

MA117 - WORKSHEET 4  
DISTRIBUTIONS  
Week 1, Friday

**Problem 1.** In triathlons, racers are often placed into age and gender groups. Leo and Mary just completed the Hermosa Beach Triathlon, where Leo competed in the “Men, Ages 30–34” group while Mary competed in the “Women, Ages 25–29” group. Leo completed the race in 1:22:28 (4948 seconds), while Mary completed the race in 1:31:53 (5513 seconds). Obviously, Leo finished faster, but they are curious about how they did within their respective groups. Can you help them? Here is some information on the performance of their groups:

- The finishing times of the “Men, Ages 30–34” group are approximately normal with mean of 4313 seconds and standard deviation 583 seconds.
  - The finishing times of the “Women, Ages 25–29” group are approximately normal with mean of 5261 seconds and standard deviation 807 seconds.
- (a) What are the  $z$ -scores for Leo’s and Mary’s finishing times? What do these  $z$ -scores tell you? Did Leo or Mary rank better in their respective groups? Explain your reasoning.
- (b) What percent of the triathletes did Leo finish faster than in his group? What percent of the triathletes did Mary finish faster than in her group?
- (c) What is the cutoff time for the fastest 5% of athletes in the men’s group?
- (d) What is the the cutoff time for the slowest 10% of athletes in the women’s group?
- (e) If the distributions of finishing times are not approximately normal, would your answers above change? Explain your reasoning.

**Problem 2.** About 13% of Americans speak Spanish at home. Suppose you take a simple random sample of 100 Americans. What is the probability that at most 10 people in your sample speak Spanish at home?

Do this problem twice: once exactly using a binominal distribution, and once approximately using a normal distribution. Make sure to verify the necessary hypothesis for using a normal approximation!

**Problem 3.** For each of the following, (i) draw a picture of a bell curve and shade in a relevant region under the curve, and then (ii) write down the R command you would use to calculate the stated quantity, using the functions `pnorm` (which takes  $z$ -scores to percentiles) and/or `qnorm` (which takes percentiles to  $z$ -scores).

- (a) The percentage of observations whose  $z$ -scores are less than 1.5.
- (b) The percentage of observations whose  $z$ -scores are greater than 1.5.
- (c) The percentage of observations whose  $z$ -scores are between  $-0.5$  and  $2$ .
- (d) The percentage of observations whose  $z$ -scores have absolute value greater than  $1.25$ .
- (e) The  $z$ -score that is larger than exactly 90% of observations.
- (f) The  $z$ -score that is less than exactly 75% of observations.
- (g) The number  $z^*$  such that 80% of the observations are contained within  $z^*$  standard deviations of the mean.

- (h) The number  $z^*$  such that 90% of the observations are contained within  $z^*$  standard deviations of the mean.

*Note.* Make sure you know how to do this problem. I promise you that something very similar will appear on quiz 2! ☺