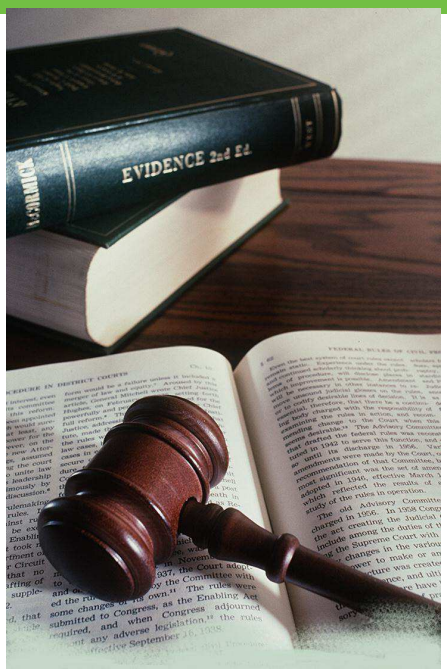
Hypothesis Testing



How good is our evidence?

What is the risk we are wrong?

UNIVERSITY of STIRLING
Management School

Topics

Curwin and Slater Ch 12 & 13

- confidence limits
- the sample mean
- hypothesis testing

What can the sample tell us about the population?

UNIVERSITY of STIRLING Management School

- Confidence limits
- Hypothesis testing - the statistical method of proof

Hypothesis testing

UNIVERSITY of STIRLING Management School

H_0 Null hypothesis

- the opposite of what we are trying to prove
- start by assuming this is true

H_1 Alternative hypothesis

- what we are trying to prove
- initially assume false

Legal Analogy

UNIVERSITY of STIRLING
Management School

The suspect is assumed innocent:

H_0 : the suspect is innocent

A Court tries to prove that the suspect committed the crime:

H_1 : the suspect is guilty

Decision Errors

UNIVERSITY of STIRLING
Management School

	H_0 true	**H_0 false**
Accept H_0	Correct decision	Type 2 error
Reject H_0	Type 1 error	Correct decision

Legal Analogy

	Innocent	Guilty
Free	Correct decision	Type 2 error
Convict	Type 1 error	Correct decision

Application in medicine

UNIVERSITY of STIRLING
Management School

	Free of cancer	Cancer
Test negative	✪	Type 2 error False -ve
Test positive	Type 1 error False +ve	Correct decision

Type 1 errors - Significance level

UNIVERSITY of STIRLING
Management School

Significance level is the probability of making a type 1 error:

- probability of rejecting H_0 when it is true
- probability of convicting an innocent person

This is usually required to be quite low e.g.. 0.05

- P<0.05 : prob. of type 1 error < 0.05
- implies requirement for convincing evidence to reject H_0 or convict.

Type 2 Errors

UNIVERSITY of STIRLING
Management School

- Prob. of accepting H_0 when it is false
- Prob. of rejecting H_1 when it is true
- Prob. of freeing a guilty person

- Also required to be low, but usually preferable to have type 2 error than a type 1 error
  - preferable to free a guilty person than to convict an innocent person
  - preferable to accept H_0 when it is in fact false rather than to reject H_0 when true.

Relationship between Type I & II errors

UNIVERSITY of STIRLING Management School

As prob. of type 1 error reduces, prob. of type 2 error increases:

- Reduce prob of rejecting H_0 , when H_0 true
- Reduce prob of convicting innocent person

- Stronger evidence is required to prove H_1
- Stronger evidence is required for conviction

- Insufficient evidence to reject H_0 when H_0 false
- Insufficient evidence to convict guilty people

Example: Car Sharing

UNIVERSITY of STIRLING Management School

- 2012 ave. no. cars in car park 220/day

- 2013 sample of 75 days
- ave. = 205 , s = 32

Question: Has the no. of cars **changed?**

- H_0: $\mu = 220$
- H_1: $\mu \neq 220$ (2 tailed)

UNIVERSITY of STIRLING
Management School

Using distribution of sample means, assuming a population mean (μ) of 220, sample mean (X) =205

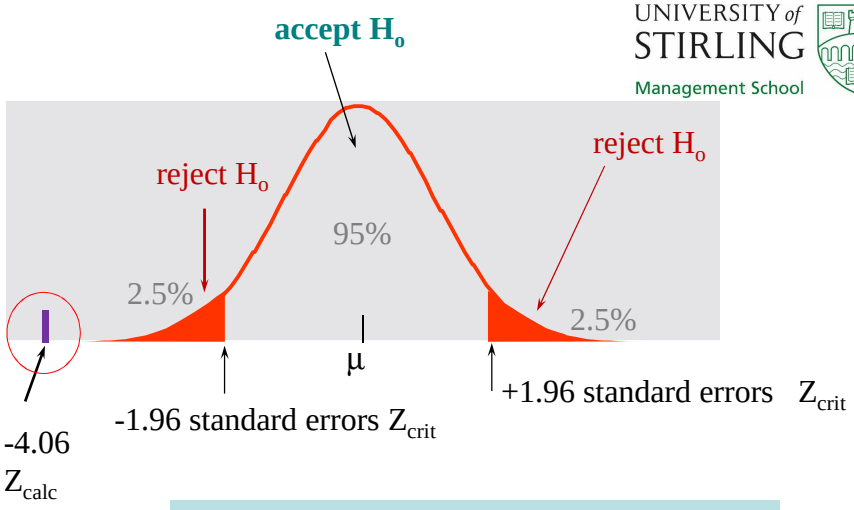
•Calculate distance of the 2013 sample (n=75) mean from the population mean

•Standardise this to obtain Z. (ie measure distance in standard error units (Z_{calc}))

$$Z_{calc} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$Z_{calc} = \frac{205-220}{\frac{32}{\sqrt{75}}} = -4.06$$

UNIVERSITY of STIRLING
Management School



$Z_{calc} < Z_{crit}$ reject H_o

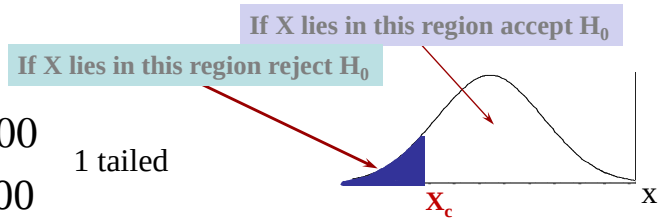
The average no. of cars has changed

1 tailed test

UNIVERSITY of STIRLING
Management School

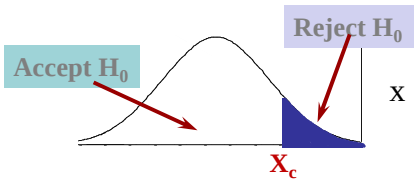
$H_0 : X \geq 200$
$H_1 : X < 200$

1 tailed



$H_0 : X \leq 200$
$H_1 : X > 200$

1 tailed



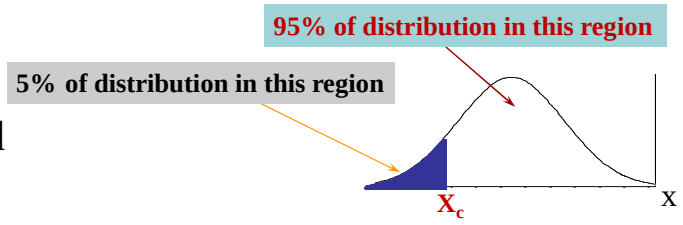
X_c Critical Value of X

5% Significance

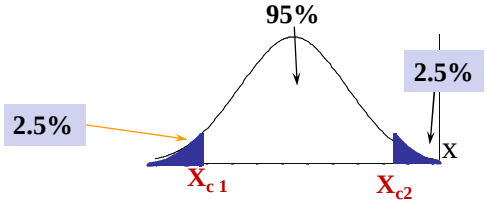
UNIVERSITY of STIRLING
Management School

Choose critical values of X:

1 tailed



2 tailed

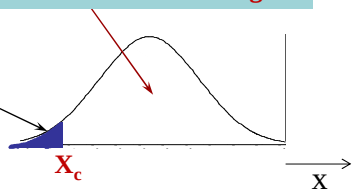


X_c is the Critical Value of X

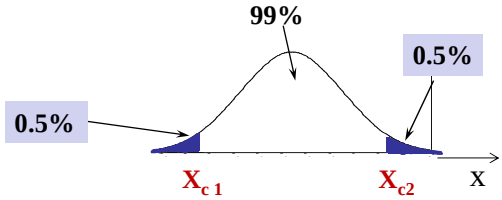
1% Significance

UNIVERSITY of STIRLING
Management School

1 tailed



2 tailed



X_c is the Critical Value of X

Critical values with the normal distribution

UNIVERSITY of STIRLING
Management School

	1 tailed	2 tailed
5% signif	1.64	1.96
1% signif	2.32	2.575

Example 2: Car Sharing

UNIVERSITY of STIRLING Management School

- 2012 ave. no. cars in car park 220/day
- 2013 sample of 75 days; ave. = 205 , s = 32

Question: Has the no. of cars reduced?

- H_0: μ => 220
- H_1: $\mu <$ 220 (1 tailed test)

UNIVERSITY of STIRLING Management School

•Using distribution of sample means, assuming a population mean (μ) of 220, sample mean (x) =205

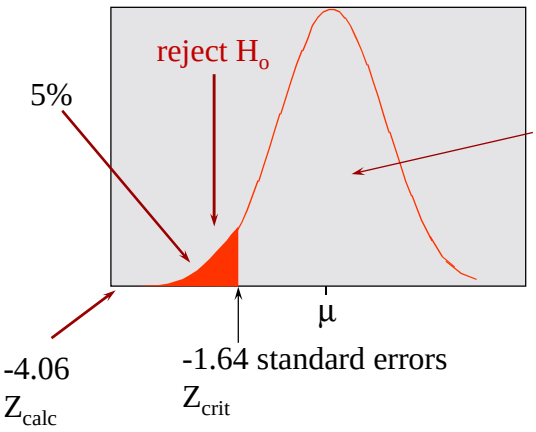
•Calculate distance of the 2013 sample mean from the population mean

•Standardise this to obtain Z (ie measure distance in standard error units (Z_{calc}))

$$Z_{calc} = \frac{X - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$Z_{calc} = \frac{205-220}{\frac{32}{\sqrt{75}}} = -4.06$$

UNIVERSITY of STIRLING
Management School



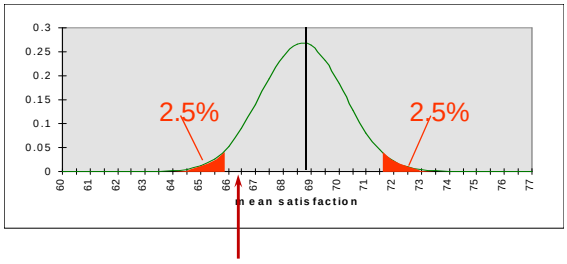
$$Z_{calc} < Z_{crit}$$

reject H_o

The average no. of cars has decreased

Example

Customer satisfaction has been measured and the long term average score is **68.7.**



Queries:

1. Is a single sample (n=50) with a mean of 66.4 significant evidence of a change in quality?

2. Does this sample mean lie within the 95% confidence limits?
   (ie. no evidence of any fundamental change).

3. Or does it lie outside suggesting a variation which is unlikely (5%) to be explained as chance?
   (ie. evidence of a shift in the true mean and a real change in customer satisfaction)

Testing a single sample

UNIVERSITY of STIRLING
Management School

- H_0: mean is still 68.7 (null hypothesis)
- H_1: mean has changed (alternative hypothesis)
- consider:

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

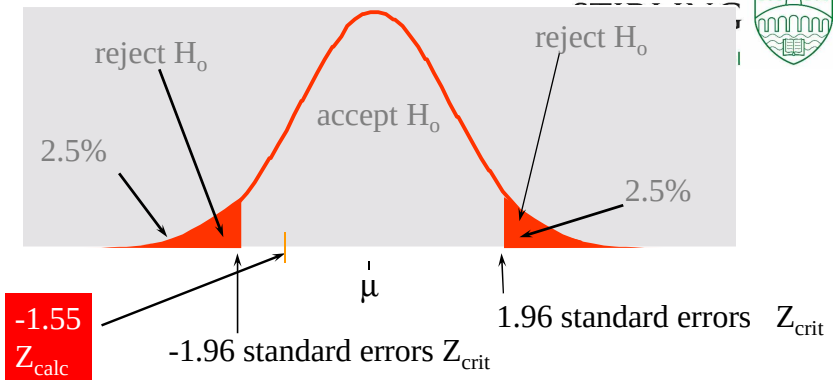
$$= \frac{66.4 - 68.7}{10.5 / \sqrt{50}}$$

$$= -1.55$$

|z| < 1.96 hence not significant at the 5% (=100-95%) level and accept the null hypothesis.

ie. population mean is unchanged and no convincing evidence of a deterioration in customer satisfaction

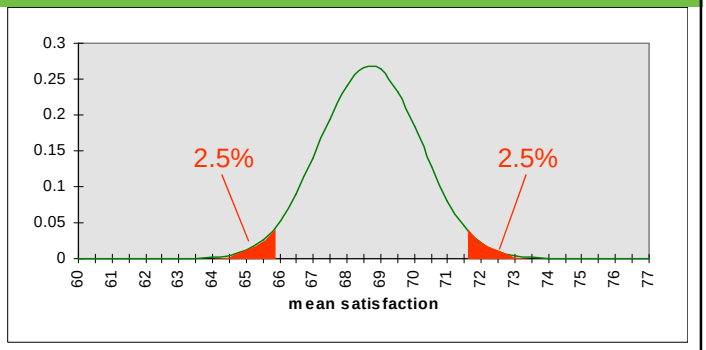
UNIVERSITY of



$|z_{calc}| < 1.96$ hence not significant at the 5% level, accept the null hypothesis.

ie. population mean is unchanged and no convincing evidence of a deterioration in customer satisfaction

or,



Calculate 95% CL = mean +/- 1.96 * standard error
= 68.7 +/- 1.96*10.5/√50
= 68.7 +/- 2.9 = 65.8 ---71.6

Sample mean= 66.4 is within 95% CL, therefore accept H_0,
ie. population mean is unchanged and no convincing evidence of a deterioration in customer satisfaction

Rules (C&S examples Ch13)

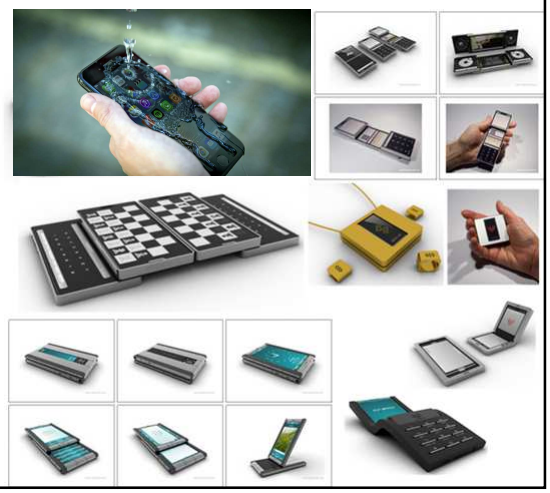
UNIVERSITY of STIRLING
Management School

• State Hypotheses:	H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$
• State Significance Level:	5%
• Determine Critical Values:	Zcrit = ± 1.96
• Calculate the test statistic:	Zcalc
• Compare:	Zcalc to Zcrit
• Conclude:	Accept or Reject H_0
• Statement of conclusion	Relationship is/is not significant at the 5% level, therefore accept/reject the null hypothesis.

UNIVERSITY of STIRLING
Management School

The Chi Squared Test

Non-Parametric Statistics



UNIVERSITY of STIRLING
Management School

Topics

- Curwin & Slater, Chapter 13 - Chi Square tests
- analysing contingency tables
- goodness-of-fit
- degrees of freedom (p.265)

UNIVERSITY of STIRLING
Management School

Non Parametric Statistics

- Is there any relationship between type of car driven and vegetarianism?
  - No continuous variable – only classes of response.
  - All data must be in raw frequencies – not percentages.
  - Observed frequencies cannot be too small

UNIVERSITY of STIRLING
Management School

Chi Squared test - What is it?

Typically, the hypothesis tested with chi square is whether or not two different samples –

(of people, texts, whatever)

are different enough in some characteristic or aspect of their behaviour that we can generalize from our samples that the populations from which our samples are drawn are also different in the behaviour or characteristic.

Chi Squared test – example

UNIVERSITY of STIRLING Management School

- Example: is there a significance between males and females in their main use of mobile phones?
- Categories:
  - Text messaging
  - Voice calls
  - Music/video playing/streaming
  - Other (data services, games, apps, picture/video messages, camera etc)

Raw frequencies

Random Samples

	Text	Voice	Music/Vid	Other
FEMALE	39	78	12	10
MALE	10	62	41	21

Characteristics of Chi Squared Test

UNIVERSITY of STIRLING Management School

- Categorical data only – no derived statistics (including percentages).
- Tabular data.
- No distribution assumed.
- More likely to be have Type II errors than the parametric tests (z,t).
  - A non-parametric test, like chi square, is a rough estimate of confidence; it accepts weaker, less accurate data
- Test needs adjusting if more than 20% of expected frequencies are <5

Complaining Patients at the Orthopaedic Centre

UNIVERSITY of STIRLING Management School

	complaints	no complaints	total
Brunel	31	429	460
Hyde	16	284	300
Picasso	13	227	240
total	60	940	1000

- Number of complaints about each of our three consultants
- Brunel seems to be the worst
- Is it just chance that are some consultants appear worse than others?
- Or, is it worth investigating the different consultants' in detail?

What would we expect if all consultants were equal?

UNIVERSITY of STIRLING Management School

observed

	complaints	no complaints	total
Brunel	31	429	460
Hyde	16	284	300
Picasso	13	227	240
total	60	940	1000

60/1000 * 460

940/1000 * 460

expected

	complaints	no complaints	total
Brunel	27.6	432.4	460
Hyde	18.0	282.0	300
Picasso	14.4	225.6	240
total	60.0	940.0	1000

Compare the observed and expected

UNIVERSITY of STIRLING Management School

observed-expected = 31-27.6

	complaints	no complaints	
Brunel	3.4	-3.4	
Hyde	-2.0	2.0	
Picasso	-1.4	1.4	

observed-expected = 13-14.4

observed-expected = 16-18.0

Square & weight

UNIVERSITY of STIRLING Management School

$(obs - exp)^2/ exp = 3.4^2/27.6$

	complaints	no complaints	
Brunel	0.419	0.027	
Hyde	0.222	0.014	
Picasso	0.136	0.009	

$(obs - exp)^2/ exp = 2.0^2/18.0$

$(obs - exp)^2/ exp = 1.4^2/14.4$

Summarising

UNIVERSITY of STIRLING

observed	expected	O-E	$(O-E)^2$	$\frac{(O-E)^2}{E}$
31	27.6	+3.4	11.560	0.419
16	18.0	-2.0	4.000	0.222
13	14.4	-1.4	1.960	0.136
429	432.4	-3.4	11.560	0.027
284	282.0	+2.0	4.000	0.014
227	225.6	+1.4	1.960	0.009
		0		0.827

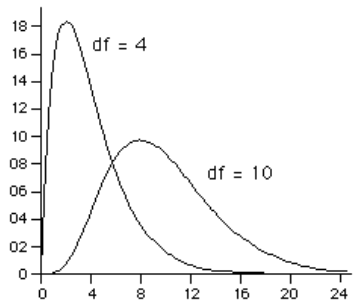
chi square

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(obs_i - \exp_i\right)^2}{\exp_i}$$
$$= 0.827$$

Is 0.827 good or bad?

UNIVERSITY of STIRLING
Management School

- if perfect agreement, χ^2=0.000
- if very poor agreement, χ^2 is large
- should reflect the number of predictions
- “degrees of freedom”



How many predictions were made?

UNIVERSITY of STIRLING Management School

observed

	complaints	no complaints	total
Brunel	31	429	460
Hyde	16	284	300
Picasso	13	227	240
total	60	940	1000

expected

	complaints	no complaints	total
Brunel	27.6	432.4	460
Hyde	18.0	282.0	300
Picasso	14.4	225.6	240
total	60.0	940.0	1000

χ^2 test

UNIVERSITY of STIRLING Management School

- H_0: no difference between consultants
- H_1: difference due to consultants

  χ^2 (calculated) =0.827
- degrees of freedom = 2
- signifigance level = 5%

  χ^2 (critical) = 5.99 (from tables)

The chi square table

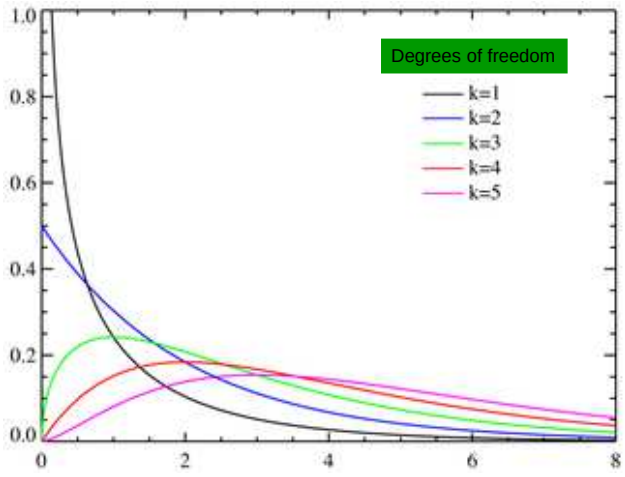
UNIVERSITY of STIRLING
Management School

Critical values of chi square

Pr	0.25	0.1	0.05	0.01	0.005	0.001
d.f.						
1	1.32	2.71	3.84	6.63	7.88	10.8
2	2.77	4.61	5.99	9.21	10.6	13.8
3	4.11	6.25	7.81	11.3	12.8	16.3
4	5.39	7.78	9.49	13.3	14.9	18.5
5	6.63	9.24	11.1	15.1	16.7	20.5
6	7.84	10.6	12.6	16.8	18.5	22.5
7	9.04	12.0	14.1	18.5	20.3	24.3
8	10.2	13.4	15.5	20.1	22.0	26.1
9	11.4	14.7	16.9	21.7	23.6	27.9
10	12.5	16.0	18.3	23.2	25.2	29.6
11	13.7	17.3	19.7	24.7	26.8	31.3
12	14.8	18.5	21.0	26.2	28.3	32.9
13	16.0	19.8	22.4	27.7	29.8	34.5
14	17.1	21.1	23.7	29.1	31.3	36.1
15	18.2	22.3	25.0	30.6	32.8	37.7
16	19.4	23.5	26.3	32.0	34.3	39.3
17	20.5	24.8	27.6	33.4	35.7	40.8
18	21.6	26.0	28.9	34.8	37.2	42.3
19	22.7	27.2	30.1	36.2	38.6	43.8
20	23.8	28.4	31.4	37.6	40.0	45.3
21	24.9	29.6	32.7	38.9	41.4	46.8
22	26.0	30.8	33.9	40.3	42.8	48.3
23	27.1	32.0	35.2	41.6	44.2	49.7
24	28.2	33.2	36.4	43.0	45.6	51.2
25	29.3	34.4	37.7	44.3	46.9	52.6
26	30.4	35.6	38.9	45.6	48.3	54.1
27	31.5	36.7	40.1	47.0	49.6	55.5
28	32.6	37.9	41.3	48.3	51.0	56.9
29	33.7	39.1	42.6	49.6	52.3	58.3
30	34.8	40.3	43.8	50.9	53.7	59.7
40	45.6	51.8	55.8	63.7	66.8	73.4
50	56.3	63.2	67.5	76.2	79.5	86.7
60	67.0	74.4	79.1	88.4	92.0	99.6
70	77.6	85.5	90.5	100	104	112
80	88.1	96.6	102	112	116	125
90	98.6	108	113	124	128	137
100	109	118	124	136	140	149

χ^2 Distribution Function

UNIVERSITY of STIRLING
Management School



χ^2 test

UNIVERSITY of STIRLING Management School

χ^2 (calculated) =0.827

χ^2 (critical) = 5.99 (from tables)

χ^2 (calculated) < χ^2 (critical)

- variation is not significant (at 5% level)
- i.e. no significant evidence that any one consultant is worse than any other

- This is only a single test - look for some other factor.

χ^2 test in general

UNIVERSITY of STIRLING Management School

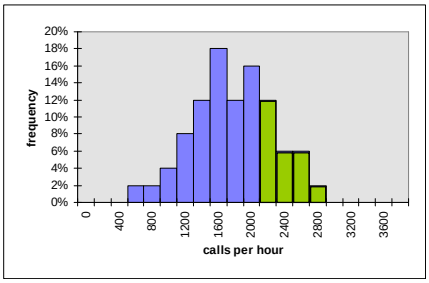
- H_0: no difference between categories
- H_1: difference due to categories

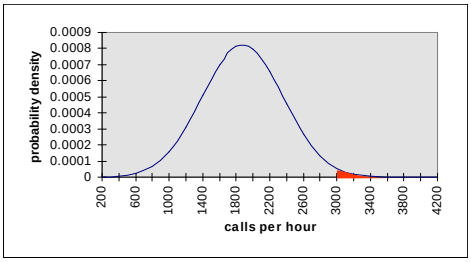
- degrees of freedom = (no. rows -1) * (no. columns-1)
- signifigance level = 5% ?

- if χ^2 (calculated) < χ^2 (critical), differences due to “chance”
- if χ^2 (calculated) > χ^2 (critical), the categories are important

Is this really a normal distribution?

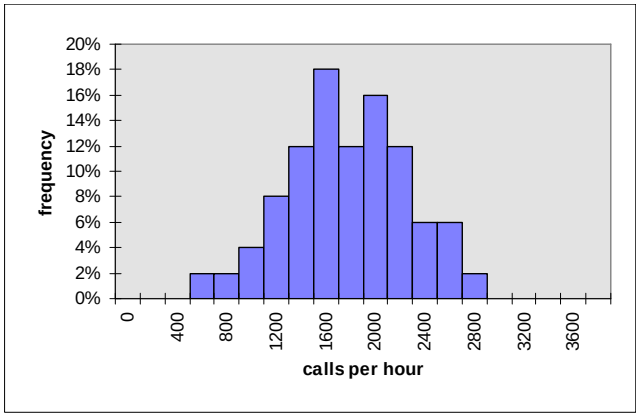
UNIVERSITY of STIRLING
Management School





Survey Results

UNIVERSITY of STIRLING
Management School



mean calls per hour = 1876
standard deviation = 484

Calculating the expected frequencies

UNIVERSITY of STIRLING

calls per hour		observed				expected
0	200	0	1.000	1.000	0.000	0.01
200	400	0	1.000	0.999	0.001	0.04
400	600	0	0.999	0.996	0.003	0.15
600	800	1	0.996	0.987	0.009	0.45
800	1000	1	0.987	0.965	0.022	1.11
1000	1200	2	0.965	0.919	0.046	2.31
1200	1400	4	0.919	0.837	0.081	4.07
1400	1600	6	0.837	0.716	0.122	6.08
1600	1800	9	0.716	0.562	0.153	7.66
1800	2000	6	0.562	0.399	0.163	8.17
2000	2200	8	0.399	0.252	0.147	7.36
2200	2400	6	0.252	0.140	0.112	5.61
2400	2600	3	0.140	0.067	0.072	3.61
2600	2800	3	0.067	0.028	0.039	1.96
2800	3000	1	0.028	0.010	0.018	0.90
3000	3200	0	0.010	0.003	0.007	0.35
3200	3400	0	0.003	0.001	0.002	0.12
3400	3600	0	0.001	0.000	0.001	0.03
3600	3800	0	0.000	0.000	0.000	0.01
3800	4000	0	0.000	0.000	0.000	0.00
4000	4200	0	0.000	0.000	0.000	0.00

expected frequency = 0.147 * 50
= 7.36

P(2000<calls<2200) = 0.399-0.252
= 0.147

P(calls>2000) =P(z>$\frac{2000-1876}{484}$)
=P(z>0.256)
=0.399

P(calls>2200) =P(z>$\frac{2200-1876}{484}$)
=P(z>0.669)
=0.252

Grouping the expected frequencies

UNIVERSITY of STIRLING

calls per hour		observed				expected	School
0	200	0	1.000	1.000	0.000	0.01	
200	400	0	1.000	0.999	0.001	0.04	
400	600	0	0.999	0.996	0.003	0.15	
600	800	1	0.996	0.987	0.009	0.45	
800	1000	1	0.987	0.965	0.022	1.11	
1000	1200	2	0.965	0.919	0.046	2.31	
1200	1400	4	0.919	0.837	0.081	4.07	
1400	1600	6	0.837	0.716	0.122	6.08	
1600	1800	9	0.716	0.562	0.153	7.66	
1800	2000	6	0.562	0.399	0.163	8.17	
2000	2200	8	0.399	0.252	0.147	7.36	
2200	2400	6	0.252	0.140	0.112	5.61	
2400	2600	3	0.140	0.067	0.072	3.61	
2600	2800	3	0.067	0.028	0.039	1.96	
2800	3000	1	0.028	0.010	0.018	0.90	
3000	3200	0	0.010	0.003	0.007	0.35	
3200	3400	0	0.003	0.001	0.002	0.12	
3400	3600	0	0.001	0.000	0.001	0.03	
3600	3800	0	0.000	0.000	0.000	0.01	
3800	4000	0	0.000	0.000	0.000	0.00	
4000	4200	0	0.000	0.000	0.000	0.00	

8.14

6.98

need to group so that (nearly) all expected frequencies >5

Summarising

UNIVERSITY of STIRLING

calls per hour		observed	expected	$\frac{(O-E)^2}{E}$
	<1400	8	8.14	0.002
1400	1600	6	6.08	0.001
1600	1800	9	7.66	0.233
1800	2000	6	8.17	0.575
2000	2200	8	7.36	0.056
2200	2400	6	5.61	0.028
>2400		7	6.98	0.000

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(obs_i - \exp_i\right)^2}{\exp_i}$$
$$= 0.896$$

χ^2 test

UNIVERSITY of STIRLING
Management School

- H_0: the distribution is normal
- H_1: the distribution is not normal

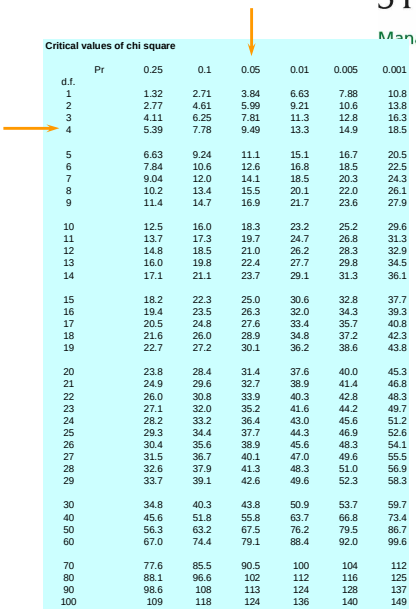
χ^2 (calculated) =0.896

- degrees of freedom
  = no. of classes - no. of pieces of data used in deriving the expected frequencies
  = 7 - 3 (mean, s.d., total) = 4

- signifigance level = 5%
  χ^2 (critical) = 9.49 (from tables)

The chi square table

UNIVERSITY of STIRLING
Management School



Critical values of chi square

Pr	0.25	0.1	0.05	0.01	0.005	0.001
d.f.						
1	1.32	2.71	3.84	6.63	7.88	10.8
2	2.77	4.61	5.99	9.21	10.6	13.8
3	4.11	6.25	7.81	11.3	12.8	16.3
4	5.39	7.78	9.49	13.3	14.9	18.5
5	6.63	9.24	11.1	15.1	16.7	20.5
6	7.84	10.6	12.6	16.8	18.5	22.5
7	9.04	12.0	14.1	18.5	20.3	24.3
8	10.2	13.4	15.5	20.1	22.0	26.1
9	11.4	14.7	16.9	21.7	23.6	27.9
10	12.5	16.0	18.3	23.2	25.2	29.6
11	13.7	17.3	19.7	24.7	26.8	31.3
12	14.8	18.5	21.0	26.2	28.3	32.9
13	16.0	19.8	22.4	27.7	29.8	34.5
14	17.1	21.1	23.7	29.1	31.3	36.1
15	18.2	22.3	25.0	30.6	32.8	37.7
16	19.4	23.5	26.3	32.0	34.3	39.3
17	20.5	24.8	27.6	33.4	35.7	40.8
18	21.6	26.0	28.9	34.8	37.2	42.3
19	22.7	27.2	30.1	36.2	38.6	43.8
20	23.8	28.4	31.4	37.6	40.0	45.3
21	24.9	29.6	32.7	38.9	41.4	46.8
22	26.0	30.8	33.9	40.3	42.8	48.3
23	27.1	32.0	35.2	41.6	44.2	49.7
24	28.2	33.2	36.4	43.0	45.6	51.2
25	29.3	34.4	37.7	44.3	46.9	52.6
26	30.4	35.6	38.9	45.6	48.3	54.1
27	31.5	36.7	40.1	47.0	49.6	55.5
28	32.6	37.9	41.3	48.3	51.0	56.9
29	33.7	39.1	42.6	49.6	52.3	58.3
30	34.8	40.3	43.8	50.9	53.7	59.7
40	45.6	51.8	55.8	63.7	66.8	73.4
50	56.3	63.2	67.5	76.2	79.5	86.7
60	67.0	74.4	79.1	88.4	92.0	99.6
70	77.6	85.5	90.5	100	104	112
80	88.1	96.6	102	112	116	125
90	98.6	108	113	124	128	137
100	109	118	124	136	140	149

χ^2 test

UNIVERSITY of STIRLING
Management School

χ^2 (calculated) =0.896

χ^2 (critical) = 9.49 (from tables)

χ^2 (calculated) < χ^2 (critical)

- variation is not significant (at 5% level)
- a normal distribution is a good fit to the observed data

Chi Squared test – example

UNIVERSITY of STIRLING Management School

- Example: is there a significance between males and females in their main use of mobile phones?
- Categories:
  - Text messaging
  - Voice calls
  - Music/Video
  - Other

Raw frequencies

Random Samples

	Text	Voice	Music/Vid	Other
FEMALE	39	78	12	10
MALE	10	62	41	21

Structuring the Problem

UNIVERSITY of STIRLING Management School

H_0: no difference between sexes

H_1: difference between sexes

- df = (no. rows -1) * (no. columns-1) = 3
- significance level: 5%

  χ^2 (critical) = 7.81 (from tables)

	Text	Voice	Music/Vid	Other	totals
Female	39	78	12	10	139
Male	10	62	41	21	134
Totals	49	140	53	31	273

Analysis: Observed v Expected

Observed	Expected	(O-E)	$\frac{(O-E)^2}{E}$
39	25	14	7.84
78	71.3	6.7	0.63
12	27	-15	8.33
10	15.7	-5.7	2.07
10	24	-14	8.17
62	68.7	-6.7	0.65
41	26	15	8.65
21	15.3	5.7	2.12
		χ^2 =	38.46

- χ^2 (calculated) >> χ^2 (critical)
- Variation is significant (at 5% level)
- Reject the null hypothesis
- Use of facilities of a mobile phone is determined by the gender of the user.

UNIVERSITY of STIRLING
Management School

Questions ?????