Lecture 8 Confidence interval

- Parameter and estimate : population mean, sample mea
- Standard error of the mean (SE) Which one?
- 95% confidence interval
- Confidence level (coefficient) 1-α
- Using z score
- Two sample problem; matched sample problem
- Illustration with Computer simulation

Standard error refers to the standard deviation of an estimator

Measurement error=

eading from an instrument - true val

One biotech company specializing microarray gene expression profiling claims they can measure the expression level of a gene with an error of size .1 (t is, after testing their method numerous times, they found the standard deviation of their measurement errors is 0.1) The distribution of errors follow norm distribution with mean 0 (unbiased).

Cells from a tumor tissue of a patient are sent to this company f croarray assay. To assure consistency, the company repeat the a imes. The result of one gene, P53 (the most well-studied tumor ppressor gene), is 1.1, 1.4, 1.5, 1.2.

timate the true level of P53. What is the SE ? Find a 95% confid erval.

Answer :

- The sample mean is $(X_1+X_2+X_3+X_4)/4 = \overline{X}$
- Standard deviation of each X random variable is .1
- So SD($\overline{X}$) is .1/sqrt (4) = 0.05, this is the SE

- Two SE is 0.05 times 2 = .1
- So 95% confidence interval runs from 1.3-.1
- To 1.3 + .1; that is 1.2 to 1.4

- Another cell sample is prepared from a healthy person. The assay results are

1.5 1.6 1.4 1.7.

Question : Does the tumor cell have a lower P53 level? Find a 95% confidence interval for The difference.

Answer

- The estimator is the difference $\bar{X} - \bar{Y}$,
- Where the Y bar is average of 4 random variables again, so it should have the same

standard deviation as SD of X bar.

Now SD($\bar{X}$-$\bar{Y}$) = sqrt (var($\bar{X}$) + var ($\bar{Y}$))

= sqrt ($.05^2 + .05^2$) = 0.0707

So SE is 2 times .0707 = 0.1414

95% confidence interval goes from (1.3-1.55)-.1414

To (1.3-1.55) +.1414; that is from -.2914 to -0.0586

Using 95% confidence interval, there is a statistically significant reduction in P53 expression

Rationale behind

- Box A , true mean=1.35 SD=.1
- Box B true mean= 1.50 SD=.1
- Generate sample from box A
- Generate sample from box B
- Using computer
- Find the difference of mean, check 2 SE interval to see if covering the true difference
- Repeat it many times to see how often the interval covers the true difference

Changing the confidence level

- For a 95% confidence interval, use 2SE rule
- This is because of normal distribution - central area P{ -2<Z<2} is about .95
- For 80% confidence interval, you look for
- P{-c<Z<c}=.80; equivalently P {Z<c}=.80+.10=.90 so c must be 1.28