University of BRISTOL

spring semester 2018

Hypothesis Testing (2) - Lecture 4

APPLIED STATISTICS - EMAT 30007

Nikolai Bode and Ksenia Shalonova

Department Of Engineering Mathematics

Likelihood ratio test

In some cases we need to perform a hypothesis test to compare two models: big ”general” model (M_B) and small ”simple” model (M_S) nested into the bigger model.

H_O: M_S fits the data.
H_A: M_S does not fit the data and M_B should be used instead.

We need to verify if M_B fits data `significantly` better.

Measure how well a model fits the data: the fit of any model can be described by the maximum possible likelihood for that model: $L(M) = \max P(data|model)$
Calculate the maximum likelihood estimates of all unknown parameters and insert them into the likelihood function.

Work out likelihood ratio: $R = \frac{L(M_B)}{L(M_S)} \geq 1$
Big values of R suggest that M_S does not fit as well as M_B.

Work out log of likelihood ratio: $log(R) = l(M_B) - l(M_S) \geq 0$
Big values of R suggest that M_S does not fit as well as M_B.

Likelihood ratio test - when to use it?

Example 1

There are a number of defective items produced on a production line in 20 days that follow $Poisson(\lambda)$ distribution: 1 2 3 4 2 3 2 5 5 2 4 3 5 1 2 4 0 2 2 6.

M_S: the sample comes from $Poisson(2)$
M_B: the sample comes from $Poisson(\lambda)$

Example 2 (MATLAB session)

Clinical records give the survival time for 30 people : 9.73 5.56 4.28 4.87 1.55 6.20 1.08 7.17 28.65 6.10 16.16 9.92 2.40 6.19 ...
In a clinical trial of a new drug treatment 20 people had survival times of: 22.07 12.47 6.42 8.15 0.64 20.04 17.49 2.22 3.00 ...
Is there any difference in survival times for those using the new drug?

M_S: both samples come from the same $Exponential(\lambda)$ distribution
M_B: the first sample comes from $Exponential(\lambda1)$ and the second sample comes from $Exponential(\lambda2)$

Likelihood ratio test - main steps

Definition

If the data come from $L(M_S)$, and $L(M_B)$ has k more parameters than $L(M_S)$ then $X^2 = 2log(R) = 2(l(M_B) - l(M_S)) \approx \chi^2(kdf)$

- Work out maximum likelihood estimates of all unknown parameters in M_S.
- Work out maximum likelihood estimates of all unknown parameters in M_B.
- Evaluate the test statistic: $\chi^2 = 2(l(M_B) - l(M_S))$
- The degrees of freedom for the test are the difference between the numbers of unknown parameters in two models. The p-value for the test is the upper tail probability of the $\chi^2(kdf)$ distribution given the test statistic.
- Interpret p-value: small values give evidence that the null hypothesis - M_S model - does not hold.

Likelihood ratio test - Poisson example (1)

There are a number of defective items produced on a production line in each of 20 days that follow $Poisson(\lambda)$ distribution: 1 2 3 4 2 3 2 5 5 2 4 3 5 1 2 4 0 2 2 6.

H_O: $\lambda = 2$ small model M_S

H_A: $\lambda \neq 2$ big mode M_B

Log-likelihood for the Poisson distribution: $l(\lambda) = (\sum_{i=1}^{20} X_i) log\lambda - n\lambda + K$

$\underline{M_B}$

MLE for unknown parameter:

$\hat{\lambda} = \frac{\sum x_i}{n} = 2.9$

Maximum possible value for the log-likelihood:

$l(M_B) = 58log(2.9) - 20 \times 2.9 = 3.7532 + K$

$\underline{M_S}$

MLE for unknown parameter: NO unknown parameters

Maximum possible value for the log-likelihood:

$l(M_B) = 58log(2) - 20 \times 2 = 0.2025 + K$

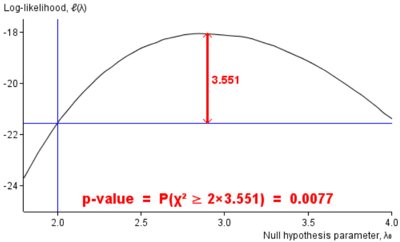
Likelihood ratio test - Poisson example (2)

Likelihood ratio test: Test statistic $\chi^2 = 2(l(M_B) - l(M_S)) = 7.101$

It should be compared to $\chi^2(1df)$ since the difference in unknown parameters is equal to 1.

The p value is 0.008 (the upper tail probability above 7.101).

Interpreting p value: p-value is very small and we conclude that there is strong evidence that M_B model fits the data better than M_S model: $\lambda \neq 2$.



Comparing two means in normal distributions

Note that a 2-sample test and a paired t-test are two different tests!

A 2-sample t-test should be used to compare group means when you have independent samples.

Drinking water containing high concentrations of trace metals can be a health hazard. An experiment was conducted to assess whether zinc concentration is different on the surface than in bottom water. 30 samples of water were obtained from different locations and these were randomly separated into groups of 15. In one group, zinc concentration was measured at the surface, whereas it was measured at the bottom of the other sample.

A paired t-test is needed when each sampled item in one group is associated with an item sampled from the other group.

Researchers were interested in whether giving newborn lambs an injected vitamin supplement increases their weight gain. 30 sets of twin lambs were used in an experiment. In each set of twins one lamb was randomly selected to get the vitamin supplement; the other got no extra vitamin. We want to assess whether the vitamin affects lamb weight.

Two-sample t-test (1)

We can carry out a hypothesis test to verify if the two means are equal:
H_0: $\mu_1 = \mu_2$
H_A: $\mu_1 \neq \mu_2$
The corresponding one-tailed alternative also holds.

Definition

If $\bar{X}_1$ and $\bar{X}_2$ come from $Normal(\mu_1, \sigma^2)$ and $Normal(\mu_2, \sigma^2)$ with sample sizes n_1 and n_2 then $T = \frac{\bar{X}_1 - \bar{X}_2}{SE(\bar{X}_1 - \bar{X}_2)} \approx t(n_1 + n_2 - 2df)$ provided $\mu_1 = \mu_2$

For relatively large sample sizes (Central Limit Theorem) we can use Z-test instead of t-test.

Two-sample t-test (2)

Example

A botanist is interested in comparing the growth response of dwarf pea stems to two different levels of the hormone indoleacetic acid (IAA). The botanist measured the growths of pea stem segments in millimetres for (0.5×10^{-4}) IAA level: 0.8 1.8 1.0 0.1 0.9 1.7 1.0 1.4 0.9 1.2 0.5 and for (10^{-4}) IAA level: 1.0 1.8 0.8 2.5 1.6 1.4 2.6 1.9 1.3 2.0 1.1 1.2. Test whether the larger hormone concentration results in greater growth of the pea plants.

We have two independent samples with $n_x = 11$ and $n_y = 13$.

H_0: $\mu_x = \mu_y$; H_A: $\mu_x < \mu_y$

The pooled estimate assumes that the variance is the same in both groups:

$S^2 = \frac{10\times(S_x)^2+12\times(S_y)^2}{22} = 0.3033$

Test statistic: $t = \frac{1.027-1.662}{\sqrt{0.3033(1/11+1/13)}} = -2.81$

p-value for 22 degrees of freedom in t-distribution: $P(t \leq -2.81) = 0.0051$

Interpreting p-value: There is very strong evidence that the mean growth of the peas is higher at the higher hormone concentration.

Paired t-test (1)

Testing whether two paired measurements X and Y have equal means is done in terms of the differences $D = Y - X$.

The hypothesis
H_0: $\mu_x = \mu_x$
H_A: $\mu_x \neq \mu_y$

can be re-written as

H_0: $\mu_d = 0$
H_A: $\mu_d \neq 0$
This can reduce the paired data set to a univariate data set of differences. The hypothesis can be assessed using t-test: $t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$.
Z-test can be used for relatively large sample sizes.

Paired t-test (2)

A researcher studying congenital heart disease wants to compare the development of cyanotic children with normal children. Among the measurement of interest is the age at which the children speak their first word.

Pair of siblings	Cyanotic sibling	Normal sibling	Difference (cyanotic - normal)
1	11.8	9.8	2.0
2	20.8	16.5	4.3
3	14.5	14.5	0.0
4	9.5	15.2	-5.7
5	13.5	11.8	1.7
6	22.6	12.2	10.4
7	11.1	15.2	-4.1
8	14.9	15.6	-0.7
9	16.5	17.2	-0.7
10	16.5	10.5	6.0

The researcher wants to test whether cyanotic children speak their first word later on average than children without the disease.

H_0: $\mu_d = 0$

H_A: $\mu_d > 0$

$t = \frac{\bar{d}-0}{s_d/\sqrt{n}} = 0.880.$

The p-value is well above zero (0.201), so there is no evidence that the cyanotic children learn to speak later.

Paired t-test (3)

The blood pressure of 15 college-aged women was measured before starting to take the pill and after 6 months of use.

	Blood pressure	
Subject	Before pill	After pill
1	70	68
2	80	72
3	72	62
4	76	70
5	76	58
6	76	66
7	72	68
8	78	52
9	82	64
10	64	72
11	74	74
12	92	60
13	74	74
14	68	72
15	84	74

A two-tailed test is used as the pill might either increase or decrease blood pressure.

H_0: $\mu_d = 0$

H_A: $\mu_d \neq 0$

$t = \frac{\bar{d}-0}{s_d/\sqrt{n}} = -3.105.$

The p-value (0.008) is very small that gives strong evidence that blood pressure has changed.

What does the negative t value suggest?

Testing proportions - example of a one-tailed test

Example

There was a double-blind experiment to assess if aspirin helped stroke patients. 155 patients were randomised to to receive either aspirin or a placebo for 6 months (77 took placebo and 78 - aspirin). A researcher classified the progress of each patient as favourable or unfavourable. Proportion with favourable progress taking placebo is 0.558 and taking aspirin is 0.808. Does aspirin increase the probability of a stroke patient making favourable progress after 6 months?

H_0: $p_a = p_p$
H_A: $p_a > p_p$

$$Z = \frac{\hat{p_a} - \hat{p_p}}{\sqrt{\frac{\hat{p_a}(1-\hat{p_a})}{n_a} + \frac{\hat{p_p}(1-\hat{p_p})}{n_p}}} = 3.46$$

p-value = tail area < 0.001

Conclusion: There is a very strong evidence that stroke patients are more likely to make favourable progress after 6 months if they take aspirin.