Announcements

- Make sure to sign in on the google form (I send a list of what section questions are useful for what pset questions afterwards)
- Pset 4 due Friday 2/24

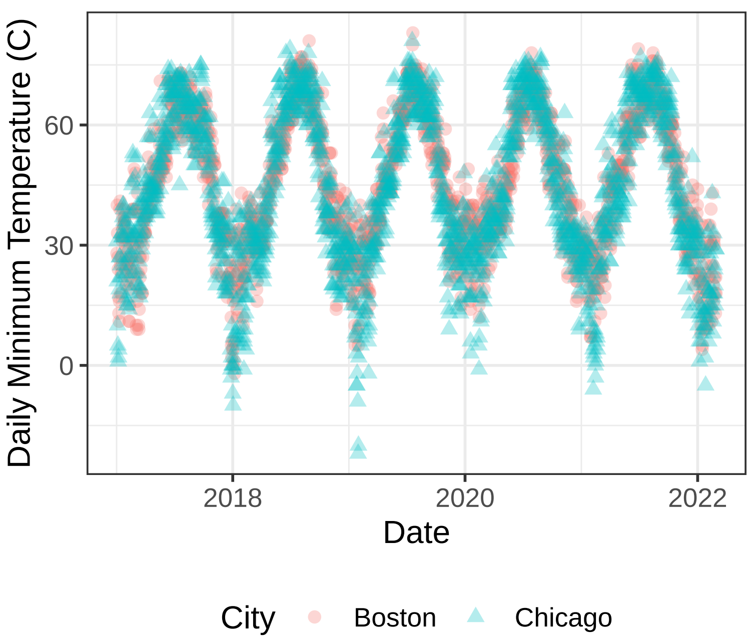
Kahneman (Warm-up)

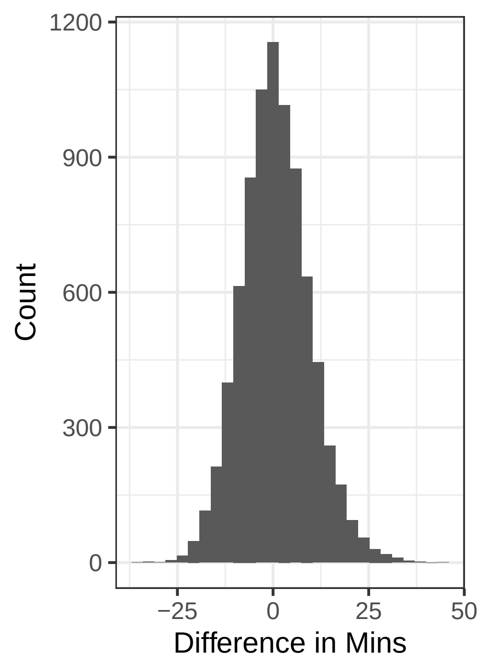
Without looking anything up, give 90% confidence intervals for the following quantities: (These are reproduced from Russo and Schoemaker 1989 but the answers are updated.)

Statement	Lower	Upper
Martin Luther King Jr.'s age at death		
Length of the Nile River		
Number of countries that are members of OPEC		
Number of books in the Catholic Old Testament		
Diameter of the moon		
Weight of an empty Boeing 747		
Year in which Wolfgang Amadeus Mozart was born		
Gestation period (in days) of an Asian elephant		
Air distance from London to Tokyo		
Deepest recorded point in the oceans		
Proportion of people taking this quiz in the room who got at least 8/10		

Brr

The following questions deal with the 2000-2022 temperatures of Boston and Chicago available here.





In this problem, we'd like to determine whether daily minimum temperatures are significantly different between Boston and Chicago. To do this, we'll explore the student-t distribution and confidence intervals. Our strategy will be to find the null distribution of some statistic assuming the true difference is 0 and then see how likely we are to have observed the crystallized version of that statistic under the null.

1. Let $Y_1, ..., Y_n \sim \mathcal{N}(0, \sigma^2)$ i.i.d. with $n = 8091$. Find the distribution of $\bar{Y}$ assuming σ^2 is known, and use that to give a standardized distribution for $\bar{Y}$.

2. Let $S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$ be the sample variance. Then, by 10.4.3 in the Stat 110 book, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. Show that the sample variance is independent of the sample mean by using facts about Multivariate Normals and the vector $(\bar{Y}, Y_1 - \bar{Y}, ..., Y_n - \bar{Y})$. (Hint: In a MVN vector, zero covariance implies independence. Also, a function of a random vector independent of a random variable is also independent of the random variable.)

3. Using the results above, write a function of $\bar{Y}$ and S^2 that has the t_{n-1} distribution (this is our pivot). Recall that the t_n distribution is defined as $\frac{Z}{\sqrt{V/n}}$ where $Z \sim \mathcal{N}(0, 1)$ and $V \sim \chi^2_n$ are independent.

4. In terms of a CDF F, determine $P\left(\frac{\bar{Y}}{\sqrt{S^2/n}} > \tau\right)$ for a fixed τ. Describe what this probability means in the context of the problem.

5. Using the pivot, find a 95% confidence interval for 0 and interpret what this means.

6. Using the data, compute this interval.

```
# TODO: Compute the interval
```

7. Based on your interval, comment on whether it seems likely that the data follows the stated distribution (i.e. that the true mean difference is 0).

8. Show that $t_n \xrightarrow{d} Z$ with $Z \sim \mathcal{N}(0, 1)$ as $n \to \infty$. (Hint: write the denominator as a sum of squared random variables and apply asymptotic tools.)

9. How close is the t_{n-1} interval above to an interval using the Standard Normal?

```
# TODO: Calculate the Standard Normal interval
```

10. Now, assume we have $Y_i \sim \mathcal{N}(\mu, \sigma^2)$ with σ^2 known. We have seen before that $\bar{Y}$ is the unbiased MLE for μ. Does $\bar{Y}$ achieve the Cramér-Rao lower bound? (The Cramér-Rao lower bound is the reciprocal of the Fisher information for the dataset for μ.)

Uniformity

1. One fast way of computing Normal-like random variables is to take the sum of 12 $\text{Unif}(0, 1)$ random variables and subtract 6. Find the expectation and variance of the resulting distribution and plot draws from it.

```
# TODO: Show draws
```

2. Now, suppose you have n $\text{Unif}(\alpha - \beta/2, \alpha + \beta/2)$ i.i.d. random variables. Write the likelihood function for α and β.

3. For the case of $\alpha = 10$ and $\beta = 5$, simulate $n = 100$ uniform random variables and use the `optim` function in R to estimate α and β from starting guesses of $(20, 100)$ by maximizing the likelihood function. How do these compare to the true values?

```
set.seed(111)
alpha = 10
beta = 5
n <- 100
x <- runif(n, alpha - beta/2, alpha + beta/2)

loglikelihood <- function(params) {
  # TODO: Write log likelihood in a way that can be optimized
}

# Initial guess
params <- c(20, 100)

# Run the optimization
optimization <- optim(params, loglikelihood, method="L-BFGS-B",
                      control=list(fnscale=-1), lower=c(-Inf,0), upper=c(Inf,Inf))
optimization$par
```

4. Find method of moments estimators for α and β and see how well these perform on the simulated data.

```
# TODO: Compute these for the sample
```

5. Another set of estimators is $\hat{\alpha} = Y_{(n/2+1/2)}$ and $\hat{\beta} = 2(Y_{(n/2+1/2)} - Y_{(1)})$. Describe the logic of these estimators and then find their biases and variances. Assume n is odd. (Note that $U_{(k)} - U_{(j)} \sim \text{Beta}(k-j, n-(k-j)+1)$ for Standard Uniforms.)

6. Make the $\hat{\beta}$ estimator unbiased.