Proportions and Normality

Grinnell College

October 15, 2025

Warm up

Suppose I have a population of turtles where the true average length of their shell is $\mu = 5.15$ inches. Further, suppose I have collected two samples with the following properties:

- **Sample 1:** $\overline{x}_1 = 5.7$, $\hat{\sigma}_1 = 0.6$, $n_1 = 25$
- **Sample 2:** $\overline{x}_2 = 4.95$, $\hat{\sigma}_2 = 0.58$, $n_2 = 25$

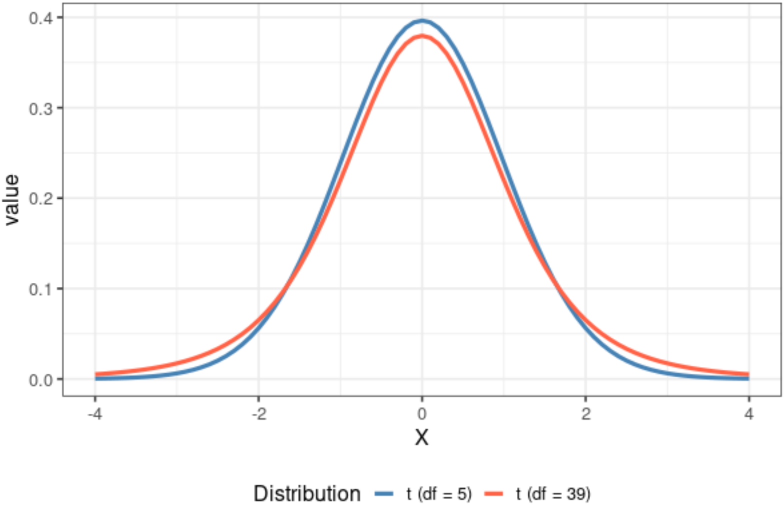
Compute the t-statistic for each variable. Based on this value, which statistic should be considered further away from the true mean?

Warm up 2

Suppose I have a population of turtles where the true average length of their shell is $\mu = 5.15$ inches. Further, suppose I have collected two samples with the following properties:

- **Sample 1:** $\overline{x}_1 = 5.7$, $\hat{\sigma}_1 = 0.6$, $n_1 = 6$
- **Sample 2:** $\overline{x}_2 = 4.95$, $\hat{\sigma}_2 = 0.58$, $n_2 = 40$

Compute the t-statistic for each variable. What is the *sampling distribution* for each statistic? Based on the value and associated sampling distribution, which statistic should be considered further away from the true mean?


0.4
0.3
0.2
0.1
0.0
value
-4
-2
0
2
4
X
Distribution
t (df = 5)
t (df = 39)

Review

- Relation of t-distribution to normal distribution
- Distributional parameters of t-distribution
- Relating sampling distribution to critical values

Birb Example

Suppose I have collected a sample of sharp-shinned hawks from a population where the average Hallux length is $\mu = 11.84$mm

Sample:

- $\overline{x} = 11.515$mm
- $\hat{\sigma} = 1.09$mm
- $n = 20$

1. Create a 90% confidence interval based on the sample
2. How *extreme* is my sample statistic relative to the true mean?

Proportions as Means

There is an interesting relationship between means and proportions

For example, consider taking a coin and flipping it 10 times. How many heads would you expect to see?

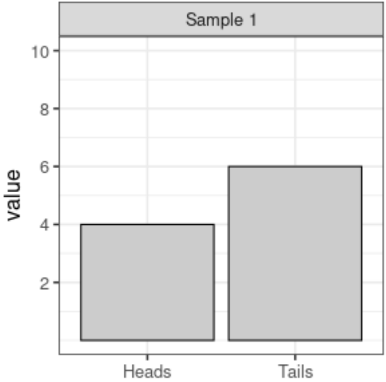
$$\text{Flips} = \{H, H, T, T, H, T, H, T, T, T\}$$
$$X = \{1, 1, 0, 0, 1, 0, 1, 0, 0, 0\}$$

We can find the *proportion* of heads from our sample by simply taking the total number of heads and dividing by the total number of flips, giving

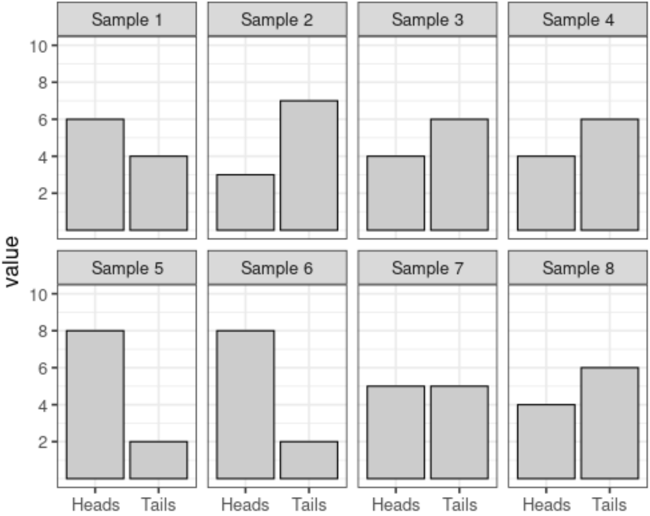
$$\hat{p} = \frac{4}{10}$$

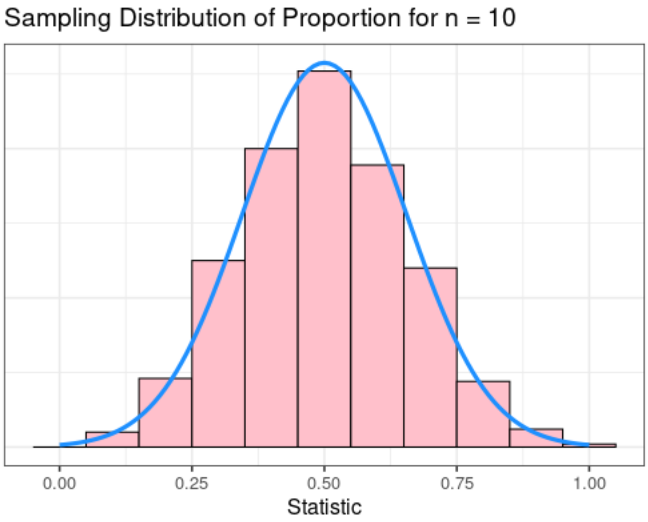
However, if we consider X, which defines H as 1 and T as 0, we can also find the sample mean:

$$\begin{aligned} \overline{x} &= \frac{1}{10} \sum_{i=1}^{n} x_i \\ &= 0.4 \end{aligned}$$



Repeated Samples for $n = 10$




Sampling Distribution of Proportion for n = 10
0.00
0.25
0.50
0.75
1.00
Statistic

Central Limit Theorem

For a sample with one proportion, the sampling distribution of our proportion statistic, $\hat{p}$ is approximately

$$\hat{p} \sim N\left(p,\ \sqrt{\frac{p(1-p)}{n}}\right)$$

There a few rules of thumb relating to the size and the proportion:

1. $n \times p \geq 10$
2. $n \times (1-p) \geq 10$

In particular, it is often difficult to estimate proportions precisely that are near the boundaries (0 and 1)

t-statistics

Just as with the sample mean, we can use our estimate of the standard error to create a t-statistic

$$t = \frac{p - \hat{p}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Also as before, this statistic has a sampling distribution, $t \sim t(df = n - 1)$

Example

In a study conducted by Johns Hopkins University researchers investigated the survival of babies born prematurely. They searched their hospital's medical records and found 39 babies born at 25 weeks gestation (15 weeks early), 31 of these babies went on to survive at least 6 months. With your group:

1. Use the appropriate t-distribution to construct a 90% confidence interval estimate for the true proportions of babies born at 25 weeks gestation that are expected to survive
2. An article on Wikipedia suggests that 70% of babies born at a gestation period of 25 weeks survive. Is the Johns Hopkins study consistent with this claim?

Example – Critical Values

As we are looking for an 90% confidence interval with $n = 39$, we need to use the `qt()` with $df = n - 1 = 38$

We can start by finding our sample mean and standard error:

$$\hat{p} = \frac{31}{39} = 0.795$$

$$SE = \sqrt{\frac{0.795(1 - 0.795)}{39}} = 0.065$$

To find our critical values, we could use R or we can use our Critical Value Sheet to find the critical value C for 90% when $df = 39 - 1$. This gives us

$$C_{90} = 1.68$$

Example – Confidence Interval

Together, we find an 90% confidence interval

$$\hat{p} \pm C \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

with

$$0.795 \pm 1.686 \times 0.065 = (0.685, 0.905)$$

Example – Checking Consistency

We are also asked if the statistic derived from our sample is consistent with prior studies showing that $p = 0.7$. We can relate our statistic to this parameter with a t-statistic

$$t = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

$$= \frac{0.795 - 0.7}{0.065} = 1.4615$$

How *extreme* it is will again depend on the distribution $t(df = n - 1)$.

Example

Suppose we collect a sample of $n = 32$ cars with front wheel drive where the number of vehicles with manual transmission is 14

1. Find the sample proportion and standard error
2. Create a 90% confidence interval for the true value of the proportion
3. Assume that the true proportion of front wheel drive cars with manual transmission is $p = 0.27$. Based on this, how *extreme* is my sample statistic relative to the true mean?

Key Takeaways

- Proportions share the same properties as the mean
- CLT and t-distribution for proportion
- Confidence intervals and test statistics computed the same way