Lecture 11

Last time

Confidence Intervals

We can calculate a statistic from a sample but how can we make inferences about the population with it. Confidence intervals gives us a range of values around our statistic that guarantees our population parameter will be in that interval in a proportion of our samples

$$POINT\ ESTIMATE \pm MARGIN\ OF\ ERROR \tag{1}$$

We just need a random sample and a normal distribution of our statistic for the confidence interval to be valid.

We set our level of confidence, C, and find the corresponding Z^* which tells us how many standard deviations away from our statistic we need to go to create this confidence interval.

****Remember** The confidence level is NOT the probability of the population parameter being in the interval. We interpret the confidence level as in repeated sampling of size n from the same population, we expect the population parameter to be within C% of the confidence intervals computed from each sample.**

Testing Hypotheses

Example:
Astrologers believe that the positions of the planets and the moon at the moment of your birth determine your personality traits. But have you ever seen any scientific evidence that astrology works? One scientific test of astrology (from *Nature* vol.318, 1985) used the following experiment:

Each of 116 volunteers were asked to give their dates and times of birth. From this information, an astrologer prepared each subject's horoscope based on the positions of the planets and the moon at the moment of birth. Each volunteer also filled out a California personality Index survey. Then the birth data and horoscope for one subject, together with the results of the personality survey for that individual and for two other participants randomly selected from the experimental group, were given to an astrologer. The astrologer was asked to predict which of the three personality charts matched the birth data and horoscope for the subject.

Let p denote the probability of a correct prediction by an astrologer. Suppose an astrologer actually has no special predictive powers, as would be expected by those who view astrology as "quack science". The predictions then merely correspond to random guessing, that is, picking one of the three personality charts at random, so $p = \frac{1}{3}$. However, the participating astrologers claimed that $p > \frac{1}{3}$ since they felt they could predict better than random guessing.

How do we decide if the astrologer really has psychic powers? When she has a probability just greater than $\frac{1}{3}$?

Just like in lecture 9 and 10 , we made confidence intervals for the true value of a parameter. Now we will test the possibility of a parameter being equal to some set value or if it is really greater than or less than it. Using these methods we can answer all types of questions:

- Does a proposed diet truly result in weight loss, on average?

  Is there evidence of discrimination against women in promotion decisions?

- Does one advertising method result in better sales, on average, than another advertising method?

The main goal in many research studies is to check whether the data support certain statements or predictions. These statements are **hypotheses** about a population. They are usually expressed in terms of population parameters for variables measured in a study. A **hypothesis test** is a method of using data to summarize the evidence about a hypothesis. For instance in the above example our *hypothesis is that the probability p that an astrologer can correctly predict which of three personality charts applies to that person equals $\frac{1}{3}$ using a person's birth day and time. That is, an astrologers' predictions correspond to random guessing.* If a high proportion of the astrologers' predictions are correct then the data might provide strong evidence against the hyposthesis that $p = \frac{1}{3}$ in favor of an alternative hypothesis representing the astrologers' claim that $p > \frac{1}{3}$.

The steps of hypothesis testing

Step 1: Check your assumptions

Each test has certain assumptions or conditions that must be met in order to make the test valid and accurate. Of course, the test assumes that the data production used randomization. Additionally, most tests assume an adequate sample size and specific shape of distribution (for us, normal usually).

Step 2: Make your hypotheses

Each hypothesis test has two hypotheses about a population parameter:

> The **null hypothesis**(H_o) is a statement that the parameter takes a particular value
>
> The **alternative hypothesis**(H_a) states that the parameter falls in some alternative range of values not the null hypothesis

The value of the null hypothesis usually represents no effect or the neutral state. The value in the alternative hypothesis then represents an effect of some type. You always formulate these hypotheses BEFORE viewing or analyzing the data. An analogy for the hypotheses may be found in a legal trial in a courtroom, in which a jury must decide the guilt or innocence of a defendant. The null hypothesis is that the defendant is innocent. The alternative hypothesis is that the defendant is guilty. The jury presumes the defendant is innocent unless proven guilty "beyond a reasonable doubt". The "burden of proof" id on the prosecutor to convince the jury that the defendant is guilty.

For the astrologer example we would write our null and alternative hypotheses as

$$H_o : p = \frac{1}{3} \tag{2}$$

$$H_a : p > \frac{1}{3} \tag{3}$$

Step 3: Compute the test statistic

The parameter to which the hypotheses refer has a point estimate from the sample data. A **test statistic** describes how far that estimate falls from the parameter value given under the null hypothesis, usually measured by the number of standard error between the two.

Using the astrology example, the null hypothesis $H_o : p = \frac{1}{3}$ states that the probability p that an astrologer can correctly predict which of three personality charts applies to that person equals $\frac{1}{3}$. The study found $\hat{p} = \frac{40}{116} = .345$ meaning that 40 out of her 116 predictions were correct. The test statistic compares this value to the null hypothesis ($p = \frac{1}{3}$) using a Z-score that measures the number of standard deviations that the estimate falls from $\frac{1}{3}$.

Step 4: Find the P-value

To interpret a test statistic value we use a probability summary of the evidence against the null hypothesis, H_o. Here's how we get it:

> We presume that H_o is true since the "burden of proof" is on the alternative H_a.
>
> Then we consider the sorts of values we'd expect to get for the test statistic, according to its sampling distribution.
>
> If the test statistic computed from the data falls well out in a tail of the sampling distribution (meaning it that value has a low probability of occurring), it is far from what H_o predicts. So assuming H_o states the true parameter value, the computed test statistic would be very unusual.
>
> When sampling there are a large number of possible outcomes, any single one may be unlikely, so we summarize how far out in the tail the test statistic falls by the probability of that value and values even more extreme. This probability is called the **P-value**. The smaller the P-value, the stronger the evidence is against H_o.

In the astrology example, suppose a P-value is small, say 0.01. This means that if H_o were true (that an astrologer's predictions were just as good as random guessing), it would be very unusual to get sample data like we got. It is so unusual that it makes more sense (and is more probably) that the null hypothesis is not really true so there is strong evidence against the null hypothesis of random guessing and support of the astrologer's claim. On the other hand, if the P-value was much larger and not near 0, then the data are consistent with H_o. So if the P-value was 0.26 then this indicates that if the astrologer were actually randomly guessing, the observed data would not be unusual.

Step 5: Make a conclusion

All we need to do now is report the P-value and interpret what it means in terms of the situation. Based on the P-value, make a decision about H_o.