Sampling Distributions

Sampling Distributions

Recall: A statistic is a number that describes a sample. A parameter is a number that describes a population. We use a statistic (computed from a sample) to estimate a parameter.

Questions:

1. What proportion of all U.S. adults have high blood pressure? Take a sample and compute the proportion of adults in the sample with high blood pressure (a statistic) and use that as an estimate.

2. What is the mean body weight for an adult male? Take a sample and compute the mean weight for the sample (a statistic) and use that as an estimate.

Different samples give different estimates. A sampling distribution is a probability distribution of a statistic based on all possible simple random samples of the same size from the same population.

Sampling Distribution of a Sample Proportion

We estimate a population proportion p by using a sample proportion $\hat{p}$.

The sample proportions vary from sample to sample. It turns out that the sampling distribution of $\hat{p}$ is approximately normal with a mean of $\mu_{\hat{p}} =$ p and a standard deviation of $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

Requirement: For this to hold, np > 5 and n(1-p) > 5.

Note: The sampling distribution of $\hat{p}$ is closely related to the binomial distribution. With the binomial distribution, we're interested in the number of successes in a sample, and with sampling distribution of $\hat{p}$, we're interested in the proportion of successes in a sample.

Example: Suppose that 40% of individuals in a population are known to have some characteristic. If random samples of size 50 are selected, what is the distribution of $\hat{p}$?

Solution: The set of sample proportions is normally distributed with a mean of $\mu_{\hat{p}} = 0.40$ and a standard deviation of $\sigma_{\hat{p}} = \sqrt{\frac{(0.4)(0.6)}{50}} = 0.0692820323$.

Example: Suppose that 70% of all households have a personal computer. In a simple random sample of 100 households, what is the probability that more than 80% of households surveyed have a personal computer?

Solution: We want to find $P(\hat{p} > 0.80)$.

The set of sample proportions is normally distributed with a mean of $\mu_{\hat{p}} = 0.70$ and a standard deviation of $\sigma_{\hat{p}} = \sqrt{\frac{(0.70)(0.30)}{100}}$ = 0.0458257569

Compute the z score: z = (value – mean)/(standard deviation)

= (0.80 – 0.70)/0.0458257569 = 2.18

$P(\hat{p} > 0.80) = P(Z > 2.18) = normalcdf(2.18, 9999, 0, 1) = 0.0146$

Example: Suppose that 48% of all elderly men suffer from impaired hearing. If a sample of 400 men is selected, what is the probability that between 45% and 55% of the men in the sample suffer from impaired hearing?

Solution: We want to find $P(0.45 < \hat{p} < 0.55)$.
The set of sample proportions is normally distributed with a mean of 0.48 and a standard deviation of $\sigma_{\hat{p}} = \sqrt{\frac{(0.48)(0.52)}{400}}$ = 0.024979992

Compute the z scores:

z = (value - mean)/(standard deviation) = (0.45 – 0.48)/0.024979992 = -1.20

z = (value - mean)/(standard deviation) = (0.55 – 0.48)/0.024979992 = 2.80

$P(0.45 < \hat{p} < 0.55) = P(-1.20 < Z < 2.80) = normalcdf(-1.20, 2.80, 0, 1) = 0.8824$

Sampling Distribution of a Sample Mean (Assuming the Population Distribution is Normal)

We estimate a population mean μ by using a sample mean $\bar{X}$.

The sample means vary from sample to sample. It turns out that if X is normally distributed, the sampling distribution of $\hat{X}$ is normal with a mean of $\mu_{\hat{X}} = \mu$ and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Example: Suppose X is a normally distributed random variable with a mean of 20 and a standard deviation of 5. If simple random samples of size 25 are selected, what is the distribution of $\bar{X}$?

Solution: Because the original distribution is normal, the set of sample means is normally distributed with a mean of $\mu_{\bar{X}} = 20$ and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1$

Example: Suppose that the total amount spent on food and lodging per day by all families vacationing for per day in Hawaii is normally distributed with a mean of $650 and a standard deviation of $120. What is the probability that a simple random sample of 10 families will spend an average of between $600 and $700?

Solution: We want to find $P(600 < \bar{X} < 700)$.

The set of sample means is normally distributed with a mean of $\mu_{\hat{X}} =$ $650 and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{120}{\sqrt{10}}$ = 37.94733192.

z = (value - mean)/(standard deviation) = (600 – 650)/37.94733192 = -1.32

z = (value - mean)/(standard deviation) = (700 – 650)/37.94733192 = 1.32

$P(600 < \bar{X} < 700) = P(-1.32 < Z < 1.32) = normalcdf(-1.32, 1.32, 0, 1) = 0.8132$

The Central Limit Theorem

In the previous example, the original population had a normal distribution. However, it can be shown that if the sample size is large enough (n > 30), no matter how the original population is distributed, the sampling distribution of $\hat{X}$ is normal with a mean of $\mu_{\hat{X}} = \mu$ and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Example: Suppose that the distribution of X is unknown with a mean of 55 and a standard deviation of 12. If simple random samples of size 60 are selected, what is the distribution of $\bar{X}$?

Solution: Because the sample size is greater than 30, the Central Limit Theorem applies. The set of sample means is normally distributed with a mean of $\mu_{\bar{X}} = 55$ and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{60}} = 1.5$

Example: Suppose that the average salary in a certain area is $30,000 with a standard deviation of $7,000. If a simple random sample of 100 people is selected, what is the probability that the average salary for the sample is below $29,000?

Solution: We want to find $P(\bar{X} < 29{,}000)$.

The Central Limit Theorem applies because the sample size is greater than 30.

The set of sample means is normally distributed with a mean of $\mu_{\hat{X}} =$ $30,000 and a standard deviation of $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{7000}{\sqrt{100}}$ = 700.

Z = (value - mean)/(standard deviation) = (29,000 – 30,000)/700 = -1.43

P($\bar{X}$ < 29,000) = P(Z < -1.43) = normalcdf(-9999, -1.43, 0, 1) = 0.0764

Example: If a simple random sample of 100 people is selected, what is the probability that the average salary for the sample is between $28,500 and $30,500?

Solution: We want to find P(28,500 < $\bar{X}$ < 30,500).

z =(value - mean)/(standard deviation = (28,500 – 30,000)/700 = -2.14

z = (value - mean)/(standard deviation) = (30,500 – 30,000)/700 = 0.71

P(28,500 < $\bar{X}$ < 30,500) = P(-2.14 < Z < 0.71) = normalcdf(-2.14, 0.71, 0, 1) = 0.7450