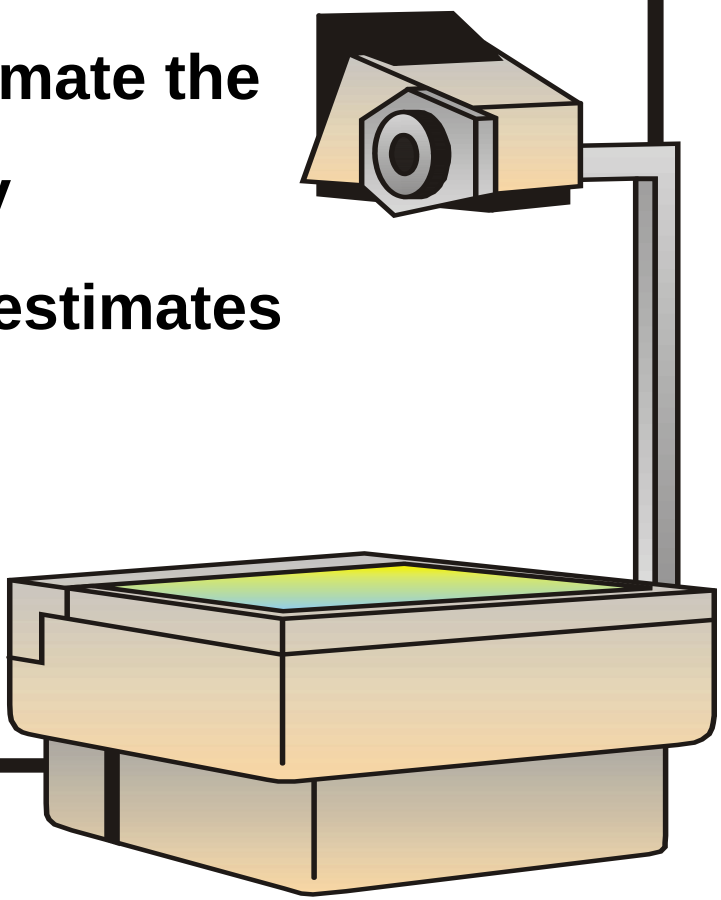
# "Statistical Inference and the Expected Date of Delivery"

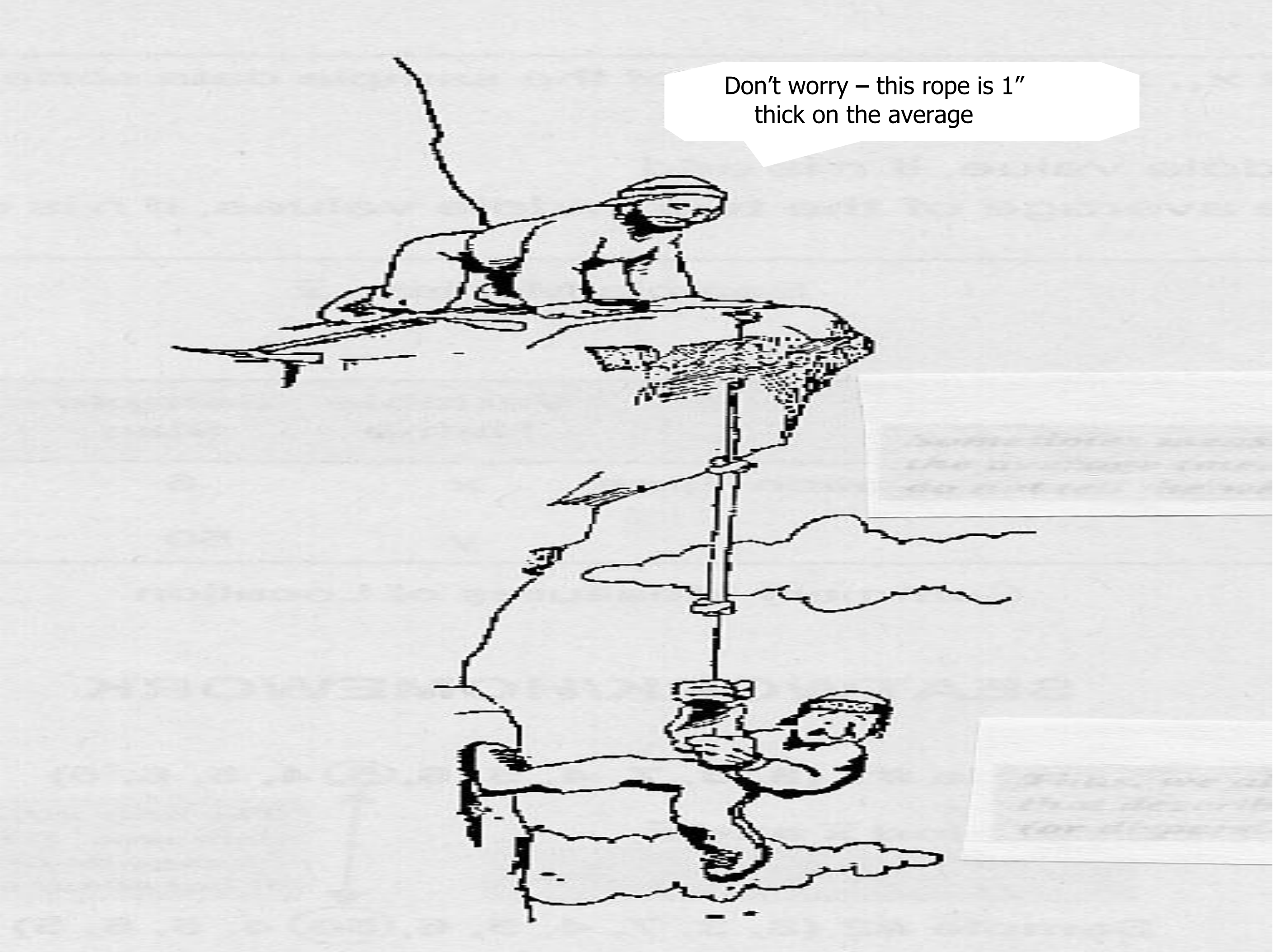
By Dr. Joe DeSimone

# What Joe will cover

- Fundamentals of Statistical Inference
- Historically - How we estimate the Expected Date of Delivery
- How can we make better estimates using statistics



# The Claims we make !



June 2, 1996

**North Little Rock, Arkansas - (AP)**

**When a jealous husband found a man's billfold in his car, he immediately drove to the address shown on an identification card in the wallet.**

**He rang the bell and, when the man answered, gave him a sound beating and a warning. Monday, the husband was fined and given a suspended 30-day jail sentence in Municipal Court . . .**

**. . . The owner of the billfold had moved from the address on his identification card.**

# Assumptions

There is an (ass)umption of certainty underlying this kind of decision-making process.

How easy it is to (ass)ume knowledge that one does not have!

**The demand for certainty is one which is natural to mankind, but is nevertheless an intellectual vice.**

**But so long as men are not trained to** withhold judgment in the absence of evidence, they will be led astray ...

... To endure uncertainty is difficult, but so are most of the other great virtues.

**- Bertrand Russell**

**But when is anything 100% certain?**

# Worried in San Diego

Dear Abby:

You wrote in your column that a woman is pregnant for 266 days. Who said so? I carried my baby for 10 months and 20 days, and there is no doubt about it because I know the exact day my baby was conceived. My husband is in the Navy and it couldn't possibly have been conceived at any other time because I saw him only once for an hour, and I didn't see him again until after the baby was born.

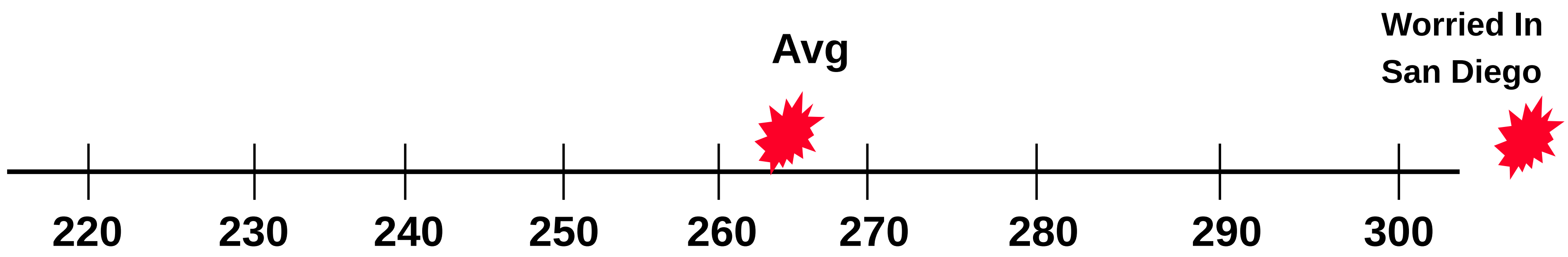
I don't drink or run around, and there is no way this baby isn't his, so please print a retraction about the 266-day carrying time because otherwise I am in a lot of trouble!

- Worried in San Diego

What would you say to her? To her Husband?

# Analyzing The Issue

- Average carrying time is 266 days
- If she reports a 260 day pregnancy, would you be suspicious?
- If she reports a 400 day pregnancy, would you be suspicious?
- At what point would you start to become suspicious? Make a mark



# Some History…….

The Estimated Date of Confinement (EDC), also known as expected date of delivery/estimated due date (EDD) or simply *Due Date*, is a term describing the estimated delivery date for a pregnant woman.

Confinement is a traditional term referring to the period of pregnancy when an upper-class, Nobel or royal woman would withdraw from society in medieval and Tudor times be confined to their rooms with midwives, ladies-in-waiting and female family members only to attend them. Back then, the EDC was often determined based on feelings, beliefs, incomplete data and much hearsay.

Nowadays, important business decisions are often made the same way!

# Rodney says…..

“When I was born…..
the doctor slapped
my mom“

# Modern Estimation

Gestational age:  Directly calculating the days since the beginning of the last menstrual period.

Normally, women are given a date for the likely delivery of their baby that is calculated as 280 days after the onset of their last menstrual period. Yet only four percent of women deliver at 280 days and only 70% deliver within 10 days of their estimated due date, even when the date is calculated with the help of ultrasound.

# Interesting Numbers….

"We found that the average time from ovulation to birth was 280 days" said Dr Anne Marie Jukic, a postdoctoral fellow in the Epidemiology Branch at the National Institute of Environmental Health Sciences (Durham, USA)

"However, even after we had excluded six pre-term births, we found that the length of the pregnancies varied by as much as 24 days."

"We were a bit surprised by this finding. We know that length of gestation varies among women, but some part of that variation has always been attributed to errors in the assignment of gestational age."

# The Assumption of Normality

“It is impossible to tell the difference between errors in calculations and natural variability without being able to measure correctly the gestational age.”

Births rarely occur *on* a due date, but they are clustered *around* due dates.

**Dr Jukic** concluded: "I think the best that can be said is that natural variability may be greater than we have previously thought.”

# Other methods include….

**Parikh's formula** is a calculation method that considers cycle duration by adding 9 months to the start of the last menstrual period then – 21 days…….

**Naegele's rule** assumes an average cycle length of 28 days, which is not true for everyone.

# Pregnancy Humor

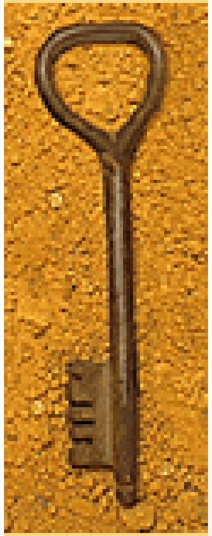
# Confidence Intervals

In estimating the parameter of a population we typically establish some Interval that will include that parameter with some specified degree of Certainty. That internal is called the **Confidence Interval.** The upper And lower limits of the interval are called the **Confidence Limits,** and The degree of certainty is called the **Confidence Level.**

Example: With 95% confidence, the true mean of the population will be included in the interval:

$$.498 \leq \mu \leq .500$$

# Hypothesis Testing



## What's this all about?

- Hypothesis
  - **An educated guess**
  - **A claim or statement about a property of a population**
- The goal in Hypothesis Testing is to analyze a sample in an attempt to distinguish between population characteristics that are likely to occur and population characteristics that are ***unlikely*** to occur.

# Hypothesis Testing

- The process of taking a practical problem and translating it to a statistical problem.

- Because we are using samples (and relatively small ones at that) to estimate population parameters, there is always a chance that we can select a “weird” sample for our experiment that may not represent a “typical” set of observations.

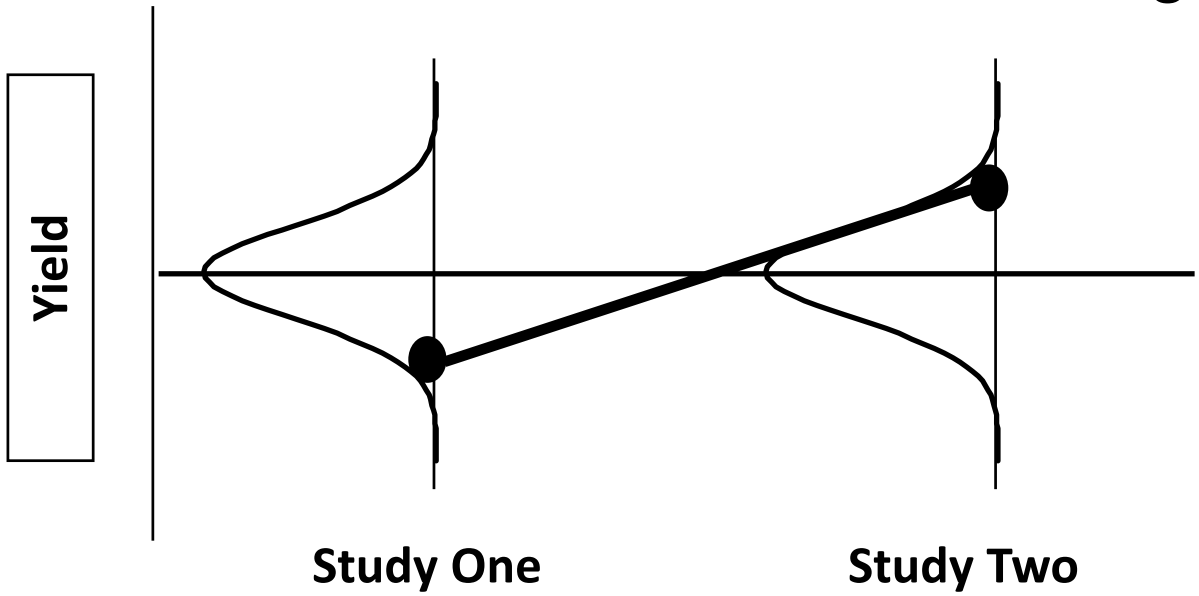
- Because of this, inferential statistics, with some assumptions, allows us to estimate the probability of getting a “weird” result purely due to chance.

- For example, if we wanted to know a coin was “fair”, we could flip it a number of times and track how many heads we saw. By chance we would expect about 50% of the flips to be heads by chance.

- If we flipped the coin 10 times and got 10 heads, we would be fairly confident the coin is not fair. There is one chance out of 1000 that we could have gotten 10 heads with a fair coin. Therefore, we would say we are willing to take a 0.1% chance of being wrong about our “unfair” coin.

# In the Real World

- We can catch a good process on a bad day
- We can catch a bad process on a good day
- In either case, we can make the wrong inference




We say we made an improvement in the process and the results were just a function of sampling.

# Hypothesis Testing

- Let's take a look at a manufacturing example. Suppose we have modified one of two reactors. We want to see if we have "significantly" improved the yield with these modifications before we modify all reactors.

- Let's look at the resulting data. In this case, Reactor B is the newly modified reactor.

| Reactor A | Reactor B |
|---|---|
| 89 | 84 |
| 81 | 86 |
| 84 | 83 |
| 84 | 91 |
| 87 | 86 |
| 79 | 79 |
| 85 | 82 |
| 81 | 89 |
| 83 | 83 |
| 84 | 88 |

# Hypothesis Testing

**Question:** **Will the modifications on Reactor B improve yield when compared to the current process, represented by Reactor A?**

**Descriptive Statistics**

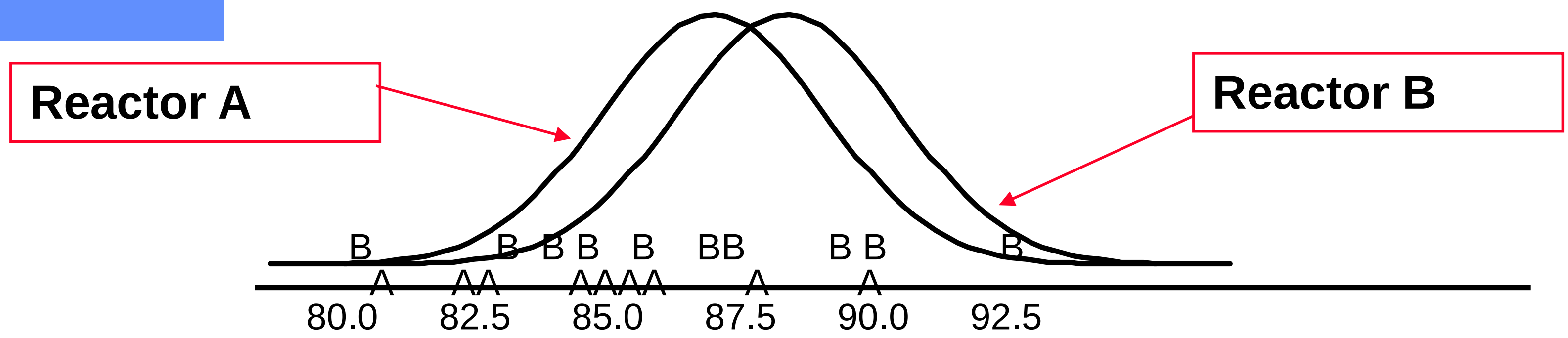
| Variable | N | Mean | Median | StDev |
|---|---|---|---|---|
| ReactorA | 10 | 84.400 | 85.000 | 2.91 |
| ReactorB | 10 | 85.60 | 85.50 | 3.72 |

**The Delta = 1.2%**

**The statistical question is:**

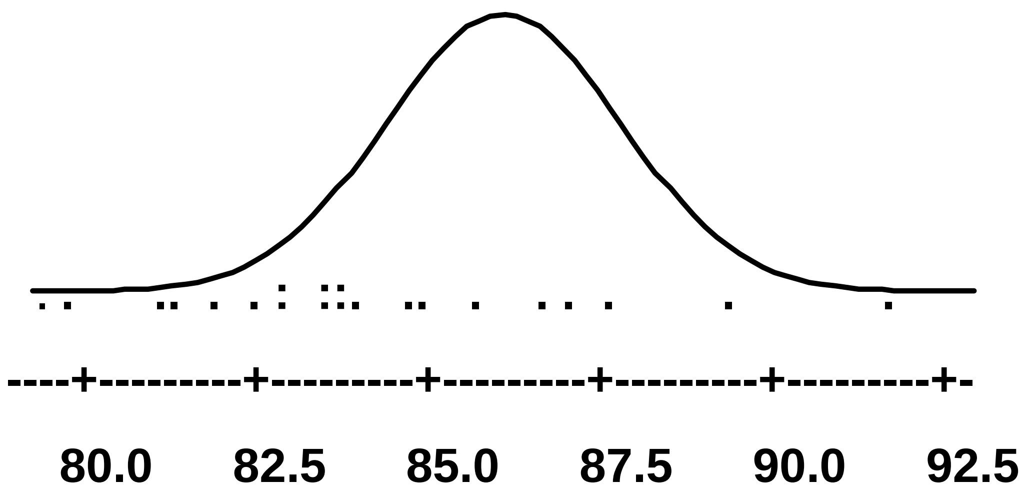
**Is the mean for Reactor B (85.6) different enough from the mean for Reactor A (84.4) to be considered important? Or, are the means close enough to have occurred just by chance and day-to-day variation?**

# Which is it?



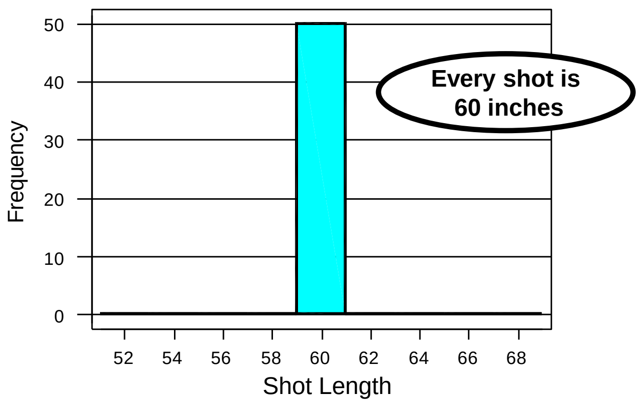

Do the Reactors represent two different Processes?

Do the Reactors represent one basic process?

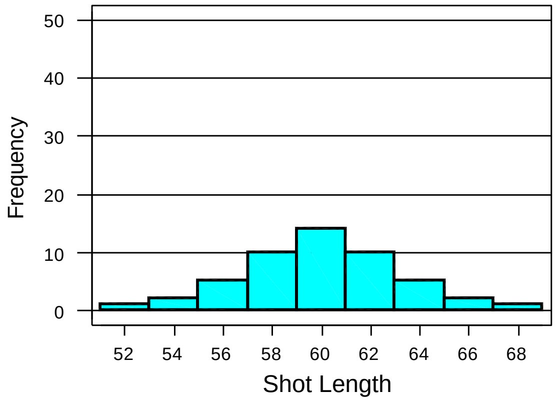


# Populations...

### *Ideal Process Output*

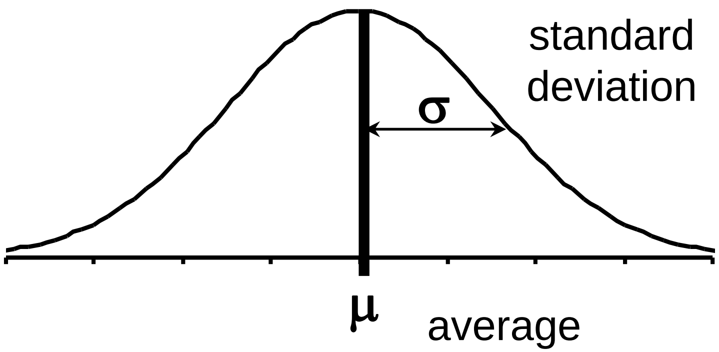



### *Actual Process Output*



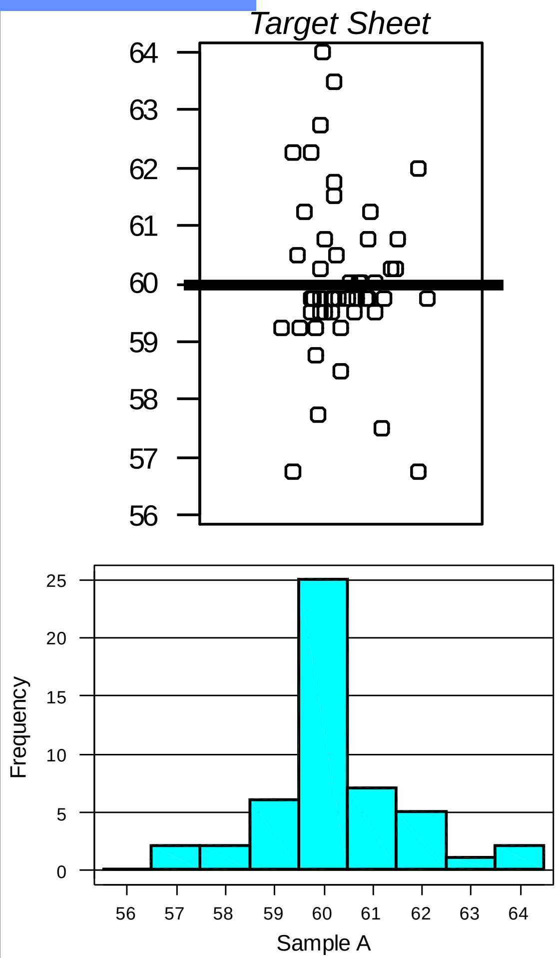

- A **population** is the entire collection of units whose characteristics are of interest.
- A **parameter** describes the "true" value of a characteristic.

### *Population Parameters*



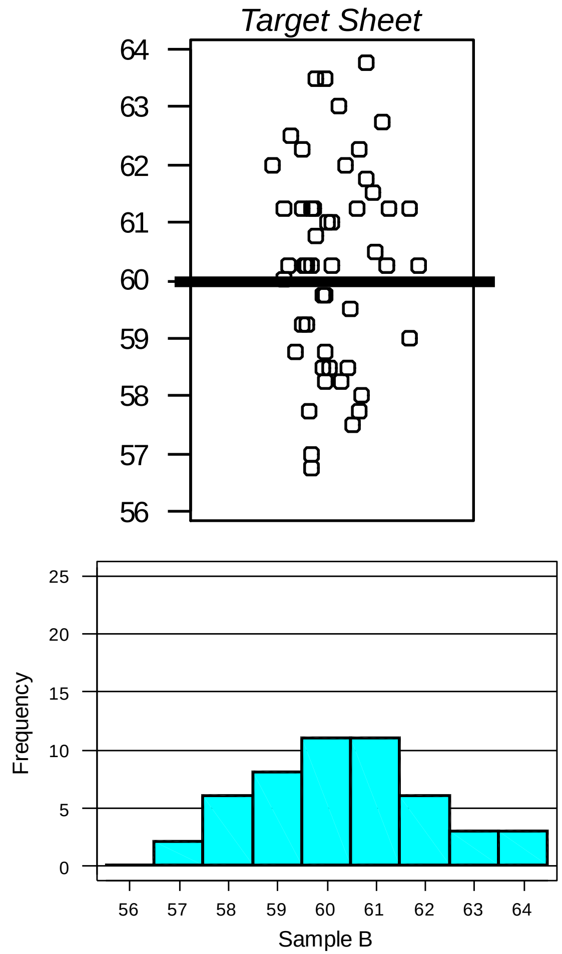

*A parameter's value is fixed but usually unknown*

# Example - Catapult Shot Length Samples



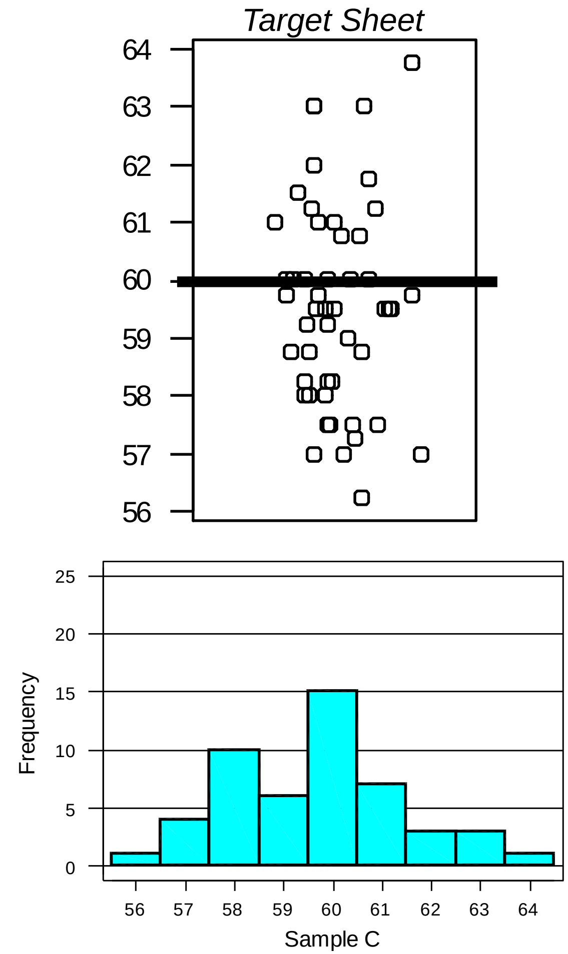

$\bar{X}_A = 60.07$

$s_A = 1.44$




$\bar{X}_B = 60.31$

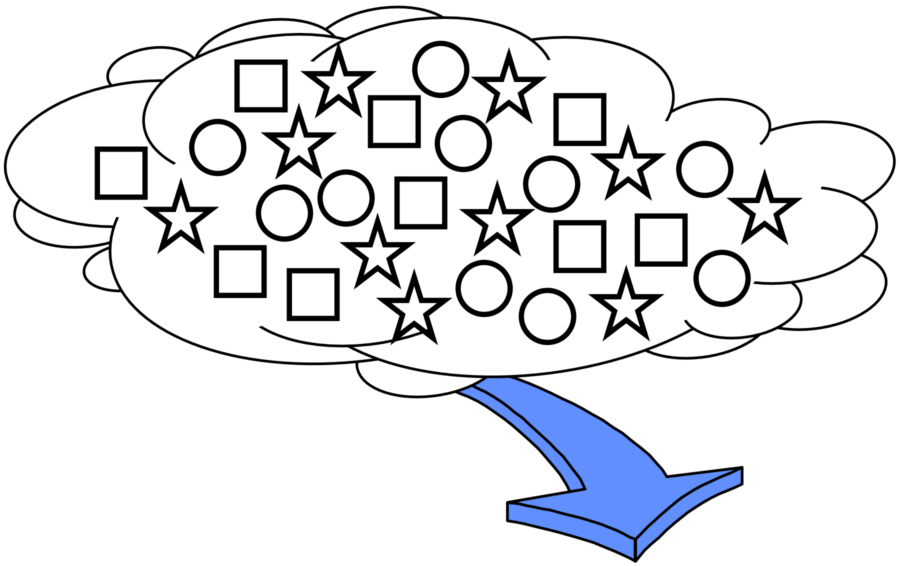
$s_B = 1.77$




$\bar{X}_C = 59.57$

$s_C = 1.76$

# ...and Samples



- A **sample** is a portion or subset of units taken from the population whose characteristics are actually recorded (measured or counted).
- A **statistic**, any number calculated from sample data, describes a sample characteristic.

| | *Sample Statistics* $\bar{X}$ | *Sample Statistics* s |
|---|---|---|
| □ = Observations taken as "sample A" | 60.07 | 1.44 |
| ☆ = Observations taken as "sample B" | 60.31 | 1.77 |
| ○ = Observations taken as "sample C" | 59.57 | 1.76 |

*A statistic's value is known for a specific sample, but usually changes from sample to sample*

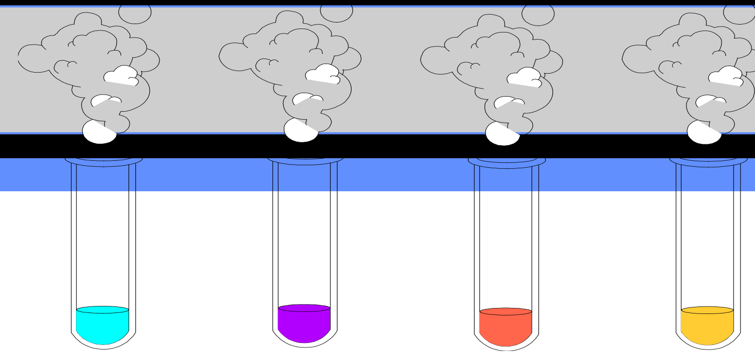
# Key Terms

**Ho = Null Hypothesis**

**Ha = Alternative Hypothesis**

**P Value = Probability Value**

# Hypothesis Testing



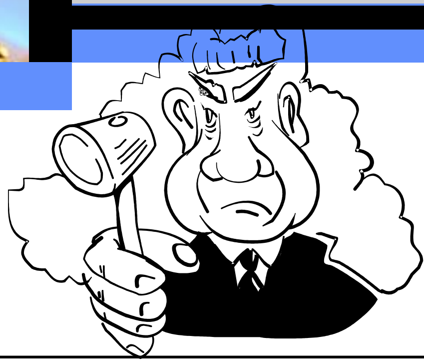
- **Real Life Hypothesis: The newly modified process will reduce defects.**
- **This is called the Alternative Hypothesis (Ha)**

- **Statistical Hypothesis: There is no difference between the old process and the improved one.**
- **This is called the Null Hypothesis (Ho)**

$$\boxed{\begin{array}{l} \text{Ho:}\quad \mu_{\mathrm{a}} = \mu_b \\ \text{Ha:}\quad \mu_{\mathrm{a}} \neq \mu_b \end{array}}$$

**We must show that the values we observed were so unlikely to come from the same process, that Ho must be wrong.**

# About The Null Hypothesis...




- The Null Hypothesis (Ho) is assumed to be true.
  This is like the defendant being presumed to be "Not Guilty".
  - **Remember: The American justice system is NOT "guilty until proven innocent."**
  - **We don't assume that our experiment has an EFFECT until the probability of "no effect" is too small to believe.**
- You are the prosecuting attorney. You must provide evidence beyond a **"reasonable doubt"**

**NOTE:**
**"Not Guilty" ≠ "Innocent"**

# Hypothesis Testing

**State a "Null Hypothesis" ($H_o$)**

**Gather evidence (a sample of reality)**

**<u>DECIDE:</u>**

**What does the evidence suggest?**

**Reject $H_o$? or Not Reject $H_o$?**

**Hypotheses of Means**

$H_o$: $\mu_o = 266$

$H_a$: $\mu_o > 266$

**Hypotheses of Standard Deviations**

$H_o$: $\sigma_A = \sigma_B$

$H_a$: $\sigma_A > \sigma_B$

# STEPS SIMPLIFIED

1. State the Hypothesis ( Ho: $\mu$ = 266, Ha: $\mu \neq$ 266 )
2. Choose a value for the type one error ( $\alpha$ = .05 )
   (Confidence level)
3. Choose the test statistic (See Summary )
4. Determine the Acceptance and Rejection regions and draw a picture. Label the curve with the *critical value* *from the test table*
5. Compute the test statistic and compare to the acceptance region and critical value
6. State the conclusion

# INFERENCE TESTS FOR AVERAGES

## Mean Tests

| Type | Test Statistic | df | Application |
|---|---|---|---|
| **Z** | $Z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ | n/a | Single sample mean.<br>Population sigma is known. |
| **t test** | $t = \dfrac{\bar{x} - \mu}{\dfrac{s}{\sqrt{n}}}$ | n – 1 | Single sample mean.<br>Population sigma unknown. |

Note: t test Is also used for a small sample size ( < 30), when
The distribution of the population is unknown or can't be estimated

# Hypothesis Testing:  How It Works

- **After data is collected, we calculate both:**
  - a Test Statistic (some form of a signal-to-noise ratio[SNR] such as a Z- or T-Score), and
  - a "P-Value".   The P value comes from Z calc:  If Z calc is 2.0:  P = 0.025

- **The "P-Value" is the probability that such results could occur when $H_o$ is true.**
  **The P-Value is based on an assumed or actual reference distribution (Normal, T-distribution, Chi-Square, F-distribution, etc.)**

- **Small "P-Value"**
- **Large "Z" or "T", etc**
- **$H_o$ is Rejected**

- **Large "P-Value"**
- **Small "Z" or "T", etc**
- **$H_o$ is Not Rejected**

# Interpreting the p-value…

Overwhelming Evidence that supports Ha
(Highly Significant)

Strong Evidence
(Significant)

Weak Evidence
(Not Significant)

No Evidence
(Not Significant)

.01 .05 .10

***"When the p is low – reject the Ho"***

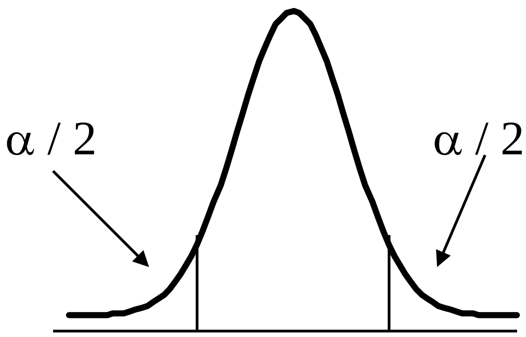
# What Is the "p-value"?

- **The p-value is the probability of** rejecting **the null hypothesis ($H_o$) when it is true (Type 1 error)**
- **Interpretation:**
  - **if this probability is low, then we conclude that the null hypothesis ($H_o$) must be wrong—that the sample** is not **from the assumed distribution**
- **Alternatively, consider the p-value to represent the risk that we are wrong if we conclude that the null hypothesis ($H_o$) is false—that we claim there is a difference when there is none.**

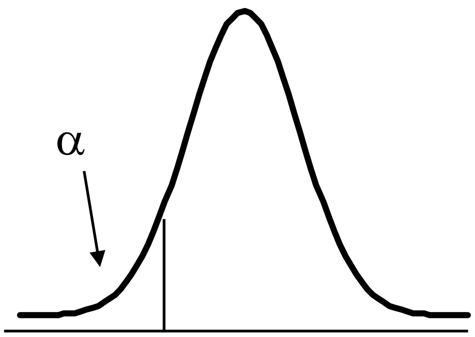
# How to Determine the Acceptance / Rejection Regions

The location of the Acc / Rej regions is always determined by the Alternative Hypothesis:

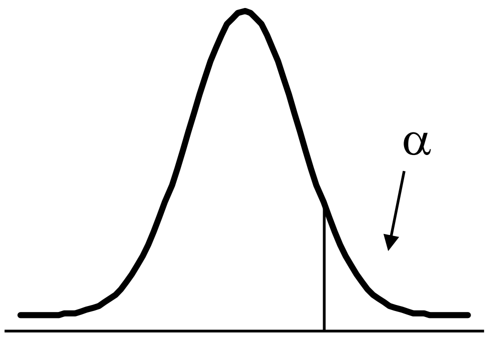
*The value of $\alpha$ is the area of the rejection region for the null hypothesis*




If Ha sign is $\neq$
two-tailed, rejection
region split




If Ha sign is $<$ ,
one-tail, rejection
region to the left

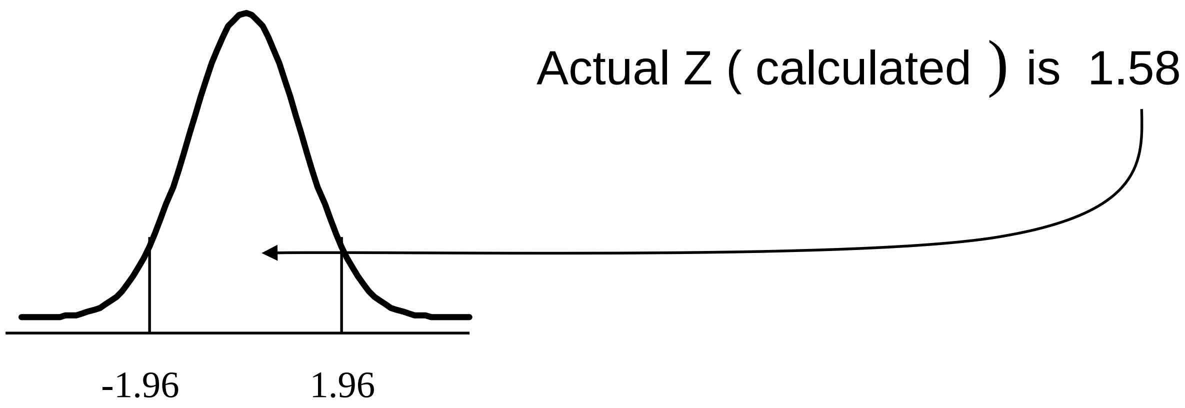



If Ha sign is $>$ ,
one-tail, rejection
region to the right

# Hypothesis Testing

Step 5: Compute the test statistic using your sample data and compare the calculated value with your critical value and

acceptance / rejection regions:




Step 6: Based on your comparison of the calculated value with the critical value ( and drawing ), you will either accept the Ho, or Reject the Ho. ***In this case what will you do ?***

# Example:  A Trial

**Jury's Decision**

| **The Truth** | **He's Not Guilty** | **He's Guilty** |
|---|---|---|
| **Actually Innocent** | **Correct** | **Type I Error** **(α-Risk)** |
| **Actually Guilty** | **Type II Error** **(β -Risk)** | **Correct** |

Consequence:

Innocent Man Goes to Jail.

Consequence: **Criminal Goes Free.**

# Conclusions of a Test of Hypothesis

- If we reject the null hypothesis, we conclude that there is enough evidence to infer that the alternative hypothesis is true.
- If we do not reject the null hypothesis, we conclude that there is not enough statistical evidence to infer that the alternative hypothesis is true.
- Remember the truth of the alternative hypothesis is what we are investigating. The conclusion focuses on the validity of the alternative hypothesis.

# Pregnancy Example

Doctors at John Hopkins Medical Center reported the following set of  n = 8 carrying times occurring during one month:

272   269   273   265   272   267   270   262

Does the sample indicate that the average carrying time has Changed from the known 266 days ?  Test at the 0.05 level of Significance.

Xbar = 268.8,  Sx = 3.8

# Use methods based on valid statistics

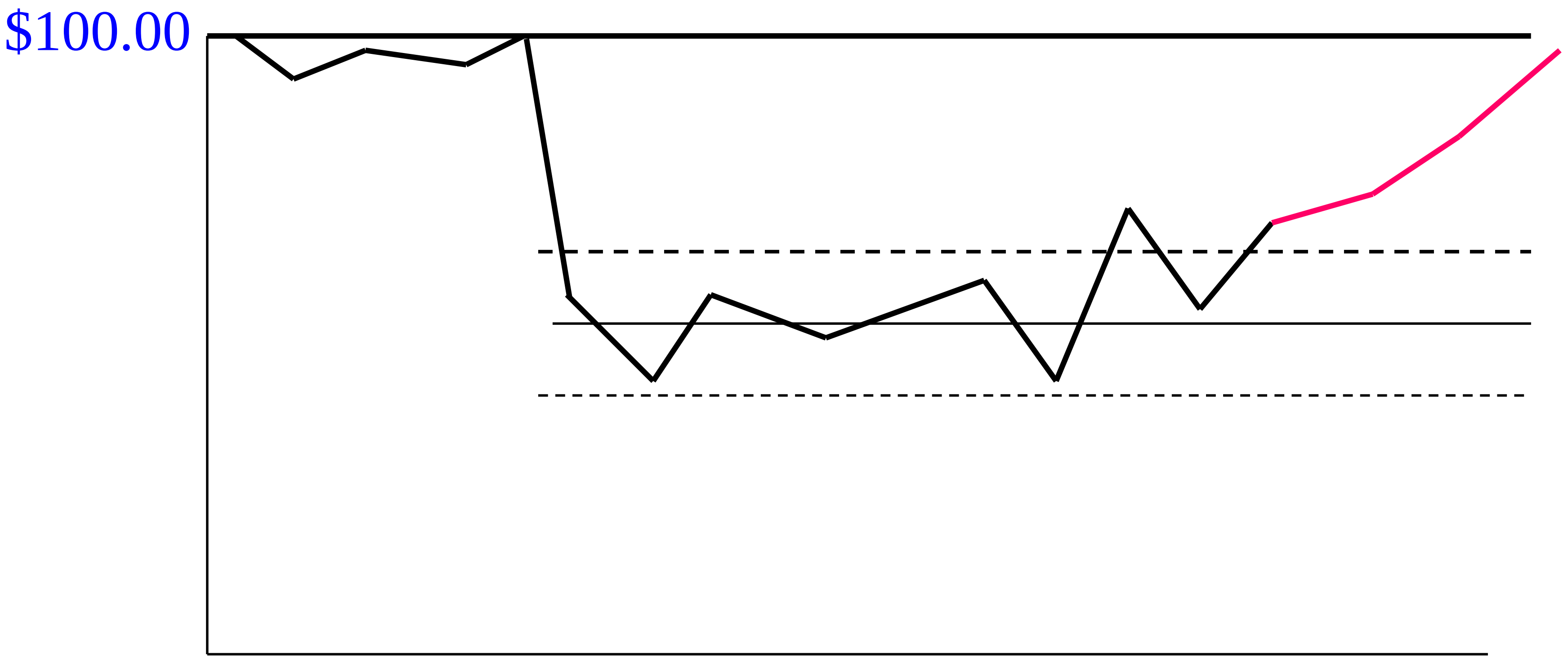
## Standard deviations are from qualified data

A more complete listing of methods is given in following table:[4]

| Method of estimating gestational age | Variability (2 standard deviations)[4] |
|---|---|
| Days from oocyte retrieval or co-incubation in in vitro fertilisation + 14 days | ±1 day |
| Days from estimated ovulation in Ovulation induction + 14 days | ±3 days |
| Days from artificial insemination + 14 days | ±3 days |
| Days from known single sexual intercourse + 14 days | ±3 days |
| Days from estimated ovulation by basal body temperature record + 14 days | ±4 days |
| First-trimester physical examination | ±2 weeks |

# Jackie's Checkbook

**A *classic* SPC Implementation story…..**



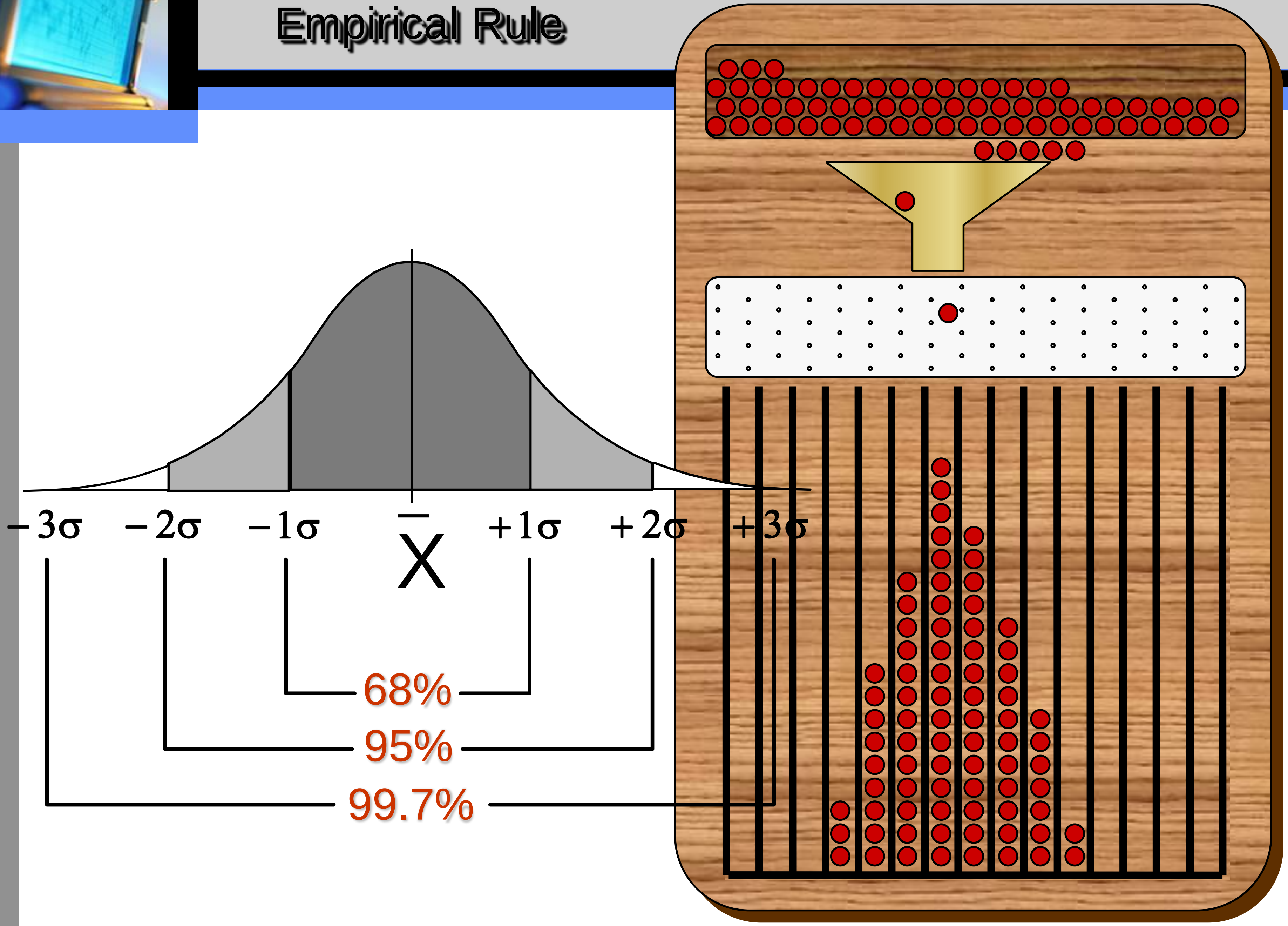
**Obstetricians have known for a long time that:**

**- Normal Distribution**

**- Avg = 266 days**

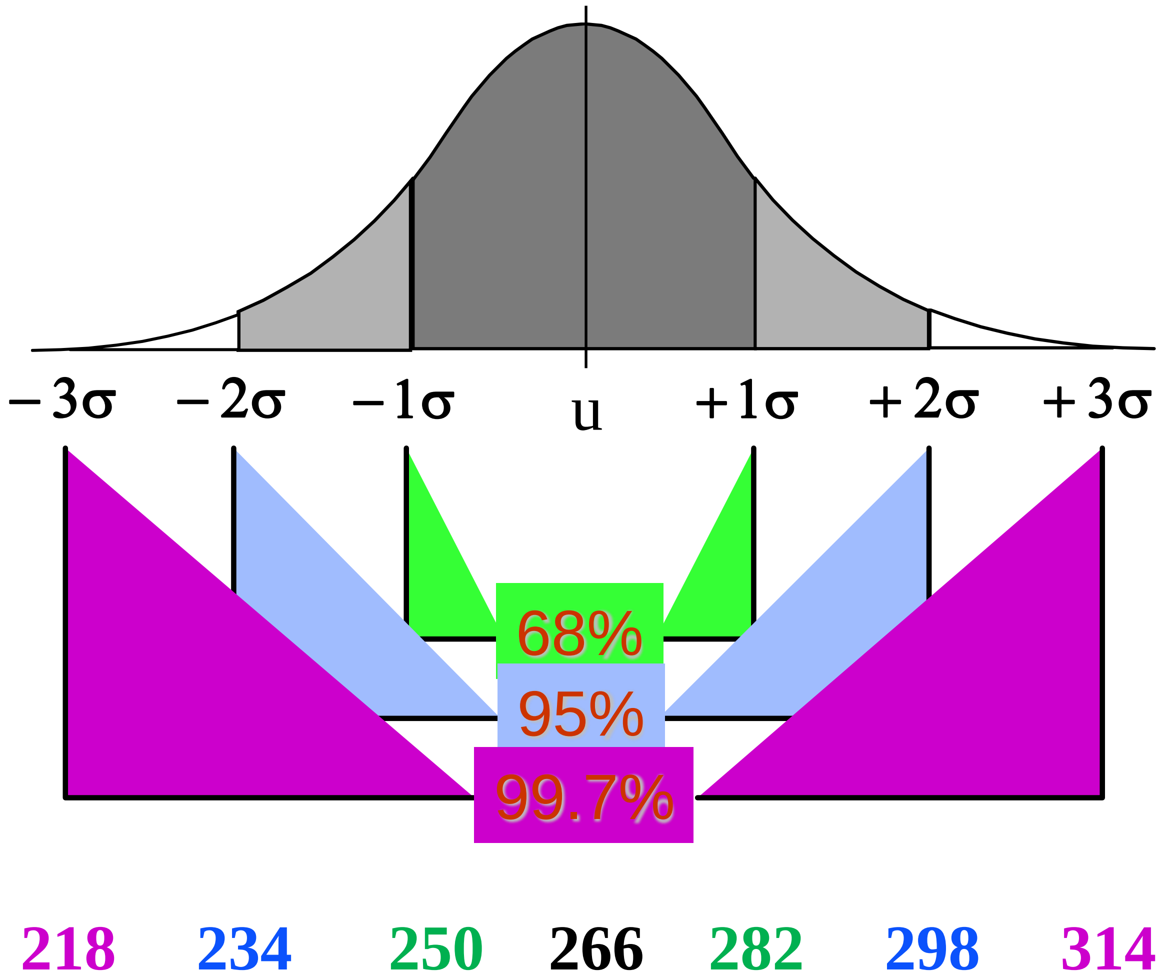
**- Std.Dev = 16 days**

# Empirical Rule



# Worried in San Diego (320)

Ob's have known for a long time: u = 266 σ = 16



# Summary of Thoughts for Inference Testing

- Use real validated data
- Test the assumptions for normality or ?
- Use confidence limits and degree of certainty
- Perform a hypothesis test and test your assumptions
- Use categories such as Demographics
- Stop guessing……… and
- Never say "I think…….

# Now…what if MEN got PREGNANT!

~ Maternity leave would last for two years….with full pay.

~There would be a cure for stretch marks.