

Principles of Inferential Statistics in Medicine

Midterm Exam – EPIB–607, October, 2002.

1. Zyban is a treatment prescribed to help smokers quit smoking. Suppose that this drug has a one year quit rate of 30%. In other words, suppose that 30% of smokers prescribed Zyban quit and are still not smoking one year post-treatment.

(a) If 10 smokers are prescribed Zyban, what is the probability that 2 or more of them will not be smoking one year post-treatment?

(b) If 1000 smokers are prescribed Zyban, what is the probability that 200 or more of them will not be smoking one year post-treatment?

2. Suppose that a researcher has just conducted a study, and found a p -value of $p = 0.10$ in testing the null hypothesis of no difference between two treatments. Encouraged by this “almost significant” result, the researcher decides to replicate the study using an identical design and sample size as the study just completed. Suppose that, unknown to the researcher, the null hypothesis is in fact true. What is the probability that the second study will provide a p -value of less than $p = 0.05$?

3. Suppose that $X \sim N(5, 25)$. In other words, suppose that X is a random variable that follows a Normal distribution with mean 5 and variance 25. What is $Pr\{X \leq 0\}$?

4. For a certain disease A, a daily dosage of 30 milligrams (mg) is prescribed for mild cases, a daily dosage of 60 mg is prescribed for moderate cases, while severe cases are prescribed 75 mg daily. Suppose that a clinical practice contains a mix of patients with disease A, such that 30% are mild cases, 50% are moderate cases, and 20% are severe cases. What should the average daily dose be among this group of patients?

5. While it is well known that increased exercise reduces the probability of cardiovascular disease (CVD), people often need to be motivated to change their exercise habits. One potential motivator may be to provide patients with information about their current risk of CVD, and compare it to what their risk would become under an exercise program. To test whether providing these two numbers to a patient increases the amount of exercise done per week, a randomized clinical trial is carried out. One hundred subjects are randomized, with 50 given their risk of CVD on and off the exercise program (treatment group), and the other 50 patients advised to exercise, but not given their

specific risk information (control group). The number of hours exercised in each group is then tracked. In the treatment group, the 50 subjects exercised an average of 4 hours per week, with a standard deviation also of 4 hours, while the control group exercised an average of 3.5 hours per week, with a standard deviation of 4 hours.

Test the null hypothesis that the risk information changes the amount of exercise done per week, versus a two-sided alternative. State the null and alternative hypotheses, carry out the test, provide a p -value, and state your conclusion.

6. In large cancer research studies, there are two ways to ascertain whether a subject has a given type of cancer or not. Method 1, chart review, is known to correctly find 70% of subjects who truly have this cancer, while falsely identifying only 1% of those truly free from cancer as having the disease. Similarly, Method 2, using a cancer registry, correctly finds 80% of subjects who truly have this type of cancer, while falsely identifying 2% of those truly free from cancer as having the disease. Suppose that the prevalence of this cancer in a given region is 1%. If a subject is identified by both methods of ascertainment as having the disease, and if these two methods operate independently from each other given the true cancer status of each patient, what is the probability that this subject truly has this type of cancer?