Point Estimation

statistical inference : reasoning that sample data may be used to draw conclusions about a population.

∴ let's say some random variable X has an unknown population mean, μ.

∴ the sample mean $\bar{X}$ is a point estimator for μ and uses the notation

$$\hat{\mu} = \bar{X}$$

∴ an actual value of $\bar{x}$ (lower-case!) is called the point estimate and is a "reasonable value" for the unknown parameter μ.

- other examples of point estimators:

$$\hat{\sigma}^2 = S^2$$

population variance

sample variance

capitalized until we have an actual value!

$$\hat{\mu}_1 - \hat{\mu}_2 = \overline{X}_1 - \overline{X}_2$$

Central Limit Theorem: as $n \uparrow$, the sample mean of an unknown probability distribution tends to be normal, with $\bar{x} \to \mu$

this is huge

$$\mu_{\bar{x}} = \mu$$

mean value of $\bar{x}$

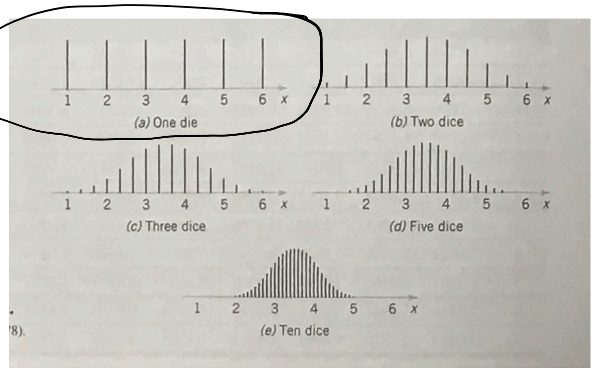
new Z formula:

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

∴ our previous z formula was for a <u>single</u> value of X. $(n=1)$

∴ note: for small sample sizes, z has a small value; large n scales accordingly

∴ Fig 7.1 p.228 (6th)



∴ probabilities associated w/ throwing dice

equally-likely outcomes; $f(x) = \frac{1}{n}$

(each # has probability of $\frac{1}{6}$)

∴ discrete <u>uniform</u> distribution

∴ note that when averaging throws of multiple dice, it <u>really</u> starts to look normal @ $n = 5$!!!

ex: resistor manufacturer <u>claims</u>

Supposed population parameters $\left\{ \begin{array}{l} \mu = 100\,\Omega \\ \sigma = 10\,\Omega \end{array} \right.$

$\mu = 100\,\Omega$ ← clearly a 100-Ω resistor

$\sigma = 10\,\Omega$ ← might reflect tolerance

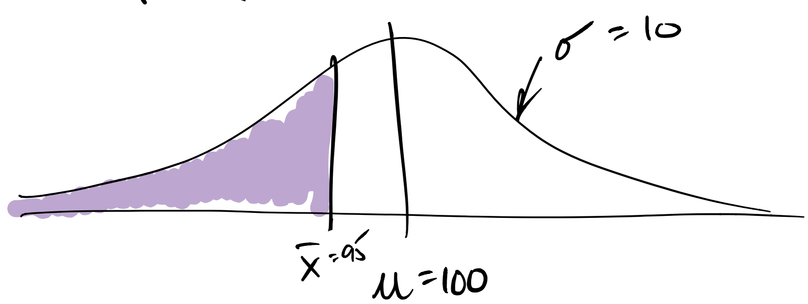
- you just measured 25 resistors, averaged the results, and got

$$\bar{x} = 95\,\Omega$$

- at first glance, ok! You're only half a standard deviation off

- but let's compute the probability of this happening, <u>if</u> μ and σ are true

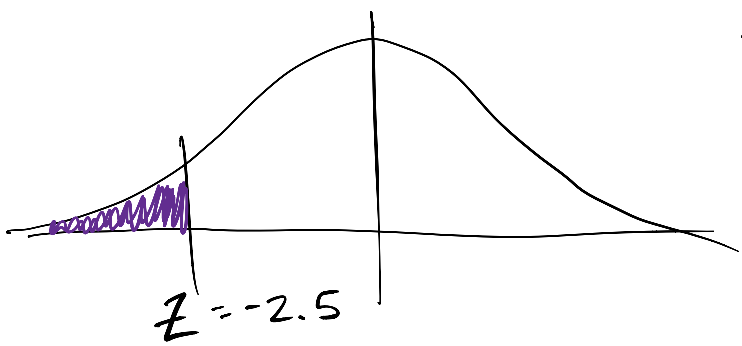
$$P(\bar{x} < 95)$$



$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{95 - 100}{10/\sqrt{25}}$$

$$Z = -2.5$$

← uh oh... that $n=25$ really blew up our Z-value to $2\frac{1}{2}$ standard deviations below the mean



from Z-table:

$$P(Z < -2.5) = 0.006210$$

or 0.6210%

∴ not very likely!

∴ in other words: the unknown parameters $\mu = 100$?? and $\sigma = 10$?? are probably not trustworthy!

.. a single measurement of 95Ω would have a z-value of

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{95 - 100}{10/\sqrt{1}} = -0.5$$

$$P(z < -0.5) = 0.308538$$

way more plausible!

.. measuring 25 and averaging, you shouldn't be that low.

.. two populations, difference in means:

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

ex: jet engine component has known mean lifespan of 5000 hours, std. dev. 40 hours

- improved component claims lifespan of 5050 hours, std. dev. 30 hours

∴ if 16 "old" components and 25 "new" components are tested, what is the probability that their sample means will differ by more than 25 hours?

$$\underbrace{\bar{X}_1 - \bar{X}_2 > 25}$$

new old

∴ $\mu_1 = 5050$ $\mu_2 = 5000$

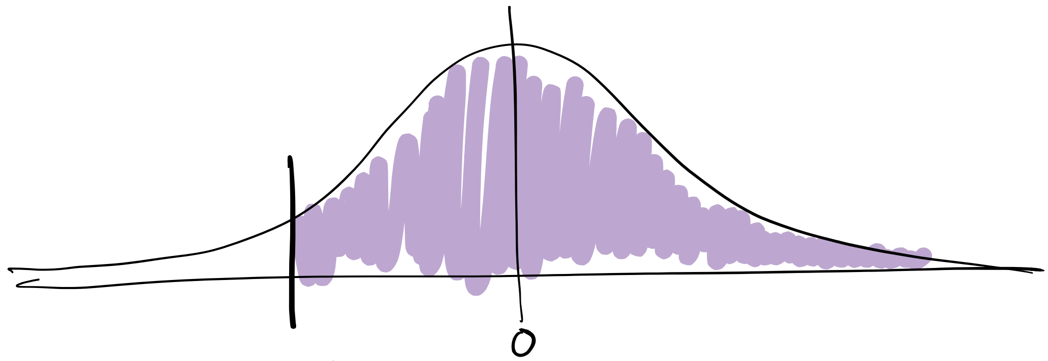
$\sigma_1 = 30$ $\sigma_2 = 40$

$n_1 = 25$ $n_2 = 16$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

$$= \frac{25 - (5050 - 5000)}{\sqrt{40^2/16 + 30^2/25}} = \frac{-25}{\sqrt{136}}$$

$$Z = -2.14$$



$Z = -2.14$

$P(Z > -2.14)$

$\because$ table is cumulative;

$= 1 - P(Z < -2.14)$

$1 - 0.016177$

$= 0.983823$

or 98.38%