PSYCHOLOGY 300A (A01)

Assignment 3
September 24, 2018

Formulas: \[ M = \frac{\sum X}{N} \quad s^2 = \frac{\sum (X-M)^2}{N-1} \quad \sigma^2 = \frac{\sum (X-\mu)^2}{N} \quad z = \frac{X-M}{s} \quad X = M + z(s) \]

1. Based on the data shown in the frequency polygon to the right, construct a frequency table (i.e., list each score value and its frequency). What percentage of the scores in this distribution have a value of 9 or higher? What is the mean of the full set of scores? Which score in the distribution has the largest squared deviation from the mean of the scores?

2. A normal distribution of scores has a certain \( M \) and \( s \). Suppose two scores are added to this distribution, one equal to the lowest score in the original distribution and one equal to the highest score in the original distribution. What will happen to \( M \) and what will happen to \( s \) when these two scores are included (stay the same, increase, decrease)?

3. Create two sets of scores (each set with 4 scores) that have equal ranges but different variances.

4. In a final exam in biology a student obtains a score of 82%, which corresponds to an A– by the percent-to-grade translation set by the course instructor. The biology class as a whole obtained scores with \( M = 78.6\% \) and \( s = 12.9\% \). In the final exam in psychology, this same student obtained a score of 78%, which corresponds to a B+ by the percent-to-grade translation set by the course instructor. The psychology class as a whole obtained scores with \( M = 70.2\% \) and \( s = 6.1\% \).
   (a) What is the \( z \) score that corresponds to the student’s biology test score?
   (b) What is the \( z \) score that corresponds to the student’s psychology test score?
   (c) Use one sentence to explain the sense in which this student did better in psychology than in biology.
   (d) On the biology test, what score (rounded to the nearest whole %) would correspond to a \( z \) score of –0.5?
   (e) On the psychology test, what score (rounded to the nearest whole %) would correspond to a \( z \) score of 2.5?

5. Given the following information based on just two members of a set of scores, determine the mean and standard deviation of the whole set of scores. One member of the set is a score of 50, which corresponds to a \( z \) score of 1.00. Another member of the set is a score of
56, which corresponds to a $z$ score of 2.00. (Hint: Draw a diagram of the distribution of scores—the shape of the distribution does not matter—and mark the location of these two scores and the mean.)

6. A set of scores forms a bimodal distribution. These raw scores are then converted to $z$ scores. What will be the shape of the distribution of these $z$ scores?

7. In a negatively skewed distribution, it is most likely that how many $z$ scores lie below zero?
   (a) half
   (b) fewer than half
   (c) more than half
   (d) none
   (e) all

8. Consider a normal distribution of scores, with $M = 64$ and $s = 14$.
   (a) What proportion of scores are greater than or equal to 48 and less than or equal to 75?
   (b) What proportion of scores are between 40 and 60, inclusive?
   (c) Suppose there are 1,432 individuals in this distribution. How many of them (to the nearest whole number) would have scores equal to or greater than 70?
   (d) What score (to the nearest whole number) corresponds to the 45th percentile?

9. A flight training school run by the armed forces to select new recruits uses a flight simulator to evaluate the potential of cadets. A large group of cadets who take this test obtain a mean of 31.6, with a standard deviation of 8.5. High scores indicate better performance. The distribution of scores on this test is normal. Only cadets who obtain a score in the upper 70% of the distribution will be selected for service in the air force. Also, cadets who score in the upper 4% of the distribution will be offered the chance to train as test pilots for newly developed aircraft. Find the raw score cutoffs (to the nearest whole number) for each of these two categories (upper 70%; upper 4%).

10. A researcher has a very large sample of normally distributed intelligence test scores with $M = 100$ and $s = 15$. The researcher wants to identify the 5% most extreme scores in the distribution, equally split between the upper and lower end of the distribution. What ranges of raw scores constitute the 5% most extreme scores?

11. Suppose that the measurement error that contributes to observed scores on some test is normally distributed with a mean of 0 and a standard deviation of 2. According to classic measurement theory, each time a person is measured on the test, the resulting observed score is an additive combination of the person’s true score and an amount of measurement error randomly drawn from the normal distribution of measurement error. What is the probability that measurement error will lead to an overestimate of a person’s true score by at least 3 points? Suppose that someone’s observed score actually overestimates their true score by 2 points. What is the probability that if that person is measured again, he or she will obtain a lower observed score? (Assume that each time a
person is measured a new measurement error value is randomly drawn from the
distribution of measurement errors.)

12. Use R to randomly sample 100 scores from a normally distributed population with \( \mu = 20 \) and \( \sigma = 4 \), with scores rounded to the nearest whole number. Recall from previous
assignments that this can be done using the `round` and `rnorm` functions. Plot a
histogram for these scores specifying 50 bars in the histogram. How would you
characterize the general shape of this distribution?

Now transform all 100 of these scores into \( z \) scores. This can be done quite simply, by
computing the mean and standard deviation for the sample using the `mean` and `sd`
functions, as in Assignment 2. Just assign the result of these functions to new data
variables, like this (assuming you put the random sample of 100 scores into a data
variable called \( X \)):

\[
\begin{align*}
> \text{meanX} &= \text{mean}(X) \\
> \text{sdX} &= \text{sd}(X)
\end{align*}
\]

Then you can transform the raw scores into \( z \) scores, putting the \( z \) scores into a new data
variable, with just one command (notice how this command captures the formula for \( z \)
scores):

\[
> zX = (X - \text{meanX})/\text{sdX}
\]

Plot a histogram of these \( z \) scores, again using 50 bars. How would you characterize the
general shape of this distribution? Compare the shape of this distribution to the shape of
the distribution of the corresponding raw scores – are these two distributions similar or
very different from one another?

Repeat the above exercise, but this time draw the random sample from a uniform
distribution with a minimum score of 8 and a maximum score of 32. The command for
drawing samples from a uniform distribution is `runif`, so you can use this command to get
your sample of 100 scores:

\[
X = \text{round}(\text{runif}(100, \text{min} = 8, \text{max} = 32))
\]

How would you characterize the shape of the distribution of raw scores and \( z \) scores based
on this new sample? Are the two distributions similar or very different from one another?
Did the \( z \) score transformation result in an approximately normal distribution of scores?

13. Use R to simulate the drawing of 100,000 random samples of size 20 from a normally
distributed population with \( \mu = 50 \) and \( \sigma = 10 \). For each sample, compute the sample
mean, and put the results into a variable so you can do two additional things. First, what
is the mean of those sample means? Second, construct a histogram of the sample means
using 100 intervals and describe its shape. The \( R \) functions needed to draw the samples
and to compute and store the sample means were presented in Assignment 2. This time,
however, instead of computing the variance of each sample using

\[
> \text{vars}=\text{replicate}(100000,\text{c}(\text{vars,}\text{var}(\text{rnorm}(20,50,10))))
\]
you will compute the mean of each sample with the following command.  
> means = replicate(100000, c(means, mean(rnorm(20, 50, 10))))