Homework 2 Austin Frownfelter

# Problem 1

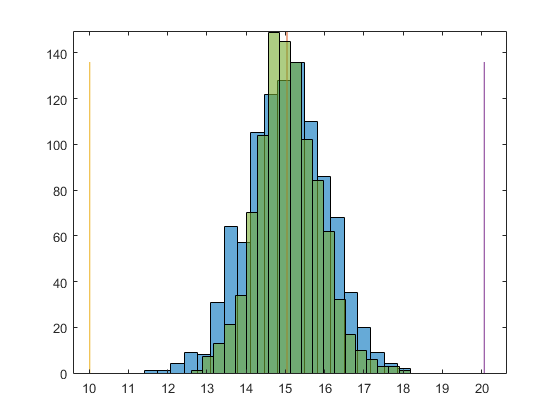
## (Part 1)

The mean and standard deviation of the mean-study-data is 15.0415 and 5.0279, respectively

## (Part 4)

The mean of subsample means of size 25 resulted in 15.0147, with a standard deviation of 0.9747. The vertical lines represent the mean and a single standard deviation of the original data. The subsample means follow a bell-shaped distribution, with the mean very close to the original data mean. With 25 data points in each subsample, it is very unlikely to get a mean out in the tail of the original dataset, which is why the standard deviation is a fraction of the data standard deviation.

## (Part 5)

The mean of subsample means of size 40 resulted in 15.0796, with a standard deviation of 0.7969. The sample-40 histogram (green) is transposed over the sample-25 to show the difference. The sample-40 is tighter than the sample-25 (with an even smaller standard deviation) because there are even fewer possibilities of getting a mean out in the tail with that many random sample points. The sample-40 is also taller, showing there is a higher probability of getting a subsample mean close to the actual mean.

# Problem 2

## (Part 2)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | 14.8675 | 14.3936 | 14.0634 | 15.0472 | 15.8946 | 15.1703 | 15.9065 | 15.6044 | 14.8419 | 14.6258 |
| Std Dev | 4.8722 | 4.8398 | 4.7843 | 5.0275 | 5.4262 | 5.0819 | 5.3552 | 4.5855 | 4.6487 | 5.4515 |

Since the dataset is evenly divisible by 10, each fold contained 100 data samples. In case the folds cannot be split evenly, the first n folds (where n is the remainder of data size divided by k folds) will be size 1 larger than the remaining folds. The mean of means comes out to exactly the mean of the data (to be expected). The standard deviation comes out to around 0.4% off of the actual standard deviation, as opposed to a maximum of 8.8% error if only a single round of training is done with one of those folds.

# Problem 3

## (A)

## (B)

## (C)

## (D)

## (E)

## (F)

## (G)

## (H)

## (J)

# Problem 4

## (A)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| P(x) | 1 | 2 | 3 | 4 | 5 | 6 | 5 | 4 | 3 | 2 | 1 |

## (B)

The expected value is 7.0

## (C)

P(no 4 in 5 games) = 0.6472

P(even in all 5 games) = 0.03125

## (D)

The mean will “shift” the distribution to the left or right (the mean is where the peak of the distribution exists.) The standard deviation affects the “height” of the distribution, where higher standard deviations produce a shorter, “fatter” distribution.