Stat 88 lec 27

Warmup 2:00 - 2:10

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.

$n = 300$

$p = 1/3$

X = # patients helped

$\mu = np = 300(1/3) = 100$

$\sigma = \sqrt{npq} = \sqrt{300\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)} = 8.2$

$X \sim \text{Binomial}(300, 1/3) \approx N(\underbrace{100}_{\mu}, \underbrace{8.2^2}_{\sigma^2})$

$P(X \geq 135) \approx 0$ since there is data > 4 SDs from the mean.

$135 = 100 + 35 = k(8.2)$, $k > 4$

$1 - \Phi\left(\frac{135-100}{8.2}\right) \approx 0.$

"large enough" means

$\mu + 3\sigma \leq 300$ and $\mu - 3\sigma \geq 0$

$\mu = 100$, $3\sigma = 24.6$

✓ ✓

last time

sec 8.3, 8.4 **Normal Approximation**

The normal distribution is famous because of the CLT which applies to sums and averages of iid RVs.

For example a binomial (n, p) is a sum of n iid 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,

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

, and use the cdf $\Phi(z)$.

Recall,

SD of sample sum

$X_1, \ldots, X_n \overset{iid}{\sim}$ with mean μ, SD σ

$S_n = X_1 + \cdots + X_n$

$E(S_n) = n\mu$

$SD(S_n) = \sqrt{n}\,\sigma$

SD of sample average

$A_n = S_n / n$

$E(A_n) = \mu$

$SD(A_n) = \frac{\sigma}{\sqrt{n}}$

Today

sec 9.1

1) Hypothesis testing

① Sec 9.1 Testing Hypotheses

eg Speed of light

You wish to test $c = 299,792.458$ km/sec is the speed of light.

Suppose 150 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: $X_1, .., X_{150}$ are iid with $E(X_1) = c = 299,792.458$

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 T.S.

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

$\bar{X}$ is approx normal.

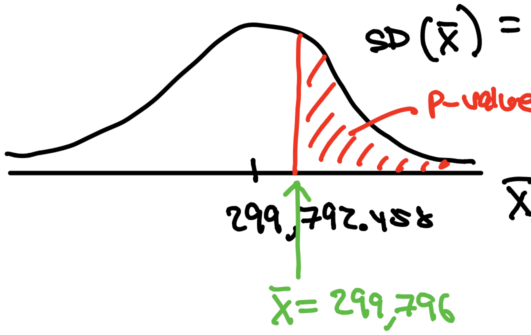
d) Find the *p*-value of the test, exactly if possible or approximately if it is not possible to get an exact answer.

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. this specifies the distribution of $\bar{X} \approx N\left(E(\bar{X}),\ SD(\bar{X})^2\right)$

$$E(\bar{X}) = E(X_1) = 299{,}792.458$$

$$SD(\bar{X}) = \frac{SD(X_1)}{\sqrt{150}} = \frac{50}{\sqrt{150}} = 4.08$$



$SD(\bar{X}) = 4.08$

p-value $= P(\bar{X} > 299{,}796)$

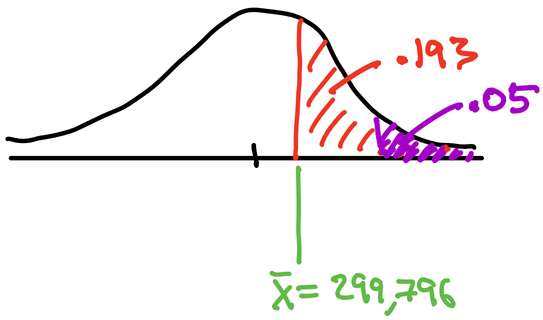
$$= 1 - \Phi\left(\frac{299{,}796 - 299{,}792.458}{4.08}\right)$$

$$\frac{299{,}796 - 299{,}792.458}{4.08} = .87$$

$$= \boxed{.193}$$

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

accept the null since $.193 > .05$



T.S. not extreme enough to reject the null.

exercise 9.5.1

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) H_0 Temperatures $X_1, \dots, X_n$ are iid with $\mu = 98.6$

b) H_A $\mu < 98.6$ (measurements are too small to be consistent w/ null).

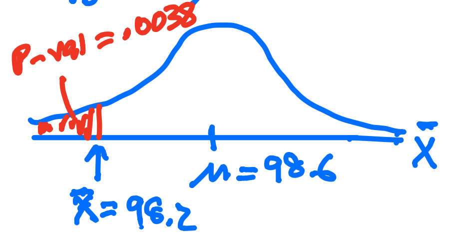
c) T.S. $\bar{X} = 98.2$, small values support H_A.

$SD(\bar{X}) = \frac{1.5}{\sqrt{100}}$

d) P-value $\text{P-val} = P(\bar{X} \le 98.2) = \Phi\left(\frac{98.2 - 98.6}{.15}\right) = .0038$

e) Conclusion (5% level)

$.0038 < .05 \Rightarrow$ reject null (accept alternative)



exercise 9.5.2

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) H_0 Mendel's model that plants are 1064 iid Bernoulli (.25) RVs is good.

b) H_A Mendel's model isn't good

c) T.S. let X = number of short plants $X = X_1 + \cdots + X_{1064}$

under H_0, $X \sim$ Binomial $(1064, .25)$, $E(X) = 1064(.25) = 266$

use X or $\bar{X}$ as T.S. (reject if $T = |X - 266|$ or $T = |\bar{X} - .25|$ is large)

d) p-val

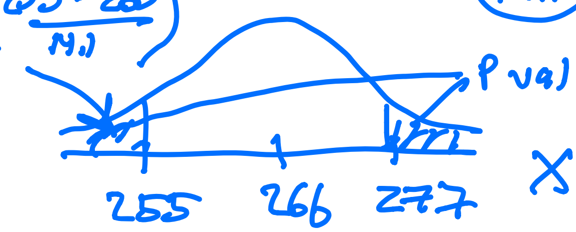
$$\text{p-val} = P(|X - 266| \geq \overset{277-266}{11})$$

$$= P(X \geq 277) + P(X \leq 255)$$

$$\approx 2\Phi\left(\frac{255-266}{14.1}\right) = \boxed{.44}$$

$SD(X) = \sqrt{1064(.25)(.75)} = 14.1$

$\Phi\left(\frac{255-266}{14.1}\right)$



e) conclusion (5% level)

accept null since $.44 > .05$

if use $T = |\bar{X} - .25|$

$$SD(\bar{X}) = \frac{\sigma}{\sqrt{1064}} = \frac{\sqrt{(.25)(.75)}}{\sqrt{1064}} = \boxed{.013}$$

$$\text{p-val} = P(|\bar{X} - .25| \geq .01)$$

$$= 2\Phi\left(\frac{.24 - .25}{.013}\right) = .44$$