

15.075 Exam 2

Instructor: Cynthia Rudin

TA: Dimitrios Bisias

October 25, 2011

Grading is based on demonstration of conceptual understanding, so you need to show all of your work.

Problem 1

You are in charge of a study that compares how two weight-loss techniques (Diet and Exercise) affect the weight loss of overweight patients. The table below shows the number of obese and severely obese people who lost significant weight (successes) and the number of people who didn't lose any weight (failures) for both of the weight-loss techniques.

	Diet successes	Diet failures	Exercise successes	Exercise failures
Obese	10	30	22	58
Severely Obese	60	20	35	5

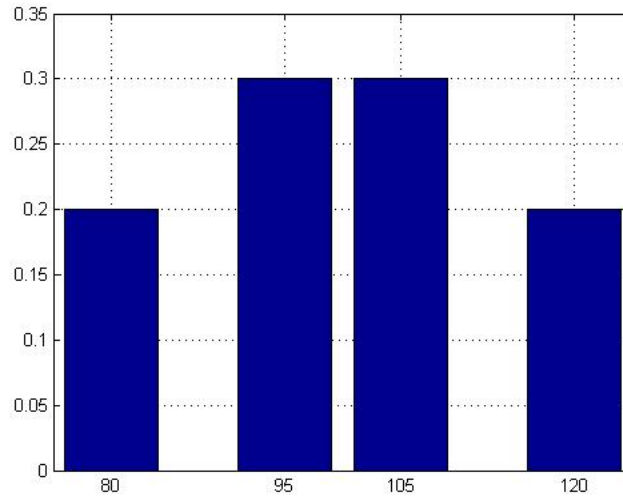
1. Within each category of obesity, compare success rates for the weight-loss techniques. What do these rates indicate?
2. Compare the total success rates for the two techniques. Explain any differences from (1).

Solution

1. For the obese category the success rate for diet is $10/40 = 25\%$ while the success rate for exercise is $22/80 = 27.5\%$. For the severely obese category, the success rate for diet is $60/80 = 75\%$ while the success rate for exercise is $35/40 = 87.5\%$. These figures show that exercise has higher success rates for both of the categories.
2. The total success rate for diet is $70/120 = 58.3\%$ while the total success rate for exercise is $57/120 = 47.5\%$. This contrasts with the result of part (1). This is an example of Simpson's paradox. The lurking variable is type of obesity. There are far less obese people in case of diet than exercise, who have the lower success rates.

Problem 2

You are the owner of a bakery selling cheesecakes. Some days you sell a lot of cheesecakes and some other days you don't, independently of other days. In fact, the distribution of the number of cheesecakes you sell each day is shown below.



If we measure cheesecakes sales for 100 days, what is the approximate probability that the average cheesecakes sales (over the 100 days) will be more than 102 cheesecakes?

Solution

Let X be the number of cheesecakes sold in one day. Then due to symmetry $E[X] = 100$ and

$$\text{var}(X) = 2 \cdot 0.2 \cdot 20^2 + 2 \cdot 0.3 \cdot 5^2 = 175$$

We are asked to find $P(\bar{X} > 102)$. From CLT we know that \bar{X} is approximately normally distributed with mean 100 and variance $\frac{175}{n} = 1.75$, therefore we have:

$$P(\bar{X} > 102) \approx P\left(Z > \frac{102 - 100}{\sqrt{175/100}}\right) = 1 - \Phi(1.51) = 0.065$$

Problem 3

Suppose that the observations X_1, X_2, \dots, X_n are iid with common mean θ and known variance σ^2 . Consider the following estimator of the mean θ :

$$\hat{\Theta}_n = \frac{X_1 + \dots + X_n}{n+1}.$$

1. What is the bias of this estimator? What happens to the bias as we increase the size of the sample?
2. What is the MSE of the estimator?

Hint: Remember to use either what you know about $Var(\bar{X})$, or what you know about the variance of the sum of independent random variables.

Solution

1. It is: $\hat{\Theta}_n = \frac{n}{n+1}\bar{X}_n$. Therefore,

$$E[\hat{\Theta}_n] = \frac{n}{n+1}E[\bar{X}_n] = \frac{n}{n+1}\theta$$

and

$$var[\hat{\Theta}_n] = \frac{n^2}{(n+1)^2}var[\bar{X}_n] = \frac{n}{(n+1)^2}\sigma^2$$

Thus, we have:

$$Bias = E[\hat{\Theta}_n] - \theta = \frac{n}{n+1}\theta - \theta = \frac{-1}{n+1}\theta$$

As $n \rightarrow \infty$ the bias goes to 0.

2. It is:

$$MSE(\hat{\Theta}_n) = Bias^2 + var(\hat{\Theta}_n) = \frac{\theta^2}{(n+1)^2} + \frac{n}{(n+1)^2}\sigma^2$$

Problem 4

The weight of an object is measured eight times using an electronic scale that reports the true weight plus a random error that is normally distributed with zero mean and variance $\sigma^2 = 4$. Assume that the errors in the observations are independent. The following results are obtained:

{4.45, 4.02, 5.51, 1.10, 2.62, 2.38, 5.94, 7.64}

Compute a 95% confidence interval for the true weight.

Solution

A 95% confidence interval for the true weight is

$$\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right] = \left[4.21 - 1.96 \frac{2}{\sqrt{8}}, 4.21 + 1.96 \frac{2}{\sqrt{8}}\right] = [2.82, 5.59]$$

Problem 5

The returns of an asset management firm in different years are independent and normally distributed with unknown mean and variance. The asset management firm claims that the standard deviation of the returns is as low as $\sigma = 2\%$ and the mean of the returns is $\mu = 22\%$. You believe that this is too good to be true. To verify your suspicion, you take the returns from the last 10 years. These are:

$$r = \{20.6, 19.2, 17, 19.1, 18.7, 22.5, 27.2, 17.9, 22.5, 21.3\}$$

(To help you with the calculations, we give you that: $\sum_{i=1}^n r(i) = 206$ and $\sum_{i=1}^n (r(i) - \bar{r})^2 = 79.34$)

1. Calculate the sample mean and standard deviation of the above random sample.
2. What is the probability that when we draw a new random sample of size 10, its sample mean will be below the one you calculated in part 1? Assume that the company's claims are correct.
3. What is the probability that when we draw a new random sample of size 10, its sample standard deviation will be larger than the one you calculated in part 1, assuming that the company's claims are correct?

Solution

1. The sample mean is: $\bar{r} = \frac{\sum_{i=1}^n r(i)}{n} = 20.6$ and the sample standard deviation is $s = \sqrt{\frac{\sum_{i=1}^n (r(i) - \bar{r})^2}{n-1}} = \sqrt{\frac{79.34}{9}} = 2.97$
2. Under the company's claims the sample mean is normally distributed with mean 22% and standard deviation $2/\sqrt{10}\%$. It is:

$$P(\bar{X} < 20.6) = P(Z < \frac{20.6 - 22}{2/\sqrt{10}}) = P(Z < -2.21) = 0.013$$

3. Under the company's claims the statistic:

$$\frac{(n-1)S^2}{\sigma^2}$$

is distributed as a chi-square with $n-1=9$ degrees of freedom. Therefore:

$$P(S > 2.97) = P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{9 \cdot 2.97^2}{4}\right) = P(\chi_9^2 > 19.85) = 0.0188$$

MIT OpenCourseWare
<http://ocw.mit.edu>

15.075J / ESD.07J Statistical Thinking and Data Analysis
Fall 2011

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.