Testing of Hypotheses

Dr. Vinay Kulkarni

Background

- Random variables
  - Their measurements show variations
  - For no reasons at all
- We are interested in
  - **Expected Values** of random variables
  - But, we may get value “off” from the expected value
  - In such situations, how to decide if the value is:
    - Within the expected / permissible deviation from the mean?
- Such situations are encountered
  - While taking many decisions

Example

- Based on a study, it has been established that:
  - On an average, blood platelet count lower than 75000 is indicative of a certain disease
- When a patient reports a blood count of 60000
  - Should the doctor start treating the patient?

Example

- Based on a study, it has been established that:
  - On an average, blood platelet count lower than 75000 is indicative of a certain disease
- When a patient reports a blood count of 60000
  - Should the doctor start treating the patient?
- Questions:
  - Is this value within expected variations?
  - Or is it not?
  - Are there any statistical tests that can help decide?

Statistical Hypothesis

- Statistical Hypothesis
  - An assumption or a statement
  - About one or two parameters
  - Involving one or more populations
  - May or may not be true
- Testing of Hypotheses
  - Based on data samples
  - Decide whether the hypothesis is true / false
- Example:
  - Hypothesis: The patient is suffering from the disease
  - Hypothesis: platelet count < 75000

The method of statistical hypotheses

- First of all we assume some hypothesis is correct
  - (Not necessarily the one we believe to be true!)
  - Example: The patient is not suffering from the disease
    - Hypothesis: Platelet count is actually >= 75000
    - This is known as the **NULL hypothesis**: H_0
    - We have to prove that the platelet count is – beyond doubt – not a random variation of the real value.
- We then carry out a test
  - The goal is to check if the result of the test is beyond the limits of believability
  - If it is, we have to reject our hypothesis.

Fundamental Concepts: Hypotheses Testing

- **The Null and alternate hypothesis**
  - H_0 = Null hypothesis
  - H_1 = Alternate hypothesis
- **Possible Decisions**
  - Reject the Null hypothesis
    - This is the Goal: to prove with high probability
  - Do not reject the Null hypothesis

- **The Test Statistic**
  - Numerical value of the test statistic leads us to make the decision
- **The Critical Region (CR) or the Rejection Region (RR)**
  - An interval determined by the selection of appropriate distributions
  - Determines the region related to the test statistic and used to decide the acceptance / rejection of hypotheses
- **Conclusion and interpretation**

The NULL hypothesis: H_0

- We choose the **NULL hypothesis H_0** to be specific enough and simple enough that we can actually compute the likelihood of any given outcome of our observations
- NULL hypothesis is something that the data is likely to reject
  - Example:
    - Assume : Average mean temperature is 98
    - A sample measurement : 99
      - Data indicates: Observed temperature is not normal
      - Hypothesis H_0 should be “Observed temperature is Normal”

The “alternate” hypothesis: H_1

- The alternate hypothesis is something that we keep in mind, and it is something that we would like to accept in case the null hypothesis gets rejected.
- The alternate hypothesis is usually something that the data will support.

- In our example:
  - H_1 : “Temperature is Not normal”

Stating the Hypothesis

Stating H_0 and H_1

1. Two tailed test
   - $H_0 : P = P_0$ versus $H_1 : P \neq P_0$

2. Left tailed test
   - $H_0 : P \geq P_0$ versus $H_1 : P < P_0$

3. Right tailed test
   - $H_0 : P \leq P_0$ versus $H_1 : P > P_0$

Outcome of Hypothesis Testing

- NULL hypothesis is rejected & Alternate hypothesis is accepted

OR

- NULL hypothesis is accepted & Alternate hypothesis is rejected

Type I Error

- Consider the following situation:
  - We reject the null hypothesis
  - However, in reality, the null hypothesis is indeed true
  - Therefore, we have rejected the null hypothesis, when, in reality it is true
  - This is called a **TYPE I ERROR**
- **TYPE I** errors ➔ <u>False positives</u>
- Probability of Type I error = α (the confidence level)
  - Also known as "<u>level of significance</u>"
- Example: A healthy person is considered sick …

Type I Error

$$\begin{aligned}\alpha &= P\,(\text{committing a type I error}),\\ &= P\,(\text{rejecting } H_0 \text{ when } H_0 \text{ is true}),\\ &= P\,(\text{rejecting } H_0 \text{ when } H_1 \text{ is false}).\end{aligned}$$

Most common values for α

- 0.01, 0.05, 0.1
- Corresponding with 99%, 95% and 90% confidence levels

Type II Error

- Type II Error
  - Failing to reject the NULL hypothesis even when in reality it is false
- Probability of Type II Error: β
  - It is very difficult to calculate this probability
  - It depends on a variety of unknown parameters

$$
\begin{aligned}
\beta &= \mathrm{P}\,(\text{committing a type II error}),\\
&= \mathrm{P}\,(\text{not rejecting } \mathrm{H}_0 \text{ when } \mathrm{H}_0 \text{ is false}),\\
&= \mathrm{P}\,(\text{not rejecting } \mathrm{H}_0 \text{ when } \mathrm{H}_1 \text{ is true}).
\end{aligned}
$$

Type I and Type II errors

- We would like to reduce both types of errors
- However, <u>when we reduce probability of Type I error we increase the probability of Type II error</u>
  - By making the test more stringent, we reduce the risk of falsely rejecting the null hypothesis BUT increase the risk of failing to reject it when we should
- That's why the de facto value of 0.05 is so popular
- Note:
  - Selecting a larger sample size minimizes both types of errors

Size and Power of a test

- The *size* of a test is given by

  α

- Power of a test is given by

  $(1 - \beta)$

- Ideal we want our test to have:
  - Low size AND
  - High power
- The practice of computing the *power* of a test is known as
  - **Power Analysis**

Hypothesis Testing: Types

- About one parameter
  - One proportion
  - One mean
  - One standard deviation
- About two parameters
  - Two proportions
  - Two means
  - Two standard deviations

Steps in Hypotheses Testing

Classical Method

1. Determine and state H_0 and H_1
2. Decide the significance level α and the critical region
3. Based on the parameter, choose the <u>test statistic</u>
4. Using available data compute the test statistic
5. Make the statistical **Accept** or **Reject** decision based on
   a) Computed value of the test statistic
   b) The critical region identified in step 2

Hypothesis Testing about one proportion

- Characteristics of the proportion
  - Best estimate $\hat{p} = x/n$
  - Standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
  - Conditions
    - The sample is a simple random sample
    - Sample values are independent of each other
    - $np(1-p) \geq 10$

Confidence levels and Critical Regions

- The most used values for α are
  - 0.01, 0.05, 0.1
  - Distribution relevant to proportions

- Test statistic for proportions

$$Z = \frac{\hat{p} - p_0}{\sqrt{[p_0(1 - p_0)/n]}} = \frac{x - np_0}{\sqrt{np_0(1 - p_0)}}$$

Example / Exercise

- Known from past surveys
  - 35% of country's citizens invest abroad
- Current Survey
  - 800 adults were surveyed
  - 320 were found to hold foreign assets
- Government wants to know
  - If the foreign investment is still > 35%
  - With 10% significance level

Solution (Classic Method)

- $H_0 : P <= 0.35$ versus $H_1 : P > 0.35$
  - Right tailed test
- Significance level 10% = 0.1
- Z value associated with 0.1 = 1.28 (Normal Dist)
- Test to be done is as follows:
  - Since it is right tailed test
  - Hypothesis H_0 can be accepted if calculated Z value is <= 1.28 (rejection region Z > 1.28)
- Calculated Z value
  - $Z = \frac{\hat{p} - p_0}{\sqrt{[p_0(1-p_0)/n]}} = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$ = 2.965
  - Since Zcal > 1.28, Hypothesis is rejected
- Result: Foreign investment still exceeds 35%

The *p-value*

- The computed probability of getting the observed result, or any result at least as extreme in its difference from what the null hypothesis would imply – is called the *p-value*
- A *p-value* of 0.05 is the de facto standard cut-off between significant and non-significant results
- If this de facto value is used as the critical value, it will result in wrong results 5% of the time

Steps in Hypotheses Testing: *p-value method*

p-value method

1. Determine and state H_0 and H_1
2. Decide the significance level α
3. Based on the parameter, choose the test statistic
4. Using available data compute the test statistic and the p-value
   - How to calculate p-value?
5. Make the statistical **Accept** or **Reject** decision based on
   - α and p-value
     a) p-value less than α should reject H_0
     b) p-value greater than α should not reject H_0

Solution: p-value method

- For two tailed tests
  - P-value = 2 * $P(Z < Z_{cal})$
- For Left-tailed tests
  - P-value = $P(Z < Z_{cal})$
- For Right-tailed tests
  - P-value = $P(Z > Z_{cal})$

- In all these cases, if the computed p-value is < significance level α the hypothesis is rejected
  - Else, hypothesis is not rejected

Solution Based on p-value

- H_0 : P <= 0.35 versus H_1 : P > 0.35
  - Right tailed test
- Significance level 10% = 0.1
- Calculated Z value
  - $Z = \dfrac{\hat{p} - p_0}{\sqrt{[p_0(1-p_0)/n]}} = \dfrac{x - np_0}{\sqrt{np_0(1-p_0)}}$ = 2.965
  - From the Normal tables 2.965 corresponds to 0.99848
- P(Z > Zcal) = 1 – 0.99850 = 0.0015
- P(Z > Zcal) = 0.0015 < significance level 0.1
  - Therefore hypothesis P <= 0.35 is rejected
- Result: Foreign investment still exceeds 35%

Hypothesis testing : Single Parameter

- The Mean, when variance is known
  - Test statistic $$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$
- The Mean, when variance is unknown
  - Large sample size $$Z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$ (normal dist)
  - Small sample size $$T = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$ (t-distribution)
- The Variance
  - Test statistic $$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$ (chi-squared)
- The Proportion
  - Test statistic $$Z = \frac{\hat{p} - p_0}{\sqrt{[p_0(1-p_0)/n]}} = \frac{x - np_0}{\sqrt{np_0(1-p_0)}}$$

$\hat{p} = x/n;$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Exercise: 1

- In a medical test, rats were subjected to unit dose of a drug and recording their response times. Based on prior studies, it is known that the rats with no drugs given have a mean response time of 1.2 seconds. The new studies, after drugs, involving 100 rats throw up a mean response time of 1.05 seconds with a standard deviation of 0.5 seconds. Can we conclude that the drug has any effect on the response time? (Level of significance 0.01)

Solution: 1

- mu = 1.2
- xbar = 1.05
- n = 100
- s = 0.5s

- Data indicates: reduction in time
- H_0 : No change in mean time
- H_1 : Change in mean time

- Since we are testing equality, test is two tailed
- Since n > 30, normal distribution assumed
- At 0.01 significance levels, the z limits are
  - 0.005 and 0.995
- $z_{0.005} = -2.576$ and $z_{0.995} = 2.576$
- Calculated z = (xbar – mu)/(s/sqrt(n)) = -3
- Since -3 is less than -2.576, it lies in REJECTION region.
- Therefore H_0 is REJECTED and H_1 is accepted
  - Mean response time has changed from 1.2 seconds with the introduction of the drug
- p-value method
  - Calculated p-value: pnorm(-3) = -0.00135
  - Since |-0.00135| < 0.01 (significance) p-value is less than significance. Therefore, H_0 is rejected & H_1 accepted

Problem 2

- A major car manufacturer wants to test a new engine to determine whether it meets new air-pollution standards.
- The mean emission m of all engines must be less than 20 parts per million of carbon.
- Ten engines are manufactured for testing purposes and the emission level of each is determined to be: 15.6 16.2 22.2 20.5 16.4 19.4 16.6 17.9 12.7 13.9.
- Does the data supply sufficient evidence to allow the manufacturer to conclude this type of engine meet the pollution standard?
  - Test the hypothesis at a level a = 0.01.

Solution: 2

- n = 10
- xbar = 17.14
- s = 2.9228
- alpha = 0.01

- Data indicates: emission < 20 ppm (xbar = 17.14)
- H_0 : emission >= 20 ppm
- H_1 : emission < 20 ppm

- Since we are testing inequality (>=), the test is left tailed
- Since n < 30, t-distribution assumed with dof = 10-1 = 9
- At 0.01 significance levels, the t limit is:
  - $t_{0.01}$ = qt(0.01) = -2.82
- Calculated t = (17.14 - 20)/(2.9228/sqrt(10)) = -3.09
- Since -3.09 is less than -2.82, it lies in REJECTION region.
- Therefore H_0 is REJECTED and H_1 is accepted
  - Emission is < 20 ppm
- p-value method
  - Calculated p-value: pt(-3.09) = 0.006467
  - Since 0.006467 < 0.01 (significance level) p-value is less than significance level. Therefore, H_0 is rejected & H_1 accepted

Problem 3

- The National Science foundation, in a survey of 2237 engineering graduate students who earned PhD degrees, found that 607 were US citizens; the majority (1630) of the PhD degrees were awarded to foreign nationals. Conduct a test to determine whether the true percentage of PhD degrees awarded to foreign nationals exceeds 50% at a level a = 0.01.

Solution: 3

- n = 2237
- pcap = 1630/2237 = 0.729
- p_0 = 0.5
- alpha = 0.01

- Data indicates: pcap > 0.5 (0.729)
- H_0 : Proportion of degree to foreigners <= 0.5
- H_1 : Proportion of degree to foreigners > 0.5

- Since we are testing inequality (<=), the test is right tailed
- Since n > 30, normal distribution assumed
- At 0.01 significance levels, the z limit (for right tailed region) is:
  - $z_{(1-0.01)} = z_{(1-0.01)}$ = qnorm(0.99) = 2.3263
- Calculated z = (0.729 – 0.5)/sqrt(0.729 * (1-0.729)/2237) = 24.37
- Since 24.37 is greater than 2.3263, it lies in REJECTION region.
- Therefore H_0 is REJECTED and H_1 is accepted
  - Therefore: Proportion of degree to foreigners > 0.5
- p-value method
  - Calculated p-value: 1-pnorm(24.37) which is approximately “0”
  - Since p-value is less than significance level, therefore, H_0 is rejected & H_1 accepted