UNIVERSITY OF TEXAS AT AUSTIN, DEPARTMENT OF MATHEMATICS
M358K - Applied Statistics

THE MOCK IN-TERM TWO

Problem 2.1. (5 points) A political scientist is interested in the effect of government type on economic development. She wants to use a sample of 30 countries evenly represented among the Americas, Europe, Asia, and Africa to conduct her analysis. What type of study should she use to ensure that countries are selected from each region of the world?

(a) Observational - simple random sample
(b) Observational - cluster sampling
(c) Observational - stratified sampling
(d) Observational - multiphase sampling
(e) None of the above.

Solution: (c)

Problem 2.2. (5 points) **Toddler Bribery!**

Dr. P. Piagette, a developmental psychologist, is trying to figure out what percentage of parents of small children resort to offering candy to their offspring to quiet them down. Realizing that parents might not answer truthfully to an outright question about bribing their little ones, she decides to use the randomized-response method.

She sets up a computer to display the question

"Have you ever offered candy to appease your toddler?"

with probability 0.6. The rest of the time, a virtual spinner spins on the screen. Half of the time, the spinner lands on red, a third of the time, the spinner lands on blue, and one sixth of the time, the spinner lands on yellow. The parent is asked

"Did the spinner land on red?"

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 74% of the subjects answered "yes". Give an estimate of the proportion of *parents who bribe their children* 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 $\mathbb{P}[Yes] = 0.74$. Our goal is to figure out $p = \mathbb{P}[Yes \,|\, Q]$ with the conditioning event Q given by

$$Q = \{\text{the subject was asked the bribery question}\}.$$

We are given that $\mathbb{P}[Q] = 0.6$.

By the *Law of Total Probability*,

$$\begin{aligned}\mathbb{P}[Yes] = \mathbb{P}[Yes \cap Q] + \mathbb{P}[Yes \cap Q^c] &= \mathbb{P}[Yes \,|\, Q]\mathbb{P}[Q] + \mathbb{P}[Yes \,|\, Q^c]\mathbb{P}[Q^c] \\ &= p(0.6) + 0.5(0.4) = 0.6p + 0.2.\end{aligned}$$

So,

$$0.6p = 0.74 - 0.2 = 0.54 \quad \Rightarrow \quad p = 0.9.$$

Problem 2.3. (5 points) A medical researcher thinks that adding calcium to the diet will help reduce blood pressure. She believes that the effect is different for men and women. 20 men and 20 women are willing to participate in the study. The researcher chooses 10 of the men and 10 of the women at random. These chosen 20 men and women take a calcium pill every day. The other 20 men and women take a placebo. This is a ...

a.: stratified random sample design.
b.: simple random sample design.
c.: randomized block experimental design.
d.: completely randomized experimental design.
e.: None of the above is correct.

Solution: c.

Problem 2.4. (5 points)

To estimate a population mean, our resident statistician Martyn Rivera plans to pick two simple random samples, each of size 100, from the population. He also plans to calculate the confidence interval with level C for each sample. What is the probability that **exactly one** of his confidence intervals will cover the population mean?

a.: C^2
b.: $1 - C^2$
c.: $2C(1 - C)$
d.: $1 - (1 - C)^2$
e.: None of the above

Solution: c.

Problem 2.5. Let $Z \sim N(0,1)$. Given that Z is at most 0.3, what is the probability that Z is at most -2.5?

(a) 0.0062
(b) 0.0100
(c) 0.6117
(d) 0.6179
(e) None of the above.

Solution: (b)
From the definition of conditional probability, we are looking for

$$\mathbb{P}[Z \leq -2.5 \,|\, Z \leq 0.3] = \frac{\mathbb{P}[Z \leq 0.3, Z \leq -2.5]}{\mathbb{P}[Z \leq 0.3]} = \frac{\mathbb{P}[Z \leq -2.5]}{\mathbb{P}[Z \leq 0.3]}.$$

From the standard normal tables, we obtain

$$\mathbb{P}[Z \leq -2.5] = \Phi(-2.5) = 0.0062 \quad \text{and} \quad \mathbb{P}[Z \leq 0.3] = \Phi(0.3) = 0.6179.$$

So, our answer is

$$\mathbb{P}[Z \leq -2.5 \,|\, Z \leq 0.3] = \frac{\mathbb{P}[Z \leq -2.5]}{\mathbb{P}[Z \leq 0.3]} = \frac{0.0062}{0.6179} = 0.0100.$$

Problem 2.6. 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,

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

Problem 2.7. When a variable follows a normal distribution, what percent of observations are contained within 1.75 standard deviations of the mean?

Solution: Let $X \sim Normal(mean = \mu, sd = \sigma)$. Then,

$$\mathbb{P}[|X-\mu| \leq 1.75\sigma] = \mathbb{P}[-1.75\sigma < X-\mu \leq 1.75\sigma] = \mathbb{P}\left[-\frac{1.75\sigma}{\sigma} < \frac{X-\mu}{\sigma} \leq \frac{1.75\sigma}{\sigma}\right] = \mathbb{P}[-1.75 < Z < 1.75]$$

where $Z \sim N(0,1)$. Using the symmetry of the bell curve and the standard normal tables, we get

$$\mathbb{P}[-1.75 < Z < 1.75] = 2\mathbb{P}[Z < 1.75] - 1 = 2(0.9599) - 1 = 0.9198.$$

Problem 2.8. (5 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,

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