**Activity 7.4.1A – Exploring Distributions of Sample Means**

*Statistical inference* is the process of using sample statistics to draw conclusions about population parameters. *Sample statistics* are numerical descriptions of sample characteristics (e.g. sample mean, sample proportion). *Population parameters* are numerical descriptions of population characteristics (e.g. population mean, population proportion). When population parameters are unknown we use sample statistics to make inferences about population parameters.

To understand the process of statistical inference, we must first explore the behavior of sample statistics from random samples. This activity will focus on sample means, how they vary, and how they relate to the underlying population mean.

**High School Seniors’ Critical Reading SAT Scores**

The table on the following page displays the SAT critical reading scores for a hypothetical population of *N* = 100 high school students. The scores in this population have a bell-shaped distribution with a mean of

1. Suppose you randomly select 5 students from this population of students and calculate the sample mean. How far do you expect the sample mean will be from the population mean? Why?
2. Select a random sample of *n* = 5 students from the population. Write down your sample and calculate the sample mean. Use the symbol , pronounced *x*-bar, to denote the sample mean. Round your answer to the nearest whole number.

Sample mean

1. Find a second random sample of *n* = 5 students from the population. Calculate the sample mean. Round your answer to the nearest whole number.

**Table:** Population of 100 SAT Critical Reading Scores

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Student | Score |  | Student | Score |  | Student | Score |
| 1 | 404 |  | 35 | 547 |  | 69 | 396 |
| 2 | 501 |  | 36 | 368 |  | 70 | 486 |
| 3 | 467 |  | 37 | 478 |  | 71 | 585 |
| 4 | 491 |  | 38 | 445 |  | 72 | 544 |
| 5 | 520 |  | 39 | 704 |  | 73 | 415 |
| 6 | 486 |  | 40 | 437 |  | 74 | 354 |
| 7 | 508 |  | 41 | 550 |  | 75 | 390 |
| 8 | 432 |  | 42 | 502 |  | 76 | 519 |
| 9 | 571 |  | 43 | 650 |  | 77 | 469 |
| 10 | 547 |  | 44 | 525 |  | 78 | 423 |
| 11 | 510 |  | 45 | 424 |  | 79 | 448 |
| 12 | 335 |  | 46 | 459 |  | 80 | 634 |
| 13 | 493 |  | 47 | 433 |  | 81 | 380 |
| 14 | 641 |  | 48 | 439 |  | 82 | 454 |
| 15 | 501 |  | 49 | 529 |  | 83 | 324 |
| 16 | 492 |  | 50 | 464 |  | 84 | 403 |
| 17 | 535 |  | 51 | 396 |  | 85 | 418 |
| 18 | 528 |  | 52 | 501 |  | 86 | 506 |
| 19 | 520 |  | 53 | 401 |  | 87 | 355 |
| 20 | 483 |  | 54 | 300 |  | 88 | 540 |
| 21 | 425 |  | 55 | 460 |  | 89 | 447 |
| 22 | 604 |  | 56 | 593 |  | 90 | 428 |
| 23 | 478 |  | 57 | 542 |  | 91 | 538 |
| 24 | 386 |  | 58 | 420 |  | 92 | 569 |
| 25 | 532 |  | 59 | 659 |  | 93 | 539 |
| 26 | 494 |  | 60 | 563 |  | 94 | 451 |
| 27 | 633 |  | 61 | 432 |  | 95 | 414 |
| 28 | 635 |  | 62 | 369 |  | 96 | 639 |
| 29 | 374 |  | 63 | 474 |  | 97 | 383 |
| 30 | 293 |  | 64 | 497 |  | 98 | 540 |
| 31 | 559 |  | 65 | 442 |  | 99 | 535 |
| 32 | 385 |  | 66 | 529 |  | 100 | 427 |
| 33 | 452 |  | 67 | 536 |  |  |  |
| 34 | 619 |  | 68 | 373 |  |  |  |

*Note*: This is a hypothetical population. This population data is based on SAT critical reading scores from high school seniors in Connecticut.

1. You and your classmates have generated a distribution of sample means. Plot your sample means and your classmates’ sample means on the dot plot below.

**Distribution of Sample Means, *n* = 5**



The dot plot above is an example of an *empirical sampling distribution* – a distribution of sample statistics obtained by simulating random sampling from a population. Similar to other distributions, an empirical sampling distribution has a center, shape and spread that summarize values in its distribution.

1. Estimate or calculate the mean and standard deviation of the sample means in the empirical sampling distribution.
2. What is the shape of the empirical sampling distribution?
3. Which sample means appear to be unusual? Explain.
4. How does the mean of the empirical sampling distribution compare to the population mean
5. *Sampling variability* refers to the fact that the distribution of sample means from random samples of the same size varies in a predictable way. How has this been exemplified in this distribution?

**The Role of Sample Size**

What happens if we increase the size of the random samples? For example, what would happen to the distribution of sample means if the sample size increased from *n* = 5 to *n* = 10?

1. Make a conjecture to answer the previous question.
2. Select two random samples of *n* = 10 from the population on page 2. For each sample, calculate the sample mean. Use the symbol , pronounced *x*-bar, to denote the sample means. Round your answers to the nearest whole number.

Sample mean 1:

Sample mean 2:

1. You and your classmates have generated a distribution of sample means. Plot your sample means and your classmates sample means on the dot plot below.

**Distribution of Sample Means, *n* = 10**



1. Estimate or calculate the mean and standard deviation of the sample means in the empirical sampling distribution.
2. What is the shape of the empirical sampling distribution?
3. Which sample means appear to be unusual? Explain.
4. How did the center, shape, and variability of the distribution of sample means change when the sample size increased from *n* = 5 to *n* = 10?

The summary below highlights the key concepts that we explored in this activity.

**Key Understanding:** Sampling Distributions of Sample Means

A *sampling distribution of sample means* is the distribution of *all* possible sample means from random samples of the same size.

* The mean of a sampling distribution of sample means is approximately equal to the population mean.
* The standard deviation of a sampling distribution of sample means is called the *standard error.*
* For *any* population, as the sample size increases, the sample means become less variable and the sampling distribution of sample means approaches a bell-shaped distribution.

1. Sample means are *estimates* of the population mean. If we need to estimate a population mean with a sample mean, why would it be beneficial to select a large-sized random sample?
2. If the population mean SAT score is 484, would it be unlikely to find a random sample of

*n* = 5 scores in which the sample mean is 540 or higher? Explain.

1. If the population mean SAT score is 484, would it be unlikely to find a random sample of

*n* = 10 scores in which the sample mean is 540 or higher? Explain.