ALQUSOUR ACADEMY
اكاديمية القصور

Q.A.J.U.S.T

BIOSTATISTICS

لطلبة الصيدلة والعلوم الطبية

Subject:

Second Exam – Part Three

ALQUSOUR ACADEMY

خاص للفصل الدراسي الأول 2016/2015

Prod. Date: 15/12/2015 Pages: 10 Price: 35

Sampling Distributions

Definition:

Suppose that in a specified population, we randomly select all possible samples from this population and compute the mean($\bar{x}$) for each sample.
Then the probability distribution of $\bar{x}$ is called the sampling distribution of the mean.

To simplify, we have 2 cases:

(1) Sampling from normally distributed populations. Here, the distribution of the sampling mean is normally distributed and has the following characteristics:

1- $\mu\,\bar{x} = \mu$ of population.
2- $\sigma\,\bar{x} = \sigma/\sqrt{n}$ because $\sigma^2\,\bar{x} = \sigma^2/n$ (σ^2 is population variance)

This is true for any sample size.

(2) Sampling from not normally distributed populations. Here we refer to central limit theorem.

The Central Limit Theorem:

When sample size is large ($n \geq 30$), the sampling distribution of the sample means will be normally distributed and you can use the mean $\mu_{\bar{x}} = \mu$ and standard deviation of the sampling distribution of the mean $(\sigma_{\bar{x}}) = \sigma/\sqrt{n}$

- $\frac{\sigma}{\sqrt{n}}$ is called the standard error of the mean (SEM).
- $\frac{\sigma}{\sqrt{n}}$ is the best estimate for SEM.

Standard deviation vs SEM

- The standard deviation (s) is a measure of the variability in the population.
- The standard error of the mean (SEM) is a measure of the precision of the estimate of the mean and is dependent on sample size.
- The SEM does not describe the variability of the population.
- The variability of the sample means is smaller than the variability of the population.

To transform $\bar{x}$ to a Z-score:

$$z = \frac{\text{Value-Mean}}{\text{Standard Error}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Remember: $(\bar{x} - \mu)$ is sampling error.

Sampling Error

It is the discrepancy between sample and population.

The term sampling error does not mean a sampling mistake, it indicates that means drawn from multiple samples taken from a population will vary from each other due to random chance and therefore may deviate from the population mean.

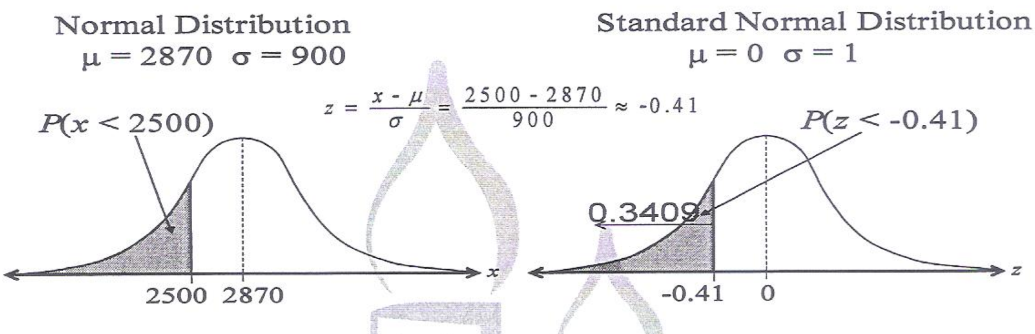
Example

A bank auditor claims that credit card balances are normally distributed, with a mean of JD 2870 and a standard deviation of JD 900.

Q1. What is the probability that a randomly selected credit card holder has a credit card balance less than JD 2500?

You are asked to find the probability associated with a value of variable *x*.

Solution:



$$P(x < 2500) = P(z < -0.41) = \mathbf{0.3409}$$

Interpretation: There is about a 34% chance that the credit card balance will be less than JD 2500.

Q2. You randomly select 25 credit card holders. What is the probability that their mean credit card balance is less than JD 2500?

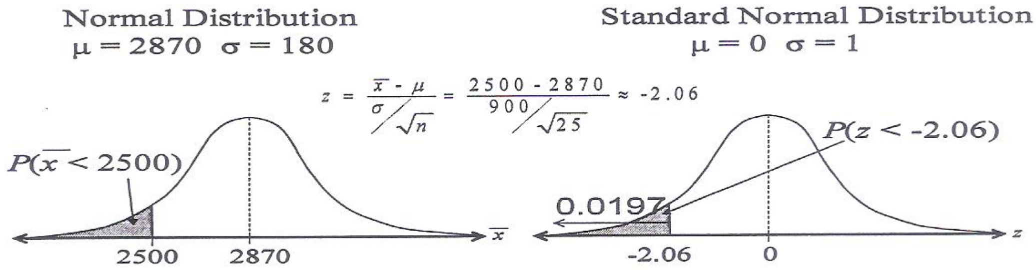
Solution:

You are asked to find the probability associated with a sample mean .

So, we are dealing with sampling distribution: then,

$$\mu_{\bar{x}} = \mu = 2870 \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{900}{\sqrt{25}} = 180$$

Solution:



$P(\bar{x} < 2500) = P(z < -2.06) = \mathbf{0.0197}$

Interpretation: There is about a 2% chance that the mean credit card balance will be less than JD 2500.

Example

What is the distribution of the sample mean of samples of size $n = 48$?

Answer:

According to central limit theorem, the sample size is large so the sample mean has approximate normal distribution.

Estimation

Remember:

(a) Statistic:
Is a descriptive measure computed from the data of a sample.

(b) Parameter:
Is a descriptive measure computed from the data of a population.

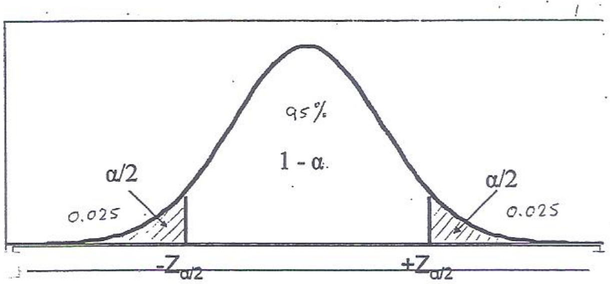
Estimation Concept:

- We estimate parameters from the sample, because we can't calculate it from population.

In estimation, no 100% . There is error.

- Level of confidence depends on the maximum allowable error, we call it α $(alpha)$.
- α is also called type 1 error or significance level, while $(1-\alpha)$ is confidence level.

If $\alpha = 0.05$, then the level of confidence = 95%.
Then α depends on confidence level.



Definitions:

A point estimate: is a single value used to estimate the population parameter.
Example: $\bar{x}$: this is our unbiased estimator of the population mean (μ) but it is probably not equal to the true mean. (in estimation we may have error)

An interval estimate: consists of two numbers defining a range that we feel includes the parameter being estimated.

Remember:
- 68 % of data lies between $\pm 1\ \sigma$ from the mean.
- 95 % of data lies between $\pm 2\ \sigma$ from the mean.
- 99.7% of data lies between $\pm 3\ \sigma$ from the mean.

Example

Blood pressure is normally distributed with mean 100 and standard deviation 10.
Find the interval in which 95% of the data lie.
Answer: 95% of data are between $\pm 2\ \sigma$ from the mean, so the interval is:
80 to 120.
Then 95% of blood pressure readings lie between 80 and 120.

Example

IQ is normally distributed with mean 80 and standard deviation 15. Find the interval in which 95% of the data lie?

Solution:

95% means interval limits are $\pm 2\sigma$ away from the mean. Then the lower limit is 50 and the upper limit is 110.
The interval is (50 to 110).

In fact, the Z-score associated with 95% area under the curve is 1.96.
(95% of all possible means are within ± 1.96 standard errors of the true mean).

So, the formula for a 95% confidence interval is:

C.I $= \bar{x} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$

General formula of Confidence interval: $\bar{x} \pm z_{\left(\frac{\alpha}{2}\right)} \times \frac{\sigma}{\sqrt{n}}$

Note: $z_{\left(\frac{\alpha}{2}\right)}$ is the Z value with the area $\left(\frac{\alpha}{2}\right)$ to the right.

Remember: In large sample size the distribution is normal (Central Limit Theorem) so we use Z value in the interval.

Example

Let $\sigma = 21$, $\bar{x} = 28$, and $n = 49$, and suppose that we want to construct a 95% confidence interval for μ.

(a) The point estimate for the mean?

Solution: $\bar{x} = 28$.

(b) The standard error?

Solution: $\frac{\sigma}{\sqrt{n}} = \frac{21}{\sqrt{49}} = 3$.

(c) What is the 95% confidence interval for μ?

Solution: C.I. is: 28 ± 1.96 (3) = 28 ± 5.88, or in the form (22.12, 33.88).

<u>Interpretation of the above 95% C.I:</u>

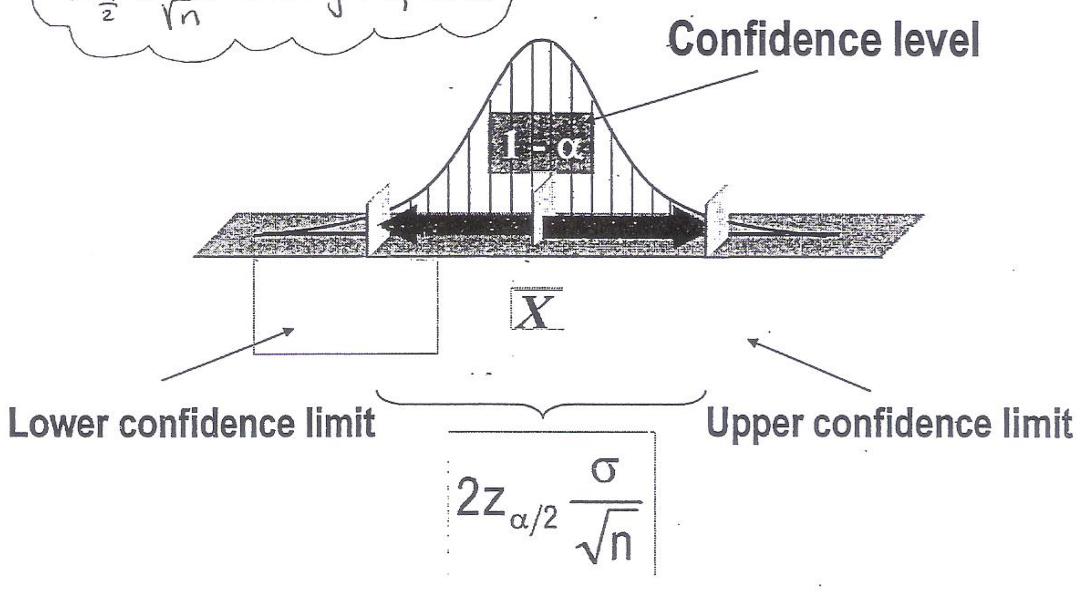
<u>Correct:</u>

- We are 95% confidence that the true population mean (of all people in population) lies within the interval (22.12 to 33.88).
- 95% of the time, in repeated sampling, the interval calculated from the same sample size will include the true μ.

<u>Incorrect:</u>

- The probability that the mean lies between 22.12 and 33.88 is 0.95.
- We are 95% confident that the <u>sample</u> mean for the observed sample lies in the interval.
- We are 100% confident that the sample mean is true with no error margin.

$Z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}$ = margin of error



$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

- Confidence interval estimation gives information about closeness to unknown population parameter.
- Confidence interval estimation is stated in terms of probability (never 100% sure).
- Increasing the sample size will make the confidence interval narrower (more precise) because SE will be small.
- Confidence intervals provide a measure of the precision of the estimate of the mean from one sample.
- Different values of the mean $\bar{x}$ changes the location of the interval.
- The interval does not predict what might happen for another sample.

Questions

In order to estimate the average waiting time in outpatient clinics at Prince's Basma Hospital, data were collected for a random sample of 81 patients over a one week period. The sample mean is 4. Assume the population standard deviation is 6 hours. Answer questions 1 and 2:

1) **The standard error of the mean is:**
   a. 0.777
   b. 0.444
   c. 0.667
   d. 0.333

2) **With a 0.95 confidence, the confidence interval is:**
   a. 1.31 - 2.69
   b. 2.26 - 5.22
   c. 1.99 - 3.22
   d. 2.69 - 5.31

3) 99% confidence interval for the mean age of Jordanians was computed to be (29.8 to 41.3). What is the interpretation attached to this interval?

a. We are 99% confident that the mean age of Jordanians is between 29.8 and 41.3.

b. 99% of the residents in our sample had ages between 29.8 and 41.3.

c. We are 99% confident that the mean age of Jordanians in our sample is between 29.8 and 41.3.

d. All of the above are valid interpretations.

Answers Key

	Answer
1	c. 0.667
2	d. 2.69 - 5.31
3	a. We are 99% confident..