

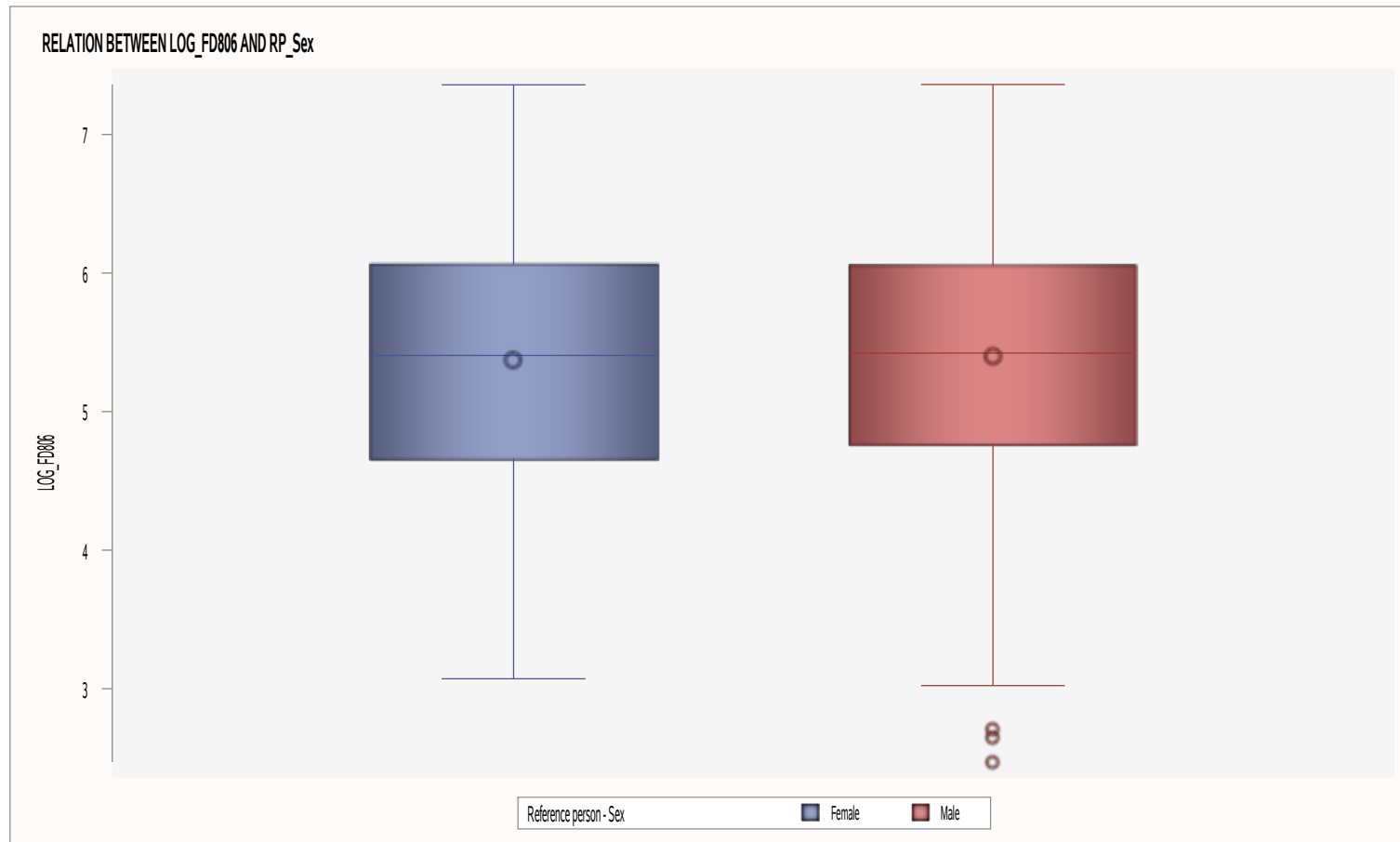
BIVARIATE ANALYSIS OF RP_Sex AND LOG_FD806 FOR ANA.MODEL2

RELATION BETWEEN LOG_FD806 AND RP_Sex

14:29 Sunday, November 21, 2021 1

The MEANS Procedure

Analysis Variable : LOG_FD806														
Reference person - Sex	N Obs	N	N Miss	Minimum	Lower Quartile	Median	Mean	Upper Quartile	Maximum	Quartile Range	Coeff of Variation	Lower 95% CL for Mean	Upper 95% CL for Mean	Skewness
Female	4262186	4262186	0	3.07	4.65	5.41	5.37	6.06	7.36	1.41	17.81	5.37	5.37	-0.08
Male	3866690	3866690	0	2.47	4.76	5.42	5.40	6.05	7.36	1.30	17.14	5.40	5.40	-0.23



T-test - Road map:

14:29 Sunday, November 21, 2021 2

Null hypothesis: There's no difference in means
Assumptions:
1.Sample distribution must be normal:
e.g:Shapiro (null hypothesis: sample has a normal distribution)
CLT :
a.If it looks normal and each group have more than 30 observations
b.If moderately skewed, each group must have more than 100 observations
2.Groups are independent of one another.
3.There are no major outliers.
4.A check for unequal variances will help determine which version of an independent samples t-test is most appropriate:
(e.g:Levene's test, null hypothesis: equal variances)
a.If variances are equal, then a pooled t-test is appropriate
b.If variances are unequal, then a Satterthwaite (also known as Welch's) t-test is appropriate

Normal Distribution?
Null hypothesis: sample has a normal distribution
CLT :
a.If it looks normal and each group have more than 30 observations
b.If moderately skewed, each group must have more than 100 observations
*rule of thumb: If skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed.
*if the sample size is over 2000, the Kolmogorov test should be used. If the sample size is less than 2000, the Shapiro test is better.

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Sex=Female

Moments			
N	4262186	Sum Weights	4262186
Mean	5.36600208	Sum Observations	22870898.9
Std Deviation	0.95566873	Variance	0.91330272
Skewness	-0.0842012	Kurtosis	-0.7315702
Uncorrected SS	126617956	Corrected SS	3892665.17
Coeff Variation	17.8096974	Std Error Mean	0.0004629

Basic Statistical Measures			
Location		Variability	
Mean	5.366002	Std Deviation	0.95567
Median	5.405197	Variance	0.91330
Mode	6.488141	Range	4.28666
		Interquartile Range	1.40972

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	11592.03	Pr > t 	<.0001
Sign	M	2131093	Pr >= M 	<.0001
Signed Rank	S	4.542E12	Pr >= S 	<.0001

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Sex=Female

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.040905	Pr > D	<0.0100
Cramer-von Mises	W-Sq	1488.016	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	11385.81	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.35843
99%	7.28909
95%	6.90938
90%	6.65368
75% Q3	6.06101
50% Median	5.40520
25% Q1	4.65129
10%	3.99046
5%	3.84289
1%	3.32215
0% Min	3.07177

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Sex=Female

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
3.07177	840	982	7.33995	1345	215
3.07177	347	748	7.34003	776	641
3.08374	1372	138	7.34357	669	63
3.08374	3250	56	7.35361	491	703
3.10727	381	1177	7.35843	4907	405

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Sex=Male

Moments			
N	3866690	Sum Weights	3866690
Mean	5.40003416	Sum Observations	20880258.1
Std Deviation	0.92565149	Variance	0.85683067
Skewness	-0.2322971	Kurtosis	-0.3289953
Uncorrected SS	116067205	Corrected SS	3313097.73
Coeff Variation	17.1415857	Std Error Mean	0.00047074

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Sex=Male

Basic Statistical Measures			
Location		Variability	
Mean	5.400034	Std Deviation	0.92565
Median	5.422392	Variance	0.85683
Mode	5.958166	Range	4.89010
		Interquartile Range	1.29722

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	11471.46	Pr > t	<.0001
Sign	M	1933345	Pr >= M	<.0001
Signed Rank	S	3.738E12	Pr >= S	<.0001

Tests for Normality				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.03126	Pr > D	<0.0100
Cramer-von Mises	W-Sq	658.6328	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	5524.586	Pr > A-Sq	<0.0050

Quantiles (Definition 5)	
Level	Quantile
100% Max	7.36074
99%	7.17072
95%	6.89278
90%	6.60801

The UNIVARIATE Procedure
Variable: LOG_FD806

Freq: WeightD

Reference person - Sex=Male

Quantiles (Definition 5)	
Level	Quantile
75% Q3	6.05270
50% Median	5.42239
25% Q1	4.75548
10%	4.18510
5%	3.92691
1%	3.05777
0% Min	2.47064

Extreme Observations					
Lowest			Highest		
Value	Freq	Obs	Value	Freq	Obs
2.47064	3940	1948	7.32095	5649	1394
2.64476	15313	2157	7.33580	1566	1654
2.71337	1520	2184	7.35014	1360	2000
3.02237	4062	2091	7.35109	759	2302
3.03495	8060	2250	7.36074	43	1742

Null hypothesis: equal variances

a.If variances are equal, then a pooled t-test is appropriate

b.If variances are unequal, then a Satterthwaite (also known as Welch's) test is appropriate

The GLM Procedure

Class Level Information		
Class	Levels	Values
RP_Sex	2	Female Male

Number of Observations Read	2327
Number of Observations Used	2327
Sum of Frequencies Read	8128876
Sum of Frequencies Used	8128876

The GLM Procedure**Dependent Variable: LOG_FD806****Frequency: WeightD**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2348.109	2348.109	2648.92	<.0001
Error	8.13E6	7205762.905	0.886		
Corrected Total	8.13E6	7208111.014			

R-Square	Coeff Var	Root MSE	LOG_FD806 Mean
0.000326	17.49306	0.941510	5.382190

Source	DF	Type I SS	Mean Square	F Value	Pr > F
RP_Sex	1	2348.108969	2348.108969	2648.92	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
RP_Sex	1	2348.108969	2348.108969	2648.92	<.0001

The GLM Procedure

Levene's Test for Homogeneity of LOG_FD806 Variance ANOVA of Absolute Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
RP_Sex	1	2904.0	2904.0	10067.5	<.0001
Error	8.13E6	2344748	0.2884		

Welch's ANOVA for LOG_FD806			
Source	DF	F Value	Pr > F
RP_Sex	1.0000	2657.15	<.0001
Error	8094140		

The GLM Procedure

Level of RP_Sex	N	LOG_FD806	
		Mean	Std Dev
Female	4262186	5.36600208	0.95566873
Male	3866690	5.40003416	0.92565149

The TTEST Procedure

Variable: LOG_FD806

Frequency: WeightD

RP_Sex	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
Female		4262186	5.3660	0.9557	0.000463	3.0718	7.3584
Male		3866690	5.4000	0.9257	0.000471	2.4706	7.3607
Diff (1-2)	Pooled		-0.0340	0.9415	0.000661		
Diff (1-2)	Satterthwaite		-0.0340		0.000660		

The TTEST Procedure

Variable: LOG_FD806

RP_Sex	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
Female		5.3660	5.3651	5.3669	0.9557	0.9550	0.9563
Male		5.4000	5.3991	5.4010	0.9257	0.9250	0.9263
Diff (1-2)	Pooled	-0.0340	-0.0353	-0.0327	0.9415	0.9411	0.9420
Diff (1-2)	Satterthwaite	-0.0340	-0.0353	-0.0327			

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	8.13E6	-51.47	<.0001
Satterthwaite	Unequal	8.09E6	-51.55	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	4.26E6	3.87E6	1.07	<.0001

The TTEST Procedure

Variable: LOG_FD806

