Luke Ehrenstrom

Final

Question 1

1. With no impact from Party ID and assuming they lived in the North, the average respondent had a 41 feeling of approval for Hillary Clinton. For every increase of 1 in Party ID, respondents gained 6.09 points of approval and if they were from the South, they lost 13.53 points of approval for Hillary Clinton. The data is statistically significant because of the low p-value generated from the F-test, indicating that there is a relationship being demonstrated within the multivariate regression. However, there is still a significant amount of variance unaccounted for indicated by the low R2 value.
2. R2 = 0.24
3. n = 100
4. k = 2
5. df1 = k # number of predictors
6. df2 = n-(k+1)
8. FStatistic = (R2/(1-R2))\*(df2/df1)
9. p = (1 - pf(FStatistic, df1, df2))
10. We would need to assume that Party Identification and whether they live in the South are able to be manipulated. These can both be manipulated, but not easily; the cultural values of the region where you live are relatively immutable. Also, there are other factors that play a role in determining whether someone approved of Hillary Clinton or not. Further variables should be considered in what causes someone’s view of Hillary Clinton to be changed.
11. F-Statistic: 15.32.

The F-Statistic indicates that there is an impact from the variables tested in the linear regression.

Source on F-Statistic: <https://stats.stackexchange.com/questions/56881/whats-the-relationship-between-r2-and-f-test>

Question 2

1. With no impact from Party ID and assuming they lived in the North, the average respondent had a 36 feeling of approval for Hillary Clinton. For every increase of 1 in Party ID, respondents from the North gained 8.09 points of approval. For those from the South, they lost 10.53 points of approval from the start and also only gained 4.52 points of approval per 1 increase in Party ID. The data is statistically significant because of the low p-value generated from the F-test, indicating that there is a relationship being demonstrated within the multivariate regression. However, there is still a significant amount of variance unaccounted for indicated by the low R2 value.
2. ŷ = 36 + 8.09x1

where ŷ is equal to the approval of someone from the North relative to their Party ID.

where x1 is equal to the Party ID.

1. ŷ = 25.47 + 4.52x1

where ŷ is equal to the approval of someone from the South relative to their Party ID.

where x1 is equal to the Party ID.

This indicates that in the political world, approval for a candidate is highly dependent upon region and partisan affiliation, and that dependent on the region, the effect of party identification becomes lessened.

Question 3

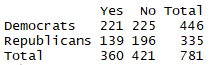
1. The causal effect of the mailers appears to be those that received the mailers and had their attitudes measured after the mailers were sent out had a more negative opinion (-3.57) of Hillary Clinton.
2. One assumption of this causal claim is that each participant had a positive chance to receive the mailers, which they did not: they were targeted by campaign officials based on demographic data.

Question 4

The 90% confidence interval is (88.67, 99.06).

1. n = 11
2. SampStdDev = 9.50
3. mean = 93.87
4. t = qt(.95, n-1)
5. MoE = SampStdDev/sqrt(n)
6. CILeft = mean - t\*MoE
7. CIRight = mean + t\*MoE
8. CI = c(CILeft, CIRight)
9. CI
10. # 88.67, 99.06

Question 5

1. 
2. n = 781
3. TableACAObserved = matrix(c(221, 225, 446, 360-221, 335-(360-221), 335, 360, 225+(335-(360-221)), 781), byrow = T, ncol = 3)
4. colnames(TableACAObserved) = c("Yes", "No", "Total")
5. rownames(TableACAObserved) = c("Democrats", "Republicans", "Total")
6. TableACAObserved
7. Cell chi-squared = 1.317
8. RepublicanNoObs = 196
9. RepublicanNoExp = 180.58
10. RepublicanNoX = ((RepublicanNoObs- RepublicanNoExp)^2)/RepublicanNoExp
11. *H*0: f(observed) = f(expected)

*H*a : f(observed) ≠ f(expected)

Df = (row-1)(col-1) = 1

Chi-Square for 95% Confidence: 3.841

Chi-Square from Observed: 5.02

The Chi-Square test statistic from the observations exceeds the Chi-Square test statistic for 95% confidence. Thus, we can reject Ho and thus there is some evidence to suggest that there is some relationship between party affiliation and whether a respondent supports the Affordable Care Act.

1. There is some evidence to suggest that there is a relationship between party affiliation and the response for approval of the Affordable Care Act.

Question 6

1. μ ₁ - μ ₂ = 3.8-3.5 = 0.3

CI: (-0.092, 0.692)

1. # Question 6
2. n1 = 288
3. n2 = 242
4. stddev1 = 2.4
5. stddev2 = 2.2
6. samplemudist = 3.8 - 3.5
7. stderror = sqrt(((stddev1^2)/n1)+((stddev2^2)/n2))
8. # both n1 and n2 exceed 30, can test normality
9. zstatistic = 1.96
10. CILeft = samplemudist - stderror\*zstatistic
11. CIRight = samplemudist + stderror\*zstatistic
12. CI = c(CILeft, CIRight)
13. CI
14. # -0.092, 0.692
15. *H*0: μ ₁ - μ ₂ = 0

*H*a : μ ₁ - μ ₂ ≠ 0

p = .1336

We fail to reject Ho, indicating that there may be some evidence to suggest that the civics class has no effect on Party ID.

* 1. p = pnorm(samplemudist, 0, sd = stderror, lower.tail = FALSE)\*2

1. No; while we may be able to make some level of a causal inference, we cannot simply assume that the results of the hypothesis test are indicative of a causal relationship because of other variables that may be playing a role.

Question 7

1. A sample distribution are the statistics associated with a single sample drawn from the population. A sampling distribution are the aggregated statistics drawn from numerous samples. A population distribution are the true parameters of a population. Using samples to form a sampling distribution, the sampling distribution can be used to estimate the parameters of a population.
2. Using spending as a dependent variable would likely have the confounding variable of inflation. If you were studying under which administrations did spending increase, then inflation would have a strong impact on whether administration policy actually increased spending or if inflation did. Another concern would be that in times of economic recession, spending would increase due to a higher number of people utilizing unemployment benefits and other entitlements. To combat these concerns, you could run a multivariate regression that includes rate of inflation and whether the country was experiencing a recession.

Question 8

1. P-Value: The probability that a hypothesis test will produce a Type-I error, or the probability that a hypothesis test will reject Ho when it is actually true.
2. Outliers: Individuals whose responses are substantially different from the rest of the data; can be measured as exceeding Q3+1.5\*IQR or is less than Q1-1.5\*IQR of a data set.
3. Autocorrelation: A comparison that tests the relationship between a variable’s current values and a variable’s past values measured over a period of time.

Source: <https://www.investopedia.com/terms/a/autocorrelation.asp>

1. Standard Error: An estimated standard deviation of the sampling distribution.
2. Counterfactual: In performing causal inference, the counterfactual is what would have happened if a key condition were different.