STAT 448 HW #2

Tianqi Wu

2018/ 9/19

Problem 1.(a)

Here is the code that produces the table:

```
proc sort data = politics;
 by descending ind_con;
run;

title 'Candidates swinging from other parties' ;
proc print data = politics;
 where swingstate = 1 and party = 'OTH';
run;
```

From the table, "Khoury Anton" with total individual contribution of $30,590.00 turns out to be the candidate with the highest contributions who are in swing state from political parties other than the Democratic and Republican parties. His candidate office state is Florida and his office is from Senate. For the political campaign platform, he said that he is willing to pay additional taxes to support the state and wants to be a independent candidate with no party affiliated. Khoury also supports an open primary system that allows more independent voters to participate. Finally, he supports Donald Trump to lead the nation since the president is a successful business man.

Candidates swinging from other parties

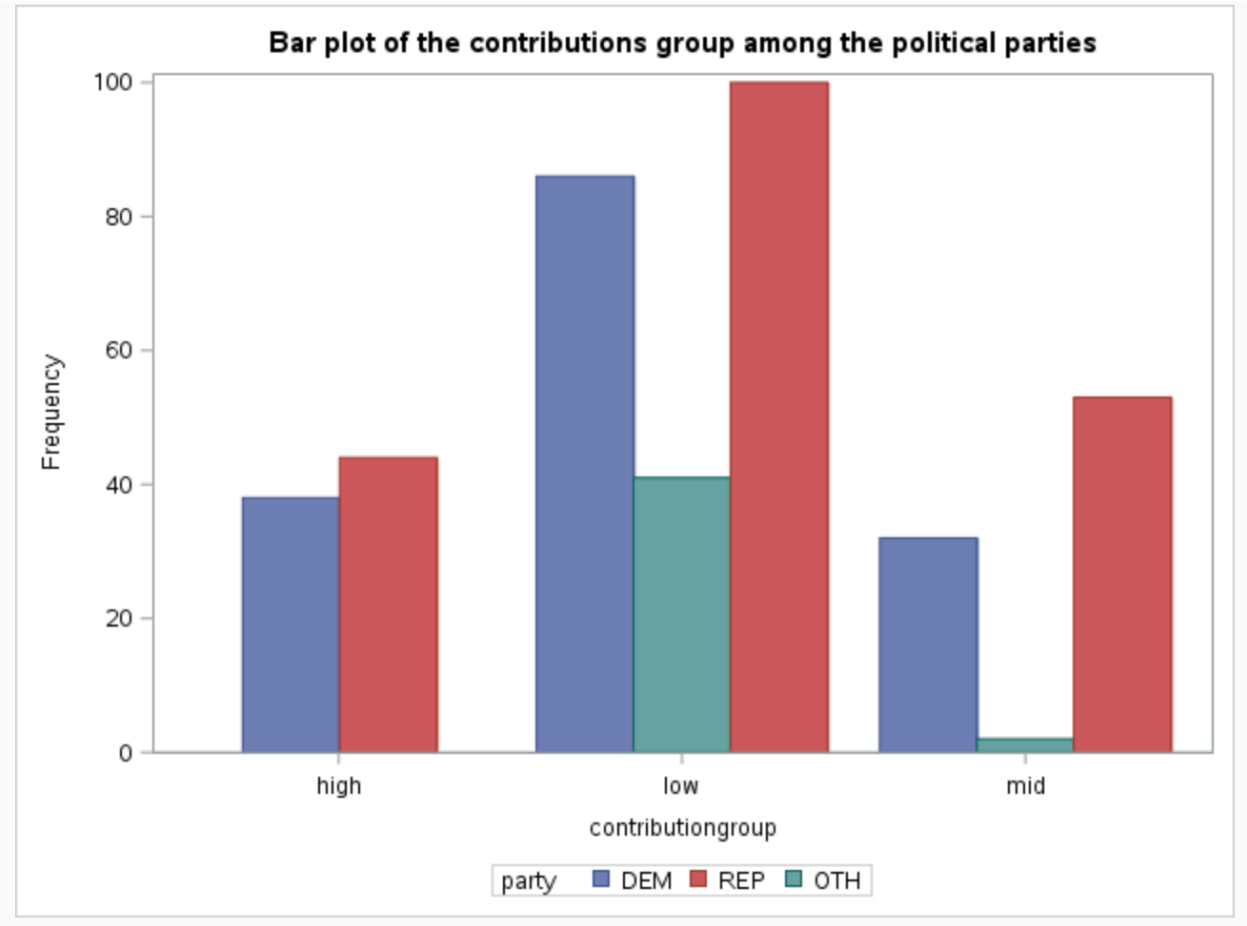
Obs	can_off	can_nam	can_off_sta	can_status	ind_con	party	swingstate	contributiongroup
249	S	KHOURY, ANTON	FL	C	$30,590.00	OTH	1	low
251	S	GUMINA, TONY DR	NV	O	$29,907.00	OTH	1	low
270	S	ANDERSON, PHILLIP NORMAN	WI	C	$19,301.00	OTH	1	low
276	S	WILLIAMS, LILY TANG	CO	C	$17,611.00	OTH	1	low
278	S	STANTON, PAUL ANTHONY	FL	C	$15,769.00	OTH	1	low
280	H	SUMMERELL, JOSEPH JOHN MR.	NC	C	$15,571.00	OTH	1	low
288	S	MENCONI, ARNOLD MICHAEL	CO	C	$13,328.00	OTH	1	low
305	S	INVICTUS, AUGUSTUS SOL	FL	C	$9,094.00	OTH	1	low
320	S	DEMARE, JOSEPH ROSARIO	OH	C	$5,996.00	OTH	1	low
343	S	NATHAN, BRUCE	FL	C	$2,971.00	OTH	1	low
350	H	CRAIG, ANDY	WI	C	$1,910.00	OTH	1	low
353	S	LUICK-THRAMS, ALIZA MICHAEL DR	IA	C	$1,696.00	OTH	1	low
360	S	CONNORS, THOMAS WILLIAM	OH	C	$1,050.55	OTH	1	low
365	S	WILLIAMS, JARROD M.	NV	O	$901.00	OTH	1	low
366	S	HUDOCK, BRANDON GENE	PA	C	$835.00	OTH	1	low
373	S	STERN, EVERETT ALEXANDER	PA	C	$500.00	OTH	1	low
380	S	GUTHRIE, SEAN PATRICK	FL	C	$235.00	OTH	1	low
381	S	JONES, THOMAS FRANKLIN	NV	O	$215.00	OTH	1	low
394	H	SCHREY, ELIZABETH ANNE	FL	O	$15.00	OTH	1	low

Problem 1.(b)

Here is the code that produces the bar plot of the contributions group among the political parties:

```
title 'Bar plot of the contributions group among the
political parties';
proc sgplot data = politics;
 vbar contributiongroup / group = party
groupdisplay=cluster;
run;
```

From the bar plot, nearly 230 of 396 candidates are from low contribution group. Among the contribution groups, there are more candidates from Republican than Democratic and other political parties. There is no candidate from other political parties classifying as high contribution group.

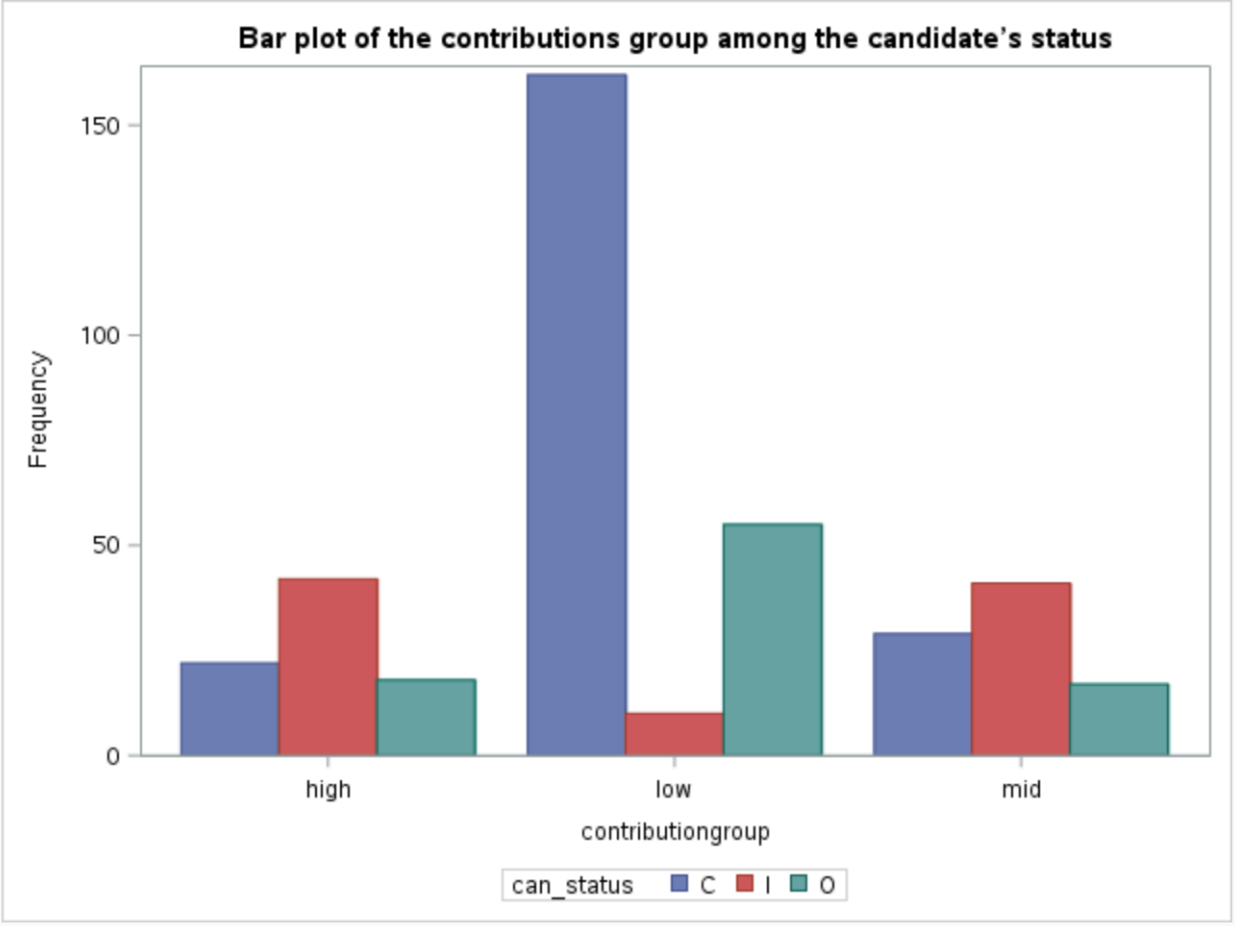


Problem 1.(c)

Here is the code that produces the bar plot of the contributions group among the candidate's status:

```
title 'Bar plot of the contributions group among the
candidate's status';
proc sgplot data = politics;
 vbar contributiongroup / group = can_status
groupdisplay=cluster;
run;
```

From the bar plot, nearly 220 of 396 candidates are from low contribution group and 160 of them are challenger. There are approximately same number of candidates from high and mid contribution group and most of them are incumbent.



Problem 1.(d)

Here is the code that constructs the contingency table with the swing states and political office variables:

```
proc freq data = politics;
 tables swingstate*can_off / expected riskdiff;
run;
```

From the risk difference test, for both column 1 and column 2 risk estimates, the confidence interval includes 0. It indicates that there is **no** strong evidence showing that there is a difference between the probability of Senatorial and House candidates going to swing states.

The FREQ Procedure

Frequency Expected Percent Row Pct Col Pct	Table of swingstate by can_off		
	can_off		
swingstate	H	S	Total
0	127 128.5 32.07 49.42 64.14	130 128.5 32.83 50.58 65.66	257 64.90
1	71 69.5 17.93 51.08 35.86	68 69.5 17.17 48.92 34.34	139 35.10
Total	198 50.00	198 50.00	396 100.00

(problem d-f shares the table above)

Statistics for Table of swingstate by can_off

Column 1 Risk Estimates

	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.4942	0.0312	0.4330	0.5553	0.4315	0.5570
Row 2	0.5108	0.0424	0.4277	0.5939	0.4247	0.5965
Total	0.5000	0.0251	0.4508	0.5492	0.4497	0.5503
Difference	-0.0166	0.0526	-0.1198	0.0865		

Difference is (Row 1 - Row 2)

Column 2 Risk Estimates

	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.5058	0.0312	0.4447	0.5670	0.4430	0.5685
Row 2	0.4892	0.0424	0.4061	0.5723	0.4035	0.5753
Total	0.5000	0.0251	0.4508	0.5492	0.4497	0.5503
Difference	0.0166	0.0526	-0.0865	0.1198		

Difference is (Row 1 - Row 2)

Sample Size = 396

Problem 1.(e)

Here is the code that constructs the contingency table with the swing states and political office variables:

```
proc freq data = politics;
 tables swingstate*can_off / expected or(cl=wald);
run;
```

From the odds ratio test, the odds ratio is 0.9356. It means that the odds of House candidates having swing state = 0 is 0.9356 times the odds of them having swing state = 1. The confidence interval includes 1 and it indicates that the variables can_off and swing state are **independent**.

Statistics for Table of swingstate by can_off

Confidence Limits for the Odds Ratio		
Odds Ratio = 0.9356		
Type	95% Confidence Limits	
Wald	0.6192	1.4138

Sample Size = 396

Problem 1.(f)

Here is the code that constructs the contingency table with the swing states and political office variables:

```
proc freq data = politics;
 tables swingstate*can_off / expected chisq;
run;
```

From the table, the chi-square value is 0.0998(small) with probability 0.7521(>0.05). Hence, there is no strong evidence to reject the null hypothesis that there is no association between the variables political office and swing states. Phi, Contingency and Cramer's V coefficients are -0.0159, 0.0159 and -0.0159. The small coefficients indicate that there is nearly **no** association between the variables.

Statistics for Table of swingstate by can_off

Statistic	DF	Value	Prob
Chi-Square	1	0.0998	0.7521
Likelihood Ratio Chi-Square	1	0.0998	0.7521
Continuity Adj. Chi-Square	1	0.0443	0.8332
Mantel-Haenszel Chi-Square	1	0.0995	0.7524
Phi Coefficient		-0.0159	
Contingency Coefficient		0.0159	
Cramer's V		-0.0159	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	127
Left-sided Pr <= F	0.4166
Right-sided Pr >= F	0.6631
Table Probability (P)	0.0798
Two-sided Pr <= P	0.8333

Sample Size = 396

Problem 1.(g)

Here is the code that constructs a contingency table with the political office and contribution group variables:

```
proc freq data = politics;
 tables can_off*contributiongroup / expected chisq;
run;
```

From the table, the chi-square has value 39.5612 with p-value <.0001. It means the there is strong evidence to reject the null that there is no association between the variables political office and contribution group. Since the table is not 2*2, using the chi-square test is appropriate here. Phi, Contingency and Cramer's V coefficients are 0.3161, 0.3014, 0.3161. It indicates that there is some **weak** association between the variables.

The FREQ Procedure

Frequency
Expected
Percent
Row Pct
Col Pct

Table of can_off by contributiongroup

can_off	contributiongroup: high	low	mid	Total
H	19 41 4.80 9.60 23.17	117 113.5 29.55 59.09 51.54	62 43.5 15.66 31.31 71.26	198 50.00
S	63 41 15.91 31.82 76.83	110 113.5 27.78 55.56 48.46	25 43.5 6.31 12.63 28.74	198 50.00
Total	82 20.71	227 57.32	87 21.97	396 100.00

Statistics for Table of can_off by contributiongroup

Statistic	DF	Value	Prob
Chi-Square	2	39.5612	<.0001
Likelihood Ratio Chi-Square	2	41.3618	<.0001
Mantel-Haenszel Chi-Square	1	38.7389	<.0001
Phi Coefficient		0.3161	
Contingency Coefficient		0.3014	
Cramer's V		0.3161	

Sample Size = 396

Problem 1.(h)

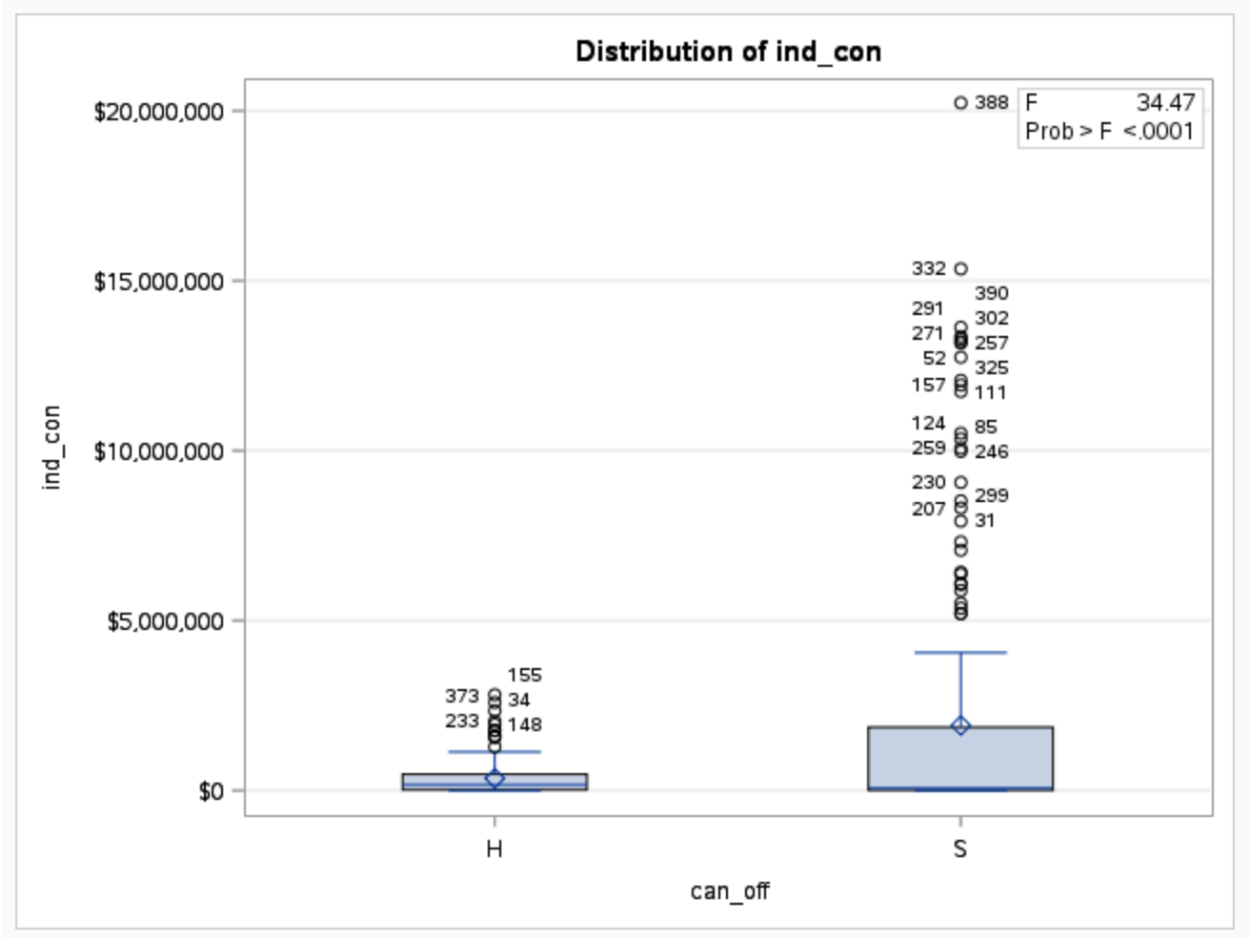
Here is the code that produces a one-way ANOVA for the total individual contributions as a function of the political office:

```
proc anova data = politics;
 class can_off;
 model ind_con = can_off;
 means can_off / hovtest welch;
run;

proc npar1way data = politics;
 class can_off;
 var ind_con;
 ods select KruskalWallisTest;
run;
```

From the box plot given by the parametric one-way anova, the mean and median differ significantly and the distribution is not symmetric with a lot of outliers for the Senate candidates. Also, the assumption of equal variance is violated since the p-value for Levene's and Welch's test is <.0001 which rejects the null of equal variance. Hence, the assumptions of normality and equal variance are violated and the nonparametric one-way anova is used.

From the Kruskal-Wallis Test result, the chi-square has value 0.0009 with p-value 0.9758(>0.05). Hence, there is **no** strong evidence to reject the null hypothesis that the group means for the main effect are equal.



The ANOVA Procedure

Levene's Test for Homogeneity of ind_con Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
can_off	1	1.768E28	1.768E28	25.86	<.0001
Error	394	2.694E29	6.838E26		

Welch's ANOVA for ind_con			
Source	DF	F Value	Pr > F
can_off	1.0000	34.47	<.0001
Error	204.2		

The NPAR1WAY Procedure

Kruskal-Wallis Test	
Chi-Square	0.0009
DF	1
Pr > Chi-Square	0.9758

Problem 1.(i)

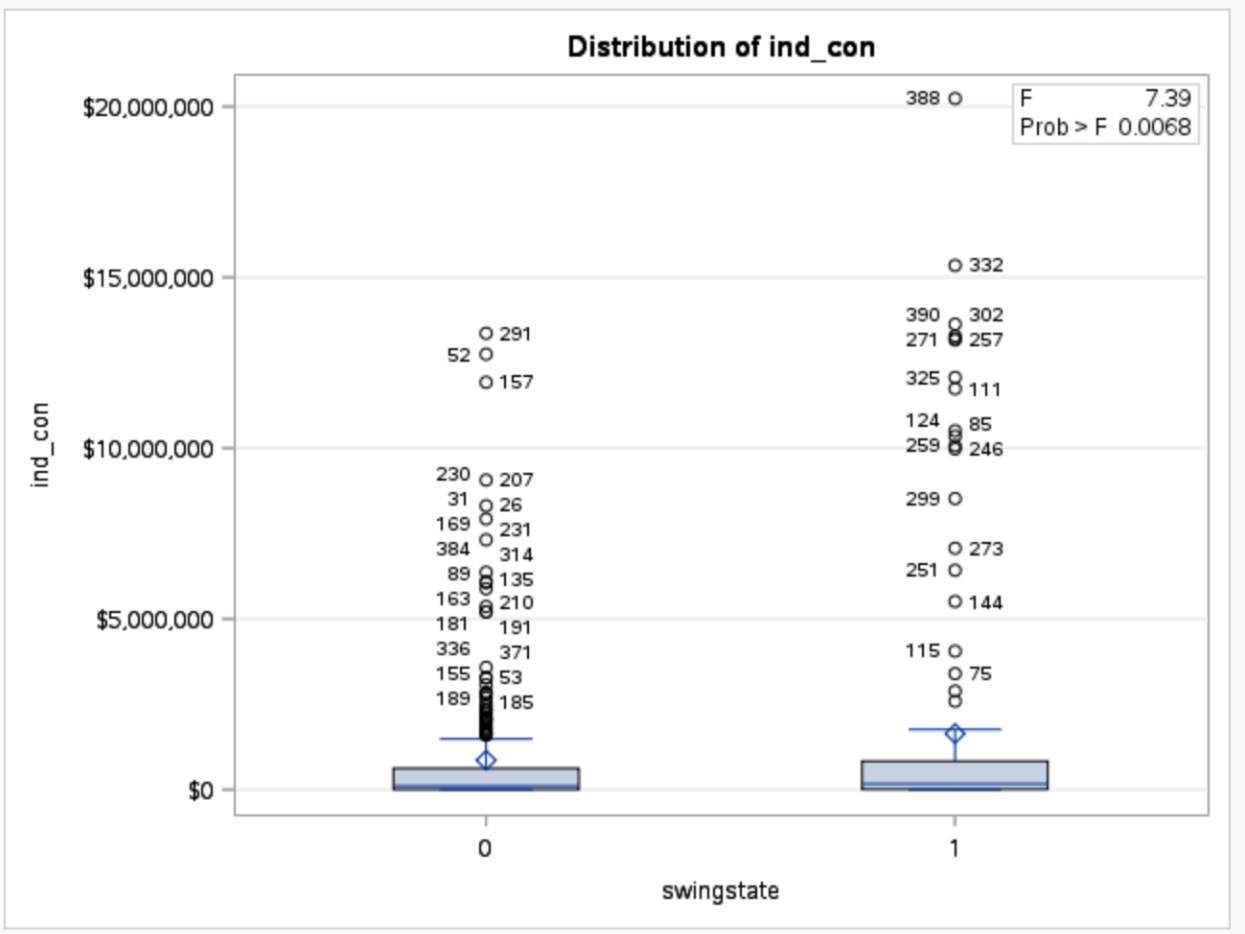
Here is the code that produces a one-way ANOVA for the total individual contributions as a function of swing state:

```
proc anova data = politics;
 class swingstate;
 model ind_con = swingstate;
 means swingstate / hovtest welch;
run;

proc npar1way data = politics;
 class swingstate;
 var ind_con;
 ods select KruskalWallisTest;
run;
```

From the box plot given by the parametric one-way anova, the mean and median differ significantly and the distribution is not symmetric with a lot of outliers for both levels of swing state. Also, the assumption of equal variance is violated since the p-value for Levene's and Welch's test is <0.05 which rejects the null of equal variance. Hence, the assumptions of normality and equal variance are violated and the nonparametric one-way anova is used.

From the Kruskal-Wallis Test result, the chi-square has value .08002 with p-value 0.3710(>0.05). Hence, there is **no** strong evidence to reject the null hypothesis that the group means for the main effect are equal.



The ANOVA Procedure

Levene's Test for Homogeneity of ind_con Variance ANOVA of Squared Deviations from Group Means					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
swingstate	1	9.402E27	9.402E27	11.49	0.0008
Error	394	3.224E29	8.183E26		

Welch's ANOVA for ind_con			
Source	**DF**	**F Value**	**Pr > F**
swingstate	1.0000	5.23	0.0234
Error	179.4		

The NPAR1WAY Procedure

Kruskal-Wallis Test	
Chi-Square	0.8002
DF	1
Pr > Chi-Square	0.3710

Problem 1.(j)

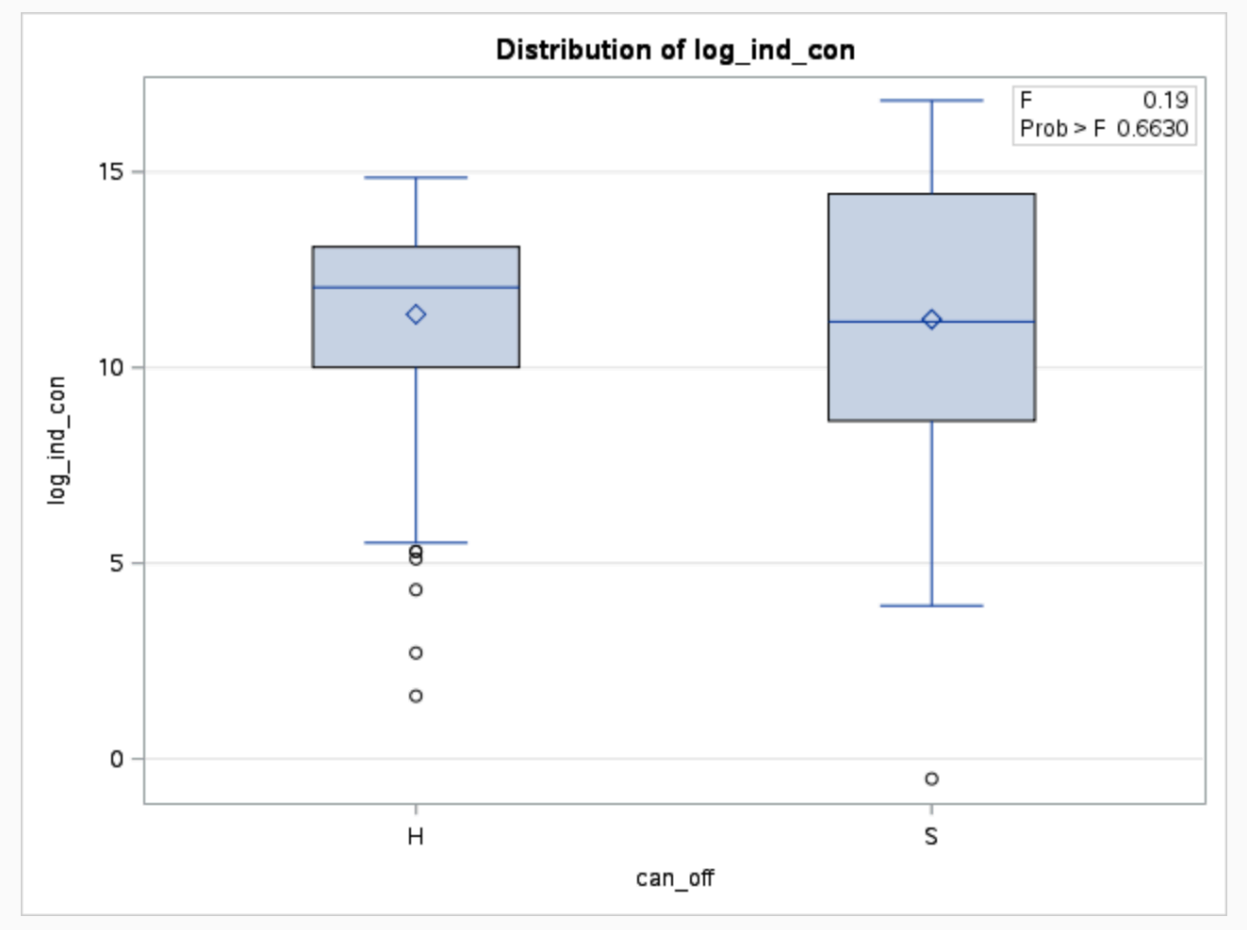
Here is the code that performs log transformation of total individual contribution and one-way anova test:

```
data log_politics;
 set politics;
 log_ind_con = log(ind_con);
run;

proc anova data = log_politics;
 class can_off;
 model log_ind_con = can_off;
 means can_off / hovtest welch;
run;
```

After the log transformation, there is not enough strong evidence to rule the data as not being normal since the mean and median do not differ much and the distribution can be roughly seen as symmetric with no outliers for both levels of candidate office. Since the p-value of Levene's test is <.0001, there is evidence that the equal variance assumption may be violated.

Overall, the normality of the distribution can be assumed but the equal variance assumption may be violated. Hence, using a log transformation of total individual contributions **allows** for a parametric one-way ANOVA with Welch correction to be done for the total individual contributions as a function of the political office.



The ANOVA Procedure

Dependent Variable: log_ind_con

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.713231	1.713231	0.19	0.6630
Error	394	3549.754972	9.009530		
Corrected Total	395	3551.468203			

R-Square	Coeff Var	Root MSE	log_ind_con Mean
0.000482	26.57374	3.001588	11.29532

Source	DF	Anova SS	Mean Square	F Value	Pr > F
can_off	1	1.71323057	1.71323057	0.19	0.6630

The ANOVA Procedure

Levene's Test for Homogeneity of log_ind_con Variance
ANOVA of Squared Deviations from Group Means

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
can_off	1	4869.7	4869.7	30.75	<.0001
Error	394	62385.2	158.3		

Welch's ANOVA for log_ind_con

Source	DF	F Value	Pr > F
can_off	1.0000	0.19	0.6631
Error	341.7		