	8	BEST.	Weight
2	Black	Num	8	BEST.	Black
3	Married	Num	8	BEST.	Married
4	Boy	Num	8	BEST.	Boy
5	MomAge	Num	8	BEST.	MomAge
6	MomSmoke	Num	8	BEST.	MomSmoke
7	CigsPerDay	Num	8	BEST.	CigsPerDay
8	MomWtGain	Num	8	BEST.	MomWtGain
9	Visit	Num	8	BEST.	Visit
10	MomEdLevel	Num	8	BEST.	MomEdLevel

Metadata description

| Variable	Description
CigsPerDay	Number of cigarettes mother smokes per day.
MomWtGain	How much does mother gained of decreased weight during pregnancy. 0 stands for no change in weight where as –ve value states decrease in weight and +ve value states increase in weight.
Visit	Number of prenatal visits by mother.
MomEdLevel	How much educated is the mother this is divided into 4 levels i.e. 0,1,2,3. Where 0 being the least educated and 3 being the most educated.

Variables in Creation Order

#	Variable	Type	Len	Format	Label
1	Weight	Num	8	BEST.	Weight
2	Black	Num	8	BEST.	Black
3	Married	Num	8	BEST.	Married
4	Boy	Num	8	BEST.	Boy
5	MomAge	Num	8	BEST.	MomAge
6	MomSmoke	Num	8	BEST.	MomSmoke
7	CigsPerDay	Num	8	BEST.	CigsPerDay
8	MomWtGain	Num	8	BEST.	MomWtGain
9	Visit	Num	8	BEST.	Visit
10	MomEdLevel	Num	8	BEST.	MomEdLevel

Increase, Decrease and same weight of the mother during pregnancy.

- I divided the variable into 3 different parts
    1. Increased (indicates increase in weight)
    2. Decreased (indicates decrease in weight)
    3. Same(no change in weight)

We can see in the below Bar Diagram that Decreased and Increased Category are almost same with slight difference in increased category where as Same(no change in weight) have the least number of mothers.
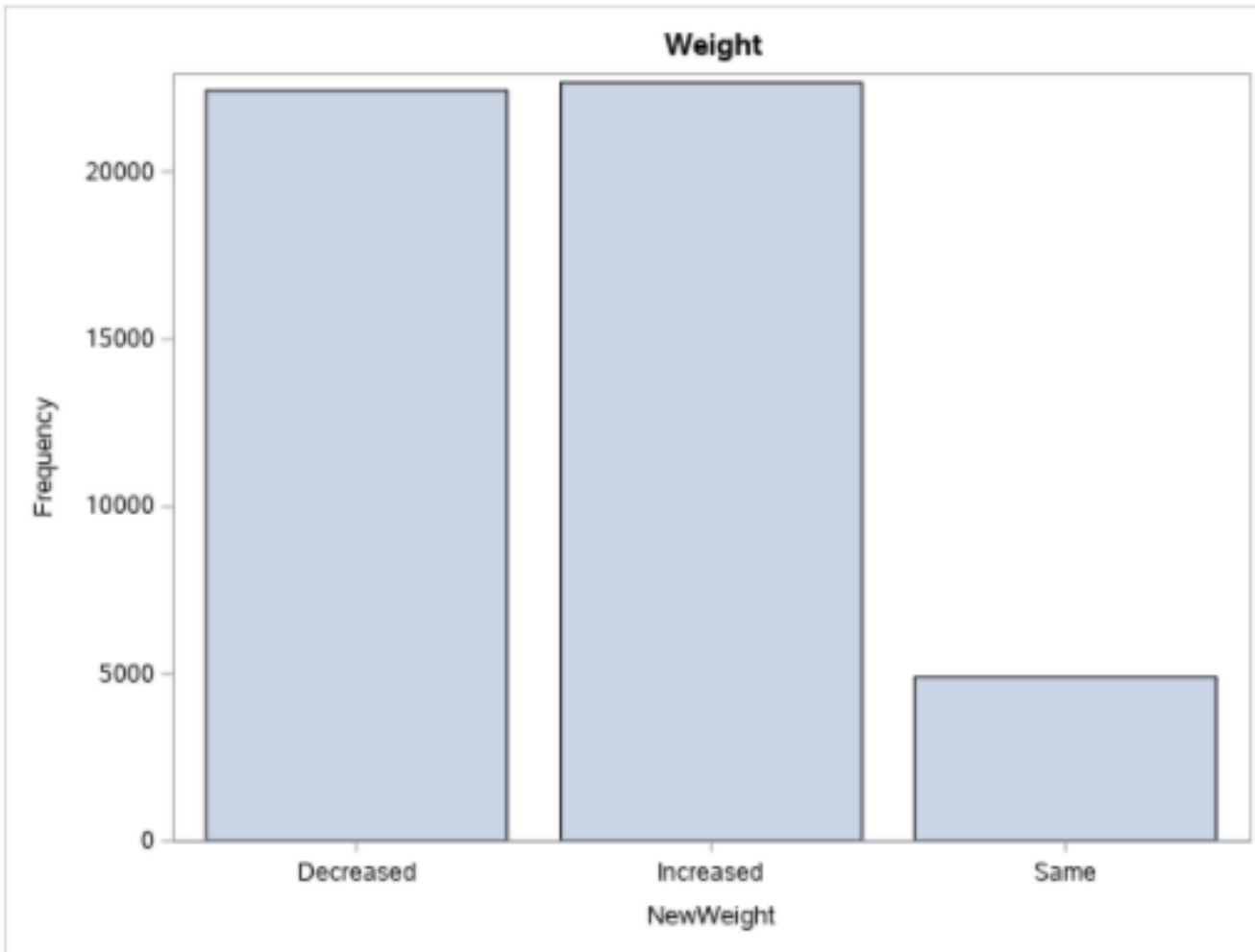
Out of 50 Thousand Data:
Decreased = 22432
Increased = 22665
Same = 4903

We can see that 90.19% of mother experience change in their weight during pregnancy.

Conclusion: Change in Weight of mother during pregnancy is normal.

Average Age of mother giving birth: using proc means

The MEANS Procedure

Analysis Variable : MomAge MomAge				
N	Mean	Std Dev	Minimum	Maximum
50000	0.42	5.73	-9.00	18.00

Considering all 50000 observations we can see that the smallest age is 18 years that is 27-9=18 and the biggest age is 45 years.
And the average age of mother giving birth is 27 years considering all 50000 obs.
That is the reason why age is centered to 27.

Age distribution of mothers

Mom Age

- Below Average 22,929
- Average 2,974
- Above average 24,097

Mom Age ■ Below Average ■ Above average ■ Average
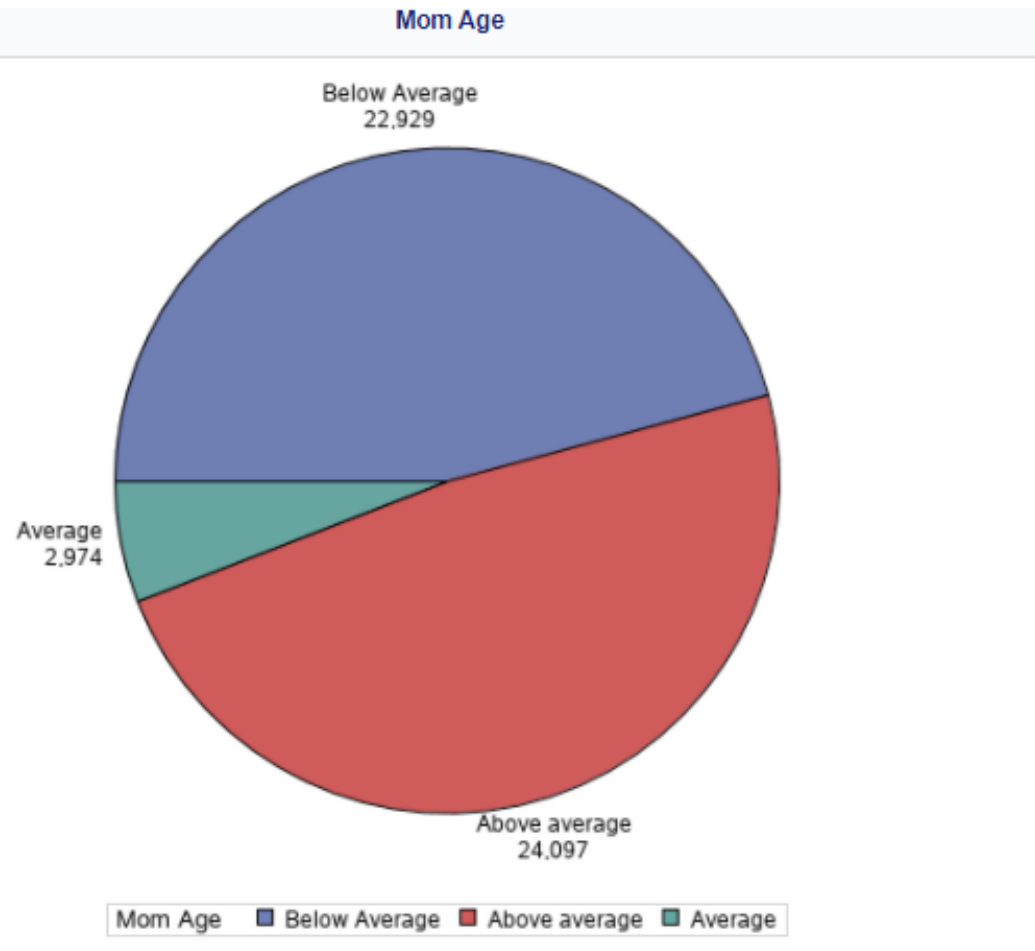
- 45.80% Mothers are below average age.
- 48.19% Mothers are above average age.
- 6.01% Mothers are of the average age.

Mom education levels
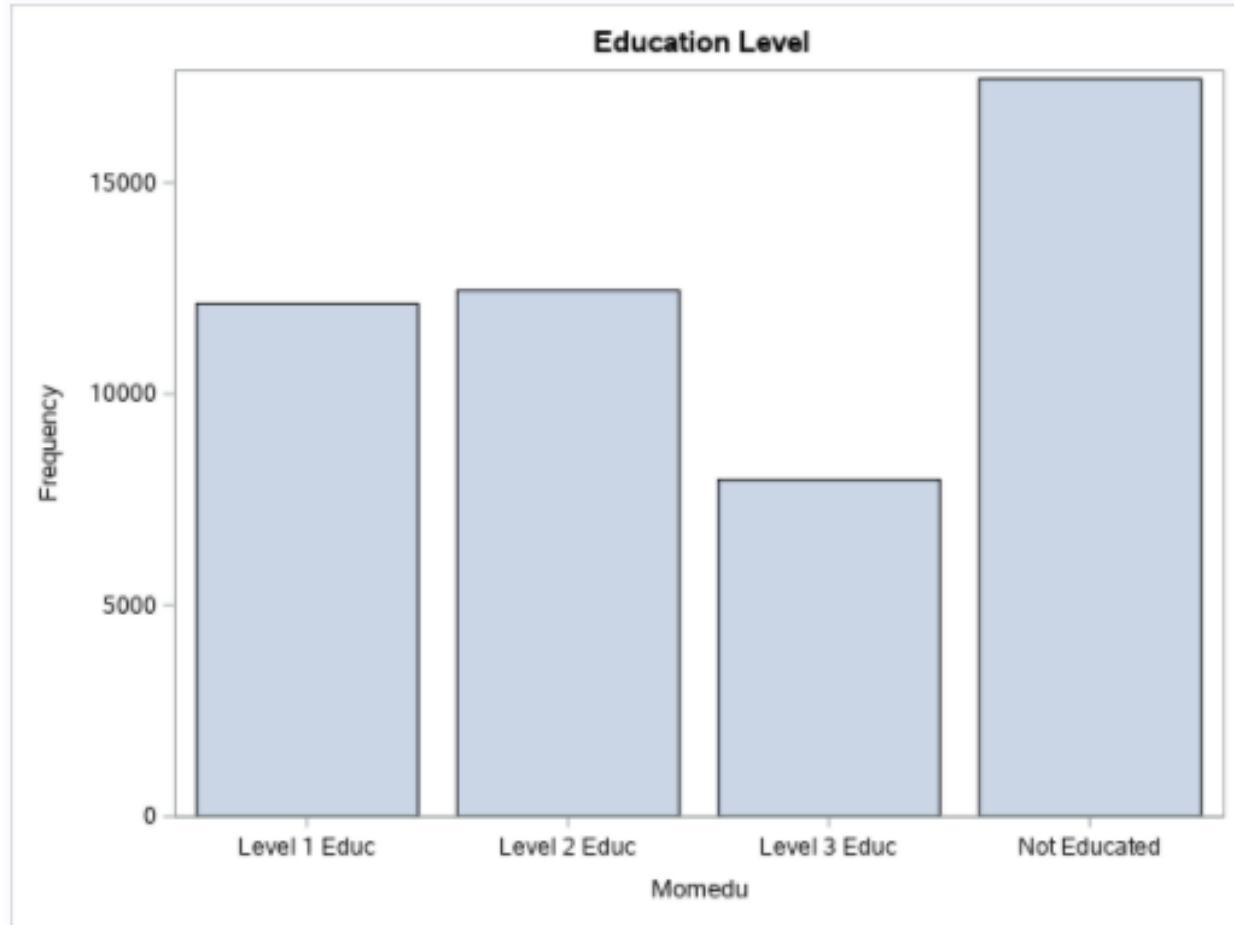
- From the given graph we can analyse that majority of the mother is not educated
- Out of 50000 observations 17449 mothers are not educated and only 7973 mothers are educated up to level 3.

# Probability of having an underweight boy baby (less than 3000 grams).

Predicted Probabilities for Underweight = 1
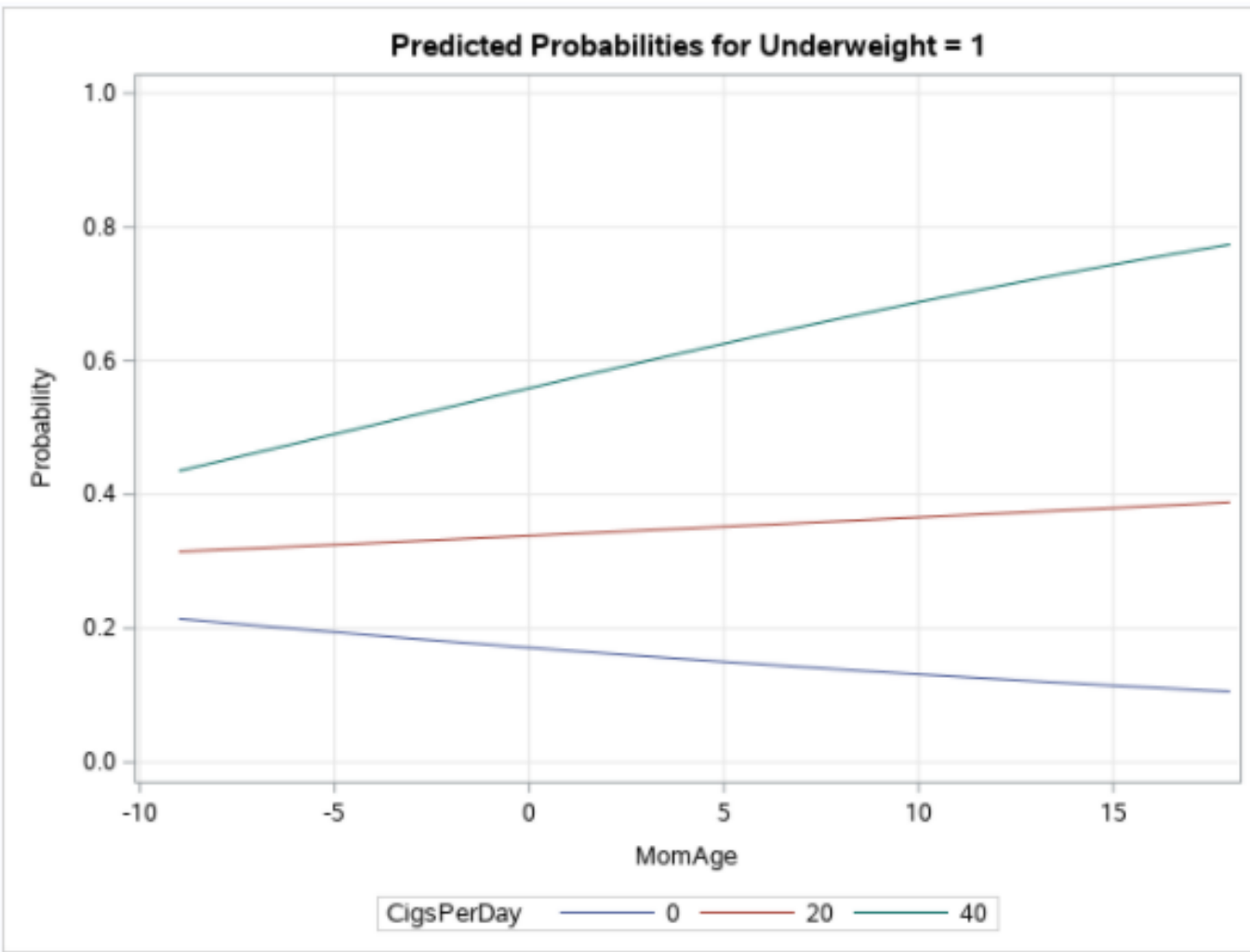
(Remember: the age is centered at 27 years).
In this example, output shows the predicted probability of having an underweight boy baby as a function of the mother's relative age. The panel enables you to see how the predicted response changes with increasing cigarette use.

The above graph shows that:

- The probability of an underweight boy is very low in *non-smoking* mothers, regardless of the mother's age.
- In smoking mothers, however, the probability of having an underweight boy increases with age.
- For mothers of a given age, the probability of an underweight boy increases with the number of cigarettes smoked.

Summary of Weight of infants in percentage.

Distribution of Weight
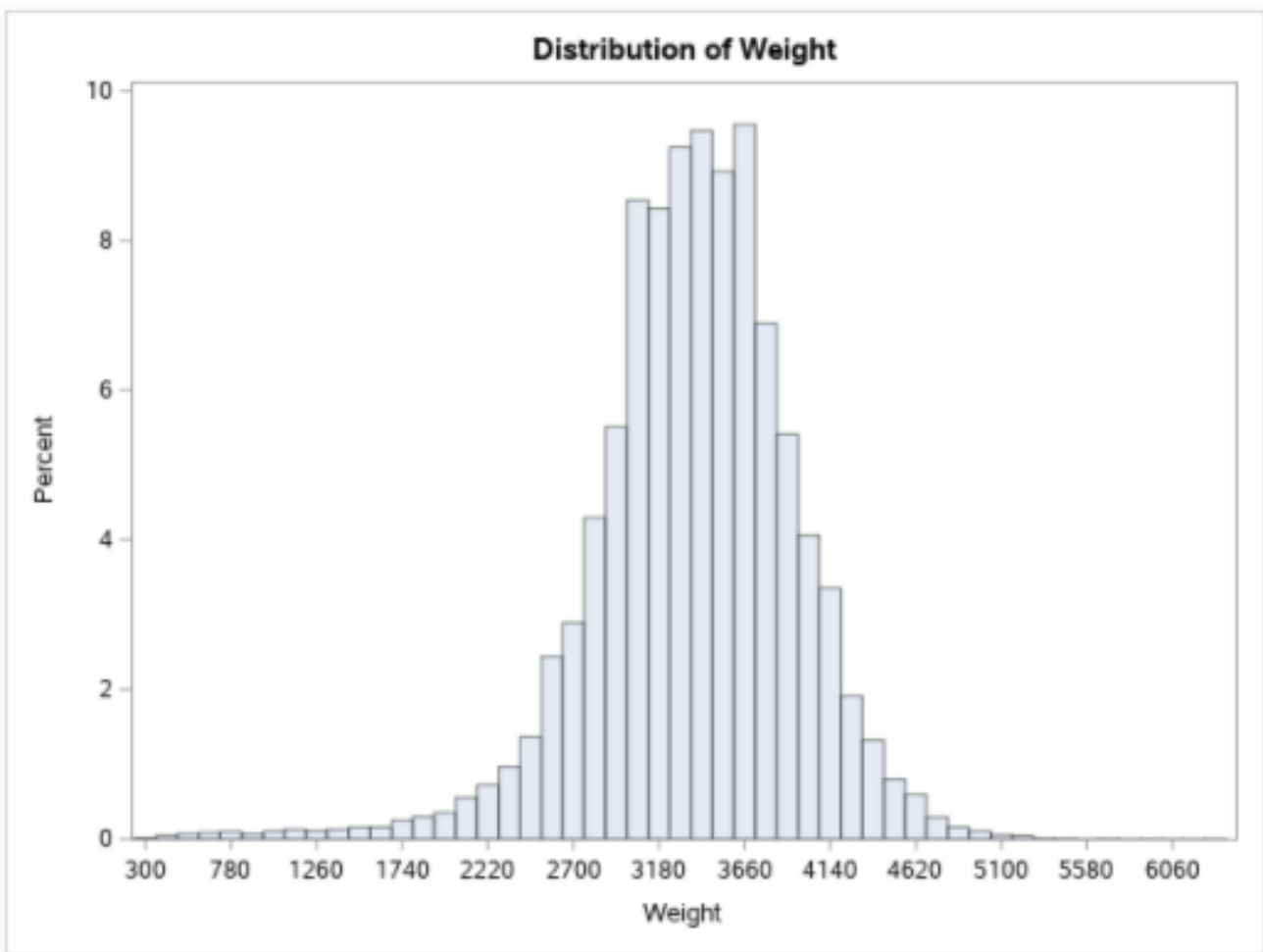
We can see that the maximum percentage of infants are weighing between 3062 grams (Q1) and 3720 grams (Q3).

Quantiles (Definition 5)	
Level	Quantile
100% Max	6350
99%	4605
95%	4224
90%	4026
75% Q3	3720
50% Median	3402
25% Q1	3062
10%	2722
5%	2466
1%	1559
0% Min	240

Prenatal visits
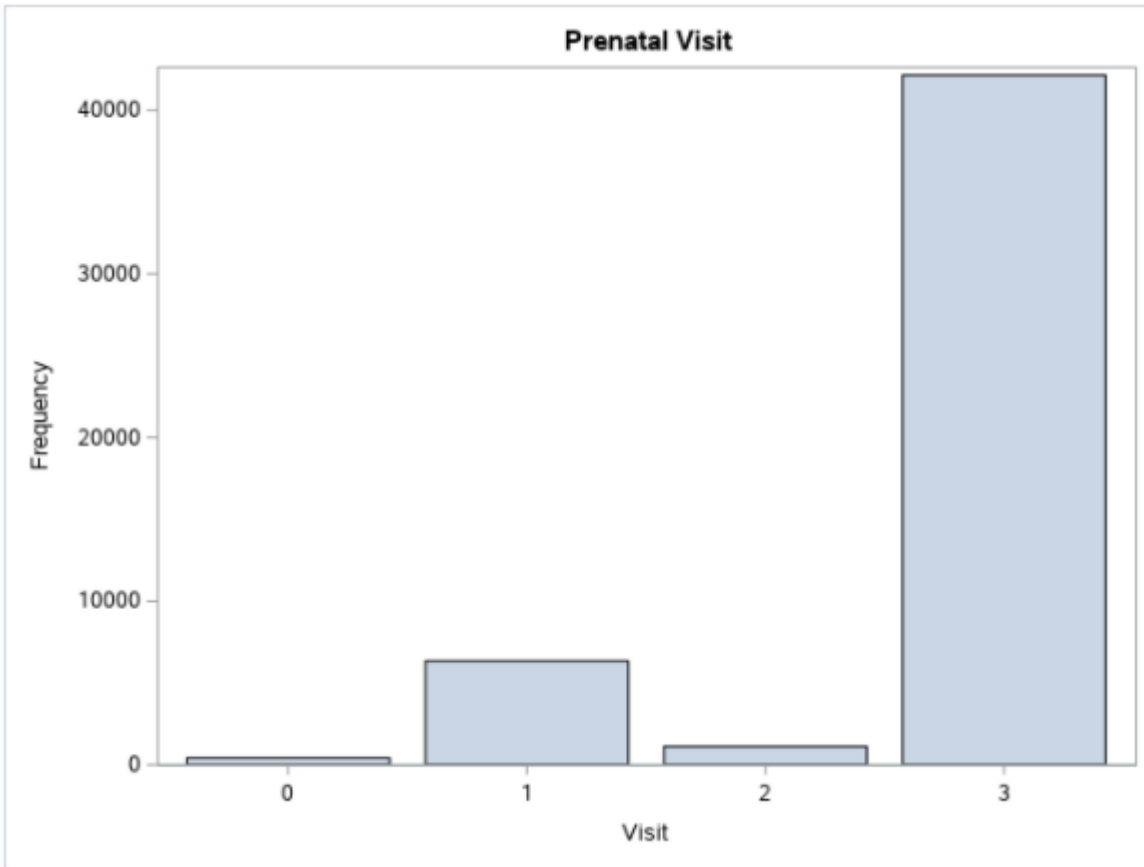
- 42144 Mothers took 3 Prenatal Visits out of 50000 observations.
- We can see that majority of mothers take 3 prenatal visit.