Lecture 22

Warmup Example: Cyber Security

Based on information from the National Cyber Security Alliance, 93% of computer owners believe they have antivirus programs installed on their computers.

In a random sample of 400 scanned computers, it is found that 380 of them (or 95%) actually have antivirus software programs.

Use the sample data from the scanned computers to test the claim that 93% of computers have antivirus software.

Requirements

1. The 400 computers were randomly selected (check!)
2. There is a fixed number of independent trials, two possible outcomes (check!)
3. Is $np_0 \geq 10$? Is $n(1 - p_0) \geq 10$?

$$np_0 = (400)(0.93) = 372$$
$$n(1 - p_0) = (400)(1 - 0.93) = 28$$

Check!

Hypotheses

Write the hypotheses:

$$\mathbf{H_0} : p = 0.93 \qquad \text{versus} \qquad \mathbf{H_A} : p \neq 0.93$$

Significance Level, Test Statistic

Since we didn't have a specified level, choose $\alpha = 0.05$. We are testing a claim about a **population proportion**, so we will use a normal approximation:

Sample statistic : only $\hat{p}$

$$z_{\text{test}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{380}{400} - 0.93}{\sqrt{\frac{0.93(0.07)}{400}}}$$

SE

```r
z_test <- ( 380/400 - 0.93 ) / sqrt( 0.93 * 0.07 / 400 )
z_test
```

```
## [1] 1.567724
```

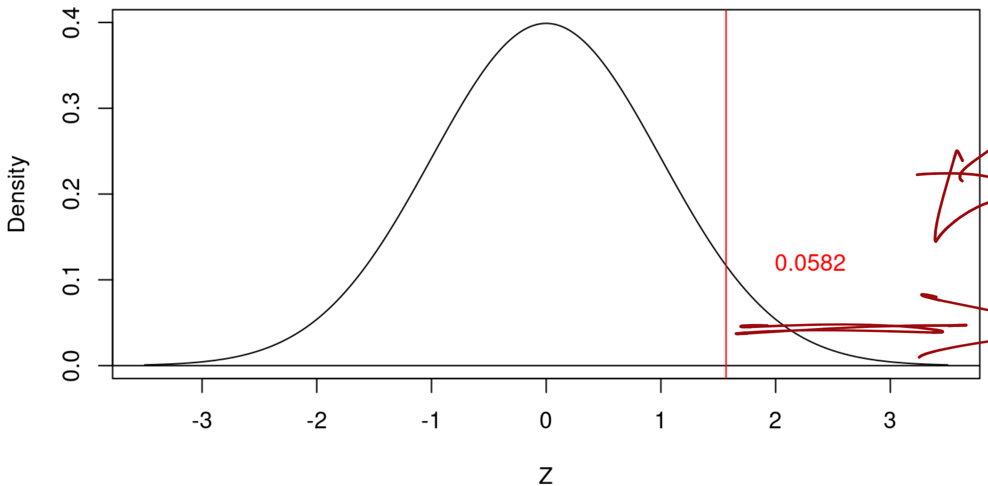
test statistic

If CI on its own: then $\hat{p} \longrightarrow SE$

The p-value

```
pnorm(z_test, lower.tail = FALSE) * 2
```

```
## [1] 0.1169457
```



Conclusion

Thus, since $p > \alpha$, we do not have evidence at the 95% level to conclude that the population proportion of computers having antivirus software is not 93%. In other words, there is not sufficient evidence to warrant rejection of this claim.

Solving Problems

Three Kinds of Problems

There are three main kinds of problems we've learned about, each with one or two sub-types.

- Normal distribution, question about means: confidence intervals (Z) and hypothesis tests (Z)
  - Requires **either** known sigma and normality **or** normality and 30+ samples

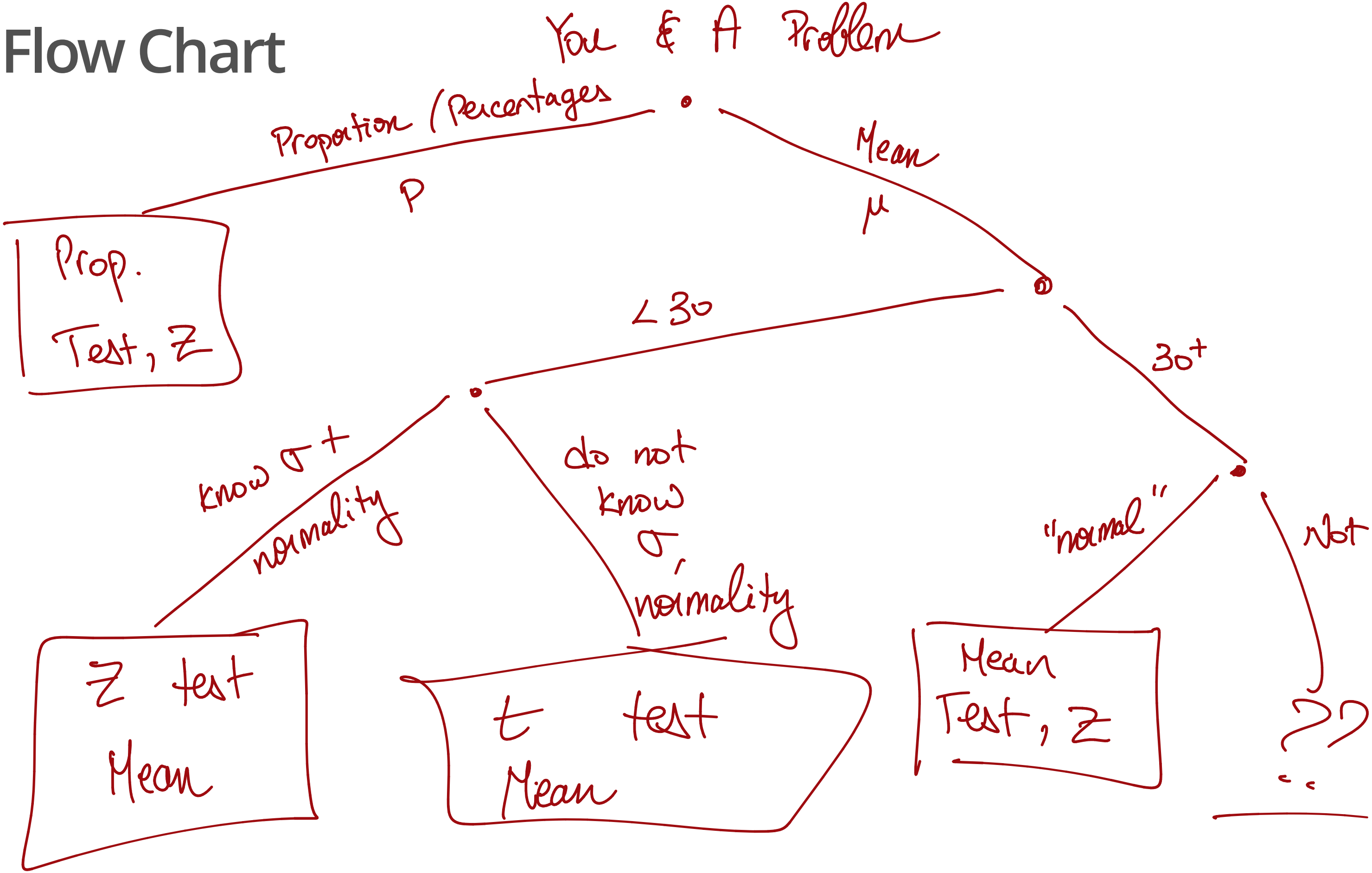
    don't care about n

    σ

    S

    do care about n
- t distribution, question about means: confidence intervals (t) and hypothesis tests (t)
  - The rest of the cases: less than 30 samples **and** sigma not known
  - Technically not all the cases: you'll discuss this more if you take 1052H
- Normal distribution, question about proportions: confidence intervals (Z) and hypothesis tests (Z)

Flow Chart



You & A Problem
Proportion / Percentages
p
Mean
μ
Prop.
Test, Z
< 30
30+
know σ + normality
do not know σ, normality
"normal"
Not
Z test
Mean
t test
Mean
Mean
Test, Z
??

Examples

Problem 1: Discarded Plastics

hypothesis

n s $\alpha = 0.05$ $\bar{x}$

A sample of 62 households had their recycling audited. The weights of their weekly recycled plastic had sample mean 1.911 pounds, with sample standard deviation 1.065 pounds. At the 95% level, test the claim that the mean weight of recycled plastic from the population of households is greater than 1.80 pounds.

alternative

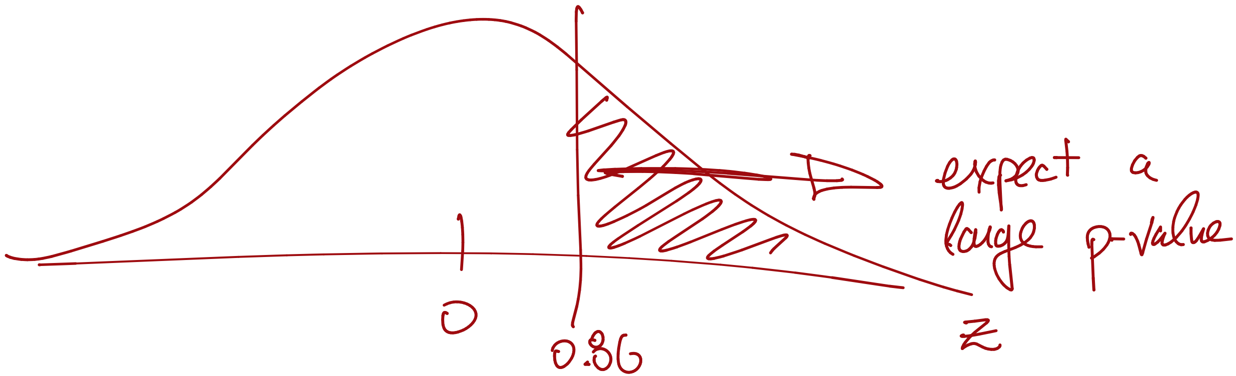
Also find a 95% confidence interval for the mean weight of recycled plastic per week.

$H_0: \mu = 1.80$ versus $H_A: \mu > 1.80$

$\mu \leq 1.80$

$n > 30 \implies$ doesn't say "not normal" $\longrightarrow Z$

$$Z = \frac{\bar{X} - \mu}{SE} = \frac{\bar{X} - \mu}{s/\sqrt{n}} = 0.36$$



p-val = 0.20592

$\therefore$ $p > \alpha$: $0.20592 > 0.05$. Thus $\longrightarrow$

We do not have evidence at the 95% level (or the $\alpha = 0.05$ level) to reject the null: we cannot conclude that the mean weight of weekly recycled plastic is more than 1.80 pounds.

Problem 2: The YSORT Trial

number of "successes"

n

$\alpha = 0.01$

The Genetics and IVF Institute conducted a clinical trial of the YSORT method designed to increase the probability of conceiving a boy. In this trial, 291 babies were born to parents using the YSORT method, and 239 of them were boys. Use a 99% significance level to test the claim that the YSORT method is effective at increasing the likelihood that a baby will be a boy.

Also find a 95% confidence interval for the true underlying proportion of babies born using the YSORT method who will be boys.

alternative

$H_0: p = 0.50$ versus $H_A: p > 0.5$

$p \leq 0.50$

$\alpha = 0.01$

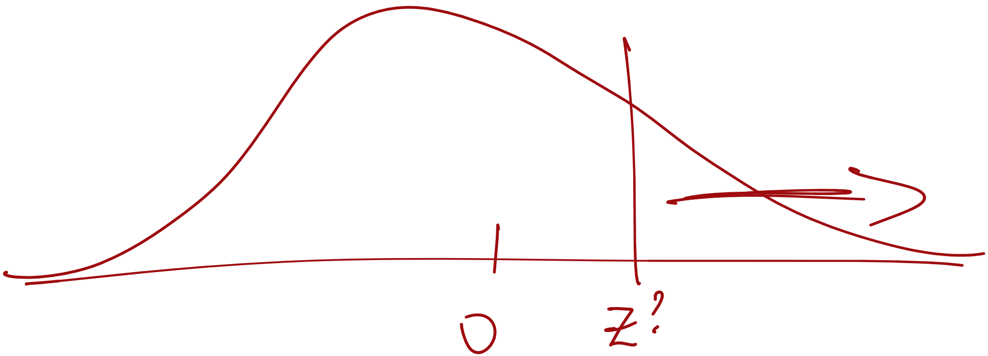
$n = 291$

$\hat{p} = 239/n = 239/291$

$p_0 = 0.5$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

$$= \frac{\frac{239}{291} - 0.5}{\sqrt{\frac{0.5(0.5)}{291}}}$$



p-val is TINY $\Rightarrow$ $p < \alpha$ $\Rightarrow$ Reject

$\therefore$ we do have evidence at the $\alpha = 0.01$ level to reject the null and conclude that the YSORT trial is associated with an increase in boys being born.

CI :

$\alpha = 0.05$

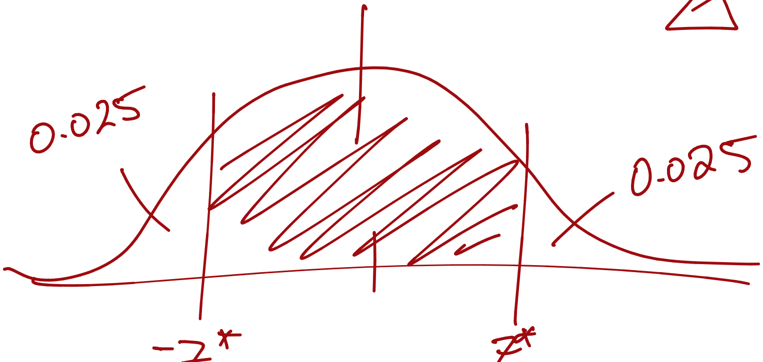
$\hat{p} = 0.8213$

$n = 291$

point $\pm$ "star" $\cdot$ SE

$\hat{p}$ $\qquad$ z^* $\qquad$ $\sqrt{\frac{p(1-p)}{n}}$

0.95 two-sided



$z^* = \text{qnorm}(0.975)$

$= 1.959964$

Problem 3: Highway Speeds

Southbound traffic on the I-280 highway near Cupertino, California had its speed monitored at 3:30pm on a Wednesday. The sample of 12 cars had mean 97.6 km/hr with standard deviation 6.56 km/hr. Test the highway patrol's claim that the average speed on this highway at this time of day is lower than the speed limit of 105 km/hr.

Also compute a 99% confidence interval for the mean speed.

Hypothesis:

$H_0: \mu = 105$ versus $H_A: \mu < 105$

$\mu \geq 105$

$n = 12$

$\bar{x} = 97.6$

$s = 6.56$

$\boxed{t}$

Problem 4: Weights of Pennies

Before 1983, US pennies were made with 97% copper and 3% zinc. After 1983, they were converted to 3% copper and 97% zinc to make them cheaper to manufacture. A simple random sample of 35 post-1983 pennies had an average weight of 2.49910g, with standard deviation 0.01648g. The US Mint specifies that post-1983 pennies should be manufactured with mean weight 2.500g. At a 95% level, do you believe that pennies are actually being manufactured with mean weight of 2.500g?

Compute a 99% confidence interval for the mean weight of post-1983 pennies.

$H_0: \mu = 2.500$ versus $H_A: \mu \neq 2.500$

$n = 35$

$\bar{x} = 2.49910$

$s = 0.01648$

$\boxed{Z}$

95%

$\boxed{99\% \text{ Z CI}}$

Problem 5: Cell Phones and Cancer

In a study of 420,095 Danish cell phone users, 125 subjects developed cancer of the brain or nervous system (Journal of the National Cancer Institute). Test the claim of the belief that such cancers are affected by cell phone use. That is, test the claim that cell phone users develop cancer of the brain or nervous system at a rate that is different from the rate of 0.0340% for people who do not use cell phones. Use a 99.5% significance level.

1% ⇒ 0.01

Z proportion test

p_0 — 0.000340

np_0 $\quad$ $n(1-p_0)$

Problem 6: Insomnia and Zopiclone

null n

A clinical trial was conducted to test the effectiveness of the drug Zopiclone for treating insomnia in older subjects. Before treatment with Zopiclone, 16 subjects had a mean wake time of 102.8 minutes. After treatment with Zopiclone, the 16 subjects had a mean wake time of 98.9 minutes, and a standard deviation of 42.3 minutes (JAMA). Assume that the 16 sample values appear to be from a normally distributed population, and test the claim that after treatment with Zopiclone, subjects have a reduced mean wake time.

alternative Xbar S

Also find a 95% confidence interval for the mean wake time.

$H_0: \mu = 102.8$ vs. $H_A: \mu < 102.8$

$\mu \geq 102.8$

What kind of test?

t