Stat 260 Lecture Notes
Set 20 - Sampling Distributions, Random Samples, and Linear Combinations of Random Variables

Say we perform an experiment and record some observations:

$$x_1 = 6.12, \quad x_2 = 4.16, \quad x_3 = 3.19, \quad x_4 = 1.86.$$

From this we can get:

$$\overline{x} = 3.8325 \quad \text{and} \quad s = 1.79.$$

from Sx button on Calculator

And if we do the experiment again we might get observations:

$$x_1 = 5.08, \quad x_2 = 6.79, \quad x_3 = 4.43, \quad x_4 = 2.15.$$

From this we can get:

$$\overline{x} = 4.6125 \quad \text{and} \quad s = 1.92.$$

Before the experiment we didn't know the values of x_1, x_2, x_3, x_4. We could treat these as random variables X_1, X_2, X_3, X_4. We could also treat the mean as a random variable $\overline{X}$ and the standard deviation as a random variable S. (Note that we use capital letters for random variables, and lower case letters for the measured values of the random variables.)

A ***point estimate*** is a single-valued statistic that estimates a population parameter.

Sample   population

Either measured values of $\overline{x}$ give a point estimate for μ.
Either measured values of s give a point estimate for σ.

μ = population / actual / true / theoretical value of the mean.
σ = population / actual / true / theoretical value of the standard deviation.

$\widetilde{x}$ (the median) is also a point estimate for μ.

There are ways to tell which point estimate is a better one, you will talk about this more in Stat 261. As it turns out, $\overline{x}$ is a better than $\widetilde{x}$ to estimate μ.

This is also why we use

$$s^2 = \frac{\sum(x_i - \overline{x})^2}{n-1} \quad \text{instead of} \quad \frac{\sum(x_i - \overline{x})^2}{n}.$$

better estimate
for population
varience σ^2

Since $\overline{X}$ is a random variable, we could find the mean / expected value of $\overline{X}$, $E(\overline{X})$.
Suppose $X_1, X_2, \ldots, X_n$ all come from the same population, so $E(X_1) = E(X_2) = \ldots = E(X_n) = \mu$ and $V(X_1) = V(X_2) = \ldots = V(X_n) = \sigma^2$.

$$E(\overline{X}) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = E\left(\frac{X_1}{n} + \frac{X_2}{n} + \frac{X_3}{n} + \ldots + \frac{X_n}{n}\right)$$

$$= \frac{1}{n}E(X_1) + \frac{1}{n}E(X_2) + \ldots + \frac{1}{n}E(X_n)$$

$$= \frac{1}{n}\mu + \frac{1}{n}\mu + \ldots + \frac{1}{n}\mu$$

$$E(\overline{X}) = \mu = E(X)$$

$$= \cancel{n} \cdot \frac{1}{\cancel{n}}\mu = \mu$$

Also, if $X_1, X_2, \ldots, X_n$ are all independent we have

$$V(\overline{X}) = V\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = V\left(\frac{X_1}{n} + \frac{X_2}{n} + \ldots + \frac{X_n}{n}\right)$$

$$= \frac{1}{n^2}V(X_1) + \frac{1}{n^2}V(X_2) + \ldots + \frac{1}{n^2}V(X_n)$$

$$= \frac{1}{n^2}\sigma^2 + \frac{1}{n^2}\sigma^2 + \ldots + \frac{1}{n^2}\sigma^2$$

$$= n \cdot \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n} = \frac{V(X)}{n}$$

So $E(\overline{X}) = \mu$, $V(\overline{X}) = \dfrac{\sigma^2}{n}$, and $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$.

In particular, this tells us that if $\overline{X}$ follows a normal distribution, then we standardize as

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

$Z = \dfrac{\text{r.v} - \text{expected value}}{\text{s.t. dev of the r.v}}$

$Z = \dfrac{X-\mu}{\sigma}$

A **statistic** can be any function of a random variable (or of many random variables).

$\bar{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$ is a statistic

The probability distribution of a statistic is called a **sampling distribution**.

Random variables X_1, X_2, ..., X_n form a ***random sample*** if:

- All of X_1, X_2, ..., X_n have the same mean and variance.
- The distributions of X_1, X_2, ..., X_n all have the same shape.
- All the random variables X_1, X_2, ..., X_n are independent of each other.

This is also called ***independent and identically distributed (i.i.d.)***.

Rule: If X_1, X_2, ..., X_n are all normal random variables then any linear combination of X_1, X_2, ..., X_n is a normal random variable. That is, if X_1, X_2, ..., X_n are all normal then $c_1X_1 + c_2X_2 + \ldots + c_nX_n$ is a normal random variable.

This tells us that if X_1, X_2, ..., X_n are normal random variables then

$$\overline{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

is a normal random variable.

So if $X_1, X_2, \ldots, X_n$ are normal, this tells us $\overline{X}$ is also normal

$Z = \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$

Example 1: Suppose X, Y, W are independent normal random variables with

$E(X) = \mu$

$$\mu_X = 10, \quad \sigma_X = 4, \quad \mu_Y = -2, \quad \sigma_Y = 2, \quad \mu_W = 3, \quad \sigma_W = 1.$$

Find $P(X + 2Y - W > 7)$.

W, Y, X are normal
↳ X+2Y-W is also normal

* check for indep. *

$$Z = \frac{\text{r.v} - \text{expected value}}{\text{st. dev. of r.v.}}$$

$$\begin{aligned} E(X+2Y-W) &= E(X) + 2E(Y) - E(W) \\ &= 10 + 2(-2) - 3 \\ &= 10 - 4 - 3 \\ &= 3 \end{aligned}$$

$$\begin{aligned} V(X+2Y-W) &= V(X) + 4V(Y) + V(W) \\ &= 16 + 4(4) + (1) \\ &= 16 + 16 + 1 \\ &= 33 \end{aligned}$$

$$\sigma_{X+2Y-W} = \sqrt{33}$$

$$\begin{aligned} P(X+2Y-W > 7) &= P\left(Z > \frac{7-3}{\sqrt{33}}\right) \\ &= P(Z > 0.70) \\ &= 1 - P(Z \le 0.70) \\ &= 1 - 0.7580 \\ &= 0.2420 \end{aligned}$$

Example 2: The manufacturing of a certain component requires three different machining operations. The amount of time each operation requires (that is, the operation time) is normally distributed with a mean of 12 minutes and a standard deviation of 5 minutes. The three operation times are independent. Suppose the cost for the first machining operation is \$1 per minute, and that the cost for the second and third operations are \$2 per minute and \$3 per minute, respectively.

What is the probability that the total cost for making the next component is more than \$114?

X_1, X_2, X_3 = time needed for machines 1, 2, 3

Want $P(X_1 + 2X_2 + 3X_3 > 114)$

X_1, X_2, X_3 are normal $\Rightarrow X_1 + 2X_2 + 3X_3$ is also normal

$$Z = \frac{\text{r.v.} - \text{expect. value}}{\text{St. dev. of r.v}}$$

$$\begin{aligned} E(X_1 + 2X_2 + 3X_3) &= E(X_1) + 2E(X_2) + 3E(X_3) \\ &= 12 + 2(12) + 3(12) \\ &= 72 \end{aligned}$$

$$\begin{aligned} V(X_1 + 2X_2 + 3X_3) &= V(X_1) + 4V(X_2) + 9V(X_3) \\ &= 25 + 4(25) + 9(25) \\ &= 350 \end{aligned}$$

$$\sigma_{X_1 + 2X_2 + 3X_3} = \sqrt{350}$$

$$\begin{aligned} P(X_1 + 2X_2 + 3X_3 > 114) = P\left(Z > \frac{114 - 72}{\sqrt{350}}\right) &= P(Z > 2.24) \\ &= 1 - P(Z \le 2.24) \\ &= 1 - 0.9875 \\ &= 0.0125 \end{aligned}$$