

HYPOTHESIS TESTING



An explorer claims that the population of rabbits on a remote island are healthy and have a mean weight of 15kg.

An expedition to the island captures 11 adult rabbits and records their weights in kg: 9, 8, 11, 19, 12, 22, 16, 7, 11, 10, 8

We suspect that the mean weight is actually less than 15 and propose to put our theory to the test using statistical methods.

We have a suspicion about the population mean but we will have to estimate the variance using

$$\frac{\sum(x-\bar{x})^2}{n-1} = \frac{\sum(x^2) - n\bar{x}^2}{n-1}$$

$$\bar{x} = \frac{133}{11} = \mathbf{12.1}, \quad \hat{\sigma}^2 = \frac{1845 - (11 \times 12.1^2)}{10} = \mathbf{23.4} \quad \text{So the estimate of the population}$$

standard deviation is $\sqrt{23.4} = \mathbf{4.8} = \hat{\sigma}$

Because the sample size n is small, we use the **t distribution** with $(n-1)$ **10 degrees of freedom**.

$$\frac{\hat{\sigma}}{\sqrt{n}} = \frac{4.8}{\sqrt{11}} = \mathbf{1.45}$$

The distribution was first studied by W. S. Gossett, who was a statistician to the Guinness Brewery and wrote under the pen name “student”.

The distribution is known as “Student’s t distribution”.

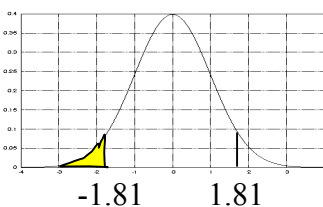
Our assumption about the mean is stated as the **null hypothesis**: $H_0: \mu = 15$

The suspicion is stated as the alternate hypothesis: $H_1: \mu < 15$

$$\text{Our test statistic } t = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} = \frac{12.1 - 15}{1.45} = \mathbf{-2}$$

This t value has to be well below the mean for us to reject the null hypothesis.

The critical value for this rejection is found from t tables.



The t value found from the sample would have to be less than -1.81 in order that we reject our null hypothesis.

Find the value in t-tables row 10 column 0.95

$$t_{95\%}(10) = 1.81 \text{ but}$$

we need to look at the left hand tail.

Since the graph is symmetrical the critical value will be -1.81 .

Since our value of -2 is less than the critical value -1.81 **we reject H_0 at the 5% significance level and conclude that the mean is less than 15.**

A 95% confidence interval for μ is now $\bar{x} \pm t_{95\%}(10) \frac{\hat{\sigma}}{\sqrt{n}}$
 $= 12.1 \pm 1.81 \times 1.45 = (9.5, 14.7).$

Confidence intervals for a proportion.

In the following example it is not the mean but the population proportion that is unknown.

In simple probability questions we are told that the probability of some event is $\frac{a}{b}$, **but how do they know that?**

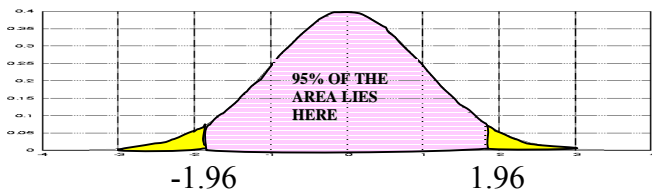
How do they know that the probability that the fish in the lake has some fishy-type disease?

Suppose they take a sample of 500 fish and find that 120 have the particular disease. It would be

easy to say that “The probability of a fish having the disease is $\frac{120}{500}$. And no disease $\frac{380}{500}$.

But how confident are we of the statement considering that we made it on the strength of a sample which would be different on another day?

It is better to give a range for that probability with a stated degree of confidence. So we give a confidence interval for the proportion.



95% confidence interval (0.203, 0.277)

99% confidence interval (0.191, 0.289) **a little wider.**

The confidence interval is quoted directly from the sample proportion and using the normal distribution. It is similar to a confidence interval for a normal distribution.

With the sample proportion $p = \frac{120}{500} = \mathbf{0.24}$

and $q = \frac{380}{500} = \mathbf{0.76}$

A 95% confidence interval for the population

proportion is $\frac{120}{500} \pm \mathbf{1.96} \sqrt{\frac{0.24 \times 0.76}{500}}$

For a 99% confidence interval replace the 1.96 by 2.58.