# ****MODEL 3 CENTRAL LIMIT THEOREM - Introduction****

**1. Pengantar melalui contoh dadu**:

* Saat melempar satu dadu, hasilnya akan berada antara 1 dan 6, dengan distribusi **uniform**. Artinya, semua angka memiliki peluang yang sama untuk muncul.
* Ketika dua dadu dilempar dan dijumlahkan, hasilnya bisa dari 2 hingga 12, tetapi angka di tengah seperti 6 akan lebih sering muncul karena lebih banyak kombinasi yang bisa mencapainya (misalnya, 1+5, 2+4, dll).
* Jika jumlah dadu yang dilempar bertambah, distribusi hasil penjumlahannya mulai mendekati **distribusi normal**

**2. Fenomena CLT:**

* Awalnya distribusi dadu adalah **uniform**, namun setelah banyak lemparan, distribusi hasil penjumlahan mendekati distribusi **normal**.
* CLT menunjukkan bahwa dengan cukup banyak percobaan, bahkan jika distribusi awalnya tidak normal, penjumlahan atau rata-rata variabel acak akan mendekati distribusi normal.

**3. Detail teknis CLT:**

* Jika X1,X2,...,XnX\_1, X\_2, ..., X\_nX1​,X2​,...,Xn​ adalah variabel acak iid dengan mean μ dan deviasi standar σ, maka:
  + Jumlah dari variabel-variabel acak tersebut Sn akan terdistribusi normal dengan mean n.μ (n kali mu) dan deviasi standar σ.√n (sigma kali akar n).
  + Rata-rata X̅ dari variabel acak tersebut akan terdistribusi normal dengan mean μ dan deviasi standar σ/√n.
* **Catatan**: Distribusi mendekati normal biasanya ketika nnn besar (lebih dari 30), namun kadang-kadang distribusi normal dapat muncul bahkan jika sampel lebih kecil dari 30, tergantung pada distribusi aslinya.

**4. Mengapa CLT penting?:**

* CLT **tidak mempedulikan** apa bentuk distribusi asli dari variabel acak. Bahkan jika distribusi asli tidak normal, jumlah atau rata-rata variabel acak tersebut akan mengikuti distribusi normal jika jumlahnya cukup besar.
* Hal ini sangat penting dalam **membentuk interval kepercayaan** dan **melakukan uji hipotesis**, sehingga distribusi normal sering digunakan dalam praktik statistik.

**5. Contoh penerapan CLT:**

* Pelemparan dadu memberikan contoh praktis bagaimana hasil dari penjumlahan variabel acak yang independen bisa menghasilkan distribusi normal meskipun distribusi awalnya tidak normal.

6. Kesimpulan:

Poin-poin utama ini menyoroti bagaimana Central Limit Theorem memungkinkan penerapan distribusi normal secara luas dalam statistik, meskipun data asli tidak mengikuti distribusi normal.

## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. Why is the central limit theorem so important?

Ans:

* It is able to disregard the distribution that some underlying X follows.
* The distribution of a sum approaches the normal distribution. This occurs while the distribution of terms in the underlying distribution are not necessairly normal.
* It can be used for making confidence intervals.
* It is commonly accepted and used in practice.

Explanation:

The central limit theorem is important because of all of the answers above. This is why it is commonly accepted and used in practice.

# ****MODEL 3 CENTRAL LIMIT THEOREM – Quick Review of Normal Distribution****

**1. Tinjauan Umum Distribusi Normal**

* Distribusi normal adalah distribusi probabilitas yang paling sering digunakan dalam statistik, terutama di bidang manajemen inventori dan rantai pasokan. Distribusi ini simetris terhadap rata-rata (μ) dan memiliki simpangan baku (σ) yang menentukan sebarannya.
* Penulisan distribusi normal umumnya adalah ~N(μ, σ), di mana:
  + μ adalah rata-rata.
  + σ adalah simpangan baku.

2. Distribusi Normal Standar (Unit Normal)

* Distribusi normal standar (ditulis sebagai ~N(0, 1)) memiliki rata-rata 0 dan simpangan baku 1.
* Transformasi ke distribusi normal standar dilakukan menggunakan statistik z:
  + Z =

di mana X adalah nilai variabel acak, μ adalah rata-rata, dan σ adalah simpangan baku.

3. Tiga Pertanyaan Utama dalam Masalah Distribusi Normal

* Ada tiga tipe utama pertanyaan yang biasa diselesaikan menggunakan distribusi normal:
  + Probabilitas bahwa X kurang dari suatu nilai k (P(X ≤ k)):
    - Ini dapat dihitung menggunakan fungsi Excel:
      * P(X ≤ k) =NORM.DIST(k, μ, σ, 1)
    - Contoh: Diberikan ~N(24,8), probabilitas bahwa X ≤ 30 adalah:
      * P(X ≤ 30) =NORM.DIST(30, 24, 8, 1) = 0.773
    - Alternatifnya, kita bisa menggunakan distribusi normal standar:
      * Z = = 0.75
      * P(Z ≤ 0.75) =NORM.S.DIST(0.75, 1) = 0.773
  + 2) Probabilitas bahwa X lebih besar dari suatu nilai k (P(X > k)):
    - Digunakan pendekatan yang mirip dengan menghitung probabilitas kurang dari k, tetapi kita mengurangi hasilnya dari 1 karena luas total distribusi adalah 1:
      * P(X > k) = 1 - NORM.DIST(k, μ, σ, 1)
      * Contoh: Probabilitas bahwa X > 30 untuk distribusi ~N(24, 8):
        + P(X > 30) = 1 - NORM.DIST(30, 24, 8, 1) = 0.227
  + 3) Probabilitas bahwa X berada di antara dua batas, yaitu antara (μ - b) dan (μ + b):
    - Probabilitas ini dihitung dengan mengurangi probabilitas di bawah batas bawah dari probabilitas di bawah batas atas:
      * P(μ- b ≤ X ≤ μ+b) =NORM.DIST(μ+ b, μ, σ, 1) - NORM.DIST(μ- b, μ, σ, 1)
      * Contoh: Untuk distribusi ~N(24,8) dan b = 4, kita menghitung probabilitas bahwa X berada di antara 20 dan 28:
        + P(20 ≤ X ≤ 28) =NORM.DIST(28, 24, 8, 1) - NORM.DIST(20, 24, 8, 1) = 0.69 - 0.31 = 0.38
        + Jadi, 38% dari probabilitas terletak di antara 20 dan 28.

**4. Menentukan Batas untuk Probabilitas Tertentu**

* Jika kita ingin mengetahui batas b di mana sejumlah persen probabilitas berada dalam interval antara (μ - b) dan (μ + b), kita bisa menggunakan fungsi NORM.S.INV untuk menghitung nilai z, lalu mengalikannya dengan simpangan baku:
  + b = z . σ
  + Contoh: Untuk distribusi ~N(24, 8) dan probabilitas 90%, kita ingin mengetahui batas b di mana terdapat 90% probabilitas di antara μ - b dan μ + b.
    - Nilai z untuk 90% adalah 1.64 (dapat dari NOMR.S.INV(0.95, kenapa 0.95? karna setelah mendapatkan z = 1.64 otomatis sebelah kiri 1.64 juga alias -1.64. jadi kalau p = 95% maka NORM.S.INV(0.975)), sehingga:
      * b = 1.64 . 8 = 13.12
      * Batas atas = 24 + 13.12 = 37.12
      * Batas bawah = 24 - 13.12 = 10.88
      * Jadi, ada probabilitas 90% bahwa variabel acak X akan berada di antara 10.88 dan 37.12.

## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. How does one convert from a normal distribution to a standard unit normal distribution using the Z statistic?

Ans:

Z =

Q2. Which formula would you use to determine the probability that some value X falls between μ and ±b for any distribution?

Ans:

P(X<μ+b) – P(X<μ-b)

# ****MODEL 3 APPLYING CENTRAL LIMIT THEOREM****

* Contoh Pemilihan Pesanan I:
  + Bagian ini memperkenalkan skenario di mana Anda mengelola pusat pemenuhan e-commerce. Tim Anda terdiri dari pemetik (pickers) yang merakit pesanan untuk dikirim. Rata-rata, mereka bisa memetik 4 pesanan per menit, dengan deviasi standar 5 pesanan. Ada selalu pesanan dalam antrian, jadi pekerjaan tidak akan habis. Masalah yang harus diselesaikan adalah memperkirakan kemungkinan bahwa tim Anda akan memproses setidaknya 260 pesanan dalam satu jam.
    - Informasi utama:
      * Rata-rata (μ) = 4 pesanan per menit
      * Deviasi standar (σ) = 5 pesanan per menit
      * Ada 60 menit dalam satu jam
      * Total pesanan yang diharapkan dalam satu jam adalah E[S60] = 604 = 240 pesanan, dan deviasi standarnya adalah σ60 = √(60) 5 = 38,7

**CATATAN:**

Standar Deviasi Total=Standar Deviasi Per Menit×√(Jumlah Menit​)

* + - * Menggunakan Teorema Limit Sentral, ini bisa diasumsikan mengikuti distribusi normal N(240, 38,7)
      * Untuk mencari probabilitas pemetikan setidaknya 260 pesanan, kita menghitung area di bawah kurva normal untuk nilai yang lebih besar dari 260. Ini dilakukan dengan fungsi norm.dist:
        + P(X ≥ 260) = 1 - P(X < 260) = 1 - NORM.DIST(260, 240, 38.7, 1) = 1 - 0.697 = 0.303
      * Jawaban: Sekitar 30% kemungkinan tim Anda akan memetik setidaknya 260 pesanan dalam satu jam.
* Contoh Pemilihan Pesanan II):
  + Masih menggunakan kondisi yang sama, di mana pesanan yang diharapkan dipetik dalam satu jam terdistribusi secara normal dengan rata-rata 240 pesanan dan deviasi standar 38,7. Kali ini, pertanyaannya adalah mencari probabilitas bahwa tim Anda akan memetik antara 220 dan 260 pesanan dalam satu jam.
    - Distribusi normal bersifat simetris. Karena kita sudah mengetahui bahwa P(X ≥ 260) = 30%, maka probabilitas P(X ≤ 220) juga sama, yaitu 30%. Jadi, probabilitas untuk memetik antara 220 dan 260 pesanan adalah 40%.
    - Jawaban: Sekitar 40% kemungkinan tim Anda akan memetik antara 220 dan 260 pesanan dalam satu jam.
* Menggunakan Nilai Z):
  + Perhitungan probabilitas dijelaskan dengan lebih detail menggunakan nilai Z. Dengan menggunakan formula Z = , kita bisa menghitung nilai Z untuk batas atas dan bawah, yaitu 260 dan 220 pesanan.
    - Z untuk batas atas (260 pesanan): Z = = 0.517
    - Z untuk batas bawah (220 pesanan): Z = = -0.517
    - Selanjutnya, kita cari probabilitasnya menggunakan distribusi normal standar:
      * Probabilitas untuk Z < 0.517 (=NORM.S.DIST(0.157,1)) adalah 0.697
      * Probabilitas untuk Z < -0.517 adalah 0.303
    - Jadi, probabilitas untuk memetik antara 220 dan 260 pesanan adalah 0.697 - 0.303 = 0.394 atau sekitar 40%.
    - Jawaban akhir: Sekitar 40% dari waktu, tim Anda akan memetik antara 220 dan 260 pesanan dalam satu jam.

## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. Assuming normal distribution, suppose you work in a trucking company where the average truck travels for 15 hours with a standard deviation of 7 hours:

Part 1

If you add up the travel time for the next three trucks arriving at your warehouse, what is the expected value of these three trucks?

Ans:

E [S3] = 153 = 45

Explaination:

The expected value for the next three trucks is the mean of the trucks times the number of trucks or simply 45.

Part 2

If you add up the travel time for the next three trucks arriving at your warehouse, what is the expected standard deviation of these three trucks?

Ans:

StdDev[S3] = 7 √3 = 12.1

Explaination:

The standard deviation of these three trucks is given as √(σ2  ntrucks)

### PRACTICE PROBLEMS

PP1 – Chiquita Republic

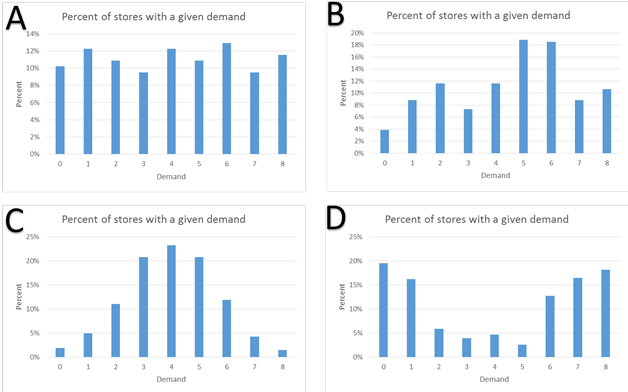
Chiquita Republic's Fire Red Shirt

Chiquita Republic is a chain of one thousand apparel stores located across Europe selling high-end clothing to a young demographic, with appealing designs at reasonable prices. You have been hired by Chiquita Republic as marketing manager for some of their products. One of these products, affectionately known as the Fire Red Shirt and identified with the SKU 20139, seems to be performing well, so you have been keeping an eye on its sales across Europe.

Based on data from earlier in the year, you know that demand for SKU 20139 in any one of Chiquita Republic's one thousand stores across Europe seems to be uniformly distributed between 0 and 8 units of the Fire Red Shirt sold in that store on that business day. This means that, at the end of any given business day, any one of the one thousand stores in Europe is equally likely to report 0, 1, 2, 3, 4, 5, 6, 7 or 8 units of SKU 20139 sold that day in that store.

Question 1

This morning, when you got to the office, you decided to print out four charts, summarizing yesterday's sales of four different SKUs in all of Chiquita Republic's one thousand stores across Europe. One of these charts pertains to SKU 20139, and the other three are about other SKUs. Each graph tells you the relative frequency with which a given number of units of that SKU were sold across the one thousand stores. For example, the first bar in the top left graph says that approximately 10% of Chiquita Republic's stores (or about 100 stores) sold 0 units of that SKU during the business day yesterday.



Unfortunately, you forgot to label the charts, and you tripped on the way back from the printer, and the charts got shuffled, so you are no longer certain about which one of the four charts corresponds to yesterday's sales of the Fire Red Shirt. Determined to figure this out through the magic of statistical insight, you lay out the four charts in front of you in the desk, and write a letter on the upper left corner of each graph: A, B, C and D.

Given what you know about the nature of the demand of the Fire Red Shirt in stores across Europe, which one of the four charts is most likely to be the chart for yesterday's sales of SKU 20139 at the store level?

Ans:

Graph A

Explanation:

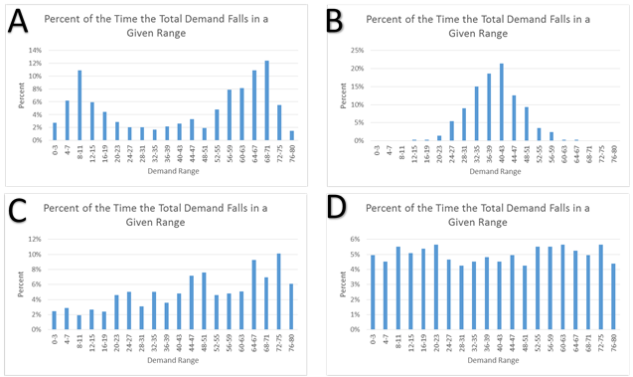
The fire red shirt is equally likely to sell any number of shirts from 0 to 8. This should be a uniform distribution. The graph closest to a uniform distribution is graph A.

Question 2

You are having a terrible day in terms of balance, because - while coming back from the printer a second time - you tripped again, and a new batch of charts that you were carrying got mixed up. This time, the charts involved in the incident were not about demand at the store level, but the demand at the distribution center (DC) level of four SKUs, one of them the Fire Red Shirt.

Chiquita Republic has several dozen distribution centers spread around Europe, and each DC serves approximately several dozen stores every day. The daily demand for a given SKU as seen by a DC is an aggregate of the daily demands of that SKU in the stores served by that DC.

Since you know what the demand for SKU 20139 looks like at the store level, you think you can apply your wide statistical knowledge to the crucial question of figuring out which one of the four charts corresponds to the demand for the Fire Red Shirt at the DC level.



Which one of these four charts is most likely to correspond to the demand for SKU 20139 at the DC level?

Ans:

Graph B

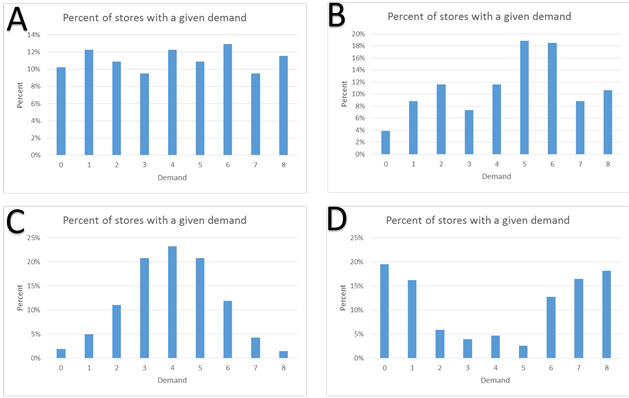
Explanation:

Using the central limit theorem, the sum of independent random variabes tends to move toward the normal distribution. The graph that most closely represents the normal disribution is graph B.

Question 3

Another one of Chiquita Republic's products that is under your supervision is SKU 20142, the world-famous Paradise Breeze Sandals. The sales of Paradise Breeze Sandals skyrocket when the weather is warm and breezy, and plummet when the weather is cold and rainy. When you realized the relationship between the weather and sales, a few months ago, you asked the stores to start including in their daily reports a note on their local weather. Yesterday, half of Chiquita Republic's stores reported warm and breezy weather, while the other half reported cold and rainy weather.

Now, you remember the four graphs with store level demand that you mixed up when you tripped on your way back from the printer the first time? One of them was for the Fire Red Shirt, but another one was for the Paradise Breeze Sandals. Below are these four charts again.



Which one of these charts is most likely to be the one for demand for SKU 20142 at the store level?

Ans:

Graph D

Explanation:

You know that the demand for this sku is highly dependent on the weather. Since half of Chiquita Republic's stores had bad weather yesterday and the other half had good weather, you should expect to see a lot of low sales and a lot of high sales, but not much in between. The graph that most closely represents this bimodal distribution is graph D.

# ****MODEL 3 SAMPLING****

**Sampling and Confidence Intervals: Understanding Through the MoonDoe Café Example**

**Sampling: The Basics**

Sampling is a statistical method used to make inferences about a population when gathering data from every individual in the population is impractical. For example, MoonDoe Café has over a million daily customers, and we need to understand customer behavior, particularly how long they spend in the café.

1. Population vs. Sample:
   * Population: This refers to all 1 million daily customers at MoonDoe Café. The population's characteristics (mean and standard deviation) are denoted by μ (mean) and σ (standard deviation).
   * Sample: Instead of analyzing the entire population, we take a smaller, random sample of customer visits (say, 50 visits). The sample characteristics (mean and standard deviation) are denoted by x̄ (mean) and s (standard deviation).
2. Why Random Sampling?
   * A random sample ensures that every customer visit has an equal chance of being selected, which helps produce an unbiased estimate of the population parameters.
3. Objective: The primary goal is to estimate the true population parameters (such as the average visit duration) using the sample data.

**Example: MoonDoe Café Sampling**

* Sample of 50 Customer Visits:
  + In this example, the sample mean x̄ is 24.6 minutes, and the sample standard deviation s is 10.7 minutes.
* Challenges:
  + The results will vary if we take another sample of 50 customer visits. The sample mean may be slightly different. For instance, a second sample might yield a mean of 23.5 minutes with a standard deviation of 8.9 minutes. This variability leads us to consider confidence intervals.

**Confidence Intervals (CI)**

Confidence intervals provide a range of plausible values for the true population mean based on the sample data.

1. How to Construct a Confidence Interval:
   * We use the sample mean and sample standard deviation to estimate the population parameters. For example, for a 90% confidence interval, the formula is:

CI = x̄ ± z

Where:

* x̄ = sample mean (e.g., 24.6 minutes)
* z = z-score corresponding to the confidence level (1.64 for 90%)
* s = sample standard deviation (e.g., 10.7)
* n = sample size (e.g., 50)

1. Example Calculation:

* For MoonDoe Café, with a sample size of n = 50, sample mean x̄= 24.6 minutes, and sample standard deviation s = 10.7, the 90% confidence interval would be:

CI = 24.6 ± 1.64

Simplifying:

CI= 24.6 ± 2.48 or 22.1, 27.1

This means that we are 90% confident that the true average duration of customer visits lies between 22.1 and 27.1 minutes.

**Key Takeaways:**

1. Random Sampling: This is essential for ensuring that the sample is representative of the population.

* Random Sample ensures that each customer has an equal chance of being selected, which helps avoid biases in the data.

1. Variability in Samples: Every sample will produce a slightly different mean and standard deviation, which is why we need confidence intervals to express the uncertainty around our estimates.
2. Confidence Intervals: The range provided by a confidence interval gives us an estimate of where the true population parameter likely falls.

* Larger sample sizes yield narrower confidence intervals, which means more precise estimates.
* Higher confidence levels (e.g., 95%, 99%) result in wider intervals, as we need to be more inclusive to achieve greater confidence.

1. Impact of Sample Size: The larger the sample size, the smaller the margin of error. For MoonDoe Café, sampling 50 customers gave a confidence interval of [22.1, 27.1]. If the sample size increased to 500, this interval would shrink, making the estimate more precise.

By understanding the relationships between sample size, variability, and confidence intervals, you can make better inferences from your data. Confidence intervals allow you to express uncertainty in a meaningful way, helping to make more informed decisions based on sample data from a large population like MoonDoe Café’s customers.

## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. You run a small coffee shop. You want to take a sample of your customers, but you have to explain it to your employees. In simple terms, how would you describe a sample of your customers?

Ans:

A set of customers randomly chosen from your customer population

Explanation:

A sample from a population is a randomly chosen subset of that population.

# ****MODEL 3 CONFIDENCE INTERVAL I****

**Confidence Intervals and Their Role in Statistical Analysis**

When dealing with data, such as the MoonDoe Café example, the goal is often to make inferences about the population based on a sample. Confidence intervals are key in this process as they provide a range where we expect the true population parameter (like the mean) to fall. Let’s explore the concept using both the MoonDoe Café case study and general principles of confidence intervals.

**Key Concepts:**

1. Confidence Interval (CI)

* A confidence interval gives us a range of plausible values for the population parameter (mean or proportion). For instance, a 90% confidence interval suggests that if we repeated the same sampling process multiple times, 90% of the calculated intervals would contain the true population mean.
* The formula for the CI for the mean is:

CI = x̄ ± z

Where:

* + x̄ = sample mean
  + z = z-score corresponding to the confidence level (1.64 for 90%)
  + σ = population standard deviation (or sample standard deviation \(s\) if \(\sigma\) is unknown)
  + n = sample size

1. Influencing Factors on Confidence Interval

* Sample Size (n): As the sample size increases, the margin of error decreases, which tightens the confidence interval, making our estimate more precise.
* Standard Deviation (σ or s): A larger standard deviation increases the margin of error, making the confidence interval wider.
* Confidence Level (e.g., 90%, 95%): Higher confidence levels (like 95% or 99%) yield wider intervals because you need to account for more variability to be more confident that the interval includes the true mean.

**MoonDoe Café Case Study:**

MoonDoe Café serves over a million customers daily, and you are tasked with estimating the expected duration of customer visits and providing confidence intervals.

1. 90% Confidence Interval Calculation (Page 24 of the PDF):

* Sample size: n = 50
* Sample mean: x̄= 24.6 minutes
* Sample standard deviation: s = 10.7
* Confidence value for 90%: z = 1.64

Plugging these into the CI formula:

CI = 24.6 x̄ 1.64

This simplifies to:

CI = 24.6 x̄ 2.48 or 22.1, 27.1

Thus, you are 90% confident that the true population mean for the duration of customer visits falls between 22.1 and 27.1 minutes.

**Step-by-Step Understanding of Confidence Interval Calculations:**

1. Starting with the Sample:

* You have a sample, for instance, the 50 customer visits at MoonDoe Café.
* The sample mean \( \bar{x} \) is the average of these observations, and \( s \) is the standard deviation, measuring how spread out the data is.

1. The Role of the Confidence Level:

* For a 90% confidence interval, you're choosing a z-score that corresponds to 90% coverage under the normal distribution, which is 1.64. This reflects that you want to capture the middle 90% of the data.

1. Impact of Sample Size:

* As n (sample size) increases, the term shrinks, meaning your confidence interval narrows. With more data, you're more certain of your estimate.

1. Understanding Random Variables:

* The sample mean x̄ is a random variable because it changes with each new sample you take. However, the population mean μ is not random—it’s a fixed number you are estimating.

**Key Insights for Effective Use of Confidence Intervals:**

1. Interpretation: A 90% confidence interval means that if we repeated this sampling method many times, 90% of those intervals would contain the true population mean.
2. Sample Size Tradeoff: Increasing sample size reduces uncertainty and narrows the confidence interval, making your estimate more precise.
3. Standard Error: The standard error of the mean (calculated as \( \frac{\sigma}{\sqrt{n}} \)) is essential for determining how much the sample mean might differ from the population mean.

**Application Beyond MoonDoe Café:**

Confidence intervals are widely used, from predicting sales revenue to estimating population parameters in surveys. They help quantify the uncertainty around estimates and give decision-makers a clear range to consider for planning and evaluation.

**Summary:**

Confidence intervals provide a structured way to express uncertainty in estimates. In the MoonDoe Café case, with a sample of 50 visits, you determined that the 90% confidence interval for the mean customer visit duration is between 22.1 and 27.1 minutes. The key factors influencing this interval were the sample size, standard deviation, and the selected confidence level. By understanding and applying these concepts, you can make informed inferences from sample data to the population level.

## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. When calculating the confidence interval of a mean, what does σx̄ equal?

Ref:



Ans:

Explanation:

When calculating a confidence interval, σx̄ =

# ****MODEL 3 Finding Confidence Intervals for MoonDoe I****

## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. If a warehouse has taken a sample and says that their average order service time is between 2.78 hours and 3.92 hours with a confidence level of 95%, what does this mean?

Ans:

* If the warehouse repeats their sampling and confidence procedure n times, they would expect 95% of these samples to capture the true average.
* If the warehouse plans to take a similar sample, there is a 95% probability that the resulting confidence interval will capture the true mean.

Explanation:

If we repeat a sampling procedure n times and create a 95% CI, our (random) intervals will capture the true population mean, on average, .95n times out of n.

If the warehouse increases their sample size, the resulting confidence interval would likely shrink since more data would allow for a better estimate.

Note that the true population mean is not a random variable. It is what it is, and there is no probability that it will change. Also note that once a sample has been taken and a confidence interval has been calculated, there is no probability that the confidence interval will change for that sample. Since, there is no way for the confidence interval to change and no way for the true mean to change, then there is no longer a probability that a computed CI captures the true mean. The confidence interval either does capture the true mean or it doesn't.

Before taking a sample, there is still uncertainty and randomness in the values to be sampled. because of this, there is a 95% probability that the CI from the sample will cover the true mean. Once again, once the sample has been taken, the CI will not change and there is no longer a probability.

# ****MODEL 3 Confidence Intervals II smaller sample size****

**Confidence Intervals with Small Samples: MoonDoe Café Case Study**

**Context of Small Sample Sizes in Sampling**

In many real-world situations, collecting large sample sizes may not be feasible due to limitations like time or budget. In such cases, smaller samples are used, but these require a slightly different approach for confidence intervals.

* **Example**: At MoonDoe Café, instead of 50 customer visits, a smaller sample of only 15 visits was collected.
  + Sample mean x̄ = 19.6 minutes
  + Sample standard deviation s = 10.5 minutes

For small samples (less than 30 observations), we use the Student t-distribution rather than the normal distribution to account for greater variability and uncertainty.

**Why Use the Student t-Distribution?**

The Student t-distribution is similar to the normal distribution but has “fatter tails,” meaning it accounts for more data variability. This distribution is used with small samples because it provides a more accurate range for the confidence interval, adjusting for the uncertainty due to a smaller sample.

1. Degrees of Freedom (df): Defined as n - 1, where n is the sample size. Here, n = 15, so df = 14.
2. Critical Value (t): For a 90% confidence level, the t-distribution provides a critical value based on df = 14 and a two-tailed test (since the interval includes both ends of the distribution).

**Calculating the Confidence Interval**

1. Formula for Confidence Interval:

CI = x̄ ± t

Where:

* x̄ = 19.6
* t = t-score corresponding to 90% confidence for df = 14 (approximately 1.76)
* s = 10.5
* n = 15

1. Applying the Formula:

CI = 19.6 ± 1.76

This simplifies to:

CI = 19.6 ± 4.77

Therefore, the 90% confidence interval is:

[14.83, 24.37]

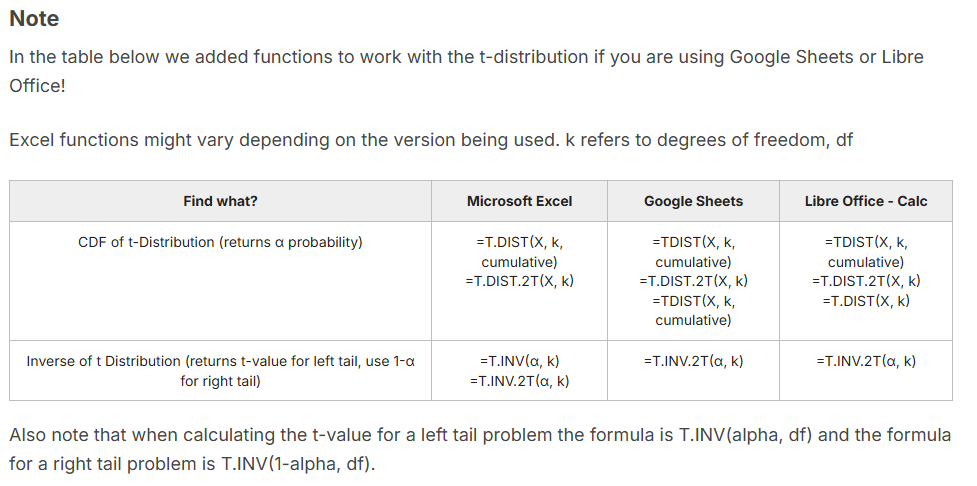
This interval means we are 90% confident that the true mean duration of customer visits lies between 14.8 and 24.4 minutes.

**Key Insights and Comparisons**

1. Wider Intervals with Smaller Samples:
   * This confidence interval is wider than what we observed with larger samples (e.g., when n = 50. Smaller samples require a larger range to maintain the same confidence level.
2. Increasing Confidence Levels:
   * If we wanted a 95% confidence interval, we’d use a larger critical t-value, widening the interval further.
3. Comparison to Large Sample Confidence Intervals:
   * With large samples n ​≥ 30, we use the normal distribution. For small samples, the t-distribution provides a more reliable interval by adjusting for greater variability.

**Conclusion**

When working with small samples, the Student t-distribution is essential for constructing accurate confidence intervals. This approach accounts for the increased variability and provides a reliable estimate for the population mean, even with fewer data points. This method ensures that decision-making is based on realistic confidence levels, especially in cases like MoonDoe Café, where only limited sample data may be available.



## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. Select all true answers relating to the t-distribution and the normal distribution?

Ans:

* The t-distribution has fatter tails than the normal distribution
* As n approaches infinity, the t-distribution approaches the normal distribution
* The t-distribution should be used if n<30

Explanation:

All of the aforementioned answers are true except that the normal distribution should not be used if n<25. The normal distribution should only be used if n>30.

# ****MODEL 3 Confidence Intervals III****

1. Prediction intervals are always going to be wider because confidence intervals are for the sample means and the prediction intervals are for individual observations.

## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. Which of the following statements are correct regarding the tradeoffs between interval, sample size, and confidence level?

Ans:

* If n is fixed, using a higher confidence value c leads to a wider interval L.
* If both n and c are fixed, reducing variability s leads to smaller interval L.

Explanation:

If c is fixed, increasing sample size n leads to smaller interval L.

### PRACTICE PROBLEMS

**PP1 - Prediction Intervals**

Part 1

Food Chemicals Inc. (FCI) produces industrial chemicals used in many food products. Some of their products are also sold to fuel companies to prevent oxidation in fuel. One of their primary products is Butylated hydroxytoluene, simply known as BHT. BHT is produced through the reaction of p-cresol and isobutylene when catalyzed by sulfuric acid.

Every shipment of p-cresol that arrives gets tested and receives a specialized rating from FCI. Due to some of FCI's specialized products, they had to create this rating system themselves.

Shipments of p-cresol from CHEMI-Cal, one of FCI's larger suppliers, receive ratings that fall in a continuous uniform distribution ranging from 7 to 15. Given this distribution, what is the probability that a delivery from CHEMI-Cal receives:

Q1. A rating of above 12.7?

Ans:

0.288

Explanation

This can be calculated in a few ways. An easy way to calculate this is to take the probability of being less than 12.7 and subtract it from 1. In this way, you can capture the probability of being greater than 12.7. As given in the lesson, the probability of being less than 12.7 is . Subtracting this from 1 should yeild an answer of .288

Q2. A rating between 8.2 and 9.5?

Ans:

0.163

Explanation:

This can be calculated in a few ways. An easy way to calculate this is to take the probability of being less than 8.2 and subtract it from the probability of being less than 9.5. In this way, you can capture the probability of being between 8.2 and 9.5. As given in the lesson, the probability of being less than 8.2 is and the probability of being less than 9.5 is . Subtracting the first from the second should yeild an answer of .163

Part 2

Martinez Chemicals in Mexico also supplies FCI with p-cresol. Shipments from Martinez Chemicals always receive ratings above 6 and below 15. Most of the time, Martinez Chemicals shipments of p-cresol receive a rating of 11. Using the triangle distribution, what is the probability that a shipment from Martinez Chemicals:

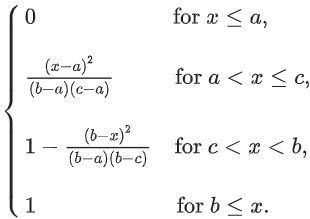
Q1. receives a rating above 9?

Ans:

0.8

Explanation:

Remember that the CDF of a triangular distribution is given as:

To calculate the the probability in the problem, you simply take one minus the CDF that p-cresol ratings are below 9. In this case, the answer that you should get is .800

Q2. receives a rating of above 8 and below 14?

Ans:

0.883

Explanation:

To calculate the the probability of being between 8 and 14, you simply need to subract the CDF of being less than 8 from the CDF of being less than 14. In this way, you can capture the probability of being between 8 and 14. Doing this should yeild an answer of .883

Part 3 A

Gaisburg International in Germany is another supplier of p-cresol to FCI. Shipments of p-cresol from Gaisburg international receive an average rating of 13 with a standard deviation of 2.3. Using the normal distribution, what is the probability that a shipment from Gaisburg Chemicals:

Q1. has a rating of 12 or higher?

Ans:

0.668

Explanation:

The probability that a p-cresol rating is above 12 is simply one minus the probability that the p-cresol rating is less than 12. This can be calculated by using the CDF of a normal distribution with x = 12, μ = 13, σ = 2.3. Your answer should be .668

Q2. Has a rating under 15 but above 9.8?

Ans:

0.726

Explanation:

The probability that a p-cresol rating is under 15 and above 9.8 is simply the probability that the p-cresol rating is less than 15 minus the probability that the p-cresol rating is under 9.8. In this way, the probability of being between the two ratings can be captured. This can be calculated by using the CDF of a normal distribution with x = 15, μ = 13, σ = 2.3 and with x = 9.8, μ = 13, σ = 2.3 . Your answer should be .726

Part 3 B

Create a 95% prediction interval for p-cresol ratings from Gaisburg International centered on the mean.

Q1. What is the upper bound of this interval?

Ans:

17.51

Q2. What is the lower bound of this interval?

Ans:

8.49

Explanation:

Creating a 95% prediction interval for p-cresol ratings from Gaisburg International can be very easy. You simply need to find where the data is centered on the mean and achieves a 95% total probability. In this case it is when the probabilities equal .025 and .975 for a CDF. The total probability (area) between these two points is equal to 95% and is centered on the mean. Using an inverse CDF you can easily calculate this with probability = .975, μ = 13, σ = 2.3 and probability = .025, μ = 13, σ = 2.3. The two points that come from this should be 17.51 and 8.49

# ****MODEL 3 Hypothesis Testing Basics****

mutually exclusive adalah kita pilih yang ini atau yang itu, tidak boleh keduanya

collectively exhaustive adalah pilihan ini sudah menyangkup semua pilihan

1. Mutually Exclusive (Saling Eksklusif):
   * Kejadian-kejadian tidak bisa terjadi bersamaan.
   * Jika Anda berada di salah satu kejadian, Anda tidak mungkin berada di kejadian lain.
   * Contoh: Dalam pelemparan dadu, hasil 1 dan 2 adalah mutually exclusive—jika hasilnya 1, maka tidak mungkin hasilnya juga
2. Collectively Exhaustive (Secara Kolektif Menyeluruh):
   * Kejadian-kejadian mencakup semua kemungkinan yang ada, sehingga tidak ada opsi lain di luar kejadian-kejadian tersebut.
   * Salah satu dari kejadian tersebut pasti akan terjadi.
   * Contoh: Dalam pelemparan koin, "kepala" dan "ekor" adalah collectively exhaustive—hasilnya pasti kepala atau ekor, tidak ada hasil lain.

## QUESTIONS, ANSWERS & EXPLANATIONS

Q1. Which of the following are correct regarding the Hypothesis Testing?

Ans:

* If measuring “is greater than” or “is less than”, then use one tailed.
* If measuring “any difference” or “is not the same as”, then two tailed.

Explanation:

If measuring “is greater than” or “is less than”, then use one tailed and if measuring “any difference” or “is not the same as”, then two tailed.

# ****MODEL 3 Hypothesis Testing Basics****

Sebelumnya mean nya 23 dengan jumlah sampel sekian. Sekarang jumlah sampelnya 100, meannya sekian, std nya sekian. Apakah mereka ada perbedaan dari lama durasinya

# ****MODEL 3 Introduction of p‐value****

**Intinya dengan kita mencari p value, kita gak perlu setting alpha dengan nilai menebak nebak, seperti 1 atau 5%**

**In pages 44–45 of the PDF, the discussion centers on using the p-value to determine statistical significance in hypothesis testing. Here’s how this fits into the context of evaluating changes in visit duration at MoonDoe Café.**

****Overview of p-Values and Application****

**A p-value represents the probability of obtaining a test result at least as extreme as the one observed, assuming the null hypothesis is true. It helps assess the exact level of significance needed to reject the null hypothesis without having to choose a specific threshold, such as 1% or 5%, from the outset.**

**For example, in MoonDoe’s case, when testing if a new store layout affects visit duration, we found that for a 99% confidence level (α = 1%), the population mean of 23 minutes fell within the confidence interval, so we couldn't reject the null hypothesis. However, at a 95% confidence level (α = 5%), the population mean of 23 was outside the interval, leading to rejection of the null hypothesis, suggesting some difference in visit duration at this lower confidence level**

****Practical Use of p-Value****

**The p-value indicates at what significance level we are exactly able to reject the null hypothesis:**

**1. Calculate the p-value directly: It shows the smallest significance level at which the observed data would allow us to reject the null hypothesis.**

**2. Compare to significance thresholds: If the p-value is below a predetermined level (e.g., α = 0.05), we reject the null hypothesis. For instance, if MoonDoe’s test yields a p-value of 0.026 (2.6%), it indicates a statistically significant result at the 5% level but not at the 1% level**

**This approach is valuable in determining the actual level of significance and is especially useful for interpreting results across various statistical tests, including regression analysis and other inference methods.**

## ****QUESTIONS, ANSWERS & EXPLANATIONS****

Q1. Which of the following is the probability, under the assumption of hypothesis, of obtaining a result equal to or more extreme than what was actually observed?

Ans:

p-value

Explanation:

p-value is the probability, under the assumption of hypothesis, of obtaining a result equal to or more extreme than what was actually observed.

Penjelasan:

**Memahami Hipotesis Nol dan P-value**

1. Hipotesis Nol (H0):
   * Dalam konteks ini, hipotesis nol menyatakan bahwa obat tidak efektif. Ini adalah pernyataan yang ingin kita uji.
2. P-value:
   * P-value adalah probabilitas mendapatkan hasil yang sama atau lebih ekstrem dari yang kita amati, dengan asumsi bahwa hipotesis nol benar.

**Interpretasi P-value**

* Jika p-value < 0.05 (atau 5%):
  + Ini menunjukkan bahwa ada kurang dari 5% kemungkinan kita akan melihat hasil yang sama atau lebih ekstrem jika hipotesis nol benar. Dalam hal ini, kita cenderung menolak hipotesis nol.
  + Artinya, hasil menunjukkan bahwa ada bukti yang cukup untuk mengatakan bahwa obat itu mungkin efektif.
* Jika p-value ≥ 0.05 (atau 5%):
  + Ini menunjukkan bahwa ada 5% atau lebih kemungkinan kita akan melihat hasil yang sama atau lebih ekstrem jika hipotesis nol benar. Dalam hal ini, kita tidak menolak hipotesis nol.
  + Ini berarti kita tidak memiliki cukup bukti untuk menyatakan bahwa obat itu efektif. Namun, ini tidak membuktikan bahwa hipotesis nol benar; kita hanya tidak memiliki cukup bukti untuk menolaknya.

Contoh Kasus

* P-value = 4% (0.04):
  + Ini berarti kita memiliki bukti yang cukup untuk menolak hipotesis nol. Kita dapat menyimpulkan bahwa obat itu kemungkinan besar efektif.
* P-value = 6% (0.06):
  + Ini menunjukkan bahwa kita tidak memiliki cukup bukti untuk menolak hipotesis nol. Kita tidak dapat menyimpulkan bahwa obat itu efektif.

Kesimpulan

* P-value bukanlah bukti absolut bahwa hipotesis nol benar atau salah. P-value hanya memberi tahu kita seberapa kuat bukti yang ada terhadap hipotesis nol. Dengan kata lain, hipotesis nol tidak dapat dibuktikan, tetapi bisa ditolak atau tidak ditolak berdasarkan nilai p yang diperoleh.

# ****MODEL 3 Hypothesis Testing – More Formally****

## ****QUESTIONS, ANSWERS & EXPLANATIONS****

**Q1.If you reject a null hypothesis that was actually correct, what type of error is this?**

**Ans:**

**Type I**

**Q2. If you fail to reject a null hypothesis that is incorrect, what type of error is this?**

**Ans:**

**Type II**

# ****MODEL 3 Hypothesis Testing – More Formally****

# ****MODEL 3 Hypothesis Testing Examples****

## ****QUESTIONS, ANSWERS & EXPLANATIONS****

Conduct a t-test to see if the mean of this data is above 27. Use .

31

24

27

27

31

26

30

31

Q1. What is the p-value of your t-test?

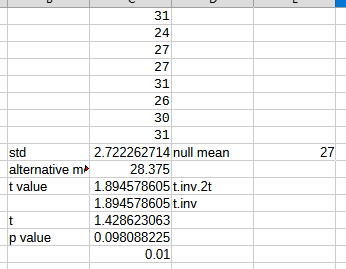
Ans:

0.098

Explanation:

Remember that this is a one tailed test. You are testing to see if your data is above the mean listed. In spreadsheets, the p value is calculated as 1-t.dist(t-value,n-1,1) where the test-statistic is (sample mean - test mean)/(sample stdev/sqrt(n)).

caranya:

std =stdev.s(c1:c8)

alternative mean = average(c1:c8)

t value (critical value) = t.inv.2t(2\*0.05,7)

t value (cara kedua) = t.inv(1-0.05,7)

t (t-value sbnrnya) =(alternative mean – null mean)/(std/sqrt(jlh sample)

jlh sample = 8

p-value = 1-t.dist(t,7,1)

correction:

* yang benar itu yang didalam kurung, kayak critical value, t-value

Q2. Can you say that the mean of the data is greater than 27?

Ans:

No

Explanation:

Since your a(alpha) > .05 you can not reject the null hypothesis. You are unable to say with confidence that the mean of the data is greater than 27.

* correction:
  + Since your p value > .05 you can not reject the null hypothesis. You are unable to say with confidence that the mean of the data is greater than 27.

# ****MODEL 3 Chi-Square Tests****

Paling cepat menemukan p value, pakai rumus =CHITEST([rentang]observed value;[rentang]expected value) .untuk ngebandingin apakah mau di reject atau tidak null hypothesis nya, bandingkan nilai alpha menggunakan degree of freedom dengan chi square distribution (x^2)

**CATATAN**:

kita bisa bilang “ I'm 99% confident that they are not equally favored.”

## ****QUESTIONS, ANSWERS & EXPLANATIONS****

You repeat the MoonDoe survey on 100 new customers. Your new data is below.

Testvalue Expected Value

Blue 32 25

Red 27 25

Green 18 25

Yellow 23 25

Q1. Run the chi-square test. What is the p-value of your ChiSq Test?

Ans:

0.237

Explaination:

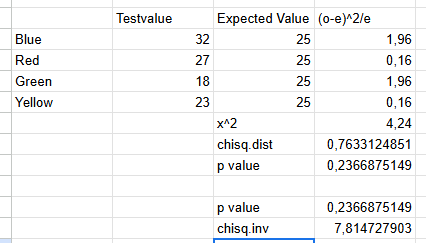
In your favorite spreadsheet program, simply use the Chi-Square function.

In excel: =CHISQ.TEST

In Google Sheets: =CHITEST

In Libre Office: =CHISQ.TEST

Then select the appropriate data.

Caranya  


X^2 itu tinggal dijumlahin aja lagi

Chisq.dist (rumusnya: =chisq.dist(x^2;3(degree of freedom);1) itu cara untuk menemukan p value, karena

P value = 1 – chisq.dist

Cara yang bawah lebih ringkas

P value =CHITEST(observed\_range, expected\_range)

Chisq.inv (critical value) =CHISQ.INV(probability(ex:0.99,0.95);degree of freedom(ex disini: 3))

Q2. Using a = 0.05 can you say that your customers are significantly different than the expected distribution?

Ans:

No

Explanation:

Since your p value > .05 you are unable to reject the null hypothesis. You are unable to say that your customers are different from the expected distribution with confidence.

# ****MODEL 3 Chi-Square Test Example II****

Liat di excel “M3U3\_Chi\_Square\_Test”, cara untuk mendapatkan nilai expected value dari norm.dist. Ada buckets 4 dan 5. itu artinya kayak perbedaannya 4/5 jaraknya. nilai range nya dari 10, kemudian 15 (untuk bucket 5)

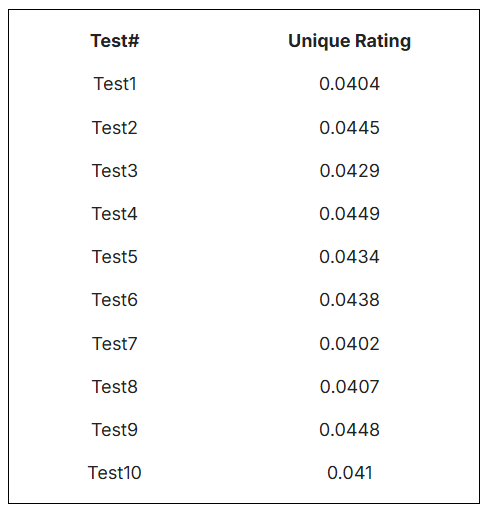
### ****PRACTICE PROBLEMS****

****PP1 - Crimson Laboratory****

Part 1

Crimson Laboratory is manufacturing a new, durable metal for customers to use in various projects. However, because molecular deformations in the metal are not uniform throughout the sheet prior to cutting, each final sheet of metal has a slightly different yield strength.

Crimson Laboratory’s QA (Quality Assurance) department has randomly selected ten different samples of its newest metal to undergo a prolonged-stress test. Crimson uses a unique rating system that they developed to test their metals. The results of this test for each of the samples are as follows:



You have been hired to review this data. First, you are tasked with creating a 95% confidence interval for the mean of the data set using a t-distribution.

Q1. What is the upper limit of this confidence interval?

Ans:

0.044

Q2. What is the lower limit of this confidence interval?

Ans:

0.0413

Explanation

A confidence interval for the mean of a sample using the t distribution is given as CI = x̄ ± t

Where:

x̄ is the mean of the sample = 0.04266

c is the t value that covers 95% of the two tailed t-distribution (or 97.5% of the one tailed t-distribution) with n-1 df

S is the Sample Standard Deviation = 0.0019

n is the sample size = 10

caranya:

yang t value =t.inv.2t(0.05,9), cara satu lagi =t.inv(0.025,9)

Part 2

Customers are asking for metals that test above a unique rating of 0.043. Test if your new metal has a mean greater than 0.043.

H0 = Your metal has a mean unique rating that is less than or equal to 0.043.

H1 = Your metal has a mean unique rating that is greater than 0.043.

Q1. Using a t-test, what is the p-value of your sample?

Ans:

0.7069

Explanation:

To solve this problem, you should conduct a right-tailed t-test. Your test statistic is caculated as

Where:

x̄ is the sample mean

μ is the test mean

S is the sample standard deviation

n is the number of observations

You will get a negative test statistic (-0.565881265503815). Your P value is then found by using a right-tailed t distribution with your test statistic and n-1 df. In excel you would use 1-T.DIST(test statistic,n-1,True) =1-t.dist(-0.565881265503815, 9,1) (because T.Dist(...) returns the left tail). Your P value should be 0.7069. The solution to this problem follows the same recipe as shown in Module 3 Video 3 for the smaller-sample example in the second part of the lecture.

Q2. The QA department requires that you be 95% sure that the metal achieves customer specifications. Given this constraint, would you reject your H0?

Ans:

Fail to Reject H0: You are not 95% confident that customer specifications will be met with this metal.

Explanation:

If your p value is less than .05, you should reject your H0. Otherwise you should fail to reject H0.

**PP2 - Lily Ambulance Services**

Lily ambulance service, which is based in Boston, attends to emergency medical calls in the city. They have a 24-hour call center which receives all calls, takes notes about location and patient’s status and assigns the nearest ambulance vehicle to the call. The ambulance vehicle gets a call notifying them of the same. They dispatch to the location after receiving the call and help the patient. The wait time, which is the average time from when the call is received to when the ambulance arrives at the scene, is 17.5 minutes.

Lily Ambulance Services realized that technology intervention may reduce their operating cost and hence introduced an app where people can request for an ambulance using an app (by putting in some of the important information like location and patient condition). Ambulance vehicles directly receive a notification, reducing the need for a call center. Lily Ambulance Services decided to keep the call center while testing the app functionality.

You have been hired to evaluate if the app had any impact on service.

By studying a sample of the data with 150 app requested ambulances, it was determined that the wait times have a mean of 16 minutes with a standard deviation of 6 minutes.

You decide to conduct a hypothesis test to evaluate if the wait times using the app based requests are equal to that of the regular call based requests. You assume a significance level of 5%:

H0 : Average wait times in app requests are equal to call based requests.

H1 : Average wait times in app requests are not equal to call based requests.

Part 1

Q1. Determine the type of hypothesis test:

Ans:

Two-Tailed

Explanation:

Lily Ambulance wants to know if there is any significant change to service. Since the hypothesis test checks for similarity/differences (not if greater than or less than), it is a two tailed test.

Part 2

Q1. What is the corresponding critical value (the z-value) of the app based requests?

Ans:

1.96

Explanation:

Since n >30 we are using the normal distribution. Remember that in a two-tailed test we split the confidence level to cover both tails. To find the critical value can use the formula in excel =NORM.S.INV(1-(0.05/2)).

Part 3

Q1. What is the upper limit of this confidence interval?

Ans:

16.96

Q2. What is the lower limit of this confidence interval?

Ans:

15.04

Explanation:

A confidence interval for the mean of a sample using the normal distribution is given as CI = x̄ ± z

Where:

x̄ is the mean of the sample = 16

c is the z value that covers 95% of the two tailed distribution (or 97.5% of the one tailed distribution) = 1.96

S is the Sample Standard Deviation = 6

n is the sample size = 150

Part 4

Q1. Compare the test statistic to the critical value and determine the hypothesis test result.

Ans:

Reject H0: You are 95% confident that average wait times in app requests are not similar to call based requests.

Explanation:

You have already calculated the confidence interval for the mean at a 95% confidence level in part 3. Your H0 states that population mean and sample mean are equal. To see if you can reject this null hypothesis all is left to do is check, if the population mean is within this confidence interval of the sample mean. If the population mean is within the confidence interval you cannot reject H0. In this case, because the population mean (17.5) is outside the confidence interval you reject the null hypothesis. (See Video 5)

**PP3 - Chi-Square Test**

Chi-Square Test

You sell 5 types of TVs at your electronics store. Your store is a small part of a larger chain. Over the past few months, corporate has developed a distribution of customer preferences for these TV types. You decided to interview 185 local customers because you don't think that corporate expectations will hold at your particular store. Conduct a Chi-Square test to see if your customers are in a different distribution than the one corporate expects them to be in. Your survey results and expected values are below.

Testvalue Expected Value

Product1 18 28

Product2 39 47

Product3 56 48

Product4 38 35

Product5 34 27

Q1. What is the p-value of your ChiSq Test?

Ans:

0.080

Explanation:

In your favorite spreadsheet program, simply use the Chi-Square function.

In excel: =CHISQ.TEST

In Google Sheets: =CHITEST

In Libre Office: =CHISQ.TEST

Q2. Using a = .05 can you say that your customers are significantly different from corporate's expected distribution?

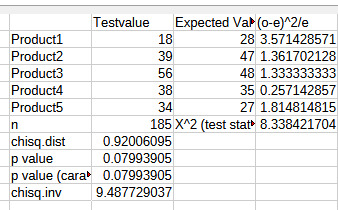
Ans:

No

Explanation:

Since your p value is greater than your alpha value of .05, you can not say that the distributions are significantly different.

Caranya:

jadi intinya,

kalau p value < alpha maka, maka tolak H0

kalau test statistic > critical value, maka tolak H0

ex:

0.07 > 0.05, maka terima H0

8.39 < 9.49, maka terima H0

# ****MODEL 3 Working with Multiple Random Variables****

pakai pivot table ni, data field nya count salah satu dari mereka

## ****QUESTIONS, ANSWERS & EXPLANATIONS****

Create the probability tables of Zippy Bright Sales for every combination of sales between:

XP219 and Whitener & Floss and Whitener

Q1. What is the probability that 4 units of XP219 and 1 unit of Whitener are sold?

Ans:

0.04

Explanation:

In the table attached, filter for instances where XP219 = 4 and then filter again where Whitener is 1 among the 1st filtered set. You should get two instances. This divided by the total week count of 50 will give you 0.04 which is the answer.

This means that 2 out of 50 weeks have XP219 sales of 4 and 1 sale of Whitener.

Q2. What is the probability that 3 units of Floss and 5 units of Whitener are sold?

Ans:

0.16

Explanation:

Follow the same method as above. The probability that 3 units of Floss and 5 unit of Whitener are sold is equal to the total count of this event occurring divided by the total week count. Your answer should be .16