STAT20060 - Statistics and Probability
Handout 5 - Statistical Inference

Damien McParland

Motivating Example

A Formula 1 racing team are interested in the distance covered (km) by a particular tyre, under reasonable weather conditions, until performance times suffer due to tyre wear. It is assumed that this distance follows an exponential distribution with unknown mean μ. The team test 50 sets of tyres and record the distance covered until tyre wear affects performance times.

- The team would like to estimate the unknown population mean parameter μ.

Sample Statistics

- Often interested in estimating a descriptive property of a population
- Estimate the property using a sample statistic calculated from a sample from the population.

Parameters:

- Fixed values of population characteristics.
  e.g. Mean (μ), variance(σ^2)

Sample Statistics:

- Any quantity calculated from the sample values.
  e.g. Sample mean ($\bar{X}$), sample variance (s^2)

- Sample statistics can be used to estimate population parameters
- A sample statistic is a descriptive measure of a sample from the population.

Sampling Distribution

- Sample statistics are random variables since their value will vary from one sample to the next.
- A **sampling distribution** is the probability distribution of a sample statistic calculated from a sample of size n.
- The sampling distribution describes how the statistic varies from one sample to the next.
- Motivating example: The sample mean, $\bar{X}$ could be used to estimate the population mean, μ.
- Illustrate this sampling distribution in R ...

Comparing Estimators

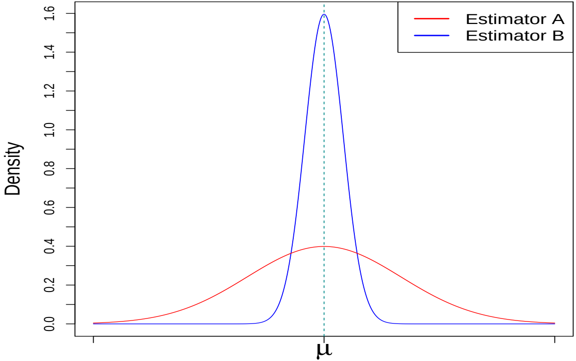
- A statistic is an unbiased estimator if the mean of the sampling distribution is equal to the quantity of interest.

- The standard error of a statistic is the standard deviation of the sampling distribution.

- Sampling distributions are helpful for choosing between estimators of a quantity.

- Generally speaking it is desirable for an estimator to be unbiased and to have a small standard error.

Comparing Estimators



- Both estimators are unbiased.
- Estimator B is preferable due to its smaller standard error.

Example Estimators:

- $\bar{X}$ is an unbiased estimator of the population mean μ.
- s^2 is an unbiased estimator of the population variance σ^2.
- Thus for the formula 1 example:
    - $\bar{X}$ provides an unbiased estimator of the mean time until tyre degradation.
    - s^2 provides an unbiased estimate of the variance of the time until degradation.

Central Limit Theorem

- The central limit theorem makes inferences about the sample mean easy.

Central Limit Theorem (CLT)

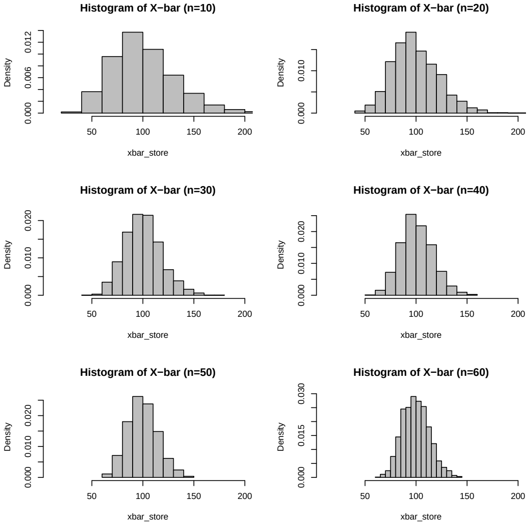
Given a sample of n independent observations from a population with mean μ and variance σ^2 then for n sufficiently large:

$$\bar{X} \;\dot{\sim}\; N\left(\mu, \frac{\sigma^2}{n}\right)$$

- The larger the sample size n the better the approximation.
- As a rule of thumb, for $n \geq 30$ the normal approximation will be reasonable.

Formula 1 Example

$\mu = 100$:



Central Limit Theorem

- Note that the CLT applies when X follows any distribution. Thus for n sufficiently large $\bar{X}$ is approximately normal even if the underlying sample is not.

- The CLT justifies the assumption of a normal distribution for any random variable which can be viewed as the sum of a large number of of independent and identically distributed quantities.

- For example measurement error is often assumed to be normally distributed since each error may be thought of as the sum of many smaller errors.

Example: CD Manufacture

Suppose that a sample of 50 items is taken from several batches of CDs produced in a particular factory and each is tested for defects. It is found that 10 CDs are defective.

- Give an unbiased estimator of the proportion of CDs produced by this factory which are defective.
- The company claim that less than 10% of their CDs are defective. If this is true what is the probability of observing more than 9 defective items in this sample.
- In light of this calculation comment on the company's claim.

Sampling Distribution of Sample Proportion

- In the previous example we used $\bar{X}$ to estimate the probability of a CD being defective.
- Thus by the CLT the sampling distribution of population proportion is approximately normal:

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

where $\hat{p} = \frac{\sum X_i}{n}$.

Confidence Intervals

- A confidence interval for a population parameter is an interval which almost certainly contains the true parameter.
- Almost certainly usually means 95% certain.

If we are given a large sample of values from a population then a $100(1-\alpha)\%$ confidence interval for the population mean, μ, is:

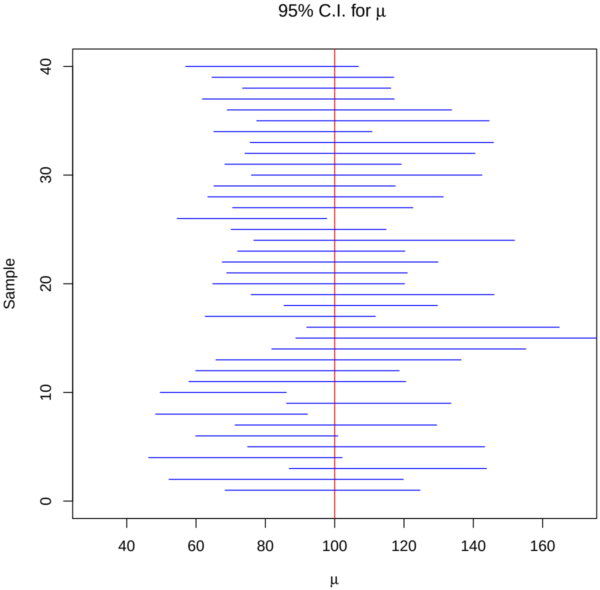
$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

- $Z_{\frac{\alpha}{2}}$ is the value of the standard normal random variable Z such that $\mathbb{P}(Z > Z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$.

Confidence Intervals

- To form a 95% confidence interval $\alpha = 0.05$ and $Z_{\frac{\alpha}{2}} = 1.96$. This value is obtained from the standard normal percentage points table.
- The confidence level of the interval can be adjusted by changing the value of α and hence the value of $Z_{\frac{\alpha}{2}}$. A higher the degree of confidence for a particular sample will lead to a wider interval.
- Notice that the confidence interval defined on the previous slide will vary from sample to sample. 95% of such intervals will contain the true mean value.
- The formula above is a direct result of the central limit theorem.

Confidence Intervals



Example: C.I. for non-normal population

A random sample of 100 observations from a non-normally distributed population possesses a mean of 83.2 and a standard deviation of 6.4.

- Find a 95% confidence interval for the population mean μ and interpret the interval.
- Find a 99% confidence interval for μ.
- Comment on the width of the intervals.

Confidence Interval for Population Proportion

- As was illustrated above, sampling distribution of a sample proportion can be obtained from the CLT. This leads to the following confidence interval for $\hat{p}$.

Large sample confidence interval for p

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

where $\hat{p}$ is the sample proportion.

- The sample is large enough if the number of observed successes and the number of observed failures are both greater than 5.

Example: CD manufacture ... revisited.

Calculate a 95% confidence interval for the proportion of defective CDs produced by the factory in the example earlier. What can you say about the company's claim that only 10% of CDs they produce are defective?

Hypothesis Testing

- Hypothesis testing is one of the most widely used statistical procedures.
- The question which hypothesis tests seek to answer is:

Is the relationship observed in the sample clear enough to be called statistically significant, or could it have been due to chance?

Basic Steps for Testing Hypotheses

1. Determine the null and alternative hypotheses.
2. Collect the data and summarise them with a single number called a test statistic.
3. Determine how unlikely the test statistic would be if the null hypothesis were true.
4. Make a decision.

Step 1: Determine the null and alternative hypotheses.

There are always two hypotheses.

- The first is the null hypothesis (H_0). This usually says that nothing is happening.
  e.g. That there is no relationship or that the relationship is due to chance.
- The alternative hypothesis (H_A) is the research hypothesis. The researcher suspects that the status quo belief is incorrect and that there is indeed a relationship.

The researcher needs to be quite sure before they reject the null hypothesis in favour of the alternative.

e.g. In a trial situation the hypotheses are:

$$H_0: \text{ Defendant is innocent. vs. } H_A: \text{ Defendant is guilty.}$$

Step 2: Collect & summarise the data with a test statistic

- The decision in a hypothesis test is based on a single number summary of the observed data. This summary is called the test statistic.

- There are many different test statistics and the one used depends on the situation.

Step 3: Determine how unlikely the test statistic would be if the null hypothesis were true.

- In order to decide if the results could be due to chance the following question is asked:

If the null hypothesis is true, how likely are we to observe a test statistic at least as extreme as the one we have observed just by chance?

- We usually need statistical tables to answer this question. Which table depends on the test statistic used.

p-value

The *p*-value is computed by assuming the null hypothesis is true, and then asking how likely we would be to observe results at least as extreme as we have under that assumption.

- The p-value does **not** give the probability that the null hypothesis is true

Step 4: Make a decision.

- Once we know how unlikely the observed test statistic is we face two choices.

Choice 1: The *p*-value is not small enough to convincingly rule out chance so we fail to reject the null hypothesis.

Choice 2: The *p*-value is small enough to convincingly rule out chance so we reject the null hypothesis and accept the alternative hypothesis.

- A *p*-value of less than a cut off point called the level of significance (α) is considered small enough to reject the null hypothesis. The standard level of significance is $\alpha = 0.05$.

- Courtroom example:

Choice 1: There is not enough evidence to prove the defendant is not innocent so he/she is not guilty

Choice 2: There is enough evidence to rule out the possibility the defendant is innocent so he/she is guilty.

Large Sample Hypothesis Test for Population Mean

- 1. Specify the null hypothesis. This will be that the population mean is equal to some prior supposed value

$$H_0 : \mu = \mu_0$$

- 2. Specify the alternative hypothesis. There are 3 possibilities here.
    1. $H_A : \mu > \mu_0$, this is a one-tailed (upper tailed) test.
    2. $H_A : \mu < \mu_0$, this is a one-tailed (lower tailed) test.
    3. $H_A : \mu \neq \mu_0$, this is a two-tailed test.

- 3. The test statistic in this situation is a Z-score:

$$Z_\mu = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

  where s is the sample standard deviation and n is the sample size.

Note that μ_0 is the symbol for the numerical value assigned to μ under the null hypothesis.

Large Sample Hypothesis Test for Population Mean

- 4. Determine the rejection region. This is the region where the *p*-value is less than the significance level α and it depends on the alternative hypothesis.
  1. Rejection region: $Z_\mu > Z_\alpha$.
  2. Rejection region: $Z_\mu < -Z_\alpha$.
  3. Rejection region: $Z_\mu > Z_{\frac{\alpha}{2}}$ or $Z < -Z_{\frac{\alpha}{2}}$, i.e. $|Z_\mu| > Z_{\frac{\alpha}{2}}$

- 5. If Z_μ is in the rejection region we reject the null hypothesis and conclude that the alternative is true. If Z_μ is not in the rejection region we say we have insufficient evidence to reject the null hypothesis.

Condition: The sample size is large. (i.e. $n \geq 30$)

Example: Student Rent

A property developer claims that the average rental income per room in student accomodation is at most €5,000 per year. A random sample of 50 students were asked how much their annual rent was and the average was €5200. The sandard deviation of the sample was €735. Do the sample results support the investors claim? (use $\alpha = 0.05$)

Example: Road Usage

A stretch of road is to be upgraded due to the volume of heavy freight traffic using this route. The local corporation say that the average number of heavy-duty vehicles using this road per hour is 71. The engineers believe this number underestimates the true number. To test this, 50 one hour periods are selected over the course of the month and the number of heavy vehicles using the road in each hour are counted. The average number per hour is 74.1 and the sample standard deviation is 13.3. Does the data support the engineers?

1. Test using a significance level of $\alpha = 0.1$.
2. Test using a significance level of $\alpha = 0.01$.

Rejection regions and p-values

- The observed significance level or p-value for a specific statistical test is the probability (assuming H_0 is true) of observing a value of the test statistic that is at least as contradictory to the null hypothesis as the actual one computed from the sample data.

- Determining the rejection region for a test and finding that the test statistic is in it is equivalent to calculating the p-value for the test and finding that it is less than the desired confidence level α.

Large Sample Hypothesis Test for Population Proportion

- 1. The null hypothesis in this case will be

$$H_0 : p = p_0$$

- 2. The alternative hypothesis will be one of:
  1. $H_A : p > p_0$, this is a one-tailed (upper tailed) test.
  2. $H_A : p < p_0$, this is a one-tailed (lower tailed) test.
  3. $H_A : p \neq p_0$, this is a two-tailed test.

- 3. The test statistic used in this situation is also a Z-score:

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

Large Sample Hypothesis Test for Population Proportion

- 4. The rejection regions corresponding to each of the alternative hypotheses respectively are:
  1. Rejection region: $Z_p > Z_\alpha$.
  2. Rejection region: $Z_p < -Z_\alpha$.
  3. Rejection region: $Z_p > Z_{\frac{\alpha}{2}}$ or $Z < -Z_{\frac{\alpha}{2}}$, i.e. $|Z_p| > Z_{\frac{\alpha}{2}}$

- 5. As before we reject the null hypothesis if Z_p is in the rejection region and fail to reject otherwise.

Conditions:

- The sample must come from a binomial population.
- The sample size n must be large. That is it must contain at least 5 failures and at least 5 successes. Equivalently we must have: $n\hat{p} \geq 5$ and $n(1 - \hat{p}) \geq 5$.

Example: Weathering Steel

Controversy surrounds the use of weathering steel in the construction of bridges. Critics say there are serious corrosive problems with this type of steel and urging the government to ban its use in bridge construction. The steel company say that 95% of such bridges have no major corrosive problems. Engineers select 60 such bridges at random and find that 54 show no major corrosive damage. Is there evidence at a significance level of $\alpha = 0.05$ that more than 5% of bridges have major corrosive damage?

Example: Airline Departure Times

A budget airline claims that 96% of its flights depart on time. A researcher working for a competitor records deperture times of 80 randomly selected flights and discovers that 5 departed late. Test the airline's claim at the 1% significance level.

Large Sample Inference for the Difference Between Two Means

- Sometimes it is of interest to compare the means of two populations to see if they are significantly different.

- e.g. Tyre wear. Two types tyre composed of different types of rubber are compared to see which is more durable. It is of interest to see if there is a significant difference between the means of the populations (denoted μ_1 and μ_2 respectively).

- Hypothesis tests and confidence intervals are available for this type of comparison. We focus on the case where we have large sample sizes.

Large Sample Confidence Interval for $(\mu_1 - \mu_2)$

- By the CLT the sampling distribution of ($\bar{X}_1$ - $\bar{X}_2$) is approximately:

$$(\bar{X}_1 - \bar{X}_2) \ \dot{\sim} \ N\left[(\mu_1 - \mu_2), \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)\right]$$

- From this we can derive the form of a confidence interval.

Large Sample Inference for $(\mu_1 - \mu_2)$

If we are given large samples of values from two populations then a $(1 - \alpha)\%$ confidence interval for the difference between the population means, $(\mu_1 - \mu_2)$, is:

$$(\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Large Sample Hypothesis Test for $(\mu_1 - \mu_2)$

- 1. The null hypothesis in this case will be

$$H_0 : (\mu_1 - \mu_2) = D_0$$

- 2. The alternative hypothesis will be one of:
  1. $H_A : (\mu_1 - \mu_2) > D_0$, this is a one-tailed (upper tailed) test.
  2. $H_A : (\mu_1 - \mu_2) < D_0$, this is a one-tailed (lower tailed) test.
  3. $H_A : (\mu_1 - \mu_2) \neq D_0$, this is a two-tailed test.

- 3. The test statistic used in this situation is also a Z-score:

$$Z_{(\mu_1 - \mu_2)} = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Large Sample Hypothesis Test for $(\mu_1 - \mu_2)$

- 4. The rejection regions corresponding to each of the alternative hypotheses respectively are:
  1. Rejection region: $Z_{(\mu_1-\mu_2)} > Z_\alpha$.
  2. Rejection region: $Z_{(\mu_1-\mu_2)} < -Z_\alpha$.
  3. Rejection region: $Z_{(\mu_1-\mu_2)} > Z_{\frac{\alpha}{2}}$ or $Z < -Z_{\frac{\alpha}{2}}$, i.e. $|Z_{(\mu_1-\mu_2)}| > Z_{\frac{\alpha}{2}}$

- 5. As before we reject the null hypothesis if $Z_{(\mu_1-\mu_2)}$ is in the rejection region and fail to reject otherwise.

Conditions: The sample sizes must be large (i.e $n_1 \geq 30$ and $n_2 \geq 30$).

Example: Manual Dexterity and Sport

In a study to investigate the relationship between manual dexterity and sport, the manual dexterity of two groups of children was tested. The first group consisted of a random sample of 37 children who do not play sport and the sample mean and standard deviation of this group was 31.68 and 4.56 respectvely. The sample mean and standard deviation of the second group was 32.19 and 4.34 respectively. The second group consisted of a random sample of 37 children who do play sport. Test the hypothesis that there is no difference between the mean manual dexterity scores ($H_0 : \mu_1 = \mu_2$) versus the alternative that that those who participate in sport have a higher average score ($H_A : \mu_1 < \mu_2$). Use $\alpha = 0.05$.

Notes:

- There is a correspondence between confidence intervals and hypothesis tests.
- Suppose we have a 95% confidence interval for a population parameter θ. This interval is equivalent to rejecting a null hypothesis that θ lies outside this interval at the 5% level.
- There are many other hypothesis tests. For further examples see any introductory statistics text.
- Each test has conditions which must be satisfied for the test to be valid.