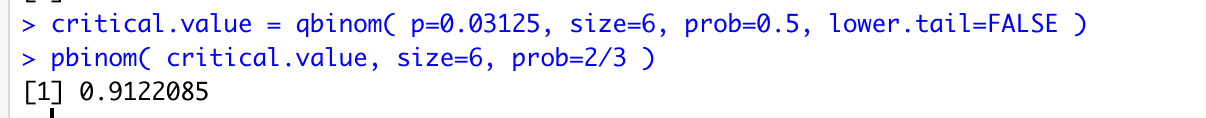
**Question 1:**

1. What is the type 1 error rate? (p-value)
   1. There are two ways to get 0 or 6 as a result over 64 total outcomes
   2. P-value: .03125
2. What is the power of your test for the alternative hypothesis p = ¾?



**Question 2:**

What are the implications of using a paired t-test on a 5 point Likert scale?

Assumptions of paired t-test:

1. A & B have a metric scale with same units
   1. Implications: If the 5 point Likert scale does not associate to a variable that would have an equal distance between the values, then the CLT cannot apply because the values are subjective to greater distances than captured in the 1-5 scale
2. Natural pairing (this is fine)
3. Each pair is i.i.d.
4. Sufficiently normal - violated using ordinal; ordinal does not have a central tendency

**Question 3:**

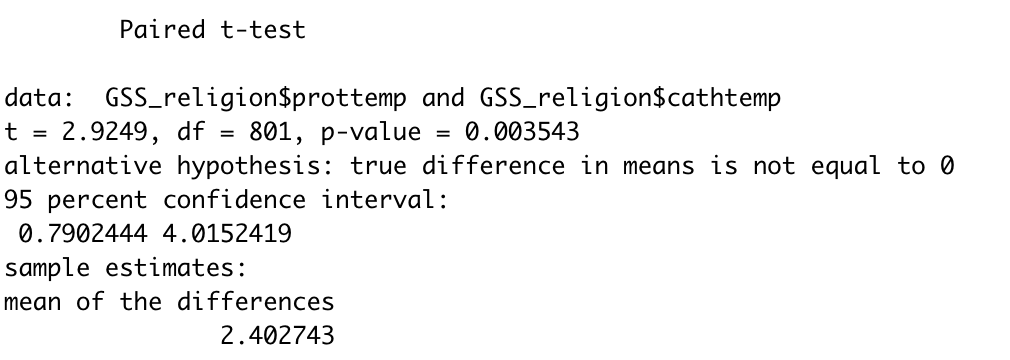
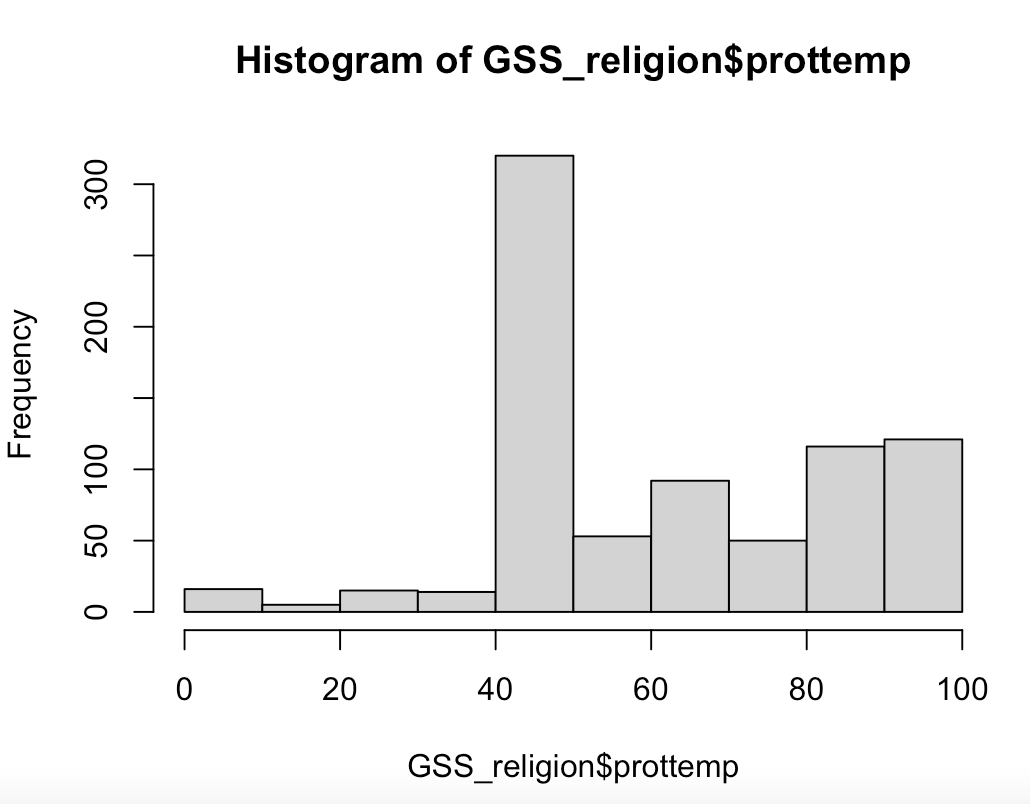
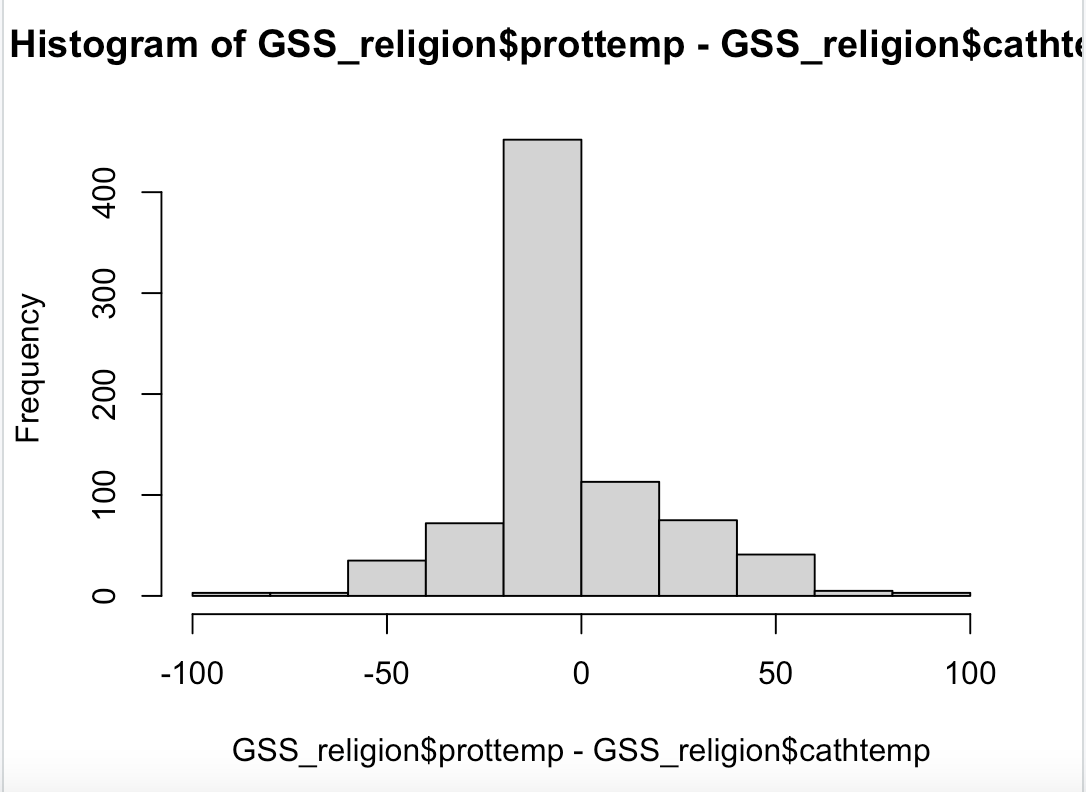
Metric assumption: A variable can be treated as scale (continuous) when its values represent ordered categories with a meaningful metric, so that distance comparisons between values are appropriate.

1. List all assumptions for a two-sample t-test and evaluate them
   1. Metric scale - not valid for variables which only have an ordinal structure: Would the rating system be metric? Seems subjective, how do you know the distance for any one individual is equal between these ratings of 0 to 10?
   2. IID data - I would challenge the IID, the GDP of countries is dependent on other countries
   3. No major deviations from normality, considering the sample size - not valid for very highly skewed distributions at higher sample size - If it does not meet metric scale, then the CLT does not apply so one cannot expect a normal distribution
2. List all assumptions for a Wilcoxon rank-sum test and evaluate them
   1. Metric scale - Age can be used as metric when the value represents an ordered category so that the distance comparisons between values is appropriate.
   2. IID - As long as age is referenced once per senator, I see no issue with iid
   3. The distribution of X is the same as of Y - same shape but changed by a constant - believe age meets this criteria
3. List all assumptions for a signed-rank test and evaluate them

Not sure about this one, is the input countries ranked by wine consumption and output the levels of death by heart or liver disease?

* 1. Metric scale - if measuring wine - could be metric (amount consumed) if related to heart or liver disease
  2. IID -
  3. The distribution of the difference X-Y is symmetric around some mean

1. List all assumptions for a paired t-test and evaluate them
   1. Metric scale - The presented scale seems like it could present some challenges in using the metric scale assumption. The classification between 100 - 50, 50, 50-0 for positive feelings or negative feelings are subjective
   2. IID - The excerpt doesn´t explain how the data collection process occurred. Were these randomly selected people or were they groups that showed up together at a survey location. The first would support IID but the second could allude to some dependency between groups. For example, if the respondents came from the same household and responded together.
   3. The distribution of the difference between measurements has no deviations from normality, considering the sample size



**Part 2: Statistical Analysis**

Question: Did Democratic voters or Republican voters experience more difficulty voting in the 2020 election?

**Independent Variable**

**Need to determine how to classify respondents into Democrat or Republican.**

* Pre -Party of registration (V201018)
* Who they voted for in Primary (V201021)
* Pre: For Whom does R intend to vote for president (V201033)
* Pre-summary: Party of Pre-election presidential vote / intent / preference (V201075x - this is a summary statistic, need to look at how this is reported)
* Pre-summary: Party of Pre-election house vote / intent / preference (V201076x - this is a summary statistic, need to look at how this is reported)
* Pre-summary: Party of Pre-election senate vote / intent / preference (V201077x - this is a summary statistic, need to look at how this is reported)
* Pre-summary: Party of Pre-election gubernatorial vote / intent / preference (V201078x - this is a summary statistic, need to look at how this is reported)
* Party ID (V201231) - agree with Tres that we should use this one for independent variable and associate those independents that lean one way or another into that category. We can reference the document shared “Measuring party support: Leaners are not Independents” petrocik

**Variables related to difficulty voting: Dependent variable**

* Post: How difficult was it for R to vote? (V202119) - could only focus on highlighted
  + -9. Refused
  + -7. No post-election data, deleted due to incomplete interview -6. No post-election interview
  + -5. Interview breakoff (sufficient partial IW)
  + -1. Inapplicable
  + 1. Not difficult at all
  + 2. A little difficult
  + 3. Moderately difficult
  + 4. Very difficult
  + 5. Extremely difficult

