Announcements

Make sure to sign in on the google form.

Pset 4 due. . .

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.
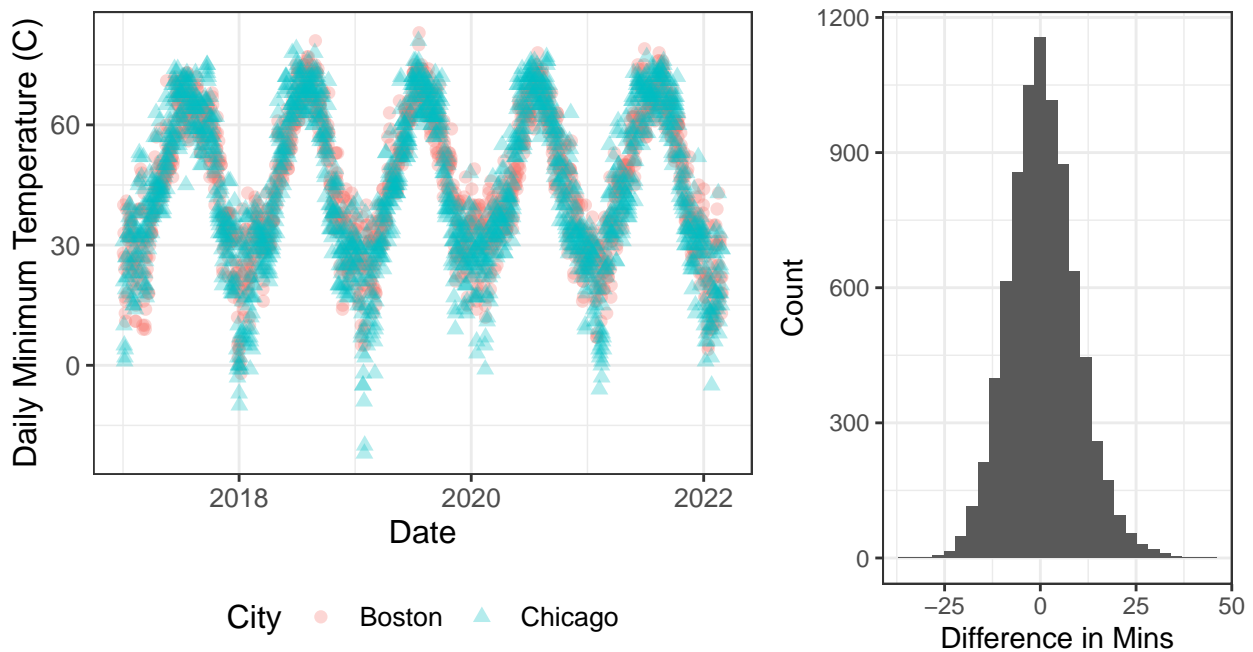
In this problem, we'd like to determine whether daily minimum temperatures are significantly different between Boston and Chicago. To do this, we'll explore the student-t distribution and confidence intervals. Our strategy will be to find the null distribution of some statistic assuming the true difference 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. The following is the previous interval computed from the data. Based on the interval, comment on whether it seems likely that the data follow the stated distribution (i.e. that the true mean difference is 0).

```
## Lower bound Upper bound
##   -0.9713260  -0.5763196
```

7. 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.)

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

```
## Lower bound Upper bound
##   -0.9712964  -0.5763491
```

9. 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.
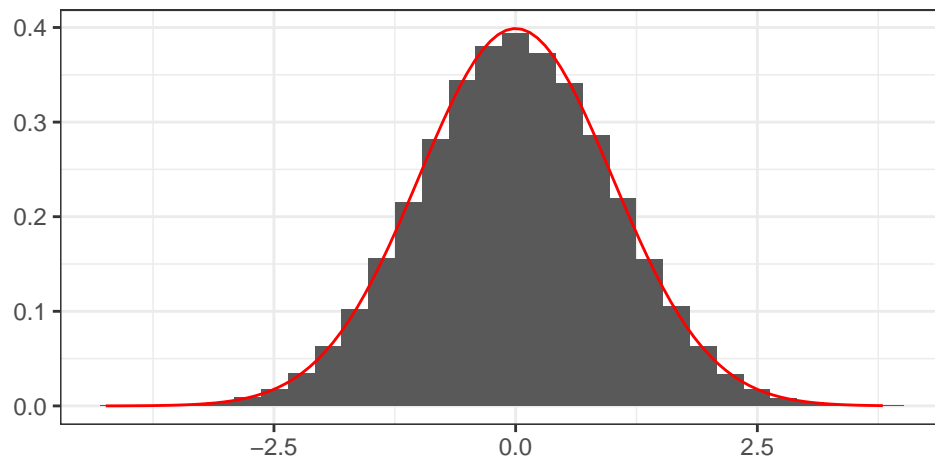
2. Now, suppose you have n $\text{Unif}(\alpha - \beta/2, \alpha + \beta/2)$ i.i.d. random variables. Write the likelihood function for α and β.

3. Consider trying to fine the MLE estimates with numerical optimization; that is, choose a starting point and iteratively evaluate the gradient of the log-likelihood and take a small step in that direction until the parameters stop updating. What might go wrong computationally?

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 $U_{(k)} - U_{(j)} \sim$ Beta$(k - j, n - (k - j) + 1)$ for Standard Uniforms.)

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