Class 14: Hypothesis Tests for Means: Using Z-Values

Statistical Inference

We take a sample to learn about a population.

- Last time we made estimates using **confidence intervals.**
- Now we do **Hypothesis Testing**: We use the information from a sample to make a "yes/no" decision.
  For example, *clinical drug trials*: are used to decide "Does the Drug Work?"

Example: Testing a Vaccine for Malaria

Malaria, carried by mosquitoes, kills about 1 million people a year, many of them children under five.[1] Ninety percent of the cases occur in Africa.[2] There is currently no vaccine against malaria, only preventative drugs, such as quinine, which must be taken continuously to work. Malaria cannot be cured; people who have the disease face recurrences throughout their lives.

In 2004, experiments were done in Mozambique, Africa, on a new drug,[3] perhaps available soon. The drug does not prevent malaria, but it does seem to reduce the severity of the disease. The experiment measured the effect of the drug on:

- The number of children who were infected
- The length of time until infection

Ex: Which variable is quantitative and which is categorical?
Number getting infected is Categorical; Length of time is Quantitative.

Ex: Does the drug extend the length of time till infection? Before the drug trial, the average length of time before a child in the population got malaria was 89 days with standard deviation 41 days. During the experiment, 724 children were treated with the drug and their average length of time until infection of was 97.5 days.

Ex: What is the population and what is the sample? Population is all children in Mozambique who could get malaria. The sample is the 724 children who were treated with the drug.

***Strategy*: The question is whether the difference between 89 and 97.5 days is large enough to be significant. Or could it be just the result of random variation?**

We start by assuming that the drug does ***not*** work; this assumption is called the **null hypothesis**. We see what the average times till infection we are predicted to be likely for a sample of 724 and compare with sample data.

Null hypothesis: H_0: Mean length of time till infection in treated population is still 89 days.

$$H_0: \mu = 89.$$

Alternate hypothesis: **H_a**: Mean time till infection in treated population is more than 89 days.

$H_a: \mu > 89.$

---

[1] For comparison, there were 2.1 million AIDS deaths worldwide in 2007. http://www.un.org/ecosocdev/geninfo/afrec/vol21no4/214-aids-declining.html

[2] "Uganda's misplaced health millions" Daniel Halperin, Harvard, http://news.bbc.co.uk/2/hi/africa/8275713.stm and "Putting a plague in perspective" http://www.nytimes.com/2008/01/01/opinion/01halperin.html?_r=1

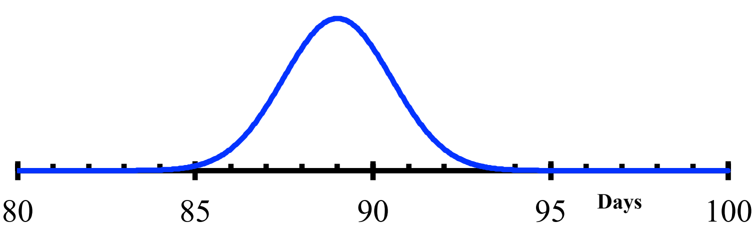
[3] "Efficacy of the RTS,S/AS202A vaccine against *Plasmodium falciparum* infection and disease in young African children: randomized controlled trial" by P. Alonso et al, *The Lancet*, Oct 16, 2004.

What does the null hypothesis assumption tell us about what lengths of time to infection are likely?
If the null hypothesis is true, the mean time to infection of the sample of 724 infected children will vary as predicted by the Central Limit Theorem. That is: the mean lengths of time will be

- approximately normally distributed,
- with mean 89 and
- standard error $41 / \sqrt{724} = 1.52$ days.

<u>Compare the Sample Statistic with the Sampling Distribution</u>
The average length of time to get sick is the mean of the sample is 97.5 days; this is called the **sample statistic**. The picture shows the sampling distribution of the means of samples of size 724. Does 97.5 look likely? No; 97.5 looks unlikely.

Distribution of sample means:
Mean = 89, Std Dev =1.52



<u>Calculate the probability of seeing the sample statistic: p-value</u>
Assuming the null hypothesis is true, we calculate the probability of seeing a sample with mean as far from 89 as 97.5.
We first find the z-value:

$$z = \frac{97.5 - 89}{1.52} = 5.6.$$

The z-value is big, so the p-value is very small. From the table, we see it is less than the smallest value there (0.03%); with Excel or a calculator you find it is about $1 \cdot 10^{-8}$.

This is probability is called the ***p*-value**. It is the probability of seeing an average length of time to get sick as high as 97.5 days if the drug does not work.

<u>Conclusion:</u>
The p-value tells us that if the null hypothesis is true, the chance of seeing a sample as extreme, or more extreme, as the one observed, is incredibly small. So there are two possibilities:

- We just saw an incredibly rare event, or
- The null hypothesis is not true

Which do we believe?
A very small p-value is evidence against the null hypothesis, so we conclude that it is probably not true. *We* ***reject*** *the null hypothesis and accept the alternate hypotheses. We conclude that the drug lengthens the time to infection.*

How small a *p*-value is needed to reject the null hypothesis?
This is a matter of choice; usually reject if p-value is less than 5%. This 5% cut off is called the **significance level**.

Steps for Hypothesis Test

Step 1: Choose null and alternate hypotheses and significance level
Step 2: Construct test statistic from the sample, calculate the *z*-value assuming null hypothesis
Step 3: Calculate *p*-value
Step 4: Make a Decision: Reject or not to Reject

Note: When choosing the null hypothesis, it must contain an equal sign (because the sampling distribution is based on it). The alternate hypothesis is generally what someone doing the experiment wants to show is true.

For the null hypothesis H_0: Mean of population $= \mu$, we use

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Ex: The malaria drug example

Step 1: Null hypothesis: Mean time to infection of all treated children is 89 days. Written $\mathrm{H}_0: \mu = 89$.
Alternate hypothesis: Mean time to infection of all treated children is greater than 89 days. Written $\mathrm{H}_a: \mu > 89$.

Step 2: Test statistic is 97.5 days. To find the *z*-value, find the mean (89 days) of the sampling distribution and its standard error $\left(41/\sqrt{724} = 1.52\right)$, assuming the null hypothesis is true. Then

$$z = \frac{97.5 - 89}{1.52} = 5.6.$$

Step 3: We found a very small *p*-value = $1 \cdot 10^{-8}$

Step 4: Because we found a small *p*-value (less than the significance level of 5%), we reject null hypothesis and conclude that the drug lengthens the time to infection.

Example: Testing Gas Mileage

A gasoline additive claims to increase the average mileage of a certain type of car from the usual 24 miles per gallon, with standard deviation 2.3 miles per gallon. A store owner wants to test this claim, and if there is convincing evidence that the mean is greater than 24, he will stock this additive. He finds that the mean mileage of a sample of 45 cars using the additive is 24.7 miles per gallon. Using a significance level of 0.05, decide whether the store should the store stock this additive.

Ex: What is the population and what is the sample? How is the mean gas mileage for each group written?
Population is all cars of this type using the additive. Mean is $\boldsymbol{\mu}$
Sample is the 24 cars using the additive. Mean is $\bar{x}$

Step 1:
Null Hypothesis: Mean mileage with additive is 24 mpg: H_0: $\boldsymbol{\mu = 24}$.
Alternate hypothesis: Mean mileage of cars with additive is greater than 24 mpg. : $\mathbf{H}_a: \boldsymbol{\mu > 24}$
Significance level: 5%

Step 2: The test statistic is 24.7 mph. Assuming the null hypothesis is true, the distribution of mean mpgs is normally distributed, with mean 24 mpg and standard deviation $\frac{2.3}{\sqrt{45}} = 0.3429$ mpg.
We have

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{24.7 - 24}{0.3429} = 2.04$$

Step 3: From the table or a calculator, the probability of a z-values above 2.04 is the p-value, which is $100 - 97.93\% = 2.07\%$.

Step 4: Since the p-value is small, we have either drawn a *very* unusual sample, or the null hypothesis is not true. Thus we have evidence that the mean is greater than 24 mpg, so we reject the null hypothesis and accept the alternate hypothesis. The store should stock the additive.

One-Sided and Two Sided Tests

In the examples we have done, we were interested in the probability that the length of time till infection was *more than* 89 days and the gas mileage was *more than* 24 mpg. What if we were worried that the drug might be harmful, or that the additive might make the gas mileage worse? In this case we do a **two-sided test** instead of the **one-sided test** we did before.

Ex: Malaria: Test whether the drug alters the length of time to infection.

Now we are asking whether the length of time to infection is different than before, not whether it is longer.

Step 1: Null hypothesis: This is the same as before: Mean time to infection of all treated children is 89 days. Written $H_0: \mu = 89$.
Alternate hypothesis: This is different. Mean time to infection of all treated children is different from 89 days. Written $H_a: \mu \neq 89$.

Step 2: This calculation is the same. Test statistic is 97.5 days. To find the z-value, find the mean (89 days) of the sampling distribution and its standard error $(41/\sqrt{724} = 1.52)$ days, assuming the null hypothesis is true. Then

$$z = \frac{97.5 - 89}{1.52} = 5.6.$$

Step 3: Here is whether the change comes: We must multiple the value from the table or calculator by 2, as we are looking for z-values at either end. Before we found a very small p-value $= 1 \cdot 10^{-8}$; it is now $= 2 \cdot 10^{-8}$, which is still very small.

Step 4: Because we found a small p-value (less than the significance level of 5%), we reject null hypothesis and accept the alternate hypothesis. We conclude that the drug alters the time to infection. In order to say whether it increases or decreases the time, we look at the data—which shows the time is increased.

Ex: Gas Additive: Test whether the additive alters the gas mileage.

Step 1: Null Hypothesis: Mean mileage with additive is 24 mpg: $H_0: \boldsymbol{\mu = 24}$.
Alternate hypothesis: Mean mileage of cars with additive is different from 24 mpg. : $\mathbf{H}_{\boldsymbol{a}}: \boldsymbol{\mu \neq 24}$

Step 2: The test statistic is 24.7 mph. Assuming the null hypothesis is true, the distribution of mean mpgs is normally distributed, with mean 24 mpg and standard deviation $\frac{2.3}{\sqrt{45}} = 0.3429$ mpg.
We have

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} = \frac{24.7 - 24}{0.3429} = 2.04$$

Step 3: In the previous example, the probability of a z-values above 2.04 was 2.07%.
Now the z-value is $2(2.07\%) = 4.14\%$.

Step 4: Since the *p*-value is again smaller than 5%, we reject the null hypothesis and accept the alternate hypothesis. The additive changes the gas mileage.

Other Significance Levels
Significance levels of 1% and 0.1% are also used. In the two sided examples, we conclude the malaria drug works at the 5% and 1% and 0.1% levels because the p-value is smaller than these significance levels. For the gas additive, the change in mpg is significant at the 5% level, but not at the 1% or 0.1% levels.