Bayesian Methods for Data Analysis in Software Engineering
ICSE 2010 – Tutorial

Mohan Sridharan* Akbar Siami Namin**

*Learning Agents and Stochastic Reasoning Laboratory
Department of Computer Science
Texas Tech University

**AdVanced Empirical Software Testing and Analysis Research Group
Department of Computer Science
Texas Tech University

May 3, 2010 - Cape Town

A Quick Review of Statistical Concepts

Describing Data from A Single Sample

- Mode
- Median
- Mean
- Percentiles
- Data Variation

Describing Data from A Single Sample
Mode – A Measure of Central Tendency

- A measurement that occurs with greatest frequency
- E.g. Number of methods in 12 classes developed by a programmer
  - c(3, 5, 6, 3, 5, 8, 2, 8, 5, 6, 3, 5)
  - Ordered: 2 3 3 3 5 5 5 6 6 8 8
  - Modes: 3 and 5
- Graphical tool: Histogram

Describing Data from A Single Sample
Median – A Measure of Central Tendency

- For odd number of measurements:
  - The middle measurement when the measurements are arranged in order of magnitude
- For even number of measurements:
  - The average of the two middle observations when the measurements are arranged in order to magnitude
- E.g. c(3, 5, 6, 3, 5, 8, 2, 8, 5, 6, 3, 5)
  - median(c(3, 5, …, 3, 5) = 5

Describing Data from A Single Sample
Mean – A Measure of Central Tendency

- The sum of a set of measurements divides by the number of measurements in the set
- E.g. c(3, 5, 6, 3, 5, 8, 2, 8, 5, 6, 3, 5)
  - mean(c(3, 5, …, 3, 5) = 4.916
- Notations
  - $\overline{X}$: Sample mean, $\overline{X} = \frac{\sum X}{n}$
  - μ :Population mean (needs estimation)

Describing Data from A Single Sample
Percentiles – A Measure of Location

- Determine the fraction or percentage of measurements above or below specified values
- Can be used to locate an observation in relation to the remaining ones
- More formal definition:
  - Let $X_1, X_2, \ldots, X_n$ be a set of measurements arranged in order of magnitude. The Pth percentile is the value of X such tat percent of the measurements are less than value of X and (100-p) percent are greater

Describing Data from A Single Sample
Percentiles – A Measure of Location

- Lower quartile
  - The 25^{th} percentile (i.e. 25% of the measurements fall below the lower quartile and 75% above it)
- Upper quartile
  - The 75^{th} percentile (i.e. 75% of the measurements lie below the upper quartile and 25% above it)

Describing Data from A Single Sample

Data Variation – Spread of Data

- Range
  - The difference between the largest and smallest measurements
    - E.g. max(c(3, 5, 6, 3, 5, 8, 2, 8, 5, 6, 3, 5)) – min(c(3, 5, 6, 3, 5, 8, 2, 8, 5, 6, 3, 5)) = 6
- Average deviations
  - Average of the absolute values of the deviations from their mean

$$AD_{\overline{X}} = \frac{\sum |X - \overline{X}|}{n}$$

  - The mean can be replace by median

Describing Data from A Single Sample
Data Variation – Spread of Data

- Variance
  - The sum of the squared deviations of the measurements about their mean, divided by n-1
  - Sample variance $S^2 = \frac{\sum (X - \overline{X})^2}{n-1}$
  - Population variance σ^2
  - E.g. sd(c(3, 5, 6, 3, 5, 8, 2, 8, 5, 6, 3, 5)) = 1.92
- Standard deviation
  - The positive square root of the variance

Estimation
Point and Interval Estimators

- Estimator
  - A rule that tells us how to calculate an estimate using sample information
    - E.g. Sample mean
- Point estimator
  - A single number used to estimate a population parameter
- Interval estimator
  - An estimation of a population parameter formed by two numbers that determine an interval within which we expect the parameter to fall

Estimation
Biased vs. Unbiased Estimators

- Unbiased estimator
  - If the expected value is equal to true value
    - E.g. mean
- Biased estimator
  - Does not satisfy the above property

Hypothesis Testing

- Statistical testing
- One-sample
- Two samples
- More than two samples

Hypothesis Testing
Statistical Testing

- The concept
- Elements of a statistical test
- Type I and II errors
- Selection of null and alternative hypothesis
- The choice of α

Hypothesis Testing
Statistical Testing

- Decide whether a particular set of observations X is too far away from the population mean μ
  - Measured by probability
- The probability that X will lie more than 1.96 standard deviations values of X away from μ is 0.5

Hypothesis Testing – Statistical Testing
Elements of a statistical test

- Null hypothesis
  - An hypothesis about one or more parameters of the population
  - Decision to reject is made by the value of some quantity computed from the sample
- Alternative hypothesis
  - Accept if the null hypothesis is rejected
- Test statistic
  - The observed of computed quantity functions as a statistical decision maker
- Rejection region
  - A set of values of the test statistics which are contradictory to the null hypothesis

Hypothesis Testing – Statistical Testing
Test Statistics

- Difference between the Sample means
  - scaled to number of standard deviations (standard errors) from the null difference of 0 for the Population means:

$$z = \frac{\overline{y}_1 - \overline{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Hypothesis Testing – Statistical Testing
Rejection Region

- Set of values of the test statistic that are consistent with *research hypothesis*, such that the probability it falls in this region when *null hypothesis* is true is α

$$z_{obs} \geq z_{\alpha} \qquad \alpha = 0.05 \Rightarrow z_{\alpha} = 1.645$$

Hypothesis Testing – Statistical Testing
Type I and II Errors

- Type I error
  - Reject the null hypothesis when it is true
  - α : The probability of a type I error
- Type II error
  - Failure to reject the null hypothesis
  - β : The probability of a type II error
- Decision table

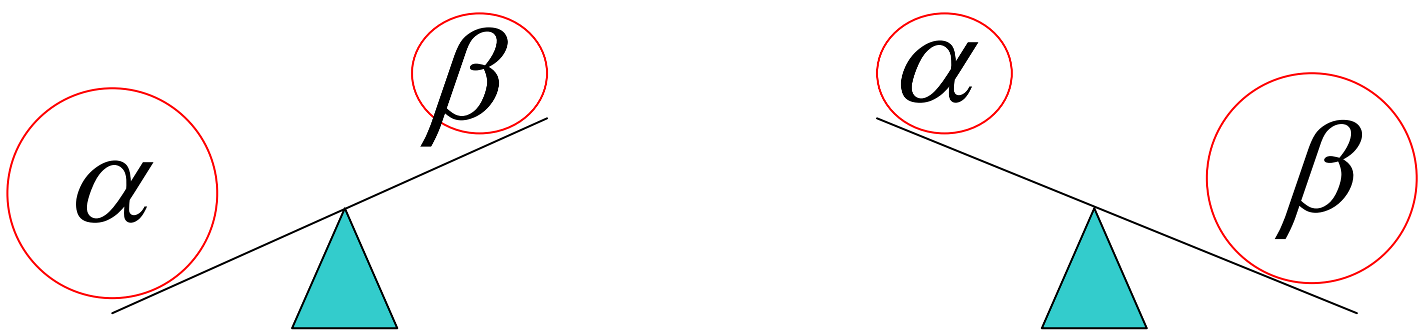
	Null Hypot	**hesis**
Decision	False	True
Reject the null hypothesis	Correct	Type I error
Failure to reject the null hypothesis	Type II error	Correct

Hypothesis Testing – Statistical Testing

Type I and II Errors – The Relationship

$$\alpha = P(Type\ I\ Error) \quad \beta = P(Type\ II\ Error)$$

Traditionally: $\alpha = 0.05 \quad \beta \le 0.20$



Hypothesis Testing – Statistical Testing
Selection of Null and Research Hypotheses

- Research (alternative) hypothesis
  - The research hypothesis that we think is true
- Null hypothesis
  - An hypothesis that is contradictory to the research hypothesis
  - Nullifies the research hypothesis
- Test procedure
  - Designed to verify the research hypothesis by showing that the null hypothesis is false
  - A proof by contradiction

Hypothesis Testing – Statistical Testing

Test of an Hypothesis Concerning mean

- Null hypothesis: $\mu = \mu_o$
- Research hypothesis: $\mu \neq \mu_o$
- Test statistic: $\overline{X}$
- Rejection region:
  - If $\overline{X}$ lies more than 1.96 standard deviations away from μ_o

Hypothesis Testing – Statistical Testing
Selection of Null and Research Hypotheses

- E.g. A software company used a trial version of a testing tool for the entire last year to check whether the tool had increased the fault exposition. One hundred applications were tested (n = 5). State the research and null hypotheses.
  - Research hypothesis
    - The tool has improved the fault detection scores
  - Null hypothesis
    - There is no significant differences between scores

Hypothesis Testing – Statistical Testing
Selection of Null and Research Hypotheses

Last year	This year
12	13
3	5
24	21
8	7
2	6

- E.g.
  - Means
    - Last year=9.8=$\overline{X}$
    - This year= 10.4
- Research hypothesis
  - Mean(last)!=mean(this)
- Null hypothesis
  - Mean(last)=mean(this)
- Rejection region
  - If the observed value of $\overline{X}$ lies more than 1.96 standard deviation away from 10.4

Hypothesis Testing
One and Two-tailed Tests

- One-tailed
  - The research hypothesis states that the parameter under test is less than (or greater than) a specified value.
- Two-tailed
  - The research hypothesis states that the parameter under test is not equal to a specified quantity.

Hypothesis Testing – One-Sample
Goodness-of-fit Test, Chi-square Test

- Null Hypothesis
  - Each of probabilities is equal
- Research Hypothesis
  - At least one of the probabilities differs from the hypothesized value
- Test Statistic $\chi^2 = \sum \frac{(O-E)^2}{E}$
  - O: Observed, E: Expected
- Rejection Region
  - Reject the null hypothesis if χ^2 exceeds the tabulated value of χ^2 for specified α and degrees of freedom.

Hypothesis Testing
Two Samples

- Comparing two population means
- Goodness-of-fit test

Hypothesis Testing – Two Samples
Comparing Two Population Means

- Null hypothesis: $\mu_1 = \mu_2$
- Research hypothesis $\mu_1 \neq \mu_2$
- Test statistic

$$Z = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

- Rejection region
  - Reject the null hypothesis if $z < z_{-\frac{\alpha}{2}} or, z > z_{\frac{\alpha}{2}}$
  - Use s_1^2 and s_2^2 to estimate standard deviations for both populations.

Hypothesis Testing – Two Samples
Comparing Two Population Means

- E.g. Functional vs. Structural
  - The number of faults found through using functional and structural testing
  - The effectiveness of two techniques

	No. Faults Exposed Functional	No. Faults Exposed Structural
A	15	23
B	19	12
C	15	16
D	29	19

Hypothesis Testing – Two Samples
Comparing Two Population Means

- Null hypothesis: $\mu_1 = \mu_2$, 19.5 = 17.5
- Research hypothesis $\mu_1 \neq \mu_2$, 19.5 != 17.5

- Student t.test(F, S)
- t = 0.4949, df = 5.389, p-value = 0.6402

Hypothesis Testing – Statistical Testing
Choice of α

- In large samples, the difference in two sample means is approximately normally distributed:

$$\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2 \, , \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right)$$

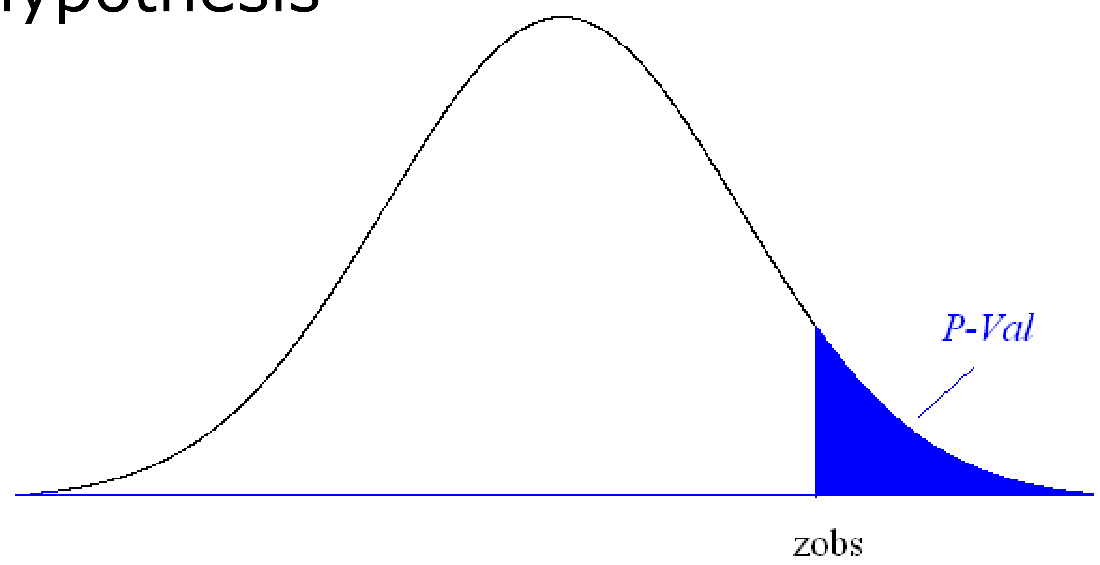
- Under the null hypothesis, μ_1-μ_2=0 and:

$$Z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

- σ_1^2 and σ_2^2 are unknown and estimated by s_1^2 and s_2^2

p-values

- Measure of the strength of evidence the sample data provides against the null hypothesis



$$p - val : p = P(Z \geq z_{obs})$$

Student's t-test

- A hypothetical test concerning a population mean
  - Appropriate for small samples
- A hypothetical test concerning of a difference in means
  - Paired t-test
  - Non-paired t-test
- Degree of freedom

Student's t-test
Why t distribution instead of z

- Z distribution
  - Appropriate for large-sample (>30)
- t distribution
  - Appropriate for small-sample (<30)
  - As the sample size gets larger, the t distribution becomes a standard normal (z) distribution

Student's t-test
Test for Mean

- Null hypothesis: $\mu = \mu_0$
- Research hypothesis: $\mu \neq \mu_0$
- Test statistic: $$t = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$
- Rejection region:
  - Reject if the computed t value (absolute) is greater than the tabulated value.
    - The tabulated value is based on (n-1) degree of freedom with $a = \alpha / 2$