TASK 1: CONFIDENCE INTERVAL ESTIMATIONS

a) Given the information, construct a 95% confidence interval for "Approing" rate for each president in your dataset and comment your findings.

	Column1
Mean	40.38297872
Standard Error	0.262607063
Median	40
Mode	38
Standard Deviation	3.118286101
Sample Variance	9.723708207
Kurtosis	0.507595279
Skewness	0.685344911
Range	15
Minimum	34
Maximum	49
Sum	5694
Count	141
Confidence Level(95.0%)	0.519188271
lower limit	39.86379045
upper limit	40.90216699

Figure 1: Trump 95% confidence interval for Approving rate

	Column1
Mean	47.96889952
Standard Error	0.262607319
Median	47
Mode	46
Standard Deviation	5.36901933
Sample Variance	28.82636857
Kurtosis	1.684423712
Skewness	1.213709111
Range	27
Minimum	40
Maximum	67
Sum	20051
Count	418
Confidence Level(95.0%)	0.516199107
lower limit	47.45270041
upper limit	48.48509863

Figure 2:Obama 95% confidence interval for Approving rate

	Column1
Mean	51.34751773
Standard Error	0.945246058
Median	50.5
Mode	32
Standard Deviation	15.87338058
Sample Variance	251.9642109
Kurtosis	-0.558889046
Skewness	0.453670948
Range	65
Minimum	25
Maximum	90
Sum	14480
Count	282
Confidence Level(95.0%)	1.860662127
lower limit	49.4868556
upper limit	53.20817986

Figure 3:GWBush 95% confidence interval for Approving rate

	Column1
Mean	45.58333333
Standard Error	1.328228452
Median	42
Mode	41
Standard Deviation	6.50696394
Sample Variance	42.34057971
Kurtosis	-1.001554028
Skewness	0.84237061
Range	19
Minimum	38
Maximum	57
Sum	1094
Count	24
Confidence Level(95.0%)	2.747649896
lower limit	42.83568344
upper limit	48.33098323

Figure 4:Biden 95% confidence interval for Approving rate

Comments:

The 95% confidence interval for Approving rate for Biden – (42.83568344,48.33098323)

The 95% confidence interval for Approving rate for Trump – (39.86379045, 40.90216699)

The 95% confidence interval for Approving rate for Obama – (47.45270041, 48.48509863)

The 95% confidence interval for Approving rate for GWBush – (49.4868556, 53.20817986)

1) GW Bush has a higher mean approval rating than the other three presidents. It signifies that average approval rating is more for him.

2) Compared to the other three presidents, Biden values the standard error more. It implies that the estimate is more accurate.

3) GW Bush has a higher standard deviation number, which shows that the data are more evenly distributed.

4) The skewness of all presidents is positive; we can conclude that they are positively skewed distribution.

5) All presidents' confidence intervals, except for Biden, do not overlap, showing that their approval ratings differ from one another.

b) Which president has the narrowest/widest confidence intervals and what factors (i.e., sample size, sample variance, etc.) might contribute to this outcome.
Obama has the narrowest confidence interval, while Biden has the biggest. The other variables, such as sample size, influence the confidence interval. The confidence interval is smaller when there are more samples, while the confidence interval is larger when there are fewer samples. Obama had more samples in this instance, which lowers the level of confidence.

The other components sample variance also impacts the confidence interval. Higher sample variance leads to largest confidence interval while lesser sample variance leads to narrowest confidence interval. The interval is also impacted by the confidence level. The interval is larger, the higher the confidence level. The confidence intervals are also influenced by this issue. The precision of confidence interval is lower for Biden approval rate because of difference between upper and lower limit is more than other presidents.

TASK 2: Hypothesis Testing

a) Use a one-sample t test to determine if "President Obama has a higher approval rating in the first 1000 days of his presidency than the last 1000 days"

In your answer, specify the competing hypothesis, the value of test statistics, and the p-value. You should clearly state the conclusion of your test (reject the null or fail to reject the null) at 5% significance level and the criteria you used to reach that conclusion.

TASK 2	
n for last 1000 days	143
Mean for last 1000 days	47.51748252
sample standard deviation for last 1000days	4.30664921
Approval rate for last 1000 days	0.332290088
n for first 1000 days	144
Mean fo first 1000 days	50.02777778
sample standard deviation for first 1000 days	6.841813488
Approval rate for first 1000 days	0.347415123
Pooled Standard Deviation	5.7065013
test static value	0.148178044
p value	0.441153521

Null Hypothesis: The approval rate of President Obama in his presidency for first 1000 days is less than or equal to approval rate in last 1000days.

Alternate Hypothesis: The approval rate of President Obama in his presidency for first 1000 days is greater than approval rate in last 1000days.

At 5% significance level, the p value is greater than 0.05. So, we do not reject the null hypothesis and Null hypothesis is true. There is no significance evidence to tell that approval rate for President Obama in first 1000 days is greater than in last 1000days.

b) Use a one-sample t test to determine if "President Trump has a same approval rating in the first 500 days of his presidency than the last 500 days"

In your answer, specify the competing hypothesis, the value of test statistics, and the p-value. You should clearly state the conclusion of your test (reject the null or fail to reject the null) at 1% significance level and the criteria you used to reach that conclusion.

Null Hypothesis: The approval rate of President Trump in his presidency for first 500 days is not equal to approval rate in last 1000days.

Alternate Hypothesis: The approval rate of President Trump in his presidency for first 500 days is equal to approval rate in last 1000days.

TASK 2	
n for last 500 days	27
Mean for last 500 days	43.18518519
sample standard deviation for last 500 days	3.961711334
Approval rate for last 500 days	1.599451303
n for first 500 days	72
Mean for first 500 days	38.875
sample standard deviation for first 500 days	2.207142694
Approval rate for first 500 days	0.539930556
Pooled Standard Deviation	2.778929548
test static value	0.437633
p- value	0.662625247

At 1% significance level, the p value is greater than 0.01. So, we do not reject the null hypothesis and Null hypothesis is true. There is significance evidence to tell that approval rate for President Trump in first 1000 days is not equal to approval rate in last 1000days and there is no significance evidence to tell that approval rate for President Trump in first 1000 days is equal to approval rate in last 1000days

TASK 3: Statistical Inference Concerning Two populations

Count of Q3a	Column Labels		
Row Labels	**Female**	**Male**	**Grand Total**
Having a close relationship to Germany	241	254	495
Having a close relationship to Russia	90	100	190
Grand Total	**331**	**354**	**685**

		Female	Male	Total
Having a close relationship to Germany	Observed	**241**	**254**	495
	Expected	239.19	255.81	495.00
Having a close relationship to Russia	Observed	**90**	**100**	190
	Expected	91.81	98.19	190.00
Total	Observed	331	354	685
	Expected	331.00	354.00	685.00

.10 chi-square
1 df
.7572 p-value

.012 Cramér's V

a) Fill out the following 2 by 2 tables by including both the expected and observed frequencies for each cell.

Observed Frequencies	**Female**	**Male**	**Grand Total**
Having a close relationship to Germany	241	254	495
Having a close relationship to Russia	90	100	190
Grand Total	331	354	685

Expected Frequencies	**Female**	**Male**	**Grand Total**
Having a close relationship to Germany	239.19	255.81	495.00
Having a close relationship to Russia	91.81	98.19	190.00

Grand Total	331.00	354.00	685.00

b) State the Null and the Alternate Hypothesis:

Null Hypothesis: There is no significant connection between genders in sex column and having a close relationship to Germany or Russia in Q3a column.

Alternate Hypothesis: There is significant connection between genders in sex column and having a close relationship to Germany or Russia in Q3a column.

c) Conduct a Chi-squared test and fill out the following table from your excel output:

	Chi-square test statistics
Degrees of freedom (df)	1
Value	0.10
P-value	0.7572

Based on results, we cannot the reject the null hypothesis at 5% significance level because the P-value is greater than significance value. Now we can say that there is no significant connection between genders in sex column and having a close relationship to Germany or Russia in Q3a column. In other words, null hypothesis is true.

d) Calculate the Cramer's V value and interpret it.

The Cramer's V value ranges from zero to one. From this range, we can infer that there is a weak to strong correlation between the variables, ranging from 0 to 1. So, here the value is **0.012**. It is very close to zero saying that the null hypothesis is true, and we cannot reject the null hypothesis.

TASK 4: Chi-Square Test

Count of partyIn	Column Labels		
Row Labels	**Republican**	**Democratic**	**Grand Total**
Likely	194	221	415
Unlikely	131	139	270
Grand Total	**325**	**360**	**685**

		Republican	Democratic	Total
Likely	Observed	**194**	**221**	415
	Expected	196.90	218.10	415.00
Unlikely	Observed	**131**	**139**	270
	Expected	128.10	141.90	270.00
Total	Observed	325	360	685
	Expected	325.00	360.00	685.00

.21	chi-square
1	df
.6500	p-value
.017	Cramér's V

Q. How would you rate the likelihood of the current rivalry between China and the United States escalating into a confrontation resembling the cold war?

By responding to the questions listed below, I'd rate the possibility of the current conflict. We must first state the null and alternative hypotheses. They help us clearly grasp what we need to look for. Calculate specified values, such as observed values, expected values, p values, etc. in

accordance with the hypotheses. These criteria can be used to evaluate how people feel about the US-China Cold War. We can determine whether to reject the null hypothesis at a specific significance level. With the aid of opinions from democratic and republican people, we may use this to estimate the likelihood of conflict between China and the US.

a) Generate a two-by-two table in Excel, put Q4 answers (Likely, Unlikely) in rows and political leanings (Republican, Democratic) in columns. In your table, make sure both the frequency counts and expected values are listed. Insert your tables below.

Observed	Republican	Democratic	Total
Likely	194	221	415
Unlikely	131	139	270
Total	325	360	685

Expected	Republican	Democratic	Total
Likely	196.90	218.10	**415.00**
Unlikely	128.10	141.90	**270.00**
Total	**325.00**	**360.00**	**685.00**

b) State the Null and the Alternate Hypothesis:

Null Hypothesis: There is no significant connection between the people with two different political leanings and their views on likelihood of cold war happening between China and US.

Alternate Hypothesis: There is significant connection between the people with two different political leanings and their views on likelihood of cold war happening between China and US.

c) Conduct a Chi-squared test and fill out the following table from your Excel Output:

	Chi-square test statistics
Degrees of freedom (df)	1
Value	0.21
P-value	0.6500

d) Based on your results, would you reject the null hypothesis at 5% significance level?

Based on results, we cannot the reject the null hypothesis at 5% significance level because the P-value is greater than significance value. Now we can say that there is no significant connection

between the people with two different political leanings and their views on likelihood of cold war happening between China and US. In other words, null hypothesis is true.

TASK 5: Statistical Inference Concerning Two populations

Count of partyIn	Column Labels		
Row Labels	Republican	Democratic	Grand Total
High-educated	182	210	392
Grand Total	**182**	**210**	**392**
$\hat{p}^{Rep}_{High-educated}$	0.464285714		
$\hat{p}^{Dem}_{High-educated}$	0.535714286		
Point estimate of our parameter of interest	-0.071428571		
standard error for the point estimate	0.050507627		
z	-1.644853627		
lower limit	-0.154506225		
upper limit	0.011649082		

a)Calculate the point estimate of our parameter of interest, $p^{Rep}_{High-educated} - p^{Dem}_{High-educated}$

The point estimate of our parameter of interest is -0.71428571

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, $p^{Rep}_{High-educated} - p^{Dem}_{High-educated}$

Standard error for the point estimate is 0.050507627.

c) Construct a 90% confidence interval for proportion differences:

By constructing a confidence interval in excel shown above, we got below confidence interval values for upper limit and lower limit.

The 90% confidence interval is (-0.154506225, 0.011649082)

d) Based on the 90% confidence interval, 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 consider as highly educated differ from their democratic leaning counterpart? Please explain.

No, we cannot decide the research question based on 90% confidence interval. First point, the zero can include between width of interval. If it is zero, there is no significant difference between Republican highly educated people and Democratic highly educated people. So, we there is no significant evidence to say that there is difference between Republican leaning American and Democratic leaning Americans who are considered as highly educated.