Solution

Is there a significant difference between Canadian men and women in their feelings towards immigrants? The 2015 CES asked respondents both their sex (sex_r), and how they feel about immigrants (p_like_immg).

i). What survey questions were asked to capture these variables (refer to the Canadian Election Study 2015 Codebook (available on avenue))? How were answers to the survey questions coded (i.e. what values can a case score)?

According to the survey codebook. The questions asked were :

sex_r : Are you?

 – (1) Male – (5) Female

The response to variable sex_r could be male or female.

– p_like_ How do you feel about the following countries and groups?

(Web version: Slide the slider to any number from 0 to 100.) Zero means you really DISLIKE the country or group, and 100 means you really LIKE the country or group.

The response to variable p_like_immig could be between 0-100.

ii). Generate a frequency table **(2 marks)**, descriptive statistics **(2 marks)**, and a histogram **(2 marks)** for the variable named "p_like_immg". Describe the output. Are there any scores that look like extreme outliers? If so, how have they affected the distribution of the variable **(4 marks)**?

Descriptives

			Statistic	Std. Error
p_like_immg	Mean		90.34	2.427
	95% Confidence Interval for Mean	Lower Bound	85.58	
		Upper Bound	95.10	
	5% Trimmed Mean		65.98	
	Median		69.00	
	Variance		25082.094	

Std. Deviation	158.373	
Minimum	0	
Maximum	1000	
Range	1000	
Interquartile Range	35	
Skewness	5.412	.038
Kurtosis	28.243	.075

The descriptive statistics of the variable "p_like_immg" were calculated. The mean response of the respondents was 90.34 with a standard deviation of 158.373. The values ranged from 0 to 1000 and variance was found to be 25082.094.

Extreme Values

			Case Number	Value
p_like_immg	Highest	1	37	1000
		2	284	1000
		3	403	1000
		4	404	1000
		5	420	1000[a]
	Lowest	1	7497	0
		2	7452	0
		3	7395	0
		4	7386	0
		5	7362	0[b]

a. Only a partial list of cases with the value 1000 are shown in the table of upper extremes.
b. Only a partial list of cases with the value 0 are shown in the table of lower extremes.

Extreme value analysis showed that upper extreme values are 1000 and lower extreme value was 0. Since, the possible values should not exceed 100, the value of 1000 is considered an outlier.
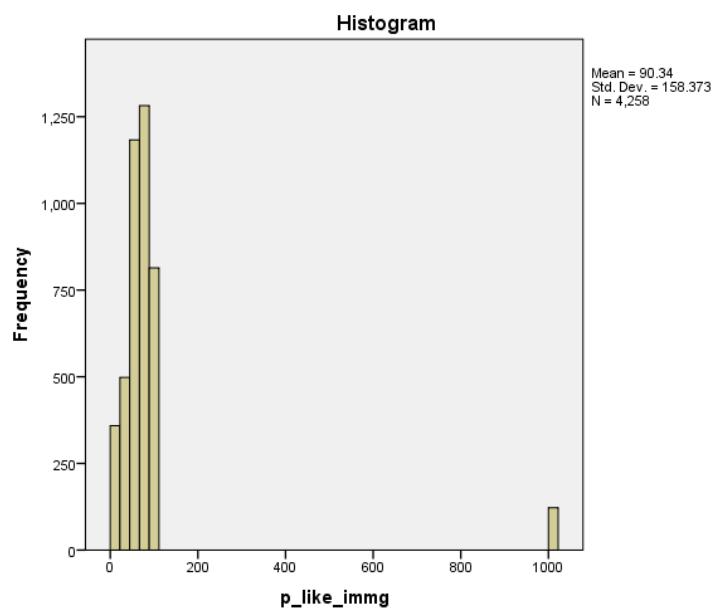
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.

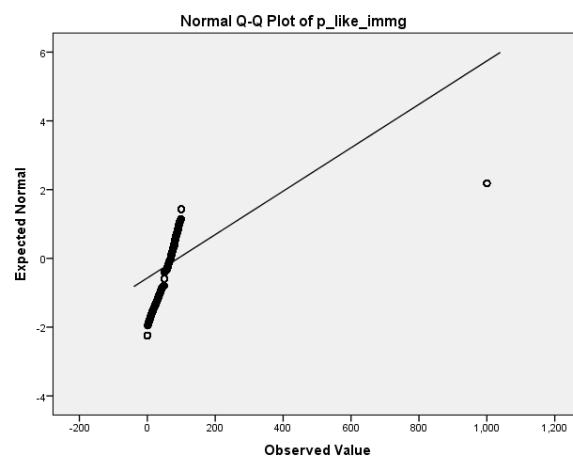| p_like_immg	.447	4258	.000	.286	4258	.000

a. Lilliefors Significance Correction

The p-value of shapiro-wilk test is less than 0.05, hence we can conclude that data is not normally distributed.

p_like_immg

The histogram shows that data is right skewed with outliers at 1000. The data is concentrated to the right of mean which makes it negatively biased.

The normal Q-Q plot reveals that data is indeed right skewed with outliers at 1000.
iii). Enter the Variable View tab. Locate "p_like_immg" in the "Name" column. Click on the "Missing" cell. Click the "Discrete missing values" button. Enter the value of any extreme outlying scores in the boxes. Click OK. Re-generate a frequency table, descriptive statistics, and a histogram for the variable "p_like_immg". Describe the output. **How has it changed?**

The missing values were set to 750, 650 and 1000 and descriptive statistics re calculated as follows:

Descriptives

			Statistic	Std. Error
p_like_immg	Mean		90.34	2.427
	95% Confidence Interval for Mean	Lower Bound	85.58	
		Upper Bound	95.10	
	5% Trimmed Mean		65.98	
	Median		69.00	
	Variance		25082.094	
	Std. Deviation		158.373	
	Minimum		0	
	Maximum		1000	
	Range		1000	
	Interquartile Range		35	
	Skewness		5.412	.038
	Kurtosis		28.243	.075

The descriptive statistics of the variable "p_like_immg" were calculated. The mean response of the respondents was 90.34 with a standard deviation of 158.373. The values ranged from 0 to 1000 and variance was found to be 25082.094.

Extreme Values

			Case Number	Value
p_like_immg	Highest	1	37	1000

	2	284	1000
	3	403	1000
	4	404	1000
	5	420	1000[a]
Lowest	1	7497	0
	2	7452	0
	3	7395	0
	4	7386	0
	5	7362	0[b]

a. Only a partial list of cases with the value 1000 are shown in the table of upper extremes.
b. Only a partial list of cases with the value 0 are shown in the table of lower extremes.

Extreme value analysis showed that upper extreme values are 1000 and lower extreme value was 0. Since, the possible values should not exceed 100, the value of 1000 is considered an outlier.
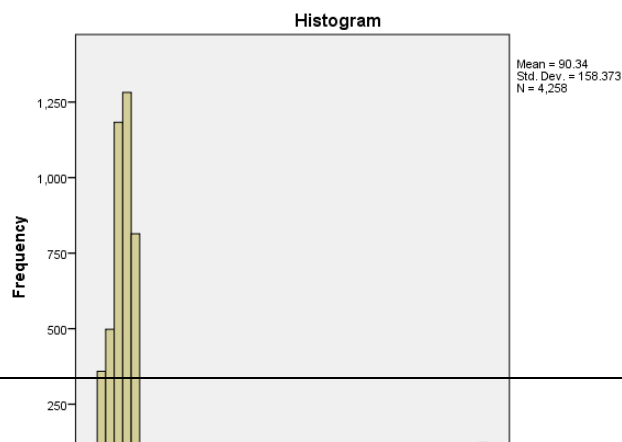
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
p_like_immg	.447	4258	.000	.286	4258	.000

a. Lilliefors Significance Correction

The p-value of shapiro-wilk test is less than 0.05, hence we can conclude that data is not normally distributed.

p_like_immg

Histogram

Mean = 90.34
Std. Dev. = 158.373
N = 4,258

No changes were seen in the histogram or descriptive statistics after adding missing values.

iv). Compare men and women on the variable "p_like_immg" using an independent samples t-test.

T-Test

Group Statistics

	Are you? [This will Auto-Advance on choice]	N	Mean	Std. Deviation	Std. Error Mean
p_like_immg	Male	2082	93.92	172.183	3.774
	Female	2164	87.08	144.238	3.101

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
p_like_immg	Equal variances assumed	11.823	.001	1.404	4244	.160	6.832	4.868	-2.711	16.375

Equal variances not assumed			1.399	4059.338	.162	6.832	4.884	-2.743	16.408

An independent t-test was performed to test the difference in gender wise response to the question : How do you feel about the word immigrant.
The t-test reveals that there is no statistically significant difference between male 3.77 ± 172.18) and female (3.101 ± 144.238) response to the question, $t(1.404) = 4244$, $p = 0.16$. Therefore, we can conclude that men and female think in similar way regarding the word immigrant.

v). Select a second variable on which to compare men and women (select a dependent variable that is either interval-ratio, or ordinal (with at least 4 possible scores)) **(2 marks for picking an appropriate dependent variable)**. Follow the same procedure as in the first example (i.e. repeat steps i-iv as necessary).

The dependent variable chosen is : ptfeel_cons

The question associated with this variable is

How do you feel about the Conservative Party?

 0 = Really Dislike, 100= Really l

Descriptives

			Statistic	Std. Error
How do you feel about the Conservative Party? 0 = Really Dislike, 100= Really l	Mean		160.97	3.700
	95% Confidence Interval for Mean	Lower Bound	153.72	
		Upper Bound	168.22	
	5% Trimmed Mean		123.37	
	Median		50.00	
	Variance		103471.624	
	Std. Deviation		321.670	

Minimum	0	
Maximum	1000	
Range	1000	
Interquartile Range	67	
Skewness	2.195	.028
Kurtosis	2.891	.056

The descriptive statistics of the variable "p_like_immg" were calculated. The mean response of the respondents was 160.97 with a standard deviation of 321.670. The values ranged from 0 to 1000 and variance was found to be 103471.624.

Extreme Values

			Case Number	Value
How do you feel about the Conservative Party? 0 = Really Dislike, 100= Really l	Highest	1	15	1000
		2	21	1000
		3	30	1000
		4	76	1000
		5	159	1000[a]
	Lowest	1	7520	0
		2	7512	0
		3	7498	0
		4	7479	0
		5	7471	0[b]

a. Only a partial list of cases with the value 1000 are shown in the table of upper extremes.
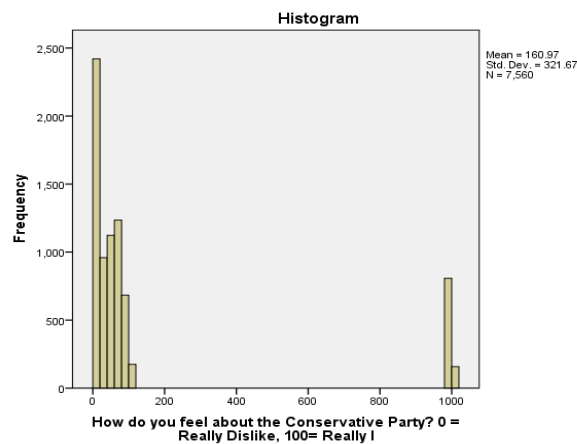b. Only a partial list of cases with the value 0 are shown in the table of lower extremes.

Extreme value analysis reveals an outlier at 1000.

Tests of Normality

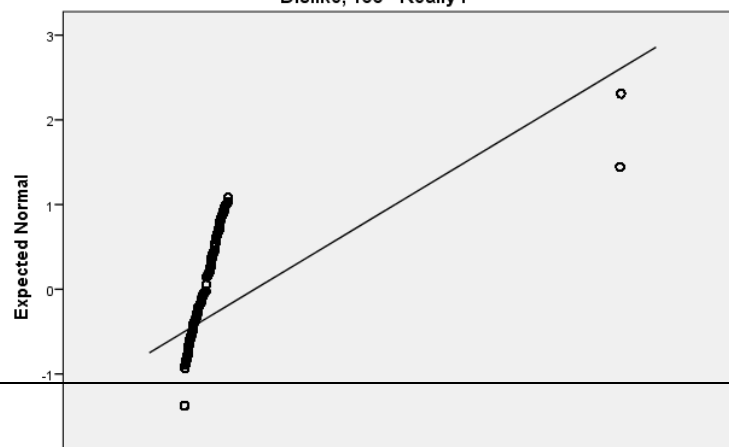|	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
How do you feel about the Conservative Party? 0 = Really Dislike, 100= Really I	.448	7560	.000

a. Lilliefors Significance Correction

The p-value of test for normality is less than 0.05 hence, the data is not normally distributed.
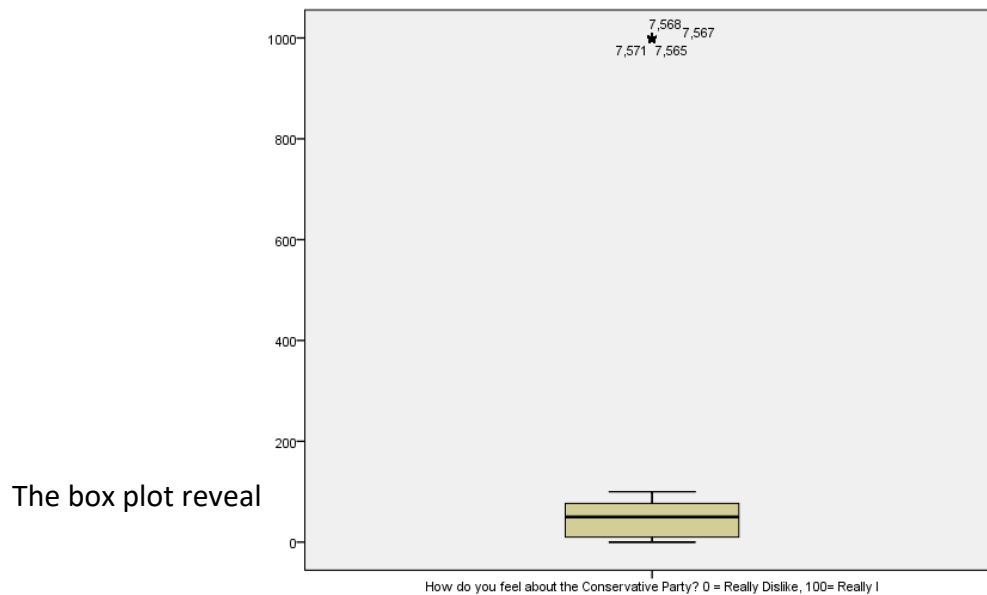

Histogram

The histogram reveals that data is not normally distributed and contains outliers at 1000. More values are concentrated below the mean making the data left skewed with positive bias.

Normal Q-Q Plot of How do you feel about the Conservative Party? 0 = Really Dislike, 100= Really I

Normal QQ plot reveals that data is left skewed with outliers at a score of 1000

The box plot reveal

How do you feel about the Conservative Party? 0 = Really Dislike, 100= Really I

T-Test

Group Statistics

	Are you? [This will Auto-Advance on choice]	N	Mean	Std. Deviation	Std. Error Mean
How do you feel about the Conservative Party? 0 = Really Dislike, 100= Really I	Male	3635	127.85	277.287	4.599
	Female	3883	186.14	350.075	5.618

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
How do you feel about the Conservative Party? 0 = Really Dislike, 100= Really I	Equal variances assumed	303.718	.000	-7.968	7516	.000	-58.288	7.315	-72.628	-43.948
	Equal variances not assumed			-8.028	7317.831	.000	-58.288	7.260	-72.521	-44.056

An independent t-test was performed to test the difference in gender wise response to the question : How do you feel about the conservative party.

The t-test reveals that there is a statistically significant difference between male (127.85 ± 277.28) and female (186.14 ± 350.057) response to the question, t(303.718, 7516) = -7.968 p = 0.00. Therefore, we can conclude that men and female think differently regarding the conservative party.