جامـــــعة المـــــستقبل
AL MUSTAQBAL UNIVERSITY

قســـم الامـــــن الـــــــسيبرانـــــــي

Department of Cyber Security

Subject:

Inferential Statistics

Class:

first

Lecturer:

Asst. lect. Mustafa Ameer Awadh

Lecture: (10)

Lecture: Inferential Statistics

Introduction

Inferential statistics allows us to make predictions or inferences about a population based on data collected from a sample. Unlike descriptive statistics, which summarize data, inferential statistics help us draw conclusions and test hypotheses.

---

1. Sampling Distributions

Definition:

A sampling distribution is the probability distribution of a statistic (e.g., mean, proportion) obtained from repeated random samples of the same size.

Key Concepts:

- **Population vs. Sample:** The population includes all individuals, while a sample is a subset.
- **Sampling Variability:** Different samples can lead to different estimates.
- **Central Limit Theorem (CLT):**
  - If the sample size is large ($n \geq 30$), the sampling distribution of the sample mean approaches a normal distribution, regardless of the original population's distribution.
  - The mean of the sampling distribution equals the population mean ().
  - The standard deviation of the sampling distribution (standard error) is given by: where is the population standard deviation and is the sample size.

---

2. Confidence Intervals (CIs)

Definition:

A confidence interval provides a range of values that likely contain the true population parameter with a certain level of confidence.

Formula for Confidence Interval (Mean):

For a population mean (when is known): where:

- = sample mean
- = critical value from the standard normal table (e.g., 1.96 for 95% confidence)
- = standard error

Common Confidence Levels:

Confidence Level	Z-Score
90%	1.645
95%	1.96
99%	2.576

Interpretation:

A 95% confidence interval means that if we repeated sampling many times, 95% of the intervals would contain the true population parameter.

---

3. Hypothesis Testing

Definition:

Hypothesis testing is a statistical method used to decide whether to accept or reject a hypothesis about a population.

Steps in Hypothesis Testing:

1. **State the Hypotheses:**
   - **Null Hypothesis ()**: Assumes no effect or difference.
   - **Alternative Hypothesis ()**: Assumes an effect or difference exists.
2. **Select the Significance Level ()**
   - Typical values: 0.05 (5%) or 0.01 (1%).
   - A smaller reduces the risk of a false positive.
3. **Choose the Test Statistic**
   - **Z-test** (when population standard deviation is known and sample size is large)
   - **T-test** (when is unknown and sample size is small)

4. **Compute the p-value**
   - The probability of obtaining a test statistic as extreme as the observed one, assuming is true.
5. **Compare p-value with**
   - If **p-value** ≤ → Reject (significant result)
   - If **p-value** > → Fail to reject (not significant)

Example:

A company claims that the average weight of their product is 500g. A sample of 30 products has a mean weight of 495g with a standard deviation of 10g. Test at .

-
-
- Compute Z-score:
- Find p-value: (less than 0.05, so reject)
- Conclusion: The product weight is significantly different from 500g.

---

4. p-values and Significance Levels

Definition:

The p-value measures the probability of observing the given test statistic under .

Interpretation:

- **Low p-value (≤ 0.05):** Strong evidence against ; reject it.
- **High p-value (> 0.05):** Weak evidence against ; fail to reject it.
- **p-value = 0.05:** Borderline case.

Level	Confidence Level	Decision Rule
0.10	90%	Less strict
0.05	95%	Standard
0.01	99%	Very strict

Conclusion

- **Sampling distributions** help estimate population parameters.
- **Confidence intervals** provide a range where the true value lies.
- **Hypothesis testing** determines whether there is enough evidence to support a claim.
- **p-values** help assess the strength of evidence against the null hypothesis.