M358K: Applied Statistics.

October 7^{th}, 2020.

Sample Mean: The Normal Sample.

Linear Combinations of Normals.

For any X and Y normal and any constants α and β, the random variable

$$\boxed{\alpha X + \beta \cdot Y}$$

is a random variable which is also NORMALLY DISTRIBUTED.

$\Rightarrow$ Any linear combination of any number of normals is normal.

The General Sample Mean.

Say that we are modelling a particular phenomenon in our population.

We plan to gather a well-chosen random sample

$$X_1, X_2, \ldots, X_n$$

where n is the sample size.

The random variables $X_1, X_2, \ldots, X_n$ are all:

- INDEPENDENT.

and also

- IDENTICALLY DISTRIBUTED

(since they share the population dist'n).

We write i.i.d.

Define the sample mean: it's

$$\overline{X}_n = \overline{X} := \frac{1}{n}\left(X_1 + X_2 + \cdots + X_n\right)$$

and it's a random variable. (a point estimator).

Generally speaking, we don't know the exact dist'n of this r.v.

Q: What is the expected value of the sample mean?

$\rightarrow$: $\mathbb{E}[\overline{X}] = \mathbb{E}\left[\frac{1}{n}(X_1 + \cdots + X_n)\right]$

$$= \frac{1}{n}\left(\mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n]\right)$$

$\uparrow$ linearity of expectation

$$= \frac{1}{n}\left(n \cdot \mathbb{E}[X]\right) = \mathbb{E}[X] \; (=: \mu_X)$$

$\uparrow$ X_i are i.d. $\qquad$ $\uparrow$ a representative of X_i

Q: What's its variance?

$\rightarrow$: $Var[\overline{X}] = Var\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right]$

$$= \frac{1}{n^2}\left(Var[X_1 + X_2 + \cdots + X_n]\right)$$

$$= \frac{1}{n^2}\left(\underline{Var[X_1]} + \underline{Var[X_2]} + \cdots + \underline{Var[X_n]}\right)$$

$\uparrow$ X_i independent

$$= \frac{1}{n^{\cancel{2}}}\left(\cancel{n} \cdot Var[X]\right) = \frac{1}{n} \cdot Var[X]$$

$\uparrow$ X_i are i.d.

$$\Rightarrow \quad SD[\bar{X}] = \frac{1}{\sqrt{n}} SD[X] = \frac{1}{\sqrt{n}} \sigma_X$$

University of Texas at Austin

Problem Set # 5

Sample mean: The normal sample.

Problem 5.1. The scores of individual students on the Advanced Dark Arts Exam are modeled as normally distributed with a mean of 19.6 and a standard deviation of 5.0. At Voldemort High, 64 seniors take the test. Assume the individual scores at this school are modeled using the same distribution as national scores. What is the sampling distribution of the sample average score for this random sample of 64 students?

State the **name** and the **parameter value(s)** of this distribution.

Problem 5.2. The *"Aristocratic Hog"* chocolate bars are all labeled to weigh 4.0 ounces. The distribution of the actual weights of these chocolate bars is modeled as normal with a mean of 4.0 ounces and a standard deviation of 0.1 ounces. Bernard, the quality control manager and principal taster, initially plans to take (and weigh) a simple random sample of size n from the production line. Then he reconsiders and decides that a sample twice as large is needed. By what factor does the standard deviation of the sampling distribution of the sample average change?

Problem 5.3. The individual students' scores in the ACT exam are modeled using the normal distribution with an unknown mean (say, it varies from year to year) and with the **known** standard deviation of 6.

You take a SRS of students who took the ACT this year. The intention is to use their sample average to estimate (infer) the population mean.

You want the standard deviation of your statistic $\bar{X}_n$ to be at most 0.10. What is the least number of students you need to sample?