Name: _____ **Date:** _____

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.

   Solution: Let p denote the proportion of US citizens that believe climate change is real.

   (b) State the null and alternative hypotheses.

   Solution:

   $$\begin{bmatrix} H_0: & p = 0.72 \\ H_A: & p > 0.72 \end{bmatrix}$$

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

   Solution: First note that $\widehat{p} = (80/100) = 0.8$. Thus,

   $$\text{ts} = \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.8 - 0.72}{\sqrt{\frac{(0.72)\cdot(1-0.72)}{100}}} \approx 1.78$$

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

   Solution: We first check:

   1) $np_0 = (100)(0.72) = 72 \geq 10$ ✓

   2) $n(1 - p_0) = 100(1 - 0.72) = 28 \geq 10$ ✓

Since both of these conditions are met, we conclude that

$$\text{TS} \overset{H_0}{\sim} \mathcal{N}(0,\ 1)$$

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

Solution: Recall that, for an upper-tailed test at an α level of significance, the critical value is the $(1-\alpha) \times 100^{\text{th}}$ percentile of the standard normal distribution. Since $\alpha = 0.05$, we can use either a table or Python to see that this value is around 1.645.

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

Solution: We reject an upper-tailed test when the value of the test statistic exceeds the critical value. In this case, $\text{ts} = 1.78 > 1.645$ meaning we *reject* the null:

> At a 5% level of significance, there was sufficient evidence to reject Yale's claims that 72% of US citizens believe in climate change, against the alternative that the true proportion of US citizens that believe in climate change is greater than 72%.

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.

Solution: Let μ denote the true average weight of a male kitten.

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

Solution: The first question we ask is whether or not the population (or, more specifically, values in the population) are normally distributed. The problem states that weights of adult male kittens (i.e. weights in the population) are normally distributed; hence, by our flowchart from Lecture 12, we know that we will use the standard normal distribution.

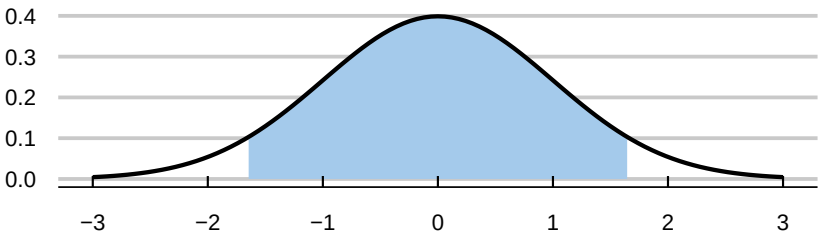
(c) Now, construct an 81% confidence interval for the true average weight of a male kitten.

Name:_____ **Date:**_____

Solution: A confidence interval, using the normal distribution, for a population mean μ will take the form

$$\overline{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}$$

where z^* is the confidence coefficient, n is the sample size, and σ is the population standard deviation. We know $\sigma = 0.82$ (given in the problem statement) and we know $n = 10$; as such, all that remains is to find the value of z^*. Since we want an 81% confidence interval, we seek the value z^* such that the blue shaded area below is 81%:



This tells us that the tails, separately, must have area $(1 - 0.81)/2 = 0.095$ meaning we seek either negative one times the 9.5^{th} percentile, or the $(100-9.5) = 90.5^{\text{th}}$ percentile. Either way we find $z^* \approx 1.31$, meaning our confidence interval becomes

$$(5.2) \pm (1.31) \cdot \frac{0.82}{\sqrt{10}} \approx [4.86\,,\ 5.54]$$

One interpretation of this interval is as follows:

> We are 95% confident that the true average weight of male kittens is between 4.86lbs and 5.54 lbs.

(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?

Solution: If the population were *not* normally distributed (or, if we couldn't assume that it was), we would then have to ask whether our sample size was "large enough". By the conventions in this class, "large enough" means $n \geq 30$; in this case, $n = 10 < 30$, so we would actually not be able to proceed constructing a confidence interval at all!

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?

Solution: Let X denote the amount of yarn contained in a randomly-selected ball of *GauchoKnit*-brand yarn; then, from the problem statement, we have $X \sim \mathcal{N}(800,\ 100)$. We seek $\mathbb{P}(750 \leq X \leq 900)$, which we find using our standard procedure:

$$\begin{aligned}
\mathbb{P}(750 \leq X \leq 900) &= \mathbb{P}(X \leq 900) - \mathbb{P}(X \leq 750) \\
&= \mathbb{P}\left(\frac{X-800}{100} \leq \frac{750-800}{100}\right) - \mathbb{P}\left(\frac{900-800}{100} \leq \frac{750-800}{100}\right) \\
&= \mathbb{P}\left(\frac{X-800}{100} \leq -0.5\right) - \mathbb{P}\left(\frac{900-800}{100} \leq 1\right) \\
&= 0.8413 - 0.3085 = \boxed{0.5328}
\end{aligned}$$

(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!

Solution: Let Y denote the number of balls of yarn, in the sample of 12, that contain between 750 and 900 yards of yarn. We suspect Y may be binomially distributed; to verify, we check the four Binomial Conditions:

1) **Independent trials?** Yes, since balls are selected with replacement.
2) **Fixed number of trials?** Yes; $n = 12$.
3) **Well-defined notion of "success"?** Yes; "success" = "ball contains between 750 and 900 yards of yarn"
4) **Fixed probability of success?** Yes; $p = 0.5328$, as computed in the previous part.

Since all four conditions are satisfied, we have that $Y \sim \text{Bin}(12,\ 0.5328)$, and so

$$\mathbb{P}(Y = 7) = \binom{12}{7}(0.5328)^7(1-0.5328)^{12-7} \approx \boxed{0.2148}$$

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.

> **Solution:** Let p denote the proportion of all US citizens that have enrolled in *TSA PreCheck*.

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

> **Solution:** Let $\widehat{P}$ denote the proportion of people in a representative sample of 169 residents that have enrolled in *TSA PreCheck*.

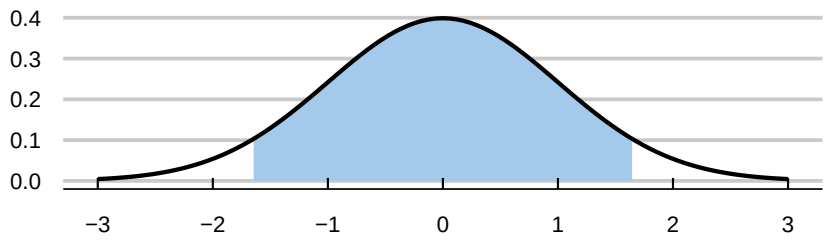
(c) What is the distribution of $\widehat{P}$? Be sure to check any relevant conditions. Your answer may need to be left in terms of the parameter p.

> **Solution:** Note that we do not know the value of p; all we know is $\widehat{p} = (123/169) \approx 0.7278$. Hence, we check the substitution approximation to the success-failure conditions:
>
> **1)** $n\widehat{p} = (169) \cdot (123/169) = 123 \geq 10$ ✓
>
> **2)** $n(1-\widehat{p}) = (169) \cdot (46/169) = 46 \geq 10$ ✓
>
> Since both conditions are satisfied, we can invoke the Central Limit Theorem for Proportions to conclude
>
> $$\widehat{P} \sim \mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{169}}\right)$$

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

> **Solution:** We know that our confidence interval will take the form
>
> $$\widehat{p} \pm z^* \cdot \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$$
>
> Since we wish to utilize a 90% confidence level, we select z^* to be the value such that the blue shaded area below is 90%:

Name:_____ **Date:**_____



This tells us that the tails, separately, must have area $(1 - 0.90)/2 = 0.05$ meaning we seek either negative one times the 5^{th} percentile, or the $(100-5) = 95^{\text{th}}$ percentile. Either way we find $z^* \approx 1.645$, meaning our confidence interval becomes

$$\left(\frac{123}{169}\right) \pm 1.645 \cdot \sqrt{\frac{\left(\frac{123}{169}\right) \cdot \left(1 - \frac{123}{169}\right)}{169}} = [0.6715\,,\ 0.7841]$$

One interpretation of this interval is:

> We are 90% confident that the true proportion of US citizens enrolled in *TSA PreCheck* is between 67.15% and 78.41%.

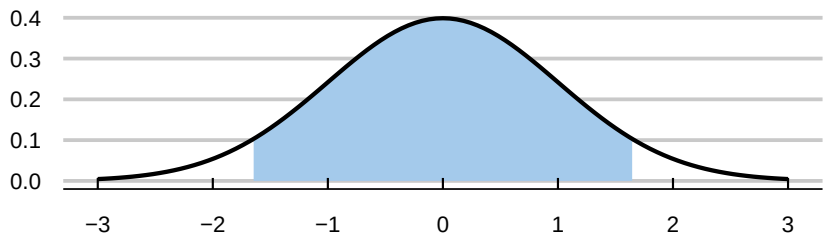
(e) Would a 98% confidence interval for p be wider or narrower than your confidence interval from part (d) above? Explain briefly.

Solution: We know that higher confidence levels correspond to wider confidence levels (again, think in terms of the fishing analogy from Lecture 11). Since 98% is larger than 90%, we would expect a 98% confidence interval to be wider than the 90% interval constructed in the previous part.

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

Solution: The only difference between our work for this part and our work for part (d) above is that the confidence coefficient z^* will change. Specifically, since we are now utilizing 98% confidence level we select z^* to be the value such that the area shaded below is 98%:

Name: _____ **Date:** _____



This tells us that the tails, separately, must have area $(1 - 0.98)/2 = 0.01$ meaning we seek either negative one times the first percentile, or the $(100 - 1) = 99^{\text{th}}$ percentile. Either way we find $z^* \approx 2.33$, meaning our confidence interval becomes

$$\left(\frac{123}{169}\right) \pm 2.33 \cdot \sqrt{\frac{\left(\frac{123}{169}\right) \cdot \left(1 - \frac{123}{169}\right)}{169}} = [0.6480\ ,\ 0.8076]$$

One interpretation of this interval is:

> We are 98% confident that the true proportion of US citizens enrolled in *TSA PreCheck* is between 64.8% and 80.76%.

As a quick sanity check, note that this interval is indeed wider than our interval from part (d) above.

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?

Solution: We know that the probability values in a p.m.f. must sum to 1. This gives

$$(0.3) + (0.1) + a + (0.2) + (0.3) = 1$$

which in turn gives

$$a = 1 - (0.3 + 0.1 + 0.2 + 0.3) = 0.1$$

(b) What is the state space of X?

Solution: The state space is the first row of the p.m.f. table (definitionally, it is the set of values for which the p.m.f. is nonzero):

$$S_X = \{-2,\ -1,\ \ 0,\ \ 1,\ \ 2\}$$

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

Solution: Using direct computation, we would compute

$$\begin{aligned} F_X(0.5) &= \mathbb{P}(X \leq 0.5) \\ &= \mathbb{P}(X = -2) + \mathbb{P}(X = -1) + \mathbb{P}(X = 0) \\ &= 0.3 + 0.1 + 0.1 = 0.5 \end{aligned}$$

We could have also used the complement rule:

$$\begin{aligned} F_X(0.5) &= \mathbb{P}(X \leq 0.5) \\ &= 1 - \mathbb{P}(X > 0.5) \\ &= 1 - [\mathbb{P}(X = 1) + \mathbb{P}(X = 2)] \\ &= 1 - (0.2 + 0.3) = 0.5 \end{aligned}$$

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

Solution: By direct computation,

$$\begin{aligned} \mathbb{P}(X \geq 0) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) \\ &= 0.1 + 0.2 + 0.3 = 0.6 \end{aligned}$$

Or, we could have (again) used the complement rule:

$$\begin{aligned} \mathbb{P}(X \geq 0) &= 1 - \mathbb{P}(X < 0) \\ &= 1 - [\mathbb{P}(X = -2) + \mathbb{P}(X = -1)] \\ &= 1 - (0.3 + 0.1) = 0.6 \end{aligned}$$

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

Name:_____ **Date:**_____

Solution:

$$\begin{aligned}\mathbb{E}[X] &= \sum_{\text{all } k} k \cdot \mathbb{P}(X = k) \\ &= (-2) \cdot \mathbb{P}(X = -2) + (-1) \cdot \mathbb{P}(X = -1) + (0) \cdot \mathbb{P}(X = 0) + (1) \cdot \mathbb{P}(X = 1) \\ &\qquad + (2) \cdot \mathbb{P}(X = 2) \\ &= (-2) \cdot (0.3) + (-1) \cdot (0.1) + (0) \cdot (0.1) + (1) \cdot (0.2) + (2) \cdot (0.3) \\ &= 0.1\end{aligned}$$

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

Solution: We first need to find $\text{Var}(X)$, the variance of X. Recall that we have two formulas at our disposal for computing the variance. Using the second formula for variance, we would first compute

$$\begin{aligned}\sum_{\text{all } k} k^2 \cdot \mathbb{P}(X = k) &= (-2) \cdot \mathbb{P}(X = -2)^2 + (-1)^2 \cdot \mathbb{P}(X = -1) + (0)^2 \cdot \mathbb{P}(X = 0) \\ &\qquad + (1)^2 \cdot \mathbb{P}(X = 1) + (2)^2 \cdot \mathbb{P}(X = 2) \\ &= (-2)^2 \cdot (0.3) + (-1)^2 \cdot (0.1) + (0)^2 \cdot (0.1) + (1)^2 \cdot (0.2) \\ &\qquad + (2)^2 \cdot (0.3) = 2.7\end{aligned}$$

and so

$$\text{Var}(X) = \left(\sum_{\text{all } k} k^2 \cdot \mathbb{P}(X = k) \right) - (\mathbb{E}[X])^2 = 2.7 - (0.1)^2 = 2.69$$

If, instead, we used the first formula for variance, we would have

$$\begin{aligned}\text{Var}(X) &= \sum_{\text{all } k} (k - \mathbb{E}[X])^2 \cdot \mathbb{P}(X = k) \\ &= (-2 - 0.1) \cdot \mathbb{P}(X = -2)^2 + (-1 - 0.1)^2 \cdot \mathbb{P}(X = -1) + (0 - 0.1)^2 \cdot \mathbb{P}(X = 0) \\ &\qquad + (1 - 0.1)^2 \cdot \mathbb{P}(X = 1) + (2 - 0.1)^2 \cdot \mathbb{P}(X = 2) \\ &= (-2 - 0.1)^2 \cdot (0.3) + (-1 - 0.1)^2 \cdot (0.1) + (0 - 0.1)^2 \cdot (0.1) \\ &\qquad + (1 - 0.1)^2 \cdot (0.2) + (2 - 0.1)^2 \cdot (0.3) = 2.69\end{aligned}$$

Either way we have Var$9X) = 2.69$, and so

$$\text{SD}(X) = \sqrt{\text{Var}(X)} = \sqrt{2.69} \approx 1.6401$$

Name:_____ **Date:**_____

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.

   Solution: Let X denote the number of questions, out of the 10 total quesetions, that Johann gets correct.

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

   Solution: We suspect X to be binomally distributed. To verify this, we check the four Binomial Conditions:

   1) Independent Trials? Yes, since we are told that Johann's answers are independent across questions.

   2) Fixed number of Trials? Yes; $n = 10$ trials (since each trial corresponds to a given question on the exam, and checking whether or not Johann got that question correct.)

   3) Well-Defined Notion of "Success?" Yes; on any given trial, "success" = "Johann got the question correct"

   4) Fixed Probability of Success? Yes; $p = 1/5 = 0.2$, since there are 5 answer choices for each question, only one of which is correct, and since Johann is guessing randomly we can use the Classical Approach to Probability to compute the probability of Johann selecting the correct answer to be $1/5$.

   Since all 4 conditions are satisfied, we conclude that

   $$X \sim \text{Bin}(10,\ 0.2)$$

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

   Solution: We seek $\mathbb{P}(X = 5)$, which we can compute using the formula for the p.m.f. of the Binomial distributions:

   $$\mathbb{P}(X = 5) = \binom{10}{5}(0.2)^5(1 - 0.2)^{10-5} \approx 0.0264$$

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

Solution: We seek $\mathbb{P}(5 \leq X \leq 7)$, which we compute as

$$\begin{aligned}\mathbb{P}(5 \leq X \leq 7) &= \mathbb{P}(X = 5) + \mathbb{P}(X = 6) + \mathbb{P}(X = 7) \\ &= \binom{10}{5}(0.2)^5(1-0.2)^{10-5} + \binom{10}{6}(0.2)^6(1-0.2)^{10-6} + \binom{10}{7}(0.2)^7(1-0.2)^7 \\ &\approx 0.0327\end{aligned}$$

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

Solution:

$$\mathbb{E}[X] = np = (10) \cdot (0.2) = 2$$

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

Solution:

$$\text{SD}(X) = \sqrt{np(1-p)} = \sqrt{(10)(0.2)(0.8)} \approx 1.2649$$

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.

Solution: Let X denote the time (in minutes) it takes Jeannine to eat breakfast on a randomly-selected day.

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

Solution: $X \sim \text{Unif}(10,\ 25)$

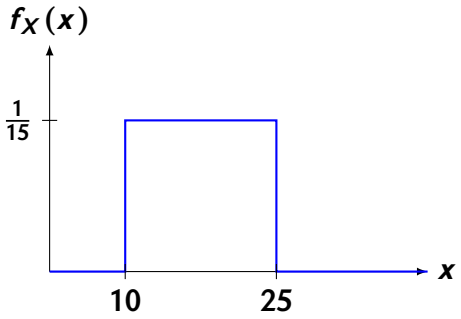
(c) What is the probability that it will take Jeannine exactly 15 minutes to each breakfast on this day?

Name:_____ **Date:**_____

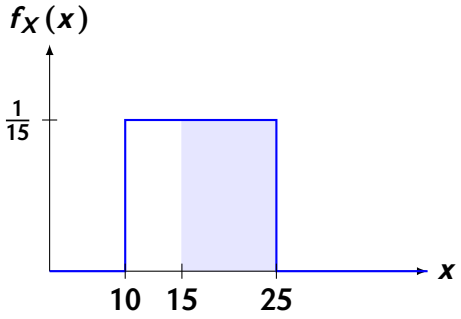
Solution: We seek $\mathbb{P}(X = 15)$. Since X is continuous, we know that $\mathbb{P}(X = k) = 0$ for any value of k; as such, the desired probability is $\boxed{0}$.

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

Solution: We seek $\mathbb{P}(15 \leq X \leq 30)$. Let's sketch a picture of the density curve of X, first:



Since the state space of X only extends to 25, the region whose area we seek is



$$\implies (25 - 15) \cdot \frac{1}{15} = \frac{10}{15} = \boxed{\frac{2}{3}}$$

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

Solution:

$$\text{SD}(X) = \frac{b-a}{\sqrt{12}} = \frac{25-10}{\sqrt{12}} = \boxed{\frac{15}{\sqrt{12}} \approx 4.33}$$

(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?

Solution: We seek $\mathbb{P}(X \geq 20 \mid X \geq 15)$. Using the definition of conditional probability, we can write this as

$$\mathbb{P}(X \geq 20 \mid X \geq 15) = \frac{\mathbb{P}(\{X \geq 20\} \cap \{X \geq 15\})}{\mathbb{P}(X \geq 15)}$$

Let's examine the numerator; the event $\{X \geq 20\} \cap \{X \geq 15\}$ says "X was greater than 20 and X was greater than 15". We see that this is equivalent to saying "X was greater than 20", meaning

$$\begin{aligned}\mathbb{P}(X \geq 20 \mid X \geq 15) &= \frac{\mathbb{P}(\{X \geq 20\} \cap \{X \geq 15\})}{\mathbb{P}(X \geq 15)} \\ &= \frac{\mathbb{P}(X \geq 20)}{\mathbb{P}(X \geq 15)} = \frac{\left(\frac{5}{15}\right)}{\left(\frac{10}{15}\right)} = \frac{1}{2}\end{aligned}$$

where the numerator and denominator of the last equation above can be computed by sketching a picture and finding a relevant area.

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`.

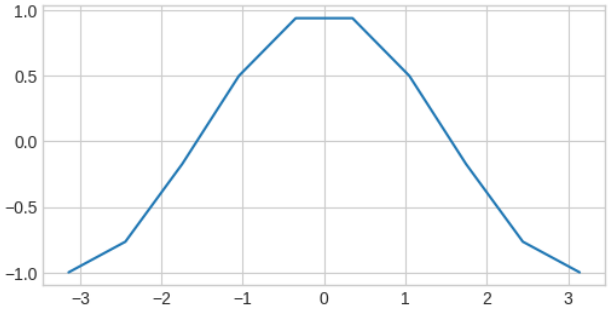
Problems 10 - 13 refer to the following situation: Markus would like to plot the function $f(x) = \cos(x)$ between $x = -\pi$ and $x = \pi$. To that end, he has written the following code (assume there is nothing written before this code):

```
%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 not 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