STUDENT NUMBER:

____

UNIVERSITY OF TORONTO
Faculty of Arts & Science

Fall 2024 MIDTERM EXAMINATION

STA130H1S Introduction to Statistical Reasoning and Data Science

Duration: 2 hours (completely unnecessary)
Aids Allowed: 1 Cheatsheet, non-programmable calculator

Exam Reminders:

- Fill out your first and last name, **UTORid**, and STUDENT NUMBER in the requested spaces on this page.
- Don't start writing the exam until the announcements have ended and the Exam Facilitator starts the exam.
- As a student, you help create a fair and inclusive writing environment. If you possess an unauthorized aid during an exam, you may be charged with an academic offence.
- Turn off and place all cell phones, smart watches, electronic devices, and unauthorized study materials in your bag under your desk. If it is left in your pocket, it may be an academic offence.
- When you are done your exam, raise your hand for someone to come and collect your exam. Do not collect your bag and jacket before your exam is handed in.
- If you are feeling ill and unable to finish your exam, please bring it to the attention of an Exam Facilitator so it can be recorded before leaving the exam hall.
- In the event of a fire alarm, do not check your cell phone when escorted outside.

Special Instructions:

All answers should be written on the exam paper itself in the space provided in the exam.

Students must hand in ALL examination materials and "cheatsheet"

0. To keep things fair, no questions about the exam will be answered during the duration of the exam. If you want to complain about what you assess to be problems with any one (or two, etc.) of the SICKSTY questions of this exam, by all means, do so here. However, what's gonna happen is I'm gonna look at how many students can correctly answer questions, type-o's and all, and I'll decide how to judge whether or not individual questions are "fine" as is or actually problematic. Don't get it twisted: there's SICKSTY questions here and essentially every one is GREAT or almost GREAT.

1. What are the defining *characteristics* of a **population** called?

   A. Arguments B. Distributions C. Parameters D. Statistics

2. What does i in $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ index?

   A. Observations B. Parameters C. Python `for` loops D. Rows

3. Which of the following best defines **inference**?

   A. Estimating a parameter using data B. Making a decision about a null hypothesis
   C. Visualizing differences using box plots D. Quantifying uncertainty about a parameter

4. Which is better and why?

   A. A 95% confidence interval because it allows you to make decisions about a null hypothesis
   B. A hypothesis test because it allows you to make decisions about a null hypothesis
   C. A hypothesis test because it allows you to make decisions about a parameter
   D. A 95% confidence interval because it quantifies uncertainty

5. Determine the most appropriate data structure for your Python program to use to manage data for the following three scenarios.

Scenario A You need to store a fixed set of coordinates (latitude, longitude) that should not be modified after initialization

Scenario B You need to maintain an ordered collection of items where duplicates are allowed and you might need to frequently add or remove items

Scenario C You need to store user profiles where each profile has a unique username as the key and their details (name, age, email) as values.

   A. Use a `tuple` for **Scenario A**, a `list` for **Scenario B**, and a `dict` for **Scenario C**
   B. Use a `list` for **Scenario A**, a `tuple` for **Scenario B**, and a `dict` for **Scenario C**
   C. Use a `dict` for **Scenario A**, a `tuple` for **Scenario B**, and a `list` for **Scenario C**
   D. Use a `tuple` for **Scenario A**, a `dict` for **Scenario B**, and a `list` for **Scenario C**

6. Which of the following gives the calculation of the **sample variance** for sample `x` an `np.array`?

   A. `(1/(len(x) - 1)) * ((x - x.mean()).sum())**2`
   B. `(((x - x.mean())**2).sum() / (len(x) - 1)).sqrt()`
   C. `((x - x.mean()) * (x - x.mean()) / (len(x) - 1)).sum()`
   D. None of the above

7. Which of the following describes the calculation of the **sample variance** for sample `x` an `np.array`?

   A. Square root of the sum of the values and divided by $n-1$

   B. Sum of differences between values and the sample mean divided by $n-1$

   C. Square root of sum of squared differences of the values from the sample mean divided by n

   D. Average of squared differences of values with sample mean multiplied by n and divided by $n-1$

8. Which of the following calculates **sample variance** with initialization `tot,xbar = 0,xs.mean()`?

   A.
   ```
   for x in xs:
       tot += x-xs.mean()
   (tot.sum()/(n-1))**2
   ```

   B.
   ```
   for x in xs:
       tot += (x-xs.mean())**2
   tot.sum()**2/(n-1)
   ```

   C.
   ```
   for x in xs:
       tot += (x-xs.mean())**2
   tot.sum()/(n-1)
   ```

   D.
   ```
   for x in xs:
       tot += x-xs.mean()
   np.sqrt(tot.sum()**2/(n-1))
   ```

9. Identify which function provides a comparable functionality to the `np.random.choice([1, 2, 3])` function which randomly selects one element from the `list [1, 2, 3]`
   (even the functionality based on `from scipy import stats` uses a slightly different output format).

   A. `stats.choice([1,2,3], p=3*[1/3])` B. `stats.binom(n=3, p=1/3).pdf(size=1)`

   C. `stats.coins(size=1).rvs(sides=3)` D. `stats.multinomial(n=1, p=3*[1/3]).rvs(size=1)`

10. Which **method** below creates samples from `scipy.stats` objects like `norm` and `gamma`?

    A. Independence B. `.rvs(size=n)` C. `.pdf(support)` D. `.random.choice()`

11. Given the following code snippets, which of the following statements is true?

```
# Code snippet 1     |   # Code snippet 2    |   # Code snippet 3 (run after others)
for x in range(5):   |   AList = "a b c"     |   for i,x in enumerate(AList.split()):
    print(x)         |   for i in AList:     |       print(i,x)
                     |       print(i)        |
```

    A. `Code snippet 1` iterates through integers 0 to 5; `Code snippet 2` prints THREE lines each with a single character; `Code snippet 3` iterates through the characters “abc” and their indexes

    B. `Code snippet 1` iterates through integers 1 to 5; `Code snippet 2` prints THREE lines each with a single character; `Code snippet 3` iterates through the characters “abc” and their indexes

    C. `Code snippet 1` iterates through integers 0 to 4; `Code snippet 2` prints FIVE lines each with a single character; `Code snippet 3` iterates through the characters “abc” and their indexes

    D. `Code snippet 1` fails since it uses `x` not `i`; `Code snippet 2` fails since `AList` is not a `list` type object; `Code snippet 3` fails if run after the others since they fail and so `AList` is never defined

12. Which of the following code snippets correctly counts how many of the words in the given paragraph are from the specified set of words using a loop?

```
paragraph = """Python is a versatile language that is widely used in data science.
               It is popular because of its simplicity and readability.
               Python syntax is clean and easy to learn."""
words_set = {"Python", "data", "syntax", "learning"}
count = 0
```

A.
```
for word.lower() in paragraph.split():
    if word in words_set.lower():
        count += 1
```

B.
```
for word in paragraph.split():
    if word in words_set:
        count += 1
```

C.
```
for word in words_set:
    if word in paragraph.split():
        count += 1
```

D.
```
for word in range(len(paragraph)):
    if word in words_set:
        count += 1
```

13. How does a chatbot generate the text that's an appropriate response to a user's input?

A. By using rule-based **logical flow control** systems of **if-else statements** to match user prompt input and conversational context to pre-written source documents using string search matching in order to determine the responses that have the best probability of answering the users query

B. By repeatedly picking **the best** next word for the response based on **conditional probabilities** of words that are likely to follow all words so far (starting with the words from the users prompt)

C. By **randomly** picking each subsequent word in the response **proportionally** to the **relative frequency** of words used to train the chatbot LLM to give answers for the topic of the prompt

D. By repeatedly **randomly** picking each subsequent word in the response **proportionally** to the **conditional probabilities** of words that are likely to follow the generated words so far (starting with the words of the prompt from the user)

14. Which of the following defines **statistical independence**?

A. $\Pr(A) = \Pr(B)$ B. $\Pr(A|B) = \Pr(A)$ C. $\Pr(A|B) = \Pr(B|A)$ D. None of these choices

15. What does " **A** | **B** " mean in the context of statements like $\Pr(A|B)$?

A. "or" as in "A or B"
B. "given" as in "A if B happened"
C. "and" as in "A and B"
D. "without" as in "A but not B"

16. What do we call a dataset that has a **longer tail** for it's **larger numeric values**?

A. A gamma distribution
B. Negatively or wrong-skewed
C. Multimodal outliers
D. Positively or right-skewed

17. We view **categorical** data using which type of figure?

    A. Bar plot B. Histogram
    C. Box plot D. Violin plot

18. Which of the following exactly tells you where the **sample median** is?

    A. Bar plot B. Histogram
    C. Box plot D. Kernel density estimate

19. What does **hypothesis testing prove**?

    A. Nothing B. The null hypothesis
    C. The alternative hypothesis
    D. The chance the null hypothesis is wrong

"You love to see it" – Remarks from "Prof Scott a Memoirs" quoted when speaking of what he loves.

Pie Chart: some Statisticians advise AGAINST the use of Pie Charts for Categorical Variables but mathematical counterexamples prove their utility.

20. For which of the following can you identify **outliers**?

    A. A histogram with a well-chosen number of bins
    B. A box plot in the context of comparing multiple box plots
    C. A kernel density estimate with a well-chosen kernel and bandwidth parameter
    D. All of the above but not all necessarily show modality

21. Which statement best describes the difference between a **histogram** and a **kernel density estimate (KDE)** or **violin plot**?

    A. A **KDE/violin plot** is more accurate than a histogram since it doesn't depend on arbitrarily choosing a number of bins
    B. A **KDE/violin plot** can provide a smooth representation of the empirical distribution of continuous data, nicely suggesting and reflecting its overall nature
    C. A **histogram** cannot be effectively used to represent continuous data but a **KDE/violin plot** is exactly for doing this
    D. Both present the same information so technically there's no difference between them

22. What is **inference** for a sample given by `x` an `np.array`?

    A. Using statistics like `x.mean()` and `x.std()` to parameterize a normal distribution
    B. The application of the `.fit(x)` method for some distribution from `scipy.stats`
    C. Reporting `np.quantile(means_bootstrapped_from_x, [0.10, 0.90])`
    D. None of these options

23. What is the essential idea of **bootstrapping**?

    A. Pretending a sample is a population   B. Simulating samples from the population distribution
    C. Increasing samples for more powerful inference   D. Visualizing estimating population samples

24. Why is a narrower confidence interval preferred for the same **confidence level**?

    A. It requires less samples   B. It is more reliable
    C. It requires more samples   D. It is more informative

25. How do we know the **actual true coverage rates** for a **bootstrapped confidence interval**?

    A. Sample size   B. Through simulation if we're willing to assume the samples population
    C. Confidence level   D. Based on the percentiles used for the `np.quantile` function

26. Which of the following will shorten the length of a **confidence interval**?

    A. Reducing the number of samples   B. Reducing the confidence level
    C. Increasing the confidence level   D. None of these options

27. Which of the following describes relationship we prefer regarding **observations** in a sample?

    A. Independence   B. Normally distributed
    B. Conditional probability   D. Multinomially distributed

28. Suppose we have a **histogram** of a sample from a **normal distribution**.
    Which of the following describes where about two-thirds (about $2/3 \times n$) of the n data points lie?

    A. In the middle 95% of the histogram bins
    B. Between the middle two-thirds of the histogram bins
    C. In bins at most couple sample standard deviations away, approximately, from the sample mean
    D. None of the above since the "two-thirds" rule is for normal *distributions* not n normal *samples*

29. What does a **box plot** with **outliers** tell you?

    A. The data is skewed
    B. The data is not normally distributed
    C. The data is not an independent sample
    D. There are data points in the data set that are more than an IQR and a half beyond a quartile

30. In a box plot, what does the **IQR (Interquartile Range)** represent?

    A. The range of the entire dataset   B. The spread of the middle fifty percent of a sample
    C. The spread of a samples fiftieth percentile   D. A 50% confidence interval for the sample mean

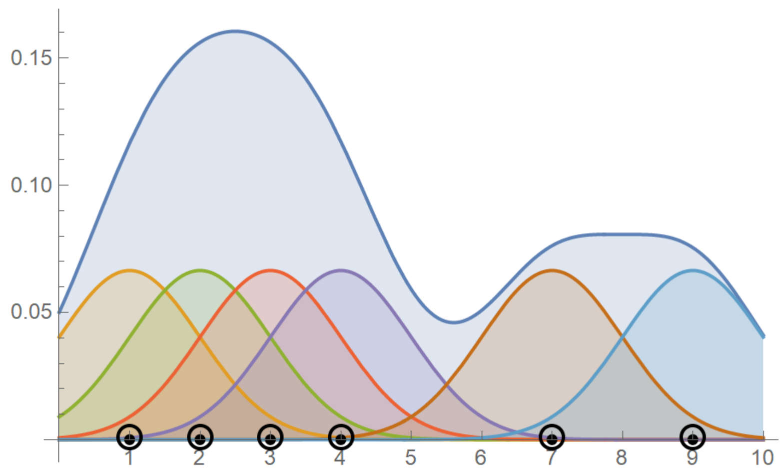
31. How can we simulate the **sampling distribution** of the sample mean?

    A. Bootstrapping   B. Theoretically with `.rvs()`   C. A and B   D. None of these option

32. Consider some data that has a long right tail. If you apply a **logarithmic transformation** to this dataset, what effect is likely to occur regarding the **mean** and **median** of the transformed data?

    A. The mean will decrease more than the median as the transformed distribution will be less skewed
    B. Change in mean and median will be similar since the tail and outliers will be scaled constantly (linearly) by a multiplicative factor
    C. The median will decrease more than the mean if the resulting data is more "normally" distributed
    D. The mean and median won't change drastically since the impact of the transformation is minimal when the data has large values

33. Which of the following options best describes the picture below.



    A. The construction of a 95% bootstrapped confidence interval given by the figure edges
    B. The spread and central tendency of a dataset as given by the Q1 through Q6 quartiles
    C. Normal distributions centered on data points additively representing an empirical distribution
    D. A violin plot with left and right sides reflecting different subsets of the data

34. Why do we use `replace=False` for **bootstrapping**?

    A. So our samples are independent
    B. So our samples are dependent
    C. So our samples reflect the uncertainty/variation at the sample size of our observed sample
    D. We do not use `replace=False` for bootstrapping

35. Why are **bootstrapped samples** the same sample size as our observed sample?

    A. So our samples are independent
    B. So our samples are dependent
    C. So our samples reflect the uncertainty/variation at the sample size of our observed sample
    D. We do not use the same sample size as our observed sample for bootstrapping

36. Which of the following can be used to do *boolean selection* with **logical conditionals**?

    A. `df[...]`
    B. `df[...]` and `df.loc[...]`
    C. `df.loc[...]` and `df.iloc[...]`
    D. `df[...]` and `df.loc[...]` and `df.iloc[...]`

37. How can **hypothesis testing** be carried out *using bootstrapped confidence intervals*?

    A. By accepting the hypothesized parameter value if it is contained in a confidence interval
    B. By rejecting a null hypothesis at a chosen confidence level if
    the associated parameter value is contained within the confidence interval
    C. By rejecting a null hypothesis at a chosen confidence level if
    the associated parameter value is not contained within the confidence interval
    D. Hypothesis testing cannot be done using confidence intervals since
    it must be based on a significance level not a confidence level

38. Assume you have run `from scipy import stats`, which of the following are **parameters** in the sense of being *characteristics* of a population, as opposed to statistics estimating parameters?

    A. `stats.gamma(shape=1, scale=1).rvs(size=n).mean()`
    B. `loc`, `scale`, `shape`, or similar arguments given some distribution from `scipy.stats`
    C. `stats.gamma(shape=1, scale=1).mean()` and `stats.gamma(shape=1, scale=1).std()`
    D. B and C

39. Which of the following is a **bootstrapped sampling distribution** for an observed sample of data?

    A.
    ```
    for i in range(N):
        xbars[i] = stats.gamma(a, loc, scale).rvs(size=n).mean()
    ```
    B.
    ```
    for i in range(N):
        xbars[i] = df.sample(n=n, replace=True)['x'].mean()
    ```
    C.
    ```
    for i in range(N):
        df['x'] = stats.gamma(a, loc, scale).rvs(size=n)
        for i in range(N):
            xbars[i] = df.sample(n=n, replace=True)['x'].mean()
    ```
    D.
    ```
    for i in range(N):
        xbars[i] = np.random.choice([0,1], size=n, replace=True).mean()
    ```

40. What does a **bootstrap confidence interval** estimate?

    A. Uncertainty regarding a sample based estimate of the true value of the population parameter
    B. The "capture" coverage rate of independently and identically distributed (i.i.d.) samples
    C. The chance the actual true parameter of a population is in the confidence interval
    D. Nothing it provides inference but it cannot be used to estimate anything

41. What is a **p-value**?

    A. The probability the null hypothesis is true
    B. The probability that the hypothesized parameter value is correct
    C. The chance that an analogous independent and identically distributed (i.i.d.) sample will
    result in an incorrect decision which wrongly rejects the null hypothesis
    D. None of the above

42. What is *the best* way to use a **p-value**?

   A. To determine the α-significance level of a hypothesis test
   B. To determine probabilities about the parameter and the null hypothesis
   C. To construct a corresponding $(1 - \alpha) \times 100\%$ confidence interval
   D. To characterize the strength of evidence against the null hypothesis

p-value	Evidence
$p > 0.1$	No evidence against the null hypothesis
$0.1 \geq p > 0.05$	Weak evidence against the null hypothesis
$0.05 \geq p > 0.01$	Moderate evidence against the null hypothesis
$0.01 \geq p > 0.001$	Strong evidence against the null hypothesis
$0.001 \geq p$	Very strong evidence against the null hypothesis

43. Which of the following is an example of a **nominal** variable?

   A. Age (in years)
   B. Happiness (on a 5-star rating scale)
   C. Gender (as a categorical variable)
   D. Height (in centimetres to one decimal place)

44. What is the primary purpose of a **box plot**?

   A. Display confidence intervals for sample statistics
   B. Show the distributional shape of a continuous variable
   C. Compare the relative prevalences of the empirical distributions of categorical data
   D. Give a quick rough comparison of distributional differences in a continuous and across the levels of a categorical variable, similarly to what can be done with the two "sides" of a violin plot

45. What's the difference between the "density" representation of a **kernel density estimator (KDE)** and the **probability density function (PDF)** of a **distribution** modelling a **population**?

   A. A **KDE** is the relative frequency of infinitely many (as opposed to n) samples from a population as a density
   B. A **PDF** is the relative frequency of infinitely many (as opposed to n) samples from a population as a density
   C. A **PDF** estimates the **empirical distribution** of a *sample* but a **KDE** estimates a *histogram*
   D. A **PDF** *estimates* a **KDE** which in turn *estimates a sample*

46. The area under the curve function estimated by a **kernel density estimate (KDE)** is guaranteed to be 1 since the **KDE** estimates the **probability density function** of a random variable based on its **empirical data distribution**. If analogously normalizing **histogram** so as to provide a **discrete approximation** of a **probability density function** based on the empirical data distribution results in bar heights of 1, 2, 3, and 4, how wide must the bins of the histogram be?

   A. 0.1 B. 1 C. 10 D. It depends

47. The % operator computes the remainder of a division. So if `i % 4 == 0` then `i` is exactly divisible by 4 (a multiple of 4). With this in mind, which of the following is true?

```
for i in range(1, 11):
    if i % 2 == 0:
        print("Wob", end=' ')
    if i % 2 == 0 and i % 3 == 0:
        print("Wobble", end=' ')
    elif i % 3 == 0:
        print("Bobble", end=' ')
    else:
        print(i, end=' ')
    print()
```

A.	B.	C.	D.
1	1	Wob	1
Wob	Wob 2	2 Wob	Wob 2
Bobble	Bobble	Bobble	Wobble
Wob	Wob 4	4 Wob	Wob 4
5	5	Wob	5
Wob	Wob Wobble	6 Wobble Bobble	Wob Wobble
7	7	Wob	7
Wob	Wob 8	8 Wob	Wob 8
Bobble	Bobble	Bobble	Wobble
Wob	Wob 10	10 Wob	10 Wob

48. If `x`, `y`, and `z` are `booleans`, what is `x and (y or z)`?

A. `x and y or z` B. `(x or y) and (x or z)`
C. `x or y and z` D. `(x and y) or (x and z)`

49. Which option below creates a column named `ticket_status` in the `pandas DataFrame` with columns

`age`	*integer* representing the person's age
`is_student`	*boolean* indicating if the person is a student
`vip_level`	*categorical* descriptor of the persons VIP status or `np.nan`

which (as in the example below) takes the value "Free Ticket" if the person is either a student and their age is under 18, or `vip_level` is not missing; but, takes the value "Regular Ticket" otherwise.

```
   age  is_student  is_vip   ticket_status
0   16        True     NaN     Free Ticket
1   25       False   Ultra     Free Ticket
2   30       False     NaN  Regular Ticket
3   22        True     NaN  Regular Ticket
```

```
df["ticket_status"] = "Regular Ticket"
```

A.
```
rule1 = (df["is_student"] & (df["age"] < 18))
rule2 = ~df["vip_level"].isna()
df.loc[ rule1 | rule2 , "ticket_status"] = "Free Ticket"
```

B.
```
rule1 = (df["is_student"] | (df["age"] < 18))
rule2 = df["vip_level"] != NaN
df.loc[ rule1 & rule2 , "ticket_status"] = "Free Ticket"
```

C.
```
rule1 = is_student and (age < 18)
rule2 = np.nonan(vip_level)
df.loc[ rule1 or rule2 , "ticket_status"] = "Regular Ticket"
```

D.
```
rule1 = df[is_student] = True
rule2 = df[age] < 18
df.loc[ rule1 | rule2 ~ df[vip_level].isnan() , "ticket_status"] = "Free Ticket"
```

50. Which of the options below are true about the following code?

```
# ‘synonyms‘ is a dict with np.array values of different lengths
# ‘synonyms_p‘ is as well with lengths matching ‘synonyms‘ for the same keys
for word in sentence:
    if word in stop_words:
        print(word)
    else:
        synonym_choice = stats.multinomial(1, p=synonyms_p[word]).rvs(1)[0] == 1
        print(synonyms[word][synonym_choice])
```

A. `sentence` and `stop_words` must be `list` objects B. `synonym_choice` is *independent* of `word`

C. `for` loop *iterations* end once `word in stop_words` is `True` D. None of these choices are true

51. Stella McStat's "Wheel of Destiny" has a "pointer" that can "land" on alternating "red" or "black" regions when the wheel is spun, presumably with a 50% chance for each of the two possible colors. Stella's setup is such that the wheel will be spun $n = 20$ times in game. Which of the following represents creating a **bootstrapped confidence interval** for the true chance wheel spins land red?

A. The proportion of red observed in the $n = 20$ spins in a game

B. An interval based on percentiles of "proportions of reds" seen over multiple $n = 20$ spin games

C. An interval based on percentiles of "proportions of reds" of $n = 20$ "spin outcomes" created by repeatedly resampling the results of the $n = 20$ spin outcomes of a single game with replacement

D. An "observed proportion plus and minus 2 standard errors" interval based on the standard error error of the bootstrapped sampling distribution of "proportion of reds" in n=20 spins

52. What would cause a bootstrapped confidence interval from the previous question to be unreliable?

A. Independent spins B. The sequential spins not affecting each others results

C. A skilled wheel spinner who can consistently "aim" their spins to land specific distances from the current wheel position and is trying to do so

D. A perfectly balanced wheel which when spun very hard has no systematic bias in final stopping location of the spin and which is always spun very hard

53. Assuming a 95% confidence interval is constructed for the the actual true chance that wheel spins land on red based on a game of n=20 spins, what does it mean?

    A. If 0.5 is not contained in the constructed interval then there's a 95% chance the wheel is not fair
    B. There's a 95% chance that actual true chance a wheel spin lands on red is in the constructed confidence interval
    C. There's a 95% chance that the next n=20 spin game will produce a proportion of red spins that is "captured" by this constructed confidence interval
    D. The 95% confidence statement suggests a 95% "capture" rate for analogously constructed confidence intervals based on n=20 spins, but without simulation this 95% chance is not guaranteed

54. Which option reflects the sampling distribution of the proportion of red spins in $n = 20$ spin games under a null hypothesis that the "Wheel of Destiny" is equally likely to land on either red or black?

    A. A histogram of "percentages of red" in n=20 spins created by repeatedly resampling the results of the $n = 20$ outcomes of a single game with replacement
    B. A histogram of "percentages of red" in $n = 20$ spins created over multiple games of n=20 spins
    C. A histogram of "percentages of heads" in $n = 20$ flips of a fair coin over multiple repeated sets of $n = 20$ flips
    D. None of the above

55. Which of the following describes a p-value for the Stella McStats "Wheel of Destiny" context for a game of n=20 spins?

    A. The number of samples generated from n=20 "fair equally like red and black" spins where the "proportion of reds" is farther from 0.5 (50%) than the "proportion of reds" observed in the game in question
    B. The number of samples generated from n=20 "fair equally like red and black" spins where the "proportion of reds" is greater than the "proportion of reds" observed in the game in question
    C. The chance of the assumption that the actual true "proportion of spins that are red" for the "Wheel of Destiny" is in fact 0.5 is actually wrong
    D. The number of bootstrapped sample proportions that are "as or more extreme" than 0.5 than the originally observed proportion

56. If we thought the "Wheel of Destiny" might actually be biased in favor of red, how might we change the nature of our null hypothesis?

    A. We could use a "one-sided" test that only considered "proportion of red spins" generated from a "fair wheel equally like to land on red and black" to be "as or more extreme" only if they were equal to or greater than the observed proportion in a game
    B. We could lower the "proportion of red spins" assumed by the null hypothesis so we are more likely to reject it and show therefore show that "proportion of red spins" is higher than 0.5
    C. We could increase our p-value by increasing the "proportion of red spins" assumed by the null
    D. We could try to reduce our p-value by using a "two-sided" test

57. How do we reject the null hypothesis that the “Wheel of Destiny” lands on red or black equally?

    A. Based on the p-value probability that this null hypothesis is wrong
    B. Based on the p-value probability that this null hypothesis is right
    C. Based on comparing the p-value to a predefined α-significance level and rejecting the null hypothesis at the α-significance level if it is smaller than α
    D. By setting the α to the p-value and rejecting the null hypothesis at the α-significance level

58. Suppose for a null hypothesis that the “Wheel of Destiny” was equally like to land on red and black we observed a p-value of 0.03. What is the nature of the evidence this provides for us?

    A. Weak evidence against the null hypothesis
    B. Moderate evidence against the null hypothesis
    C. Strong evidence against the null hypothesis
    D. Very strong evidence against the null hypothesis

59. How do we provide evidence in favor of the null hypothesis?

    A. Small p-values close to zero
    B. By failing to reject the null hypothesis
    C. Large p-values close to one
    D. We do not ever provide evidence in favor of the null hypothesis

60. Suppose we had set an α-significance level of 0.05. What would be possible about observing the previously mentioned p-value of 0.03 assuming it was generated from an experiment based on independent and identically distributed (i.i.d.) samples?

    A. We would wrongly reject the null hypothesis (making a Type I error) for 3% of analogous experiments of i.i.d. samples
    B. We would wrongly reject the null hypothesis (making a Type I error) for 5% of analogous experiments of i.i.d. samples
    C. We would wrongly fail to reject the null hypothesis (making a Type II error)
    D. We would wrongly fail to reject the null hypothesis (making a Type II error) at the $\alpha = 0.01$-significance level

What doesn’t kill me only makes me wish it had / I am gonna make it through this year if it kills me