Section 4.11: Sampling Distributions and the Central Limit Theorems

1 The Sampling Distribution of the Mean

The goal of inferential statistics is to use a sample to make an inference about a population.

A class of 50 students wants to study the average GPA at KSU.

Student number 1 collects a sample of 5 student GPA's.
$S_1 = \{3.01, 3.28, 2.97, 3.41, 3.21\}$, $\bar{x} = 3.176$
Student number 2 collects a sample of 5 student GPA's.
$S_2 = \{2.89, 3.33, 1.97, 2.59, 3.01\}$, $\bar{x} = 2.758$
Student number 3 collects a sample of 5 student GPA's.
$S_3 = \{2.93, 2.78, 3.41, 3.17, 2.81\}$, $\bar{x} = 3.02$
The remaining 47 students proceed in a similar fashion.

Problem 1 *Given 50 sample averages what might you do to estimate the true population average?*

Problem 2 *Are there differences in the variations in the single observations and the variations of the sample averages?*

Definition 1 *The **sampling distribution** of a sample statistic is the distribution of the values of the statistic created by repeated samples of n observations.*

2 The Central Limit Theorem for Means

Theorem 1 ***The Central Limit Theorem*** *(CLT for means) The mean of a random sample has a sampling distribution whose shape can be approximated by a normal model if $n \geq 30$ (no matter the shape of the original distribution). The larger the sample, the better the approximation will be. The center of the sampling distribution, μ_x, is the population mean μ. The spread of the sampling distribution, σ_x, is $\frac{\sigma}{\sqrt{n}}$. The value σ_x is called the **standard error**.*

Example 1 *When throwing three darts, the average score is 45 with a standard deviation of 11.5. What is the center and spread for the sampling distribution of the mean for the average score when each of the 40 students in class throws three darts? The center* $\mu_x = \mu = 45$ *and the spread is the standard error* $\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{11.5}{\sqrt{40}} = 1.818\,3$

Example 2 *What is the probability that the average score of the forty students in class will be less than 42 when throwing three darts? Using* $normalcdf(-\infty, 42, 45, 1.82)$ *we get* 0.0496.

Example 3 *What is the probability that the average score of the forty students in class will between 40 and 50 when throwing three darts? Using* $normalcdf(40, 50, 45, 1.82)$ *we get* 0.994.

Example 4 *What is the probability that Natasha will score between 40 and 50 when throwing three darts? We do not know the shape of the original distribution and cannot answer this question.*

Example 5 *What is the cutoff score between the worst 15% and the best 85% of groups of 40 students? This would be* P_{15} *in the sampling distribution of the mean. Using* $invNorm(.15, 45, 1.82)$ *we get* 43.11. *So, the worst 15% of all groups of 40 students have an average score less than 43.11.*

Example 6 *Is the normal model good for predicting the average score for a sample of 5 students? Explain. No, this sample size is too small to apply the Central Limit Theorem.*

Remark 2 *If the original data distribution is normal then the sampling distribution will also be normal even if* $n < 30$.

Problem 3 *The time it takes students in a cooking school to learn to prepare seafood gumbo is a random variable with a normal distribution where the average is 3.2 hours with a standard deviation of 1.8 hours.*

i. Find the probability that the average time it will take a class of 36 students to learn to prepare seafood gumbo is less than 3.4 hours.

ii. Find the probability that it takes Kevin between 3 and 4 hours to learn to prepare seafood gumbo.

iii. Would it be unusual for the average time a group of 50 students needs to learn to prepare seafood gumbo is less than two hours and thirty minutes?

iv. Is the normal model good for predicting the average score for a sample of 5 students? Explain.

Problem 4 *Grade point averages at a particular school follow a normal distribution with* $\mu = 2.89$ *and* $\sigma = 0.63$.

i. Find the probability that the average GPA for a sample of 5 students is greater than 3.0.

ii. Find the probability that the average GPA for a sample of 10 students is between 2.0 and 2.75.

iii. Find the probability that Nathan's GPA is between 2.0 and 2.75.

Problem 5 *The time it takes a baseball player to learn the team's base-running signals is a random variable with a normal distribution where the average is 8.2 hours with a standard deviation of 2.2 hours.*

i. Find the probability that the average time it will take a roster of 36 players to learn the team's base-running signals is less than 9 hours.

ii. Find the probability that the average time it will take a roster of 45 players to learn the team's base-running signals is less than 9 hours.

iii. Find the probability that it takes Jesse less than 9 hours to learn the team's base-running signals.

Problem 6 *The lifespan of drummers in Spinal Tap follows a normal distribution with an average of 10 months and a standard deviation of 1.5 months. Neil signs on to be their next drummer. What is the probability that Neil survives to see his 1 year anniversary as the new drummer in Spinal Tap?*

3 The Central Limit Theorem for Proportions

The mean is not the only statistic whose sampling distribution is normal. Many useful statistics have their own versions of the Central Limit Theorem.

Theorem 3 ***The Central Limit Theorem*** *(CLT for proportions) The proportion of a random sample has a sampling distribution whose shape can be approximated by a normal model if* $np \geq 10$ *and* $n(1-p) \geq 10$*. The larger the sample, the better the approximation will be. The center of the sampling distribution,* μ_p*, is the population proportion* p*. The spread of the sampling distribution,* σ_p*, is* $\sqrt{\frac{p(1-p)}{n}}$*. The value* σ_p *is called the* ***standard error****.*

Example 7 *As of 6/20/2012, Ben Roethlisberger has completed 63.1% of his passes. If Ben throws 40 passes in his next game, what is the likelihood that he completes at least 60% of his passes? Note that* $p = .631$*, so* $\sigma_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.631(1-.631)}{40}} = 0.076$*. Using* $normalcdf(.6, \infty, .631, .076)$ *we get* 0.658*.*

Example 8 *If Ben throws 35 passes in his next game, what is the likelihood that he completes no more than 15 of his passes? Note that* $p = .631$*, but with a different number of passes* $\sigma_p = \sqrt{\frac{.631(1-.631)}{35}} = 0.082$*. Also, we need to convert 15 completions into a* $\frac{15}{35} = 0.428\,57$ *completion rate. Now using* $normalcdf(-\infty, .428, .631, .082)$ *we get* 0.0067*.*

Problem 7 *A fair coin is flipped 100 times and the percentage of heads is recorded. We should expect to see* $50\% = .5$ *of the flips be heads. Of course, where chance is involved the actual percentage of observed heads may not be exactly as expected.*

1. Find σ_p.

2. Are the conditionals of the central limit theorem met?

3. A fair coin is flipped 100 times and the percentage of heads is recorded. What is the probability that at most 45% of the flips are heads?

Problem 8 *A fair die is rolled 72 times and the percentage of 6s is recorded. What is the probability that at most 10% of the rolls are 6s?*

Problem 9 *Shaquille O'Neal has a lifetime free throw percentage rate of* $p = .527$*. Assume free throw attempts are independent.*

1. If Shaq shoots 40 free throws what is the probability that he makes at least 50% of his free throws?

2. If Shaq shoots 50 free throws what is the probability that he makes no more than 60% of his free throws?

3. If Shaq shoots 100 free throws a day, determine the number of free throws made that separates his 10% worst shooting days from the other shooting days.

Problem 10 *It is estimated that 53% of all movie-watching, baseball fans prefer the Durham Bull's Ebby Calvin (Nuke) LaLoosh over the Cleveland Indian's Ricky (Wild Thing) Vaughn. If 150 fans vote on which player is better, what is the probability that Vaughn receives at least 50% of the vote?*

Problem 11 *A candy company claims that its candy flavor mix contains 13% dark chocolate candies. Suppose that the candies are packaged at random in large bags containing about 200 candies. What is the probability that a bag will contain more than 20% dark chocolate candies?*

4 Exercises

1. Navidi: 1-5, 7, 8, 10, 11, 13, 16