Hypothesis Tests - Tests of Significance

Confidence intervals are appropriate when one is interested in estimating unknown population parameters. Confidence intervals provide a range of possible values for the population parameter that are consistent with the sample data.

The second type of inference we will consider is the hypothesis test (or test of significance)

The hypothesis test assesses the degree to which sample data support a particular claim about the value of a population parameter.

8.1
8.3

Example – Can Dogs Smell Cancer?

1989 – British woman asked for mole to be removed from leg because dog constantly sniffed it and had tried to bite it off; woman had malignant melanoma

2001 – man reports sudden interest by dog in patch of eczema on man's leg; man had developed skin cancer

Isolated incidents or an indication that dogs have the ability to detect some forms of cancer?

Conduct experiment and perform hypothesis test to investigate the question of whether dogs can detect cancer.

Experiment – Can Dogs Detect Bladder Cancer?

Six dogs selected and trained to identify bladder cancer

For the actual test a dog would be presented with 7 urine samples, 1 belonging a bladder cancer patient and 6 non-cancer samples. Dog would lie down next to one of the urine samples to indicate presence of cancer. 54 trials in total.

How to decide whether dogs can detect cancer?

Study in British Medical Journal 2004

Experiment – Can Dogs Detect Bladder Cancer?

What would you expect if dogs are actually unable to detect cancer?

Dog chooses correct sample strictly random chance.
Success rate should be $1/7 \approx 14.3\%$

$H_0: p = 1/7 = p_0$ (null hypothesis; no effect, no difference)
↖ very specific value for parameter

What would you expect if dogs are actually able to detect cancer?

Dogs should have a higher success rate than would be predicted by chance alone.

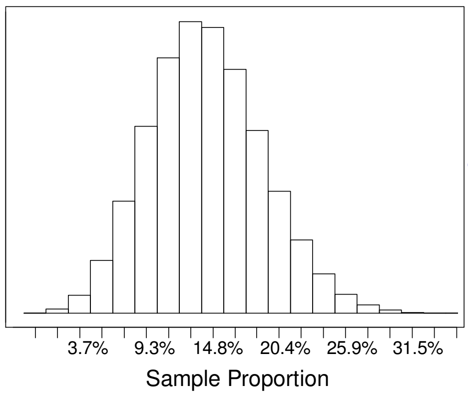
$H_1: p > 1/7$ (alternative hypothesis)
↖ not specific

hypotheses make statements about population parameter p.

5

Experiment – Can Dogs Detect Bladder Cancer?

How will you decide if dogs seem able or unable to detect cancer?



Assume H_0 is true, $p = 1/7$

Consider $\hat{p} \sim N\left(1/7, \frac{\sqrt{1/7 \cdot 6/7}}{\sqrt{54}}\right)$

$\hat{p} = \frac{X}{n}$

$X \sim Bin(n, p_0)$

If H_0 true

$X \sim Bin(54, 1/7)$

$\hat{p} = 40.7\%$

Dogs correctly detected cancer in 22 trials, or a 40.7% success rate.

6

Assessing Evidence

Two choices:

i) Dog's unable to detect cancer. True success rate is 14.3% and the sample proportion of 40.7% is the result of random chance, or

ii) Dog's are able to detect cancer. True success rate is higher than 14.3%.

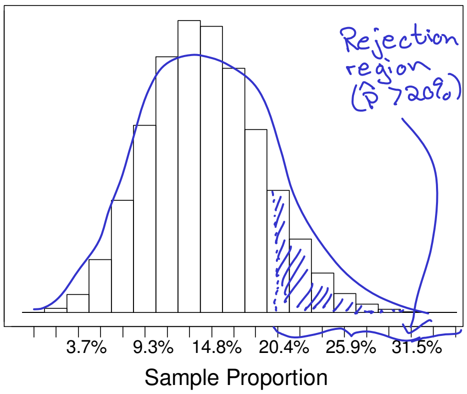
Which choice seems more reasonable?

40.7% is so much larger than what would be expected if $p = 1/7$ (H_0 true) that it seems more reasonable to believe the alternative hypothesis that dogs are able to detect cancer ($p > 1/7$)

7

Conclusion

Before collecting data we may have proposed that any success rate in excess of 20% was convincing enough to conclude that dogs are able to detect bladder cancer.



possible that Ho is in fact true but you get large enough $\hat{p}$ to reject Ho, by chance alone.

How likely?

$P(\hat{p} \geq 20\%) = P(Z \geq 1.2)$

$= 11.5\%$

$Z = \frac{20\% - 14.3\%}{4.76\%}$

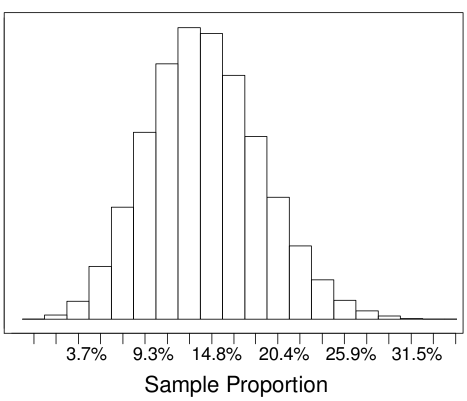
$= 1.197$

or 1.2

8

Conclusion

Alternatively, we could measure how unusual success rate of 40.7% is if dogs are actually unable to detect cancer.



$Z = \frac{40.7\% - 14.3\%}{4.76\%} = 5.55$

How unusual is 40.7% if Ho true?

$P(\hat{p} \geq 40.7\%) =$ p-value

$= P(Z \geq 5.55)$

$\approx 0\%$

0.0000000001

Seems extremely that chance alone produced the observed success rate.

Level of Significance

How much evidence is required in order to reject null hypothesis?

Before conducting experiment you could choose some **level α** that determines how much evidence is required to reject null hypothesis.

α determines how unusual or rare outcome must be to reject null hypothesis

α determines the rejection region, and the threshold for the p-value: small α means more evidence required; rejection region smaller, p-value must be smaller

How to choose α?

Standard is $\alpha = 5\%$

What can go wrong?

Conclude dogs can detect cancer, when in fact they can't

Reject null hypothesis even though null hypothesis true.

Type I error

$\alpha = P(\text{Type I error})$

Conclude dogs appear unable to detect cancer, when in fact they can

Fail to reject null hypothesis even though null hypothesis is false

Type II error

$\alpha\downarrow \rightarrow P(\text{Type I error})\downarrow \rightarrow P(\text{Type II error})\uparrow$

Steps in Hypothesis Tests

1) Hypotheses $H_0: p = 1/7 = p_0$

$H_1: p > 1/7$ (one-sided)

2) Level of significance

Need to choose α

3) Test statistic

$\hat{p} \sim N\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$ or $\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$ if H_0 true

4) Decision rule

Reject H_0 if $\underbrace{\hat{p} > ? \text{ or } z > ?}_{\text{rejection region}}$ or p-value $< \alpha$

5) Calculate

$\hat{p} = 40.7\%$

6) Conclusion

Sufficient evidence to reject H_0

Dogs appear able to detect bladder cancer.