Assignment 3: Statistical Findings Memo

Tina Wooldridge

#| label: setup #| include: true #| echo: false #| message: false #| warning: false library(tidyverse) library(magrittr) # ensures %>% is available

---

Assignment 3: "The Leap"

Conceptual Focus

This assignment covers **Chapter 14: Making Inferences**. In Week 13, we *described* our sample (e.g., "In our sample, Group A used more platforms"). Now, we must *verify* these findings. As your slides state, this is the "leap of faith" from our sample to the population.

Our core question is: "Is the difference I see in my sample 'real'… or could it just be a fluke due to random chance?" We will use **Hypothesis Testing** (and p-values) to answer this, and we'll use the correct test for each job based on your **PowerPoint (Slide 23)**.

Statistical Findings Memo

To: Alex Chen (Director of Insights)

From: [Tina Wooldridge]

Subject: Statistical Verification for 2025 Digital Landscape Brief

This memo formally tests the key patterns identified in our descriptive analysis. The goal is to verify that these findings are statistically significant and not just a fluke of our sample before we present them to the client. All tests use an alpha level of $\alpha = .05$.

```
alternative hypothesis: true difference in means between group PC and group Smartphone is not 
95 percent confidence interval:
 -0.7109869 -0.3386605
sample estimates:
        mean in group PC mean in group Smartphone
                1.338384                 1.863208
```

```
# Optional: group means for interpretation
device_data %>%
  group_by(device_type_w144) %>%
  summarise(mean_platforms = mean(platform_count, na.rm = TRUE),
            sd_platforms   = sd(platform_count, na.rm = TRUE),
            n              = n())
```

```
# A tibble: 2 x 4
  device_type_w144 mean_platforms sd_platforms     n
  <fct>                     <dbl>        <dbl> <int>
1 PC                         1.34         1.10   198
2 Smartphone                 1.86         1.10   424
```

Decision: [Use `rq2_t$p.value`. The group means clarify which device group reports higher platform_count.]

Conclusion: [People who took the survey on a PC and those who took it on a Smartphone use, on average, a **different number of social media platforms**.
The means shown in the t-test output tell us which group reports higher platform use.?]

RQ 3: Difference (Party & Platform Count)

Research Question: Is there a significant *difference* in the mean number of platforms used (`platform_count`) across our three `party_simple` groups?

Hypotheses: * **H0 (Null):** The mean `platform_count` is the same for all three `party_simple` groups in the population. * **HA (Alternative):** At least one group's mean `platform_count` is different from the others.

Results (ANOVA): [The `aov` output will appear below.]

```
# Ensure party is a factor
w144_wrangled <- w144_wrangled %>%
  mutate(party_simple = as.factor(party_simple))

# Fit ANOVA
anova_model <- aov(platform_count ~ party_simple, data = w144_wrangled)
summary(anova_model)
```

```
              Df Sum Sq Mean Sq F value   Pr(>F)    
party_simple   2   23.9  11.956   9.692 7.17e-05 ***
Residuals    619  763.6   1.234                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
4 observations deleted due to missingness
```

Results (Post-Hoc Test): [The `TukeyHSD` output will appear below.]

```r
# Post-hoc Tukey HSD
tukey_res <- TukeyHSD(anova_model)
tukey_res
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = platform_count ~ party_simple, data = w144_wrangled)

$party_simple
                                   diff        lwr         upr     p adj
Independent/Other-Democrat   -0.6397183 -1.2280348 -0.05140172 0.0292262
Republican-Democrat          -0.3605721 -0.5735772 -0.14756707 0.0002300
Republican-Independent/Other  0.2791461 -0.3105162  0.86880851 0.5068859
```

Decision: [The ANOVA p-value is $\mathbf{p < .05}$, so we **Reject H** .
This means the three political groups differ in average platform_count.]

Conclusion: [Political affiliation is meaningfully related to how many social media platforms people use.
The Tukey test clarifies which groups differ and by how much, helping us identify which political segments are the heaviest—or lightest—users.]

Final Discussion

[Taken together, the three statistical tests confirm that our descriptive observations reflect real differences rather than random chance. The chi-square test showed that TikTok use varies significantly by political party, while the t-test revealed a difference in platform adoption between PC and smartphone respondents. However, the most strategically important result comes from the ANOVA: political affiliation is strongly linked to the number of platforms people use, with clear differences across party groups. This finding is both statistically significant and practically meaningful, because it identifies which segments are the heaviest users and therefore the most reachable across multiple channels. In contrast, the device type result, though significant, is less actionable for client strategy. In short, while several patterns are statistically reliable, the party-based differences in platform count provide the most valuable insight for the client presentation. This distinction illustrates the principle of "significance versus meaningfulness": the lowest p-value is not necessarily the most useful result, and our recommendation should emphasize the findings that matter most for decision-making.]

Appendix: R Code and Commentary

1. Setup

[This section loads the packages and the cleaned dataset created in Assignment 1. The tidyverse package provides the tools needed for data wrangling, summarizing, and statistical analysis.]

```
library(tidyverse)
load("w144_wrangled.RData") # Load cleaned dataset from Assignment 1
```

2. RQ 1: Chi-Square Test

[`chisq.test()` is the correct test because both variables — **party_simple** and **uses_tiktok** — are **categorical**. According to Slide 23, when we examine **association between two categorical variables**, the chi-square test is the correct choice.]

```
# Create a contingency table (party_simple x uses_tiktok)
tiktok_party_table <- table(w144_wrangled$party_simple,
                            w144_wrangled$uses_tiktok)

# Run the Chi-Square Test of Independence
chisq.test(tiktok_party_table)
```

```

    Pearson's Chi-squared test

data:  tiktok_party_table
X-squared = 24.15, df = 2, p-value = 5.702e-06
```

3. RQ 2: Independent Samples t-test

[The t-test is appropriate because we are comparing the mean of a **numeric variable** (platform_count) across **two groups**defined by a **binary categorical variable** (device_type_w144: PC vs Smartphone). This is exactly the scenario described in Slide 23.]

```
# Filter the data to include only PC (1) and Smartphone (2) users
device_data <- w144_wrangled %>%
  filter(device_type_w144 %in% c(1, 2))

# Independent samples t-test
t.test(platform_count ~ device_type_w144, data = device_data)
```

```

    Welch Two Sample t-test
```

```
data:  platform_count by device_type_w144
t = -5.5429, df = 384.97, p-value = 5.532e-08
alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
95 percent confidence interval:
 -0.7109869 -0.3386605
sample estimates:
mean in group 1 mean in group 2
       1.338384        1.863208
```

4. RQ 3: ANOVA & TukeyHSD

[ANOVA (`aov()`) is the correct test because we are comparing the mean of a **numeric variable** (platform_count) across **three groups** (Democrat, Republican, Independent). This extends beyond the 2-group t-test from RQ2.
TukeyHSD is used to determine **which** party groups differ from each other once ANOVA tells us a difference exists.]

```r
# Run the ANOVA
anova_model <- aov(platform_count ~ party_simple, data = w144_wrangled)

# Summary of model
summary(anova_model)
```

```r
# Tukey Post-Hoc Test
TukeyHSD(anova_model)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = platform_count ~ party_simple, data = w144_wrangled)

$party_simple
                                  diff        lwr         upr     p adj
Independent/Other-Democrat   -0.6397183 -1.2280348 -0.05140172 0.0292262
Republican-Democrat          -0.3605721 -0.5735772 -0.14756707 0.0002300
Republican-Independent/Other  0.2791461 -0.3105162  0.86880851 0.5068859
```