Ch5 Key Concepts in Hypothesis Testing

Jih-Chang Yu

August 31, 2023

Motivating Example 1

- National Center for Education Statistics reports 2007 fourth-grade reading scores.
- National mean: 220.99, standard deviation: 35.73.
- Data from National Assessment of Educational Progress.
- Belief: Your school district excels in teaching reading.

Sample Data and Initial Inference

- Randomly selected 50 fourth-grade students from your district.
- Administered the same exam.
- Sample mean: 230.2.
- Seems to support your belief.
- Critic's concern: Sample luck could influence the result.

Eliminating Sampling Variability

- Limited resources allow only sample data.
- Question: Would testing all fourth graders yield a mean similar to the national average?
- Addressing sampling variability as an explanation.

Components of Statistical Inference

- Statistical inference consists of two parts:
    - Statement about parameter value.
    - Measure of statement reliability, often as probability.
- Traditional inference objectives:
    - Conducting tests of hypotheses:
        - Hypothesize specific parameter values or relationships.
        - Decisions based on sample statistics.
        - Reliability: Probability of incorrect decisions.
    - Estimating parameters:
        - Using sample statistics, usually intervals.
        - Reliability: Confidence level in the interval.

Motivating Example 2

- Company packages salted peanuts in 8-oz. jars.
- Goal: Control peanuts amount in jars from a machine.
- Power is defined as averaging 8 oz. per jar.
- Avoid consistent over- or underfilling.

Sample and Hypothesis Testing

- Monitor control with a sample of 16 jars.
- Jars randomly sampled at different time intervals.
- Contents weighed for each jar in the sample.
- Mean weight of peanuts in these 16 jars used for hypothesis testing.
- Hypothesis: Machine working properly.

Costly Adjustment

- If machine not working properly:
- Costly adjustment needed.

Definition of Null Hypothesis

Definition The null hypothesis is a statement about the values of one or more parameters. This hypothesis represents the status quo and is usually not rejected unless the sample results strongly imply that it is false.

- **Null Hypothesis (H0):** No effect, no difference, or no relationship.

Definition of Alternative Hypothesis

Definition The alternative hypothesis is a statement that contradicts the null hypothesis. This hypothesis is declared to be accepted if the null hypothesis is rejected. The alternative hypothesis is often called the research hypothesis because it usually implies that some action is to be performed, some money spent, or some established theory overturned.

- **Alternative Hypothesis (H1 or Ha):** Expected effect, difference, or relationship.

Example:

Suppose you're testing a new drug's impact on blood pressure:

$H0$:"The new drug has no impact on blood pressure."

$H1$:"The new drug has an impact on blood pressure."

Definition of Rejection Region

Definition

The rejection region is a range of values determined by a critical value or values. It represents the set of sample outcomes that would lead to the rejection of the null hypothesis in a hypothesis test. Sample results falling within the rejection region provide strong evidence against the null hypothesis.

Rejection Region in Hypothesis Testing

Consider the example of the company packaging salted peanuts:

- The null hypothesis (H_0) assumes the machine is working properly, i.e., the mean weight is 8 oz.
- The alternative hypothesis (H_1) suggests the machine is not working properly, leading to deviations from 8 oz.
- To determine whether to reject H_0, we establish a critical region.
  - In hypothesis testing, we establish a critical value denoted as C. The critical value C defines the boundary of the **rejection region**.
  - If the testing statistic T is larger than C or lesser than C, we reject the null hypothesis (H_0).

Type I and Type II Errors

- **Type I Error:** We reject H_0 when it's actually true.
  - This is like a "false positive" situation.
  - It's like accusing someone of something they didn't do.
  - The probability of making a Type I error is denoted by α
- **Type II Error:** We fail to reject H_0 when it's actually false.
  - This is a "false negative" scenario.
  - It's like missing something that's right in front of us.
  - The probability of making a Type II error is denoted by β.

Significance Level

- The significance level (α) sets the threshold for rejecting the null hypothesis.
- Common values: 0.05, 0.01.

Example:

Let's set $\alpha = 0.05$. This means we're willing to accept a 5% chance of making a Type I error (rejecting H0 when it's true).

Sample Size and Sampling

- A proper sample size ensures the ability to detect real effects.
- Random and representative sampling is crucial.

Example:

For our drug study, we need a sample size that provides enough statistical power to detect changes in blood pressure.

Type I Error: False Positive (replenish)

- Occurs when a true null hypothesis is incorrectly rejected.
- Consequences can be significant:
  - Unnecessary actions, expenses, or treatments.
  - Wrong conclusions leading to misguided decisions.

Type II Error: False Negative (replenish)

- Occurs when a false null hypothesis is not rejected.
- Consequences also important:
    - Missed opportunities, delayed actions.
    - Wrong conclusions that there is no effect when there actually is one.

Importance of Type I Errors

- In many scenarios, Type I errors are considered more important than Type II errors.
- Why?
  - Immediate negative impact: Incorrect claims, actions, or treatments.
  - Potential life-altering consequences in fields like medicine and law.

Balancing the Errors

- Both errors have trade-offs and must be balanced.
- Context matters: Consequences, goals, costs.
- Striving for a balance that minimizes both errors is crucial.

Conclusion

- Type I errors can lead to immediate negative impacts and wrong decisions.
- The balance between errors depends on the situation.
- Importance of understanding and managing these errors cannot be overstated.

Example for Type I Error:

- **Medical Testing:** A pregnancy test produces a positive result for a woman who is not actually pregnant (H0: Not pregnant).
- **Legal System:** Convicting an innocent person based on insufficient evidence (H0: Innocent).
- **Scientific Research:** Declaring a new treatment effective when it's actually no better than a placebo (H0: No effect).
- **Quality Control:** Rejecting a batch of products as defective when they are actually within acceptable standards (H0: Within standards).

Example for Type II Error:

- **Medical Testing:** A patient has a disease, but the test fails to diagnose it (H0: Disease absent).
- **Market Research:** Failing to identify a potentially successful product due to inconclusive market data (H0: No market potential).

Balancing Errors with Rejection Region

- The rejection region determines the trade-off between Type I and Type II errors.
- Smaller α reduces Type I error but might increase Type II error.
- It's important to choose an appropriate α based on the situation.

One-Sample T-Test

Definition
The one-sample t-test compares the mean of a sample to a known population mean to determine if they are significantly different.

Assumptions

1. Random sampling from the population.
2. Normally distributed population or large sample size.
3. Independence of observations.

Rejection Regions in Hypothesis Testing

Let's use the Bernoulli distribution with parameter p to calculate the rejection regions for different hypotheses.

- Consider a binary event with probability p of success.
- We're interested in testing hypotheses about p.

Rejection Region Calculation with n Bernoulli Trials and Significance Level α

- $H_0 : p > p_0$:
  - Rejection region: Any outcome with k or fewer successes out of n trials, where k is determined by the value that satisfies $P(X \leq k) \leq \alpha$.
- $H_0 : p < p_0$:
  - Rejection region: Any outcome with k or more successes out of n trials, where k is determined by the value that satisfies $P(X \geq k) \leq \alpha$.
- $H_0 : p = p_0$:
  - Rejection region: Any outcome with k or fewer or k or more successes out of n trials, where k is determined by the values that satisfy $P(X \leq k) \leq \frac{\alpha}{2}$ and $P(X \geq k) \leq \frac{\alpha}{2}$.

Rejection Region Calculation using Central Limit Theorem

Let's apply the Central Limit Theorem (CLT) to calculate the rejection regions for different hypotheses when we have n independent Bernoulli trials with probability p, considering a significance level α.

- $H_0 : p > p_0$:
  - Rejection region: Any outcome with a sample proportion $\hat{p}$ that is less than $p_0 - z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}$, where z_α is the critical z-score for significance level α.
- $H_0 : p < p_0$:
  - Rejection region: Any outcome with a sample proportion $\hat{p}$ that is greater than $p_0 + z_\alpha \sqrt{\frac{p_0(1-p_0)}{n}}$.
- $H_0 : p = p_0$:
  - Rejection region: Any outcome with a sample proportion $\hat{p}$ that is less than $p_0 - z_{\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$ or greater than $p_0 + z_{\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$.

Rejection Region Calculation using Poisson Distribution

Let's utilize $X_1, X_2, \ldots, X_n$, which are independent and identically distributed Poisson random variables, to calculate the rejection regions for different hypotheses, considering a significance level α.

- $H_0 : \lambda > \lambda_0$:
  - Rejection region: Any outcome with k or fewer events, where k is determined by the value that satisfies $P(X_1 + X_2 + \ldots + X_n \leq k) \leq \alpha$.
- $H_0 : \lambda < \lambda_0$:
  - Rejection region: Any outcome with k or more events, where k is determined by the value that satisfies $P(X_1 + X_2 + \ldots + X_n \geq k) \leq \alpha$.
- $H_0 : \lambda = \lambda_0$:
  - Rejection region: Any outcome with k or fewer or k or more events, where k is determined by the values that satisfy $P(X_1 + X_2 + \ldots + X_n \leq k) \leq \frac{\alpha}{2}$ and $P(X_1 + X_2 + \ldots + X_n \geq k) \leq \frac{\alpha}{2}$.

Distribution of Sum of iid Poisson Random Variables

If you have n independent and identically distributed (iid) Poisson random variables $X_1, X_2, \ldots, X_n$ with parameter λ, their sum follows a Poisson distribution with parameter $n\lambda$.

Mathematically:

$$X_1 + X_2 + \ldots + X_n \sim \text{Poisson}(n\lambda)$$

So, when you add up n iid Poisson random variables, the resulting distribution is a Poisson distribution with the parameter being the product of n and the individual parameter λ.

Hypothesis Testing using CLT - $H_0 : \lambda > \lambda_0$

- The CLT states that $\bar{X}$ approximately follows a Normal distribution with mean λ and standard deviation $\frac{\sqrt{\lambda}}{\sqrt{n}}$.
- Calculate the z-score: $z = \frac{\bar{X}-\lambda}{\frac{\sqrt{\lambda}}{\sqrt{n}}}$.
- Determine the critical z-value z_α for significance level α.
- If $z > z_\alpha$, reject H_0 in favor of the alternative $H_1 : \lambda > \lambda_0$, which means

$$\bar{X} > \lambda + z_\alpha \cdot \frac{\sqrt{\lambda}}{\sqrt{n}}$$

- Otherwise, fail to reject H_0.

Hypothesis Testing using CLT

$H_0 : \lambda < \lambda_0$

- If $z < -z_\alpha$, reject H_0 in favor of the alternative $H_1 : \lambda < \lambda_0$, which means
$$\bar{X} < \lambda - z_\alpha \cdot \frac{\sqrt{\lambda}}{\sqrt{n}}$$
- Otherwise, fail to reject H_0.

$H_0 : \lambda = \lambda_0$

- If $|z| > z_\alpha$, reject H_0 in favor of the alternative $H_1 : \lambda \neq \lambda_0$.
- Otherwise, fail to reject H_0.

Hypothesis Testing under Exponential Distribution

Let's explore how to perform hypothesis testing using the Central Limit Theorem (CLT) for different cases under the exponential distribution with mean λ.

Hypothesis Testing using CLT - $H_0 : \lambda > \lambda_0$

Consider a sample of n independent and identically distributed (iid) exponential random variables $X_1, X_2, \ldots, X_n$ with mean λ. We want to test the hypothesis $H_0 : \lambda > \lambda_0$.

- Calculate the sample mean $\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$.
- Under H_0, the expected value of $\bar{X}$ is λ and the standard deviation is $\frac{\lambda}{\sqrt{n}}$.
- The CLT applies when n is sufficiently large.

Hypothesis Testing using CLT - $H_0 : \lambda > \lambda_0$

- The CLT states that $\bar{X}$ approximately follows a Normal distribution with mean λ and standard deviation $\frac{\lambda}{\sqrt{n}}$.
- Calculate the z-score: $z = \frac{\bar{X} - \lambda}{\frac{\lambda}{\sqrt{n}}}$.
- Determine the critical z-value z_α for significance level α.
- If $z > z_\alpha$, reject H_0 in favor of $H_1 : \lambda > \lambda_0$, which means

$$\bar{X} > \lambda + z_\alpha \cdot \frac{\lambda}{\sqrt{n}}$$

- Otherwise, fail to reject H_0.

Hypothesis Testing using CLT

For $H_0 : \lambda < \lambda_0$, the procedure is analogous:

- If $z < -z_\alpha$, reject H_0 in favor of $H_1 : \lambda < \lambda_0$.
- Otherwise, fail to reject H_0, which means

$$\bar{X} < \lambda - z_\alpha \cdot \frac{\lambda}{\sqrt{n}}$$

For $H_0 : \lambda = \lambda_0$, we need a two-tailed test:

- If $|z| > z_{\alpha/2}$, reject H_0 in favor of $H_1 : \lambda \neq \lambda_0$.
- Otherwise, fail to reject H_0.

Hypothesis Testing of the normal distribution with Known σ

Given: $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$, σ is known.

- Null Hypothesis (H_0):
    - $H_0 : \mu < \mu_0$
    - Rejection Region: $\bar{X} < \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$
- Alternative Hypothesis (H_1):
    - $H_1 : \mu > \mu_0$
    - Rejection Region: $\bar{X} > \mu_0 + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$
- Null Hypothesis (H_0):
    - $H_0 : \mu = \mu_0$
    - Rejection Region: $|\bar{X} - \mu_0| > z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

z_α and $z_{\alpha/2}$ are critical values from the standard normal distribution.

Understanding p-value

- The p-value is a measure of evidence against a null hypothesis.
- It indicates how extreme the observed data is under the assumption that the null hypothesis is true.
- A low p-value suggests that the observed data is unlikely under the null hypothesis.
- Commonly used in hypothesis testing to make decisions about rejecting or not rejecting the null hypothesis.

Calculating p-value

- For a given test statistic, the p-value is calculated based on the probability of observing a value as extreme as the test statistic.
- In a two-tailed test, the p-value accounts for extreme values in both tails of the distribution.
- In a one-tailed test, the p-value considers only one tail of the distribution.
- Smaller p-value indicates stronger evidence against the null hypothesis.

Two-tailed Test

- In a two-tailed test, the p-value is calculated as twice the probability of observing a test statistic as extreme as the one observed.
- If p-value $< \alpha$ (usually the significance level), you may reject the null hypothesis.
- This suggests that the observed data is significantly different from what the null hypothesis predicts.

One-tailed Test

- In a one-tailed test, the p-value is calculated based on the probability of observing a test statistic as extreme as the one observed in a specific direction.
- If p-value $< \alpha$, you may reject the null hypothesis.
- This suggests that the observed data is significantly greater or smaller (depending on the direction) than what the null hypothesis predicts.

Conclusion: One-tailed vs. Two-tailed

- Use a one-tailed test when you are specifically interested in deviations in one direction.
- Use a two-tailed test when deviations in both directions are of interest.
- The choice between one-tailed and two-tailed test should be based on the research question and the nature of the hypothesis being tested.

Statistical Revolution: Gaust and Fisher

Before Gaust, statistical data were often large, analyzed using methods based on the central limit theorem and the assumption of normal distribution. K. Pearson, a statistical authority, even regarded normal distribution as divinely ordained.

However, 20th century saw a shift due to analysis of data from controlled experiments. With smaller datasets, doubts emerged about traditional methods relying solely on the central limit theorem. This is where Gaust(William Sealy Gosset) and Fisher(Sir Ronald Aylmer Fisher) entered the scene as pioneers.

Gaust and Fisher: Changing Paradigms

Gorst encountered small sample sizes, often just four or five data points. Through extensive data collection, he observed that:

The distribution of $t = \sqrt{n}(\bar{x} - \mu)/s$ differed significantly from the traditional standard normal distribution. Particularly, the tail probabilities were notably distinct.

This discrepancy led Gaust to question whether another distribution family existed. However, his statistical expertise fell short to address this challenge.

Statistical Evolution: Gaust and Fisher

Gaust's pursuit for answers led him to approach K. Pearson in 1906. He immersed himself in the study of statistics under Pearson's guidance from 1906 to 1907. The focus was on addressing statistical issues arising from the analysis of data with small sample sizes.

Despite Gaust's remarkable observations, his limited statistical knowledge hindered him from fully resolving the problems he encountered. Thus, Pearson's mentorship played a crucial role in shaping the future of statistical analysis.

Introduction to the t-Distribution

- In scenarios where the true population standard deviation (σ) is unknown, we must estimate it from the sample data.
- This necessity leads us to the utilization of the t-distribution.
- The t-distribution becomes crucial in hypothesis testing, particularly when dealing with smaller sample sizes.
- Estimating the population standard deviation introduces uncertainty, making the t-distribution an appropriate tool.

One-Sample T-Test

Hypothesis
H_0: The population mean is equal to a specified value ($\mu = \mu_0$)
H_A: The population mean is not equal to the specified value ($\mu \neq \mu_0$)

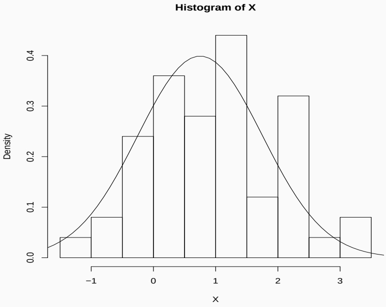

Figure 1: Here, we are interested in whether $\mu = 1$ or not?

T-Distribution Definition

The Student's t-distribution is a probability distribution used for making statistical inferences about the mean of a small sample from a normally distributed population. It is employed when the population standard deviation is unknown. The probability density function (PDF) of the t-distribution is given by:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Where:

- t is the random variable representing the standardized sample mean.
- ν is the degrees of freedom, which measures the sample size. As ν increases, the t-distribution approaches the standard normal distribution.

T-Distribution Composition

The Student's t-distribution is derived from the combination of the standard normal distribution (z-distribution) and the chi-squared distribution: Let Z be a random variable following the standard normal distribution, and let X^2 be a random variable following the chi-squared distribution with ν degrees of freedom. Then, the t-distributed random variable T is defined as

$$T = \frac{Z}{\sqrt{X^2/\nu}}$$

Where:

- Z follows the standard normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.
- X^2 follows the chi-squared distribution with ν degrees of freedom.
- ν is the degrees of freedom parameter in the t-distribution.

Deriving the t-Distribution

The formula $t = \frac{\bar{X}-\mu}{S/\sqrt{n-1}}$ in the t-distribution is derived from the properties of the sample mean ($\bar{X}$) and sample standard deviation (S). From the properties:

- Given that $\bar{X}$ is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$:
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$
- And nS^2 follows a chi-squared distribution with $n-1$ degrees of freedom, scaled by σ^2:
$$nS^2/\sigma^2 \sim \chi^2(n-1)$$
- $\bar{X}$ and S^2 are independent from each other.

We can use these properties to show that $\frac{\bar{X}-\mu}{S/\sqrt{n-1}}$ follows a t-distribution with $n-1$ degrees of freedom.

Example of one sample t-test

Listing 1: R code (Setting)

```
mu=1


n=50
X=rnorm(n,mu,1)
hist(X,prob=TRUE)
points(seq(-2,5,by=0.1),dnorm(seq(-2,5,by=0.1),
0.75,1),type="l")

t.test(X,mu=1,alternative = "two.sided")
t.test(X,mu=1,alternative = "less")
t.test(X,mu=1,alternative = "greater")
```

Recall "Confidence interval"

Example of one sample t-test (association of z-test)

Listing 2: R code (Setting)

```
n=c(1,3,10,20,30,100)
x=seq(-2,2,by=0.01)
plot(x,dnorm(x),type="l",ylab="f(x)")
for(i in 1:6){
points(x,dt(x,df=n[i]),type="l",col=1+i,lwd=1.75)
}
legend(-2, 0.4, c(paste("df=", c(2:7))),
 lty=1, col=2:7,lwd=2)
```

Two-Sample T-Test

Definition
The two-sample t-test compares the means of two independent samples to determine if they are significantly different.

Assumptions

1. Random sampling from the populations.
2. Normally distributed populations or large sample sizes.
3. Independence of observations between the two samples.

Two-Sample T-Test: Null Hypothesis

The null hypothesis (H_0) for a two-sample t-test states that the population means of the two samples being compared are equal.

$$H_0 : \mu_1 = \mu_2$$

where μ_1 represents the population mean of the first sample and μ_2 represents the population mean of the second sample.

The alternative hypothesis (H_A) is that there is a significant difference between the means of the two populations.

$$H_A : \mu_1 \neq \mu_2$$

Two-Sample T-Test Formula

- The formula for the two-sample t-test is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- $\bar{x}_1$ and $\bar{x}_2$ are the sample means, s_1 and s_2 are the sample standard deviations, n_1 and n_2 are the sample sizes.

Paired T-Test

Definition
The paired t-test compares the means of two related samples to determine if there is a significant difference between them.

Assumptions

1. Random sampling from the population.
2. Normally distributed population or large sample size.
3. Dependent or paired observations between the two samples.

Paired T-Test Formula

- The formula for the paired t-test is:

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

- $\bar{d}$ is the mean difference between paired observations, s_d is the standard deviation of the differences, and n is the number of paired observations.

Interpreting Results

- Calculate the test statistic (t-value) using the appropriate formula.
- Determine the degrees of freedom (df) based on the sample sizes and study design.
- Compare the obtained t-value with the critical t-value from the t-distribution table (or use software).
- If the obtained t-value exceeds the critical t-value (at a chosen significance level), the difference in means is considered statistically significant.
- Report the p-value to quantify the level of significance.

Conclusion

- The t-test is a powerful statistical test to compare means in different scenarios.
- The one-sample t-test compares a sample mean to a known population mean.
- The two-sample t-test compares the means of two independent samples.
- The paired t-test compares the means of paired observations.
- Assumptions regarding normality, independence, and dependence should be checked.