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

Practice for In-term exam II

Definitions.

Problem 1.1. (5 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 1.2. (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 1.3. (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 1.4. (2 points) Resident statistitian 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 1.5. (10 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 $X_i, i = 1, \dots, 30$ be random variables denoting the individual weights of the guests who boarded the elevator. Set $S_n = X_1 + \dots + X_{30}$. 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 1.6. (25 points) *Source: "Probability" by Pitman.*
A final exam consists of multiple choice problems – each problem with 5 offered answers only one of which is correct. Before the final exam the diligent student is given a practice set of multiple choice problems. Knowing that exactly 70% of the final exam will be out of the practice set, the student works out the entire practice set and gets the correct answer to each question.

When he takes the final exam, the student proceeds to answer the known questions correctly. However, for the remaining questions, he panics and chooses the answers completely at random.

(i) (5 points) What is the probability that the student answers a randomly chosen question correctly?

(ii) (5 points) **Given** that the student answered a particular question correctly, what is the probability that he was guessing at random when he was answering that question?

Let the total number of questions in the exam be 20. Let the random variable N represent the total number of questions the student answered correctly.

(iii) (5 points) What is the distribution of the random variable N?

(iv) (5 points) What is the expected value of N?

(v) (5 points) What is the standard deviation of N?

Solution:

(i)

$$0.7(1) + 0.3(0.2) = 0.76$$

(ii)

$$\frac{0.3(0.2)}{0.76} = 0.0789$$

(iii) The random variable N can be written as

$$N = 14 + X,$$

where $X \sim Binomial(n = 6, p = 1/5)$.

(iv)

$$\mathbb{E}[N] = 14 + 6 * 0.2 = 15.2$$

(v)

$$Var[N] = Var[X] = 6 * 0.2 * 0.8 = 0.96 \quad \Rightarrow \quad SD[N] = \sqrt{0.96} = 0.979796.$$

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.

Multiple-choice problems.

Problem 1.8. 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 1.9. 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 1.10. (5 points) Ahead of the school year, the KMS Functional-Fitness coach plans to track the progress of her students. She will record the number of jumping jacks her students can make before collapsing on a weekly basis. What kind of a procedure is this?

(a) A prospective observational study.

(b) A retrospective observational study.

(c) An experiment.

(d) A survey.

(e) None of the above.

Solution: (a)

Problem 1.11. (5 points) **Post-It Thievery!**

Chris P. Bacon, the office manager at a large temp agency wants to figure out what proportion of his workforce has been pilfering Post-Its. Realizing the issues with conducting a survey which outright asks: *"Have you ever committed unauthorized removal of Post-Its from the premises?"*, he decides to use the randomized-response method.

He prompts a computer to display the question

"Have you ever taken a Post-It home?"

with probability 0.75. The rest of the time, a virtual fair coin is flipped on the screen and the subject is asked

"Is the outcome heads?"

In both cases, the subject is prompted to click the button with **Yes** or **No**. The interviewer did not know the actual question asked, just the ultimate response. So, there was no real reason for the subject to lie, and we assume that the subjects responded truthfully.

It turned out that 80% of the subjects answered "yes". Give an estimate of the proportion of *Post-It thieves* in this population.

(a) 0.75

(b) 0.8

(c) 0.85

(d) 0.9

(e) None of the above.

Solution: (d)
Now, we are given that $\mathbb{P}[Yes] = 0.80$. Our goal is to figure out $p = \mathbb{P}[Yes \mid Q]$ with the conditioning event Q given by

$$Q = \{\text{the subject was asked the Post-It question}\}.$$

We are given that $\mathbb{P}[Q] = 0.75$.

By the *Law of Total Probability,*

$$\begin{aligned}\mathbb{P}[Yes] = \mathbb{P}[Yes \cap Q] + \mathbb{P}[Yes \cap Q^c] &= \mathbb{P}[Yes \mid Q]\mathbb{P}[Q] + \mathbb{P}[Yes \mid Q^c]\mathbb{P}[Q^c] \\ &= p(0.75) + 0.5(0.25) = 0.75p + 0.125.\end{aligned}$$

So,

$$0.75p = 0.8 - 0.125 = 0.675 \quad \Rightarrow \quad p = 0.9.$$