

STATISTICA, AA 2010-11, prof. Roberto D'Angiò

Parte di ANALISI STATISTICA DEI DATI con EXCEL

Modalità di svolgimento di questa parte del corso e sua modalità d'esame. Questa parte del corso si svolge interamente in aula informatica con apprendimento attivo su PC (*hands on and learning by doing*) da parte dello studente. L'esame, che è *open book and notes*, si svolge nell' unica sessione (non replicabile) del 4 Novembre 2010, e consiste in una prova su PC (*no paper*) da parte dello studente che produrrà un file Excel contenente le analisi statistiche richieste dalle domande formulate nel file Word del testo d'esame. Punteggio d'esame: da 0 a 12 trentesimi (il rimanente punteggio da 0 a 18 trentesimi è riservato alla parte d'esame in forma scritta vertente su Calcolo delle Probabilità e Statistica Inferenziale).

PROGRAMMA SVOLTO

Nozioni informatiche di base. *Data base* (o base dati), *record* (righe) di un data base e *campi* di un record (colonne) di un data base. Esempi di data base aziendali. Tipologie di accesso ad un data base: nozione di "tabulato", comandi di query (o di interrogazione) di un data base, reporting interno gestionale.

Nozioni statistiche di base. Unità osservativa, variabili con valori numerici, variabili con valori non numerici, dati grezzi, analisi statistica dei dati grezzi, informazione. Analisi statistica dei dati grezzi uni-dimensionale (o uni-variata) e bi-dimensionale (o bi-variata). Rappresentazioni grafiche nel piano cartesiano dei dati grezzi di due variabili: grafico di sequenza dei dati grezzi o grafico della serie storica dei dati grezzi (quarta icona del comando "grafico a dispersione (XY)" di Excel) e grafico di dispersione dei dati grezzi (prima icona del comando "grafico a dispersione (XY)" di Excel). Motivi per cui il grafico di sequenza non è sufficiente e si rende necessario il grafico di dispersione.

Analisi statistica uni-dimensionale (o uni-variata) dei dati grezzi.

Tabella per dati singoli (comandi tabella pivot "raggruppa/trascina/separa" per variabili numeriche, solo "trascina" per variabili non numeriche) delle frequenze assolute, relative percentuali (comando "% del Totale complessivo"), assolute cumulate (comando "Totale parziale in"), relative percentuali cumulate (comando "% del Totale parziale in"). Uso dei filtri di Excel. Uso delle formule di calcolo di Excel. Grafico a barre verticali (con tabella pivot). Significato delle frequenze relative percentuali cumulate (lettura diretta della tabella delle frequenze).

Nozione di indice statistico. Indici statistici di posizione ed indici statistici di variabilità o dispersione.

Indici statistici: minimo, massimo, range, moda, media e sue proprietà, varianza (e deviazione standard) e sue proprietà. La varianza (deviazione standard) come indice di affidabilità della media. Comandi Excel e/o formule di calcolo Excel per gli indici statistici. Quantili: decili, quartili, mediana, lettura inversa della tabella delle frequenze. Motivi e conseguenze dell'uso delle frequenze percentuali cumulate *convenzionali* *alfa%* (10%, 20%, ecc., 25%, 50%, ecc.) invece di quelle *effettive* della tabella.

Tabella per dati in intervalli (solo per variabili con valori numerici) delle frequenze assolute, relative percentuali, assolute cumulate, relative percentuali cumulate e loro significato (comando tabella pivot "raggruppa/raggruppa per *lunghezza intervallo*/trascina" e poi, senza fare "separa", si calcolano le frequenze come per la tabella per dati singoli).

Analisi statistica bi-dimensionale (o bi-variata) dei dati grezzi.

Tabella delle frequenze assolute congiunte e relative percentuali *congiunte* e loro significato (con comandi tabella pivot "trascina" nelle righe la prima variabile, "trascina" nelle colonne la seconda variabile, e poi si procede come con una sola variabile). Significato delle frequenze nella *colonna marginale* e nella *riga marginale* della tabella.

Tabella delle frequenze relative percentuali *condizionate* e loro significato (con comandi tabella pivot "% del totale di riga", "% di totale di colonna).

Analisi della regressione lineare di due variabili. Retta di regressione e metodo dei minimi quadrati. Indice "R quadro" (o "coefficiente di correlazione lineare al quadrato") e suo significato (comandi Excel "aggiungi linea di tendenza", "tipo di regressione: lineare" che è già impostato di default, "visualizza l'equazione sul grafico", "visualizza il valore di R quadro sul grafico"). Uso della retta di regressione per la stima o previsione del valore di una variabile in corrispondenza di un valore dell'altra ed uso del valore di "R quadro" per valutare l'affidabilità della stima o previsione ottenuta con la retta di regressione.