University of Texas at Austin, Department of Mathematics
M358K - Applied Statistics

Practice for In-term exam II

Definitions.

Problem 1.1. (10 points) Provide the definition of the *probability density function* of a **continuous** random variable.

Solution: The *probability density function* of a continuous random variable X is defined as

$$f_X(x) = F'_X(x) \quad \text{for all } x \text{ where the derivative exists}$$

where F_X stands for the cumulative distribution function.

Problem 1.2. (10 points) Provide the expression for the *probability density function* of a **standard normal** random variable.

Solution:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad x \in \mathbb{R}.$$

True/False Questions.

Problem 1.3. (2 points) The mean and median of any normal distribution are equal. *True or false?*

Solution: TRUE
....since both the mean and the median are equal to the parameter μ.

Problem 1.4. (2 points) The margin of error for a confidence interval for the population mean μ increases as the sample size increases. *True or false?*

Solution: FALSE

Problem 1.5. (2 points) Resident statistician Margie N. Rivera calculated a confidence interval of $[-0.56, 0.88]$. Her assistant boasts: "We should be 95% confident that the **sample average** falls in the provided interval". This is a valid statement. *True or false?*

Solution: FALSE

Free-response problems.

Problem 1.6. (15 points) *Source: "Probability" by Jim Pitman.*
A large elevator in a new hotel is deigned to carry up to about 30 people with a total weight of up to 5000 lbs. More than 5000 lbs overloads the elevator. The mean weight of the hotel guests is 150 lbs with a standard deviation of 55 lbs. Suppose exactly 30 of the hotel's guests enter the elevator. Assuming that the weights of individual guests are independent random variables, what is the approximate probability that the elevator gets overloaded?

Solution: Let $X_i, i = 1, \ldots, 30$ be random variables denoting the individual weights of the guests who boarded the elevator. Set $S_n = X_1 + \cdots + X_{30}$. Due to the central limit theorem, we have that approximately

$$S_n \sim Normal(mean = 30(150), sd = 55\sqrt{30})$$

So,

$$\mathbb{P}[S_n > 5000] = \mathbb{P}\left[\frac{S_n - 4500}{55\sqrt{30}} > \frac{5000 - 4500}{55\sqrt{30}}\right] \approx 1 - \Phi\left(\frac{500}{55\sqrt{30}}\right) = 1 - 0.9515 = 0.0484.$$

Problem 1.7. (10 points)
Source: "Probability and Statistics for Engineers and Scientists" by Walpole, Myers, Myers, and Ye.

A corrosion study was made in order to determine whether coating an aluminum metal with a corrosion retardation substance reduced the amount of corrosion. Also of interest is the influence of humidity on the amount of corrosion. Two levels of coating – no coating and chemical-corrosion coating – were used. In addition, there were two relative humidity levels at 20% relative humidity and at 80% relative humidity.

The coating is a protectant that is advertised to minimize fatigue damage in this type of material. A corrosion measurement can be expressed in thousands of cycles to failure.

There are eight aluminum specimens used.

(i) (5 points) What is the explanatory variable in the above experiment design? What are the possible values it can take? *Hint: Draw a table of possible treatment combinations!*

(ii) (2 points) What are the **experimental units/cases**?

(iii) (3 points) How would you assign the experimental units to the treatments to ensure that you are not introducing bias in your results?

Solution:

(i) The explanatory variable is the combination of coating or no coating, and 20% and 80% relative humidity. There are 4 possible treatment combinations.

(ii) The eight aluminum specimens.

(iii) Randomize the assignment of speciments to different treatment combinations.

Problem 1.8. (15 points) A fair tetrahedron (a four-sided symmetric Platonic solid) with sides dyed pink, purple, mauve, and fuchsia will be rolled 2000 times. You intend to record the color of the side the tetrahedron fell on after every roll. According to the normal approximation to the binomial, what is the approximate probability that the outcome is mauve at most 510 times?

Solution: The number of occurrences of mauve is

$$X \sim Binomial(2000, 1/4).$$

Since $2000(1/4) > 10$ and $2000(3/4) > 10$, we can use the normal approximation. We have

$$\mathbb{E}[X] = 2000(1/4) = 500 \quad \text{and} \quad SD[X] = \sqrt{2000(1/4)(3/4)} = 19.36492.$$

Hence,

$$\mathbb{P}[X \leq 510] = \mathbb{P}\left[\frac{X-500}{19.36492} \leq \frac{510.5-500}{19.36492}\right] \approx \Phi(0.5422176) = 0.7061657.$$

Problem 1.9. (10 points) Assume that the amount of liquid in a bottle of pediatric antibiotic is normally distributed with an unknown mean μ and with a known standard deviation of 4 milliliters. You perform 9 measurements of the contents of a bottle of this antibiotic. These are the values you obtain (in milliliters):

$$100, 102, 104, 98, 96, 101, 99, 95, 103.$$

What is the 98%-confidence interval you report?

Solution: The sample average is $\bar{x} = 99.77778$. The critical value of the standard normal distribution associated with the confidence level of 98% is

$$z^* = \Phi^{-1}(0.99) = 2.326348$$

So, the required confidence interval is

$$\mu = 99.77778 \pm 2.326348 \left(\frac{4}{\sqrt{9}} \right) = 99.77778 \pm 3.101797.$$

Multiple-choice problems.

Problem 1.10. Suppose a poll suggested the US President's approval rating is 45%. We would consider 45% to be ...

(a) the population proportion.
(b) the point estimate.
(c) the sample median.
(d) the sample standard deviation.
(e) the sample variance.

Solution: (b)

Problem 1.11. Let the population distribution be normal with mean μ and standard deviation σ. Let $\bar{X}$ denote the sample mean of a sample of size $n$from this population. Then, we know the following about the distribution of $\bar{X}$:

(a) $\bar{X} \sim Normal(mean = \mu, variance = \sigma^2)$
(b) $\bar{X} \sim Normal(mean = \mu, variance = \frac{\sigma^2}{n})$
(c) $\bar{X} \sim Normal(mean = \mu, variance = \frac{\sigma^2}{\sqrt{n}})$
(d) $\bar{X} \sim Normal(mean = \frac{\mu}{n}, variance = \frac{\sigma^2}{n})$
(e) None of the above are correct.

Solution: (b)
For the verification, see class notes.

Problem 1.12. (5 points) Consider a normal population distribution for a large population. You know that the standard deviation of the sampling distribution of the sample mean for samples of size 36 is 2. How large should a sample from the same population be so that the standard deviation of the sample mean becomes 1.2?

(a) 10
(b) 51
(c) 52
(d) 100
(e) None of the above.

Solution: (d)
Let σ denote the population standard deviation. Then, the standard deviation of the sample mean $\bar{X}_{36}$ of a sample of size 36 is

$$SD[\bar{X}_{36}] = \frac{\sigma}{\sqrt{36}} = \frac{\sigma}{6} = 2 \quad \Rightarrow \quad \sigma = 12.$$

Let n denote the unknown sample size for the sample whose standard deviation for the sample mean $\bar{X}_n$ needs to be 1.2. Then,

$$SD[\bar{X}_n] = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{n}} = 1.2 \quad \Rightarrow \quad \sqrt{n} = \frac{12}{1.2} = 10 \quad \Rightarrow \quad n = 100.$$