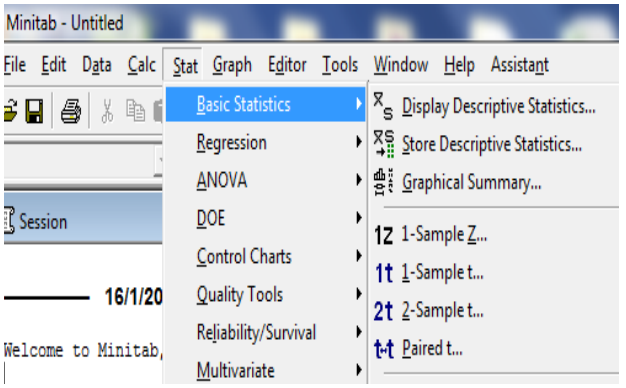
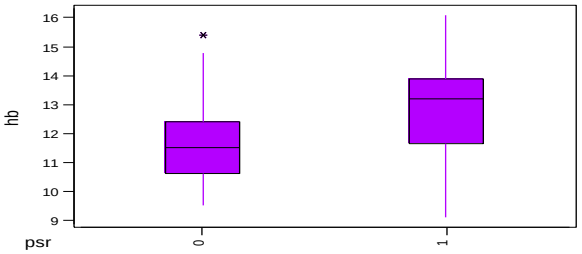
Short Course on

# INTRODUCTORY STATISTICAL DATA ANALYSIS

# 9 & 10 MARCH 2021




**Two-Sample T-Test and CI: DietA, DietB**

```
Two-sample T for DietA vs DietB
          N      Mean      StDev    SE Mean
DietA    10     10.00       3.23        1.0
DietB    10     14.70       6.78        2.1

Difference = mu DietA - mu DietB
Estimate for difference:  -4.70
95% CI for difference: (-9.88, 0.48)
T-Test of difference = 0 (vs not =):
T-Value = -1.98  P-Value = 0.071  DF = 12
```

SCHOOL OF MATHEMATICAL SCIENCES
UNIVERSITI SAINS MALAYSIA, PENANG

MINITAB
Making Data Analysis Easier
www.minitab.com

SPSS

## WHY CATEGORICAL DATA ANALYSIS?

Social scientists typically study a phenomenon involving categorical variables, such as race, gender, marital status and occupation that can be measured using only a limited number of values or categories. Until the late 1960s, association between 2 categorical variables, cross-classified in contingency table, is usually analyzed using chi-sq test. ***However, analysis involving larger tables often run into difficulties with interpretation.***

Including categorical variables as independent variables in regression models does not present any difficulties as it can be analyzed by constructing dummy variables. However, the situation changes drastically when having to use ***categorical variables as dependent variables as much of the theory and techniques of linear regression is simply inapplicable.***

This course aims to introduce a few techniques for analyzing categorical/survey data: ***FACTOR ANALYSIS, LOGISTIC REGRESSION, LOG-LINEAR MODEL***

Short Course on

# STATISTICAL ANALYSIS FOR CATEGORICAL / SURVEY DATA

# 23 & 24 MARCH 2021

# INTRODUCTORY STATISTICAL DATA ANALYSIS

## DAY 1 : WEDNESDAY, 9 MARCH 2021

| | |
|---|---|
| **8:30 AM** | **REGISTRATION** |
| **8:45 AM** | **SESSION 1 – DESCRIPTIVE STATISTICS AND EXPLORATORY ANALYSIS**<br>▪ Measure of Central, Dispersion and Location; Box-Plot and Stem-and-Leaf |
| **10:45 AM** | **TEA BREAK** |
| **11:00 AM** | **SESSION 2 – PRINCIPLES OF HYPOTHESIS TESTING**<br>▪ Distribution for Sample Mean; Significance Level, p-value, 1-sample t-test and z-test |
| **12:45 NOON** | **Q & A (SESSIONS 1 & 2)** |
| **1:00 PM** | **LUNCH** |
| **2:00 PM** | **SESSION 3 – HYPOTHESIS TESTING FOR 2 SAMPLES**<br>▪ Tests for Equality 2 variances, Independent t-test and paired t-test |
| **3:45 PM** | **TEA BREAK** |
| **4:00 PM** | **SESSION 4 – ANALYSIS OF VARIANCE**<br>▪ Hypothesis test for several means; 1-way ANOVA, Post Hocs |
| **5:45 PM** | **Q & A (SESSIONS 3 & 4)** |

## DAY 2 : THURSDAY, 10 MARCH 2021

| | |
|---|---|
| **8:30 AM** | **SESSION 5 – LINEAR REGRESSION ANALYSIS**<br>▪ Correlation, simple regression, coefficient of determination - $R^2$ |
| **10:30 AM** | **TEA BREAK** |
| **10:45 AM** | **SESSION 6 – CATEGORICAL (Survey) DATA ANALYSIS**<br>▪ Chi-sq test statistic; Tests of independence |
| **12:45 NOON** | **Q & A (SESSIONS 5 & 6)** |
| **1:00 PM** | **LUNCH** |
| **2:00 PM** | **SESSION 7 – MORE ON REGRESSION AND CATEGORICAL DATA ANALYSES**<br>▪ Multiple Linear Regression, $\bar{R}^2$; Test of homogeneity and Goodness-of-Fit |
| **3.45 PM** | **TEA BREAK** |
| **4.00 PM** | **SESSION 8 – MORE HANDS-ON EXERCISE**<br>• Recaps on 2-sample Independece t-test and 1-way ANOVA |
| **5:15 PM** | **Q & A (followed by CERTIFICATE PRESENTATION at 5.30pm)** |

## Objectives of Hypothesis Testing

- ❖ Be able to describe fundamental principles of hypothesis testing such as type of tails and p-value
- ❖ Be able to perform appropriate test in different situations
- ❖ Be able to interpret the results and making decision
- ❖ Be able to better understand analysis and results presented in research reports and journal articles

## Lead Facilitator

**Dr Zainudin Arsad** has successfully conducted more than 150 short courses from Jan. 2009 to Dec. 2019 period. He is a senior lecturer, obtaining his BSc. in 1994 and PhD. in 2002 from the Dept. of Actuarial Mathematics and Statistics, Heriot-Watt University, UK. His areas of research are Time Series Analysis and Econometric Modeling with main topic of interest in the Applications of Kalman Filter Technique in Financial Issues. Dr Zainudin is an active member of Sustainable Tourism Research Cluster (STRC) in USM that receives RM4.2million research grant, with main responsibility on investigating determinants influencing tourist satisfaction level on tourism services and products.

**LOG-LINEAR MODEL** is used for Poisson distributed data. As an extension to chi-sq independence test on 2-way table, the log-linear model relates three or more discrete/categorical variables in terms that are very similar to ANOVA. The model reflects various main effects and interactions between the categorical variables. The form of association among the variables can be described by the odds ratios.

**LOGISTIC REGRESSION (or LOGIT model)** is used to model dichotomous/categorical response variable. Since the response variable is not continuous, in the logit model, the log odds of the outcome (the likelihood of the response variable taking a certain limited-choice discrete value), is modeled as a linear combination of the explanatory variables. A simpler **binary logistic regression** has been widely used, among others, to predict the absence/presence of disease and success/failure of loan application.

# STATISTICAL ANALYSIS FOR CATEGORICAL/SURVEY DATA

## DAY 1 : WEDNESDAY, 23 MARCH 2021

| | |
|---|---|
| **8:30 AM** | **REGISTRATION** |
| **8:45 AM** | **SESSION 1 – GRAPHICAL REPRESENTATION**<br>▪ Pie chart & bar chart; boxplot |
| **10:45 AM** | **TEA BREAK** |
| **11:00 AM** | **SESSION 2 – HYPOTHESIS TESTING**<br>▪ Null & alternative, Significance level, p-value, non-parametric test |
| **12:45 NOON** | **Q & A (SESSIONS 1 & 2)** |
| **1:00 PM** | **LUNCH** |
| **2:00 PM** | **SESSION 3 – CROSS TABULATION**<br>▪ Chi-sq tests: Goodness-of-fit & Independence |
| **3:45 PM** | **TEA BREAK** |
| **4:00 PM** | **SESSION 4 – LINEAR REGRESSION**<br>▪ Spearman rank correlation, Dummy variable |
| **5:45 PM** | **Q & A (SESSIONS 3 & 4)** |

## DAY 2 : THURSDAY, 24 MARCH 2021

| | |
|---|---|
| **8:30 AM** | **SESSION 5 – LOGISTIC REGRESSION**<br>▪ Binary dependent, Goodness-of-fit, Odds-ratio, Wald Statistic |
| **10:45 AM** | **TEA BREAK** |
| **11:00 AM** | **SESSION 4 – FACTOR ANALYSIS**<br>▪ Communality, Factor rotation, VARIMAX, Cronbach-alpha |
| **12:45 PM** | **Q & A (SESSIONS 5 & 6)** |
| **1:00 PM** | **LUNCH** |
| **2:00 PM** | **SESSION 7 – LOG-LINEAR MODEL (Assoc. Prof. Dr. Tahir)**<br>▪ Poisson regression, Saturated model, Deviances, interactions |
| **3.45 PM** | **TEA BREAK** |
| **4.00 PM** | **SESSION 8 – MORE HANDS-ON & CASE STUDIES**<br>• More practices on Logistic Regression m& Factor Analysis |
| **5:15 PM** | **Q & A (followed by CERTIFICATE PRESENTATION at 5.30pm)** |

* **Assoc. Prof. Dr. Mohd Tahir** obtained his MSc from USM Penang in 2002 and PhD from Universiti Kebangsaan Malaysia in 2007. His research area is Econometrics Modeling, particularly in Wavelet Analysis and Regime Switching Model Dr. Tahir has been teaching Categorical Data Analysis and Linear Models courses for MSc. Degree.

For more information on content of courses and brochure, kindly contact:
Dr. Zainudin Arsad (013-5159571 or zainudin.arsad@usm.my) or Ms. Noor Farhana Fazil (nfarhana.stat17@gmail.com)

**Course Fee**

**Fee inclusive of SST**
Introductory Statistics: **RM400** per participant (Group 3-6: RM360 per participant)
Categorical/Survey Data: **RM450** per participant (Group 3-6: RM405 per participant)
The fees cover course materials, lunch & morning refreshments and a Certificate of Attendance.

**Postgraduate student** : **RM300** – Introductory Statistics & **RM350** – Categorical Data, requires proof of status.
**Group 3-6 students** : **RM270** – Introductory Statistics & **RM315** – Categorical Data, per student

**Accommodation**

Recommended nearby hotel is U Hotel (RM175 - RM225, reservation at 04-658-1000, only 300m walk from USM). Alternatively, stay at Vistana Hotel (RM225 - RM350, reservation at 04-646-8000, 3km from USM, 10mins by taxi). Muslims can try to get a room at the USM Pusat Islam (only RM70 per night, limited rooms, enquiry at 04-653-3753). For other listings of accommodation please visit www.penang-hotels.com (please make your own reservation).

# REGISTRATION FORM (Closing Date: 7 Mac. & 21 Mac. 2021)

# INTRO. STATISTICAL DATA ANALYSIS (9 & 10 MARCH 2021)

# ANALYSIS FOR CATEGORICAL/SURVEY DATA (23 & 24 MARCH 2021)

Please scan and email this registration form (together with a copy of LO/PO, if applicable) to :

**Noor Farhana Fazil**
School of Mathematical Sciences,
Universiti Sains Malaysia, 11800 USM PENANG.
Email : **nfarhana.stat17@gmail.com**

**NAME OF SHORT COURSE (Please tick (✓) short course(s) to be attended)**

| | | | | |
|---|---|---|---|---|
| Introductory Statistical Data Analysis (9 & 10 March 2021) | | Yes | | No |
| Statistical Analysis for Categorical/Survey Data (23 & 24 March 2021) | | Yes | | No |

*Please register the following name/names*: (Please use separate sheet, if required)

| Item | Name | Designation |
|---|---|---|
| *1. | | |
| | | |
| | | |
| | | |
| | | |
| | | |

| | | |
|---|---|---|
| Industry Sector: | | |
| Company: | | |
| Address: | | |
| | | Postcode: |
| *Primary Person: | | *Mobile Phone: |
| *Telephone No.: | *Fax No. | *Email: |

"I hereby agree that the personal data that I have provided to USAINS, whether now or in future, may be used, recorded, stored, disclosed, or otherwise processed by or on behalf of USAINS in accordance with the Personal Data Protection Act 2010 and USAINS' data protection policy (available at USAINS' website - www.usainsgroup.com), for the purpose of facilitation and organisation of this event, research and audit, and maintenance of a participant database for the promotion of this event, and such ancillary services as may be relevant."

**MODE OF PAYMENT**

| | | | | | |
|---|---|---|---|---|---|
| I enclose | | Number | Bank | No. of Participants: | |
| | Crossed Cheque | | | Postgraduate Student: | |
| | Cash on Day | | | Group Discount: | |
| | Bank Transfer | | | | |
| | LO/PO | | | Total Sum: | RM |
| | | Payment must be made payable to **'Usains Holding Sdn. Bhd.'**. | | | |

1. Telegraphic Transfer. Please note the following:

   Payee Name: Usains Holding Sdn. Bhd.
   Details: Short Course on Introductory Statistical Data Analysis OR Statistical Analysis for Categorical/Survey Data
   Name of Bank: AmBank (M) Berhad, Level 21, Menara Dion, Jalan Sultan Ismail, 50250 Kuala Lumpur.
   Account Number: 888 – 100 – 985 – 0380 Swift Code: ARBKMYKL
   **(Please SCAN & EMAIL your Bank-in Slip (write name on slip) with registration form to nfarhana.stat17@gmail.com)**

2. Proof of **Local Order (LO) or Purchase Order (PO)** must be scanned/emailed to Dr. Zainudin/Ms. Noor Farhana for confirmation and secure of place, and it must be presented during morning registration.
   The Organizer reserves the right to refrain a registered participant from taking part in the event if no proof of payment can be presented. This only applies to registered participants who have NOT paid the registration fee PRIOR to the event date.

**Cancellation / Substitution**

Cancellation must be made in writing through fax, e-mail or post **at least 10 working days** before the course. No refunds are available after this period. In the case of cancellation, **an administration charge of RM150** will be applied. However, substitute participants are welcomed at no extra charge provided written notice is given to the organizer at **least 5 working days** before the event.

| | |
|---|---|
| Date: | Company's Official Stamp |
| Signature: | |