STATISTICS

Lab5

Limit theorems and sampling distributions

Warsaw School of Economics
Winter Semester 2017

based on J.T. Mc Clave, P.G. Benson,T. Sincich: Statistics for Business and Economics, 11th Edition, 2010

De-Moivre-Laplace limit theorem

$\hat{p} = \frac{X}{n}$ sample proportion

- Where X has a binomial distribution
- When n is large (n>=100) it is diffic probabilities of the binomial rando $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
- It was shown that in this situation we can use normal probability distribution with parameters $\mu = np$ and

$$\sigma = \sqrt{np(1-p)}$$

$$X \sim N(np, \sqrt{np(1-p)})$$

This issue was proved by de-Moivre and Laplace and is called now **de Moivre-Laplace limit theorem**

Linderberg-Levy (Central) Limit Theorem

Consider a random sample of n observations selected from a population which is not **normally distributed or its distribution is unknown**

Then, when n is sufficiently large (n≥100), the sampling distribution of $\overline{X}$ will be approximately a normal with mean μ and standard deviation $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

$$\boxed{\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})}$$

or

$$\boxed{\bar{x} \sim N(\mu, \frac{S}{\sqrt{n}})}$$

If population standard deviation is unknown

The larger the sample size, the better will be the normal approximation to the sampling distribution of $\overline{X}$

This issue was proved by Linderberg and Levy and now is called **Linderberg-Levy Limit Theorem**

Sampling distribution of a mean

Population distribution	Population standard deviation	Sample size	Sample statistic
normal	known	any	$\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$
normal	unknown	>=30	$\bar{x} \sim N(\mu, \frac{S}{\sqrt{n}})$
normal	unknown	< 30	$\frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \sim t - Student$
any	known	>=100	$\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

Sampling distribution of a sum

Population distribution	Population standard deviation	Sample size	Sample statistic
normal	known	any	$\sum X_i \sim N(\mu \cdot n, \sigma \cdot \sqrt{n})$
normal	unknown	>=30	$\sum X_i \sim N(\mu \cdot n, S \cdot \sqrt{n})$
any	known	>=100	$\sum X_i \sim N(\mu \cdot n, \sigma \cdot \sqrt{n})$

Sampling distribution of a proportion

- So we learnt that if n (n>=100) is large and X has a binomial distribution we can use a normal approximation for X:

$$X \sim N(np, \sqrt{np(1-p)})$$

- Similarly we can derive a distribution for a sample proportion:

$$\hat{p} = \frac{X}{n}$$

$$\hat{p} \sim N(p, \sqrt{\frac{p(1-p)}{n}})$$

Exercise 1
De-Moivre-Laplace limit theorem

According to the insurance company calculations probability of stealing a car in Wąchock in one year period amounts to 5%.

1) Consider 5 owners of the cars (ex. neighbors). What is the probability that one of their 5 cars will be stolen during a year?
2) Consider the parking with 30 cars. What is the probability that none of the cars will be stolen during a year?
3) According to the insurance company databases 200 citizens of that town have bought anty-theft insurances for their cars. What is the distribution of the possible number of loses? What is the expected number of loses? What is the standard deviation of the number of loses?
4) Using the normal approximation calculate the probability that the percentage observed by this insurance company in Wąchock will be lower than 4%? Compare the results with the calculated exact results.
5) Suppose that the same probability of loses is observed for the whole portfolio of insurances in that company (n=100 000). What is the probability that the observed percentage of customers with loses will be higher than 5.2%?

Exercise 2 – Central Limit Theorem

The sum of goods bought by each of the customers of Carrefour supermarket is an unknown random variable with expected value EX=150 PLN and DX=90 PLN.

1) What is the probability that the total value of shopping expenditures of 100 clients will be higher than 18 000 PLN?
2) What is the probability that the average expenditure of 120 clients will be between 135 and 145 PLN?

Exercise 3
Sampling distributions of mean

Customers of a popular bar had some doubts about the honesty of a barman, who serves 0,5l pints of beer. One day the group of students decided to check if the barman are honest by trying a sample of beers. They were also aware that there is usually a margin of error (sometimes randomly a bit more or less beer than 0,5 l is in the glass). and they wanted to know the probability of random mistakes of an honest barman.

1) Suppose that the amount of beer in the glass filled by an honest barman has normal distribution N(0,5 ; 0,05). What is the probability that in the sample of 16 beers the observed average will be lower than 0,48 ?
2) Suppose that the amount of beer in the glass filled by an honest barman has normal distribution with expected value 0,5. What is the probability that in the sample of 16 beers the observed average will be lower than 0,48 if the standard deviation observed in this sample was equal to 0,05 ?
3) Suppose that we don't know the distribution of amounts of beer filled by an honest barman but we know the expected value EX= 0,5 and standard deviation in the population DX=0,05. What is the probability that in the sample of 120 beers average amount in the glass will be below 0,49 l.

Exercise 4

One measure of elevator performance is cycle time–the time between successive elevator starts. *Simulation* (Oct. 1993) published a study on the use of a micro-computer-based simulator for estimating elevators cycle times. The simulator produced an average cycle time μ of 26 seconds when traffic intensity was set at 50 persons every five minutes. Consider a sample of 200 simulated elevator runs and let $\bar{x}$ represent the mean cycle time of this sample.

a. What do you know about the distribution of x, the time between successive elevator starts? (Give the value of the mean and standard deviation of x and the shape of the distribution, if possible.)

b. What do you know about the distribution of $\bar{x}$? (Give the value of the mean and standard deviation of $\bar{x}$ and the shape of the distribution, if possible.)

c. Assume σ, the standard deviation of cycle time x, is 20 seconds. Use this information to calculate $P(\bar{x} > 26.8)$.

d. Repeat part c but assume $\sigma = 10$.

Source: J.T. Mc Clave, P.G. Benson,T. Sincich: Statistics for Business and Economics, 9th Edition, 2005

Exercise 5

Last year a company began a program to compensate its employees for unused sick days, paying each employee a bonus of one-half the usual wage earned for each unused sick day. The question that naturally arises is, "Did this policy motivate employees to use fewer sick days?" *Before* last year, the number of sick days used by employees had a distribution with a mean of 7 days and a standard deviation of 2 days.

a. Assuming that these parameters did not change last year, find the approximate probability that the sample mean number of sick days used by 100 employees chosen at random was less than or equal to 6.4 last year.

b. How would you interpret the result if the sample mean for the 100 employees was 6.4.?

Source: J.T. Mc Clave, P.G. Benson,T. Sincich: Statistics for Business and Economics, 9th Edition, 2005

Exercise 6

Researchers in a Massachusetts Institute of Technology are experimenting with melatonin, a hormone which might alleviate negative consequences of a jet lag. In order to test their hypothesis they selected a sample of 100 volunteers, gave them a dosage of melantonin, placed in a dark room at midday and told to close their eyes for 30 minutes. The researchers measured the time elapsed before each volunteer fall asleep and they found that the mean was 9 minutes.

The previous research suggests that with a placebo (no hormone) the mean time to fall asleep is 15 minutes with a standard deviation 10 minutes. If melatonin is effective it should be very unlikely that the sample mean is 9 or less. Is melatonin effective?

Source: J.T. Mc Clave, P.G. Benson,T. Sincich: Statistics for Business and Economics, 9th Edition, 2005

Exercise 5

the probability that a scanned item is priced incorrectly is $^1/_{30} = .033$.

Suppose 10,000 supermarket items are scanned. What is the approximate probability that you observe at least 100 items with incorrect prices?

Source: J.T. Mc Clave, P.G. Benson,T. Sincich: Statistics for Business and Economics, 9th Edition, 2005

Exercise 6

An article in *The International Journal of Sports Psychology* (July–Sept. 1990) evaluated the relationship between physical fitness and stress. The research revealed that white-collar workers in good physical condition have only a 10% probability of developing a stress-related health problem. What is the probability that more than 60 in a random sample of 400 white-collar employees in good physical condition will develop stress-related illnesses?

Source: J.T. Mc Clave, P.G. Benson,T. Sincich: Statistics for Business and Economics, 9th Edition, 2005

Exercise 7

According to *New Jersey Business* (Feb. 1996), Newark International Airport's new terminal handles an average of 3,000 international passengers an hour, but is capable of handling twice that number. Also, 80% of arriving international passengers pass through without their luggage being inspected and the remainder are detained for inspection. The inspection facility can handle 600 passengers an hour without unreasonable delays for the travelers.

a. When international passengers arrive at the rate of 1,500 per hour, what is the expected number of passengers who will be detained for luggage inspection?

b. In the future, it is expected that as many as 4,000 international passengers will arrive per hour. When that occurs, what is the expected number of passengers who will be detained for luggage inspection?

c. Refer to part b. Find the approximate probability that more than 600 international passengers will be detained for luggage inspection. (This is also the probability that travelers will experience unreasonable luggage inspection delays.)

Source: J.T. Mc Clave, P.G. Benson,T. Sincich: Statistics for Business and Economics, 9th Edition, 2005

Exercise 8

The next table reports the credit card industry's market share data for mid-2003. A random sample of 100 credit card users is to be questioned regarding their satisfaction with their credit card company. For simplification, assume that each credit card user carries just one credit card and that the market share percentages are the percentages of all credit card customers that carry each brand.

Credit Card	Market Share %
Visa	51.4
MasterCard	30.7
American Express	12.3
Discover	5.6

Source: U.S. Payment Card Information Network, June 2003.

a. For random samples of 100 credit card users, what is the expected number of customers who carry Visa? Discover?

b. What is the probability that half or more of the sample of credit card users carry Visa? American Express?

Source: J.T. Mc Clave, P.G. Benson,T. Sincich: Statistics for Business and Economics, 9th Edition, 2005