MATH1905 Statistics
Lecture 15-16

Lecturer: Marc Raimondo
Carslaw 817
Lectures: Mon.11am (Chem LT1), Tue.8am (Chem LT1)
Student consultations: Tuesday 2-3pm
General information, tutes, solutions etc...
http://www.maths.usyd.edu.au/u/UG/JM/MATH1905/
or

First Year Office (FYO), Carslaw 520.

Lecture 15-16 Inference

This is the third part of the course which is mainly concerned with *hypothesis testing.* We first illustrate the statistical reasoning which underpins all such tests. A coin is tossed 12 times.Although it looks like an ordinary coin, there is reason to suspect that it $might$ be a trick coin, weighted for $tails$.

(i) Suppose it comes up 12 tails. Then it is very likely it is a trick coin.WHY?If X represents the number of tails in 12 throws when a **fair**coin is tossed, then $X \sim \mathcal{B}(12, \frac{1}{2})$ and $\mathrm{P}(X = 12) = (0.5)^{12} = 0.000244$ to 6 d.p. This means that it is very unlikely (less than 3 chances in 10,000)that such an extreme result should occur by chance alone, providing very strong evidence that it is a trick coin.

(ii) Now suppose it comes up exactly 10 tails. If $\boldsymbol{X}$ represents the number of tails in 12 throws when a **fair**coin is tossed, then
$\boldsymbol{X \sim \mathcal{B}(12, \frac{1}{2})}$ and
$\mathbf{P}(\boldsymbol{X} \geq \mathbf{10}) = \mathbf{1} - \boldsymbol{P}(\boldsymbol{X} \leq \mathbf{9}) = \mathbf{0.019(3}\boldsymbol{dp}\mathbf{)}$
There is still a case that it is a trick coin, because it is so unlikely that such an extreme result as 10 (or even more) tails should occur with a fair coin (less than 2 chances in 100), providing fairly strong evidence that it is a trick coin.

(iii) Finally, suppose it comes up exactly 8 tails. If a **fair**coin is tossed, then $\boldsymbol{X \sim \mathcal{B}(12, \frac{1}{2})}$ and $\mathbf{P}(\boldsymbol{X} \geq \mathbf{8}) = \mathbf{1} - \boldsymbol{P}(\boldsymbol{X} \leq \mathbf{7}) = \mathbf{0.194(3}\boldsymbol{dp}\mathbf{)}$
There is not a very strong case now that the coin is a trick one,since if the coin is fair,12 tosses yield 8 or more tails nearly 20% of the time. Notice that the smaller the probability of a result *as unusual as* the observed one, the stronger our feeling that the coin is a trick coin.

This is an example of a *test of significance*.

Null hypothesis H_0 and alternative hypothesis H_1.

H_1 :is the hypothesis of an effect we are interested to assess.

H_0 :is called the null hypothesis because it is the hypothesis that the effect is null (No effect).

In the coin tossing example, if $p = P(Tail)$

H_1. The effect is weighted coin for Tail:$p > \frac{1}{2}$

H_0. The null hypothesis is fair coin(no effect): $p = \frac{1}{2}$.

Test statistic τ.

The test statistic is chosen so that EXTREME values of τ favours the alternative H_1. In other words, The test statistic is chosen so that EXTREME values of τ favours the effect we are interested in. By EXTREMEwe mean extremely large values of τ or extremely small value of τ.

In the coin-tossing example we can take $\tau = X$ the number of tails in the sample.

Clearly if τ is large , this means we obverse a large number of tailswhich favours the hypothesis of the weigthed coin$H_1 : p > \frac{1}{2}$

P-value.

P-value : The probability, computed assuming that H_0 is true (no effect), that the test statistic would take a value as extreme as or more extreme than actually observed.

In scenario ii) of our coin tossing example, we observe $\tau_{obs} = 10$ tails. Large values of τ favour H_1, so

ASSUMING THAT H_0 IS TRUE: $\tau = X \sim \mathcal{B}(12, \frac{1}{2})$
$P = P_{H_0}(\tau \geq \tau_{obs}) = \mathrm{P}(X \geq 10)$
$P = 1 - P(X \leq 9) = 0.019(3dp)$

Findings. The smaller the P-value, the stronger the evidence against H_0 is provided by the data. In other words, the smaller the P-value, the stronger the evidence in favour of H_1 (effect) is provided by the data.

In scenario ii of our coin tossing example, we find that

$$P = 0.019$$

Findings: $P = 0.019$ is quite small (less than 2 chances in 100).It means that it is quite unlikely that such an extreme result as 10 (or even more) tails should occur with a fair coin (less than 2 chances in 100), providing fairly strong evidence that it is a trick coin.
If $P < 0.05$ we say that the data supports (favours) H_1

If $P > \mathbf{0.05}$ we say that the data supports (favours) H_0

(Eight point check)

1. **Null hypothesis H_0(no effect)**
2. **Alternative hypothesis H_1(effect)**
3. **Test statistic τ:extreme values favours of τ favours H_1**
4. **Sampling distribution of τ: depends on τe.g. $B(n, p)$**
5. **Which observed values of τ favour H_1? large or small?**
6. **Observed value, τ_{obs}, of τ from the sample: computed from the data**
7. **P-value,**in the light of 5: $P_{H_0}(\tau \geq \tau_{obs})$ or $P_{H_0}(\tau \leq \tau_{obs})$
8. **Findings**

If the P-value is < 0.05, data supports H_1(significant evidence for H_1)
If the P-value is > 0.05, data rejects H_1(no significant evidence for H_1)

Important things to remember

There is no final proof that H_0 is true or false.

A small P-value means one of two things has happened: $either$ H_0 is true and the discrepancy is due to an unlikely event, or H_0 is false.

A large P-value simply indicates that the data are consistent with the truth of H_0.

A P-value < 0.05 is often considered a convincing argument against H_0 and we adopt this convention in this book. However a value other than 0.05 can sometimes be used.

The smaller the P-value,the stronger the evidence in favour of H_1

There is an asymmetry between H_0 and H_1. We proceed on the assumption that H_0 is true,and investigate the strength of evidencein favour of H_1.

Tests for p from $\mathcal{B}(n, p)$

Suppose we want to investigate the claim (effect) that $p > p_0$,*one-sided tests*, where p is the success probability in binomial trials. Suppose n independent trials are conducted.

Set up the hypotheses $H_0 : p = p_0, \quad H_1 : p > p_0.$

Use the variable X (denoting the number of successes) as the appropriate test statistic τ,noting that if H_0 is true, $X \sim \mathcal{B}(n, p_0)$and *large* observed values of Xfavours H_1.From the data, observe the value of X (x, say). The probability of observing a value as unusual as the observed value (when H_0 is true) is the P-value, $\mathrm{P}(X \geq x)$.

(The coin tossing example was of this form.)

Alternatively, Suppose we want to investigate the claim (effect) that $p<p_0$,
$H_0 : p = p_0, \quad H_1 : p<p_0.$ P-value$=\mathrm{P}(X \leq x)$.

Example. In 1974 a large scale study on drug abuse showed that approximately 40% of young adults smoked tobacco. It is anticipated that this proportion has since dropped. What evidence is there of a drop in the rate
(a) If a recent pilot study of 10 randomly selected young adults showed that 3 smoked tobacco.
Solution Let p be the proportion of young adults who smoke tobacco now.

We want to assess the evidence of a drop in the proportion of young adults who smoke tobacco, we set $H_0 : p = 0.4, \quad H_1 : p < 0.4$. Let X be the number of smokers in the sample. If H_0 is true $X \sim B(10, 0.4)$. The P-value is

$$P = P(X \leq 3) = 0.38 \geq 0.05$$

Providing little evidence of a drop, the data are consistent with $p = 0.4$.

(b) If 30 out of a sample of 100 smoked. $X \sim B(100, 0.4)$. The P-value is

$$P = P(X \leq 30) = 0.025(3dp) < 0.05 \quad \text{significant!}$$

Two-sided tests for a proportion

Often, the only sensible alternative to $H_0 : p = p_0$is $H_1 : p \neq p_0$. This is called a two-sided test. We will discuss *only* the case $p_0 = 0.5$.

- Set up the hypotheses $H_0 : p = 0.5$, and $H_1 : p \neq 0.5$

- Use the variable X (denoting the number of successes) as the appropriate test statistic,noting that $X \sim \mathcal{B}(n, p)$.

- It is now large *and* small observed values of X which favours H_1.

Two-sided tests for a proportion

• Assuming that H_0 is true$X \sim \mathcal{B}(n, 0.5)$.The mean value of X is $E(X) = np = \frac{n}{2}$.Large *and* small observed values of X correspond to large observed values of $|X - \frac{n}{2}|$. If the data show the observed value of X to be x then the two-sided P-value is

$$\mathrm{P}\left(|X - \tfrac{n}{2}| \geq |x - \tfrac{n}{2}|\right).$$

By symmetry of the binomial distribution with respect to $\frac{n}{2}$ when $p = 0.5$the P-value is computed as follows

$P = 2P(X \geq x)$, if $x > \frac{n}{2}$, and

$P = 2\mathrm{P}(X \leq x)$, if $x < \frac{n}{2}$.

Example.A medical team is ready to perform clinical trials on a new headache preparation, 'Migro', which may or may not be superior to the standard medication.60 patients receive both the standard medicationand Migro, recording simply their preference for one or other medication.Perform a test of significance, if the trials result in 35 preferences for Migro and 25 for the standard medication.

Solution Let p be the proportion of patients who prefer Migro.We wish to test $H_0 : p = 0.5$, and $H_1 : p \neq 0.5$

Assuming H_0 is true, $X \sim \mathcal{B}(60, 0.5)$. The observer value is $x = 35$.The mean value of X is $\frac{60}{2} = 30$, since $x = 35 > 30$ the P-value is $P = 2 \times P(X \geq 35)$, using the normal approximation with continuity correction:$P = 2 \times P(\frac{X-30}{\sqrt{15}} \geq \frac{35-0.5-30}{\sqrt{15}}) =$ $2 \times P(Z \geq 1.16) \approx 0.24$
no evidence that Migro is superior.

The sign test

The sign test is an application of the binomial test and is used in two different situations: the single sample sign test and the paired sample sign test

The single sample sign test. Suppose a sample $(x_1, \cdots, x_n)$ is taken from a *continuous* distribution with mean μ,in order to test the null hypothesis, $H_0 : \mu = \mu_0$ against one sided or two sided alternatives.

If we assume that the pdf is symmetric,then μ is also the median,so each observation is equally likely to be above or below μ.It is therefore useful to look at the signs of each of the differences $(x_i - \mu_0),\ i = 1, \ldots, n$.

We now define a variable X which counts the number of '$+$' signs and note that X is $\mathcal{B}(n, p_+)$, where p_+ is the probability of a positive difference.So the null hypothesis can be rewritten as$H_0 : p_+ = 0.5$.If there are a few observations $equal$ to μ_0, we simply $ignore$ them,and work with the reduced sample.

Example. The following data are 15 measurements of moisture retention (%) using a new sealing system. This system is expected to be better (i.e. to have greater retention) than the system previously in use, for which the mean retention was 96%.Use the sign test to analyse the data.

97.5	95.2	97.3	96.0	96.8	99.8	97.4	95.3
98.2	99.1	96.1	97.6	98.2	98.5	99.4	

Solution $sign(d_i) = sign(x_i - 96) =$
$(+, -, +, 0, +, +, +, -, -, +, +, +, ..., +)$

Let p_+ be the proportion of $+$ signs in the sample. We wish to test
$H_0 : p_+ = 0.5$ and $H_1 : p_+ > 0.5$

In total we have $3'-'$, $12'+'$ and $1Zero'$. The reduce sample has size $15 - 1 = 14$. Under $H_0 : p_+ = 0.5$, $X \sim B(14, 0.5)$, so the P-value is

$$P = P(X \geq 12) = 1 - P(X \leq 11) \approx 0.0065 < 0.05$$

significant evidence that $p_+ > 0.5$.

One Sample or Paired Sample?

One-sample: Suppose you have a sample of size n drawn from a population. You want to make (or check) some statement about the population. A typical situation was given in the previous example.

Paired-sample:Pairs of *similar* individuals or things are selected. In an experiment, one treatment is applied to one member of each pair, the other treatment to the second member.The members of a pair may be two patients of the same age and sex who have just undergone the same operation.The aim of pairing is to make the comparison more accurate by having the members of any pair as alike as possible except in the treatment difference that the investigator deliberately introduces.Sometimes, a single individual is measured on two occasions: this type of pairing is called *: self-pairing*.

The paired sample sign test

Suppose that we have a set of paired data (e.g. patients before and after a diet for obesity) and wish to test the hypothesis that the distribution of differences has mean $\mu = 0$ (e.g. the obesity diet is not effective). We can perform a sign test by first taking differences within the pairs, and then proceeding as above.

Example. A new weight-reducing plan has been developed and the before-and-after weights of 18 people who have used this plan are as follows.

Before	66	80	68	86	100	71	62	66	58
After	65	81	67	85	96	73	61	63	60
Diff	1	-1	1	1	4	-2	1	3	-2

Before	85	78	63	68	56	95	67	64	75
After	85	78	61	69	57	95	67	64	72
Diff	0	yel 0	2	-1	-1	0	0	0	3

Solution. Let p_+ be the proportion of $+$ signs in the sample. We wish to test $H_0 : p_+ = 0.5$ and $H_1 : p_+ > 0.5$ (effective diet)

In total we have $5'-'$, $8'+'$ and $5'Zeros'$. The reduce sample has size $18 - 5 = 13$. Under $H_0 : p_+ = 0.5$, $X \sim B(13, 0.5)$, so the P-value is

$$P = P(X \geq 8) = 1 - P(X \leq 7) \approx 0.29 > 0.05$$

data supports H_0: $p_+ = 0.5$ (the diest is not effective)