



STAT2012 LECTURE NOTES

STATISTICAL TEST (ADVANCED)

Contents

Course summary	8
Introduction	8
Methodology of Statistical testing.....	9
Basic concepts:.....	9
Population	9
Sample.....	9
Dataset:	9
Steps of testing:	9
1. Model and assumptions.....	9
2. Hypothesis on (the parameter in) the population	9
3. Finding the test statistics, T	10
4. Calculate the p – value (observed significance level)	10
5. Conclusion.....	11
Review of important results:	11
Sample mean:	11
Sample variance:	12
For $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$ Random variables:.....	12
Useful R commands:	12
Tests for Population mean:.....	12
Normal population	13
t – test for one sample.....	13
Paired sample.....	16
Large sample size tests for population mean	18
Theory: Central Limit theorem.....	21
Significance level: Decision rule	23
Significance level α	23
Test for mean supply	23
Critical value and rejection region	23
Decision rule	23
Performances of statistical tests.....	27
Types of errors	27
Classifications.....	27
Probability of errors	28
Power of a test.....	28
Remarks:	28

Power curves.....	30
Remarks: (increasing power of a test)	30
Choice of sample size	30
Example 5.3:.....	31
Confidence Intervals	31
Definition of Confidence interval.....	32
Remarks:	32
Confidence intervals for the mean μ	32
Assumption 1:	32
Relationship between testing procedure and CI	33
Assumption 2:	33
One sided CI and duality	34
Example 6.1 (beer)	34
Non parametric testing.....	35
Binomial model	35
In general:	35
Sign test for median:.....	36
The sign test:.....	36
Hypothesis 1:.....	36
Hypothesis 2:.....	37
Test statistic:	37
P value:.....	37
Conclusion:.....	37
The sign test: in general	37
Hypothesis:	37
Test stat:.....	37
P value.....	37
Rank:	38
Definition	38
Non parametric tests	38
Wilcoxon sign rank test.....	38
Wilcoxon sign rank test.....	38
Transformation of data to symmetry	42
Transformation of data to symmetry	42
Skewed data definition	42
Transformations for right skewed data:	43

Median	43
Definition	43
Tests for median	44
Two sample t test	46
One sample and paired sample (Review)	47
One sample t test.....	47
Paired sample.....	47
Two sample problems.....	47
General two sample t test.....	47
Wilcoxon rank-sum test	52
Introduction	52
Wilcoxon rank-sum test:	52
Remarks on Wilcoxon rank sum test: (testing distribution functions)	55
In general: Testing distribution functions that are shifted	56
Review of χ^2 distribution and introduction to F distribution.....	56
χ^2 (chi squared) distribution	56
Definition	56
F distribution.....	57
Definition of F distribution	57
Distributions derived from normal distribution (review)	58
Standard normal	59
Student's T distribution.....	59
Chi squared	59
F distribution	59
Testing Hypothesis concerning variances: $H: \sigma^2 = \sigma_0^2$	59
In general of variance testing:	60
Construct a test (chi squared).....	60
Fixed level test	61
Comparing the variance of 2 normal distributions	62
Constructing the test	62
Fixed level F test.....	62
Considerations in design of experiments.....	64
Elements affecting power	64
Designing experiments to increase accuracy of a test.....	64
Generally	65
Paired sample or independent sample?	65

ANOVA	67
One way ANOVA	67
Example: (4 kinds of teaching techniques)	68
In general:	68
Anova for one way variance.....	69
F test.....	70
The ANOVA table	70
Remarks on one way ANOVA.....	72
Anova in R	72
Example (marks in different classes)	72
Transformations for comparison	74
Example (iron).....	75
Individual and multiple comparisons.....	79
Multiple Comparrisons	81
The Bonferroni method	81
Multiple Comparrisons in R.....	83
Multiple Comparrisons in Splus (not R)	85
Kruskal- Wallis Test (multiple Comparrisons)	87
Design and additive decomposition.....	90
Block design:	91
Method 1: 1 way layout	91
Method 2: 2 way layout (randomised block design)	91
Two layout method in general:.....	92
Adding m experiments:.....	92
Additive decomposition	93
Parameters.....	94
Fitted values and additive models	94
Fitting by median: Additive decomposition.....	99
Transforming the data and interaction.....	99
Two way ANOVA	106
Statistical model for two way data:	107
Single Replicates	107
Friedman test: (non parametric method of 2 way anova).....	112
Friedman test:.....	112
Ranked data	113
Freidman test in R:	115

Replicated Two Way ANOVA	116
Factorial design (data setup)	116
Statistical model for 2 way data layout with replicated	117
Hypothesis for 2 way layout replicates:.....	118
Regression analysis: Least squares	130
Introduction:	130
EXAMPLE:	131
Example 24.1: prediction of heights	131
Linear regression model.....	132
Parameters of regression models.	132
Fitting the model:.....	132
Estimation theory in the SLR model.....	138
Formulas:	138
Properties of the LSE:.....	138
Distribution of the estimators α, β and S^2	140
SLR statistical inference:	141
Tests about the slope β	142
F test for linear regression	148
Prediction at a specified $x = x_0$	151
Correlation	155
Regression Analysis: Robust Method.....	161
Example: fuel frame (from Lecture 25).....	162
Remarks about regression diagnostics:	164
General Robust Fit regression:.....	164
Goal of Robust Regression:	165
Robust regressions:.....	165
1. Least absolute deviation (LAD) regression (L_1)	165
2. Least trimmed squares regression L_2	166
3. M –estimates of regression:.....	167
4. Robust MM regression.....	168
Kernel smoothing.....	170
Idea of smoothing:	171
Definition:	171
Remarks:	171
Connection with least squares:.....	171
Kernel smoothing:.....	171

Remarks:	172
Commonly used kernel functions	172
Kernal smoothing in R	174
k –NN (k Nearest neighbour) Regression	177
Set up:	178
kNN vs smoothing:	178
Remarks:	178
Locally Weighted Scatterplot smoothing (LOWESS/LOESS)	180
Set up of LOWESS:.....	181
Comparing smoother functions	183
Tests for normality:	185
Testing goodness of fit:.....	185
Kolmogorov-Smirnov test	185
Analysis of Categorical Data.....	191
Introduction to categorical data:	192
The multinomial distribution:	192
In general:	193
Chi squared goodness of fit test (χ^2)	197
Construction of the test:	197
Test for homogeneity in Categorical Data	202
Tests for homogeneity	203
In general:	203
Test for Independence:	205
In general:	206
Fisher's Exact test:	209
Hypergeometric distribution:	210
Revision:.....	211
Methodology of statistic tests:	211
One sample (paired sample)	211
T test:	211
Sign test: (non parametric)	212
Wilcoxon sign-rank test.....	212
Fixed level α tests	213
Confidence interval:.....	214
One sample: performance of statistical tests	214
Type 1 error:.....	214

Type 2 error:.....	214
Power of a fixed level test:.....	215
Considerations in design of experiments:	215
Two samples:	215
Two sample t test:.....	215
Comparing variances of 2 normal distributions:.....	215
Wilcoxon rank sum test:	215
ANOVA: one way data.....	216
One way ANOVA:	216
Kruskal-Wallis test (nonparametric)	218
ANOVA: two way data	218
The two –way ANOVA table.....	218
Regression analysis:	219
Simple linear regression:.....	219
Robust regression and Kernel smoothing:	221
Analysis of Categorical Data.....	222
χ^2 (chi squared) test: Goodness of fit test	222
Test for homogeneity:.....	223
Test for independence:	223

STAT2012 STATISTICAL TESTS LECTURE NOTES

Lecture 1. Monday, 25 July 2016

Course summary

1 quiz 14 Sep (5)

2 assignments (10) (Aug 24, Oct 12)

Weekly computer labs (10%)

1 hour comp exam (open book) (10) (26 Oct)

2 Hour final (65%)

HAND IN COMP LABS BY 5PM THE DAY OF THE LAB

Introduction

Assumptions: sample $X_1, X_2, \dots, X_N$

Hypotheses: H vs H_A

Test statistics: $T = T(X_1, X_2, \dots, X_n)$

p – value: $p = P_H(T \geq t)$

Conclusion: by “interpreting” the p value

- Often come across data in various sizes/formats. Data is the raw material for statistics

Example 1.1 Beer labelling

A brand claims beer content is 375 mL on the label. A sample of 40 bottles gives a sample average of 373.9 and a standard deviation of 2.5

- Is the labelling correct?

Eg 1.2: Height comparison

Is there evidence that girls are taller than boys on their 10th birthday?

- The nature of data is different to the first one; as there are 2 samples here (boys and girls) compared to just the beer bottles

Eg 1.3: Temperature

With a bunch of observed temperatures; is there evidence it is NOT normally distributed?

Aim of course:

- Attempt to answer these yes/no questions. Introduce methods of applied stats including parametric and non parametric; for the analysis of data and techniques of statistical inference for 1,2 or several samples

Methodology of Statistical testing

*Basic concepts:**Population*

Collection of all possible results that are target of interest, described by random variable X

Sample

Is a subset of the population; $X_1, \dots, X_n$

- We require that each element in a dataset is **representative** of the population; this is equivalent to saying that $X_1, \dots, X_n$ are required to be iid RVs

Dataset:

$x_1, \dots, x_n$ is an observed value of the sample $X_1, \dots, X_n$. Sometimes also sample or sample value

*Steps of testing:*1. *Model and assumptions*

The model represents our belief about the probability distribution F describing the population; with $X \sim F_\theta(x)$ where $F_\theta(x)$ is a specific distribution function only depending on the unknown parameter θ .

Eg: $X \sim N(\mu, \sigma^2)$; $X \sim B(n, p)$ with parameters μ, σ^2, p these corresponding statistical analysis therefore are called **parametric**

A statistical analysis is called **nonparametric** if our assumption on the population distribution F do not specify a particular class of probability distribution.

Parametric hypothesis testing:

$X_1, \dots, X_n$ iid RV with $X_i \sim F_\theta(x) \forall i \in (1, n)$

Non parametric hypothesis testing

$X_1, \dots, X_n$ are iid random variables with distribution function F

Note: in general we also require F to not be heavy (long)-tailed; require certain moment conditions. When the F is heavy tailed, the boxplot of the dataset should be skewed. For skewed data, we can sometimes transform them into symmetric. (Discussed later)

2. *Hypothesis on (the parameter in) the population**Null hypothesis*

The statement against which we are searching for evidence is called the null hypothesis, denoted H_0 . The null hypothesis is often a “no difference” statement.

Alternative hypothesis

The statement we will consider if H_0 is false is called the alternative hypothesis, H_1

Eg: if $X \sim F_\theta(x)$ we may have

$$H_0: \theta = \theta_0 \text{ (against)}$$

$$H_1: \theta > \theta_0 \text{ (right sided alternative)}$$

$$\theta < \theta_0 \text{ (left sided alternative)}$$

$$\theta \neq \theta_0 \text{ (two sided alternative)}$$

Example 1.1 (cont) Beer labels

Assume claim amount from $X \sim N(\mu, \sigma^2)$ then null hypothesis is:

$$H_0: \mu = 375$$

$$H_1: \mu < 375 \text{ (as seeing if less beer than labelled)}$$

Eg 1.2: children height

Assume girls $X \sim N(\mu_1, \sigma_1^2)$ and boys $Y \sim N(\mu_2, \sigma_2^2)$ then null hypothesis $H_0: \mu_1 - \mu_2 = 0$; $H_1: \mu_1 - \mu_2 > 0$

Eg 1.3 Temperature normally distributed

$$H_0: X \sim N(\mu, \sigma^2); H_1: X \not\sim N$$

3. Finding the test statistics, T

A function $T := T(X_1, \dots, X_N)$ of the sample $X_1, \dots, X_n$ is called a test statistics. To be a useful statistics for testing H_0 , T is chosen such that:

- The distribution of T is completely determined when H_0 is true; $\theta \in \Theta_0$
- The particular observed values of T , called rejection region, can be taken as evidence of poor agreement with the assumption that H_0 is true.

A **test statistic** is a statistic (a quantity derived from the sample) used in statistical hypothesis testing.^[1] A hypothesis test is typically specified in terms of a test statistic, considered as a numerical summary of a data-set that reduces the data to one value that can be used to perform the hypothesis test. In general, a test statistic is selected or defined in such a way as to quantify, within observed data, behaviours that would distinguish the null from the alternative hypothesis, where such an alternative is prescribed, or that would characterize the null hypothesis if there is no explicitly stated alternative hypothesis.

4. Calculate the p – value (observed significance level)

For a given data $x_1, \dots, x_n$ we can calculate $t_{obs} = T(x_1, \dots, x_n)$, the observed value of the test statistic T .

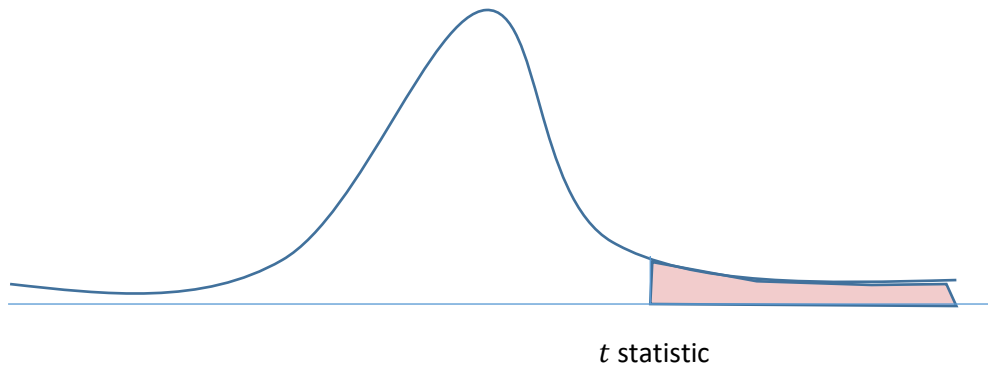
The p value, or observed significance level, is defined formally as the smallest α level, such that its corresponding rejection region R contains t_{obs}

Eg:

$$H_0: \mu = \mu_0$$

$$H_1: \mu < \mu_0$$

Test statistic: $t = 2.594$ with $n = 7$



Eg: assume large values of T support the alternative, that is: R has the form $[r, \infty)$ where r can be calculated

- The smaller the p value is equal to the probability under H_0 that the test statistica T is as extreme (or more) as t_{obs} ; quantifies evidence against H_0 in the following sense:
 - If p is small, then either H_0 is true and poor agreement due to an unlikely event; or H_0 is false, therefore **the smaller the p value, the stronger the evidence against H_0 in favour of H_1**
 - A larger p value does not mean that there is evidence H_0 is true, only that th test detects no inconsistency between predictions of H_0 and true experiment

5. Conclusion

The p value is a probability, $p \in [0,1]$. We can assess the significance of the evidence provided by the data against H_0 by interpreting the p – value.

Although such rules are arbitrary, we use the following convention in this course:

$p > 0.1$ data consistent with H_0

$p \in [0.05,0.1)$ borderline against H_0

$p \in [0.025,0.05)$ reasonable strong evidence against H_0

$p \in [0.01,0.025)$ strong evidence against H_0

$p < 0.01$ very strong evidence against H_0

Review of important results:

Sample mean:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Sample variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

For $X_1, \dots, X_n$ iid $N(\mu, \sigma^2)$ Random variables:

For $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ iid continuous RV's:

$$\begin{aligned} \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} &\sim N(0,1) \\ \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} &\sim \chi_n^2 \\ \frac{(n-1)S^2}{\sigma^2} &\sim \chi_{n-1}^2 \\ \frac{\bar{X} - \mu}{\left(\frac{S}{\sqrt{n}}\right)} &\sim t_{n-1} \end{aligned}$$

$\bar{X}$ and S^2 are independent random variables

Useful R commands:

`pnorm(x,m,s)` $P(X \leq x)$ with $X \sim N(m, s^2)$

`pt(x,d)` $P(X \leq x); X \sim t_d$

`pchisq(x,d)` $P(X \leq x); X \sim \chi_d^2$

`qnorm(p,m,s)` is a number x such that $P(X \leq x) = p$ where $X \sim N(m, s^2)$

Lecture 2. Tuesday, 26 July 2016

Tests for Population mean:

- Normal population with unknown variance
- One sample and paired sample
- One sample t test and paired sample t test

Reminder of yesterday:

1. Assumptions: sample $X_1, \dots, X_n$
2. Hypotheses H_0 vs H_A
3. Test statistics $T = T(X_1, \dots, X_n)$
4. p value
5. *conclusion*

Normal population

One sample: suppose we have a sample $(X_1, \dots, X_n)$ of the size n drawn from a normal population with unknown variance. We want to make some statement about the population mean μ .

t – test for one sample

Assumptions

X_j : are iid random variables with $X_j \sim N(\mu, \sigma^2)$ where σ^2 is unknown.

Hypothesis:

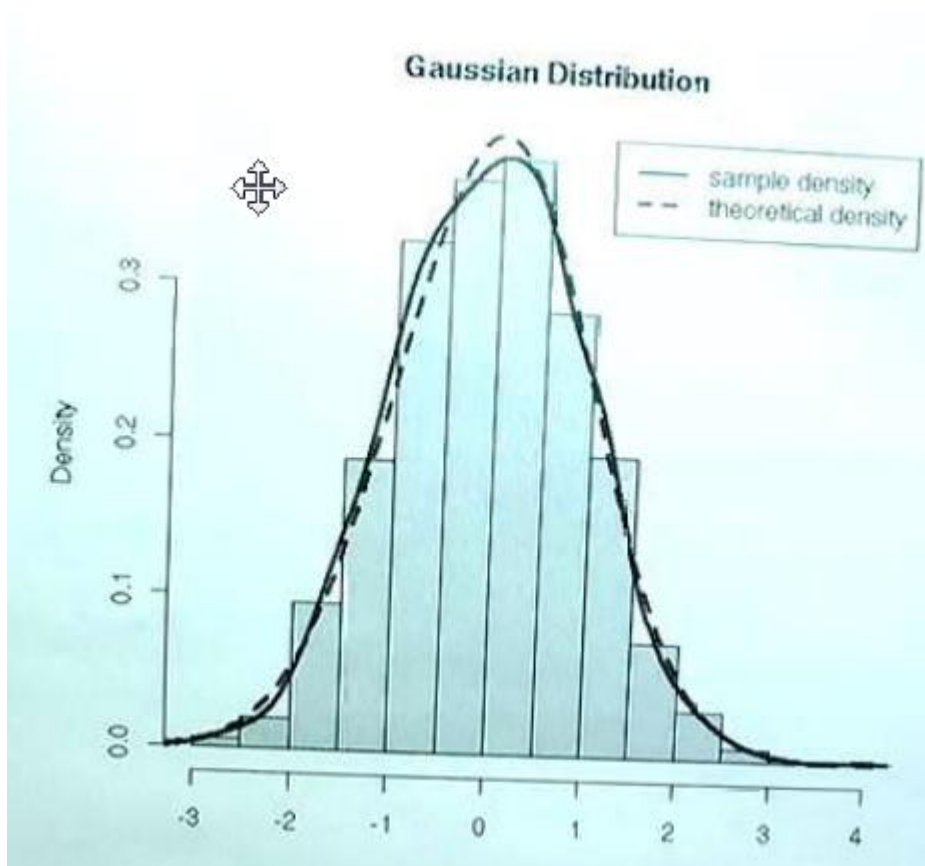
$$H: \mu = \mu_0$$

Vs

$$H_A: \mu > \mu_0 \text{ or}$$

$$\mu < \mu_0 \text{ or}$$

$$\mu \neq \mu_0$$



Test statistics (Student's t – statistics)

$$T_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$$

Where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

Note that $T_n \sim t_{n-1}$ under $H: \mu = \mu_0$

p value

Let $(x_1, x_2, \dots, x_n)$ be a dataset and t be the observed value of T_n ; ie.

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$$

Where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

We then want to find the p value as:

$$p = P(t_{n-1} \geq t) \text{ for } H_A: \mu > \mu_0$$

$$p = P(t_{n-1} \leq t) \text{ for } H_A: \mu < \mu_0$$

$$p = 2P(t_{n-1} \geq |t|) \text{ for } H_A: \mu \neq \mu_0$$

Decision:

We want to interpret our p value.

$$p > 0.1 \text{ data consistent with } H_0$$

$$p \in [0.05, 0.1) \text{ borderline against } H_0$$

$$p \in [0.025, 0.05) \text{ reasonable strong evidence against } H_0$$

$$p \in [0.01, 0.025) \text{ strong evidence against } H_0$$

$$p < 0.01 \text{ very strong evidence against } H_0$$

The p value is the probability; so it is between $[0, 1]$. We can assess the significance of the evidence provided by the data against the null hypothesis by “interpreting” the p value.

Although such rules are arbitrary, in this course we shall use:

P values for STAT2912:

$$p > 0.1 \text{ Data consistent with } H_0$$

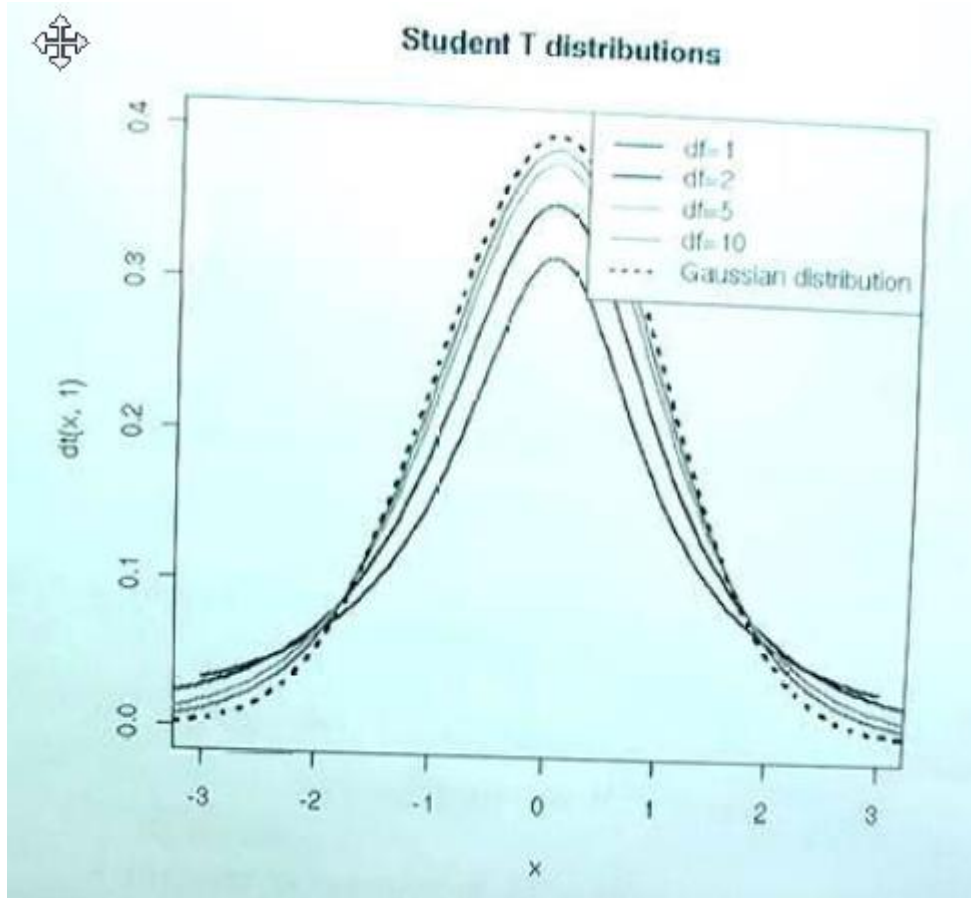
$$p \in [0.05, 0.1) \text{ Borderline evidence against } H_0$$

$$p \in [0.025, 0.05) \text{ Reasonable strong evidence against } H_0$$

$$p \in [0.01, 0.025) \text{ Strong evidence against } H_0$$

$$p < 0.01 \text{ Very strong evidence against } H_0$$

Student's T distribution



- Has a wider tail than the normal; so used if we are not sure if the sample is big enough.

Remarks on t test:

We have $T_n \sim t_{n-1}$ under $H_0: \mu = \mu_0$

$\bar{X}$ is an ideal estimator of μ , which means that $\bar{X}$ is near μ_0 if $H: \mu = \mu_0$ is true. Therefore it is reasonable to believe that we should reject H in favour of H_A if $t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$ (observed value of T_n) is large (negative large or both depending on H_A). On the other hand, if t is large, then p value is small. This coincided with our 'interpretation' of the p value.

Note that $P(t_{n-1} \geq t) = P(t_{n-1} \leq -t)$ for $t \geq 0$. And that $P(|t_{n-1}| \geq |t|) = 2P(t_{n-1} \geq |t|)$.

In R:

`pt(t,n-1)` gives $P(t_{n-1} \leq t)$

*Examples:**Example (beer)*

Beer contents in a six pack is ; is the mean content of the beer less than 375 as labelled?

Solution:

Consider sample $X_1, \dots, X_6$ normally distributed; $X_j \sim N(\mu, \sigma^2)$ with σ^2 unknown

Hypothesis:

$$\mu = 375$$

$$H_1: \mu < 375$$

For the data:

$$\bar{x} = 374.87; s^2 = 0.087$$

And the observed value of the test statistics is:

$$T_n = \frac{\sqrt{n}(\bar{X} - 375)}{s} = \frac{\sqrt{6}(374.87 - 375)}{0.087} = -1.1094$$

The p value is:

$$p = P(t_5 \leq t) = 0.1589 \text{ using the table of } pt(t,5).$$

So the data is consistent with the claim on the label; as $p > 0.1$

In R:

`t.test(x, alternative = "?", mu =  $\mu_0$ )`

is used to construce the one-saple t test, where x is the data, ?? can be less, greater or two.sided depeding on what you want

Remarks:

- For small sample size, ($n \leq 20$) the t test is sensitive to the assumption that the sample is from a normal population. It is recommended that all data be checked for shape by looking at a boxplot and a normal qq-plot of the data (in R: boxplot and qqnorm command)
- For small sample size ($n \leq 20$) it is required that the boxplot of the data should be as symmetric as possible and the points in qqplot be nearly around a straight line. (note boxplot and qqplot may be misguided if n is very small ($n \leq 10$)).
- If the sample size is large ($n \geq 20$); as long as the data does not come from a long-tail distribution, the t test should be fine. (discussed lecture 3)

Paired sample

Suppose we have a sample of paired observations (eg: before/after data or twin data). We want to make some statement about the mean difference.

Example 2/2 (Blood samples from smokers)

Blood samples from indivisuals before/after smoking are used to measure aggregation of blood platelets:

Question: is the aggregations affected by smoking?

Analysis of paired tests:

Let X_i and Y_i be the blood platelets before/after for each individual. We get a sample of paired observations $(x_1, y_1), \dots, (x_n, y_n)$

(note that X, Y are dependent). However, we want to know the difference)

$$d_i = X_i - Y_i$$

As d is from different individuals, we can assume that d_i is ind RV's with $d_j \sim N(\mu, \sigma^2)$ with unknown σ^2

And answer the question by testing the following:

$$H: \mu = 0; H_A: \mu \neq 0$$

This reduces the paired sample problem to a single sample (of a difference), with testing $\mu = 0$

In general;

Let $(x_1, y_1), \dots, (x_n, y_n)$ be a sample of paired observations. We wish to test

$$H: \mu_x = \mu_y$$

Vs

$$H_A: \mu_x >, <, \neq \mu_y$$

Which is the same as $d = X_j - Y_j$

$$H: \mu_d = 0 \text{ vs } H_A \mu_d >, <, \neq 0$$

Paired sample t test:

Assumptions:

$d_i = X_i - Y_i$ are a random sample from a normal population with unknown σ^2 ; that is $d_i \sim N(\mu, \sigma^2)$

Test statistics:

$$T_n = \frac{\sqrt{n}\bar{d}}{s_d} \sim t_{n-1} \text{ under } H$$

P value:

Same as before

In R:

```
t.test(x,y,alternative = "?", mu=9,paired = TRUE)
```

Eg: 2.2

We have

Before: 25 25 27 44 30 67 53 53 52 60 28;
 After: 27 29 37 36 46 82 57 80 61 59 43.

$x \leftarrow c(25, 25, 27, 44, 30, 67, 53, 53, 52, 60, 28)$ > $y \leftarrow c(27, 29, 37, 36, 46, 82, 57, 80, 61, 59, 43)$ >

$t.test(x, y, alternative = "two.sided", mu = 0, paired = T)$

Paired t-Test data: x and y t = -2.9065, df = 10, p-value = 0.0157

alternative hypothesis: true mean of differences is not equal to 0 95 percent confidence interval: -14.93577 -1.97332

sample estimates: mean of x - y -8.454545

the p value value is 0.0156. hence strong evidence against H_0 (so strong evidence people affected by smoking)

Remarks

- Recommended that the data difference shape checked in a boxplot/ normal qqplot
- For small sample, boxplot should be as symmetric as possible; and qqplot as near a straight line.
- For large sample a t test is fine as long as it is not from a long tail distribution

Lecture 3. Wednesday, 27 July 2016

Large sample size tests for population mean

If sample size is large ($n \geq 20$)

Test for population mean:

Suppose we want to test a population mean μ based on a random sample $X_1, \dots, X_n$; where the sample size n is considered to be large enough ($n \geq 20$)

Let $(x_1, x_2, \dots, x_n)$ be an observed value

If the sample does not come from a heavy tail distribution (too many outliers in boxplot), we may use the following procedures to test hypothesis about population mean μ .

Common large sample tests:

Hypothesis:

$$H: \mu = \mu_0; \text{ vs } H_A: \mu >, <, \neq \mu_0$$

Test statistics:

Students t statistic: $T_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$ where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$; $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

(note that we do not have $T_n \sim t_{n-1}$)

Or Z statistic: (given population variance σ^2) $Z_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$ where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

P value

Let t be the observed value of T_n or Z_n , that is

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s} \quad \text{or}$$

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sigma} \quad (\text{with given } \sigma)$$

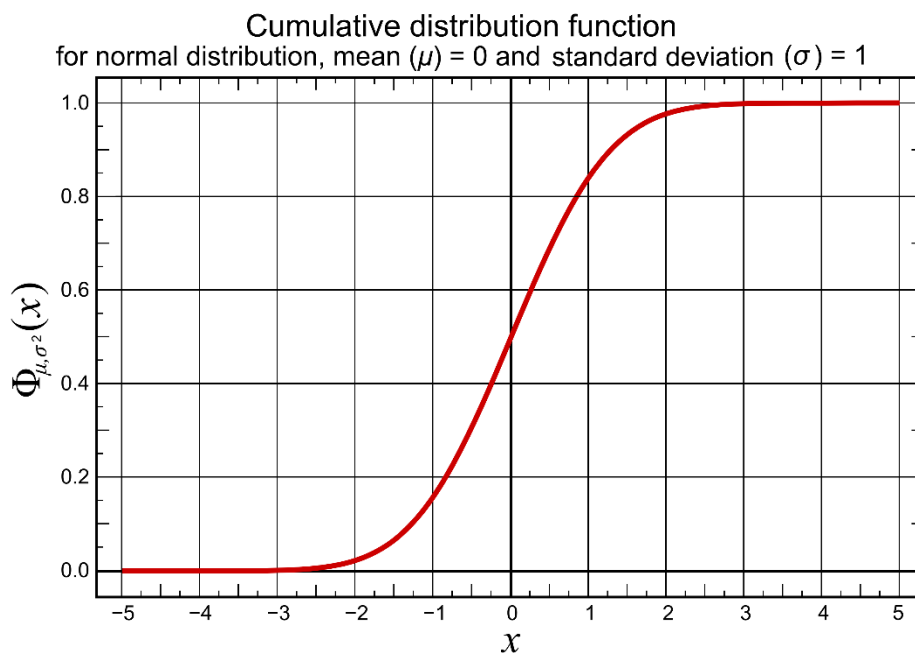
If the sample size n is large enough, the p value is approximately given by

$$p \approx 1 - \Phi(t) \quad \text{for } H_A: \mu > \mu_0$$

$$p \approx \Phi(t) \quad \text{for } H_A: \mu < \mu_0$$

$$p \approx 2(1 - \Phi(|t|)) \quad \text{for } H_A: \mu \neq \mu_0$$

Where $\Phi(t)$ is the CDF of the normal $N(0,1)$ distribution.



Remarks:

- The paired sample may be discussed in a similar manner
- The choice of test statistic is based on the fact that $\bar{X}$ is a point estimate of μ . It means that $\bar{X}$ is close to μ_0 if $H: \mu = \mu_0$ is true. If in reality, $\mu \neq \mu_0$, $|\bar{X} - \mu_0|$ is more likely to be large.
- If the sample is from a normal population with given variance σ^2 , the test procedure for μ (with statistic Z_n) is still applicable for small sample size ($n \leq 20$). Note that under $H: \mu = \mu_0$; $Z_n = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma} \sim N(0,1)$ for each $n \geq 1$ in this case.
- Calculation of the p value is based on the CENTRAL LIMIT THEOREM
- Large sample size for μ is equivalent to t test if the sample size is large enough.

- Note: for large sample size, we don't need to check the box plot/qq plot to check for normality

Example 3.1: (vice president)

A VP of sales claims that people are averaging no more than 15 sales each week. He samples 24 people, and the number of contacts they made recorded.

18, 17, 18, 13, 15, 16, 21, 14, 24, 12, 19, 18,
17, 16, 15, 14, 17, 18, 19, 20, 13, 14, 12, 15

$$H_0: \mu \leq 15; H_A: \mu > 15$$

We get that:

$$t = \frac{\sqrt{24}(\bar{x} - 15)}{s}$$

$$p = 1 - \Phi(t)$$

In R:

```
> x<-c(18,17,18,13,15,16,21,14,24,12,19,18, 17,16,15,14,17,18,19,20,13,14,12,15)
> barx<-mean(x) >
> s<-sqrt(var(x))
> t<-sqrt(24)*(barx - 15)/s
> p<-1-pnorm(t)
> p
[1] 0.007954637
```

As $p < 0.01$ there is strong evidence to indicate the VP is incorrect, that is; the average number of sales contact per week exceeds 15.

We can also do the one sample t test command in R:

```
> x<-c(18,17,18,13,15,16,21,14,24,12,19,18, 17,16,15,14,17,18,19,20,13,14,12,15)
> t.test(x, alternative="greater",mu=15)
```

One-sample t-Test

data: x t = 2.411, df = 23, p-value = 0.0121

alternative hypothesis: true mean is greater than 15 95 percent confidence interval: 15.42167 NA
sample estimates: mean of x 16.45833