# 1 Central Limit Theorem

1. Suppose the average male height in the US is 69.5 inches with a standard deviation of 3 inches. Given a random sample of 100 men, what is the chance that their average height is less than 69.2 inches?

2. During a year, an average US family uses 11,000 kWh of electricity. Suppose the standard deviation of use is 5,000 kWh. In a sample of 625 homes, what is the probability that the average electrical consumption is more than 10,600 kWh?

3. (calculator required) Suppose that the average daily rainfall in Berkeley in December is .16 inches per day with a standard deviation of .09 inches. Assume that the amount of rain each day is independent. Over the course of the month of December, what is the chance that the total rainfall is less than 4.5 inches?

4. A toy company's manufacturing plant has a 1% defective rate: this means that each toy has an independent 1% chance of being defective. What is the probability that, in a load of 10000 toys, more than 130 are defective?

# 2 Hypothesis Testing

In each of the following situations, write down what $H_a$ and $H_0$ would be. Also write down what Type I and Type II errors would be in each situation.

1. Researchers wish to determine whether a particular hormone affects the appetite of mice, so they select 49 mice and record how much the mouse eats after being given the hormone. Regular healthy mice eat 10 g of food per day with a standard deviation of 3 g per day.

2. A particular disease has an average recovery time of 27 days with a standard deviation of 2 days when someone receives the usual antibiotics. Researchers wish to test a new kind of antibiotic to see if it speeds up the recovery. They test it on 36 random participants with the disease and record how long it takes them to recover.

3. A particular kind of tree grows an average of 1 inch per year with a standard deviation of .4 inches. After adding a chemical supplement to the soil around a crop of 100 of these trees, researchers measure the average growth over the course of a year and wish to determine if the supplement affected the growth rate.

4. (calculator required) Researchers wish to know if more than 30% of people approve of President Obama so they sample 500 random Americans and ask them if they approve of President Obama.

For each situation in the previous part, use the following collected data to answer the researchers' questions using $\alpha = .05$. Write down what the $p$-value is.

1. The average appetite of the 49 mice is 10.43 grams per day.

2. The average recovery time of the 36 participants is 26 days.

3. The average growth of the 100 trees is 1.02 inches in the year.

4. Of the 500 americans, 170 say they approve of President Obama.

**Some Solutions**

Central Limit Theorem #1: Let $X$ be the average height of the men in the sample (in inches). Then by the central limit theorem, $X \sim N(69.5, 3/\sqrt{100}) = N(69.5, .3)$. And

$$P(X < 69.2) = P\left(\frac{X - 69.5}{.3} < \frac{69.2 - 69.5}{.3}\right) = P(Z < -1) \approx .16.$$

Hypothesis testing #1: Let $f$ be the amount of food a mouse eats in a day. Then $H_a$ is $f \neq 10$ and $H_0$ is $f = 10$. A type I error would be (recall this is when $H_0$ is true but you reject it) when the hormone doesn't change the mouse's appetite but your data suggests it does. A type II error would be (recall this is when $H_a$ is true but your data doesn't give you the evidence saying so) when the mouse appetites are affected by the hormone but your data doesn't give you evidence for this. Let $X$ be the average amount eaten by the sample of 49 mice. Then the CLT says that $X \sim N(10, 3/\sqrt{49}) = N(10, 3/7)$. Since the alternative is two-sided, the $p$-value is

$$P(X > 10.43) + P(X < 9.57) = P\left(\frac{X - 10}{3/7} > \frac{10.43 - 10}{3/7}\right) + P\left(\frac{X - 10}{3/7} < \frac{9.57 - 10}{3/7}\right)$$
$$\approx P(Z > 1) + P(Z < -1) \approx .32.$$

Since the $p$-value is bigger than .05 we don't have evidence to show that the hormone affects the mouse diets.

Hypothesis testing #2: Let $r$ be the recovery time from the disease with the new antibiotics. Then $H_a$ is $r > 27$ and $H_0$ is $r = 27$. I'll let you figure out the type I and type II errors. Let $X$ be the average recovery time of a sample of 36 individuals with the new antibiotic. By the CLT, $X \sim N(27, 2/\sqrt{26}) = N(27, 1/3)$. Because this is a one-sided alternative then $p$-value is

$$P(X < 26) = P\left(\frac{X - 26}{1/3} < \frac{26 - 27}{1/3}\right) = P(Z < -3) \approx .003.$$

Since this $p$-value is less than $\alpha = .05$ we have evidence that the alternative hypothesis is true, i.e. that the new antibiotic does speed up recovery.