Today's plan:

- Section 4.4: Statistical Inference.
- Section 4.4.1: Repeated Two-Outcome Experiments

Section 4.4: Statistical Inference

Here is the setup for a statistical inference problem:

- We have some population (fish), and some variable of interest (percent tagged).

- We have some population (fish), and some variable of interest (percent tagged). The value of the variable is called the **parameter**.

- We have some population (fish), and some variable of interest (percent tagged). The value of the variable is called the **parameter**.
- Pick a sample and get the data from the sample.

- Find the value of the variable for the sample.

- Find the value of the variable for the sample. This is called the **statistic**.

- Find the value of the variable for the sample. This is called the **statistic**.
- From the statistic, infer conclusions about the parameter.

Example

A diaper manufacturer wants to know the percentage of leaky diapers in a batch they just made.

Solution

- *Testing a diaper for leakage ruins the diaper.*

Solution

- *Testing a diaper for leakage ruins the diaper.*
- *So they can't test them all.*

Solution

- *Testing a diaper for leakage ruins the diaper.*
- *So they can't test them all.*
- *Instead, take a random sample of 100 diapers, and test those.*

Solution

- *It turns out 20 of the diapers in the sample are defective.*

Solution

- *It turns out 20 of the diapers in the sample are defective.*
- *One might be tempted to infer that 20% of all diapers are defective.*

A closer look:

A closer look:

- The population here is

A closer look:

- The population here is all diapers in the batch.

A closer look:

- The population here is all diapers in the batch.
- The parameter is the percent p of leaky diapers in the whole batch.

A closer look:

- The population here is all diapers in the batch.
- The parameter is the percent p of leaky diapers in the whole batch.
- The sample is 100 random diapers.

- In the sample, 20 diapers tested positive for leakage, so

- In the sample, 20 diapers tested positive for leakage, so
- the statistic is $\hat{p} = 20\%$ defective diapers in the sample.

Notation

We denote the statistic with the same symbol as the parameter, but with a hat ^ on top.

Note: It's unlikely the leakage percent in the sample (the statistic) is exactly the percent in the whole population (the parameter).

Definition

The difference between the statistic and the parameter is called the **sampling error**.

Definition

The difference between the statistic and the parameter is called the **sampling error**.

It's probably not zero!

Section 4.4.1: Repeated Two-outcome experiments

The diaper example is a random experiment:

- Pick a diaper, test for leaks

The diaper example is a random experiment:

- Pick a diaper, test for leaks
- The two outcomes are: *defective*, *good*

The diaper example is a random experiment:

- Pick a diaper, test for leaks
- The two outcomes are: *defective*, *good*

(Just like flipping a weighted coin.)

Definition

Repeat the experiment n times and count the number k of successes (versus failures). This measurement has a **binomial** distribution.

When n is greater than 30, a binomial distribution is **near-normal**, so we can use those tools.

Question

What are the mean and the standard deviation of a binomial distribution?

Question

What are the mean and the standard deviation of a binomial distribution?

Since binomial distributions are near-normal, the mean and the standard deviation tell us everything.

Binomial Principle

Run a two-outcome random experiment n times ($n \geq 30$), and count the number k of successes.

If the two-outcome experiment has probability p of success

then the variable k will have a near-normal distribution with

- $\mu = n \cdot p$
- $\sigma = \sqrt{n \cdot p \cdot (1 - p)}$

Example

- Doctored coin with probability of heads $p = 0.7$.

Example

- Doctored coin with probability of heads $p = 0.7$.
- Flip $n = 50$ times and count the number k of heads.

Example

- Doctored coin with probability of heads $p = 0.7$.
- Flip $n = 50$ times and count the number k of heads.
- What can we say about k?

According to the binomial principle, *k* has a **near-normal** distribution with

- $\mu = 50 \times 0.7 = 35$

According to the binomial principle, *k* has a **near-normal** distribution with

- $\mu = 50 \times 0.7 = 35$
- $\sigma = \sqrt{50 \times 0.7 \times (1 - 0.7)} \approx 3.24$

In particular,

- 68% of the time the number k of heads will be between

$$\mu-\sigma = 31.76 \quad \text{and} \quad \mu+\sigma = 38.24$$

In particular,

- 68% of the time the number k of heads will be between

$$\mu - \sigma = 31.76 \quad \text{and} \quad \mu + \sigma = 38.24$$

Remark: Since k is an integer, really 68% of the time k will be between 32 and 38.

- 95% of the time the number k of heads will be between

$$\mu - 2\sigma = 28.52 \quad \text{and} \quad \mu + 2\sigma = 41.48$$

- 95% of the time the number *k* of heads will be between

  $\mu - 2\sigma = 28.52$ and $\mu + 2\sigma = 41.48$

  So between 29 and 41.

- 95% of the time the number k of heads will be between

$$\mu - 2\sigma = 28.52 \quad \text{and} \quad \mu + 2\sigma = 41.48$$

  So between 29 and 41.
- Finally, 99.7% of the time k will be between

$$\mu - 3\sigma = 25.28 \quad \text{and} \quad \mu + 3\sigma = 44.72$$

- 95% of the time the number k of heads will be between

  $$\mu - 2\sigma = 28.52 \quad \text{and} \quad \mu + 2\sigma = 41.48$$

  So between 29 and 41.
- Finally, 99.7% of the time k will be between

  $$\mu - 3\sigma = 25.28 \quad \text{and} \quad \mu + 3\sigma = 44.72$$

  So between 26 and 44.

Remarks:

- Despite the fact the coin has a 70% chance of landing heads, we can't guarantee that the number k of heads will be 35 (70% of 50).

Remarks:

- Despite the fact the coin has a 70% chance of landing heads, we can't guarantee that the number k of heads will be 35 (70% of 50).
- All we can say is that k will be...

- between 32 and 38, with somewhat low certainty; only 68% confidence.

- between 32 and 38, with somewhat low certainty; only 68% confidence.
- between 29 and 41, with a reasonable 95% confidence.

- between 32 and 38, with somewhat low certainty; only 68% confidence.
- between 29 and 41, with a reasonable 95% confidence.
- between 26 and 44; almost certainly, 99.7% confidence.

The previous example is somewhat unrealistic. We just magically know the probability p of getting heads?

More realistic: we want to estimate the probability of heads p from a particular run of the experiment (from k or $\hat{p} = \frac{k}{n}$). To do so, we need to reverse our equations, since we're reversing whether we know p or $\hat{p}$.

More realistic: we want to estimate the probability of heads p from a particular run of the experiment (from k or $\hat{p} = \frac{k}{n}$). To do so, we need to reverse our equations, since we're reversing whether we know p or $\hat{p}$.

$$p \longleftarrow \hat{p}$$

Example

- We suspect that a coin has been doctored.

Example

- We suspect that a coin has been doctored.
- We perform a run of $n = 100$ flips to determine whether it is

Example

- We suspect that a coin has been doctored.
- We perform a run of $n = 100$ flips to determine whether it is
- Of the flips, $k = 58$ come up heads.

Example

Questions:

(a) Estimate the probability p of heads for this coin.

Example

Questions:

(a) Estimate the probability p of heads for this coin.

(b) Is the coin doctored?

Remarks:

- A quick naive answer would be that yes it's doctored, because $k \neq 50$ so $\hat{p} \neq 50\%$.

Remarks:

- A quick naive answer would be that yes it's doctored, because $k \neq 50$ so $\hat{p} \neq 50\%$.
- This is naive because what we actually want to know is whether $p = 50\%$.

Solution

By the binomial principle k has a near-normal distribution.

Solution

By the binomial principle k has a near-normal distribution.
So with 68% certainty

$$\mu - \sigma \leq k \leq \mu + \sigma$$

Solution

Plugging in $\mu = p \cdot n$ *and* $\hat{p} = \frac{k}{n}$ *this eventually becomes:*

$$\hat{p} - \frac{\sigma}{n} \leq \ p \ \leq \hat{p} + \frac{\sigma}{n}$$

So, we have window in which to expect p, namely within $\frac{\sigma}{n}$ of $\hat{p}$.

So, we have window in which to expect p, namely within $\frac{\sigma}{n}$ of $\hat{p}$. Danger:

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

So, we have window in which to expect p, namely within $\frac{\sigma}{n}$ of $\hat{p}$. Danger:

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

which involves p itself! Bad.

So, we have window in which to expect p, namely within $\frac{\sigma}{n}$ of $\hat{p}$.
Danger:

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)}$$

which involves p itself! Bad.
But assuming $p \approx \hat{p}$, we can say

$$\sigma \approx \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}.$$

In our example,

$$\sigma \approx \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}$$

$$= \sqrt{100 \times \frac{58}{100} \times \left(1 - \frac{58}{100}\right)} = 4.94,$$

so with 68% confidence we can say

$$\frac{58 - 4.94}{100} \leq p \leq \frac{58 + 4.94}{100}$$

so with 68% confidence we can say

$$\frac{58 - 4.94}{100} \leq p \leq \frac{58 + 4.94}{100}$$

$$0.5306 \leq p \leq 0.6294.$$

so with 68% confidence we can say

$$\frac{58 - 4.94}{100} \leq p \leq \frac{58 + 4.94}{100}$$

$$0.5306 \leq p \leq 0.6294.$$

- 0.5 is not in this interval

so with 68% confidence we can say

$$\frac{58 - 4.94}{100} \leq p \leq \frac{58 + 4.94}{100}$$

$$0.5306 \leq p \leq 0.6294.$$

- 0.5 is not in this interval
- looks like a doctored coin!

so with 68% confidence we can say

$$\frac{58 - 4.94}{100} \leq p \leq \frac{58 + 4.94}{100}$$

$$0.5306 \leq p \leq 0.6294.$$

- 0.5 is not in this interval
- looks like a doctored coin!
- but we have only 68% confidence in this result

- We can increase the confidence of our result if we use 2σ of σ.

- We can increase the confidence of our result if we use 2σ of σ.
- Repeating the above calculations, now with a 95% confidence:

$$0.4812 \leq p \leq 0.6788$$

- We can increase the confidence of our result if we use 2σ of σ.
- Repeating the above calculations, now with a 95% confidence:

$$0.4812 \leq p \leq 0.6788$$

- This time the possibility of $p = 0.5$ is there

Conclusion:

- the coin looks suspicious

Conclusion:

- the coin looks suspicious
- but we cannot conclude that it has been doctored; not with a decent confidence level.

Confidence Intervals

Suppose we perform a two-outcome random experiment n times ($n \geq 30$) and count the number k of successes.

Confidence Intervals

Suppose we perform a two-outcome random experiment n times ($n \geq 30$) and count the number k of successes.

- denote by $\hat{p} = \dfrac{k}{n}$ the percent of successes

Confidence Intervals

Suppose we perform a two-outcome random experiment n times ($n \geq 30$) and count the number k of successes.

- denote by $\hat{p} = \dfrac{k}{n}$ the percent of successes
- and use the approximation $\sigma \approx \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}$

We can estimate the probability p:

We can estimate the probability p:

- With a 68% confidence level:

$$\hat{p} - \frac{\sigma}{n} \leq p \leq \hat{p} + \frac{\sigma}{n}$$

We can estimate the probability p:

- With a 68% confidence level:

$$\hat{p} - \frac{\sigma}{n} \leq p \leq \hat{p} + \frac{\sigma}{n}$$

We call $\left[\hat{p} - \frac{\sigma}{n}, \hat{p} + \frac{\sigma}{n}\right]$ the **68% confidence interval**.

- With a 95% confidence level we can say that:

$$\hat{p} - 2\frac{\sigma}{n} \leq p \leq \hat{p} + 2\frac{\sigma}{n}$$

- With a 95% confidence level we can say that:

$$\hat{p} - 2\frac{\sigma}{n} \leq p \leq \hat{p} + 2\frac{\sigma}{n}$$

We call $\left[\hat{p} - 2\frac{\sigma}{n}, \hat{p} + 2\frac{\sigma}{n}\right]$ the **95% confidence interval**.

- With a 99.7% confidence level we can say that:

$$\hat{p} - 3\frac{\sigma}{n} \leq p \leq \hat{p} + 3\frac{\sigma}{n}$$

- With a 99.7% confidence level we can say that:

$$\hat{p} - 3\frac{\sigma}{n} \leq p \leq \hat{p} + 3\frac{\sigma}{n}$$

We call $\left[\hat{p} - 3\frac{\sigma}{n}, \hat{p} + 3\frac{\sigma}{n}\right]$ the **99.7% confidence interval**.

- Note the expression $\dfrac{\sigma}{n}$ in all of the confidence intervals.

- Note the expression $\frac{\sigma}{n}$ in all of the confidence intervals.
- This is the standard deviation as a percentage of the sample size.

Definition

We call $\frac{\sigma}{n}$ the **standard error**

$$\textbf{st.err.} = \frac{\sigma}{n} \approx \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Let's consider the diapers example again:

Example

- Pick 100 diapers and test them.

Example

- Pick 100 diapers and test them.
- 20 of the diapers in the sample are defective.

Example

- Pick 100 diapers and test them.
- 20 of the diapers in the sample are defective.
- Estimate the percentage of defective diapers in the production line

- Decide on the confidence level.

- Decide on the confidence level.
- In most situations 95% is appropriate, so let's do that.

- Decide on the confidence level.
- In most situations 95% is appropriate, so let's do that.
- So we get the confidence interval by going 2 standard errors away from $\hat{p}$.

Solution

We have $n = 100, \; k = 20$ *so*

Solution

We have $n = 100, \ k = 20$ *so*

- $\hat{p} = \frac{20}{100} = 0.2$

Solution

We have $n = 100,\ k = 20$ *so*

- $\hat{p} = \frac{20}{100} = 0.2$
- $\sigma \approx \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})} = \sqrt{100 \cdot 0.2 \cdot 0.8} = 4$

Solution

We have $n = 100, \ k = 20$ *so*

- $\hat{p} = \frac{20}{100} = 0.2$
- $\sigma \approx \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})} = \sqrt{100 \cdot 0.2 \cdot 0.8} = 4$

The standard error is

$$\textbf{st.err.} = \frac{\sigma}{n} = \frac{4}{100} = 0.04$$

Solution

Therefore the 95% confidence interval is $[0.12, 0.28]$, *i.e.*

$$0.12 \leq p \leq 0.28$$

with probability 95%

Conclusion:

Conclusion:

- We're pretty confident that not more than 28% of all diapers are leaky.

Conclusion:

- We're pretty confident that not more than 28% of all diapers are leaky.
- This is quite a bit higher than the 20% just from the sample.

Conclusion:

- We're pretty confident that not more than 28% of all diapers are leaky.
- This is quite a bit higher than the 20% just from the sample.
- To gain more confidence we'd need a bigger sample.

Next time: Section 4.4.2: Capture-Recapture method revisited