Sessions 6-7

Hypothesis Testing

Managerial Decisions (revisited)

We will need 400 more person hours to finish this project.

The average spending per new customer will be greater than INR 8000

The retail market will grow by 50% in the next 5 years.

Our quality will not improve after the consulting project.

We will be able to rationalize the number of flights to 80% of the current level.

Our potential customers do not spend more than 60 minutes on the web every day.

Less than 5% clients will default on their loans.

Learning Objectives

- How and when to formulate hypotheses about population parameters?
- How to quantify the strength of the evidence?
- What are Type I and Type II errors ?

Hypothesis testing: formulation of null and alternative hypothesis

- Hypotheses are claims on population parameters, that are open to test and rejection in light of strong evidence against them

- The initial claim is called the null hypothesis (H_0)
  - Generally the status quo
  - Action taken: Do nothing

- A competing claim to the null hypothesis is called the alternative hypothesis (H_A, H_a, H_1)
  - Often a claim to be tested or a change to be detected
  - Evidence is presented against the null hypothesis
  - Action taken: Do something

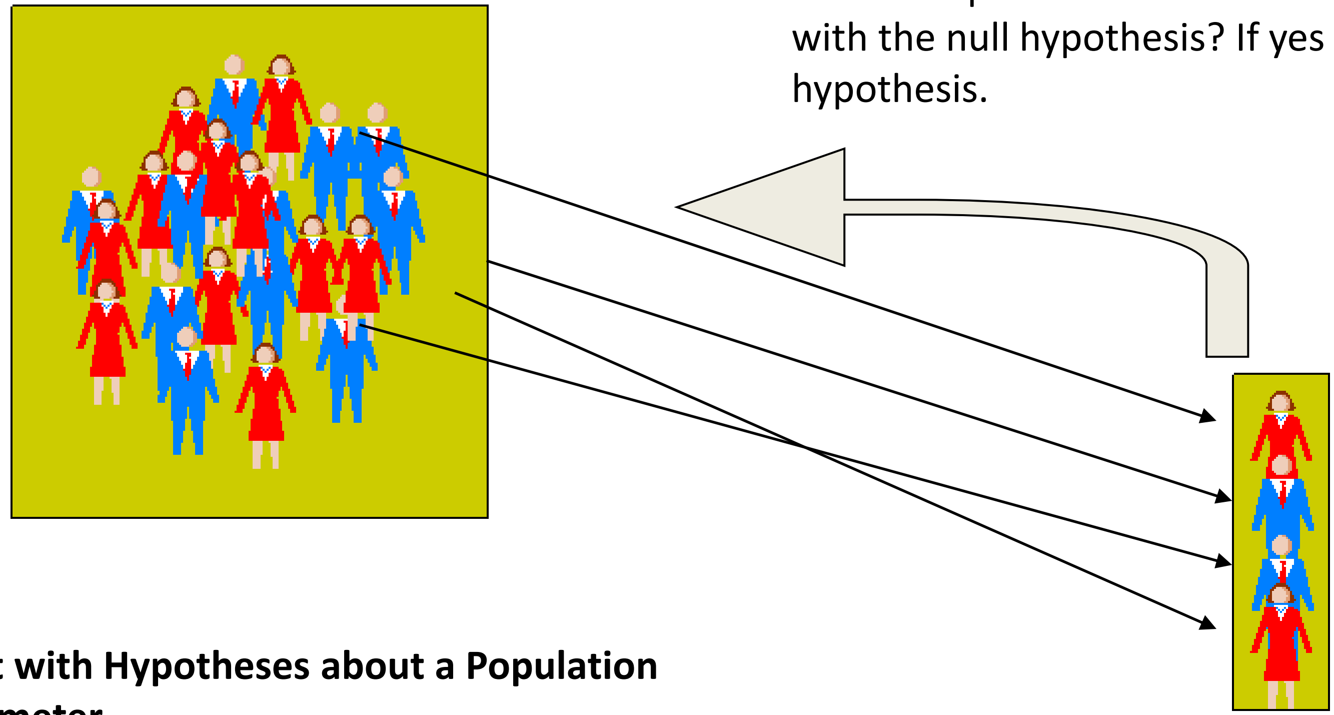
- The two hypotheses are
  - Mutually exclusive
  - Collectively exhaustive

Examples: formulation of null and alternative hypothesis

- After graduating from ISB, you started your own business. Now to get customers, you decided to do online advertising. But somehow it is not working – you get on an average about 100 hits on the site per day. An IT consulting company that is known to be good at search engine optimization proposes to increase your click through rate and get the number of visits up to 200 on an average. You would hire the consulting company if the consulting company can get more clicks than by yourself. Which hypotheses should you test?

- A truck company wants on-time delivery for 98% of the parts they order from a metal manufacturing plant. They have been ordering from Hudson Manufacturing but will switch to a new, cheaper manufacturer (Steel-R-Us) unless there is evidence that this new manufacturer cannot meet the 98% on-time goal. As a test the truck company purchases a random sample of metal parts from Steel-R-Us, and then determines if these parts were delivered on-time. Which hypotheses should they test?

Hypothesis Testing: collecting evidence



3. **Reject/Do Not Reject Hypothesis**
   Is the sample information strongly inconsistent with the null hypothesis? If yes then the reject hypothesis.

1. **Start with Hypotheses about a Population Parameter**
   Parameter could be mean, proportion or something else.

2. **Collect Sample Information**
   Collect information from a randomly chosen sample and calculate the appropriate sample statistic.

Example: Supermarket Loyalty Program

- A supermarket plans to launch a loyalty program if it results in an average spending per shopper of more than $120 per week

- A random sample of 80 shoppers enrolled in the pilot program spent an average of $130 in a week with a standard deviation of $40

- Should the loyalty program be launched?

Hypothesis testing: strength of the evidence

Question: Assuming null is true, how unlikely is it to observe the evidence obtained? In other words, how strong is the evidence against null hypothesis?

- p-value: The chance of obtaining a sample statistic that is as extreme or more extreme than the observed value of the sample statistic, given that the null hypothesis is true.

The lower the p-value → the stronger the evidence against H_0

The Testing Process

1. State the null and alternative hypothesis. Begin by assuming that H_0 (typically status quo) is true
   - e.g. I believe that the spending will be less than or equal to $120.
2. Quantify what is meant by a “strong enough evidence” to reject H_0
   - e.g. p-value should be less than 0.05

Management

3. Collect the evidence
   - e.g. A pilot resulted in average spending of $130 in a sample of 80 customers
4. Calculate the strength of the evidence or p-value
   - e.g. The probability of getting a sample average of $130 or more under H_0 is 0.01

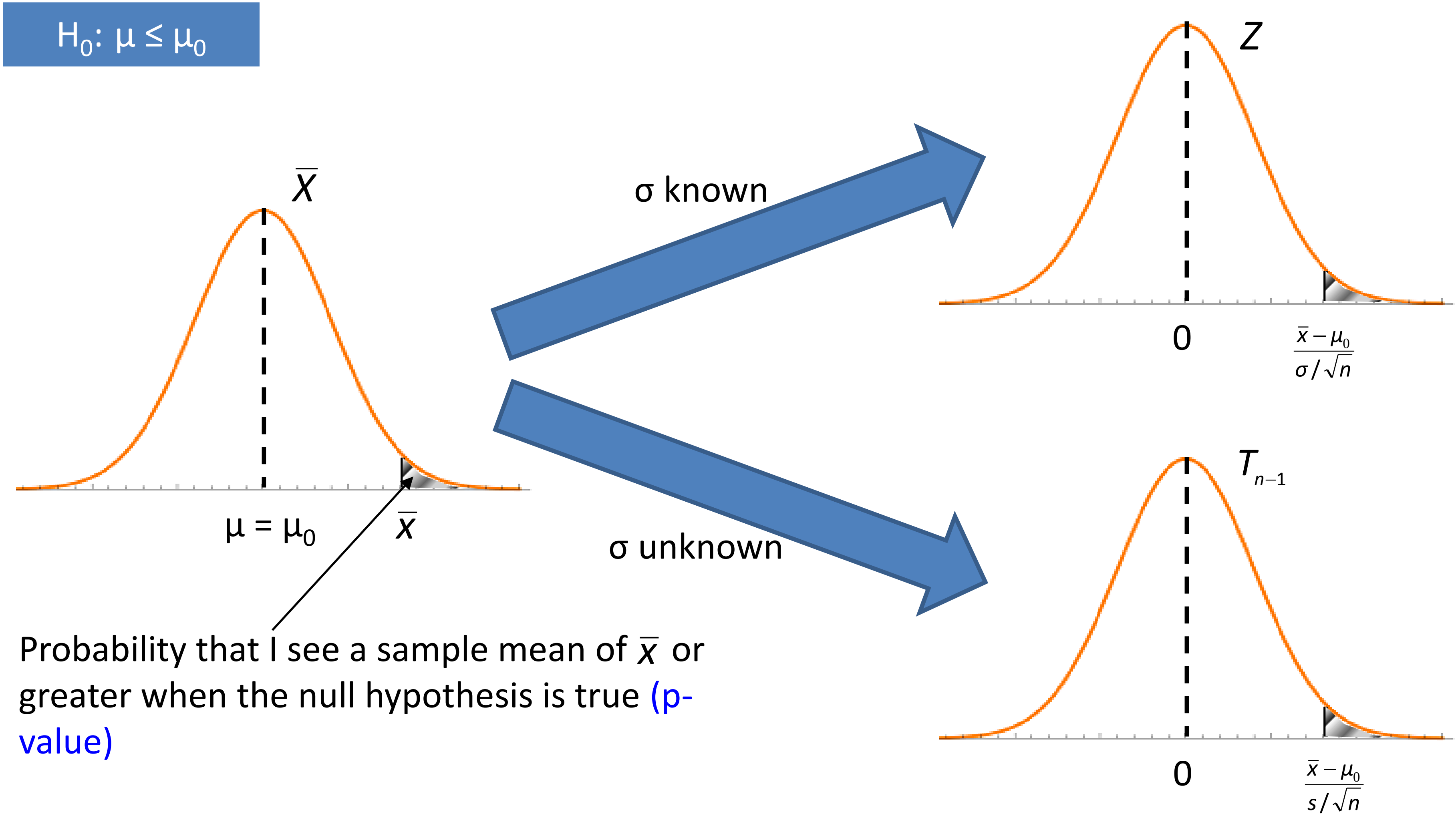
Statistics

5. Conclude and take appropriate action
   - e.g. The evidence is strong enough (0.01 < 0.05) to reject H_0; launch the card

Management

Calculating the strength of the evidence

H_0: $\mu \leq \mu_0$


$\bar{X}$
$\mu = \mu_0$
$\bar{x}$
σ known
σ unknown
Z
0
$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$
T_{n-1}
0
$\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

Probability that I see a sample mean of $\bar{x}$ or greater when the null hypothesis is true (p-value)

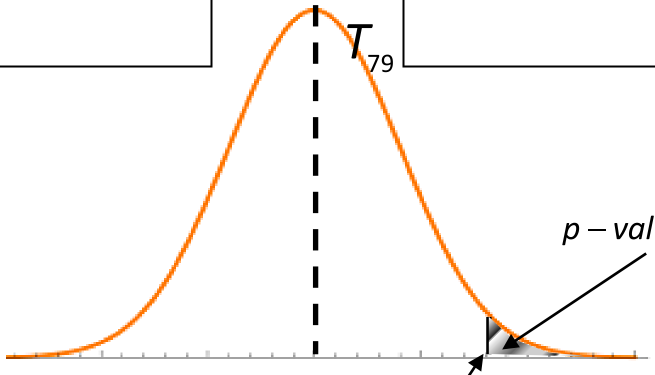
Right-tailed hypothesis test (sample mean of 130)

Null Hypothesis: The average spending is less than or equal to $120

$H_0 : \mu \leq \mu_0$ (120)

Acceptable level of type-I error is 0.05

($\alpha = 0.05$)



$p-value = 0.01$

$$t = \frac{130-120}{40/\sqrt{80}} = 2.236$$

Probability of seeing a sample mean of 130 or more if $\mu \leq 120$ is 0.01

OR

The evidence 130 is strong enough to reject the null hypothesis

OR

The average spending of 130 is unlikely to occur by chance in a sample of size 80

Launch the loyalty card

You can make two types of errors!

Decision Reality	Do not reject H_0	Reject H_0
H_0 is true	Correct decision	Type I error
H_A is true	Type II error	Correct decision

- P(Type-I error) = α (also called the significance level)

- P(Type-II error) ≠ 1 – P(Type-I error) because these errors occur in different versions of reality

Be cautious in your conclusion!

- First state your statistical conclusion from the hypothesis test
  - Reject or Fail to reject H_0 at a significance level of α

Reject the null hypothesis

- You have strong enough evidence to reject the null and go with the alternate
- Does not mean that the alternate hypothesis is true → you could have committed a Type I error
- Take the action associated with the rejection of null hypothesis

"Fail to reject" the null hypothesis

- You are saying that you do not have strong enough evidence to reject the null
- Does not mean that the null hypothesis is true → you could have committed a Type II error
- Continue with the status quo

Example: Spam Filter

- A small company is considering buying a commercial filtering software costing $15,000 per annum that the vendor claims will significantly reduce spam.

- The company believes that if the software can reduce spam to less than 20% of the mails in the inbox, then it is worth the cost because of improved employee productivity.

- The company collected a simple random sample of 100 mails in the inbox with the new filtering software in place and found that 11 of them were spam.

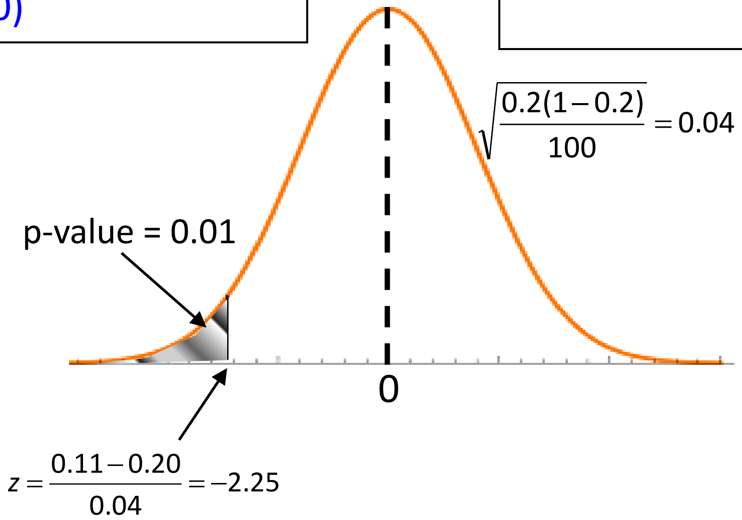
- Should it buy the software? Assume that the significance level is $\alpha = 0.05$.

Left-tailed hypothesis test (sample proportion of 0.11)

Null Hypothesis: The software cannot reduce the spam filter rate to less than 20%
$H_0: \pi \geq \pi_0$ (0.20)

Acceptable level of type-I error is 0.05

($\alpha = 0.05$)



Probability of seeing a sample of 0.11 or less if π ≥ 0.20 is 0.01

OR

The evidence 0.11 is sufficiently low and seems to favor H_A

OR

The evidence 0.11 is highly unlikely to occur by chance in a sample of size 100 if π ≥ 0.20

Buy the spam filter software

Example: Process Control at a Call Center

- Performance of a call center is monitored by the average call duration
- Data from 18 months shows that on the days when the process runs normally (μ = 4 min, σ = 3 min)
- Cannot monitor each and every call due to limited resources; so randomly sample 50 calls per day

Day	Mean Call Duration
1	3.7
2	4.1
3	3.5
4	4.2
5	3.9
6	4.1
7	4.2
8	3.8
9	3.7
10	4.6
11	3.7
12	4.6
13	4.0
14	4.2
15	3.8
16	4.4
17	5.3
18	6.1
19	7.2
20	6.5

ISB

Process control: Two sources of variability

- We already know that sample mean every day will be different – inherent variability

- But when should you be alarmed and conclude that the system is not behaving normally – external variability

- A pragmatic approach is to say:
  - I believe that the process is unchanged, i.e., μ = 4
  - Bring strong enough evidence to make me to change my mind, i.e., $\overline{x}$ is very different from μ
  - I am looking for deviations on either side of μ as evidence

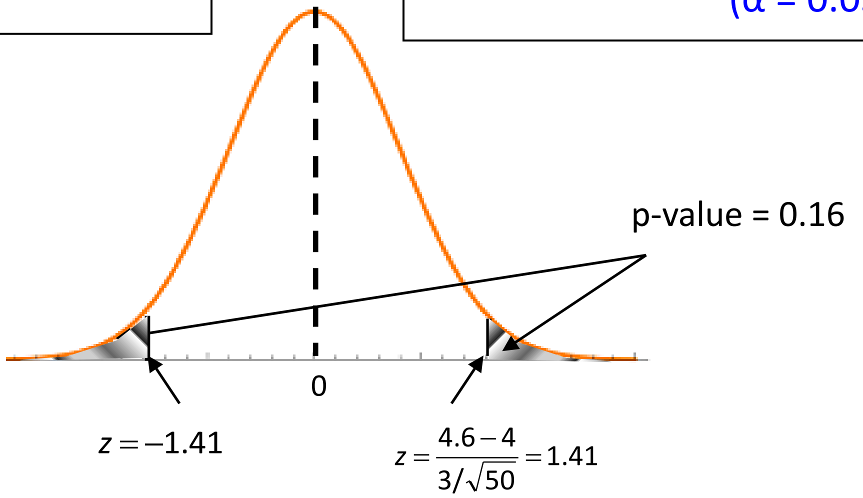
Two-tailed hypothesis test (sample mean of 4.6)

Null Hypothesis: The mean call duration is 4 minutes

H_0: μ = μ_0 (4)

Acceptable level of type-I error is 0.05

(α = 0.05)



Probability of seeing a sample mean of ≥ 4.6 or ≤ 3.4 is 0.16

OR

The evidence 4.6 is not strong enough to reject H_0

OR

The sample mean 4.6 occurred probably just due to chance if μ = 4

I cannot conclude that the process has changed → Do not investigate

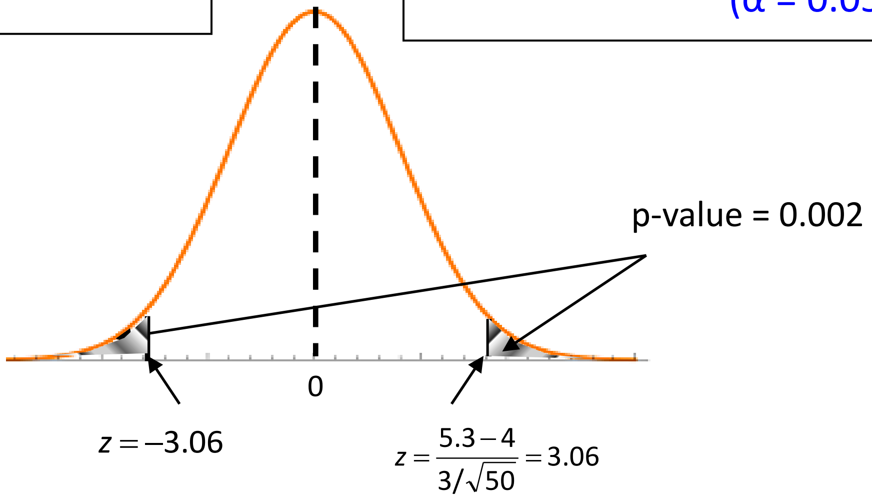
Two-tailed hypothesis test (sample mean of 5.3)

Null Hypothesis: The mean call duration is 4 minutes

H0: $\mu = \mu_0$ (4)

Acceptable level of type-I error is 0.05

($\alpha = 0.05$)



Probability of seeing a sample mean of ≥ 5.3 or ≤ 2.7 is 0.002

OR

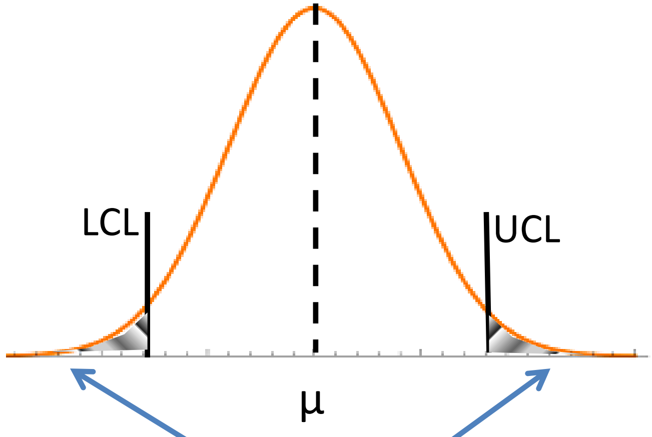
The evidence 5.3 is sufficiently high and seems to favor H_A

OR

An average call duration of 5.3 is highly unlikely to occur by chance if $\mu = 4$

I conclude that the process has changed → Investigate

“Control limits” to routinize hypothesis testing



Type I error (α)

- In some contexts, routinely calculating the p-value for sample means and comparing to α-value is tedious
- Instead, managers calculate “target sample mean” corresponding to α-value and compare it with sample mean at hand

$$UCL = \mu + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad LCL = \mu - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

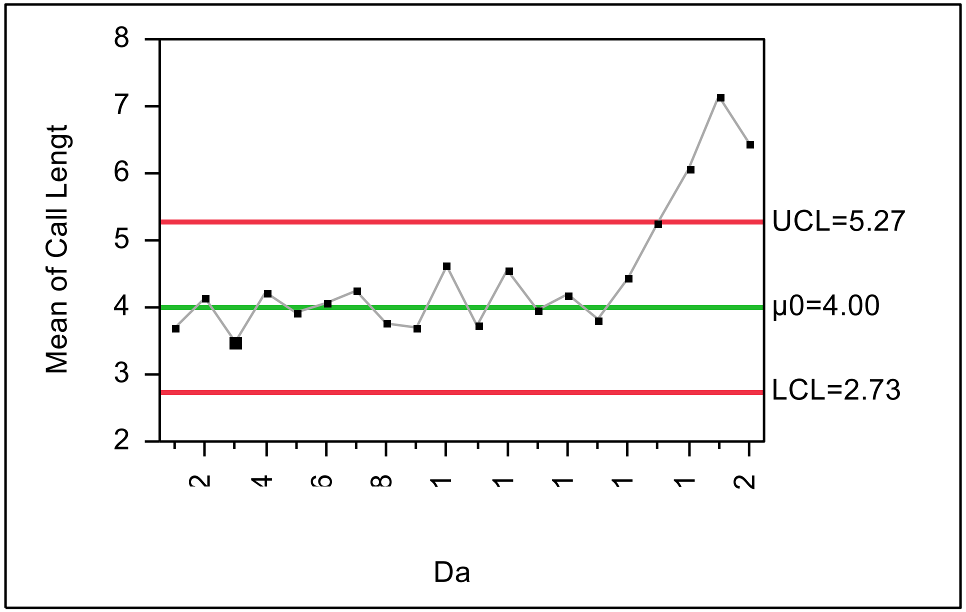
- These target sample means are called control limits
- Standard practice in industry
  - $z_{\alpha/2}$ = 3, which corresponds to 1-α = 99.7%,
  - UCL = 5.27 and LCL = 2.73

How to set control limits (i.e., decide α)?

- Key decision of the process manager
- Control limits do not depend on the sample data
- For a given level of variability in the process, there is a trade-off between two types of errors

	Type I error	Type II error
Cause	Process has not changed but got an unusual sample	Process has changed but the sample appears usual
Consequence	Out of pocket cost of investigation into the root cause	Opportunity cost of allowing an inappropriate process to run

Control Charts: Implementing Control Limits in Practice



- This chart is called "Xbar chart" because it tracks sample mean ($\bar{X}$)
- Visual tool to control the process
- Any deviation outside the band will attract investigation
- The longer run goal is to reduce σ and tighten the control limits (process improvement)

Statistical Vs. Practical Significance

- There could be a difference between a statistically significant result and a practically important one

- Large sample sizes often give statistically significant results, even if it has low economic value

- E.g. A recent study of productivity of radiologists in telemedicine
  - n = 2.8 million “readings”
  - Every 1000 additional cases reduced the “reading time” by 13.83 seconds (p-value < 0.001)

Summary of Session VI-VII

1. When to perform hypothesis test about population parameters?

- Hypothesis is an assumption about a population parameter that is subject to a test and rejection based on evidence
- Hypothesis test is applicable when the manager has specific position on a population parameter which needs to be rejected in order to take action

2. How to conduct hypothesis test for population mean and population proportion?

- A manager typically targets a certain type-I error called level of significance.
- If the p-value for a given sample is less than α-value (the level of significance) under null hypothesis, the manager rejects the null hypothesis and takes appropriate action.

3. What are Type I and Type II errors ?

- Given the uncertainty, we can never reject or not reject a hypothesis with certainty
  - Type I error – Null hypothesis is true but you reject it based on “unusual” sample
  - Type II error – Null hypothesis is not true but you accept it based on “usual” sample

Pop Quiz!

- The management of a chain of hotels avoids intervention in the local management of a property owned by franchisees unless problems become far too common. Specifically, it will intervene only if the fraction of customers satisfied with the hotel's service drops below 33%. A survey of 80 guests in a hotel found that 16 of them would return to the hotel next time they are in the city.
  - State the null and alternate hypothesis
  - Describe Type-I and Type-II errors are and the associated costs in this context.
  - Do the data supply evidence to intervene if the acceptable level of significance is 0.025.

Pop Quiz!

A skydiver is preparing to jump out of a plane using a new parachute. Under the null hypothesis that the parachute will open, the Type I error means:

A. The person jumps and the parachute will not open

B. The person jumps and the parachute opens

C. The person does not jump and the parachute would have opened

D. The person does not jump and the parachute would not have opened

Note: The skydiver does not wish to commit suicide