**Sampling Distribution and Interval Estimation**

**Standard Error of the Means**

**Intro**

Standard error of the mean is a measure of dispersion of the distribution of sample means and is similar in concept to the standard deviation in a frequency distribution and it measures likely deviation of a sample mean from the grand mean of the sampling distribution.

Recollect the five children data. Taking all the sample means as calculated before we can calculate standard deviation of the sampling distribution thus;

|  |  |  |
| --- | --- | --- |
| **X̅** | **μ** | **(X̅ – μ)^2** |
| 3 | 6 | 9 |
| 4 | 6 | 4 |
| 5 | 6 | 1 |
| 6 | 6 | 0 |
| 7 | 6 | 1 |
| 8 | 6 | 4 |
| 9 | 6 | 9 |

**Standard error = 2.**

Standard deviation of the population was 2.83. Standard error will always be less than the standard deviation for the simple reason that the values of sample means, are much closer to each other than the values of variable (X) in the population. Hence the deviations of x bar from x double bar would be much smaller leading to a smaller value of standard error.

The formal relationship between standard deviation of the population and the standard error is given below

Standard error = Standard deviation/SQRT(sample size)

**Example**

The IQ scores of college students are normally distributed with a mean mu of 120 and standard deviation sigma of 10.

1. What is the probability that the IQ score of any one student chosen at random is between 120 and 125.
2. If a random sample of 25 students is taken, what is the probability that the mean of this sample will be between 120 and 125.

Use the standardized normal distribution formula we get

Z = (x – μ)/sigma

= 125-120/10 = 5/10 =0.5

Area is 19.5%.

This means that there is a 19.15% chance that a student picked up at random will have an IQ score between 120 and 125.

b.

With the sample of 25 students, it is expected that the sample mean will be much closer to the population mean, and hence it would be highly likely that the sample mean would be between 120 and 125.

The formula to be used in the case of standardised normal distribution for the sampling distribution

**Example 2**

The average account receivable in a ledger is $125 with a standard deviation of $20. A random sample of 25 receivable accounts is selected from this ledger.

What is the probability that the average of this sample will be less than $115.

Solution:

If all possible samples of size 25 are selected from this ledger, then the sampling distribution of the mean of these samples would be normally distributed with a mean (X-double bar) OF 125 and a standard error of the mean = s.d/root(n).

Z = (x̅– μ)/S.E.M where S.E.M = S.D/root(n) = 20/root(25) = 4

Z = (115-125)/4 = -2.5

Area to the left of point 115: 0.5000 – 0.4938 = 0.0062

P(x̅ <= 115) = 0.0062

**Example 3**

The average IQ score of students in a school of gifted children is 165 with a standard deviation of 27. A random sample of 36 students is taken. What is the probability that the sample mean will be either less than 170 or more than 175

Solution

Area to the left of 170 = 0.8665

Are to the right of 175 = 0.0132

Total required area = 0.8665 + 0.0132 = 0.8797

**Confidence Interval**

**Interval Estimate of the Population mean (page 193)**

Our real interest is to draw conclusions about the population. We take a random sample of a given size from a population and compute a *statistic* which is a characteristic of a sample. And this becomes an estimate of the similar characteristic of the population. There are two major types of estimates namely point estimate and interval estimates.

A point estimate uses a single sample value to estimate the desired population parameter. For example, a sample mean x-bar is considered a point estimate of the population mean mu. Since a population parameter is always inferred from sample statistic, it is necessary and important that such sample statistic should be as highly reliable as an estimator for population parameter as possible.

Point estimator though simplistic in nature has some drawbacks. It may not exactly locate the population parameter resulting in some margin of uncertainty. For instance the average may or may not be equal or close to the average of the population. To overcome this difficulty, statisticians use another type of estimation known as interval estimation. In this method we first find a point estimate. Then we use the estimate to construct an interval on both sides of the point estimate, which we can be reasonably confident that the true parameter will lie.

Since the sample means are normally distributed with a mean of mu and standard deviation of sem, it follows that sample means follow normal distribution characteristic.

Then transforming the sampling distribution of sample means into the standard normal distribution we get

Z = x̅ – μ/s.e.m

x̅ – μ = Z x s.e.m

μ = x̅ ± Z x s.e.m

*Diagram here*

**Interpretation**

1. If all possible samples of size n were taken, then on the average 95% of these samples would include the population mean within the interval around that sample means bounded by X1 and X2.
2. If we took a random sample of size n from a given population, the probability is 0.95 that the population mean would lie between interval X1 and X2 around the sample mean as shown.
3. If a random sample of size n was taken from a given population, we can be 95% confident in our assertion that the population mean will lie around the sample mean in the interval bounded by values of X1 and X2.

**Example**

The sponsor of a television programme targeted at the children’s market wants to find out the average amount of time children spend watching television.

A random sample of 100 children indicated the average time spent by these children per week was 27.2 hours.

From previous experience, the population SD is known to be 8 hours. A confidence level of 95% is considered to be adequate.

Solution

x̅ = 27.2

Z =1.96

SD = 8

n = 100

S.E.M = SD/root(n) = 8/10 = .8

Then.

X1 = x̅ – Z x S.E.M = 27.2 – 1.96 x 8 = 25.632

X2 = 27.2 + 1.96 x 8 = 27.2 + 1.568 = 28.768

1. How would the confidence interval limits change if the confidence level was increased from 95% to 99%

Ans

X1 = 27 – 2.58 x .5 = 27 – 1.29 = 25.71

X2 = 27 + 1.29 = 28.29

**Problem**

Ms Anne is a student in the college who wants to buy a used car. She has decided that she will buy a Buick Century of 1989 year model in a reasonably good condition.

She has randomly selected 100 sale advertisements from the local newspaper. She found the average price to be $ 4500. She also knows that the standard deviation of such used car prices in the area is $ 520.

Establish a 95% confidence interval estimate of the true average price for all used cars in this category.