Warm-Up for Day 21

Prof. Jordan C. Hanson

August 2, 2022

1 Formula Area

1. Normal distribution PDF: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}(x-\mu)^2/\sigma^2\right)$

2. The *z-score* of a particular event x drawn from a normal distribution is $z = (x - \mu)/\sigma$.

3. Probabilities for a normal distribution: $p(-1\sigma < x < 1\sigma) = 0.68$, $p(-2\sigma < x < 2\sigma) = 0.95$, and $p(-3\sigma < x < 3\sigma) = 0.997$.

4. Notation for normal distribution: $X = N(\mu_x, \sigma_x)$, where μ_x and σ_x are the population mean and standard deviation, respectively.

5. **Central Limit Theorem:** Suppose you have a continous random variable X, with mean μ_x and standard deviation σ_x. Suppose you draw n values from its distribution (whatever that distribution is), and calculate the average, $\bar{x}$. If you repeat this process, the *distribution of the averages* $\bar{x}$ will be $N(\mu_x, \sigma_x/\sqrt{n})$. That is, the mean of the distribution of averages will be the mean of X, and the standard deviation of the distribution of averages will be the standard deviation of X divided by $\sqrt{n}$.

2 Normal Distribution, Central Limit Theorem

1. According to the Internal Revenue Service, the average length of time for an individual to complete (keep records for, learn, prepare, copy, assemble, and send) IRS Form 1040 is 10.5 hours (without any attached schedules), and the standard deviation is 1.5 hours. The distribution is unknown. Suppose we sample 64 taxpayers, and compute the mean time-to-complete the 1040, $\bar{t}$ and the standard deviation in the mean, σ_t. What are t and σ_t?

2. Suppose the world-average marathon time is 145 minutes for professional athletes, and the standard deviation is 15 minutes. We do not know the underlying distribution of times. If we collected times of 100 runners' races, (a) What mean should we expect to find across these 100 times? (b) What would the standard deviation of the mean be? (c) If a runner in our set of 100 measusrements ran a marathon in 130 minutes, how many standard deviations is this time below the *population* mean? (d) How far below the *sample* mean is this time, given the error in the mean?

3. The cost of unleaded gasoline in the Bay Area once followed an unknown distribution with a mean of \$4.50 per gallon and a standard deviation of \$0.16 per gallon. (a) How many standard deviations below the mean is the price \$3.94? (b) Sixteen gas stations from the Bay Area are randomly chosen. We are interested in the average cost of gasoline for the 16 gas stations. What is the average of the 16 prices, and what is the standard deviation of the prices? (c) Graph the distribution of all gasoline prices in the Bay Area *assuming* it follows a normal distribution.