SA1 Tutorial I

Manjusha Kancharla

20/04/2019

Agenda

Review and Practice on:

- Random Variables
- Probability Distributions
- Properties of Normal distributions
- Sampling Distributions
- Central Limit Theorem (CLT)
- Interval Estimation

Practice Problem-1

Worker participation in management is a new concept that involves employees in corporate decision making. The following data are the percentages of employees involved in worker participation programs in a sample of firms. Draw a Box-plot of the data and comment on the plot.

5, 32, 33, 35, 42, 43, 42, 45, 46, 44, 47, 48, 48, 48, 49, 49, 50, 37, 38, 34, 51, 52, 52, 47, 53, 55, 56, 57, 58, 63, 78, 43, 56, 47, 57, 61, 44, 60, 38, 52, 54, 51, 62, 35, 45, 46, 49, 42, 32, 59, 63, 33, 55, 34, 36, 53, 39, 48, 37, 50, 41

Practice Problem-1: Solution

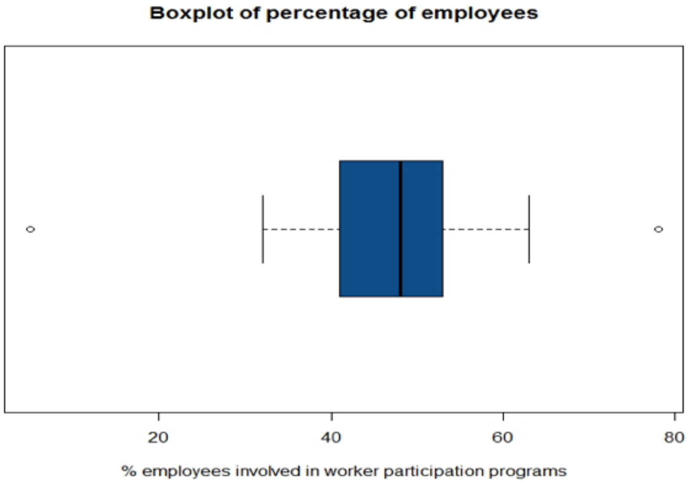

Figure 1: Box Plot

Practice Problem-2

The time required for putting together a food order at a restaurant is normally distributed with $\mu = 45$ min and $\sigma = 8$ min. The restaurant manager plans to have work begin on an order 10 minutes after the order has been placed and the customer is told that their food will be ready within 1 hour from order. What is the probability that the restaurant manager cannot meet his commitment?

Practice Problem-2: Solution

- Let T be the time it takes to work on an order
- We have that $T \sim N(45, 8^2)$
- The work begins 10 minutes after getting the order
- So we need the order to be completed in $t \leq 50$ minutes

Question: Why am I using 't' instead of 'T' here?

Practice Problem-2: Solution

What is the probability that the restaurant manager cannot meet his commitment?

- Convert this into math!

Practice Problem-2: Solution

What is the probability that the restaurant manager cannot meet his commitment?

- Convert this into math!
- $P(T \leq 50)$

What do we do next?

Practice Problem-2: Solution

What is the probability that the restaurant manager cannot meet his commitment?

- Convert this into math!
- $P(T \leq 50)$

What do we do next?

- Convert to Z (Standard Normal)
- $Z = \frac{T-\mu}{\sigma} \sim N(0,1)$
- $P(T \leq 50) = P(Z \leq \frac{50-45}{8}) = P(Z \leq 0.625)$
- In excel try function **=1 – NORMSDIST(0.625)**
- Or **=1 – NORMDIST(50, 45, 8, 1)**

Practice Problem-2: Solution

What is the probability that the restaurant manager cannot meet his commitment?

$$P(T \leq 50) = 1 - P(\geq 50) = 0.2659$$

Practice Problem-3

PALCO Industries, Inc., is a leading manufacturer of cutting and welding products. One of the company's products is an acetylene gas cylinder used in welding. The amount of nitrogen gas in a cylinder is a normally distributed random variable with mean 124 units of volume and standard deviation 12. We want to find the amount of nitrogen x such that 10% of the cylinders contain more nitrogen than this amount.

Practice Problem-3: Solution

- Let X: Amount of nitrogen gas in a cylinder
- Given that $X \sim N(124, 12^2)$

We want to find the amount of nitrogen x such that 10% of the cylinders contain more nitrogen than this amount.

- Translate this to math

Practice Problem-3: Solution

- Let X: Amount of nitrogen gas in a cylinder
- Given that $X \sim N(124, 12^2)$

We want to find the amount of nitrogen x such that 10% of the cylinders contain more nitrogen than this amount.

- Find 'x' such that $P(X > x) = 0.10$.
- Is this equalivalent to finding 'z' such that $P(Z > z) = 0.10$?

Practice Problem-3: Solution

- Let X: Amount of nitrogen gas in a cylinder
- Given that $X \sim N(124, 12^2)$

We want to find the amount of nitrogen x such that 10% of the cylinders contain more nitrogen than this amount.

- Find ‘x’ such that $P(X > x) = 0.10$.
- Is this equalivalent to finding ‘z’ such that $P(Z > z) = 0.10$?
- YES!
- Find z and obtain x value using $x = \mu + z\sigma$.

Practice Problem-3: Solution

- Find 'z' such that $P(Z > z) = 0.10$?
- Look up 'z' value in Normal table

Practice Problem-3: Solution

- Find 'z' such that $P(Z > z) = 0.10$?
- Look up value in Normal table
- z = 1.28
- Then x value reqiured is
  $x = \mu + z\sigma = 124 + 1.28 * 12 = 139.36.$

Practice Problem-4

If $X_1 \sim N(\mu, \sigma^2)$ and $X_2 \sim N(\mu, \sigma^2)$ are iid normal random variables, then what's the difference between $2X_1$ and $X_1 + X_2$? Discuss both their distributions and parameters.

Practice Problem-4: Solution

Refer to lecture notes 1 slide # 27-28.

- $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$
- X_1, X_2 are independent
- $Y = aX_1 + bX_2$
- $E(Y) = a\mu_1 + b\mu_2$
- $V(Y) = a^2\sigma_1^2 + b^2\sigma_2^2$
- $Y \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

Use this for our problem: $X_1 \sim N(\mu, \sigma^2)$ and $X_2 \sim N(\mu, \sigma^2)$ are iid normal and the Ys are $2X_1$ and $X_1 + X_2$.

Practice Problem-4: Solution

Consider $Y_1 = 2X_1$

- $E(Y_1) = 2 * E(X_1) = 2\mu$
- $V(Y_1) = 2^2 * V(X_1) = 4\sigma^2$
- $Y_1 \sim N(2\mu, 4\sigma^2)$

Practice Problem-4: Solution

Consider $Y_2 = X_1 + X_2$

- $E(Y_2) = E(X_1) + E(X_2) = 2\mu$
- $V(Y_2) = V(X_1) + V(X_2) = 2\sigma^2$
- $Y_2 \sim N(2\mu, 2\sigma^2)$

What is the difference between Y_1 and Y_2?

Practice Problem-4: Solution

- $Y_1 \sim N(2\mu, 4\sigma^2)$
- $Y_2 \sim N(2\mu, 2\sigma^2)$

What is the difference between Y_1 and Y_2?

- Both have normal distributions, but the second one has lower variance.

Practice Problem-5

Auditors at a small community bank randomly sample 100 withdrawal transactions made during the week at an ATM machine located near the bank's main branch. Over the past 2 years, the average withdrawal amount has been \$50 with a standard deviation of \$40. Since audit investigations are typically expensive, the auditors decide to not initiate further investigations if the mean transaction amount of the sample is between \$45 and \$55. What is the probability that in any given week, there will be an investigation?

Practice Problem-5: Solution

We have:

- X: Withdrawal amount
- $\mu = 50, \sigma = 40, n = 100$

We are interested in the probability of an investigation. Which is same as the probability that the sample mean is beyond \$45 and \$55.

How would you go about findin this? What do we need to find a probability?

Practice Problem-5: Solution

We have:

- X: Withdrawal amount
- $\mu = 50, \sigma = 40, n = 100$

We are interested in the probability of an investigation. Which is same as the probability that the sample mean is beyond \$45 and \$55.

How would you go about findin this? What do we need to find a probability?

- A probability distribution!
- We don't have one!

Practice Problem-5: Solution

CLT to the rescue.

The distribution of the sample mean:

- will be normal when the distribution of data in the population is normal
- will be approximately normal even if the distribution of data in the population is not normal, under some conditions:
- Each data point in the sample is independent of the other
- The sample size is large enough (30 usually considered large enough)
- If data is quite symmetric and has few outliers, even smaller samples are fine. Otherwise, we need larger samples

If these conditions hold:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

- The standard deviation here, $\frac{\sigma}{\sqrt{n}}$, is called the Standard Error of the Mean, $SE(\bar{X})$

Practice Problem-5: Solution

- From Central Limit Theorem, we know that $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.
- The $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}} = 4$
- $\bar{X} \sim N(50, 4^2)$
- Find $P(45 \leq \bar{X} \leq 55) = P(-1.25 \leq Z \leq 1.25) = 0.79$
- But we want probability outside of this range
- Probability that the sample mean is beyond \$45 and \$55 is 1-$0.79 = 0.21$!

Practice Problem-6

A book publisher monitors the size of shipments of its textbooks to university bookstores. For a sample of texts used at various schools, the 95% confidence interval for the size of the shipment was 250 ± 45 books. Which, if any, of the following interpretations of this interval, are correct?

1. All shipments are between 205 and 295 books
2. 95% of shipments are between 205 and 295 books.
3. The procedure that produced this interval generates ranges that hold the population mean for 95% of samples.
4. If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295.
5. We can be 95% confident that the range of 160 to 340 holds the population mean.

Practice Problem-6: Solution

All shipments are between 205 and 295 books

- Incorrect. The interval describes, with 95% confidence, the location of the average shipment size μ, not the sizes of individual shipments

95% of shipments are between 205 and 295 books

- Incorrect. The interval does not describe individual shipments

The procedure that produced this interval generates ranges that hold the population mean for 95% of samples.

- Correct. 95% of intervals created in this fashion contain the true population mean.

Practice Problem-6: Solution

If we get another sample, then we can be 95% sure that the mean of this second sample is between 205 and 295.

- Incorrect. The interval does not describe the mean of another sample.

We can be 95% confident that the range of 160 to 340 holds the population mean.

- Incorrect. The interval does not correspond to a 95% confidence level.

Practice Problem-7

A survey of 5,250 business travelers worldwide conducted by OAG Business Travel Lifestyle indicated that 91% of business travelers consider legroom the most important in-flight feature. (Angle of seat recline and food service were second and third, respectively.) Give a 95% confidence interval for the proportion of all business travelers who consider legroom the most important feature.

Practice Problem-7: Solution

Identify the random variable here.

Practice Problem-7: Solution

Identify the random variable here.

Proportion of business travelers who consider leg-room to be the most important in-flight feature.

Practice Problem-7: Solution

We want a 95% CI for a proportion.

- Let π be the population proportion
- An estimate is the sample proportion, $p =?$

Practice Problem-7: Solution

We want a 95% CI for a proportion.

- Let π be the population proportion
- An estimate is the sample proportion, $p = 0.91$

The $(1 - \alpha)$ *100% CI for π is :

$$p \pm z_{\alpha/2} \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

Practice Problem-7: Solution

The $(1-\alpha)$ *100% CI for π is :

$$p \pm z_{\alpha/2}\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

- $\alpha = 0.05$, $z_{0.025} = 1.96$, n $= 5250$

$$0.91 \pm 1.96 * \frac{\sqrt{0.91 * (1-0.91)}}{\sqrt{5250}} = (0.90, 0.92)$$

Practice Problem-8

A university wants to know more about the knowledge of students regarding international events. The are concerned that their students are uninformed in regards to new from other countries. A standardized test is used to assess students knowledge of world events (national reported mean=65, S=5). A sample of 30 students are tested (sample mean=58, Standard Error=3.2). Compute a 99 percent confidence interval based on this sample's data. How do these students compare to the national sample?