

GOVT 6019: Homework 6 — Inference

Fall 2026

Melissa Sands (Instructor)

Izza Malik (Teaching Assistant)

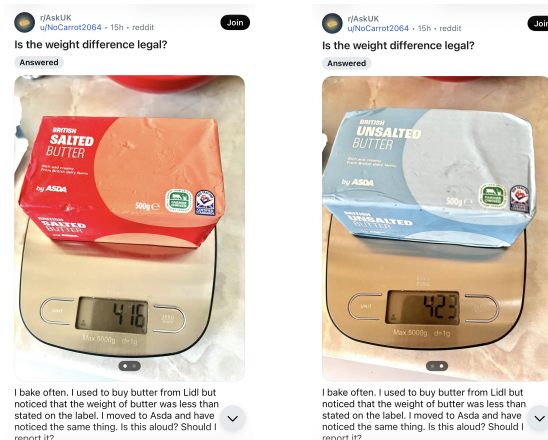
Question 1: True or false

State whether each claim is true or false. Give a one-sentence reason.

- Two 95% confidence intervals for the means of two independent groups overlap, so the difference between the group means is not statistically significant at $\alpha = 0.05$.
- Group A's 95% confidence interval contains group B's sample mean, so the difference between the two means is not statistically significant at $\alpha = 0.05$.
- A two-sided test gives $p = 0.03$, so there is a 3% chance that H_0 is true.
- With $n = 10$, a 95% confidence interval for μ that uses 1.96 in place of the t_9 critical value covers μ less than 95% of the time.
- A researcher runs a two-sided test and gets $p = 0.08$. After seeing that the estimate is positive, they switch to $H_1 : \mu > \mu_0$ and report $p = 0.04$. The chance of a Type I error is still 5%.
- Lowering α from 0.05 to 0.01 lowers the chance of a Type I error and raises the chance of a Type II error.

Question 2: The great British butter shortfall

Britain is in a cost-of-living crisis, and the price of the weekly shop is on everyone's mind. Reddit user u/NoCarrot2064 bought two 500 g blocks of Asda butter, one salted and one unsalted, to bake a Victoria sponge, a traditional British cake made with equal weights of butter, sugar, flour and eggs. Their kitchen scale disagreed with the packaging.



Assume the following.

- The scale is properly calibrated, and it weighs only the butter.
 - The weight of a block that Asda labels 500 g is a draw from a normal distribution with mean μ_S for salted butter and μ_U for unsalted butter.
 - The population standard deviation of block weights is known: $\sigma = 40$ g for both kinds.
 - Each block is one observation, so $n = 1$ for each kind.
- a. Write the null and alternative hypotheses for a two-sided test of whether the mean weight of Asda's salted butter differs from the 500 g on the package. Write the same pair of hypotheses for the unsalted butter.
 - b. Write the test statistic for each test. Why is this a z -test and not a t -test? Could you run a t -test with $n = 1$?
 - c. Compute the test statistic for each block. Use `qnorm()` to find the critical value at $\alpha = 0.05$. State the decision for each test.
 - d. Use `pnorm()` to compute the two-sided p -value for each test. State each conclusion in words about Asda's butter.
 - e. A shopper does not care whether Asda overfills its butter. Write the one-sided null and alternative hypotheses for each kind of butter that match that concern. Redo (c) and (d) for the one-sided tests. Which decisions change? Why?
 - f. UK law sets limits on short packs. Under the Weights and Measures (Packaged Goods) Regulations 2006, a 500 g pack has a *tolerable negative error* (TNE) of 15 g. A pack below 485 g is "non-standard." Trading Standards guidance says that no more than 2.5% of a packer's packs should be non-standard.
 - i. Suppose Asda fills its butter exactly as labelled, so $\mu = 500$, with $\sigma = 40$ g. Use `pnorm()` to find the share of blocks below 485 g. Would Asda meet the 2.5% guideline?
 - ii. Find the largest σ for which a packer with $\mu = 500$ meets the guideline.
 - iii. Redo the two-sided tests from (c) and (d) with that value of σ . Which decisions change? What does the change tell you about the role of the assumed σ in a z -test?
 - g. The Regulations add a second rule: no single pack may be short by more than twice the TNE. For a 500 g block, that means no block may weigh less than 470 g. Does either of u/NoCarrot2064's blocks break this rule? Do you need a hypothesis test to answer that question? Explain how this question differs from the question your tests in (a)–(e) answered.

Question 3: Working 9 to 5?

The 40-hour work week has been the U.S. standard since the Fair Labor Standards Act of 1938. The 2024 ANES asked respondents who had worked at some point in the past 12 months: "During the past 12 months, in the weeks worked, how many hours did you usually work each week?" The file `anes2024_hours.csv` holds one row per respondent: an ID and the answer. Empty cells are respondents who were not asked or who did not answer. Its codebook is `anes2024_hours_codebook.md`.

```
library(tidyverse)
hours <- read_csv("data/anes2024_hours.csv") |> # adjust to your folder layout
drop_na()
```

- a. How many respondents gave a number? Plot a histogram of `hours` and describe its shape. Then look at the largest answers. Are any of them impossible? Are any implausible? Explain. Keep every answer for the rest of this question.
- b. Compute the sample mean, the sample standard deviation and the estimated standard error of the mean. Use `qt()` to construct a 95% confidence interval for μ , the mean weekly hours of U.S. citizens of voting age who worked in the past 12 months. Report the margin of error.
- c. Interpret the interval in one or two sentences. What does the “95%” refer to?
- d. Test whether the mean work week differs from 40 hours. Go through each step:
 - i. State H_0 and H_1 .
 - ii. Choose $\alpha = 0.05$ and use `qt()` to find the critical value.
 - iii. Compute the t -statistic.
 - iv. Use `pt()` to compute the two-sided p -value.
 - v. State the decision and the conclusion in words.
 - vi. Check your answers with `t.test()`.
- e. Repeat steps (i)–(v) of (d) with the null value 39 in place of 40. Do your decisions in (d) and (e) match whether 40 and 39 lie inside the interval from (b)? State the general rule that connects a two-sided test at level α to a $1 - \alpha$ confidence interval.
- f. Suppose you had only these 15 respondents, drawn at random from the file:

52 10 32 42 10 25 80 45 50 38 50 48 24 30 40

Compute the 95% confidence interval for μ from these 15 answers, once with `qt()` and once with 1.96. Which interval is wider? Why? Then test $H_0: \mu = 40$ at $\alpha = 0.05$. This sample mean is farther from 40 than the full-sample mean. Explain why the decision differs from the one in

(d) anyway.

Question 4: Paired or independent?

In lecture we compared the 2024 ANES feeling thermometer ratings of Kamala Harris and Donald Trump with a two-sample t -test:

```
t.test(x = harris_ft, y = trump_ft)
```

Welch Two Sample t-test

```
data: harris_ft and trump_ft
t = 10.008, df = 10985, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 5.883011 8.748770
sample estimates:
mean of x mean of y
47.21404 39.89815
```

This test treats the Harris ratings and the Trump ratings as two independent samples. In the ANES, each respondent rated both candidates. The file `anes2024_ft.csv` holds one row per respondent: an ID and the respondent's Harris and Trump ratings (0-100). Empty cells are nonresponse. Its codebook is `anes2024_ft_codebook.md`.

```
library(tidyverse)
ft <- read_csv("data/anes2024_ft.csv") |> # adjust to your folder layout
drop_na()
```

- a. How many respondents rated both candidates? Compute the correlation between their Harris and Trump ratings.
- b. For respondent i , let H_i and T_i be the Harris and Trump ratings and $D_i = H_i - T_i$. The paired test works with the mean difference $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$. Show that $\bar{D} = \bar{H} - \bar{T}$, in steps:
 - i. Write $\bar{D}$ as a sum, substituting $H_i - T_i$ for D_i .
 - ii. Split that sum into two sums, one over the H_i and one over the T_i .
 - iii. Recognize each of the two terms as a sample mean.
 - iv. Check the result in R: compare `mean(ft$harris - ft$trump)` with `mean(ft$harris) - mean(ft$trump)`.

What does this result imply about the point estimates of the paired test and the two-sample test?

- c. The two tests differ in the variance they assign to that estimate.
 - i. For any two random variables, $\mathbb{V}[X - Y] = \mathbb{V}[X] + \mathbb{V}[Y] - 2\text{Cov}(X, Y)$. Apply the identity to $\mathbb{V}[\bar{D}] = \mathbb{V}[\bar{H} - \bar{T}]$.
 - ii. With n independent pairs, $\mathbb{V}[\bar{H}] = \sigma_H^2/n$, $\mathbb{V}[\bar{T}] = \sigma_T^2/n$ and $\text{Cov}(\bar{H}, \bar{T}) = \text{Cov}(H_i, T_i)/n$. Substitute these to write $\mathbb{V}[\bar{D}]$ in terms of σ_H^2 , σ_T^2 , $\text{Cov}(H_i, T_i)$ and n .
 - iii. Write the variance of the difference in means, $\mathbb{V}[\bar{H} - \bar{T}]$, in the two-sample (non-paired) test, with sample sizes n_H and n_T .
- d. State H_0 and H_1 for the paired test. Run the paired t -test in R. Report the estimated difference, its standard error (the `stderr` element of the result), the t -statistic and the 95% confidence interval. State the decision at $\alpha = 0.05$. Compare these to the two-sample (non-paired) output above. That output does not print a standard error: recover it as the estimated difference divided by t . Which quantities changed, and which did not?
- e. Using the sample standard deviations and the correlation from (a), compute both standard errors for the respondents in `ft`: the paired one from your formula in (c), and the two-sample one. Which is larger, and by what factor? Explain the gap in terms of the covariance term.
- f. The paired example in lecture compared respondents' Harris ratings before and after the election, where the correlation was 0.89. In that example, would a two-sample test overstate or understate the standard error of the mean difference? State the general rule: how does the sign of the correlation between paired measurements determine whether ignoring the pairing overstates or understates the standard error?
- g. Suppose an analyst runs a two-sample test on paired data whose correlation is negative. Which error rate is then larger than the analyst believes, Type I or Type II? Explain. For the Harris and Trump ratings, does the conclusion of the hypothesis test change?