Ch. 26: Tests of Significance

Ex: Jim, Jane, and Peter have been playing a game of Monopoly for several hours....

Jim (shouting):

"The dice are not fair – they are biased. There are far more double sixes than there should be!"

Jane (worried):

"It seems you have studied a little bit too much of Stat 1040 – why should those dice not be fair? We just bought the game – the dice have not been used before and nobody had a chance to manipulate them!"

Jim (not convinced at all) takes a pen and paper and starts to record the outcome of the next 720 rolls....

After several hours ....

Jim (jumps up and shouts):

"Here it is – the dice are not fair! During the last 720 rolls, there have been 32 double sixes. We all know, the chance for a double six should be _____. Thus, the expected number of double sixes after 720 rolls is _____."

Jane (a bit sceptical):

"But that could just be due to chance error, couldn't it?"

Jim (convincingly):

"Maybe, but if the dice really are fair, there is only a small chance of getting 32 or more double sixes. Peter – can you help me and calculate that exact chance?"

Peter:

"Sure the chance is: ..."

Jane (now convinced):

"Hmmmm... That's indeed a pretty small chance. We better use another set of dice when we play Monopoly next time."

Conclusion:

In the previous game of Monopoly, there is only a very small chance that we will see this many double sixes in 720 rolls. This means:

- The players have been extremely lucky, or
- The dice are really biased: The true probability of double sixes is higher than 1/36.

Tests of Significance

The previous example shows the general pattern of ______________________________ or ______________________________.

We want to show that some observed difference is real (i.e., the difference is in the population, not just in the sample).

- A skeptical person sees a difference and says the difference is merely due to chance variation.
- A credulous person sees the difference and says there is a real difference.
- A statistically literate person will find out who is right, i.e., how likely an observed difference is with nothing to explain it *but* chance variation.

The Null and the Alternative

In significance testing, we compare two __________:

- A ________________ which says that the difference is due to chance variation.

- And an ________________, which says that the difference is real. This is usually what we wish to prove.

A null hypothesis must be formulated as a box model.

Note:
Hypotheses are about the numbers in a box, not in a sample.

Null and Alternative for Monopoly

Ex: In our Monopoly example, it is:

Null Hypothesis:

Alternative Hypothesis:

Test Statistic

Once we have our null and alternative hypothesis formulated as a box model, we can calculate a ________________.

A common test statistic for averages and percentages is the ________________:

$$z = \frac{\text{observed - expected}}{\text{SE}}.$$

A test using the z-score is called a ________________.

Note:

The test statistic z represents how many standard units an observed value is from its expected value, calculated from the null hypothesis.

z-score for Monopoly

Ex: The z-score for the Monopoly example is:

Note:

The z-score is not very meaningful itself. So we use the normal curve to convert z into an ____________________________, commonly called ______________ (for probability).

P-values

The P-value is the chance of getting a sample value at least as extreme as the observed one, given that the null hypothesis is true.

Ex: In our Monopoly example the P-value is:

Conclusion

If the P-value is small (less than 5%), there is strong evidence against the null hypothesis. This means, we ______________ the null hypothesis.

If the P-value is fairly large (greater than 5%), the evidence against the null hypothesis is weak. This means, we ______________________ the null hypothesis.

We do not say that we ______________ the null hypothesis, because we may just not have enough data to see that the null hypothesis is wrong.

Ex: In our Monopoly example, the conclusion is:

Statistical Significance

If the P-value is less than 5%, the result is
______________________________.

If the P-value is less than 1%, the result is
____________________________________.

Note:
The P-value is not the chance that the null hypothesis is true – it is the chance of seeing such a weird sample if the null hypothesis were true.

So, if the P-value is small, we think the null hypothesis is not true.

Four-Step Procedure for Hypothesis Testing

1. State the null and the alternative hypotheses in words and in terms of a box model.

2. Find the test statistic (i.e., standard units).

3. Calculate the P-value (area under the curve).

4. State conclusions (rejecting the null hypothesis or not rejecting, and in your own words).

Ex: Bottles of orange juice are supposed to have 16 fluid ounces. A random sample of 100 bottles from a large batch contains an average of 15.7 ounces with an SD of 0.2 ounces. Test the hypothesis that the bottles are being filled correctly, against the alternative that they are not full enough.

Ex: Researchers separately asked 153 husbands and wives to state the highest school grade completed by the wife. For each couple, they recorded

$$X = \text{husband's answer - wife's answer.}$$

X averaged 0.32, with an SD of 1.1. Does this suggest an average difference which is greater than 0?

t-tests

The t-test is used instead of the z-test when:

- the number of draws is small (less than 30), and
- the SD of the box is unknown (i.e., we must bootstrap), and
- the histogram for the tickets in the box is close to the normal curve.

The t-test is just like the z-test, but:

1. We use SD_+ instead of SD, where

$$SD_+ = \text{SD of measurements} \times \sqrt{\frac{\text{number measurements}}{\text{number measurements - 1}}}.$$

2. We still calculate our test statistic as

$$\frac{\text{Observed - Expected}}{\text{SE}},$$

but we do not use the normal table to find our P-value.

Instead, we use a ________________, using the line with ______________________ equal to the number of measurements - 1.

The Gauss Model for Measurement Error

If our measurement technique is unbiased,

individual measurement =

______________ + ______________

The _________________ says:

- Chance error is like a draw from a box called the "_________________"

- The error box has an average of zero provided the measurements are made with no bias.

Ex: Sugar packets are supposed to have a weight of 1500 grams. A student buys 5 packets and finds:

$$1473, 1489, 1525, 1585, 1513$$

Average:

SD:

Is there more sugar in the packets than it is supposed to be? Assume the Gauss model with no bias.

When to use the t-test?

Number of observation is

Histogram of the box is

SD of the box is