Ch. 21: The Accuracy of Percentages

So far, we assumed we knew what was in the box (at least, the average and SD), e.g.:

- Games of chance

- Sampling from a known population (known average and SD)

In practice, usually, we don't know the population (i.e., the box).

That's why we sample!!

Ex: We want to estimate the percentage of voters in a large population who support the Republican candidate and obtain a SE of this percentage.

HOW?

- Estimate the population percentage using the sample percentage.

- Estimate the SE by pretending that the population is just like the sample and calculate $SE_{\%}$.

This is called ________________.

Suppose we sample 1600 voters from this population, and find out that 56% of the sample support the Republican candidate.

1. Estimate the percentage of voters in the whole population who support the Republican candidate.

2. Find the standard error of this percentage.

Confidence Intervals

The normal approximation tells us that there is a 95% chance that a sample sum or percentage will be within 2 SE's of the EV.

Therefore, if we consider a range

$$\text{sample } \% \quad \pm \quad 2 \cdot \text{SE},$$

there is about a 95% chance that this will include the EV (the population percentage).

We call this range a __________________ (CI).

For this interval to be valid, the sample size should be large and the sample percentage should not be too close to 0% or 100%.

Ex: What is the 95% confidence interval for the percentage of voters who support the Republican candidate?

Ex: Lemons are premium–grade, table–grade, or juice–grade. A farmer takes a random sample of 500 lemons from a large crop and finds that 75 are juice–grade. Find a 95% confidence interval for the percentage of juice–grade lemons in the crop.

Confidence Intervals (ctd.)

We can construct different confidence intervals. Our intervals are based on the normal approximation and the ____________________ we desire.

$$\begin{array}{lcll} \text{sample \%} & \pm & 1\cdot\text{SE} & \rightarrow\ 68\%\ \text{CI} \\ \text{sample \%} & \pm & 2\cdot\text{SE} & \rightarrow\ 95\%\ \text{CI} \\ \text{sample \%} & \pm & 3\cdot\text{SE} & \rightarrow\ 99.7\%\ \text{CI} \end{array}$$

Ex: The 68% CI for the percentage of voters who support the Republican candidate is:

Ex: The 99.7% CI for the percentage of juice–grade lemons in the crop is:

Confidence Intervals (ctd.)

When we want some other confidence interval, the multiplier comes from the normal curve.

E.g., for a 77% confidence interval, we use

$$\text{sample } \% \quad \pm \quad \underline{\qquad\quad} \cdot \text{SE}$$

Ex: A health inspector takes a random sample of 300 ten–year–olds in a city and finds that 73% of them have had chicken–pox. Find a 90% confidence interval for the percentage of 10–year–olds in the city who have had chicken–pox.

Interpretation of Confidence Intervals

The 90% confidence interval for the percentage of children who have had chicken–pox is:

Ex: True or False and explain: There is a 90% chance that the population percentage is between ___ % and ___ %.

Interpretation

We say: "We are 90% confident that the population percentage is between ___ % and ___ %".

What does this mean?

- 90% of all such intervals contain the population percentage

or

- Before you sample, there is a 90% chance the sample you get will give an interval containing the population percentage.

Ex: Suppose we have a box containing a very large number of marbles, 80% red and 20% blue. If we have 100 people each sample 2,500 marbles and construct a confidence interval as

$$\text{percent reds in sample} \quad \pm \quad 2 \cdot \text{SE},$$

then about 95 (95%) should include the true value (80%) in their intervals.

In a computer simulation of the process, 96 of 100 confidence intervals include 80%. The other 4 do not.