Problem Set 4

Francisco Brady

04 October, 2023

Public Policy 529

Fall 2023: Problem Set #4

Due Wednesday, October 4th

Note for grader: I made this using Rmarkdown, which can render LaTeX. It is useful including text, equations, and including code output/calculations all in one file. Let me know if you have any questions!

1. Suppose that, in the population, the mean amount of time spent watching television has a normal distribution. The mean is 20 hours per month and the standard deviation is 8 hours.

(a) What are the z scores associated with the following amounts of time spent watching television: 14, 24, and 33?

```
z1 <- (14 - 20)  / 8
z2 <- (24 - 20) / 8
z3 <- (33 - 20) / 8
```

The z-score for 14 is: -0.75
The z-score for 24 is: 0.5
The z-score for 33 is: 1.625

(b) What percentage of people watch more than 33 hours of television per month? What percentage of people watch less than 14 hours? What percentage of people watch less than 24 hours?

```
z33 <- (33 - 20) / 8
z14 <- (14 - 20) / 8
z24 <- (24 - 20) / 8
```

To find these percentages, calculate the z-score and look up the results in a z score table:

The z-score for 33 is 1.625. Looking up that value in the table, it is between 1.62 and 1.63, so the area given by the table is 0.0516 and 0.526, meaning that between 5.16% and 5.26% watch more than 33 hours.

The z-score for 14 is -0.75. Looking up that value (0.75) in the table, we get: 0.2266, meaning about 22.66% watch less than 14 hours. Note that we used the positive value instead of the negative value, but they are the same because of the symmetry of the distribution.

The z-score for 24 is 0.5. Looking up that value in the table, we get: 0.3085. Subtracting 1 from that value gives 0.6915, meaning around 69.15% watch less than 24 hours.

(c) What percentage of people watch between 14 and 33 hours?
First find the z-scores for 14 and 33 hours.

```
z14 <- (14 - 20) / 8
z33 <- (33 - 20) / 8
```

Look up the values in the z table to find the areas under the curve 0.2266 for 14, and 0.0526 for 33. Since the two areas represent the tails on each side, you can add them and subtract from 1 to get their range.

$$1 - (0.2266 + 0.0526) = 0.7208$$

So around 72.8% of people watch between 14 and 33 hours.

(d) A person's z score for TV watching is -0.25. How many hours of television is that?
To find a value from a z-score, we can rearrange the z-score formula:

$$Z = \frac{Y - \mu}{\sigma}$$
$$Y = \sigma Z - \mu$$

Using our values:

$$Y = 8 \cdot (-0.25) + 20 = 18$$

2. We learned the definition of percentiles in the Descriptive Statistics lecture. Find the z scores corresponding to the percentiles below. If necessary, provide a range rather than a specific z score.

(a) The 15th percentile of the normal distribution.
Z-score between 1.03 and 1.04

(b) The 70th percentile of the normal distribution.
Z-score between 0.52 and 0.53

(c) The 95th percentile of the normal distribution.
Z-score between 1.64 and 1.65

3. As the previous questions illustrate, a z score is a threshold above which, and below which, is a specific proportion of the area of a normal distribution. These proportions represent the probabilities with which the value of a variable falls above/below the value associated with that z score.

(a) Suppose that the parameter α (alpha) represents the area in the upper tail of the normal distribution above some threshold z. What is the z score associated with $\alpha = .01$?
To find the value above the given parameter, subtract 1 - $\alpha = 0.99$. In the z table, a z-score between 2.32 and 2.33 yields a value of 0.98983 and 0.99010.

(b) Now suppose α is the area in the lower tail below z. What is the z score associated with $\alpha = .05$?
In the z table, the z-score between -1.64 and -1.65 gives an area below z between 0.04847 and 0.0505.

(c) Now suppose α is the total area associated with two, equally-sized areas in the upper and lower tails of the normal distribution. In other words, each area is of size $\frac{\alpha}{2}$. If $\alpha = .05$, what are the z scores that define these areas?

$\frac{0.05}{2} = 0.025$

In the z table, the z-score of -1.96 gives an area of 0.02500 for the lower tail, and 1.96 gives an area of 0.97500.

4. The distribution of household income in the United States has a strong positive skew. Suppose the mean (μ) is \$97,941, and the standard deviation (σ) is \$76,021. You collect a random sample of size n = 400 and measure household income.

(a) As we have learned, the mean of your sample comes from a probability distribution over all possible sample means (i.e. a sampling distribution). According to the Central Limit Theorem, what value is at the middle of this distribution? Provide a specific number.

The value in the middle of the sampling distribution should approach \$97,941.

(b) Would this sampling distribution have a positive skew? Why or why not?

No, the central limit theorem states that the sampling distribution will be an approximately normal distribution, even if the underlying distribution is skewed, because the sample size is larger than 30.

(c) According to the Central Limit Theorem, what is the size of random sampling error for the sample mean? Provide an interpretation of this number.

The size of the random sampling error for the sample mean is the typical deviation from the population parameter, called the standard error.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

For this sample, it would be: $\frac{76{,}021}{\sqrt{400}} = 3801.05$

```
76021/ sqrt(400)
```

```
## [1] 3801.05
```

(d) Using what you found above, what range includes 95 percent of all sample means?

95% of all observations would be between \$75,648.50 and \$76,393.50

```
lower <- 1.96 * (3801.05/ sqrt(400))
upper <- -1.96 * (3801.05/ sqrt(400))
76021 - lower
```

```
## [1] 75648.5
```

```
76021 - upper
```

```
## [1] 76393.5
```

(e) What is the probability that your sample mean would be greater than \$101,742?

Using for formula for the sample z-score:

$$Z = \frac{x - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{101{,}742 - 97{,}941}{\frac{76{,}021}{\sqrt{400}}}$$

```r
(101742 - 97941) / (76021 / sqrt(400))
```

```
## [1] 0.9999868
```

Looking up 0.99 in the z table, it returns 0.1611, meaning there is around a 16.11% chance that you could find a sample mean greater than $101,742

5. Suppose that, with probability .32, a randomly selected person will agree that astrology has scientific truth. Also, suppose that there is a probability of .89 that a person will agree that science brings opportunities for the next generation. Finally, suppose that there is a probability of .29 that a person will agree with both.

Using your choice of formulas, a table, a probability tree, or a Venn diagram, find the following.

A: Astrology has scientific truth
B: Science brings opportunities
P(A) = .32
P(B) = .89
P(A and B) = .29

	A	~A	Total
B	.29	.60	.89
~B	.03	.08	.11
Total	.32	.68	1.0

(a) What is the probability that a person agrees with neither statement?

P(~A and ~B) = 0.08

(b) What is the probability that person will agree with the statement that science brings opportunities for the next generation given that they do not agree that astrology has scientific truth?

$P(B| \sim A) = \frac{P(B \text{ and } \sim A)}{P(\sim A)}$
$P(\sim A) = 1 - P(A) = .68$ $P(B \text{ and } \sim A) = .60$

$$P(B| \sim A) = \frac{.60}{.68} = 0.882$$