PSTAT 5A: MT2 Practice Problems

Please also take a look at the past exams posted to the GitHub site, for additional practice problems. I also recommend you revisit past homework and quiz problems.

1. A recent study by Yale claims that, nationally (in the United States) 72% of people believe that climate change is a real phenomenon. To test these claims, Siobhan takes a representative sample of 100 US citizens and notes that 80 of these citizens believe climate change is real. Suppose that Siobhan wishes to use a 5% level of significance to test Yale's claims against an upper-tailed alternative.

   (a) Define the parameter of interest.

   (b) State the null and alternative hypotheses.

   (c) Compute the observed value of the test statistic.

   (d) Assuming the null is correct, what distribution does the test statistic follow? Be sure to check any relevant conditions.

   (e) What is the critical value for this test?

   (f) Conduct the test, and phrase the conclusions in the context of the problem.

2. The weight of a male kitten is found to be well-modeled by a normal distribution with unknown average but known standard deviation of 0.82lbs. A representative sample of 10 kittens is taken; these 10 kittens have an average weight of 5.2lbs. Suppose that we are interested in performing inference on the true average weight of a male kitten.

   (a) Define the parameter of interest.

   (b) What distribution would we use when constructing confidence intervals for the true average weight of a male kitten?

   (c) Now, construct an 81% confidence interval for the true average weight of a male kitten.

   (d) How would your answer to part (b) change (if at all!) if the weight of male kittens was *not* known to follow a normal distribution?

3. The length (in yards) of yarn contained in a randomly-selected ball of *GauchoKnit*-brand yarn is found to follow a normal distribution with mean 800 yards and a standard deviation of 100 yards.

   (a) What is the probability that a randomly-selected ball of *GauchoKnit*-brand yarn will contain between 750 and 900 yards of yarn?

(b) A sample of 12 balls of *GauchoKnit*-brand yarn is taken with replacement, and the number of these balls that have between 750 and 900 yards of yarn is recorded. What is the probability that this sample contains exactly 7 balls of yarn that contain between 750 and 900 yards of yarn? Make sure to clearly define any additional random variables you might need, and to check any relevant conditions!

4. The Transportation Security Administration (TSA) offers a service called *TSA PreCheck* which grants participants shorter wait times at airport security lines, along with a few other perks. Sam is interested in performing inference on the true proportion of travelers that have *TSA PreCheck*. To that end, they take a representative sample of 169 travelers and note that 123 of these travelers are enrolled in *TSA PreCheck*.

   (a) Define the parameter of interest, and call it p.

   (b) Define the random variable of interest, and call it $\widehat{P}$.

   (c) What is the distribution of $\widehat{P}$? Be sure to check any relevant conditions. Your answer may need to be left in terms of the parameter p.

   (d) Construct a 90% confidence interval for p, and interpret this interval in the context of the problem.

   (e) Would a 98% confidence interval for p be wider or narrower than your confidence interval from part (d) above? Explain briefly.

   (f) Construct a 98% confidence interval for p, and interpret this interval in the context of the problem.

5. Consider a random variable X with the following probability mass function (p.m.f.):

k	-2	-1	0	1	2
$\mathbb{P}(X = k)$	0.3	0.1	a	0.2	0.3

   where a is some as-of-yet unknown constant.

   (a) What is the value of a?

   (b) What is the state space of X?

   (c) If $F_X(x)$ denotes the cumulative distribution function (c.d.f.) of X at x, what is the value of $F_X(0.5)$?

   (d) What is $\mathbb{P}(X \geq 0)$?

   (e) What is $\mathbb{E}[X]$, the expected value of X?

   (f) What is $\mathrm{SD}(X)$, the standard deviation of X?

6. A multiple-choice quiz contains 10 questions. Johann has not studied for the quiz, and decides to guess the answers to these 10 questions. Assume there are 5 possible answer choices for each question, and that Johann's answers are independent across questions. The number of questions on the quiz that Johann gets correct is recorded.

   (a) Define the random variable of interest, and call it X.

   (b) What distribution does X follow? Be sure to check any/all relevant conditions!

   (c) What is the probability that Johann gets exactly half of the questions correct?

   (d) What is the probability that Johann gets between 5 and 7 questions (inclusive on both ends) correct?

   (e) What is the expected number of questions Johann will get correct?

   (f) What is the standard deviation of the number of questions Johann will get correct?

7. The amount of time it takes Jeannine to eat breakfast is uniformly distributed between 10 minutes and 25 minutes. A day is selected at random, and the amount of time it takes Jeannine to eat breakfast on this day is recorded.

   (a) Define the random variable of interest, and call it X.

   (b) Using proper notation, state the distribution of X. Be sure to include any/all relevant parameter(s)!

   (c) What is the probability that it will take Jeannine exactly 15 minutes to each breakfast on this day?

   (d) What is the probability that it will take Jeannine between 15 minutes and 30 minutes to eat breakfast on this day?

   (e) What is the standard deviation of the time (in minutes) it takes Jeannine to eat breakfast on a randomly-selected day?

   (f) (**Challenge**) If we know that it has taken Jeannine more than 15 minutes to eat breakfast on this day, what is the probability that it took her more than 20 minutes to eat breakfast on this day?

Multiple Choice Questions

8. **True or False:** The expected value of a random variable must be an element of the random variable's state space.

   A. True

   B. False

9. Indrani would like to compute the standard deviation of the list `[1, 2, 3]` using our familiar formula for standard deviation

$$s_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2}$$

To that effect, she imports all functions from the `numpy` module (and does **not** import the `numpy` module with any nickname) and runs `std([1, 2, 3])`. Will this give her the desired output?

    A. Yes, the code will yield the desired output.

    B. No, because Indrani needs to write `numpy.std([1, 2, 3])`. The code, as she currently has it written, will result in an error.

    C. No, because Indrani needs to write `std(1, 2, 3)`. The code, as she currently has it written, will result in an error.

   D. No, because Indrani needs to include `ddof = 1` in her call to `std`.

Problems 10 - 13 refer to the following situation: Markus would like to plot the function $f(x) = \cos(x)$ between $x = -\pi$ and $x = \pi$. To that end, he has written the following code (assume there is nothing written before this code):

```
%matplotlib inline
import matplotlib
import matplotlib.pyplot as ___Blank 1___
plt.style.use('seaborn-v0_8-whitegrid')

___Blank 2___ numpy ___Blank 3___ np

x_grid = np.linspace(-np.pi, np.pi, 10)
plt.plot(x_grid, np.cos(x_grid));
```

This code, after filling in the blanks appropriately, has resulted in the following output:
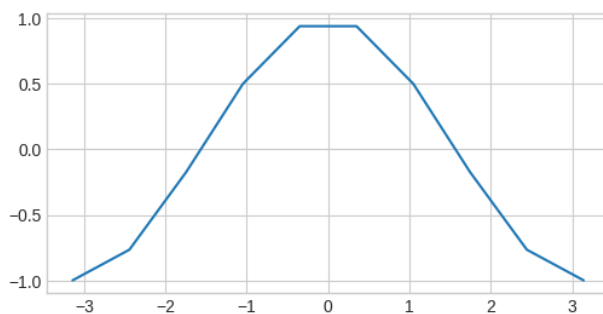
10. What should go in Blank 1?

A. `matplotlib.pyplot`

B. `pyplot`

C. `plt`

D. `mtpltlbpplt`

E. None of the above

11. What should go in Blank 2?

     A. **import**

     B. **load**

     C. **store_module**

     D. `*`

     E. None of the above

12. What should go in Blank 3?

     A. `*`

     B. **as**

     C. **if**

     D. **elif**

     E. None of the above

13. Note that the resulting plot is quite "jagged." Markus would like to fix that, and make the resulting plot smoother without changing the $x-$ and $y-$limits. Which of the following will achieve that?

     A. Change the `1` in his call to `np.linspace()` to a larger number; e.g. `100`.

     B. Change the `0` in his call to `np.linspace()` to a larger number; e.g. `100`.

     C. Change the `10` in his call to `np.linspace()` to a larger number; e.g. `100`.

     D. None of the above

....

14. What will be the result of running `scipy.stats.t.ppf(0.05, df = 17)`? Assume all functions from the `scipy.stats` module have been imported, and that the `scipy.stats` module has <u>not</u> been imported with any nickname.

     A. 1.33

  B. 1.74

  C. 2.11

  D. 2.57

  E. 2.90

15. Suppose $Z \sim \mathcal{N}(0, 1)$. What is the value of c such that $\mathbb{P}(Z > c) = 0.0594$?

  A. -2.01

  B. -1.56

  C. 1.56

  D. 2.01

  E. None of the above