Analysis deck

BAN 100 NCC ASSIGNMENT 2
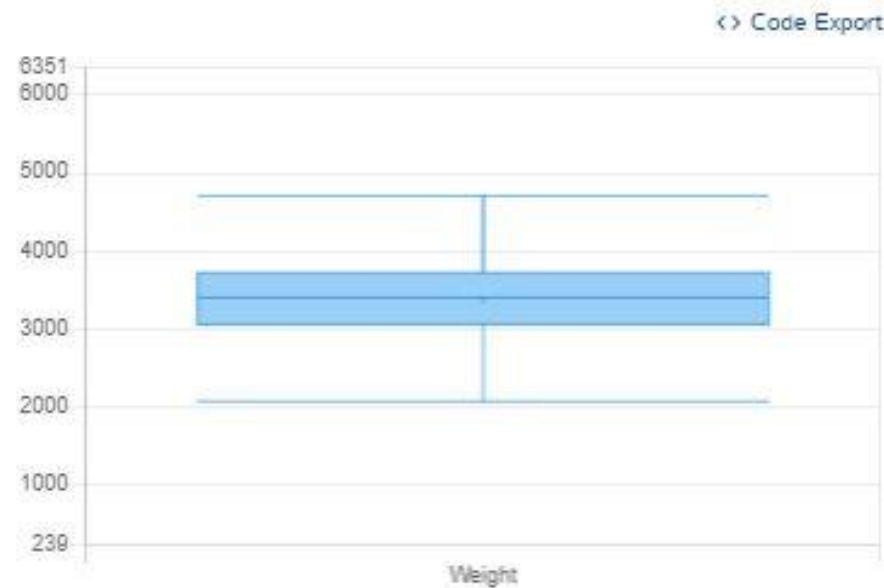
By

Tejaskumar Aghara

Description of Weight Column

Weight (float64)

Describe | Histogram | Categories

<> Code Export

- Total Rows: 50000
  - Count (non-nan): 50,000
  - Count (missing): 0
  - % Missing: 0
- 25%: 3,062
- 50%: 3,402
- 75%: 3,720
- kurt: 2.6901
- max: 6,350
- mean: 3,370.7566
- median: 3,402
- min: 240
- mode: 3,402
- sem: 2.533
- skew: -0.7554
- std: 566.3851
- sum: 168,537,832
- unique: 2,430
- var: 320,792.0312

- We are focusing on Weight column for the analysis
- Further slides will show the TTEST for Weight vs other categorical variables.
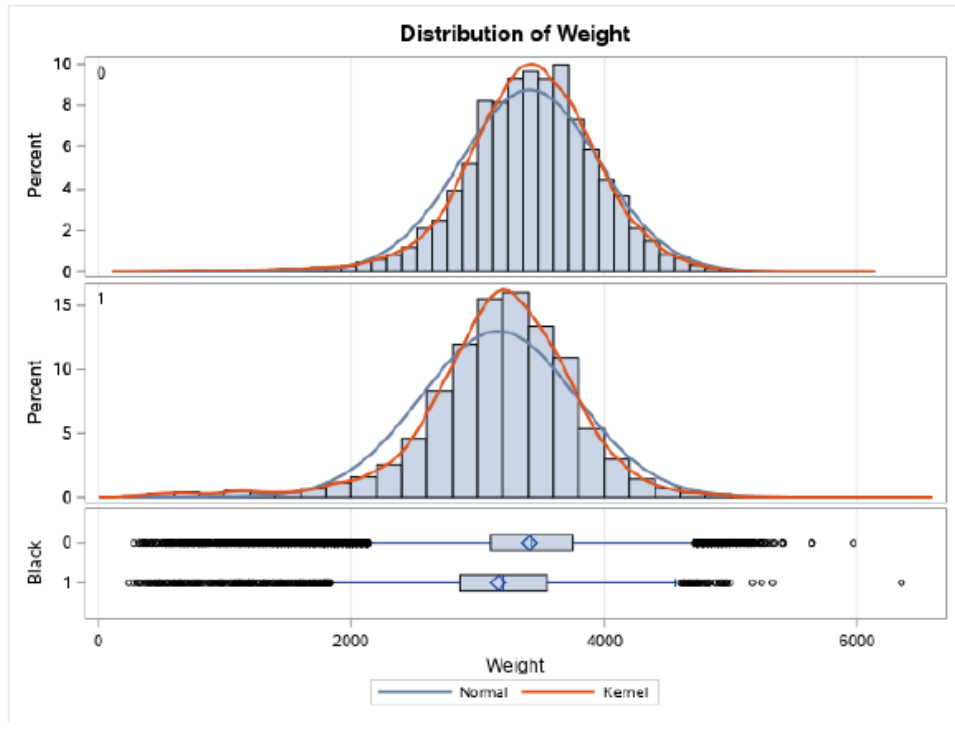
Pearson Correlation of all Variables

- As we can see below, Variable weight is not highly correlated to any other variable

	Weight	Black	Married	Boy	MomAge	MomSmoke	CigsPerDay	MomWtGain	Visit	MomEdLevel
Weight	N/A	-0.16	0.15	0.10	0.10	-0.14	-0.13	0.21	0.07	0.00
Black	-0.16	N/A	-0.36	0.00	-0.12	-0.04	-0.06	-0.05	-0.12	-0.05
Married	0.15	-0.36	N/A	0.01	0.36	-0.18	-0.14	0.00	0.22	0.02
Boy	0.10	0.00	0.01	N/A	0.00	0.00	0.00	0.03	0.00	0.00
MomAge	0.10	-0.12	0.36	0.00	N/A	-0.09	-0.05	-0.06	0.14	0.05
MomSmoke	-0.14	-0.04	-0.18	0.00	-0.09	N/A	0.82	-0.02	-0.09	0.01
CigsPerDay	-0.13	-0.06	-0.14	0.00	-0.05	0.82	N/A	-0.03	-0.08	0.02
MomWtGain	0.21	-0.05	0.00	0.03	-0.06	-0.02	-0.03	N/A	0.05	-0.03
Visit	0.07	-0.12	0.22	0.00	0.14	-0.09	-0.08	0.05	N/A	-0.04
MomEdLevel	0.00	-0.05	0.02	0.00	0.05	0.01	0.02	-0.03	-0.04	N/A

TTEST: Weight vs Black

- As we can see in the table P is 0.0001
- We are rejecting the null Hypothesis
- Means are different for both the classes

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

Distribution of Weight
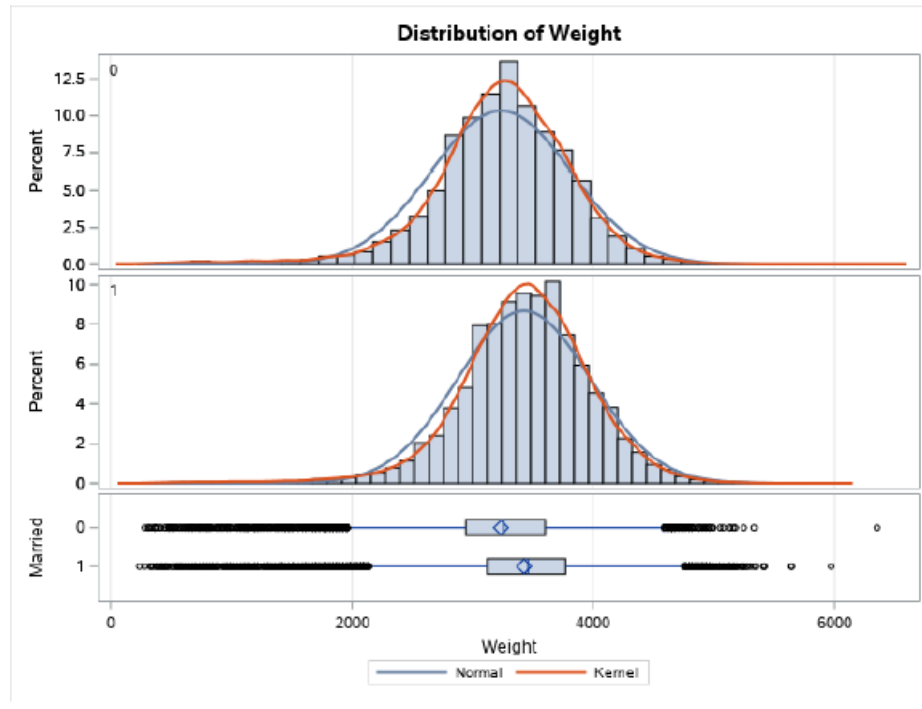
TTEST: Weight vs Married

- As we can see P <= 0.05 We are rejecting the null Hypothesis
- Means are different for both the classes married and non married

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

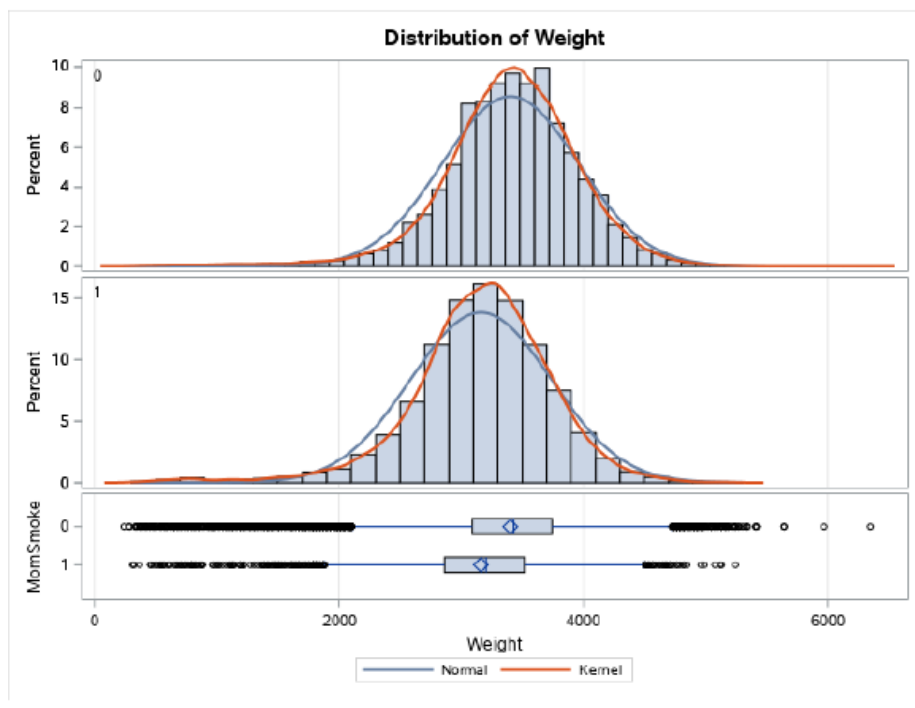| Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	14368	35630	1.10	<.0001

Distribution of Weight

TTEST: Weight vs Smoke

- As we can see in the table P <= 0.05
- We are rejecting the null Hypothesis
- Means are different for both the classes Mom Smoke =1 or =0

MomSmoke	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
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

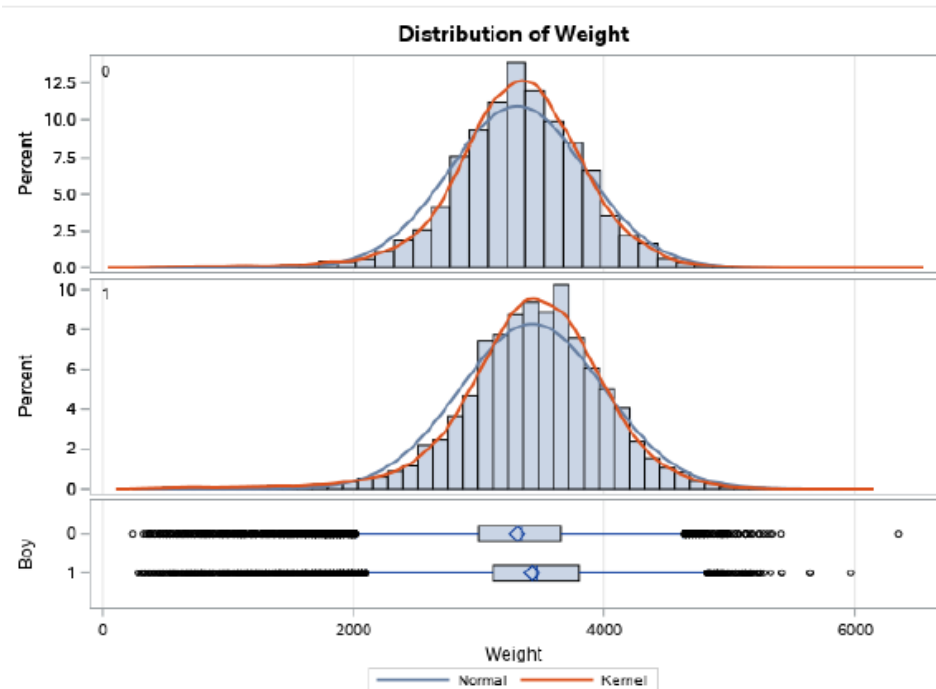| Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	6532	43466	1.07	0.0004

Distribution of Weight

TTEST: Weight vs Boy or Girl

- As we can see in the table P <= 0.05
- We are rejecting the null Hypothesis
- Means are different over the boy and girl class

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

Distribution of Weight

Scatter Plot: Weight vs Mom Weight Gain

One can observe from bellow stated scatter plot that:
- Less observation below mean for the Mom who gained weight
- More observations below mean for the mom who lost weight or gained less weight.

MomWtGain by Weight