Winter 2023 Final Exam
Problem 16

Problem 16

We collect data on the play times of 100 games of *Chutes and Ladders* (sometimes known as *Snakes and Ladders*) and want to use this data to perform a hypothesis test.

Problem 16.1

Which of the following pairs of hypotheses can we test using this data?

Option 1: **Null Hypothesis:** In a random sample of Chutes and Ladders games, the average play time is 30 minutes. **Alternative Hypothesis:** In a random sample of Chutes and Ladders games, the average play time is not 30 minutes.

sample

Option 2: **Null Hypothesis:** In a random sample of Chutes and Ladders games, the average play time is not 30 minutes. **Alternative Hypothesis:** In a random sample of Chutes and Ladders games, the average play time is 30 minutes

Option 3: **Null Hypothesis:** A game of Chutes and Ladders takes, on average, 30 minutes to play. **Alternative Hypothesis:** A game of Chutes and Ladders does not take, on average, 30 minutes to play.

population

Option 4: **Null Hypothesis:** A game of Chutes and Ladders does not take, on average, 30 minutes to play. **Alternative Hypothesis:** A game of Chutes and Ladders takes, on average, 30 minutes to play.

- Option 1
- Option 2
- Option 3
- Option 4

Problem 16.2

We use our collected data to construct a 95% CLT-based confidence interval for the average play time of a game of *Chutes and Ladders*. This 95% confidence interval is [26.47, 28.47]. For the 100 games for which we collected data, what is the mean and standard deviation of the play times?

pop

sample

dist. of sample means

$\boxed{\text{mean of } 27.47 \quad \text{SD of } 5}$

Problem 16.3

Does the CLT say that the distribution of play times of the 100 games is roughly normal?

Key word! $\boxed{\cancel{\text{mean}}}$

jump $= \frac{1}{2}$

$$\text{SD of dist of sample means} = \frac{\text{SD of sample}}{\sqrt{\text{sample size}}}$$

- Yes
- No

Problem 16.4

Of the two hypotheses you selected in part (a), which one is better supported by the data?

- Null Hypothesis : takes 30 min on avg
- Alternative Hypothesis : $\neq$ 30 min on avg

$$\frac{1}{2} = \frac{x}{\sqrt{100}}$$

Winter 2024 Final Exam
Problem 9

did in 9am

Problem 9

In 2024, the Olympics will include breaking (also known as breakdancing) for the first time. The breaking competition will include **16 athletes**, who will compete in a single-elimination tournament.

In the first round, all 16 athletes will compete against an opponent in a face-to-face “battle". The 8 winners, as determined by the judges, will move on to the next round. Elimination continues until the final round contains just 2 competitors, and the winner of this final battle wins the tournament.

The table below shows how many competitors participate in each round:

Round	Competitors
1	16
2	8
3	4
4	2

After the 2024 Olympics, suppose we make a DataFrame called `breaking` containing information about the performance of each athlete during each round. `breaking` will have one row for each athlete's performance in each round that they participated. Therefore, there will be $16 + 8 + 4 + 2 =$ **30 rows** in `breaking`.

In the `"name"` column of `breaking`, we will record the athlete's name (which we'll assume to be unique), and in the other columns we'll record the judges' scores in the categories on which the athletes will be judged (creativity, personality, technique, variety, performativity, and musicality).

Problem 9.1

How many rows of `breaking` correspond to the winner of the tournament? Give your answer as an integer.

Problem 9.2

How many athletes' names appear exactly twice in the `"name"` column of `breaking`? Give your answer as an integer.

Problem 9.3

If we merge `breaking` with itself on the `"name"` column, how many rows will the resulting DataFrame have? Give your answer as an integer.

Hint: Parts (a) and (b) of this question are relevant to part (c).

Problem 9.4

Recall that the number of competitors in each round is $16, 8, 4, 2$. Write one line of code that evaluates to the array `np.array([16, 8, 4, 2])`. You **must use** `np.arange` in your solution, and you **may not use** `np.array` or the DataFrame `breaking`.

Winter 2024 Final Exam
Problem 11

Problem 11

Aladár Gerevich is a Hungarian fencer who is one of only two men to win Olympic medals 28 years apart. He earned 10 Olympic medals in total throughout his career: 7 gold, 1 silver, and 2 bronze. The table below shows the distribution of medal types for Aladár Gerevich, as well as a few other athletes who also earned 10 Olympic medals.

TVD

Athlete	Gold	Silver	Bronze
Aladár Gerevich	0.7	0.1	0.2
Katie Ledecky	0.7	0.3	0
Alexander Dityatin	0.3	0.6	0.1
Franziska van Almsick	0	0.4	0.6

0.2
0.5
0.7

Problem 11.2

Among the other athletes in the table above, whose medal distribution has the largest total variation distance (TVD) to Aladár Gerevich's distribution?

furthest

Problem 11.3

Suppose Pallavi earns 10 Olympic medals in such a way that the TVD between Pallavi's medal distribution and Aladár Gerevich's medal distribution is as large as possible. What is Pallavi's medal distribution?

TVD

Athlete	Gold	Silver	Bronze
Aladár Gerevich	0.7	0.1	0.2
Pallavi	x 0	y 1	z 0

0.9

Problem 11.4

ex.) [0.7, 0.1, 0.2]

More generally, suppose `medal_dist` is an array of length three representing an athlete's medal distribution. Which of the following expressions gives the maximum possible TVD between `medal_dist` and any other distribution?

[0 1 0]

- `medal_dist.max()`
- `medal_dist.min()`
- `1 - medal_dist.max()`
- `1 - medal_dist.min()`
- `np.abs(1 - medal_dist).sum()/2`

TVD 0.9

[0.6, 0.2, 0.2]

[0 1 0]

TVD 0.8