Stat 88 lec 28

Warmup: 2:00–2:10

A population distribution is known to have an SD of 20. Determine the p-value of a test of the hypothesis that the population mean is equal to 50, if the average of a sample of 64 observations is 55.

$H_0: \mu = 50$

$H_1: \mu \neq 50$

T.S. is $T = |\bar{X} - 50|$ — we reject null if this is large.

$T = |55 - 50| = 5$

$\text{P-val} = P(T \geq 5) = P(|\bar{X} - 50| \geq 5)$

$= P(\bar{X} \geq 55) \text{ or } P(\bar{X} \leq 45)$

$= 2\Phi\left(\frac{45-50}{2.5}\right) = \boxed{.0455}$

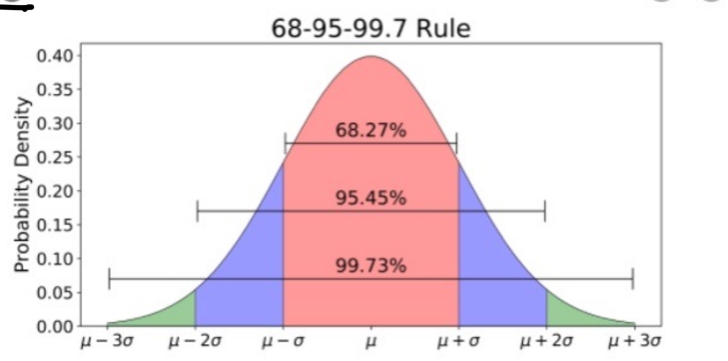
$SD(\bar{X}) = \frac{\sigma}{\sqrt{64}} = \frac{20}{8} = 2.5$

Conclusion of test at level .05 is ?

$.0455 < .05 \Rightarrow$ reject the null.

Last time

Sec 8.3 Empirical rule:



Sec 9.1 Hypothesis testing

Hypothesis testing has 5 steps:

1) H_0 — a specific distribution

ex $X_1, \dots, X_{100}$ are iid Bernoulli(.5)

2) H_A — one or two sided depending on the context of the problem.

ex $X_1, \dots, X_{100}$ are i.i.d. Bernoulli(p) for $p > .5$

3) T.S. — often $\bar{X}$ or a RV whose distribution we know so we can compute a p-value.

4) p-value — Chance of being as or more extreme than the value of our T.S. in the direction of the alternative.

5) conclusion accept null if p-value $\geq$ level of test.
reject null if p-value $<$ level of test

Today Sec 9.2 A/B testing: Fisher Exact Test
Sec 9.3 Confidence Interval

① sec 9.2 **A/B testing: Fisher Exact Test**

A/B test is shorthand for a simple controlled experiment

A = control grp }
B = treatment grp. }

ex (study for treatment of chronic back pain)

Sample of 31 patients with back pain

15 → treatment (botulinum toxin A) (9 of 15 pain relief)

16 → control grp (2 of 16 pain relief)

Is the treatment effective?

a) H_0 – treatment has no effect. (same as randomly assigning cured patients to each of 2 grps)

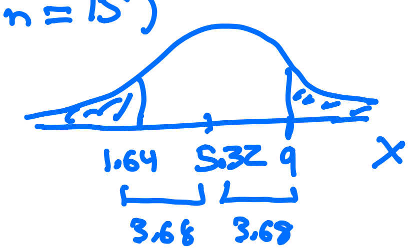
b) H_A – treatment has an effect on back pain (good or bad)

c) T.S. – $T = |X - 5.32|$ where

X = # of pain relieved treatment patients

$X \sim HG(N=31, G=11, n=15)$

$E(X) = 15 \cdot \frac{11}{31} = 5.32$



d) P-value we assume H_0.

$T = |X - 5.32| = |9 - 5.32| = 3.68$

$P\text{-val} = P(T \geq 3.68)$

$= P(X \le 1.64) + P(X \ge 9)$

$= P(X \le 1) + P(X \ge 9)$

$$= \sum_{g=0}^{1} \frac{\binom{11}{g}\binom{20}{15-g}}{\binom{31}{15}} + \sum_{g=9}^{11} \frac{\binom{11}{g}\binom{20}{15-g}}{\binom{31}{15}} = \boxed{.0092}$$

e) Conclusion (5% level)

$.0092 < .05 \Rightarrow$ reject null.

This is called the Fischer Exact Test.

exercise 9.5.9

9. A randomized controlled trial was conducted as part of a effort to encourage high school students from under-resourced communities to apply for college. The trial had 200 participants. A simple random sample of 95 participants received special coaching for the ACT. The remaining participants received no intervention.

At the end of the experiment, the participants got to decide whether or not they would take the ACT. Among the 95 students in the treatment group, 75 decided to take the test. Among the 105 students in the control group, 70 decided to take it.

Is the difference statistically significant? Answer this question by performing a test of whether or not the intervention had any effect.

a) H_0 treatment did nothing

b) H_A treatment did something good or bad.

c) T.S, $X = \#$ test taken in treatment group

assuming H_0, $X \sim HG(N=200, G=145, n=95)$

$E(X) = 95\left(\frac{145}{200}\right) = 68.875$

T.S is $T = |X - 68.875|$

d) P-val

we got $T = |75 - 68.875| = 6.125$

so p-val $= P(T \geq 6.125) = P(X \leq \underbrace{68.875 - 6.125}_{= 62.75})$

$+ P(X \geq 75)$

$$\Rightarrow \text{P-val} = \sum_{g=0}^{62} \frac{\binom{145}{g}\binom{55}{95-g}}{\binom{200}{95}} + \sum_{g=75}^{95} \frac{\binom{145}{g}\binom{55}{95-g}}{\binom{200}{95}}$$

e) Conclusion

$= \boxed{.285}$

$.285 > .05$ so accept null.

② sec 9.3 Confidence interval (CI) of μ

Suppose $X_1, \dots, X_n$ is iid with mean μ and SD σ

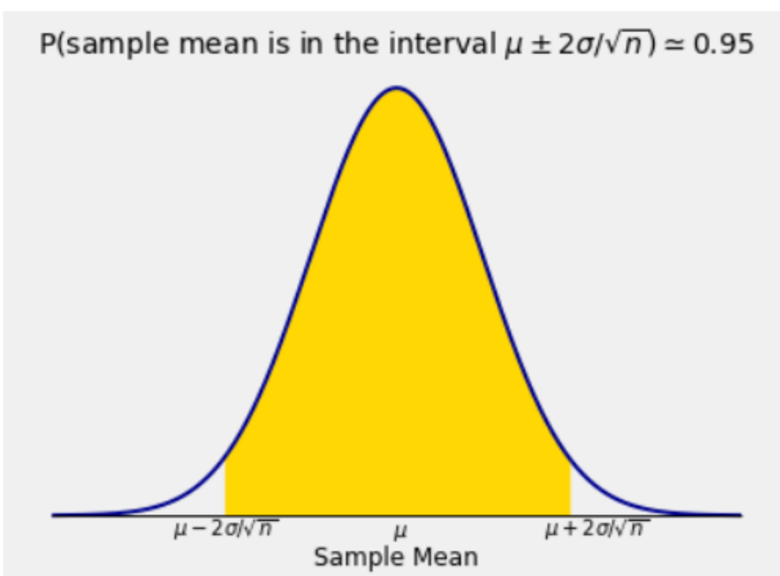
$\bar{X}$ is an unbiased estimator of μ (i.e $E(\bar{X}) = \mu$).

and $SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ is a measure of the average spread of $\bar{X}$.

A 95% CI is

$$\left(\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}\right)$$ and

may write as $\bar{X} \pm 2\frac{\sigma}{\sqrt{n}}$



Since $\bar{X} \approx N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right)$

$$P\left(\mu - 2\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 2\frac{\sigma}{\sqrt{n}}\right) \approx .95$$

We can rewrite this to get an interval of estimates of μ.

$$P\left(\mu - 2\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 2\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

see appendix of this lecture.

$$\Leftrightarrow P\left(\bar{X} - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

$$\Leftrightarrow P\left(\mu \in \left(\bar{X} - 2\frac{\sigma}{\sqrt{n}}, \bar{X} + 2\frac{\sigma}{\sqrt{n}}\right)\right) \approx 0.95$$

95% CI for μ.

ex (from warmup)

A population distribution is known to have an SD of 20. You test if the population mean is equal to 50. The average of a sample of 64 observations is 55. What is your 95% confidence interval for the population mean?

$\frac{\sigma}{\sqrt{n}} = 2.5$, $\bar{X} \pm 2\frac{\sigma}{\sqrt{n}} = 55 \pm 2(2.5)$
$= (50, 60)$

ex (Proportion of undecided voters)

In a simple random sample of 400 voters in a state, 23% are undecided about which way they will vote. Find a 95% CI for the proportion of undecided voters in the state.

$X_1, \ldots, X_{400}$ iid with mean $p = .23$

and $\sigma = \sqrt{.23(.77)}$

$= .44$

95% CI is $.23 \pm 2\left(\frac{.44}{\sqrt{400}}\right)$

$= (.186, .274)$

Confidence level

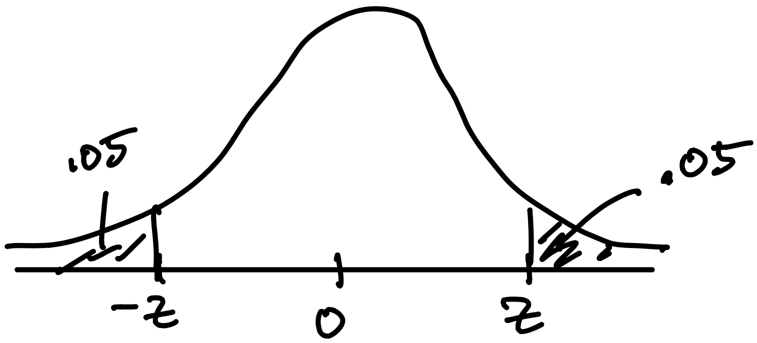
In above problem find 99.7% CI

99.7% CI is $.23 \pm 3\left(\frac{.44}{\sqrt{400}}\right)$

$= (.164, .296)$

Notice greater certainty requires a wider interval.

To find 90% CI



$$z = \Phi^{-1}(.95) = \text{stats.norm.ppf}(.95)$$
$$= 1.64$$

so 90% CI is

$$\left(\bar{X} - 1.64\frac{\sigma}{\sqrt{n}},\ \bar{X} + 1.64\frac{\sigma}{\sqrt{n}}\right)$$

$\bar{X} = .23$, $\frac{\sigma}{\sqrt{n}} = \frac{.44}{20}$, $\bar{X} = .23$

$$= (.194, .266)$$

Appendix

Show

$$P\left(\mu - 2\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 2\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

$$\Longleftrightarrow \quad P\left(\bar{X} - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 2\frac{\sigma}{\sqrt{n}}\right) \approx 0.95$$

Pf/

The event

$$\mu - 2\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 2\frac{\sigma}{\sqrt{n}}$$

$$\Longleftrightarrow$$

$$-\mu + 2\frac{\sigma}{\sqrt{n}} > -\bar{X} > -\mu - 2\frac{\sigma}{\sqrt{n}}$$

$$\Longleftrightarrow$$

$$\bar{X} + 2\frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - 2\frac{\sigma}{\sqrt{n}}$$

$$\Longleftrightarrow \quad \bar{X} - 2\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{2\sigma}{\sqrt{n}}$$

so the probabilities are the same. $\square$