17.0 Significance Tests

- Answer Questions
- Review Confidence Intervals
- The Gauss Model
- Genetics
- Significance Tests

17.1 CI Review

The general formula for a two-sided C% confidence interval is:

$$L, U = pe \pm se * z_C.$$

Here z_C is the critical value from a normal table—it is the value that has area C in the middle.

If n is large compared to the population, use $se * fpcf$ in place of the se

The se is $sd/\sqrt{n}$ for an average, $sd * \sqrt{n}$ for a sum, and $\sqrt{\hat{p}(1-\hat{p})/n}$ for a proportion.

Where do CIs come from?

The CLT says:

$$z = \frac{\bar{X} - EV}{sd/\sqrt{n}} \dot{\sim} N(0,1)$$

so

$$\mathbf{P}[-z_C \leq \frac{\bar{X} - EV}{sd/\sqrt{n}} \leq z_C] \approx C/100$$

where z_C is the value from a normal table that has area C shaded in the middle.

Now we can use ordinary algebra to manipulate the terms inside the probability statement to solve for L and U.

$$
\begin{aligned}
\frac{C}{100} &\approx \mathbf{P}[-z_C \le \frac{\bar{X} - EV}{sd/\sqrt{n}} \le z_C] \\
&= \mathbf{P}[-\frac{sd}{\sqrt{n}} * z_C \le \bar{X} - EV \le \frac{sd}{\sqrt{n}} * z_C] \\
&= \mathbf{P}[-\frac{sd}{\sqrt{n}} * z_C - \bar{X} \le -EV \le \frac{sd}{\sqrt{n}} * z_C - \bar{X}] \\
&= \mathbf{P}[\frac{sd}{\sqrt{n}} * z_C + \bar{X} \ge EV \ge \bar{X} - \frac{sd}{\sqrt{n}} * z_C]
\end{aligned}
$$

So

$$
\begin{aligned}
L &= \bar{X} - \frac{sd}{\sqrt{n}} * z_C \\
U &= \bar{X} + \frac{sd}{\sqrt{n}} * z_C
\end{aligned}
$$

17.2 The Gauss Model

The Gauss model was first used (by Gauss) to find the orbit of Ceres, the first asteroid to be discovered. But before its observation could be confirmed by other astronomers, it disappeared behind the sun.

Gauss used his model, and an incredibly complex calcualtion involving nine coordinate systems, to predict where and when astronomers should look to recapture it when its orbit carried it from behind the sun.

The Gauss model says that measurement errors in an observation are like draws from a box.

The measurement is the true value plus whatever error is marked on the ticket:

$$\textbf{measurement} = \textbf{true value} + \textbf{random error}.$$

The EV of the box is zero (otherwise, the measurement process is biased).

The standard deviations of the box is usually, but not always, unknown.

17.3 Genetics

Gregor Mendel was an Augustinian monk in charge of the monastery's truck garden. He noted that several traits in pea plants, e.g.:

- color
- height
- wrinkled pods

seemed to be inherited from the parent plants in a predictable way.

To study color Mendel got inbred strains, whose progeny were always yellow or always green. Then he did experiments in which those inbred strains were crossed, and he observed the colors of the offspring.

Recall from biology: Mendelian theory says that each plant has two genes for color, and each parent contributes one of those genes, at random, to the progeny. Thus:

$$
\begin{aligned}
GG \times GG &\Rightarrow GG \\
YY \times YY &\Rightarrow YY \\
YY \times GG &\Rightarrow YG \\
YY \times YG &\Rightarrow YY, YG \\
GG \times YG &\Rightarrow GG, GY \\
YG \times YG &\Rightarrow GG, YY, YG, GY
\end{aligned}
$$

Yellow is dominant. Any plant that has a yellow gene provides only yellow peas.

The inbred plants were GG or YY. When crossing these, the first generation all had yellow peas (because of dominance) even the though genetic composition of each plant was YG.

The second generation was formed by crossing the first generation plants. This cross was:

$$YG \times YG \Rightarrow GG, YY, YG, GY$$

and it gave plants such that 3/4 had yellow peas, 1/4 had green peas.

Mendel slowly developed his theory that each plant had two genes, contributed at random. He could predict that among, say, 100 second generation offspring, about 25 should bear green peas. So Mendel made many such crosses and found that his predicted numbers were close to those observed.

But how can Mendel prove his theory?

He had no statistical way to show that his observed counts of yellow and green pea plants matched well to the predictions from his model. All he could do was present his predictions, his counts, and wave his hands.

So he (probably) faked his data in order to get better agreement and thus to present a stronger case. We believe this because his reported counts were too good to be true—they were closer to his predictions than could happen under his model.

17.4 Significance Tests

A significance test is a way to decide whether the data strongly support point of view or another.

There are many kinds of significance tests, but all involve:

- a null and alternative hypothesis
- a test statistic
- a significance probability (P).

The handout gives an overview of many tests.

Step 1: Pick the null and alternative hypotheses.

If possible, put what you want to prove as the alternative. Or make the alternative hypothesis the one that leads to new action.

Example 1:

H_0: The mean lifetime with the new drug ≤ 72 years.

H_A : The mean lifetime with the new drug > 72 years.

Example 2:

H_0 : The mean annual global temperature $= 56\ F^\circ$.

H_A : The mean annual global temperature $\neq 56\ F^\circ$.

Step 2: Calculate the test statistic.

The test statisic is a one-number summary of all the information in the sample regarding the correctness of the alternative hypothesis. Different kinds of hypothesis tests (e.g., about means, proportions, differences of means, differences of proportions, etc.) require different test statistics.

The handout lists many standard cases.

Step 3: Find the P-value (or significance probability).

Use a table to find the P-value. This is "the probability of observing data that is as or more supportive of the alternative hypothesis when the null hypothesis is correct."

This interpretation of the P-value is a bit subtle—you will probably need to think about it.

Example: Suppose you have a new drug that may extend life expectancy. You give it to 36 random people and find that their average lifespan is 76 years, and the standard deviation in their lifespan is 12. You know that the average U.S. lifespan is 72, and hope to show that your drug improves that.

The first step is to choose the null and alternative hypotheses. For this problem, the following are the right ones. This is case I.2 of the handout:

H_0: The mean lifetime with the new drug ≤ 72 years.

H_A : The mean lifetime with the new drug > 72 years.

The second step is to find the test statistic. We are in case II.a of the handout, so:

$$ts = \frac{\bar{X} - \mu_0}{sd/\sqrt{n}} = \frac{76 - 72}{12/\sqrt{36}} = 2$$

The third step finds the significance probability. If you use hypotheses I.2, then you use rule III.2.

The significance probability, or P-value, is

$$\mathbf{P}[z > 2] = (1 - .9545)/2 = .02275.$$

So if the null hypothesis is true and the drug does not help, then you have only about 2.275 chances in 100 of observing the result in your experiment. This is pretty persuasive that your drug helps.

If you had a larger sample, then you might get a significance probability that is even smaller, say .000001. In this case, there is only 1 chance in a million that you would get such strong support when the drug is not helpful.