

AE 17: CLT II- Inference

Dav King

3/15/22

Coming up

- Lab due on Friday at 11:59 PM.

Main ideas

- Apply the ideas from last class to use the CLT to carry out hypothesis tests.
- Learn how to use the `t_test` function.

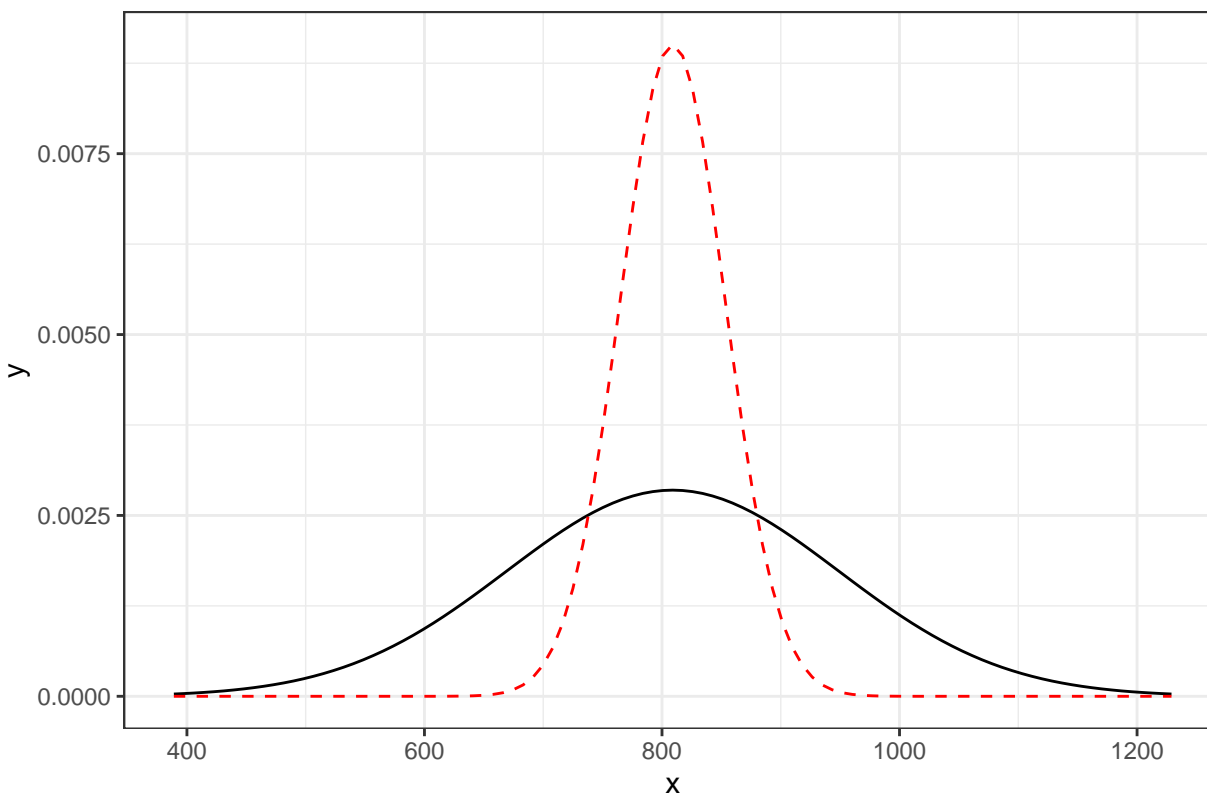
CLT-based inference with Pokemon (Gotta hypothesis test them all)

```
library(tidyverse)
library(tidymodels)
```

Plotting our Example from Last Class

```
library(tidyverse)
ggplot(data = data.frame(x = c(809 - 140*3, 809 + 140*3)), aes(x = x)) +
  stat_function(fun = dnorm, args = list(mean = 809, sd = 140),
              color = "black") +
  stat_function(fun = dnorm, args = list(mean = 809, sd = 140/sqrt(10)),
              color = "red", lty = 2) + theme_bw() +
  labs(title = "Black solid line = population dist., Red dotted line = sampling dist.")
```

Black solid line = population dist., Red dotted line = sampling dist.



Recall

The Central Limit Theorem

For a population with a well-defined mean μ and standard deviation σ , these three properties hold for the distribution of sample average \bar{X} , assuming certain conditions hold:

1. The mean of the sampling distribution is identical to the population mean μ ,
2. The standard deviation of the distribution of the sample averages is σ/\sqrt{n} , or the **standard error** (SE) of the mean, and
3. For n large enough (in the limit, as $n \rightarrow \infty$), the shape of the sampling distribution of means is approximately *normal* (Gaussian).

Conditions

What are the conditions we need for the CLT to hold?

- **Independence:** The sampled observations must be independent. This is difficult to check, but the following are useful guidelines:
 - the sample must be random

- if sampling without replacement, sample size must be less than 10% of the population size
- **Sample size / distribution:**
 - if data are numerical, usually $n > 30$ is considered a large enough sample, but if the underlying population distribution is extremely skewed, more might be needed
 - if we know for sure that the underlying data are normal, then the distribution of sample averages will also be exactly normal, regardless of the sample size
 - if data are categorical, at least 10 successes and 10 failures.

CLT results: \bar{X} , \hat{p}

Assuming the conditions for the CLT hold, \bar{X} approximately has the following distribution:

$$\text{Normal}(\mu, \sigma/\sqrt{n})$$

Equivalently, we can define the quantity Z , such that $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, where Z has the following distribution:

$$\text{Normal}(0, 1)$$

Assuming the conditions for the CLT hold, \hat{p} approximately has the following distribution:

$$\text{Normal}\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

We can standardize this in a similar way and define a quantity Z that is normally distributed with a mean of 0 and a standard deviation of 1.

The hypothesis testing framework

1. Start with two hypotheses about the population: the null hypothesis and the alternative hypothesis.
2. Choose a (representative) sample, collect data, and analyze the data.
3. Figure out how likely it is to see data like what we observed, **IF** the null hypothesis were in fact true.
4. If our data would have been extremely unlikely if the null claim were true, then we reject it and deem the alternative claim worthy of further study. Otherwise, we cannot reject the null claim.

The “errors”

Suppose we test a certain null hypothesis, which can be either true or false (we never know for sure!). We make one of two decisions given our data: either reject or fail to reject H_0 .

We have the following four scenarios:

Decision	H_0 is true	H_0 is false
Fail to reject H_0	Correct decision	Type II Error
Reject H_0	Type I Error	Correct decision

It is important to weigh the consequences of making each type of error.

CLT-based testing

Testing comparison

What changes now that we plan to use a CLT-based approach in doing our testing?

We no longer have to simulate the null distribution. The Central Limit Theorem gives us an approximation for the distribution of our point estimate under the null hypothesis.

Rather than work directly with the sampling distribution of the point estimates, we'll use standardized versions that we'll call **test statistics**.

For tests of μ :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

where μ_0 is the value of μ under the null hypothesis.

For tests of p :

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

where p_0 is the value of p under the null hypothesis.

Test statistic and p-value

Recall step 3 of our testing framework: Figure out how likely it is to see data like what we observed, **IF** the null hypothesis were in fact true.

To do this:

1. Compute the test statistic's value - all information is obtained from the sample data or value of the parameter under the null hypothesis.
2. To quantify how likely it is to see this test statistic value given the null hypothesis is true, compute the probability of obtaining a test statistic as extreme or more extreme than what we observed. This probability is calculated from a known distribution.

The Data

We will be using the `pokemon` dataset, which contains information about 45 randomly selected Pokemon (from all generations). You may load in the dataset with the following code:

```
pokemon <- read_csv("pokemon.csv")
```

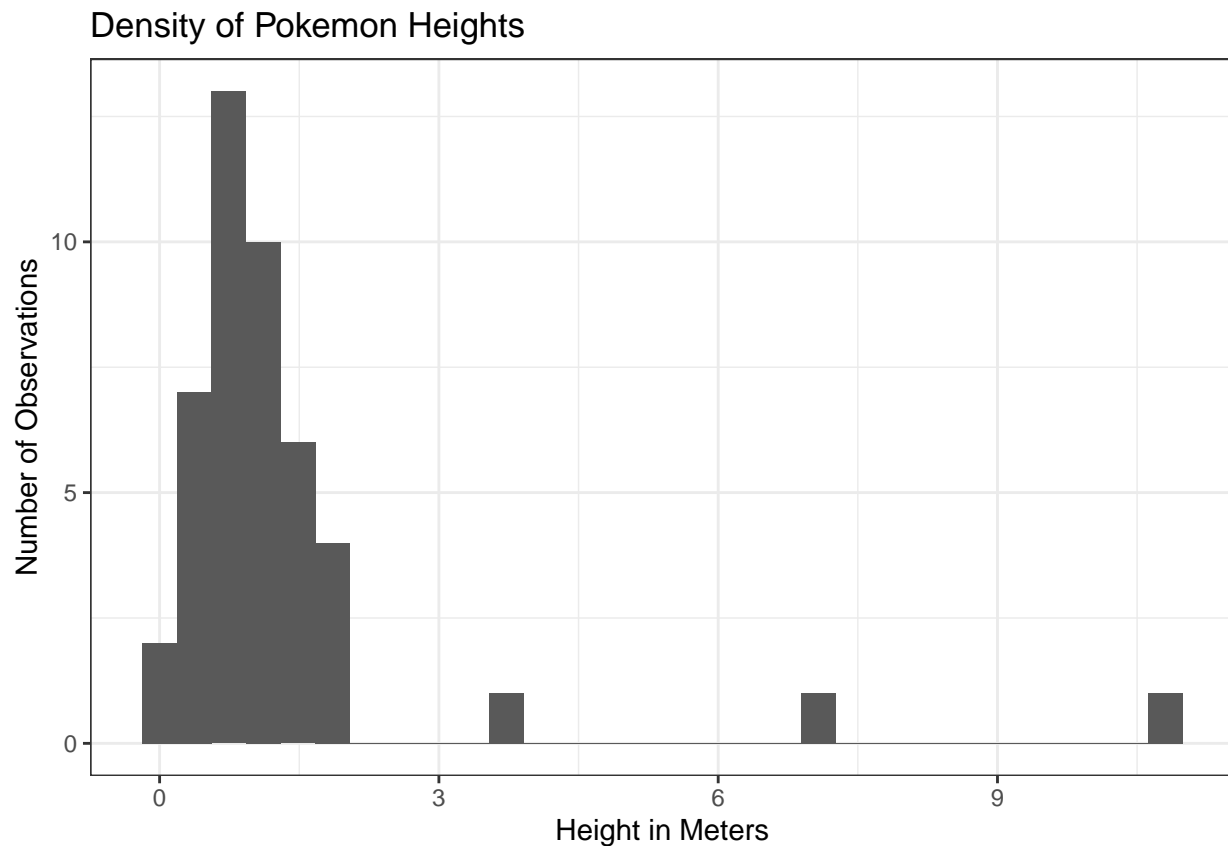
Exercise 1

Let's start by looking at the distribution of `height_m`, the typical height in meters for Pokemon, using a visualization and summary statistics.

Please make a histogram and then find summary statistics.

```
ggplot(pokemon, aes(x = height_m)) +
  geom_histogram() +
  labs(title = "Density of Pokemon Heights", x = "Height in Meters",
        y = "Number of Observations") +
  theme_bw()
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



```
summary(pokemon$height_m)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.100   0.600   1.000   1.344   1.400  10.900
```

```
sd(pokemon$height_m)
```

```
## [1] 1.820284
```

In the previous lecture, we were given the mean, μ , and standard deviation, σ , of the population. That is unrealistic in practice (if we knew μ and σ we wouldn't need to do statistical inference!).

Today we will use our sample data and the Central Limit Theorem to draw conclusions about the μ , the mean height in the population of Pokemon.

- What is the point estimate for μ , i.e., the “best guess” for the mean height of all Pokemon?

1.344

- What is the point estimate for σ , i.e., the “best guess” for the standard deviation of the distribution of Pokemon heights?

1.820

Exercise 2

Before moving forward, let's check the conditions required to apply the Central Limit Theorem. Are the following conditions met:

- Independence?

Yes - it was sampled randomly and we can reasonably assume 45 is less than 10% of the total population.

- Sample size/distribution?

Yes - the sample size is larger than 30.

Exercise 3

Construct and interpret a 95% confidence interval for the mean height in meters (`height_m`) of Pokemon species by using the Central Limit Theorem.

The formula for a t- confidence interval is:

$$\bar{x} \pm t_{n-1}^* * s / \sqrt{n}$$

```
t.test(pokemon$height_m)
```

```
##
## One Sample t-test
##
## data:  pokemon$height_m
## t = 4.9546, df = 44, p-value = 1.12e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.7975705 1.8913183
## sample estimates:
## mean of x
##  1.344444
```

```
#Yes, I understand how to do this the slow and tedious way.
```

We are 95% confident that the true population mean height in meters of Pokemon species lies on the interval (0.798, 1.891).

Exercise 4

The average height of a human is 1.65 meters. Evaluate whether a randomly selected Pokemon species has a *different* mean height by using the Central Limit Theorem.

In doing so, state your null and alternative hypotheses, the distribution of your test statistic under the null hypothesis, your p-value, decision, and conclusion in context of the research problem. Please use the `t_test` function here.

$$H_0 : \mu = 1.65 \quad H_a : \mu \neq 1.65$$

Under H_0 , $\mu = 1.65$ and σ is unknown.

```
t_test(pokemon, response = height_m, mu = 1.65)
```

```
## # A tibble: 1 x 7
##   statistic t_df p_value alternative estimate lower_ci upper_ci
##   <dbl> <dbl> <dbl> <chr>          <dbl>    <dbl>    <dbl>
## 1    -1.13   44  0.266 two.sided      1.34    0.798    1.89
```

At $\alpha = 0.05$, we fail to reject H_0 . We do not have significant evidence to suggest that a randomly selected Pokemon species has a different mean height.

Exercise 5

Now evaluate whether a randomly-selected Pokemon species has a *lower* mean height by using the Central Limit Theorem. In doing so, state your null and alternative hypotheses, the distribution of your test statistic under the null hypothesis, your p-value, decision, and conclusion in context of the research problem.

$$H_0 : \mu = 1.65 \quad H_a : \mu < 1.65$$

Under H_0 , $\mu = 1.65$ and σ is unknown.

```
t_test(pokemon, response = height_m, alternative = "less", mu = 1.65)
```

```
## # A tibble: 1 x 7
##   statistic t_df p_value alternative estimate lower_ci upper_ci
##   <dbl> <dbl> <dbl> <chr>          <dbl>    <dbl>    <dbl>
## 1    -1.13   44  0.133 less          1.34    -Inf    1.80
```

At $\alpha = 0.05$, we fail to reject H_0 . We do not have significant evidence to suggest that a randomly selected Pokemon species has a lower mean height than humans.

Exercise 6

Suppose the true mean height among Pokemon species is 1.4 meters. In your conclusions from Exercises 2 and 3, did you make the correct decision, a Type 1 error, or a Type 2 error? Explain.

Arguably, we made a type 2 error - that is, that H_0 was false but that we failed to reject it. This could be fixed by lowering the confidence level from 0.95. However, given just how high our p_values were (0.266 and 0.133, respectively), we made the correct decisions with these samples even if they may not have been representative of the population. Given that 1.4 and 1.65 are not huge differences for such a small sample size, though this is technically a type 2 error, I have no issue with it and consider it broadly correct.