**Large Sample tests of hypothesis**

Main points in this chapter

1. Standard method to test research questions.
2. Discussion - risks involved when decision based on the test is incorrect.
3. Detailed discussion of the standard method.
4. Application of standard method for research questions using large samples.

**Recall: chem lab or your biology lab in high school and the Scientific Method**

**Observations:** A good scientist is observant and notices thing in the world around him/herself. (S)he sees, hears, or in some other way notices whats going on in the world, becomes curious about whats happening, and raises a question about it.

**Hypothesis:** This is a tentative answer to the question: an explanation for what was observed. The scientist tries to explain what caused what was observed (hypo = under, beneath; thesis = an arranging).

**Testing:** 1. Hypotheses are possible causes. An hypothesis is not an observation, rather, a tentative explanation for the observation.

2. Hypotheses reflect past experience with similar questions (educated propositions about cause).
3. Multiple hypotheses should be proposed whenever possible. One should think of alternative causes that could explain the observation (the correct one may not even be one that was thought of!)
4. Hypotheses should be testable.
5. Hypotheses can be proven wrong/incorrect, but can never be proven or confirmed with absolute certainty. Someone in the future with more knowledge may find a case where the hypothesis is not true.

**Statistical method**

1. Observe the economy, raise a question or set of questions.
2. Prepare an answer in the form of a hypothesis ($H_0$) (also known as the null hypothesis).
3. Prepare counter responses ($H_a$ - alternative) if the null is proven wrong or incorrect.
4. Collect data.
5. Specify a statistical test.

6. Determining the critical regions to reject.
7. Obtain the findings, prepare the results.

**My modest example using Spam and eggs**

Implications from theory - what the theory predicts with the respect the differences in the proportion of income spent on a good.

$$\gamma = \frac{\text{Income spent on good i}}{\text{Total income}}$$

Let's look at the value of $\gamma$ spent by high income $\gamma_{High}$ and the the value of $\gamma$ spent by low income $\gamma_{Low}$

$\gamma_{Low} - \gamma_{High} < 0 \ \rightarrow$ Luxury good
$\gamma_{Low} - \gamma_{High} = 0 \ \rightarrow$ Normal good
$\gamma_{Low} - \gamma_{High} > 0 \ \rightarrow$ Inferior good

**Test #1 - is Spam a luxury good?**

1) Hypothesis to reject: $\rightarrow \gamma_{Low} - \gamma_{High} \geq 0$

I wish to reject the notion that the good could be either a normal or inferior good. Logically, rejecting this hypothesis implies that I fail to reject that it is a luxury good.

**Failing to reject $\neq$ Accepting an outcome**

Accepting an outcome implies that you have accepted the theory.

**Full bank of tests for "Is Spam a luxury good?"**

1a) Null (Hypothesis to reject): $\rightarrow \gamma_{Low} - \gamma_{High} \geq 0$
1b) Alternative hypothesis: $\rightarrow \gamma_{Low} - \gamma_{High} < 0$
2a) Null (Hypothesis to reject): $\rightarrow \gamma_{Low} - \gamma_{High} = 0$
2b) Alternative hypothesis: $\rightarrow \gamma_{Low} - \gamma_{High} \neq 0$
3a) Null (Hypothesis to reject): $\rightarrow \gamma_{Low} - \gamma_{High} \leq 0$
3b) Alternative hypothesis: $\rightarrow \gamma_{Low} - \gamma_{High} > 0$

The researcher will want to reject Tests #1 and #2 and fail to reject #3.

**Further comment regarding the three tests**

Two are one tailed tests, one is a two tailed test.

**Specifics: Making Assumptions**

1. Type variable (categorical, interval)
2. Type of population (binomial, normal, normal given the Central Limit Theorem)
3. Type of analysis (univariate, bivariate)
4. Large or small sample
5. Differences in variances - has an impact on differences in means test using a small sample.
6. Null hypothesis ($H_0$), and alternative hypothesis ($H_a$) - one or two tailed test.

## Specifics: Sampling Distribution

1. Standard normal (z scores).
2. Student's t distribution.
3. $\chi^2$ distribution.
4. F distribution.

## Specifics: One tailed or two tailed

If the hypothesis is an inequality (e.g., $\mu > 0, \mu < 1$), we can use a one tail test. If we are testing if $\mu$ is a specific value, the alternative hypothesis is that $\mu$ is not this value and can be any value in the distribution. For this case, we use a two tail test.

## Specifics: Choosing a critical region

Describes rejection area. Answers the questions: what are we willing to risk in being wrong?

**Three scenarios-two tailed test**

**Scenario #1 $\alpha = 20\%$ $\alpha/2 = 10\%$ (or p value = 0.10)** - means that when we reject the hypothesis, we reject it with a confidence level of 80%.

**Scenario #2 $\alpha = 10\%$ $\alpha/2 = 5\%$ (or p value = 0.05)** - means that when we reject the hypothesis, we reject it with a confidence level of 90%.

**Scenario #3 $\alpha = 2\%$ $\alpha/2 = 1\%$ (or p value = 0.01)** - means that when we reject the hypothesis, we reject it with a confidence level of 98%.

**Notion of Significance - table on page 346**

**Do researchers only report results that are significant?**

**Risk**

There are two types.

1. **Type I error** rejecting a hypothesis when in fact, it is true.
2. **Type II error** failing to reject a hypothesis when one should reject it.

**Probability of making a Type I error.**

Significance level (p value) tells you the probability of making a Type I error.

**Amount of risk for the three scenarios**

Highest risk of making a Type I error - lowest confidence level taken.

Lowest risk of making a Type I error - highest confidence level taken.

**Probability of making a Type II error.**

The probability of a Type II error is $\beta$.

Power of your statistical test is given as (1-$\beta$).

**Computing $\beta$ and (1-$\beta$)**

Suppose your hypothesis test is that

$$\text{HO: } \mu_0 = A$$

You want to compute a power test to determine the probability of rejecting HO when the alternative mean $\mu_a$ = C.

1. Compute the two confidence interval values. The book uses the margin of error values but the example (9.8) assumes that the significance level is 5%. My instructions apply for all significance levels. These values are the endpoints of the Type II region. The formula for the confidence interval is

   $$\mu_0 \pm z_{\frac{\alpha}{2}} SE$$

   Note, function uses $\mu_0$.

   From this you have the left boundary value (LBV) and right boundary value (RBV).

2. Draw the two graphs. The left and right boundary points are points around $\mu_0$. Determine where $\mu_a$ is relative to these boundaries and determine the rejection area of the new distribution overlapping the acceptance region of the old distribution.

3. Compute z scores for the two values using the following functions.

$$z(\text{left boundary}) = \frac{LBV - \mu_a}{SE}$$

$$z(\text{right boundary}) = \frac{RBV - \mu_a}{SE}$$

Note: this set of functions uses $\mu_a$.

4. Given the drawing above, determine the p(accepting $h_a$ when $\mu = \mu_a$).

The power of the test or the probability of correctly rejecting $H_0$ given that $\mu$ is $\mu_a$ is (1-$\beta$).

## Relationships between Type I & II probabilities and power test

1. Increasing the significance level reduces the confidence interval thus increasing the probability of a Type II error and reducing the power of the test.

2. Increasing the sample size decreases the standard error. This decreases the probability of a Type II error and increases the power of the test.

3. If $\mu_a$ is very close to $\mu_0$, it weakens the power test.