9/5일 첫 수업에서

0. 2 시간 수업으론 전체 수업을 다 할 순 없습니다. 따라서 중요한 부분만 건드리면 진행 할 것입니다. 여러분은 첫 시간엔 컴퓨터를 끄고 집중해야 합니다. 차근차근 나가면 앞 부분만 건드리게 됩니다. 필요한 내용을 다루어서 일단 진행해 나가는 것이 중요하며 나중에 통계학이 필요할 때 다시 볼 수 있는 정도를 바랍니다(연결고리, 포기하지 않는 것 미워하지 않는 것)

1. 한 시간은 이론이나 실기 설명을 하고 10분 쉰 후 1시간은 엑셀 연습합니다. 수업 시간에 미쳐 하지 못한 연습 문제 등을 과제물로 그 주 일요일까지 완성해서 메일에 파일 첨부하여 제출합니다. 한 번이라도 거르게 되면 다음 수업과 연결이 끊어지고 포기하게 됩니다. 꼭 자신의 힘으로 이리저리 해서 제출하기 바랍니다.

2. 수업일정을 보니 책 2장과 3장 내용은 오늘 해야 합니다.

2장은 그래프를 이용한 자료 보여주기이고

3장은 수치로 보여주는 통계 요약입니다.

수업 중엔 2장의 100명 학생의 몸무게 자료에서 데이터분석 => 기술통계법=>요약통계량 을 해보았습니다(교과서 55페이지). 먼저 엑셀 추가기능(분석도구, 분석도구-VBA)을 설치해야 합니다. (교과서 10페이지)

산술평균은 sum( )/count, 분산은 (xi-xm)^2 의 합 나누기 자료수(count( ))입니다.자료들이 평균값에서 얼마나 널리(멀리) 퍼져있나를 알아보는 값입니다. 표준편차(stdev)는 분산에 제곱근입니다. 제곱근을 하면 단위가 같아지고(자료와) 또 그 크기(scale)이 맞게 됩니다(제곱하고 다시 제곱근하니).

변동계수는 책을 읽어보세요. 표준편차를 평균으로 나눈 값(또는 곱하기 100, %)으로 평균에 대해 상대적인 분포(퍼져있음)을 나타냅니다.

중앙값(med)은 크기 순으로 놓인 자료의 가운데 값으로 자료 수가 짝수이면 중앙 순위의 두 값의 중간값으로 합니다(1 2 3 6 이면 2.5).

최빈수(mod)는 출현 빈도가 가장 놓은 자료값입니다.

상관계수는 하지 않았고 첨도, 왜도는 자료 분포가 뽀족한 정도(몰린 정도)와 대칭이 찌그러진 정도입니다.

3. 1장에는 통계(학)이란 이야기가 잠깐 나옵니다. 어렵게 정의를 내리지 않아 매우 편하고 좋습니다. 첫 페이지 표지에 나온 말입니다.

(1) 많은 자료를 간단히 요약(정리)하여 보여주는 것. 그러니까 많은 자료 => 적은 요약?, 여기엔 기술적 통계란 말을 쓰고 그래프로 하는 것과 수치로 하는 이야기가 나왔습니다.

(2) 적은 (표본)자료를 가지고 많은 모집단의 특성을 알아내는 것이 통계이다. 란 말이 나오고 따라서 여긴엔 확률과 오차 문제가 나옵니다. 2페이지에 나오는 통계학은 오차에 대한 학문이다란 말은 공감하면서 부연하여 수학이 오차를 인정하는 학문이고 인생이 바로 오차를 잘 관리하는 것이다. 병도 돈도 사랑도 결혼도 인간관계도 모두 오차가 있고 이를 잘 (통제할 수 있는 범위를 설정하고 이 범위 내에서) 관리하는 것이다.

1장 6페이지에 자료 종류 이야깁니다.

3 가지가 있고

(1) 수치형태(양적, 계량적 자료) ; 여기에도 2 종류가 있어 비율자료(곱셈, 덧셈이 가능)와 구간자료(온도, 시험점수 등으로 덧셈이 되지 않고 비율적이지 않는 자료)가 있습니다.

수치 아닌 자료로는

(2)명목형 자료와

(3)순서형자료가 있습니다.

순서형 자료는 매우좋다. 좋다. 보통이다. 나쁘다. 아주 나쁘다와 같이 순서가 있고 이를 가중치를 넣어 수치형 자료 형태로 잔화할 수도 있습니다(그럼 평균이나 표준편차 등도 구할 수 있고 자료를 분석하는데 도움이 됩니다).

2011-09-06, 곽