BUILDING SCIENCE
IN AFRICA

AIMS VIDEO COURSES
**SUPPORTING BOOKLET**

# PROBABILITY & STATISTICS

WITH
**PROF DAVID SPIEGELHALTER**

AIMS
SOUTH AFRICA

African Institute for Mathematical Sciences
6 MELROSE ROAD | MUIZENBERG | CAPE TOWN 7945 | SOUTH AFRICA
TEL: +27 (0)21 787 9320 | FAX: +27 (0)21 787 9321
EMAIL: info@aims.ac.za | WEB: www.aims.ac.za
AIMS SOUTH AFRICA

## AIMS Online Courses

The mission of the AIMS academic programme is to provide an excellent, advanced education in the mathematical sciences to talented African students in order to develop independent thinkers, researchers and problem solvers who will contribute to Africa's scientific development.

Teaching at AIMS is based on the principle of learning and understanding, rather than simply listening and writing, during classes, and on creating an atmosphere of increasing our knowledge through class discussions, through small group discussions, by formulating conjectures and assessing the evidence for them, and sometimes going down wrong paths and learning from the mistakes that led us there. The essential features of the classes at AIMS are that, in contrast to formal lecture courses, they are highly interactive, where the students engage with the lecturer throughout the class time, are encouraged to learn together in a journey of questioning and discovery, and where lecturers respond to the needs of the class rather than to a pre-determined syllabus. AIMS teaching philosophy is to promote critical and creative thinking, to experience the excitement of learning from true understanding, and to avoid rote learning directed only towards assessment.

Leading international and local experts offer the courses at AIMS, which are three weeks long (each module consisting of 30 hrs) and collectively form the coursework for a structured masters degree which also includes a research component. The advertised content is a guide, and the lecturers are encouraged, and indeed expected, to adapt daily to meet the current needs of the students.

Over the past ten years AIMS has achieved international recognition for this innovative and flexible approach. It has been the starting point for the remarkable success of our students and alumni and we all benefit from the support of many who have "witnessed the AIMS-magic and keep coming back for more."

This year we have decided to film selected courses and to make them available to a larger audience as an online facility. African universities may choose to use these courses to supplement and enhance their own postgraduate programmes. We believe this would be best achieved through engagement with AIMS. One way for this to happen, would be for AIMS to suggest or nominate a specialist tutor to spend time at the university, guiding students who follow the online programme. Where possible expert lecturers who have taught at AIMS may visit the university to give a short introduction to the course. We would welcome this interaction as well as the contribution our online courses will make to the growth of the mathematical sciences ecosystem in Africa.

Barry Green
Director & Professor of Mathematics
African Institute for Mathematical Sciences
January 2013

PROBABILITY & STATISTICS
2012

PROF DAVID SPIEGELHALTER

**DAY 9**

AIMS
SOUTH AFRICA

# Hypothesis Testing

Basic idea:

- Set up a *null hypothesis* denoted $H_0 : \theta = \theta_0$ for some parameter
- Suppose we want to compare $H_0$ with the alternative hypothesis that $\theta > \theta_0$.
- We want to see whether our observed data $\underline{x}$ provide strong evidence against $H_0$
- Calculate a 'test statistic' $T(\underline{x})$, often an observed estimate $\hat{\theta}(\underline{x})$, where high values of $T(\underline{x})$ indicate evidence against $H_0$
- Calculate the probability of observing such an extreme result, if $H_0$ were really true; i.e. $P(T(\underline{X}) \geq T(\underline{x})|H_0)$. This is known as the 'one-sided P-value'
- Declare the result 'statistically significant at the 5% level' if $P < 0.05$.

# Hypothesis Testing

## IMPORTANT NOTES

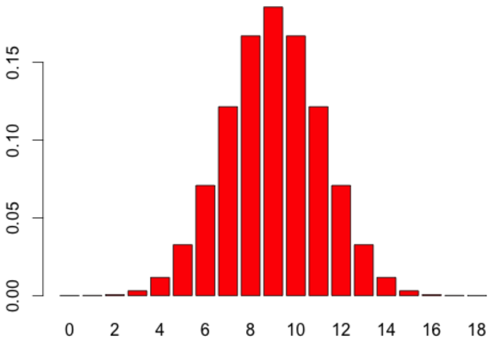
- We may be also want to reject $H_0$ for *low* values of $T(\underline{x})$ - this is known as a *two-sided* test of $H_0 : \theta = \theta_0$ against the alternative that $\theta \neq \theta_0$
- The usual technique is to calculate the one-sided P-value, and double it to turn it into a 2-sided test
- If a 95% interval does not contain 0, this is essentially equivalent to saying $P < 0.05$ in a 2-sided test
- $P < 0.05$ is the traditional test, often indicated by '*', with '**' indicating $P < 0.01$ and '***' indicating $P < 0.001$
- If the result is not statistically significant, it does NOT mean the null hypothesis is true. It just means there is insufficient evidence to be confident it is false
- If we do lots of tests, bound to get a significant result!

# Example

$x = 14$ out of $n = 18$ (78%) females could roll their tongues

Is this evidence against $\theta_0 = 0.5$, i.e. that it is just 50:50?

Under $H_0$, $X \sim \text{Binomial}(18, 0.5)$. What is the probability that $X \geq 14$?



# Example

```
sum(dbinom(14:18,  18,  0.5))
```

P=0.015

We can double this to give 2-sided P-value = 0.03 , i.e. significant at 5% level

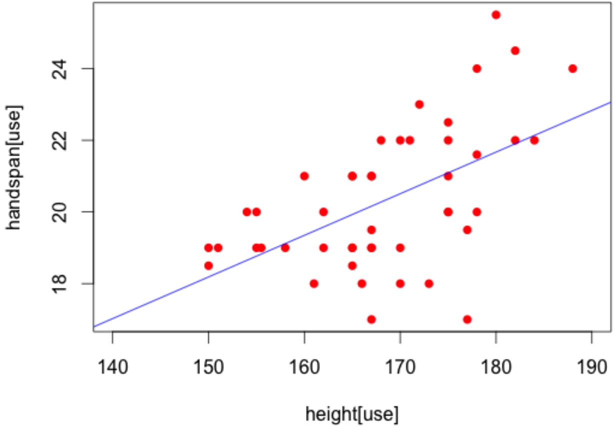
# Example

Can use 'prop.test' to get confidence intervals and test for a single proportion

```
> prop.test(14, 18, p=0.5)

1-sample proportions test with continuity correction

data:  14 out of 18, null probability 0.5
X-squared = 4.5, df = 1, p-value = 0.03389
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.5191861 0.9262769
sample estimates:
        p
0.7777778
```

95% interval excludes 0.5, so $P < 0.05$ in a 2-sided test

# Significance of a gradient



# Significance of a gradient

```
Coefficients:
             Estimate   Std. Error    t value   Pr(>|t|)
(Intercept)  0.76245    4.49710     0.170      0.866
height[use]  0.11616    0.02669      4.352    7.9e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

The observed gradient is 0.116 (expected change in handspan per cm extra height)

There is 7.9e-05 (0.00008) chance of observing such an extreme gradient (either $>$ 0.116, or $<$ -0.116) if there is really no relationship between height and handspan

Therefore we can be 99.99% confident there is a relationship.

[Note the intercept in R is the expected value for 'handspan' for someone with height 0!]

# Significance of a gradient

Remember: if we are given an estimate $\hat{\theta}$ with a standard error $\sqrt{V}$, then

- An approximate 95% confidence interval for $\theta$ is $\hat{\theta} \pm 1.96\sqrt{V}$
- A standardised test statistic is $z = \frac{(\hat{\theta} - \theta_0)}{\sqrt{V}}$ (may be called a 't value')
- Under $H_0$, $z$ has arisen from a standard Normal(0,1) distribution
- An approximate 2-sided P-value for testing $H_0: \ \theta = \theta_0$ is $2 \times \min(\Phi(z), \Phi(-z))$

```
# Suppose I have an estimate t=3 with standard error 1.7.
# I wish to test whether it comes from a distribution with mean 0
t=3
s=1.7
z=(t-0)/s
Phi=pnorm(z)    # norm is Phi cumulative distribution function of s
#     Phi(-z)   = 1-Phi
P=2 * min(Phi,1-Phi) # converts to 2-sided P-value whether t is hig
P=0.078
```

# Significance of a gradient

Can also see if confidence interval includes 0

```
> confint(fitted)
                 2.5 %   97.5 %
(Intercept) -8.30085273 9.825761
height[use]  0.06237379 0.169948

> confint(fitted,level=0.9999)
                0.005 %   99.995 %
(Intercept) -18.47741692 20.0023255
height[use]   0.00197997  0.2303418
```

99.99% interval just excludes 0

# Memory test (not to be marked!)

Take a clean sheet of paper

Try to remember the numbers I read out

Write them down *when I tell you to*

No cheating!

Hand in your paper (anonymous)

# Significance of a difference in proportions

In the class, we observed $X = 14$ out of $n = 18$ (78%) females could roll their tongues, compared to $Y = 22$ out of $m = 31$ (71%) males.

Can arrange as 2-way 'contingency table'

| | Can roll tongue | Can't roll | |
|---|---|---|---|
| female | 14 | 4 | 18 |
| male | 22 | 9 | 31 |
| | 36 | 13 | 49 |

# Chi-squared statistic

Calculate 'Expected' table under null hypothesis of 'no association', i.e. equal proportions.

Overall proportion $= 36/49 = 0.735$

Applying this proportion to both the men and women, we would 'expect' to see $18 \times 36/49 = 13.2$ females who could roll their tongue. Complete 'expected' table

| | Can roll tongue | Can't roll | |
|---|---|---|---|
| female | $E_1$=13.2 | 4.8 | 18 |
| male | 22.8 | 8.2 | 31 |
| | 36 | 13 | 49 |

Compare the Observed and Expected counts in all 4 cells, $i = 1, 2, 3, 4$, using the 'Chi-squared statistic'

$$X = \sum_i \frac{(O_i - E_i)^2}{E_i} =$$

$(14-13.2)^2/13.2+(4-4.8)^2/4.8+(22-22.8)^2/22.8+(8.2-8)^2/8.2 = 0.21$

# Likelihood ratio test

Let $\hat{\theta}_0$ is the MLE of the parameters under $H_0$, of length $p_0$, which maximises $f(\underline{x}|\theta_0)$

Let $\hat{\theta}_1$ is the MLE of the parameters under $H_1$, of length $p_1$, which maximises $f(\underline{x}|\theta_1)$

Then

$$X = -2 \log \left[ \frac{f(\underline{x}|\hat{\theta}_0)}{f(\underline{x}|\hat{\theta}_1)} \right]$$

is the *likelihood ratio statistic*.

If $H_0$ is true, then for large samples $X$ has a $\chi^2_{p_1 - p_0}$ distribution

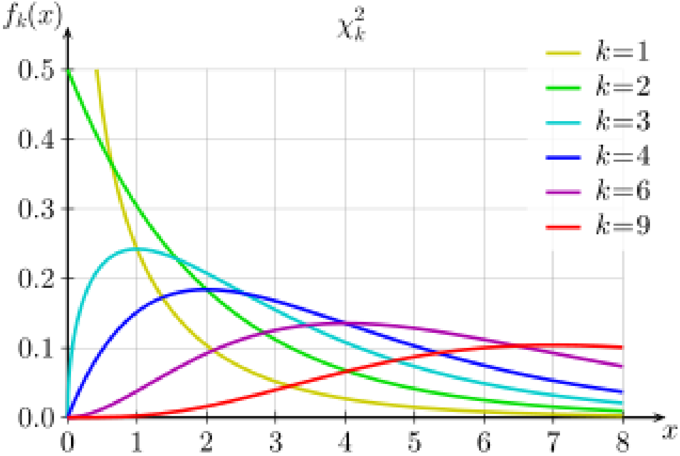
The likelihood ratio is a 2-sided test.

# Chi-squared distribution

If $Z_1, Z_2, .., Z_k$ are independent Normal$(0, 1)$ variables, then $X = \sum_i Z_i^2$ is said to have a $\chi_k^2$ distribution on $k$ degrees of freedom

- Denoted: Chi-square$(k)$
- Density: $f(x|k) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{k/2}\Gamma\left(\frac{k}{2}\right)},\ x > 0$
- Expectation (mean): $\mathbb{E}[X|k] = k$
- Variance: $\mathbb{V}[X|k] = 2k$
- Standard deviation $= \sqrt{2k}$

# Chi-squared distribution



# Likelihood ratio test for comparing two proportions

The 'Chi-squared statistic'

$$X = \sum_i \frac{(O_i - E_i)^2}{E_i} =$$

is an approximation to the likelihood ratio test for comparing two proportions

$p_1 = 2$ (two proportions to estimate)

$p_0 = 1$ (one common proportion)

So $p_1 - p_0 = 1$, and the test statistic should have a $\chi^2_1$ distribution if $H_0$ is true.

# Chi-squared distribution

Therefore the Chi-squared statistic for a $2 \times 2$ table has an approximate $\chi_1^2$ distribution under the null hypothesis.

$X = 0.21$ for tongue-rolling and gender

Therefore we should calculate $P(\chi_1^2 > 0.21) = 0.64$

```
> pchisq(0.21,1, lower.tail=FALSE)
[1] 0.6467674
```

So no evidence that tongue-rolling is associated with gender - the observed difference in proportions could just have been due to chance.

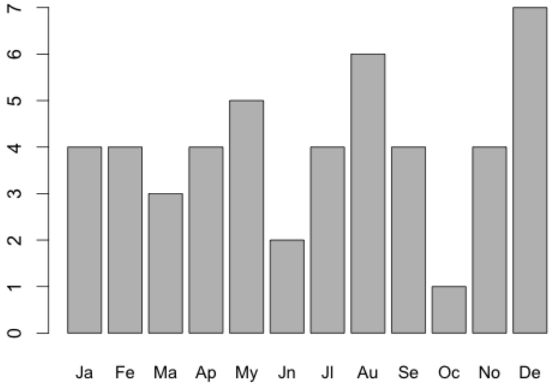
# Test of a difference in proportions in R

```
> prop.test(c(14,22),c(18,31))

2-sample test for equality of proportions with continuity correctio

data:  c(14, 22) out of c(18, 31)
X-squared = 0.0342, df = 1, p-value = 0.8533
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.2256423  0.3618430
sample estimates:
   prop 1    prop 2
0.7777778 0.7096774

Warning message:
In prop.test(c(14, 22), c(18, 31)) :
  Chi-squared approximation may be incorrect
```

Slightly different P-value due to slightly different method of calculating Chi-squared statistic

# 'Goodness-of-fit tests'

```
month=factor(birthmonth,levels=c("January","February","March","Apri
barplot(table(month),cex.names=0.9)
```



# 'Goodness-of-fit tests'

Is there really a difference between the months?

$H_0 : p = 1/12$ for each month, so expect $E_i = 50/12 = 4.17$ in each month

Observed counts $= \underline{O} = 4, 4, 3, 4, 5, 2, 4, 6, 4, 1, 4, 7$

Chi-square $X = \sum \frac{(O_i - E_i)^2}{E_i} = 6.8$

$p_1 = 11$ (12 probabilities, but must sum to 1)

$p_0 = 0$ (all the probabilities are $1/12$)

So $p_1 - p_0 = 11$, and the test statistic should have a $\chi^2_{11}$ distribution if $H_0$ is true.

$P(\chi^2_{11} > 6.8) = 0.82$

So NO evidence that there is a real difference between the months - the variability is just due to chance

```
 > chisq.test(t)
Chi-squared test for given probabilities
X-squared = 7, df = 11, p-value = 0.7991
```

# Stats practical on hypothesis testing

You are recommended to type commands into the script window (top left) and then run them by selecting the line (you just need then cursor on the line, no need to highlight it)

Do these in any order!

1. Download `class-data4.R` into RStudio
2. Download `class-data.csv` and read it in
3. Run the code testing whether 14/18 is significant evidence against $H_0 : \theta = 0.5$.
4. How would the results have differed if only 13 out of 18 women could roll their tongues?
5. (a) Use the code provided to fit handspan against height (excluding the outlier datapoint with height = 140)
   (b) In the `summary(fitted)`, check that the 't value' is the ratio of the estimate to its standard error (this makes it a test statistic of the null hypothesis that the parameter is 0)
   (c) For the gradient `height[use]`, what is the 2-sided P-value associated with a 't value' of 4.352? [Check slide 8] (This may not exactly match that given in R)
6. Test whether gender is associated with the arm that is on top when crossing arms (arm cross)? First create a 2 x 2 table [table(arm cross,gender)], and then carry out a chi-squared test [you can use prop.test]
7. When testing between months, redo the calculation assuming there were 10 times as many people born in each month (i.e. each of the entries in `t` multiplied by 10). What difference does it make?
8. For the memory test data,
   (a) look at the proportions getting the answer right for 5,7,9,11 digits.
   (b) Does there appear to be a trend?
   (c) Is there evidence for a true difference between 7 and 9 numbers?
   (d) Can you plot the estimates and 95% confidence intervals against the number of digits 5,7,9,11?

# Assignment 2

Please hand in solutions to these questions in Latex

1. The heights of men in a country have a Normal distribution with mean 167 cm and standard deviation 10.

   (a) What is the probability of a random man being more than 180 cm tall?

   (b) If I take a group of 4 men at random, and take the mean $X$ of their heights, what is the distribution of $X$?

   (c) What is the probability that $X$ will be more than 180 cm?

2. In a survey of 200 people picked at random, 114 said they were afraid to walk out at night

   (a) What is the estimated proportion of the population that are afraid to walk out at night?

   (b) Calculate a 95% confidence interval for this proportion.

   The following year, another group of 200 random people were asked the same question, and 100 said they were afraid. The police force said this showed an improvement.

   (a) What is the estimated change in the population that are afraid to walk out at night?

   (b) Calculate directly (not using R) a 99% confidence interval for this change using the formula for the asymptotic Normal approximation

   (c) Calculate the result using R

   (d) Would you say this was good evidence that the streets had got safer?

3. The number of goals scored by a football team can be assumed to be a Poisson random variable with mean $\theta$. In 5 matches, they score (3,1,5,2,3) goals.

   (a) What is the maximum likelihood estimate of $\theta$?

   (b) In terms of $\theta$, what is the probability of getting no goals in a match?

   (c) What is the maximum likelihood estimate of this probability?

optional Consider a sample $X_1, ., X_n$ from an exponential distribution with mean $1/\lambda$ [see probability Lecture 4].

   (a) Find an expression for the maximum likelihood estimate (MLE) of $\lambda$

   (b) Find the Fisher Information

   (c) Find the asymptotic variance of the MLE.

Also, complete the exercises in the Stats Practical 2 (not to be marked)