ADTA 5130 Section 100 – Fall 2023

Mini Project 2 – Group 28

Group Members: Harini Kamarthy

Nisha Bhattarai

Vaishnavi Gunna

Task 0 : Adjusting Presidential Approval Ratings

In each worksheet dedicated to a President, populate the columns named "Approving_adj_[PresidentName]" and "Disapproving_adj_[PresidentName]". Allocate the value from the "Unsure/NoData" column between the "Approving" and "Disapproving" columns in a proportional manner. Ensure that the sum of "Approving_adj_[PresidentName]" and "Disapproving_adj_[PresidentName]" totals 100.

Steps :

- For allocating the value from "Unsure/NoData" column between "Approving" and "Disapproving" columns in a proportional manner, added new columns on each president data sheet as "**Approving_adj_[PresidentName]**" and "**Disapproving_adj_[PresidentName]**", with a formula as below,
- **Approving_adj_[PresidentName] = Approving + (Unsure/NoData * Approving / (Approving + Disapproving))**
- **Disapproving_adj_[PresidentName] = Disapproving + (Unsure/NoData * Disapproving / (Approving + Disapproving))**
- Also, to make sure that both the new columns sum to total of 100, we have added a new column **"100-(Adj_Appr + Adj_Disappr)"** which should result in 0.

Same method is applied on all 4 president sheets named as "Biden", "Trump", "Obama", "GWBush" in the Excel sheet.

Please find the screenshots of the same for each president as below.

Biden

	A	B	C	D	E	F	G	H
1	Start Date	End Date	Approving	Disapproving	Unsure/NoData	Approving_adj_Biden	Disapproving_adj_Biden	100-(Adj_Appr + Adj_Disappr)
2	8/1/2023	8/23/2023	42	53	5	44.21052632	55.78947368	0
3	7/3/2023	7/27/2023	40	55	5	42.10526316	57.89473684	0
4	6/1/2023	6/22/2023	43	54	3	44.32989691	55.67010309	0
5	5/1/2023	5/24/2023	39	57	4	40.625	59.375	0
6	4/3/2023	4/25/2023	37	59	4	38.54166667	61.45833333	0
7	3/1/2023	3/23/2023	40	56	4	41.66666667	58.33333333	0
8	2/1/2023	2/23/2023	42	54	4	43.75	56.25	0
9	1/2/2023	1/22/2023	41	54	5	43.15789474	56.84210526	0
10	11/9/2022	12/2/2022	40	55	5	42.10526316	57.89473684	0
11	10/3/2022	10/20/2022	40	56	4	41.66666667	58.33333333	0
12	9/1/2022	9/16/2022	42	56	2	42.85714286	57.14285714	0
13	8/1/2022	8/23/2022	44	53	3	45.36082474	54.63917526	0
14	7/5/2022	7/26/2022	38	59	3	39.17525773	60.82474227	0
15	6/1/2022	6/20/2022	41	57	2	41.83673469	58.16326531	0
16	5/2/2022	5/22/2022	41	54	5	43.15789474	56.84210526	0

Biden | Trump | Obama | GWBush | PEW | Task 0 | Task 1 - 95% CI | Task 2a -

Trump

	A	B	C	D	E	F	G	H
1	Start Date	End Date	Approving	Disapproving	Unsure/NoData	Approving_adj_Trump	Disapproving_adj_Trump	100-(Adj_Appr + Adj_Disappr)
2	1/4/2021	1/15/2021	34	62	4	35.41666667	64.58333333	0
3	12/1/2020	12/17/2020	39	57	4	40.625	59.375	0
4	11/5/2020	11/19/2020	43	55	2	43.87755102	56.12244898	0
5	10/16/2020	10/27/2020	46	52	2	46.93877551	53.06122449	0
6	9/30/2020	10/15/2020	43	55	2	43.87755102	56.12244898	0
7	9/14/2020	9/28/2020	46	52	2	46.93877551	53.06122449	0
8	8/31/2020	9/13/2020	42	56	2	42.85714286	57.14285714	0
9	7/30/2020	8/12/2020	42	55	3	43.29896907	56.70103093	0
10	7/1/2020	7/23/2020	41	56	3	42.26804124	57.73195876	0
11	6/8/2020	6/30/2020	38	57	5	40	60	0
12	5/28/2020	6/4/2020	39	57	4	40.625	59.375	0
13	5/1/2020	5/13/2020	49	48	3	50.51546392	49.48453608	0
14	4/14/2020	4/28/2020	49	47	4	51.04166667	48.95833333	0
15	4/1/2020	4/14/2020	43	54	3	44.32989691	55.67010309	0
16	3/13/2020	3/22/2020	49	45	6	52.12765957	47.87234043	0

Biden | Trump | Obama | GWBush | PEW | Task 0 | Task 1 - 95% CI | Task 2a -

Obama

	A	B	C	D	E	F	G	H
1	Start Date	End Date	Approving	Disapproving	Unsure/NoData	Approving_adj_Obama	Disapproving_adj_Obama	100-(Adj_Appr + Adj_Disappr)
2	1/16/2017	1/19/2017	59	37	4	61.45833333	38.54166667	0
3	1/9/2017	1/15/2017	57	39	4	59.375	40.625	0
4	1/2/2017	1/8/2017	55	42	3	56.70103093	43.29896907	0
5	12/26/2016	1/1/2017	55	40	5	57.89473684	42.10526316	0
6	12/19/2016	12/25/2016	56	40	4	58.33333333	41.66666667	0
7	12/12/2016	12/18/2016	56	40	4	58.33333333	41.66666667	0
8	12/5/2016	12/11/2016	57	40	3	58.7628866	41.2371134	0
9	11/28/2016	12/4/2016	54	43	3	55.67010309	44.32989691	0
10	11/21/2016	11/27/2016	56	40	4	58.33333333	41.66666667	0
11	11/14/2016	11/20/2016	56	41	3	57.73195876	42.26804124	0
12	11/7/2016	11/13/2016	57	41	2	58.16326531	41.83673469	0
13	10/31/2016	11/6/2016	53	45	2	54.08163265	45.91836735	0
14	10/24/2016	10/30/2016	54	44	2	55.10204082	44.89795918	0

Biden | Trump | Obama | GWBush | PEW | Task 0 | Task 1 - 95% CI | Task 2a -

GWBush

	A	B	C	D	E	F	G	H
1	Start Date	End Date	Approving	Disapproving	Unsure/NoData	Approving_adj_GWBush	Disapproving_adj_GWBush	100-(Adj_Appr + Adj_Disappr)
2	1/9/2009	1/11/2009	34	61	5	35.78947368	64.21052632	0
3	12/12/2008	12/14/2008	29	67	4	30.20833333	69.79166667	0
4	12/4/2008	12/7/2008	32	61	7	34.40860215	65.59139785	0
5	11/13/2008	11/16/2008	29	66	5	30.52631579	69.47368421	0
6	11/7/2008	11/9/2008	28	68	4	29.16666667	70.83333333	0
7	10/31/2008	11/2/2008	25	70	5	26.31578947	73.68421053	0
8	10/23/2008	10/26/2008	31	66	3	31.95876289	68.04123711	0
9	10/10/2008	10/12/2008	25	71	4	26.04166667	73.95833333	0
10	10/3/2008	10/5/2008	25	70	5	26.31578947	73.68421053	0
11	9/26/2008	9/27/2008	27	69	4	28.125	71.875	0
12	9/8/2008	9/11/2008	31	65	4	32.29166667	67.70833333	0
13	9/5/2008	9/7/2008	33	64	3	34.02061856	65.97938144	0
14	8/21/2008	8/23/2008	29	66	5	30.52631579	69.47368421	0
15	8/7/2008	8/10/2008	33	61	6	35.10638298	64.89361702	0
16	7/25/2008	7/27/2008	32	65	3	32.98969072	67.01030928	0

Biden | Trump | Obama | GWBush | PEW | Task 0 | Task 1 - 95% CI | Task 2a - ••• +

Order the Presidents in descending sequence based on the average "Approving_adj" ratings during their initial 1000 days in office along with their average "Approving_adj" values.

Steps:

- Add a new column "**Time Period in Days**" and add the formula "**End Date – Start Date**" to store the information of number of days.
- Add a new column "**Cummulative Time Period from End to Start Date**", where it stores the **cumulative frequency** of the "**Time Period in Days**" column from End to Start Date.
- Add a new column "**Cummulative Time Period from Start to End Date**", where it stores the **cumulative frequency** of the "**Time Period in Days**" column from Start to End Date.
- Repeat the above 3 steps for every sheet "Biden", "Trump", "Obama", "GWBush".
- Find the average of "**Approving_adj_[PresidentName]**" for first Initial 1000 days for each president.
- Note each president average ratings in another worksheet named as "**Task 0**" and sort them in a descending order.
- Result in the descending order obtained is as follows.
  GWBush > Obama > Biden > Trump

	F	G	H	I	J	K	L	M
	Approving_adj_Biden	Disapproving_adj_Biden	100-(Adj_Appr + Adj_Disappr)	Time Period in Days	Cummulative Time Period from 2023 to 2021	Cummulative Time Period from 2021 to 2023		Average Adjusted Approval of Initial 1000 Days
	44.21052632	55.78947368	0	22	22	564		46.26654899
	42.10526316	57.89473684	0	24	46	542		
	44.32989691	55.67010309	0	21	67	518		
	40.625	59.375	0	23	90	497		
	38.54166667	61.45833333	0	22	112	474		
	41.66666667	58.33333333	0	22	134	452		
	43.75	56.25	0	22	156	430		
	43.15789474	56.84210526	0	20	176	408		
	42.10526316	57.89473684	0	23	199	388		
	41.66666667	58.33333333	0	17	216	365		
	42.85714286	57.14285714	0	15	231	348		

	A	B
1	President Name	Adjusted Approval Ratings Average for Initial 1000 Days
2	GWBush	54.04641829
3	Obama	53.63981333
4	Biden	46.26654899
5	Trump	42.33311189
6		

Task 1 : Confidence Interval Estimations

You will use data in "Biden", "Trump", "Obama", and "GWBush" worksheets to complete this task.

a. **Utilizing the given dataset, calculate a 95% confidence interval for the "Approving_adj" rate for each President. Keep in mind that this should be based on the adjusted Approval ratings you computed in Task 0.**

Steps:

- Calculate the Mean, Standard Deviation, Size for "**Approving_adj_[PresidentName]**" for each president on their respective data sheets using the formulas as below.
  Mean = AVERAGE(array)
  Standard Deviation = STDEV(array)
  Size = COUNT(array)

N	O	P
Average of Adjusted Approval for All Days	Standard Deviation	Size
46.26654899	6.500581	31

- Created a new worksheet tab named as "**Task 1 – 95% CI**", and added the info of already calculated Mean, Standard Deviation, Size of each President into it.
- Now calculate Standard Error & Margin of Error with the formulas as below, for each president data,
  Standard Error (SE) = Standard Deviation(σ) / SQRT(Size)
  Margin of Error (ME) = CONFIDENCE.NORM(α, SE, Size), where α = 0.05 as we are calculating Confidence Interval of 95%.
- Now calculate the Confidence Interval,
  CI Lower Limit = Mean – ME
  CI Upper Limit = Mean + ME

	A	B	C	D	E	F	G	H
1	President Name	Mean (μ)	Standard Deviation (σ)	Size (N)	Standard Error (SE)	Margin of Error (ME)	CI -> Lower Limit	CI -> Upper Limit
2	Obama	51.30958607	6.167146	418	0.301644971	0.028917187	51.28066888	51.33850326
3	Trump	42.36928541	3.086017	141	0.259889514	0.042897037	42.32638837	42.41218245
4	GWBush	54.04641829	16.92339	282	1.007773209	0.1176214	53.92879689	54.16403969
5	Biden	46.26654899	6.500581	31	1.167538814	0.410996924	45.85555207	46.67754591

b. Identify which President has the narrowest and widest confidence intervals. Discuss potential factors, such as sample size or sample variance, that may influence these outcomes.

Steps:

- In the same sheet add a new column "**CI Width**", with the formula,
  CI Width = CI Upper Limit – CI Lower Limit

	A	B	C	D	E	F	G	H	I
1	President Name	Mean (μ)	Standard Deviation (σ)	Size (N)	Standard Error (SE)	Margin of Error (ME)	CI -> Lower Limit	CI -> Upper Limit	CI Width
2	Obama	51.30958607	6.167146	418	0.301644971	0.028917187	51.28066888	51.33850326	0.057834373
3	Trump	42.36928541	3.086017	141	0.259889514	0.042897037	42.32638837	42.41218245	0.085794073
4	GWBush	54.04641829	16.92339	282	1.007773209	0.1176214	53.92879689	54.16403969	0.235242801
5	Biden	46.26654899	6.500581	31	1.167538814	0.410996924	45.85555207	46.67754591	0.821993848
6									

- Sort the CI Width in ascending order, and it resulted in the order of **narrowest to Widest confidence intervals** as below.
  Obama < Trump < GWBush < Biden
- By this result, here we can observe and infer that as the **Sample Size increases, Confidence Interval becomes narrow.**

Task 2 : Hypothesis Testing

You will use data in "Trump" and "Obama" worksheets to complete this task.

In Task 2, you may find it necessary to filter the dataset for hypothesis testing purposes.

a. Execute a t-test to examine whether President Obama had a higher approval rating during the initial 1000 days of his presidency compared to the final 1000 days. In your response, clearly outline the competing hypotheses, the calculated test statistics, and the p-value. Make sure to state your test conclusion unambiguously—either rejecting or failing to reject the null hypothesis—based on a 5% significance level. Also, clarify the criteria you employed to arrive at this conclusion.

Steps:

- Copy the "Trump" data sheet into the new worksheet named as "**Task 2a – Hypothesis Testing**".

Null Hypothesis (H0): **μinitial ≤ μfinal** (There is no significant difference in the approval ratings between the initial and final 1000 days of Obama's presidency).

Alternative Hypothesis (H1): **μinitial > μfinal** (Obama's approval rating during the initial 1000 days is significantly higher than during the final 1000 days).

- This is a **Right-Tailed Test**, and so we have calculated each value say mean, standard deviation on excel sheet using the formulas separately.
- For this scenario, we can also use the data of "**t-Test: Two Sample Assuming Unequal Variances**" pulled from Data Analysis Toolpak on Excel sheet.
- Manual calculations on Excel using formulas and t-Test plotted table results are matched here.

Count_Initial1000	166
Count_Last1000	167
Mean1_Initial 1000 Days	53.63981333
Mean2_Last 1000 Days	49.00057903
Standard Dev1	7.889011992
Standard Dev2	4.248126801
Standard Error SE	0.694969134
T - Statistic	**6.675453737**
T-Test (P-Value)	**7.75308E-11**
α	0.05

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	53.6398133	49.000579
Variance	62.2365102	18.0465813
Observations	166	167
Hypothesized Mean Difference	0	
df	253	
t Stat	**6.67545374**	
P(T<=t) one-tail	**7.7499E-11**	
t Critical one-tail	1.65089868	
P(T<=t) two-tail	1.55E-10	
t Critical two-tail	1.9693848	

Calculated Test Statistics = 6.67545374

p-value = 7.7499E-11

Here, we have to state the conclusion based on 5% Significance value i.e., α = 0.05.

7.7499E-11 < 0.05 i.e., p-value < α

Test Conclusion :

Reject H0, Null Hypothesis; We can conclude that "Obama's approval rating during the initial 1000 days is significantly higher than during the final 1000 days".

Here we have followed the p-Value Approach of Alternative Hypothesis HA: μ > μ0, with Right-Tail Probability P(Z>=z). And so, we reject H0, as p-value < α.

b. Perform a t-test to investigate if President Trump maintained a consistent approval rating between the first 500 days and the last 500 days of his term. Your answer should specify the competing hypotheses, the value of the test statistics, and the p-value.

Clearly state the conclusion of your test, whether you reject or fail to reject the null hypothesis, at a 1% significance level. Explain the criteria that guided your conclusion.

Steps:

- Copy the "Trump" data sheet into the new worksheet named as "**Task 2b – Hypothesis Testing**".

Null Hypothesis (H0): **μfirst 500 days = μlast 500 days** (There is no significant difference in approval ratings between the first and last 500 days of President Trump's term).

Alternative Hypothesis (H1): **μfirst 500 days ≠ μlast 500 days** (There is a significant difference in approval ratings between the first and last 500 days of President Trump's term).

- This is a **Two-Tailed Test**, and so we have calculated each value say mean, standard deviation on excel sheet using the formulas separately.
- For this scenario, we can also use the data of "**t-Test: Two Sample Assuming Equal Variances**" pulled from Data Analysis Toolpak on Excel sheet.
- Manual calculations on Excel using formulas and t-Test plotted table results are matched here.

Count_Initial500	82
Count_Last500	38
Mean1_Initial 500 Days	41.37441551
Mean2_Last 500 Days	44.49952416
Standard Dev1	2.467421571
Standard Dev2	3.54209103
Standard Error SE	0.635936019
T - Statistic	**-4.914187212**
T-Test (P-Value)	**1.48129E-07**
α	0.01

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	41.3744	44.4995
Variance	6.08817	12.5464
Observations	82	38
Pooled Variance	8.11321	
Hypothesized Mean Differen	0	
df	118	
t Stat	**-5.5908**	
P(T<=t) one-tail	7.4E-08	
t Critical one-tail	2.35837	
P(T<=t) two-tail	**1.5E-07**	
t Critical two-tail	2.61814	

Calculated Test Statistics = -5.5908

p-value = 1.5E-07

Here, we have to state the conclusion based on 1% Significance value i.e., α = 0.01.

1.5E-07 < 0.01 i.e., p-value < α

Test Conclusion :

Reject H0; We can conclude that "There is a significant difference in approval ratings between the first and last 500 days of President Trump's term".

Here we have followed the p-Value Approach of Alternative Hypothesis HA: $\mu \neq \mu 0$, with Two - Tail Probability 2P(Z>=z) if z > 0 or 2P(Z<=z) if z < 0. And so, we reject H0, as p-value < α.

Task 3 : Statistical Inference Concerning Two populations

By using the sex and Q3a variables in "PEW" worksheet, complete the following tasks.

Research hypothesis: "Male Americans are more likely to prioritize having close relationship to Russia than their female counterparts".

A. **Clearly state both the null and alternative hypotheses.**

   Null Hypothesis (H0) : pmale = pfemale ==> pmale – pfemale <= 0 ; Male and female Americans have the same likelihood of prioritizing having a close relationship with Russia.

   Alternative Hypothesis (H1) : pmale > pfemale ==> pmale – pfemale > 0 ; Male Americans are more likely to prioritize having a close relationship with Russia than their female counterparts.

B. **Differentiate between Type-I and Type-II errors in hypothesis testing.**

   Type-I Error (False Positive):
   Rejecting the null hypothesis when it is actually true. In this context, it would mean concluding that male Americans are more likely to prioritize a close relationship with Russia when, in reality, there is no difference.

   Type-II Error (False Negative):
   Failing to reject the null hypothesis when it is actually false. In this context, it would mean not concluding that male Americans are more likely to prioritize a close relationship with Russia when, in reality, there is a difference.

C. **Assuming that the sampling distribution of the differences in sample proportions is approximately normal, construct a 90% confidence interval for the difference between the two population proportions (male and female). Provide an interpretation of your findings.**

Steps:

- Initially copy the columns "**Sex**", "**Q3a**" from "**PEW**" worksheet to the new worksheet named as "**Task 3**".
- Apply the filters accordingly for proportions to calculate, and note the details of **n1, n2, x1^, x2^.**
- Calculate the values p1^ and p2^, p^ using the formulas as,
  P1^ = x1^ / n1^
  P2^ = x2^ / n2^
  P^ = (x1^+x2^) / (n1^+n2^)
- Now calculate the Standard Error SE value using the formula**,**
  Standard Error (SE) = SQRT((p^*(1-p^)) * ((1/n1)+(1/n2))).
- Here, as it is given as **90% Confidence Interval,** α = 0.10 and α/2 = 0.05 which implies **Zα/2 = 1.645.**
- Now we can calculate the Confidence Interval values using the formulas,
  CI Lower Limit = (p1^ - p2^) – (Zα/2 * SE)
  CI Upper Limit = (p1^ - p2^) + (Zα/2 * SE)

Male Count (n1)	354
Female Count (n2)	331
Russia_Male (x1)	100
Russia_Female (x2)	90
p_male (p1^)	0.28248588
p_female (p2^)	0.27190332
p^	0.27737226
p1^-p2^	0.01058255
p^(1-p^)	0.20043689
1/n1 + 1/n2	0.00584601
SE	0.03423091

At 90% CI	
Zα/2	1.645
CI_Lower	-0.04573
CI_Upper	0.066892

Therefore, **90% confidence interval** for the difference between the two population proportions (male and female) is **[-0.04573, 0.066892].**

With 90% confidence, we can now report that percentage change between the two population proportions male and female who are likely to prioritize having close relationship to Russia is -4.57% and 6.69%.

D. Compute the value of the test statistics and the corresponding p-value, assuming a 5% significance level and independent sampling. Your calculation should focus on testing the difference between the two population proportions.

Steps:

- This is the continuation of the previous problem 3(c), after calculating all those values, calculate the test statistic using the formula,
  Test Statistic Z = (p1^ - p2^) / SE
- As this is a **Right-Tailed Test**, and got a test statistic **Z** as **positive** value, for finding the p-value, use the formula as below,
  p-Value = 1 – NORM.S.DIST(Z, TRUE).

Male Count (n1)	354
Female Count (n2)	331
Russia_Male (x1)	100
Russia_Female (x2)	90
p_male (p1^)	0.28248588
p_female (p2^)	0.27190332
p^	0.27737226
p1^-p2^	0.01058255
p^(1-p^)	0.20043689
1/n1 + 1/n2	0.00584601
SE	0.03423091
Z_Test Statistic	**0.30915192**
α	0.05
p-value	**0.37860298**

Calculated Test Statistics = 0.30915192

p-value = 0.37860298

Here, we have to state the conclusion based on 5% Significance value i.e., α = 0.05.

0.37860298 > 0.05 i.e., p-value > α

E. Based on a 5% significance level, draw a conclusion from your test results. Do the findings lend support to the research hypothesis?

Test Conclusion :

Here we got **p-value >** α from the above calculations.

Do not Reject H0; We cannot conclude that "Male Americans are more likely to prioritize having a close relationship with Russia than their female counterparts".

Here we have followed the p-Value Approach of Alternative Hypothesis HA: p1 – p2 > d0, with Right-Tail Probability P(Z>=z). And so, we do not reject H0, as p-value > α.

Therefore, the findings here do not lend support to the Research Hypothesis.

Task 4 - Statistical Inference Concerning Two populations

You will use the partyln and Q4 variables in "PEW" worksheet to complete this task.

"Are people with different political leanings equally likely to say conflict akin to the Cold War will occur between the China and the US?"

A. Define the null and the alternative hypothesis.

In statistical hypothesis testing, the null hypothesis (*H*0) and the alternative hypothesis (*H*1 or *Ha*) are statements about a population or populations. They are used to formulate a testable claim and to guide the statistical analysis.

Null Hypothesis (H0) : pRepublican_likely - pDemocrat_likely = 0 ; Republicans and Democrats are equally likely to say that conflict akin to the Cold War will occur between China and the US.

Alternative Hypothesis (H1) : pRepublican_likely - pDemocrat_likely ≠ 0 ; Republicans and Democrats are not equally likely to say that conflict akin to the Cold War will occur between China and the US.

B. Calculate the value of test statistics and the p-value at 8% significance level to test your hypothesis.

Steps:

- Initially copy the columns "**PARTYLN**", "**Q4**" from "**PEW**" worksheet to the new worksheet named as "**Task 4**".
- Apply the filters accordingly for proportions to calculate, and note the details of **n1, n2, x1^, x2^.**
- Calculate the values p1^ and p2^, p^ using the formulas as,
  P1^ = x1^ / n1^
  P2^ = x2^ / n2^
  P^ = (x1^+x2^) / (n1^+n2^)
- Now calculate the Standard Error SE value using the formula**,**

Standard Error (SE) = SQRT((p^*(1-p^)) * ((1/n1)+(1/n2))).

- Calculate the test statistic using the formula,
  Test Statistic Z = (p1^ - p2^) / SE
- As this is a **Two-Tailed Test**, and got a test statistic **Z** as **negative** value, for finding the p-value, use the formula as below,
  p-Value = 2 * NORM.S.DIST(Z, TRUE).

Republican Count (n1)	325
Democratic Count (n2)	360
Republican_likely (x1)	194
Democratic_likely (x2)	221
p_Republican_likely (p1^)	0.596923077
p_Democratic_likely (p2^)	0.613888889
p^	0.605839416
p1^-p2^	-0.016965812
p^(1-p^)	0.238798018
1/n1 + 1/n2	0.005854701
SE	0.037391055
Z_Test Statistic	**-0.453739863**
p-value	**0.650016072**

Calculated Test Statistics = -0.453739863

p-value = 0.650016072

Here, we have to state the conclusion based on 8% Significance value i.e., α = 0.08.

0. 650016072 > 0.08 i.e., p-value > α

Test Conclusion :

Here we got **p-value >** α from the above calculations.

α	0.08
p-value > α	Do not Reject H0; We cannot conclude that "Republicans and Democrats are not equally likely to say that conflict akin t

Do not Reject H0; We cannot conclude that " Republicans and Democrats are not equally likely to say that conflict akin to the Cold War will occur between China and the US ".

Here we have followed the p-Value Approach of Alternative Hypothesis HA: p1 – p2 ≠ d0, with Two -Tail Probability 2P(Z>=z) if z > 0 or 2P(Z<=z) if z < 0. And so, we reject H0, as p-value < α.

C. **At the 5% level of significance, what is your conclusion to the test. Do the results support the research hypothesis?**

Calculated Test Statistics = -0.453739863

p-value = 0.650016072

Here, we have to state the conclusion based on 5% Significance value i.e., α = 0.05.

0. 650016072 > 0.05 i.e., p-value > α

Test Conclusion :

Here we got **p-value >** α from the above calculations.

α	0.05
p-value > α	Do not Reject H0; We cannot conclude that "Republicans and Democrats are not equally likely to say that conflict akin to

Do not Reject H0; We cannot conclude that " Republicans and Democrats are not equally likely to say that conflict akin to the Cold War will occur between China and the US ".

Here we have followed the p-Value Approach of Alternative Hypothesis HA: p1 – p2 ≠ d0, with Two -Tail Probability 2P(Z>=z) if z > 0 or 2P(Z<=z) if z < 0. And so, we reject H0, as p-value < α.

Therefore, the findings here do not lend support to the Research Hypothesis, either in 8% or 5% significance level.

Task 5: Statistical Inference Concerning Two populations

You will use the EDUCATION and PARTYLN variables in "PEW" worksheet to complete this task.

"Are Republican-leaning Americans more educated than Democratic-leaning Americans?"

a) **Calculate the point estimate of our parameter of interest,**
$pRepHigh\_educated - pDemHigh\_educated$

Steps:

- Initially copy the columns **"Education",** "**PARTYLN**" from "**PEW**" worksheet to the new worksheet named as "**Task 5**".
- Apply the filters accordingly for proportions to calculate, and note the details of **n1, n2, x1^, x2^.**
- Calculate the values p1^ and p2^, p^ using the formulas as,
  P1^ = x1^ / n1^
  P2^ = x2^ / n2^

- Now, calculate Point Estimate $pRepHigh\_educated - pDemHigh\_educated$, i.e., p1^-p2^

Republican Count (n1)	325
Democratic Count (n2)	360
HighlyEducated_Republican (x1)	182
HighlyEducated_Democratic (x2)	210
p_HighEducated_Republican (p1^)	0.56
p_HighEducated_Democratic (p2^)	0.583333333
p1^ - p2^ (Point Estimate)	**-0.023333333**

Therefore, calculated Point Estimate value is -0.023333333

b) **Assuming that the population parameter is normally distributed and all conditions for confidence interval calculations for proportion differences met, calculate the appropriate standard error for the point estimate, $pRepHigh\_educated - pDemHigh\_educated$.**

Steps:

- Continue the process for the above problem. Calculate the Standard Error of two populations proportions using the formula,
  SE = SQRT(([p1^(1-p1^)] / n1) + ([p2^(1-p2^)] / n2))

Republican Count (n1)	325
Democratic Count (n2)	360
HighlyEducated_Republican (x1)	182
HighlyEducated_Democratic (x2)	210
p_HighEducated_Republican (p1^)	0.56
p_HighEducated_Democratic (p2^)	0.583333333
p1^ - p2^ (Point Estimate)	**-0.023333333**
[p1^(1-p1^)]/n1	0.000758154
[p2^(1-p2^)]/n2	0.000675154
SE of Point Estimate	**0.037859057**

Therefore, calculated Standard Error is 0.037859057

c) **Construct a 90% confidence interval for proportion differences.**

Steps:

- Continue the process for the above problem. Here, as it is given as **90% Confidence Interval,** α = 0.10 and α/2 = 0.05 which implies **Zα/2 = 1.645.**

- Now we can calculate the Confidence Interval values using the formulas,
  CI Lower Limit = (p1^ - p2^) – (Zα/2 * SE)
  CI Upper Limit = (p1^ - p2^) + (Zα/2 * SE)

Republican Count (n1)	325
Democratic Count (n2)	360
HighlyEducated_Republican (x1)	182
HighlyEducated_Democratic (x2)	210
p_HighEducated_Republican (p1^)	0.56
p_HighEducated_Democratic (p2^)	0.583333333
p1^ - p2^ (Point Estimate)	**-0.023333333**
[p1^(1-p1^)]/n1	0.000758154
[p2^(1-p2^)]/n2	0.000675154
SE of Point Estimate	**0.037859057**
Zα/2 (90% Confidence Interval)	1.645
CI_Lower Limit	**-0.085611481**
CI_Upper Limit	**0.038944815**

Therefore, **90% confidence interval** for the difference between the two population proportions (Republican_HighEducated and Democratic_HighEducated) is **[-0.08561, 0.038945].**

With 90% confidence, we can now report that percentage change between the two population proportions Republic and Democratic who are more educated is -8.56% and 3.89%.

d) **Based on the 90% confidence interval values, what can you say about our research question? Do you have statistically significant evidence to say that the level of education differs across political leanings: proportion of Republican-leaning Americans who are considered as highly-educated differ from their democratic leaning counterpart? Please explain.**

Test Conclusion:

Here we have the 90% Confidence Interval [-0.08561, 0.038945].
So, Confidence Interval contains 0 value in it, which implies that we do not Reject H0.
Do not Reject H0 ; We cannot conclude at 90% Confidence Interval that "the level of education differs across political leanings: proportion of Republican-leaning Americans who are considered as highly-educated differ from their democratic leaning counterpart".

Therefore, we do not have statistically significant evidence to say that the level of education differs across political leanings: proportion of Republican-leaning Americans who are considered as highly-educated differ from their democratic leaning counterpart.