**Lecture 4 Dr. Haider Raheem**

**Introduction to Epidemiology**

**Monthly Cost of Pharmaceutical Treatment**

**Example**

The monthly cost of pharmaceutical treatment for nine patients with disease X was found to be JPY 150, 190, 220, 260, 210, 100, 130, 120, and 150K, respectively.

**Question 1.** Find the mean, variance, and median.

**Answer 1.** Mean = (150 + 190 + 220 + 260 + 210 + 100 + 130 + 120 + 150)/9 = JPY 170K

Variance = sum of squared deviations from the mean/(sample size − 1)

= {(150 − 170)2 + (190 − 170)2 + (220 − 170)2 + (260 − 170)2 + (210 − 170)2 + (100 − 170)2 + (130 − 170)2 + (120 − 170)2 + (150 − 170)2}/(9 − 1)

= 22400M/8 = JPY2800M

Median = the middle (fifth) value when arranged in numerical order, JPY 150K

**Question 2.** What is the standard deviation?

**Answer 2.** Standard deviation = square root of the variance = = 52.9K

**Question 3.** What is the standard error?

**Answer 3.** Standard error = standard deviation/= 52.9K / = 17.6K

**Question 4.** Find the 95% confidence interval for the mean.

**Answer 4.** 95% confidence interval using the t-distribution = sample mean ± 2.306 × standard error

= 170K ± 2.306 × 17.6K

= (JPY 130K, 210K)

**Question 5.** Assuming the national average of monthly cost for the pharmaceutical treatment of disease X is known to be JPY 120K in Japan, how do you interpret the treatment cost for this patient group?

**Answer 5.** The obtained 95% confidence interval does not contain the national average JPY 120K. That is, a significant difference is observed at the 5% significance level. The mean monthly treatment cost of JPY 170K reported for this patient group is therefore different from the national average.

**Question 6.** Would it be appropriate to use the national average, JPY 120K, as an estimate in the base case analysis and the 95% confidence interval determined in Question 4 as the range of variation in sensitivity analysis?

**Answer 6.** When using the national average of JPY 120K as an estimate in the base case analysis, it is not appropriate to use the range (JPY 130K, 210K) as the range of variation, since a significant difference is observed. However, when using the mean monthly treatment cost of JPY 170K reported for this patient group as an estimate in the base case analysis (assuming this is appropriate), it is acceptable to use the 95% confidence interval (JPY 130K, 210K) for the range of variation of the mean.

**Epidemiological Survey**

**Example**

In an epidemiological survey of a region, the percentage of smokers diagnosed with lung cancer was 70% in a group of 40 patients and 40% in a group of 30 subjects without lung cancer.

**Question 1.** What type of study is this?

**Answer 1.** A case-control study.

**Question 2.** What is the risk ratio?

**Answer 2.** A case-control study is a retrospective study, in which the risk of onset cannot be determined. Therefore, the risk ratio (relative risk) cannot be estimated either.

**Question 3.** What is the odds ratio?

**Answer 3.** A two-by-two table can be drawn as follows:

|  |  |  |
| --- | --- | --- |
|  | Lung cancer | No lung cancer |
| Smoker | 28 (a) | 12 (b) |
| Non-smoker | 12 (c) | 18 (d) |
|  | 40 | 30 |

Odds ratio = *a* × *d/b* × *c*

= (28 × 18)/(12 × 12) = 3.5

**Question 4.** Find the 95% confidence interval for this odds ratio.

**Answer 4.** The 95% confidence interval of the odds ratio is

OR \* *e*±1.96 V where *V*  = 1/28 + 1/12 + 1/12 + 1/18 = 0.258

= 3.5 × *e*±1.96 0.258

= (1.29, 9.47)

**Question 5.** Is it appropriate to conclude that smoking is associated with lung cancer?

**Answer 5.** Since the 95% confidence interval obtained in Question 4 does not include 1, the odds ratio of 3.5 is statistically significant at the 5% significance level. Consequently, smoking is associated with lung cancer.

However, as pointed in Question 2, the risk for developing lung cancer from smoking cannot be estimated in a case-control study, and thus the odds ratio of 3.5 should not be interpreted as a risk ratio of 3.5. It is also incorrect to look at the two-by-two table by row and understand that 28 of 40 smokers developed lung cancer, thereby estimating that the incidence of lung cancer from smoking is 70% (= 28/40). Where the prevalence is low, however, the odds ratio is approximated to the risk ratio.

Since modeling in economic evaluations generally uses approaches such as decision trees and Markov models, which are used for analysis in prospective studies, it is not appropriate, in principle, to use evidence from case-control studies.

**Sample Size Calculation (1)**

**Example**

The average monthly cost per patient for the treatment of disease X is known to be JPY 1M in the standard treatment group. You wish to demonstrate that the average monthly cost of JPY 1.1M or more per patient receiving a new treatment is statistically significantly higher.

**Question 1.** Give reasons why sample size determination is required.

**Answer 1.** Reasons may include the following: to minimize the number of subjects, minimize the cost required for the study, increase the feasibility of the study, determine the sufficient level of statistical significance (for the evaluation of both effectiveness and cost), and design a subgroup analysis.

**Question 2.** Is it possible to calculate the needed sample size?

**Answer 2.** Not possible.

**Question 3.** What additional information is required to determine the required sample size?

**Answer 3.** The following four factors need to be considered:

• What is the probability of type I error, α (significance level)?

• What is the statistical power (1 − *β*)?

• What is the difference in means between the two groups (*μ*1 − *μ*0)?

• What is the standard deviation *σ* of the population?

As the two means are already given, information on the standard deviation, significance level, and statistical power is additionally required. For instance, a question such as the following would allow for the calculation of sample size: “The average monthly cost per patient for the treatment of disease X is known to be JPY 1M in the standard treatment group, and *the standard deviation, JPY 0.24M*. You wish to demonstrate that the average monthly cost of JPY 1.1M or more per patient receiving a new treatment is statistically significantly higher. Given *the significance* *level of 5%* and *the power of 90%*, what is the sample size needed?”

**Question 4.** Based on the answer for Question 3, calculate the required sample size.

**Answer 4.** By substituting the values, *μ*0 = 1M, *σ* = 0.24M, *μ*1 = 1.1M,

*α* = 0.05 → *Z*α = 1.96,

1 − *β* = 0.90 → *β* = 10% → *Z*β = −1.28,

into the formula for sample size calculation:

Therefore, the sample size is 61.

**Sample Size Calculation (2)**

**Example**

You wish to compare the standard treatment group for a disease with that of another group treated with a new drug X in a randomized clinical trial cost-effectiveness analysis. Given that the difference in average monthly cost of treatment per patient between the two groups is considered statistically significant if the difference is JPY 50K or greater, what is the sample size needed for a statistical test?

**Question 1.** Is it possible to calculate the sample size needed?

**Answer 1.** The sample size cannot be calculated without information regarding the standard deviation, significance level, and statistical power.

**Question 2.** Given the standard deviation of JPY 120K for both groups, significance level of 5%, and power of 80%, what is the needed sample size?

**Answer 2.** The following formula can be used to determine the total sample size n needed for the two groups, based on the distribution of the difference in mean between the two groups (*Δ*):

By substituting the values, *Δ* = 50,000, *σ* = 120,000, the statistical power of 80% → *Z*β = −0.84, and the significance level of 0.05 → *Z*α = 1.96 into the above formula:

Therefore, the sample size is 91.

**Diagnostic Tests and Decision Tree for Lung Cancer**

**Example**

Assume that 18 of the 20 individuals suspected of lung cancer based on a computed tomography (CT) turned out to truly have lung cancer as a result of pathological examination, whereas the 2000 individuals diagnosed as free of lung cancer based on the same CT scan included four overlooked patients that had lung cancer. Given this, answer the following questions:

**Question 1.** Draw a two-by-two table based on the two factors, test results, and presence or absence of lung cancer.

**Answer 1.** Two by two table for lung cancer test

****

**Question 2.** What are the sensitivity and the specificity of this CT scan?

**Answer 2.** Sensitivity = probability of a positive test given that the patient has lung cancer = 18/(18 + 4) = 82%.

Specificity = probability of a negative test given that the patient is free of lung cancer = 1996/(2 + 1996) = 99.9%.

Sensitivity and specificity are key test characteristics in laboratory medicine.

**Question 3.** What are the positive and negative predictive values?

**Answer 3.** Positive predictive value = probability that the patient truly has lung cancer given a positive test = 18/20 = 90%.

Negative predictive value = probability that the patient is truly free of lung cancer given a negative test = 1996/2000 = 99.8%.

Positive and negative predictive values are key test characteristics in medical diagnosis.

**Question 4.** If you were to integrate this CT scan into the analytical modeling in an economic evaluation using a decision tree, what would the tree look like?

**Answer 4.** Positive and negative predictive values are assigned to the respective chance nodes following the bifurcation into positive and negative tests. One should be careful not to assign sensitivity and specificity here.



**Number Needed to Treat (NNT)**

**Example**

The results of a clinical trial indicated efficacy rates of 60% and 55% for a new drug and a control drug, respectively. Given this, answer the following questions:

**Question 1.** What is the number needed to treat (NNT), and what does it indicate?

**Answer 1.** The number needed to treat (NNT) is defined as the reciprocal of the difference in efficacy between the two groups. Thus, 1/(0.60 − 0.55) = 20.

This value indicates that, on average, 1 out of 20 patients will benefit from switching to the new drug. Of course, whether a drug is effective in 1 patient is a stochastic phenomenon, and cases where the drug is effective in all or none of the 20 patients are also possible in reality. Therefore, NNT expresses the difference in the efficacy rate of 5% as an outcome measure of switching in terms of the number of patients. Since a difference in efficacy of 10% gives an NNT of 10, the difference needs to be above 10% to obtain a single-digit NNT.

NNT can also be explained in terms of probability. The phenomenon of a drug being effective or not is generally modeled on the “heads or tails” outcome of tossing a coin. The binomial distribution B(*n*, *p*) is a typical probability distribution, in which the average number of heads that appear in *n* number of tosses can be obtained by *n* × *p*, where *p* is the probability of obtaining head on a given trial. Based on this binomial distribution, the average number of patients that will benefit from switching to the new drug = *n* × *p*

= NNT × (1/NNT)

= 1,

where *p* is the probability of the patient benefiting from switching to the new drug from the control drug and NNT the number of patients switching to the new drug.

**Question 2.** Does the NNT obtained in Question 1 have any clinical significance?

**Answer 2.** The smaller the NNT, the more effective the drug is compared to the control drug, although there is no absolute clinical standard for how small the NNT should be. An NNT of 30 or lower is considered “small” enough in some cases, in which the NNT of 20 for the new drug is a reasonable value. However, patients always expect the ideal NNT of 1 and may misunderstand the drug as having little effect when given a rough explanation such as “only 1 out of 20 patients treated with the new drug benefits from it.” David Sackett and colleagues, who advocated evidence-based medicine, suggested that the NNT in terms of number of patients is a better measure compared with the probability of patients understanding the benefit of switching treatments. However, it should be noted that a simple conversion of efficacy to the number of patients without the underlying concept of probability could cause confusion.

**Question 3.** Give one example of where the NNT would take the same value as the one obtained in Question 1.

**Answer 3.** Different pairs of efficacy rates with the same difference between the respective two groups would give the same NNT. As an example of lower efficacy rates, a new and control drug with efficacy rates of only 15% and 10%, respectively, would also give an NNT of 1/(0.15 − 0.10) = 20. In other words, the NNT has a limitation in that it does not reflect the baseline efficacy rate of treatment.

**Question 4**. Given that the sample size is 1000 for both groups, what is the 95% confidence interval of the NNT?

**Answer 4.** To find the 95% confidence interval of the NNT, we first calculate the 95% confidence interval of the difference in efficacy between the two groups using the following formula:

the 95% confidence interval of the difference in efficacy

between the two groups = (*p*1 – *p*2) ± 1.96 ,

where *V* = *p*1(1 – *p*1)/*n* + *p*2 (1 – *p*2)/*m*. Here, *p*1 efficacy of the new drug, *p*2 efficacy of the control drug, *n* number of patients in the new drug group, and *m* number of patients in the control drug group. Thus,

*V* = 0.6 × 0.4 /1000+ 0.55 × 0.45/1000 = 0.0004875;

95% confidence interval of the difference in efficacy between the two groups

= (0.60 – 0.55) ± 1.96

= (0.00672, 0.0933)

By taking the reciprocal of the upper and lower limits, the 95% confidence interval of the NNT is calculated to be (1/0.0933, 1/0.00672) = (11 patients, 148 patients).

**Question 5.** What would the 95% confidence interval be if the sample size were 100 for both groups?

**Answer 5.** Similar to the previous question, *V* = (0.6 × 0.4/100) + (0.55 × 0.45/100) = 0.004875

Thus,

95% confidence interval = 0.05 ± 1.96 = (–0.0868, 0.187).

In this case, the 95% confidence interval of the difference in efficacy between the two groups includes zero, meaning there is no statistically significant difference between the two groups. Therefore, there is no clinical significance in obtaining a 95% confidence interval of the NNT that includes negative values.

**Question 6.** Given that the treatment cost per patient is JPY 400K and 300K for the new and control drugs, respectively, what is the average treatment cost per patient for whom the treatment is effective in the two groups?

**Answer 6.** Given a sample size of 1000 for both groups, the new drug group requires JPY 400K × 1000 for 600 patients to benefit from the treatment, and the control drug group requires JPY 300K× 1000 for 550 patients to benefit. Thus, the average treatment cost per patient is

400,000K/600 = JPY 667K for the new drug group and

300,000K/550 = JPY 545K for the control drug group.

**Question 7.** Based on the treatment costs given in the previous question, what is the incremental cost-effectiveness ratio (ICER)?

**Answer 7.** ICER = difference in cost between the two groups/difference between the two groups in the number of patients for whom the respective treatments were effective

= (400,000K – 300,000K)/(600 − 550)

= JPY 2M per patient for whom the treatment is effective.

**Question 8.** What is the relationship between the ICER and NNT?

**Answer 8.** The ICER in Question 7

= (400,000K – 300,000K)/(600 – 550)

= (400,000 – 300,000) × {1/(0.6 – 0.55)}

= difference in treatment cost per patient between the two groups × NNT.

That is, the ICER is the product of the incremental cost per patient and NNT.