Sampling Distribution of Sample Proportions

We can measure elements of a population in terms of a qualitative property but looking at the proportion of the population that has the specific quality.

We can take samples of size n from the population $\widetilde{N}$ and measure the number of successes (elements with the specific quality). Let x be this number.

Then is the **sample proportion**

For a fixed sample size n, the collection of all possible sample proportions p' forms the population $\widetilde{N}_{p'}$ has a probability distribution called the **sampling distribution of sample proportions**. Note that there is a sampling distribution of sample proportions for each sample size n.

Notice that x has a binomial distribution with probability of success p and probability of failure $q = 1 - p$.

The variable p' is the variable x scaled by the factor $1/n$. So the probability distribution for p' should have the same shape as the corresponding binomial.

Recall that the binomial distribution is approximately normal if both np and $n(1-p)$ are at least 5.

Properties of Sampling Distributions of Sample Proportions

$\mu_{p'}$ = expected value of random variable p'

= proportion of success withing $\widetilde{N}$

= p

$\sigma_{p'}$ = standard deviation of p'

= $\sqrt{\dfrac{p \cdot (1-p)}{n}}$

Confidence Interval Estimates for p

Method: Choose a level of confidence and take a random sample of size n. The resulting sample proportion

$$p' = \frac{x}{n}$$

will be our point-estimate for p.

Again, our margin of error is given by

$$E = z(\alpha/2) \cdot \sigma_{p'}$$

The problem is that we don't know p, this is what we are estimating. Instead we use p'.

We get

$$E = z(\alpha/2) \cdot \sqrt{\frac{p'(1-p')}{n}}$$

Example: In a random sample of 225 smokers, 162 considered themselves addicts. Construct a 95% C.I.E. (confidence interval estimate) for proportion of smokers that consider themselves addicts.

Increasing Sample Size

Rearranging the error formula we see that the smallest sample size that will give a desired error is

$$n = (z(\alpha/2))^2 \cdot \frac{p(1-p)}{E^2}$$

We will use a previous estimate for p if it is available. Otherwise we will use $p = 0.5.$

Why $p = 0.5$?

Example: We would like to use a random sample to predict the outcome of the next federal election for the Liberals. What is the minimum sample size required to have a margin of error of 3% with 95% confidence?

Given that the communist party have never won more than 5% of the vote, can we reduce the sample size if we want to predict the proportion of votes that the communist party will receive in the next election?

Hypothesis Tests Concerning One Proportion p

If we want to test a hypothesis concerning the proportion of a population p with a specific quality the test statistic will be the z-score of the sample proportion $p' = x/n$ which we calculate as follows

$$z = \frac{p' - p_0}{\sigma_{p'}} = \frac{p' - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Where p_0 is the value of p claimed by the null hypothesis.

As before, we can either use a classical or p-value approach.

Types of hypothesis:

- Two tail test

- Right tail test

- Left tail test

Example: A rep for Microsoft claims that Microsoft outlook is used by at least 75% of internet users. A random sample of 300 internet users included 216 that use Microsoft outlook. Use a p-value approach to test the claim at $\alpha = 0.05.$

Example: In 1998 Amtrak trains arrived on time 78% of the time. A recent study found that 330 of 400 amtrak trains arrived on time. Does this indicate a change in lateness? (Use a classical approach to test this claim at $\alpha = 0.05$)