Math 58 / 58B - Introduction to (Bio)Statistics

solutions: Normal Probabilities and Z scores Worksheet

in class Jan 30, 2020

In class today you have a worksheet[1] that gives you practice calculating Z scores and finding percentages / probabilities associated with the normal distribution.

1. My dogs, Sequoia and Wrigley, are trying to get into college. As part of the entrance requirements, they both took a standardized test. Sequoia took the SAT and scored a 1250. Wrigley took the ACT and scored a 29.

- Explain why it doesn't make any sense to compare the raw score of 1250 to the raw score of 29.

Solution

No, the raw scores are on totally different scales. We have no idea whether one of them is a "good" score and the other isn't or vice versa.

With some additional information we find out that SAT scores have a mound-shaped (approximately normal) distribution with a mean of about 1050 and a standard deviation of about 200. ACT scores have a mound shaped (approximately normal) distribution with a mean of 21 and a standard deviation of about 5.5.

- Now can you tell who scored higher?
  - Find the Z score for each of Wrigley and Sequoia.
  - Using `xpnorm` (in the mosaic package: `library(mosaic)`), calculate the percent of test takes who are higher than each of Wrigley and Sequoia. Draw a picture. [If you don't have your computer / R, feel free to use the table in the book.]
  - Explain in words who has the higher test score (in terms of getting into college).

Solution

Sequoia is only one standard deviation above the mean, whereas Wrigley is almost a standard deviation and a half above the mean.

```
Z_w = (29-21) / 5.5
Z_w
```

```
## [1] 1.454545
```

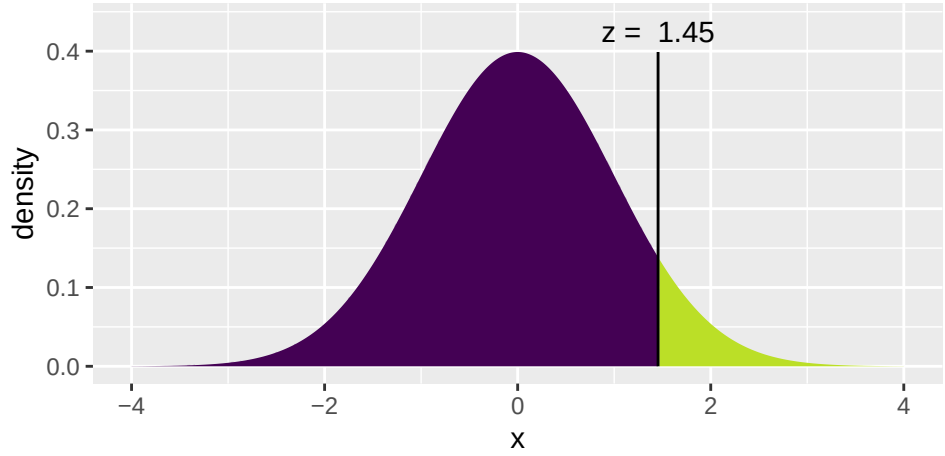
```
Z_s = (1250 - 1050)/200
Z_s
```

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

7.3% of the ACT test takers scored higher than Wrigley. 15.9% of the SAT test takers scored higher than Sequoia

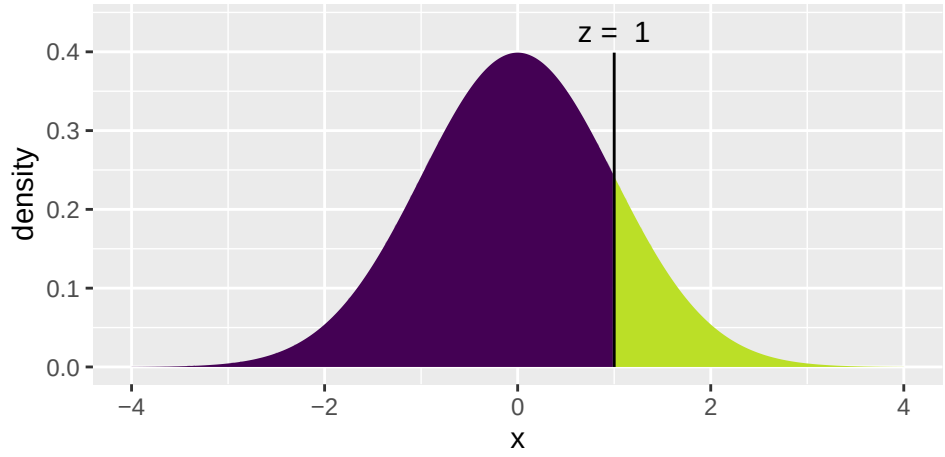
```
xpnorm(1.4545, 0, 1, plot=TRUE)
```

[1] questions taken from https://askgoodquestions.blog/2019/08/26/8-end-of-the-alphabet/ by Allan Rossman




```
## [1] 0.9270961
```

```
xpnorm(1, 0, 1, plot=TRUE)
```




```
## [1] 0.8413447
```

Given the Z scores and the corresponding percentages of test takers above each of their scores, it seems as though Wrigley's score is relatively higher than Sequoia's score.

2. Some of you may be familiar with Hans Rosling who founded the website https://www.gapminder.org/ and dedicated his life to promoting awareness of global health issues, see his Ted talks here: https://www.ted.com/playlists/474/the_best_hans_rosling_talks_yo. One question he liked to ask is:

   Has the percentage of the world's population who live in extreme poverty doubled, halved, or remained about the same over the past twenty years?

- Before you go on, answer the question. Has the extreme poverty doubled, halved, or remained about the same? What do you think?

  The correct answer is that this percentage has halved, but only 5% of a sample of 1005 U.S. adults in 2017 got this right. Rosling liked to say that chimpanzees would do better than people: With only three options, we would expect 33.33% of chimpanzees to answer correctly.

- If in fact the students are randomly guessing, how many standard deviations away from the "random guess" value is 0.05? [Hint: use proportions and not percentages in your calculations.]

note: we covered this in class on Tuesday, so it's in the notes, but the formula doesn't show up in your text until the box on page 124 in section 3.1.1.

Do not use the computer here (except as a calculator, and feel free to use a calculator or use the computer / R as a calculator). Note: you need to know how many people were asked, look above.

Solution

$$SD(\hat{p}) = \sqrt{p(1-p)/n} = \sqrt{(1/3)(2/3)/1005} = 0.0149$$

```
sqrt((1/3)*(2/3)/1005)
```

```
## [1] 0.01486999
```

How far is 0.05 from (1/3) in units of standard deviation? That's just a Z score! Yikes, the 5% value is MORE THAN 19 STANDARD DEVIATIONS BELOW RANDOM GUESSING!!!

```
Z_p = (0.05 - (1/3)) / sqrt((1/3)*(2/3)/1005)
Z_p
```

```
## [1] -19.05404
```

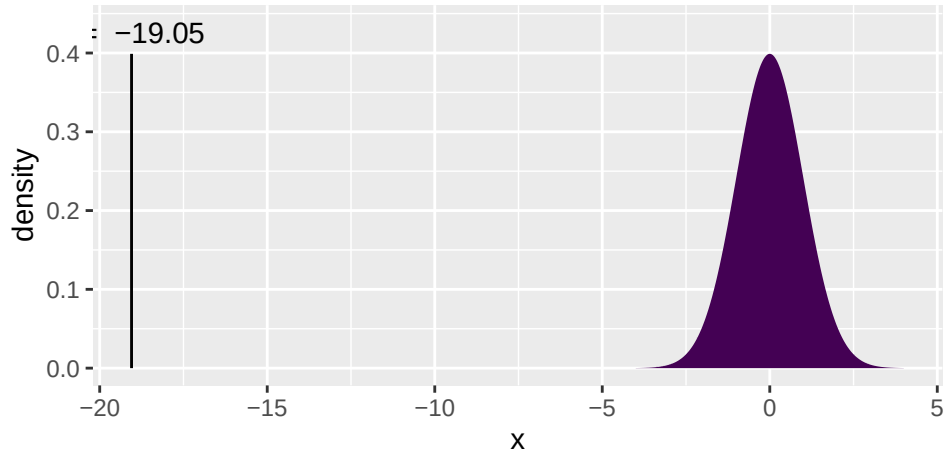
- What does this say about humans doing so much *worse* than random guessing when answering the question about poverty? (No hypothesis test here, just a reflection on the distance between the observed data and the random guess answer.)

Solution

Not only are humans *wrong*, but they are *wrong* at an extremely high rate. That is, they are wrong in such a way that they can't possibly be guessing. There must be something about the question that makes so many people get it wrong (maybe that they are all seeing the same media narrative which describes continued problems with extreme poverty?)

We could find the percent of samples that would have produced such a small $\hat{p}$ if people were indeed random guessing. Unsurprisingly, the proportion of such samples is exceedingly small:

```
xpnorm(-19.05, 0, 1, plot=TRUE)
```




```
## [1] 3.28511e-81
```

3. Suppose that my turtles, Mir & Elmer, take an exam for which the mean score is 70 and standard deviation of scores is 8. Elmer's score on the exam is 75, and Mir's score is 1.5 standard deviations above Elmer's score. What is Mir's score on the exam?

Solution

1.5 standard deviations is 12 (one and a half times of 8). 12 points higher than 75 is 87. Mir's score is 87.