ST517 Note Outline 6: Hypothesis Testing for a Mean

Lecture 6.1: Statistical Significance-Overview

Vitamin E may increase Risk of Prostate Cancer
(October 2011) A study published in the Journal of the American Medical Association this month shows that a daily vitamin E supplement may raise the risk of prostate cancer. In this study 8737 men were randomly assigned to take a daily supplement of vitamin E. another 8696 men were assigned to take a placebo. All the men in this study were over 50 years of age and had no signs of prostate cancer at the beginning of the study. The men took a daily supplement of 400 International Units of vitamin E. After three years it was found that a higher percentage of the men in the Vitamin E group had developed prostate cancer. The difference was 17% higher and the researchers found this difference was statistically significant.

Subjects:

Explanatory variable:

Response variable:

Type of study design:

Is it possible

- Some men would develop prostate cancer anyway

Statistical significance

Establishing statistical significance

- Hypothesis test

Basic steps to hypothesis testing

1. Establish null and alternative hypotheses
2. Find test statistic and null distribution
3. Make a decision about the hypotheses
4. State the conclusion in the context of the problem

Lecture 6.2: Basic Step to a Hypothesis Test

Step 1: Establish null and alternative hypothesis

Null Hypothesis: Beginning claim

- Allows establishment of sampling distribution, in this context called the:

  Notation:

Alternative Hypothesis: Another theory

  Notation:

Example: Vitamin E and Prostate cancer
Null hypothesis: Vitamin E does not make a difference in prostate cancer

Research Hypothesis: Vitamin E does make a difference in prostate cancer

Step 2: Find test statistic and null distribution

Test statistic: Numeric measure of distance from sample value to what is expected under null hypothesis

Null distribution: Distribution of the test statistic assuming

- Can be created by simulation or by theoretical model (i.e. a probability distribution)
- Sets frame of reference so we can see how the sample statistics would be expected to behave

Step 3: Make a decision about the hypotheses

<u>P-value</u>: Proportion of the null distribution that is as or more extreme than the test statistic

- Notes:

Basic idea:

- If the p-value is too small, then the statistic is unlikely given the null distribution. Perhaps something else is going on.

- How small is too small?

<u>Significance level</u>: Cutoff point for the p-value; indication of small

Notation:

Rule of thumb:

Statistical significance: Sample results not likely to have occurred just by random chance

Step 4: State the conclusion in the context of the problem

- Just telling someone "We reject the null hypothesis" or "The results are statistically significant" does not give them useful information about the problem at hand
- Conclusion needs to give enough information in the context of the problem that the reader can make an informed decision
    - Needs to be about the alternative hypothesis, because this is the hypothesis driving the research (the one we are interested in learning about)

Conclusion:

- Careful that your language is not too strong

Lecture 6.3: Hypothesis Test for a Population Mean

1. Establish Null and Alternative Hypotheses

   Null hypothesis:

   Choice of alternatives:

2. Find test statistic and null distribution

   Test statistic: $t = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}}$

   Null distribution:

   Conditions required for this test to be valid:

3. Make a decision about the hypotheses

   Find p-value under null distribution, in the direction of alternative hypothesis

   Rule of thumb: p-value ≤ α, reject H_0

4. Make a conclusion in the context of the problem
   - If you rejected H_0: Conclude that there is enough evidence to support the alternative hypothesis.
   - If you did <u>not</u> reject H_0: Conclude that there is <u>not</u> enough evidence to support the alternative hypothesis.

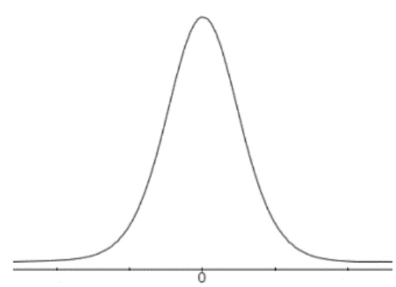
The t-distribution: Finding Probabilities

- P-values are probabilities under a null distribution; for tests of the mean, the appropriate null distribution is a t-distribution
- Technology that can be used to calculate the p-value for tests of the mean:
    - Graphing calculator (e.g. TI-84): `tcdf(lower_bound,upper_bound,df)`
    - Software, e.g.
        - SAS (see hypothesis test code posted to Moodle)
        - Excel: `t.dist(upper_bound,df,TRUE)`
    - Online calculators
        - E.g. stattrek.com/online-calculator/t-distribution.aspx
        - Fill in: *df* and t score; computer provides area to the left of that t score

Example: For the following situations find the appropriate p-value.

a. $H_0: \mu = 38, H_a: \mu < 38$ n = 13, t = -1.53

Interpretation of p-value:

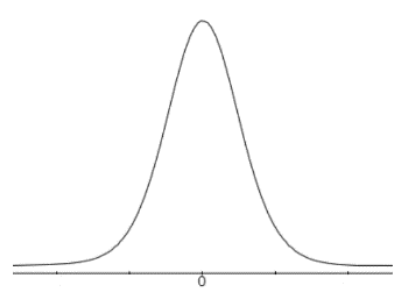


Using the technology for this example:

- Graphing calc: `tcdf(-1000,-1.53,12)`
- Excel: `t.dist(-1.53,12,TRUE)`

b. $H_0: \mu = 38, H_a: \mu > 38$ n = 13, t = 2.01: p-value =

Interpretation of p-value:



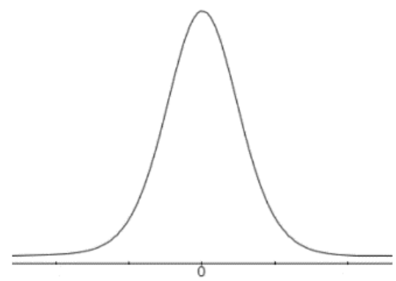
Using the technology for this example:

- Graphing calc: `tcdf(2.01,1000,12)`
- Excel: `1-t.dist(2.01,12,TRUE)`

Example (continued): For the following situations find the appropriate p-value.

c. $H_0: \mu = 38, H_a: \mu \neq 38$ n = 13, t = 2.01: p-value =

Interpretation of p-value:

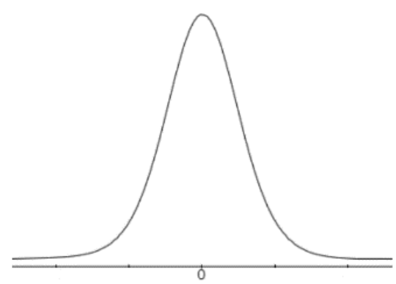


Using the technology for this example:

- Graphing calc: `2*tcdf(2.01,1000,12)`
- Excel: `2*t.dist(-2.01,12,TRUE)`

d. $H_0: \mu = 38, H_a: \mu > 38$ n = 13, t = -1.90: p-value =

Interpretation of p-value:



Using the technology for this example:

- Graphing calc: `tcdf(-1.90,1000,12)`
- Excel: `1-t.dist(-1.90,12,TRUE)`

Example: A candy company produces candy covered chocolates. The company specifications indicate that the candies should average 0.85 grams. A quality control manager for the company took a random sample of 38 candies from the production line. The resulting values were used to produce this output. Does this data indicate that the average weight has changed from the specification? Conduct an appropriate hypothesis test. Use a significance level of 5%.

Summary statistics:

Column	Mean	Std. Dev.	Min	Max	Q1	Q3	n	Median
weight	0.873	0.034	0.81	0.98	0.85	0.89	38	0.87

Lecture 6.4: Important Notes about Hypothesis Tests

Note: Statistical tests are based on randomness and assumptions.

Note: Statistical significance is not practical significance.

Note: Having more subjects in a study makes it more likely you will find statistically significant results.

Note: We cannot prove the null hypothesis.

Note: We can use a confidence interval to conduct a hypothesis test.

Example: A cookie company manufactures large sugar cookies that are then sold to a chain of coffee shops. The cookies are made by a machine that is set to produce a cookies that have an average weigh 75 grams. The company would like to know if this machine is working correctly and producing cookies that average 75 grams. They take a random sample of 30 cookies and used their weights to produce this output.

95% confidence interval results:
μ: Mean of variable

Variable	Sample Mean	Std. Err.	DF	L. Limit	U. Limit
Oatmeal	75.26	1.63	29	71.93	78.60

Based on this output would we reject the null hypothesis that $\mu = 75$ at $\alpha = 0.05$? Explain.

Lecture 6.5: Errors and Statistical Power

Errors that can occur

- What if we make the wrong decision?

- Type I error

- Significance Level (α)

- Type II error

	H_0 Actually True	H_0 Actually False
Say "Reject H_0"		
Say "Do not reject H_0"		

Power of a statistical test

- Power

Example: Malaria is a serious health concern. Early diagnosis and treatment are important tools in the fight to control this disease. We can think of a test for malaria as a hypothesis test with the following hypotheses:

Ho: patient does not have malaria   vs.   Ha: patient has malaria

If a patient tests positive for malaria, the therapy recommended by the World Health Organization is artemisinin-combination therapy or ACT, which has a known side effect of vomiting. If a patient exhibits symptoms of malaria but tests negative for the disease, this may result in the patient being treated for a different condition. Untreated malaria can lead to more severe conditions and, in most cases, death.

a. Define Type I and Type II Errors in the context of this problem.

b. Describe a real-world consequence of making each type of error in this context.

Notes:

- In practice, we want to protect the "status quo"

- Most tests have the

- For a fixed sample size *n*, there is a

- Ideally want probabilities of making a mistake to be small and power to be large
    - These probabilities are properties of the procedure

    - They are not applicable to the decision once it is made

Increasing Power

- Power increases if we increase the sample size.

- Power increases if we increase the significance level

- Power increases if there is a bigger difference to find

Calculating Power or the Probability of Making an Error

- Finding power requires setting several other values

- Recall: Power = 1 – P(Type II Error) = 1 – β
- There are several formulas for calculating β
- However, for this course we will focus on understanding rather than calculation
  - Reason through finding probabilities based on the context of the problem
  - Or use computer software

Example: There is a basket with 10 marbles inside. The marbles are either red or white, but we do not know the number of each color. We want to decide between the hypotheses:

H_0: Basket has 9 Red and 1 White vs. H_A: Basket has 4 Red and 6 White

We will select a single marble from the basket. What is the most reasonable Decision Rule?

- Reject the null hypothesis if the ball is

With this rule, what are the chances of making a mistake?

- P(Type I error) =

- P(Type II error) =

- What is the power of the test?

Example (continued): Suppose a ball is now selected from the basket and it is observed and found to be white.

- What would be the decision?

- Could a mistake have been made?

- If so, which type?

- What is the probability that this type of mistake was made?

Determining Sample Size

- In practice, issues of error and power should be considered before a study is conducted
- Limit type I error by setting the significance level the test will be conducted at

- Set an acceptable level for power (and type II error)

- Determine minimum sample size that would be needed to achieve these values

    - Again, depends on setting several other values (e.g. significance level, believed value of the mean under the alternative, believed value of σ)

Example—SAS Code:

```
proc power;
   onesamplemeans        Requests test for a single mean;
   nullmean = 2          Set null value (default = 0);
   mean = 5 10 15        Set best guess(es) for value of mean if Ha true;
   stddev = 30 50        Set best guess(es) for value of popul. std.dev;
   alpha = 0.05          Set the significance level (default = 0.05);
   power = 0.80          Set the desired power level;
   sides = 1             1-sided test (defaults to a 2-sided test);
   ntotal = .;           Requests smallest sample size needed to achieve
                         desired power under conditions specified
run;
```

Corresponding SAS Output:

The POWER Procedure
One-sample t Test for Mean

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Number of Sides	1
Null Mean	2
Alpha	0.05
Nominal Power	0.8

Computed N Total

Index	Mean	Std Dev	Actual Power	N Total
1	5	30	0.800	620
2	5	50	0.800	1719
3	10	30	0.803	89
4	10	50	0.800	243
5	15	30	0.807	35
6	15	50	0.801	93

Power analysis: Determining predicted power

- Alternatively, you could determine highest power possible for a given sample size (and other conditions specified as before)

Example—SAS Code:

```
proc power;
   onesamplemeans
   nullmean = 2
   mean = 10
   stddev = 30
   alpha = 0.05
   power = .
   sides = 1
   ntotal = 62;
run;
```

Same options as before…

Now requests achieved power under conditions specified

Corresponding SAS Output:

The POWER Procedure
One-sample t Test for Mean

Fixed Scenario Elements	
Distribution	Normal
Method	Exact
Number of Sides	1
Null Mean	2
Alpha	0.05
Mean	10
Standard Deviation	30
Total Sample Size	62

Computed Power
Power
0.667