246명의 환자들의 표본으로부터 나이 신장 체중 수입 혈압 등의 여러 변인들의 평균, 분산, 표준편차와 표준 오차를 계산중이다. 환자 100명분의 데이터를 추가로 받았을 때 위의 표본 통계량 중 어떤 것이 주로 바뀌었을까?

## UNIT 5 HOMEWORK

Question 1:

I am calculating the mean, median, standard deviation, and standard error of the mean for several variables (age, height, weight, income, blood pressure, etc.) from a sample of 246 patients. If I receive data for an additional 100 patients, which of the above statistics (mean, median, standard deviation, or standard error of the mean) would be expected to change substantially?

1. The mean
2. The median
3. The standard deviation
4. The standard error of the mean

운동 프로그램을 완료한 낭포 성 섬유증을 앓는 13 명의 아동을 대상으로 한 연구에서 운동 내구성 (운동 지속 시간)은 평균 0.77 분 향상되었습니다. 내구 시간 변화의 표준 편차는 0.83 분이었다. 이론적 분포의 평균은 무엇입니까?

Question 2:

In a study of 13 children with cystic fibrosis who completed an exercise program, exercise endurance (duration on an exercise test) improved by a mean of 0.77 minutes. The standard deviation for the change in endurance was 0.83 minutes. What is the theoretical distribution of the mean?

1. T-distribution (with 12 degrees of freedom); mean=true mean; standard error= 0.23
2. Standard normal distribution; mean=true mean; standard error= 0.23
3. T-distribution with 12 degrees of freedom; mean=true mean; standard error= 0.07
4. T Standard normal distribution; mean=true mean; standard error= 0.07
5. T-distribution (with 12 degrees of freedom); mean=true mean; standard error= 0.83

Question 3:

Thirty heart disease patients are put on an exercise regimen. Twenty patients improve on the exercise stress test and ten don't improve or get worse. Calculate the 95% confidence interval for the true proportion of heart disease patients who improve their fitness using this particular exercise regimen. Recall that proportions are normally distributed with a standard error of

$$\sqrt{(p(1-p)/n)}$$

(You may use the observed proportion to calculate the standard error.)

1. 66%
2. 66%-70%
3. 50%-84%
4. 6%-76%

30 명의 심장 질환 환자가 운동 처방을 받는다. 20 명의 환자가 운동 스트레스 테스트에서 개선되고 10 명이 개선되거나 악화되지 않습니다. 이 특정 운동 요법을 사용하여 운동 능력을 향상시키는 심장 질환 환자의 실제 비율에 대한 95 % 신뢰 구간을 계산하십시오.

표준 오차 약 8%, 표본의 확률 p = 66%, 95% 신뢰구간 - 2시그마
66-16< < 66+16

Question 4:

The following data were collected from a case-control study of breast cancer and fat intake:

| | case | control |
|---|---|---|
| High-fat diet | 10 | 5 |
| Low-fat diet | 40 | 55 |
| total | 50 | 60 |

Statistical inferences for odds ratios are based on the natural log of the odds ratio, rather than the odds ratio itself (because the distribution for an odds ratio does not follow a normal distribution). The sampling distribution of the natural log of the odds ratio (lnOR) follows a normal distribution, with standard error

$$=\sqrt{(1/a + 1/b + 1/c + 1/d)}$$

(where a, b, c, and d are the cells in the 2x2 table).

Calculate the odds ratio for breast cancer (comparing high-fat diet to low-fat diet) from the 2x2 table above.

1. 2.75

10/5 / 40/55 = 2*55/40 = 2.75

2. 0.40
3. 2.52
4. 2.40
5. 1.00

Question 5:

Take the natural log of the odds ratio that you calculated in question 4.

1. 15.6

ln 2.75 = 1.0116...

2. 2.75
3. 0.5
4. 1.01
5. 0

Question 6:

Calculate the standard error of the lnOR, according to the formula given above.

1. 15.6
2. 0.586
3. 0.09
4. 1.172
5. 1.01

Question 7:

Calculate the 95% confidence interval for the lnOR.

1. .24, 1.76
2. -.01, 1.96
3. -.14, 2.16
4. 0, -2.16
5. -.34, 3.16

Question 8:

Convert the upper and lower confidence limits that you calculated in (7) back to odds ratios by exponentiating (i.e., calculate the 95% confidence interval for the OR).

1. 0.87, 8.66
2. 1.58, 3.92
3. 1.01, 9.65
4. 0, 2.16
5. 0.51, 6.56

Question 9:

TRUE OR FALSE. The confidence interval of an odds ratio is symmetric. Why?

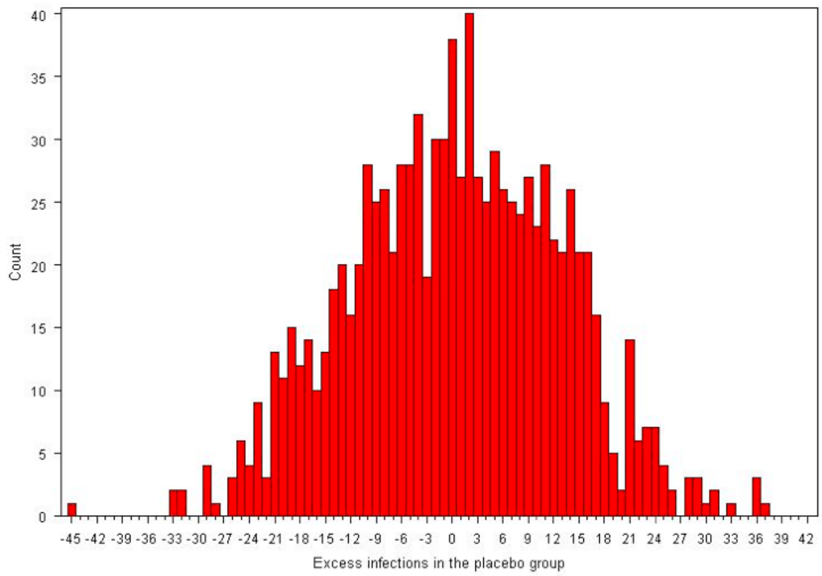
Question 10:

TRUE or FALSE. The odds ratio here is not statistically significant at the p=.05 level. Why?

Question 11:

A hypothetical HIV vaccine trial involving 20,000 participants—10,000 in the vaccine group and 10,000 in the placebo group—had the following results: 6.3 infections per 1000 in the vaccine group and 9.0 infections per 1000 in the placebo group.

I ran a computer simulation to predict possible outcomes of the trial if the null hypothesis is true—that is, if vaccinated and unvaccinated people are equally likely to contract HIV. I ran 1000 virtual trials of 20,000 people (10,000 per group) assuming that the vaccine is ineffective. Outcomes are expressed as excess infections in the placebo group. Here are the results of the 1000 virtual trials displayed as a histogram.




Question 11:

From the simulation, I learn that the statistic 'excess infections in the placebo group' follows a normal distribution with a mean of 0 and a standard deviation (standard error) of 12.3. Use this information to calculate the two-sided p-value. Round to the nearest thousandth.

Question 12:

How would the above histogram change if I ran the simulation 10,000 times rather than 1000 times?

1. It would simply be more smooth.
2. The p-value would get smaller.
3. The standard error would get smaller.
4. The distribution would be more similar to a T-distribution.

Question 13:

The Centers for Disease Control (CDC) collects yearly statistics on drinking behavior in the United States by surveying a random sample of U.S. adults. The following data display the percent of adults aged 18 years and over who had 5 or more drinks in 1 day at least once in the given year.

| Year | Percent | 95% confidence interval |
|---|---|---|

| | | |
|---|---|---|
| 1997 | 21.4 | 20.8-22.0 |
| 1998 | 20.2 | 19.6-20.8 |
| 1999 | 20.3 | 19.6-21.0 |
| 2000 | 19.2 | 18.6-19.9 |
| 2001 | 20.0 | 19.4-20.6 |
| 2002 | 19.9 | 19.2-20.5 |
| 2003 | 19.1 | 18.5-19.8 |

For which years (or year) above could you reject the null hypothesis that more than 20% of U.S. adults had 5 or more drinks in 1 day, at the .05 significance level?

There may be more than one correct answer.
1.1997 2.1998 3.1999 4.2000 5.2001 6.2002 7.2003