6 Statistical Inference

We collect samples to get data from them and to find out more about the population which the samples represent. We would like to also know how confident we can be of our inferences from the sample data.

Statistical inference provides methods for drawing conclusions about a population from sample data.

In this class, we are going to look at inferences for the population mean. There are two methods we consider for drawing inferences about the population mean from the sample. In the first case we look to give an estimate of the population mean and in the second, we try to test a guess of the population mean. In both cases we are working on the assumption that we have a random sample, since these methods are based on answering the question as to what would happen if we repeat the sampling many times.

6.1 Estimating with Confidence

Suppose we have a random variable x with unknown mean μ and known standard deviation σ. If we collect a random sample of size n, what can we tell about the population mean μ?

- The law of large numbers states that the sample mean $\bar{x}$ is close to the population mean μ as n gets large.
- The mean of all possible sample means from samples of size n is equal to the population mean μ and the standard deviation of all possible sample means is $\frac{\sigma}{\sqrt{n}}$.
- If the population is normally distributed the sampling distribution of $\bar{x}$ is also normally distributed.
- If the sample size is large, the central limit theorom tells us that $\bar{x}$ is approximately normal.

6.1.1 Statistical confidence

If we choose many samples of size n, where n is large or the population is normally distributed, then the sample mean has a normal distribution.

- The 68-95-99.7 rule states that 95% of all values fall within two standard deviations of the population mean.
- From that we can conclude that the population mean is with two standard deviations of 95% of all observations.
- We can extend this idea sample means, since we know the sampling distribution of the sample mean tends to be normally distributed under certain conditions.
- 95% of all sample means will lie within 2 standard deviations of the sample mean. Hence the population mean is within two standard deviations of 95% of all sample means.

We have 95% of all values of $\bar{x}$ lying with 2 standard deviations of the population mean μ under certain conditions. Suppose the population had standard deviation σ and either the population distribution is normal or we have a large enough sample size (so that the conditions for the central limit theorom are

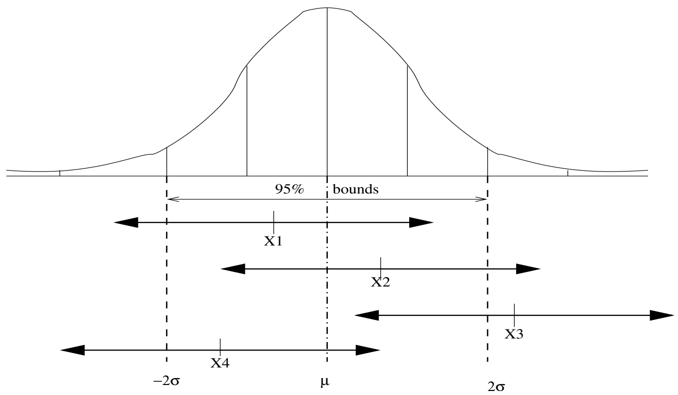

Figure 1: 4 values from a normal distribution

satisfied). Then the standard deviation of $\bar{x}$ is $\frac{\sigma}{\sqrt{n}}$ and the sampling distribution of $\bar{x}$ is approximately normal. Hence 95% of all $\bar{x}$ values lie within $2\frac{\sigma}{\sqrt{n}}$ of μ. Suppose the sample mean $\bar{x}$ of a random sample was within $\mu \pm 2\frac{\sigma}{\sqrt{n}}$. Then μ will lie within $\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}$. Since we know that 95% of all $\bar{x}$ will lie within $\mu \pm 2\frac{\sigma}{\sqrt{n}}$, 95% of all intervals $\bar{x} \pm 2\frac{\sigma}{\sqrt{n}}$ will contain μ.

A **level** C **confidence interval** for a paramter has two parts

- An **interval** calculated from the data, usually of the form **estimate** $\pm$ **margin of error** .
- A **confidence level** C, which gives the probability that the interval will capture the true parameter in repeated samples.

6.1.2 Confidence intervals for the mean μ

Draw an SRS of size n from a population having unknow mean μ and known standard deviation σ. A level C confidence interval for μ is

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

The critical value z^* is found in table C. This interval is exact when the population is normal and is approximately correct for large n in other cases.

Here are some commonly used z^* values.

Confidence Level	z^*
90%	1.645
95%	1.960
99%	2.576

A soda manufacturer wants to estimate the population mean μ of mean content of a can. Suppose the distribution of the content of a can is distributed normally with an unknown mean μ and a known standard deviation σ=3. If a random sample of size $n = 9$ gave us a sample mean $\bar{x} = 298$, construct a 95% confidence interval for μ, the mean content of a can.

6.1.3 How confidence intervals behave

The confidence interval for μ has an estimate and a margin of error. We would like to have a high confidence level, since that would indicate how confident we are of our answers and we would like a low margin of error, since that would indicate how well we have pinned down the parameter. The margin of error is $z^* \frac{\sigma}{\sqrt{n}}$. It has three parts and the margin of error gets smaller when

- z^* gets smaller. This is the same as reducing our confidence since the value of z^* increases and decreases with the confidence levels.
- σ gets smaller. The standard deviation is a measure of how far the values are from the mean. A smaller standard deviation indicates that the values are around the mean and hence its easier to find the true value.
- n gets larger. Increasing our sample size decreases the margin of error for a fixed confidence level, since we are reducing the standard deviation of the sample distribution of $\bar{x}$.

What is the margin of error for the confidence interval calculated for the soda example? What is the margin of error if our sample size n was 36?

6.1.4 Choosing the sample size

The confidence interval for a population mean will have a specified margin of error m when the sample size is

$$n = (\frac{z^* \sigma}{m})^2$$

If the soda manufacturer wanted a 95% confidence interval for the mean content of a soda can, what is the required sample size for a desired margin of error of 1?

6.2 Tests of Significance

Another way of making statistical inference is called *tests of significance* . Here we make a claim about the population and use the sample data to assess evidence for or against your claim.

6.2.1 The reasoning of tests of significance

Suppose we had a claim about the population mean. We then proceed to draw many samples and calculate the sample mean for all those samples. How many of those sample means would make us reject our claim? That is the logic behind hypothesis testing. We are interested in seeking evidence against our claim from the data. Since we cannot or do not want to collect many samples, we ask ourselves how often would we observe the data that we have if our claim was true.

Consider the soda manufacturing example. Suppose we were interested in the claim that the population mean μ was 300 ml. We have a normal population with an unknown mean and a known standard deviation σ=3. The sampling distribution of $\bar{x}$ is normal with mean μ (unknown) and standard deviation $\frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{9}} = 1$. Suppose we observe a sample mean $\bar{x} = 298$. What is the probability of observing such a value if the true mean was indeed 300 ml? If the true mean was indeed 300 ml then the sampling distribution of $\bar{x}$ is normal with mean $\mu = 300$ and standard deviation $\frac{\sigma}{\sqrt{n}} = 1$. The probability of observing a value as extreme as 298 is given by

$$P(\bar{x} < 298) = P(z < \frac{298 - 300}{1}) = P(z < -2) = 0.0228$$

There is 2.28% chance of observing a sample mean as small as 298 if the true mean was indeed 300 ml. This low probability should lead us to believe that our claim sould not be true.

6.2.2 The vocabulary of significance tests

Let us introduce some notaions to formalize the process of hypothesis testing.

Null Hypothesis H_0: The statement being tested is called the **null hypothesis.** The test is designed to assess the strength of evidence against the null hypothesis. We usually write the null hypothesis in this form $H_0 : \mu = \mu_0$, where μ_0 is the hypothesized value we are interested in.

Alternative Hypothesis H_a: The claim about the population for which we are trying to find evidence for is referred to as the alternative hypothesis. It is represented by H_a. The alternative hypothesis can be of the following forms.

- One-sided or one-tailed alternative: $H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$
- Two-sided or two-tailed alternative: $H_a : \mu \neq \mu_0$

***P*-value:** The probability, computed assuming H_0 is true, that the observed outcome would take a value as extreme or more extreme than that actually observed is called the P-value of the test. The smaller the P-value is, the stronger is the evidence against H_0 provided by the data.

A researcher is interested in testing if ST 311 students get enough sleep at night. He thinks that the students should get 7 hours of sleep at night to be fresh for classes the next day. Help him write the hypothesis to test his claim.

6.2.3 More detail: P-values and statistical significance

A significance test uses data in the form of a **test statistic** . The test statistic is based on a statistic that is used to estimate the parameter (We have μ as our parameter of interest and $\bar{x}$ as our test statistic.) We try to assess evidence against H_0 by giving a probability, P-value, of observing a statistic as extreme or more than actually observed, if H_0 were actually true. The direction of the extremeness is determined by the alternative hypothesis H_a.

A small P-value is evidence against the null hypothesis because it indicates a low probability of occurence of such outcomes if the null hypothesis were actually true. But how small is small? We can decide on a fixed value before hand and we would need that much evidence to reject H_0. The decisive value of the P-value is called the **significance level** and is referred to as α. If the P-value is as small or smaller than α then we say that the data are **statistically significant at level α.**

6.2.4 Tests for a population mean

There are three steps in carrying out a significance test.

1. State the hypothesis
2. Calculate the test statistic
3. Find the P-value

z test for a population mean.

To test the hypothesis $H_0 : \mu = \mu_0$ based on an SRS of size n from a population with unknown mean μ and known standard deviation σ, compute the **one-sample z statistic**

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

In terms of a variable Z having a standard normal distribution, the P-value for a test of H_0 against

- $H_a : \mu > \mu_0$ is $P(Z > z)$
- $H_a : \mu < \mu_0$ is $P(Z < z)$
- $H_a : \mu \neq \mu_0$ is $2P(Z > |z|)$

These P-values are exact if the population distribution is normal and are approximately correct for large n in other cases.
The researcher decides to take a random sample of 36 ST 311 students and finds the mean hours of sleep to be 6.8 hours. If the standard deviation of the numbers of hour slept by students is 1.5 hours, then help him find the P-value of the null hypothesis.

6.2.5 Tests with fixed significance level

Sometimes we require our data to be significant at level α. In such cases we require that the P-value be less than or equal to α before we reject the null hypothesis. This is done to ensure some protection against wrongly rejecting the null hypothesis. However since we want to test at a fixed significance level we can use the critical values z^* from table C to help us decide on rejecting the null hypothesis or not.

The number z^* with probability p lying to its right under the standard normal curve is called the **upper** p **critical value** of the standard normal distribution.

To test the hypothesis $H_0 : \mu = \mu_0$ based on an SRS of size n from a population with unknown mean μ and known standard deviation σ, compute the **one-sample** z **statistic**

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Reject H_0 at significance level α against a one-sided alternative
where z^* is the upper α critical value from Table C.

$$\begin{array}{lll} H_a : \mu > \mu_0 & \text{if} & z \geq z^* \\ H_a : \mu < \mu_0 & \text{if} & z \leq -z^* \end{array}$$

Reject H_0 at significance level α against a two-sided alternative
where z^* is the upper $\alpha/2$ critical value from Table C.

$$H_a : \mu \neq \mu_0 \quad \text{if} \quad |z| \geq z^*$$

What is the researcher's conclusion at $\alpha = 5\%$ level? What about $\alpha = 1\%$ level?

6.2.6 Tests from confidence intervals

A level α two-sided significance test rejects a hypothesis $H_0 : \mu = \mu_0$ exactly when the value μ_0 falls outside a level $1 - \alpha$ confidence interval for μ.

How to test a random number generator? A random number generator generates values between 0 and 1, and has a uniform distribution(i.e., each number between 0 and 1 is equally likely.). A 95% confidence interval for the mean of a sample of values generated from a random number is (0.3799,0.4930). Is the random number generator working correctly?

6.3 Making Sense of Statistical Significance

6.3.1 How small a P is convincing?

There is no universal rule to decide how small a P-value we need to decide it our data is significant. It depends on H_0 and the consequences of rejecting H_0.

- How plausible is H_0? $\rightarrow$ If H_0 is a strong assumption, then convincing evidence (small P-value) is required to reject H_0.
- What are the consequences of rejecting H_0? $\rightarrow$ If rejecting H_0 results in expensive changes, again convincing evidence (small P-value) is required to reject H_0.

6.3.2 Type I and Type II errors

If the consequences of wrongly rejecting H_0 (Type I error) are greater than not rejecting H_0 when H_0 is actually incorrect (Type II error), then we need a really small P-value to be convinced that H_0 is wrong. On the other hand if rejecting H_0 incorrectly is something that is better than incorrectly not rejecting H_0, then we would reject H_0 even if the P-value was not that small.

Consider the following two cases.

- If μ was the mean concentration of a chemical in a drug. If μ was supposed to equal 50 mg, and a slight deviation from it is lethal, then even mild evidence (relatively small P-values of the order of 0.1 to 0.2) is good enough to reject $H_0 : \mu = 50$.
- Consider H_0 : Accused is innocent v/s H_a : Accused is guilty. If we believe that someone is innocent till proven guilty (beyond reasonable doubt), then we need stron evidence to reject H_0 (Small P-values of the order of 0.01-0.05).

6.3.3 Statistical significance and practical significance

Statistical significance can be over-rated. Consider the following example.

Is mean GPA falling from previous mean of 3.5? A sample of 10,000 students gave us a mean of 3.49 and assume that the population standard deviation is 0.5. If we wanted to test $H_0 : \mu = 3.5$ v/s $H_a : \mu < 3.5$ we have

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{3.49 - 3.5}{0.5/\sqrt{10000}} = -2$$

and the P-value is $P(z < -2) = 0.0228$ and the results are significant at the 5%-level but we can see that the sample mean is just 0.01 smaller than the population mean and as such the difference maybe statistically significant, but not practially useful.