STAT 88: Lecture 27

Contents

Last time

Normal Approximation:

The normal distribution is famous because of the Central Limit Theorem which applies to sums and averages of i.i.d. random variales.

For example, a binomial(n, p) is a sum of n i.i.d. indicators and so is approximately normal for large n.

To compute areas under the normal curve, we convert the normal curve to the standard normal curve,

$$Z = \frac{X - E(X)}{\mathrm{SD}(X)},$$

and use the standard curve CDF $\Phi(z)$.

Warm up: Suppose that each of 300 patients has a probability of 1/3 of being helped by a treatment independent of its effect on the other patients. Find approximately the probability that more than 134 patients are helped by the treatment.

9.1. Testing Hypotheses

The Speed of Light You wish to test 299,792.458 km/sec is the speed of light.

Suppose 150 such measurements on the speed of light have an average of 299,796 km/sec and an SD of 50 km/sec. Are these data consistent with the model that measurements are i.i.d. with mean equal to the currently accepted value of $c = 299,792.458$? Or are they too big?

Steps

(a) State an appropriate null hypothesis in informal terms and also in terms of random variables.

Null hypothesis

H_0: $X_1, \cdots, X_{150}$ are i.i.d. with $E(X_i) = c$ (= 299,792.458)

(b) State an appropriate alternative hypothesis.

Alternative hypothesis

H_A: measurements are too big to be consistent with null, $c > 299,792.458$

(c) What test statistic do you want to use? Justify your choice.

$\bar{x}$ is a natural choice of test statistic.

You favor H_A if you get a large value of $\bar{x}$

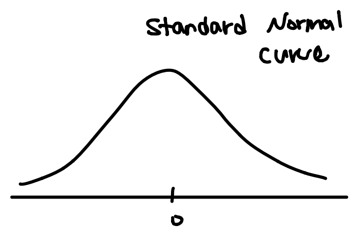
(d) Find the p-value of the test, exactly if possible of approximately if it is not possible to get an exact answer.

The p-value is the chance, assuming that the null hypothesis is true, of getting a test statistic equal to the one that was observed or even more in the direction of the alternative.

- We assume the null is true, i.e. $E(X_i) = 299\,792.458$.
- To find p-value, we need to know the distribution of test-statistic under the null !!
- By CLT, $\bar{X} \approx N(E(\bar{X}), Var(\bar{X}))$,

  Under H_0, $E(\bar{X}) = E(X_i) = 299{,}792.458$ and

  $$Var(\bar{X}) = \frac{Var(X_i)}{150} \approx \frac{50^2}{150} = 16.67$$
- Observed value $= 299{,}796.000$
- p-value $= P(\bar{X} > 299{,}796)$ under H_0



(e) At the 5% level, what is the conclusion of the test? Why?

Example: (Exercise 9.5.1) All the patients at a doctor's office come in annually for a check-up when they are not ill. The temperatures of the patients at these check-ups are independent and identically distributed with unknown mean μ.

The temperatures recorded in 100 check-ups have an average of 98.2 degrees and an SD of 1.5 degrees. Do these data support the hypothesis that the unknown mean μ is 98.6 degrees, commonly known as "normal" body temperature? Or do they indicate that μ is less than 98.6 degrees?

(a) State H_0.

(b) State H_A.

(c) Find test statistic.

(d) Find p-value.

(e) Conclusion (5% level).

Example: (Exercise 9.5.2) One of Gregor Mendel's models was about a type of pea plant that is either tall or short. His model was that each such plant is short with chance 1/4, independently of all other plants. In the plants that he bred, he observed 787 tall ones and 277 short ones. Do the data support his model? Or do they indicate that the model is not good? Make a decision in the following steps.

(a) State H_0.

(b) State H_A.

(c) Find test statistic.

(d) Find p-value.

(e) Conclusion (5% level).