Conditions for inference with the SDSP

Just like we did with the sampling distribution of the sample mean, we have to meet specific sampling conditions in order to be able to use the sampling distribution of the sample proportion to make inferences about the population proportion.

The conditions for inference that apply to the sampling distribution of the sample proportion are similar to the conditions we applied to the sampling distribution of the sample mean.

Random

Any sample we take needs to be a simple random sample. Often we'll be told in the problem that sampling was random.

Normal (large counts)

The sampling distribution of the sample proportion can only be guaranteed to be normal if $np \geq 5$ and $n(1-p) \geq 5$, where n is the sample size and p is the population proportion. If $np \geq 5$ is true, it tells us that we can expect to have at least 5 "successes" in the sample, and if $n(1-p) \geq 5$ is true, it tells us that we can expect to have at least 5 "failures."

So if our sample size is $n = 100$ and the population proportion is $p = 60\,\%$, then we multiply 100 by 0.6 and by $1 - 0.6 = 0.4$ to make sure both values are at least 5.

$$100 \cdot 0.6 = 60 > 5$$

$$100 \cdot 0.4 = 40 > 5$$

Since both values are at least 5, the sampling distribution of the proportion is approximately normal.

If we don't know the population proportion p, then we use the sample proportion $\hat{p}$ as its best estimate, and use $\hat{p}$ to check that our sample has at least 5 "successes" and at least 5 "failures."

Note that there's some debate among statisticians and statistics textbooks about whether the large counts condition should be $np \geq 10$ and $n(1-p) \geq 10$ or $np \geq 5$ and $n(1-p) \geq 5$. For our purposes, we'll stick to using the $np \geq 5$ and $n(1-p) \geq 5$ rule.

Independent (10% rule)

If we're sampling with replacement, then the 10% rule tells us that we can assume the independence of our samples. But if we're sampling without replacement (we're not "putting our subjects back" into the population every time we take a new sample), then we need keep our sample size below 10% of the total population.

If our sample meets these conditions, then we can use the sampling distribution of the sample proportion to answer questions about the probability that any given sample proportion will fall within some distance of the population proportion.

For these types of problems, we'll need to use

$$z_{\hat{p}} = \frac{\hat{p} - p}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

to determine z-scores that apply to the sampling distribution of the sample proportion. Let's work through an example.

Example

An ice cream shop claims that 40% of their $1{,}000$ customers order their ice cream in a waffle cone. We want to verify this claim, so we take a random sample of 90 customers and see whether or not they order a waffle cone.

What is the probability that our results are within 5% of the ice cream shop's 40% claim?

In our case, $n = 90$ and $p = 0.4$, which means $1 - p = 0.6$, so

$$np = (90)(0.4) = 36 \geq 5$$

$$n(1-p) = (90)(1-0.4) = (90)(0.6) = 54 \geq 5$$

and we've verified normality. We were told in the question that our sample was random, and our sample is 90 of the total population of $1{,}000$, which means it's $90/1{,}000 = 9\%$ of the population, so we're not violating the 10% rule.

With the conditions for inference out of the way, we can calculate the mean and standard deviation of the sampling distribution of the sample proportion. The mean is

$$\mu_{\hat{p}} = p$$

$$\mu_{\hat{p}} = 0.4$$

To find the standard deviation of the sampling distribution, we need to apply the finite population correction factor, given that our population is finite and we're sampling from more than $5\,\%$ of the population.

$$\sigma_{\hat{p}} = \sqrt{\frac{0.4(1-0.4)}{90}}\sqrt{\frac{1{,}000-90}{1{,}000-1}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.4(0.6)}{90}}\sqrt{\frac{910}{999}}$$

$$\sigma_{\hat{p}} \approx 0.0516(0.9544)$$

$$\sigma_{\hat{p}} \approx 0.0492$$

The question asks us for the probability that our sample proportion is within $5\,\%$ of population proportion $p = 40\,\%$. In other words, how likely is it that the sample proportion falls between $35\,\%$ and $45\,\%$?

$$P(0.35 < \hat{p} < 0.45) \approx P\left(\frac{0.35-0.4}{0.0492} < z < \frac{0.45-0.4}{0.0492}\right)$$

$$P(0.35 < \hat{p} < 0.45) \approx P\left(\frac{-0.05}{0.0492} < z < \frac{0.05}{0.0492}\right)$$

$$P(0.35 < \hat{p} < 0.45) \approx P(-1.02 < z < 1.02)$$

In a z-table, a z-value of 1.02 gives 0.8461,

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830

and a value of -1.02 gives 0.1539.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611

Which means the probability under the normal curve of the sampling distribution of the sample proportion between these z-scores is

$$P(-1.02 < z < 1.02) \approx 0.8461 - 0.1539$$

$$P(-1.02 < z < 1.02) \approx 0.6922$$

$$P(-1.02 < z < 1.02) \approx 69\,\%$$

Which means there's an approximately $69\,\%$ chance that our sample proportion will fall within $5\,\%$ of the ice cream shop's claim. In other words, approximately $69\,\%$ of our samples will produce a sample proportion that's within $5\,\%$ of the population proportion.