Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Type I and type II errors.
2. Power.
3. Estimation and CIs.
4. CIs and dog sniffing cancer example.
5. Sample size calculation.

Read chapter 4.

HW2, due Fri, May8, 1159pm. 2.3.15, 3.3.18, and 4.1.23.

These problems are on the next 3 slides.

Midterm is Tue May5 in class.

HW should be submitted BY EMAIL to STATGRADER@STAT.UCLA.EDU

The course website is http://www.stat.ucla.edu/~frederic/13/S26 .

Needles

Exercises 2.3.15 and 2.3.16 refer to the needle data.

2.3.15 Consider a manufacturing process that is producing hypodermic needles that will be used for blood donations. These needles need to have a diameter of 1.65 mm—too big and they would hurt the donor (even more than usual), too small and they would rupture the red blood cells, rendering the donated blood useless. Thus, the manufacturing process would have to be closely monitored to detect any significant departures from the desired diameter. During every shift, quality control personnel take a sample of several needles and measure their diameters. If they discover a problem, they will stop the manufacturing process until it is corrected.

a. Define the parameter of interest in the context of this study and assign an appropriate symbol to it.

b. State the appropriate null and alternative hypotheses using the symbol defined in (a).

c. Describe what a Type I error would be in this study. Also, describe the consequence of such an error in the context of this study.

d. Describe what a Type II error would be in this study. Also, describe the consequence of such an error in the context of this study.

3.3.18 Reconsider the investigation of the manufacturing process that is producing hypodermic needles. Using the data from the most recent sample of needles, a 90% confidence interval for the average diameter of needles is found to be (1.62 mm, 1.66 mm). For each of the following statements, say whether VALID or INVALID.

a. We are 90% confident that the average diameter of the sample of 35 needles is between 1.62 and 1.66 mm.

b. Based on the 90% confidence interval, there is evidence that the average diameter of needles produced by this manufacturing process is 1.65 mm.

c. Based on the 90% confidence interval, there is evidence that the average diameter of needles produced by this manufacturing process is different from 1.65 mm.

d. We are 90% confident that the average diameter of needles produced by this manufacturing process is between 1.62 and 1.66 mm.

e. About 90% of the needles produced by this manufacturing process have a diameter between 1.62 and 1.66 mm.

f. If we want to be more than 90% confident, we should take a larger sample of needles.

Colds and exercise

4.1.23 In November 2010, an article titled "Frequency of Colds Dramatically Cut with Regular Exercise" appeared in *Medical News Today*. The article was based on the findings of a study by researchers Nieman et al. (*British Journal of Sports Medicine*, 2010) that followed 1,002 people aged 18–85 years for 12 weeks, asking them to record their frequency of exercise (5 or more days a week? Yes or No) as well as incidences of upper respiratory tract infections (Cold during last week? Yes or No).

a. Identify the explanatory variable in this study. Also classify this variable as categorical or quantitative.

b. Identify the response variable in this study. Also classify this variable as categorical or quantitative.

c. Identify a confounding variable that provides an alternative explanation for the lower frequency of colds among those who exercised 5 or more days per week, compared to those who were largely sedentary.

1. Type I and Type II errors

- In medical tests:
  - A type I error is a false positive. (conclude someone has a disease when they don't.)
  - A type II error is a false negative. (conclude someone does not have a disease when they actually do.)
- These types of errors can have very different consequences.

Type I and Type II Errors

TABLE 2.9 A summary of Type I and Type II errors

		What is true (unknown to us)	
		Null hypothesis is true	Null hypothesis is false
What we decide (based on data)	Reject null hypothesis	Type I error (false alarm)	Correct decision
	Do not reject null hypothesis	Correct decision	Type II error (missed opportunity)

•

Type I and Type II errors

TABLE 2.10 Type I and Type II errors summarized in context of jury trial

		What is true (unknown to the jury)	
		Null hypothesis is true (defendant is innocent)	Null hypothesis is false (defendant is guilty)
What jury decides (based on evi-dence)	Reject null hypothesis (Jury finds defendant guilty)	Type I error (false alarm)	Correct decision
	Do not reject null hypothe-sis (Jury finds defendant not guilty)	Correct decision	Type II error (missed opportunity)

The probability of a Type I error

- The significance level is the probability of a type I error, when Ho is true.
- Suppose the significance level is 0.05. If the null is true we would reject it 5% of the time and thus make a type I error 5% of the time.
- If you make the significance level lower, you have reduced the probability of making a type I error, but have increased the probability of making a type II error.

The probability of a Type II error

- The probability of a type II error is more difficult to calculate.
- In fact, the probability of a type II error is not even a fixed number. It depends on the value of the true parameter you are estimating.
- The probability of a type II error can be very high if:
  - The effect size is small.
  - The sample size is small.

2. Power

- The probability of rejecting the null hypothesis when it is false is called the **power** of a test.
- Power = 1 – P(Type II error). It is usually expressed as a function of μ.
- We want a test with high power and this is aided by:
  - A large effect size, i.e. true μ far from the parameter in the null hypothesis.
  - A large sample size.
  - A small standard deviation.
  - A higher significance level means greater power. The downside is that you get more type I errors.

3. Estimation and confidence intervals.

Chapter 3

Chapter Overview

- So far, we can only say things like
  - "We have strong evidence that the long-run frequency of death within 30 days after a heart transplant at St. George's Hospital is greater than 15%."
  - "We do not have strong evidence kids have a preference between candy and a toy when trick-or-treating."
- We want a method that says
  - "I believe 68 to 75% of all elections can be correctly predicted by the competent face method."

Confidence Intervals

- Interval estimates of a population parameter are called **confidence intervals**.
- We will find confidence intervals three ways.
  - Through a series of tests of significance to see which proportions are plausible values for the parameter.
  - Using the standard error (the standard deviation of the simulated null distribution) to help us determine the width of the interval.
  - Through traditional theory-based methods, i.e. formulas.

Statistical Inference: Confidence Intervals

Section 3.1

Can Dogs Sniff Out Cancer?

Section 3.1

Can Dogs Sniff Out Cancer?

Sonoda et al. (2011). Marine, a dog originally trained for water rescues, was tested to see if she could detect if a patient had colorectal cancer by smelling a sample of their breath.

- She first smells a bag from a patient with colorectal cancer.
- Then she smells 5 other samples; 4 from normal patients and 1 from a person with colorectal cancer
- She is trained to sit next to the bag that matches the scent of the initial bag (the “cancer scent”) by being rewarded with a tennis ball.

Can Dogs Sniff Out Cancer?

In Sonoda et al. (2011). Marine was tested in 33 trials.

- Null hypothesis: Marine is randomly guessing which bag is the cancer specimen ($\pi = 0.20$)
- Alternative hypothesis: Marine can detect cancer better than guessing ($\pi > 0.20$)

π represents her long-run probability of identifying the cancer specimen.

Can Dogs Sniff Out Cancer?

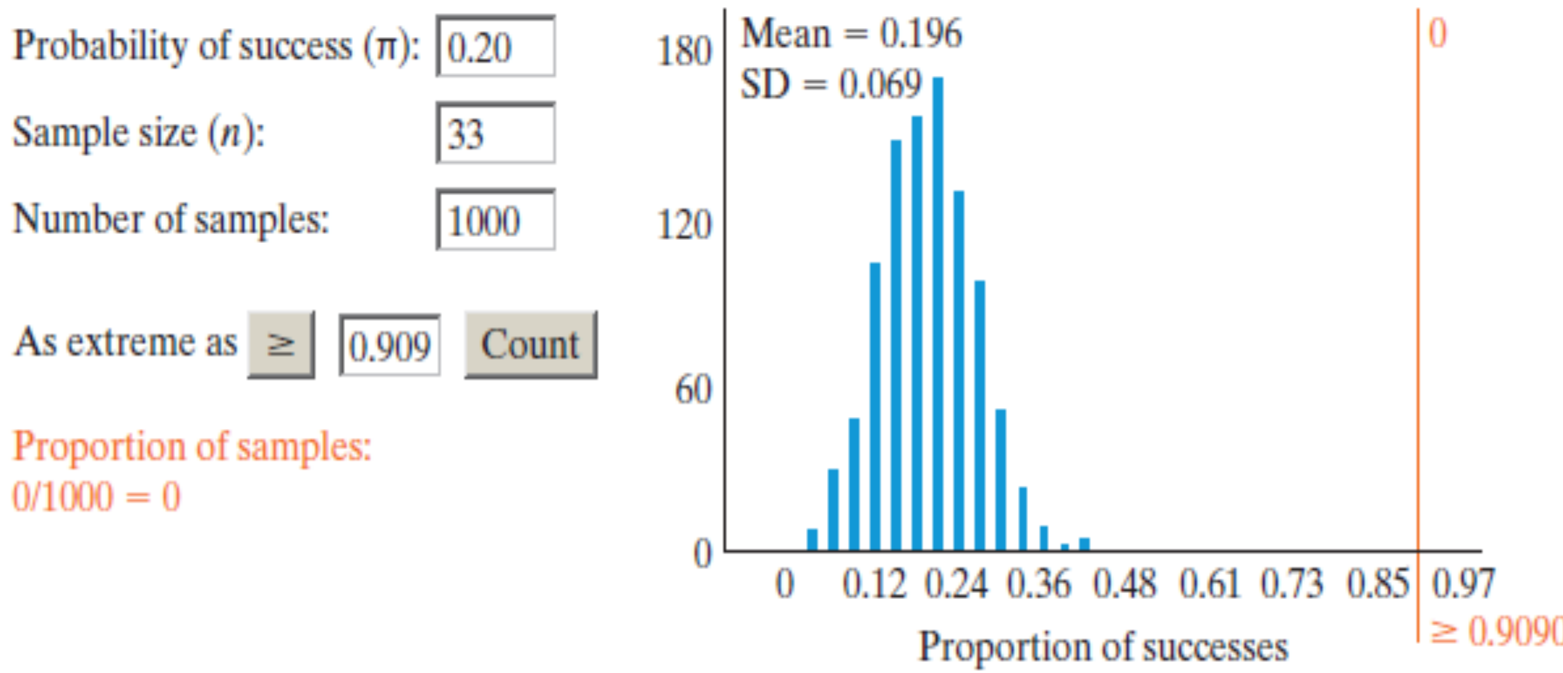
- 30 out of 33 trials resulted in Marine correctly identifying the bag from the cancer patient
- So our sample proportion is

$$\hat{p} = \frac{30}{33} \approx 0.909$$

- Do you think Marine can detect cancer?
- What sort of p-value will we get?

Can Dogs Sniff Out Cancer?

Our sample proportion lies more than 10 standard deviations above the mean and hence our p-value ~ 0.

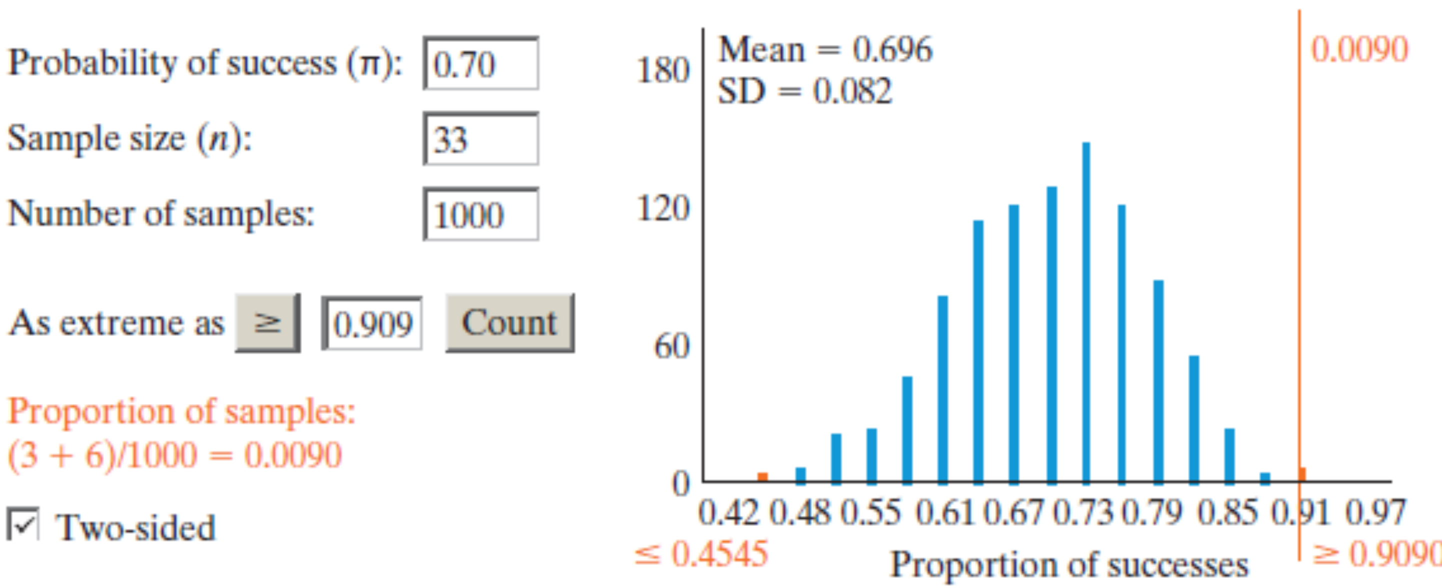


Can Dogs Sniff Out Cancer?

- Can we estimate Marine's long run frequency of picking the correct specimen?
- Since our sample proportion is about 0.909, it is plausible that 0.909 is a value for this frequency. What about other values?
- Is it plausible that Marine's frequency is actually 0.70 and she had a lucky day?
- Is a sample proportion of 0.909 unlikely if $\pi = 0.70$?

Can Dogs Sniff Out Cancer?

- H_0: $\pi = 0.70$ H_a: $\pi \neq 0.70$
- We get a small p-value (0.0090) so we can essentially rule out 0.70 as her long run frequency.

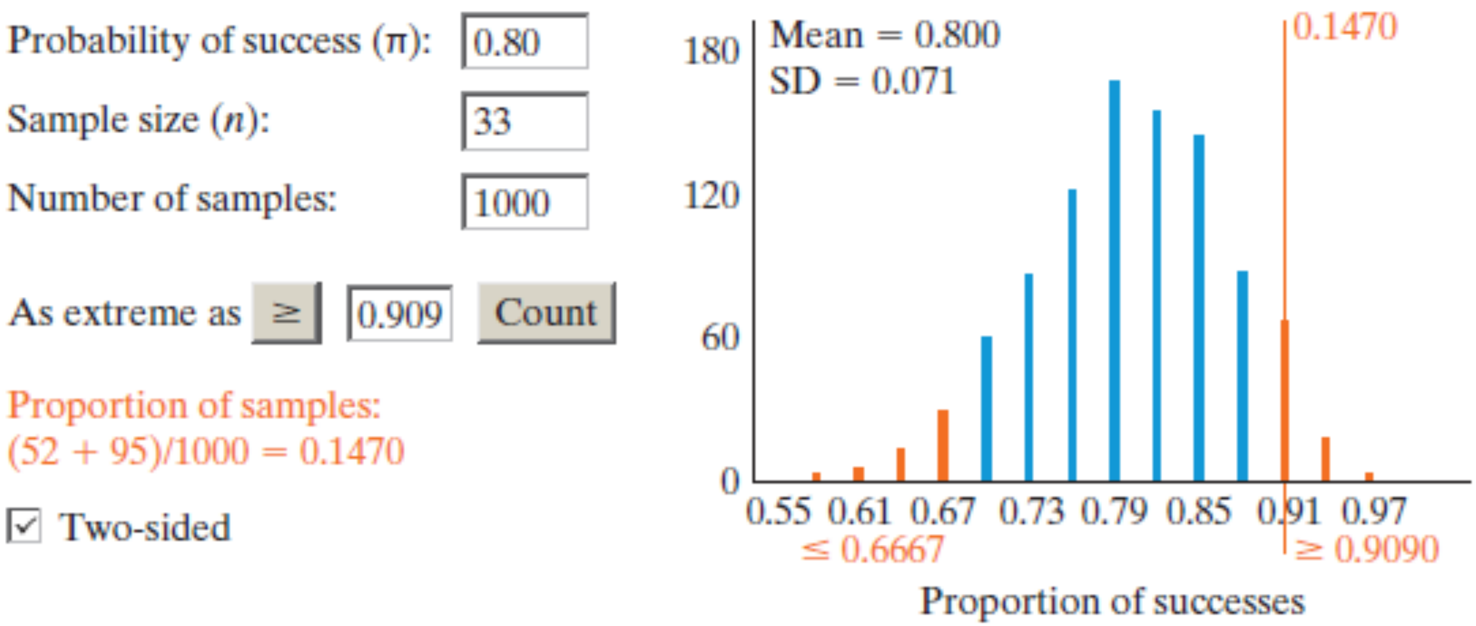


Can Dogs Sniff Out Cancer?

- What about 0.80?
- Is 0.909 unlikely if π = 0.80?

Can Dogs Sniff Out Cancer?

- H_0: $\pi = 0.80$ H_a: $\pi \neq 0.80$
- We get a large p-value (0.1470) so 0.80 is a ***plausible*** value for Marine's long-run frequency.



Developing a range of plausible values

- If we get a small p-value (like we did with 0.70) we will conclude that the value under the null is not plausible. This is when we reject the null hypothesis.
- If we get a large p-value (like we did with 0.80) we will conclude the value under the null is plausible. This is when we can’t reject the null.

Developing a range of plausible values

- One could use software (like the one-proportion applet the book recommends) to find a range of plausible values for Marine's long term probability of choosing the correct specimen.
- We will keep the sample proportion the same and change the possible values of π.
- We will use 0.05 as our cutoff value for if a p-value is small or large. (Recall that this is called the **significance level**.)

Can Dogs Sniff Out Cancer?

- It turns out values between 0.761 and 0.974 are plausible values for Marine's probability of picking the correct specimen.

Probability under null	0.759	0.760	0.761	0.762	…….	0.973	0.974	0.975	0.976
p-value	0.042	0.043	**0.063**	**0.063**		**0.059**	**0.054**	0.049	0.044
Plausible?	No	No	Yes	Yes	……… Yes	Yes	Yes	No	No

Can Dogs Sniff Out Cancer?

- (0.761, 0.974) is called a ***confidence interval***.
- Since we used 5% as our significance level, this is a 95% confidence interval. (100% – 5%)
- 95% is the ***confidence level*** associated with the interval of plausible values.

Can Dogs Sniff Out Cancer?

- We would say we are 95% confident that Marine's probability of correctly picking the bag with breath from the cancer patient from among 5 bags is between 0.761 and 0.974.
- This is a more precise statement than our initial significance test which concluded Marine's probability was more than 0.20.
- Sidenote: We do not say P{π is in (.761, .974)} = 95%, because π is not random. The *interval* is random, and would change with a different sample. If we calculate an interval this way, then P(interval contains π) = 95%.

Confidence Level

- If we increase the confidence level from 95% to 99%, what will happen to the width of the confidence interval?

Can Dogs Sniff Out Cancer?

- Since the confidence level gives an indication of how sure we are that we captured the actual value of the parameter in our interval,
  to be more sure our interval should be wider.
- How would we obtain a wider interval of plausible values to represent a 99% confidence level?
  - Use a 1% significance level in the tests.
  - Values that correspond to 2-sided p-values larger than 0.01 should now be in our interval.

5. Sample size calculation.

We previously saw that, when testing proportions, the standardized statistic Z = $\frac{\hat{p}-\pi}{SE}$,

where SE = $\sqrt{\pi(1-\pi)/n}$.

We also know that for the 2-sided Z-test, 1.96 is the cutoff for statistical significance.

If |Z| > 1.96, then p-value < 5%.

Suppose π = 50%, $\hat{p}$ = 70%, n = 10. How many more observations are needed to achieve statistical significance, if the effect size stays the same?

Z = $\frac{\hat{p}-\pi}{SE}$, SE = $\sqrt{\pi(1-\pi)/n}$.

We also know that for the 2-sided Z-test, 1.96 is the cutoff for statistical significance.

If |Z| > 1.96, then p-value < 5%.

Suppose π = 50%, $\hat{p}$ = 70%, n = 10. How many more observations are needed to achieve statistical significance, if the effect size stays the same?

We want to find n, so that 1.96 = $\frac{\hat{p}-\pi}{\sqrt{\pi(1-\pi)/n}} = \frac{.7-.5}{\sqrt{.5(1-.5)/n}}$

Squaring both sides, $1.96^2 = (.2)^2 / (.25/n) = .04n/.25$,

i.e. n = .25 * 1.96^2 / .04 = 24.01.

n needs to be *at least* 24.01, so really 25.

So we need 15 *more* observations.