

APPENDIX A: AMES HOUSING DATA

SAS Data Set Name: AMESHOUSING

Data Source: Dr. Dean DeCock

Data Description: Data set contains information from the Ames, Iowa Assessor's Office used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010. Note that the original data set has 2930 observations (houses) with 82 variables.

Acknowledgement: The authors would like to acknowledge the enormous amount of time and work that Dr. Dean DeCock spent editing and compiling the Ames Housing data and documentation. The authors strongly feel that the quality and volume of data were critical in providing the examples necessary to illustrate the methods described in this book. Thank you – your work is definitely appreciated!

Variable Details: The original data set has 82 columns and includes 23 nominal, 23 ordinal, 14 discrete, and 20 continuous variables (and 2 additional observation identifiers). For this text, the authors consider a specific group of properties; the population of interest is defined as all single-family detached, residential-only houses, with sale conditions equal to 'family' or 'normal' as defined by the SAS code below. This results in a population size of 1984 houses. The authors created additional variables, as defined by the code in the table below (variable 83 through 103), resulting in a total of 103 variables. The original documentation can be found at <http://ww2.amstat.org/publications/jse/v19n3/Decock/DataDocumentation.txt>. See also: DeCock, Dean. 2011. "Ames, Iowa: Alternative to the Boston Housing Data." *Journal of Statistics Education* 19(3):1-14. <http://ww2.amstat.org/publications/jse/v19n3/decock.pdf>.

```
IF BLDG_TYPE='1Fam';
IF MS_ZONING='RH' OR MS_ZONING='RL' OR MS_ZONING='RM'
  OR MS_ZONING='RP';
IF SALE_CONDITION='Abnorml' OR SALE_CONDITION='AdjLand'
  OR SALE_CONDITION='Alloca' OR SALE_CONDITION='Partial' THEN DELETE;
```

| Variable Number | Variable | Type | Variable Description |
|-----------------|-----------------|------|---|
| 1 | Order | Num | Observation number |
| 2 | PID | Char | Parcel identification number |
| 3 | MS_SubClass | Num | Dwelling type |
| 4 | House_Style | Char | Style of dwelling |
| 5 | FirstFlr_SF | Num | First floor square footage |
| 6 | SecondFlr_SF | Num | Second floor square footage |
| 7 | Low_Qual_Fin_SF | Num | Low quality finished square feet (all floors) |
| 8 | Gr_Liv_Area | Num | Above ground living area = var5 + var6 + var7 |
| 9 | Year_Built | Num | Original construction date |
| 10 | MS_Zoning | Char | Zoning classification of the sale – all houses are zoned as Residential |
| 11 | Lot_Frontage | Num | Linear feet of street connected to property |
| 12 | Lot_Area | Num | Lot size in square feet |
| 13 | Street | Char | Type of road access to property: Grvl = Gravel, Pave = Paved |
| 14 | Alley | Char | Type of alley access to property: Grvl = Gravel, Pave = Paved, NA = No alley access |
| 15 | Lot_Shape | Char | Lot shape: Reg = Regular, IR1 = Slightly irregular, IR2 = Moderately irregular, IR3 = Irregular |
| 16 | Land_Contour | Char | Flatness of the property: Lvl = Near Flat/Level, Bnk = Banked-Quick significant rise from street grade to building, HLS = Hillside - Significant slope from side to side, Low = Depression, |

| Variable Number | Variable | Type | Variable Description |
|-----------------|----------------|------|---|
| 18 | Lot_Config | Char | Lot configuration: Inside = Inside lot, Corner = Corner lot, CulDSac = Cul-de-sac, FR2=Frontage on 2 sides of property, FR3=Frontage on 3 sides of property |
| 19 | Land_Slope | Char | Slope of property: Gtl = Gentle slope, Mod = Moderate slope, Sev = Severe slope |
| 20 | Neighborhood | Char | Physical locations within Ames city limits: Blmngtn = Bloomington Heights, Blueste = Bluestem, BrDale = Briardale, BrkSide = Brookside, ClearCr = Clear Creek, CollgCr = College Creek, Crawfor = Crawford, Edwards = Edwards, Gilbert = Gilbert, Greens = Greens, GrnHill = Green Hills, IDOTRR = Iowa DOT and Rail Road, Landmrk = Landmark, MeadowV = Meadow Village, Mitchel = Mitchell, Names = North Ames, NoRidge = Northridge, NPkVill = Northpark Villa, NridgHt = Northridge Heights, NWAmes = Northwest Ames, OldTown = Old Town, SWISU = South & West of Iowa State University, Sawyer = Sawyer, SawyerW = Sawyer West, Somerst = Somerset, StoneBr = Stone Brook, Timber = Timberland, Veenker = Veenker |
| 21 | Condition_1 | Char | Proximity to various conditions: Artery = Adjacent to arterial street, Feedr = Adjacent to feeder street Norm = Normal RRNn = Within 200' of North-South Railroad RRAn = Adjacent to North-South Railroad PosN = Near positive off-site feature--park, greenbelt, etc. PosA = Adjacent to postive off-site feature RRNe = Within 200' of East-West Railroad RRAe = Adjacent to East-West Railroad |
| 22 | Condition_2 | Char | Proximity to various conditions (if more than on is present): Artery = Adjacent to arterial street Feedr = Adjacent to feeder street Norm = Normal RRNn = Within 200' of North-South Railroad RRAn = Adjacent to North-South Railroad PosN = Near positive off-site feature--park, greenbelt, etc. PosA = Adjacent to postive off-site feature RRNe = Within 200' of East-West Railroad RRAe = Adjacent to East-West Railroad |
| 23 | Bldg_Type | Char | Type of dwelling: 1Fam = Single-family detached (this book omits duplexes, townhouses, two-family dwellings) |
| 24 | Overall_Qual | Num | Rating of the overall material and finish of the house: 10 = Very Excellent, 9 = Excellent, 8 = Very Good, 7 = Good, 6 = Above Average, 5 = Average, 4 = Below Average, 3 = Fair, 2 = Poor, 1 = Very Poor |
| 25 | Overall_Cond | Num | Rating of the overall condition of the house: 10 = Very Excellent, 9 = Excellent, 8 = Very Good, 7 = Good, 6 = Above Average, 5 = Average, 4 = Below Average, 3 = Fair, 2 = Poor, 1 = Very Poor |
| 26 | Year_Remod_Add | Num | Year of remodel or addition (same as Year_Built if no remodel or addition) |

| Variable Number | Variable | Type | Variable Description |
|-----------------|---------------|------|---|
| 27 | Roof_Style | Char | Type of roof: Flat = Flat, Gable = Gable, Gambrel = Gabrel (Barn), Hip = Hip, Mansard = Mansard, Shed = Shed |
| 28 | Roof_Matl | Char | Roof material: ClyTile = Clay or Tile, CompShg = Standard (Composite) Shingle, Membran = Membrane, Metal = Metal, Roll = Roll, Tar&Grv = Gravel & Tar, WdShake = Wood Shakes, WdShngl = Wood Shingles |
| 29 | Exterior_1st | Char | Exterior covering on house: AsbShng = Asbestos Shingles, AsphShn = Asphalt Shingles, BrkComm = Brick Common, BrkFace = Brick Face, CBlock = Cinder Block, CemntBd = Cement Board, HdBoard = Hard Board, ImStucc = Imitation Stucco, MetalSd = Metal Siding, Other = Other, Plywood = Plywood, PreCast = PreCast, Stone = Stone, Stucco = Stucco, VinylSd = Vinyl Siding, Wd Sdng = Wood Siding, WdShng = Wood Shingles |
| 30 | Exterior_2nd | Char | Exterior covering on house (if more than one material): AsbShng = Asbestos Shingles, AsphShn = Asphalt Shingles, BrkComm = Brick Common, BrkFace = Brick Face, CBlock = Cinder Block, CemntBd = Cement Board, HdBoard = Hard Board, ImStucc = Imitation Stucco, MetalSd = Metal Siding, Other = Other, Plywood = Plywood, PreCast = PreCast, Stone = Stone, Stucco = Stucco, VinylSd = Vinyl Siding, Wd Sdng = Wood Siding, WdShng = Wood Shingles |
| 31 | Mas_Vnr_Type | Char | Masonry veneer type: BrkCmn = Brick Common, BrkFace = Brick Face, CBlock = Cinder Block, None = None, Stone = Stone |
| 32 | Mas_Vnr_Area | Num | Masonry veneer area in square feet |
| 33 | Exter_Qual | Char | Evaluates the quality of the material on the exterior: Ex = Excellent, Gd = Good, TA = Typical, Fa = Fair, Po = Poor |
| 34 | Exter_Cond | Char | Evaluates the present condition of the material on the exterior: Ex = Excellent, Gd = Good, TA = Typical, Fa = Fair, Po = Poor |
| 35 | Foundation | Char | Type of foundation: BrkTil = Brick & Tile, CBlock = Cinder Block, PConc = Poured Concrete, Slab = Slab, Stone = Stone, Wood = Wood |
| 36 | Bsmt_Qual | Char | Evaluates the height of the basement: Ex = Excellent (100+ inches), Gd = Good (90-99 inches), TA = Typical (80-89 inches), Fa = Fair (70-79 inches), Po = Poor (<70 inches), NA = No Basement |
| 37 | Bsmt_Cond | Char | Evaluates the general condition of the basement: Ex = Excellent Gd = Good TA = Typical (slight dampness allowed) Fa = Fair (dampness or some cracking or settling) Po = Poor (severe cracking, settling, or wetness) NA = No Basement |
| 38 | Bsmt_Exposure | Char | Refers to walkout or garden level walls: Gd = Good Exposure, Av = Average Exposure, Mn = Minimum Exposure, No = No Exposure, NA = No Basement |

| Variable Number | Variable | Type | Variable Description |
|-----------------|----------------|------|---|
| 39 | BsmtFin_Type_1 | Char | Rating of Type I basement finished area: GLQ = Good Living Quarters ALQ = Average Living Quarters BLQ = Below Average Living Quarters Rec = Average Rec Room LwQ = Low Quality Unf = Unfinished NA = No Basement |
| 40 | BsmtFin_SF_1 | Num | Type I basement finished square feet |
| 41 | BsmtFin_Type_2 | Char | Rating of Type 2 basement finished area (if multiple types): GLQ = Good Living Quarters, ALQ = Average Living Quarters BLQ = Below Average Living Quarters, Rec = Average Rec Room, LwQ = Low Quality, Unf = Unfinished, NA = No Basement |
| 42 | BsmtFin_SF_2 | Num | Type 2 basement finished square feet |
| 43 | Bsmt_Unf_SF | Num | Unfinished basement square feet |
| 44 | Total_Bsmt_SF | Num | Total square feet of all basement area = var40 + var42 + var43 |
| 45 | Heating | Char | Type of heating: Floor = Floor Furnace, GasA = Gas forced warm air furnace GasW = Gas hot water or steam heat, Grav = Gravity furnace, OthW = Hot water or steam heat other than gas, Wall = Wall furnace |
| 46 | Heating_QC | Char | Heating quality and condition: Ex = Excellent, Gd = Good, TA = Average/Typical, Fa = Fair, Po = Poor |
| 47 | Central_Air | Char | Central air conditioning: N = No, Y = Yes |
| 48 | Electrical | Char | Electrical system: SBkr = Standard Circuit Breakers & Romex FuseA = Fuse Box over 60 AMP and all Romex wiring (Average) FuseF = 60 AMP Fuse Box and mostly Romex wiring (Fair) FuseP = 60 AMP Fuse Box and mostly knob & tube wiring (poor) Mix = Mixed |
| 49 | Bsmt_Full_Bath | Num | Number of full bathrooms in basement |
| 50 | Bsmt_Half_Bath | Num | Number of half bathrooms in basement |
| 51 | Full_Bath | Num | Number of full bathrooms above ground |
| 52 | Half_Bath | Num | Number of half bathrooms above ground |
| 53 | Bedroom_AbvGr | Num | Number of bedrooms above ground |
| 54 | Kitchen_AbvGr | Num | Number of kitchens above ground |
| 55 | Kitchen_Qual | Char | Kitchen quality: Ex = Excellent, Gd = Good, TA = Average/Typical, Fa = Fair, Po = Poor |
| 56 | TotRms_AbvGrd | Num | Total number of rooms above ground (does not include bathrooms) |
| 57 | Functional | Char | Home functionality (assume typical if deductions are warranted): Typ = Typical Functionality, Min1 = Minor Deductions 1, Min2 = Minor Deductions 2, Mod = Moderate Deductions, Maj1 = Major Deductions 1, Maj2 = Major Deductions 2, Sev = Severely Damaged, Sal = Salvage only |
| 58 | Fireplaces | Num | Number of fireplaces |

| Variable Number | Variable | Type | Variable Description |
|-----------------|----------------|------|--|
| 59 | Fireplace_Qu | Char | Fireplace quality: Ex = Excellent - Exceptional Masonry Fireplace Gd = Good - Masonry Fireplace in main level TA = Average - Prefabricated Fireplace in main living area or Masonry Fireplace in basement Fa = Fair - Prefabricated Fireplace in basement Po = Poor - Ben Franklin Stove NA = No Fireplace |
| 60 | Garage_Type | Char | Garage location: 2Types = More than one type of garage, Attchd = Attached to home, Basment = Basement Garage, BuiltIn = Built-In (Garage part of house), CarPort = Car Port, Detchd = Detached from home, NA = No Garage |
| 61 | Garage_Yr_Blt | Num | Year garage was built |
| 62 | Garage_Finish | Char | Interior finish of the garage: Fin = Finished, RFn = Rough Finished, Unf = Unfinished, NA = No Garage |
| 63 | Garage_Cars | Num | Size of garage in car capacity |
| 64 | Garage_Area | Num | Size of garage in square feet |
| 65 | Garage_Qual | Char | Garage quality: Ex = Excellent, Gd = Good, TA = Typical/Average, Fa = Fair, Po = Poor, NA = No Garage |
| 66 | Garage_Cond | Char | Garage condition: Ex = Excellent, Gd = Good, TA = Typical/Average, Fa = Fair, Po = Poor, NA = No Garage |
| 67 | Paved_Drive | Char | Paved driveway: Y = Paved, P = Partial Pavement, N = Dirt/Gravel |
| 68 | Wood_Deck_SF | Num | Wood deck area in square feet |
| 69 | Open_Porch_SF | Num | Open porch area in square feet |
| 70 | Enclosed_Porch | Num | Enclosed porch area in square feet |
| 71 | ThreeSsn_Porch | Num | Three season porch area in square feet |
| 72 | Screen_Porch | Num | Screen porch area in square feet |
| 73 | Pool_Area | Num | Pool area in square feet |
| 74 | Pool_QC | Char | Pool quality: Ex = Excellent, Gd = Good, TA = Typical/Average, Fa = Fair, NA = No Garage |
| 75 | Fence | Char | Fence quality: GdPrv = Good Privacy, MnPrv = Minimum Privacy, GdWo = Good Wood, MnWw = Minimum Wood/Wire, NA = No Fence |
| 76 | Misc_Feature | Char | Miscellaneous feature not covered in other categories: Elev = Elevator, Gar2 = 2nd Garage (if not described in garage section), Othr = Other, Shed = Shed (over 100 SF), TenC = Tennis Court, NA = None |
| 77 | Misc_Val | Num | Value of the miscellaneous feature |
| 78 | Mo_Sold | Num | Month sold |
| 79 | Yr_Sold | Num | Year sold |

| Variable Number | Variable | Type | Variable Description |
|-----------------|-------------------|------|--|
| 80 | Sale_Type | Char | Type of sale: WD = Warranty Deed – Conventional, CWD = Warranty Deed – Cash, VWD = Warranty Deed - VA Loan, New = Home just constructed and sold, COD = Court Officer Deed/Estate, Con = Contract 15% Down payment regular terms, ConLw = Contract Low Down payment and low interest, ConLI = Contract Low Interest, ConLD = Contract Low Down, Oth = Other |
| 81 | Sale_Condition | Char | Condition of sale: This book uses only Family = Sale between family members Normal = Normal Sale (this book omits houses with sale condition equal to trade, foreclosure, short sale, adjoining land purchase, two linked properties with separate deeds, home not completed or under construction) |
| 82 | SalePrice | Num | Sale price (in \$\$) |
| 83 | Bonus | Num | Bonus=0; if SalePrice > 175000 then Bonus=1; |
| 84 | Bsmt Fin_SF | Num | Bsmt Fin_SF = BsmtFin_SF_1 + BsmtFin_SF_2; |
| 85 | Age_at_Sale | Num | Age at Sale = Yr Sold-Year Built; |
| 86 | Fullbath_2plus | Num | fullbath_2plus=0; if full_bath=. then fullbath_2plus=.; if full_bath ge 2 then fullbath_2plus=1; |
| 87 | Overall_Quality | Num | Overall_Quality=1; if Overall_Qual=. then Overall_Quality=.; if Overall_Qual=5 then Overall_Quality=2; if Overall_Qual ge 6 then Overall_Quality=3; |
| 88 | Overall_Condition | Num | Overall_Condition=1; if Overall_Cond=. then Overall_Condition=.; if Overall_Cond=5 then Overall_Condition=2; if Overall_Cond ge 6 then Overall_Condition=3; |
| 89 | TwoPlusCar_Garage | Num | TwoPlusCar_Garage=0; if Garage_Cars=. then TwoPlusCar_Garage=.; if Garage_Cars ge 2 then TwoPlusCar_Garage=1; |
| 90 | Poured_Concrete | Num | Poured_Concrete=0; if foundation=" then Poured_Concrete=.; if foundation='PConc' then Poured_Concrete=1; |
| 91 | One_Floor | Num | One_Floor=0; if house_style=" then One_Floor=.; if house_style='1Story' then One_Floor=1; |
| 92 | Fireplace_1plus | Num | Fireplace_1plus=0; if fireplaces=. then Fireplace_1plus=.; if fireplaces ge 1 then Fireplace_1plus=1; |
| 93 | Has_Fence | Num | Has_Fence=1; if fence=" then Has_Fence=.; if fence='NA' then Has_Fence=0; |
| 94 | Land_Level | Num | Land_Level=0; if Land_Contour=" then Land_Level=.; if Land_Contour='Lvl' then Land_Level=1; |
| 95 | CuldeSac | Num | CuldeSac=0; if Lot_Config=" then CuldeSac=.; if Lot_Config='CulDSac' then CuldeSac=1; |

| Variable Number | Variable | Type | Variable Description |
|-----------------|----------------------|------|---|
| 96 | Vinyl_Siding | Num | Vinyl_Siding=0; if exterior_1st="" then Vinyl_Siding=.; if exterior_1st='VinylSd' then Vinyl_Siding=1; |
| 97 | Paved_Driveway | Num | Paved_Driveway=0; if Paved_Drive="" then Paved_Driveway=.; if Paved_Drive='Y' then Paved_Driveway=1; |
| 98 | AbvGr_BR | Num | AbvGr_BR=3; if Bedroom_AbvGr=. then AbvGr_BR=.; if Bedroom_AbvGr=1 or Bedroom_AbvGr=2 then AbvGr_BR=1; if Bedroom_AbvGr=3 then AbvGr_BR=2; |
| 99 | Normal_Prox_Cond | Num | Normal_Prox_Cond=0; if condition_1="" then Normal_Prox_Cond=.; if condition_1='Norm' then Normal_Prox_Cond=1; |
| 100 | Total_Functionality | Num | Total_Functionality=0; if functional="" then Total_Functionality=.; if functional='Typ' then Total_Functionality=1; |
| 101 | High_Exterior_Qual | Num | High_Exterior_Qual=1; if Exter_Qual="" then High_Exterior_Qual=.; if Exter_Qual='TA' or Exter_Qual='Fa' or Exter_Qual='Po' then High_Exterior_Qual=0; |
| 102 | High_Kitchen_Quality | Num | High_Kitchen_Quality=1; if Kitchen_Qual="" then High_Kitchen_Quality=.; if Kitchen_Qual='TA' or Kitchen_Qual='Fa' or Kitchen_Qual='Po' then High_Kitchen_Quality=0; |
| 103 | High_Exterior_Cond | Num | High_Exterior_Cond=1; if Exter_Cond="" then High_Exterior_Cond=.; if Exter_Cond='TA' or Exter_Cond='Fa' or Exter_Cond='Po' then High_Exterior_Cond=0; |

APPENDIX B: DIABETIC HEALTH CARE DATA

Data Set Name: DIABETICS

Data Description: The data file provided with this book, DIABETICS, contains demographic, clinical, and geo-location data for 63,108 patients who have been diagnosed with diabetes. The observation of study is at the patient level, each having a total of 125 variables that fall into the following categories:

1. Demographic information, such as patient ID, gender, age, and age range
2. Date of the last doctor's visit and the general state of the patient, including height, weight, BMI, systolic and diastolic blood pressure, type of diabetes, if their diabetes is controlled, medical risk, if the patient has hypertension, hyperlipidemia, peripheral vascular disease (PVD), renal disease, and if the patient has suffered a stroke.
3. The results of 57 laboratory tests, including those tests from the comprehensive metabolic panel (CMP) which are used to evaluate the how the organs function and to detect various chronic diseases.
4. Information related to prescription medicine, including type of medication, dosage form, and the number and nature of adverse events with duration dates.
5. Geo-Location data including City and State where the patients resides, along with longitude and latitude.

The complete data dictionary with detailed descriptions is as follows:

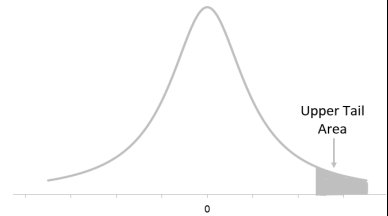
| Variable Number | Variable | Type | Variable Description |
|-----------------|-----------------|------|---|
| 1 | Patient_ID | Num | Patient ID |
| 2 | GENDER | Char | Gender |
| 3 | AGE | Num | Age |
| 4 | AGE_RANGE | Char | Age Range |
| 5 | DATE | Num | Date of Last Doctor Visit |
| 6 | DATE_YEAR | Num | Year of Last Doctor Visit |
| 7 | QUARTER | Num | Quarter of Last Doctor Visit |
| 8 | DATE_MONTH_NBR | Num | Month of Last Doctor Visit (values = 1 - 12) |
| 9 | DATE_MONTH_NAME | Char | Month of Last Doctor Visit |
| 10 | WEEKDAY_NBR | Num | Weekday of Last Doctor Visit (values = 1 - 7) |
| 11 | WEEKDAY_NAME | Char | Day of Week |
| 12 | AE_STARTDT | Num | Adverse Event Start Date |
| 13 | AE_STOPDT | Num | Adverse Event Stop Date |
| 14 | AE_DURATION | Num | Duration of Adverse Event |
| 15 | TYPE | Char | Type of Diabetes: Type 1, Type 2, Secondary, N/A |
| 16 | Type_1 | Num | Type 1 Diabetes: 1=Type 1, 0=Other |
| 17 | Type_2 | Num | Type 2 Diabetes: 1=Type 2, 0=Other |
| 18 | Secondary | Num | Secondary Diabetes: 1=Secondary, 0=Other |

| Variable Number | Variable | Type | Variable Description |
|-----------------|-----------------------|------|---|
| 19 | PRIMARY_MED | Char | Primary Medication: AG Inhibitor, Amylin Mimetic, Biguanide, DPP-4 Inhibitor, Incretin Mimetic, Meglitinide, Sulfonylurea, Thiazolidinedione |
| 20 | PRIME_DOSAGE_FORM | Char | Form of Primary Dosage: Injectable, Oral |
| 21 | SECONDARY_MED | Char | Primary Medication: Alpha-glucosidase inhibitor, Biguanide, DPP-4 Inhibitor, Meglitinide, Sulfonylurea, Thiazolidinedione |
| 22 | SECONDARY_DOSAGE_FORM | Char | Form of Secondary Dosage: Oral |
| 23 | HYPERTENSION | Num | Does patient have hypertension: 1=Yes, 0=No |
| 24 | HYPERLIPIDEMIA | Num | Does patient have high blood pressure: 1=Yes, 0=No |
| 25 | PVD | Num | Does patient have peripheral vascular disease: 1=Yes, 0=No |
| 26 | STROKE | Num | Has the patient had a stroke: 1=Yes, 0=No |
| 27 | Diabetes_Med_Risk | Num | Omit |
| 28 | RENAL_DISEASE | Num | Does patient have renal disease: 1=Yes, 0=No |
| 29 | VISIT_NBR | Num | Visit Number = 1 for all patients |
| 30 | WEIGHT | Num | Weight |
| 31 | BMI | Num | Body-Mass-Index |
| 32 | SYST_BP | Num | Systolic Blood Pressure |
| 33 | DIAST_BP | Num | Diastolic Blood Pressure |
| 34 | EFFICACY | Char | Measure of efficacy: Good, Moderate, Poor, None |
| 35 | NAES | Num | Number of adverse events |
| 36 | SEVERITY | Char | Severity of adverse event: Mild, Moderate, Severe |
| 37 | AE1 | Char | Adverse event 1: ABDOMINAL PAIN, CHEST PAIN, DIZZINESS, HALLUCINATIONS, HEADACHE, ITCHING, NAUSEA, SKIN RASH, TINNITUS, VOMITING |
| 38 | AE2 | Char | Adverse event 2: ABDOMINAL PAIN, DIZZINESS, HALLUCINATIONS, HEADACHE, ITCHING, NAUSEA, PALPITATIONS, SKIN RASH, TINNITUS, VOMITING |
| 39 | AE3 | Char | Adverse event 3: ABDOMINAL PAIN, DIZZINESS, HALLUCINATIONS, HEADACHE, ITCHING, NAUSEA, PALPITATIONS, SKIN RASH, TINNITUS, VOMITING |
| 40 | Acetoacetate | Num | Lab test |
| 41 | Alanine | Num | Lab test |
| 42 | Alcohol | Num | Lab test |
| 43 | Ammonia | Num | Lab test |

| Variable Number | Variable | Type | Variable Description |
|-----------------|-------------------------|------|----------------------|
| 44 | Amylase | Num | Lab test |
| 45 | Ascorbic_Acid | Num | Lab test |
| 46 | Aspartate | Num | Lab test |
| 47 | Bicarbonate | Num | Lab test |
| 48 | Blood_Urea_Nitrogen | Num | Lab test |
| 49 | BUN_Creatinine_Ratio | Num | Lab test |
| 50 | Blood_Volume | Num | Lab test |
| 51 | Calcium | Num | Lab test |
| 52 | Carbon_Dioxide_Pressure | Num | Lab test |
| 53 | Carbon_Monoxide | Num | Lab test |
| 54 | CD4_Cell_Count | Num | Lab test |
| 55 | Ceruloplasmin | Num | Lab test |
| 56 | Chloride | Num | Lab test |
| 57 | Cholesterol | Num | Lab test |
| 58 | Copper | Num | Lab test |
| 59 | Creatine_Kinase | Num | Lab test |
| 60 | Erythrocyte | Num | Lab test |
| 61 | Glucose | Num | Lab test |
| 62 | Hemoglobin_A1c | Num | Lab test |
| 63 | Hydroxyvitamin_25_D | Num | Lab test |
| 64 | Iron | Num | Lab test |
| 65 | Iron_Binding_Capacity | Num | Lab test |
| 66 | Lactate | Num | Lab test |
| 67 | Lactic_Dehydrogenase | Num | Lab test |
| 68 | Lead | Num | Lab test |
| 69 | Lipase | Num | Lab test |
| 70 | Magnesium | Num | Lab test |
| 71 | MCH | Num | Lab test |
| 72 | MCHC | Num | Lab test |
| 73 | MCV | Num | Lab test |
| 74 | Osmolality | Num | Lab test |
| 75 | Oxygen_Pressure | Num | Lab test |
| 76 | Oxygen_Saturation | Num | Lab test |
| 77 | Phosphatase_Prostatic | Num | Lab test |

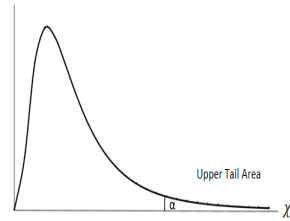
| Variable Number | Variable | Type | Variable Description |
|-----------------|-----------------------------|------|---|
| 78 | Phosphorus | Num | Lab test |
| 79 | Platelet_Count | Num | Lab test |
| 80 | Potassium | Num | Lab test |
| 81 | Prostate_Specific_Antigen | Num | Lab test |
| 82 | Prothrombin | Num | Lab test |
| 83 | Pyruvic_Acid | Num | Lab test |
| 84 | Red_Blood_Cell_Count | Num | Lab test |
| 85 | Sodium_Na | Num | Lab test |
| 86 | Total_Protien | Num | Lab test |
| 87 | TSH | Num | Lab test |
| 88 | Uric_Acid | Num | Lab test |
| 89 | Vitamin_A | Num | Lab test |
| 90 | White_Blood_Cell_Count | Num | Lab test |
| 91 | Zinc_B_Zn | Num | Lab test |
| 92 | Specific_Gravity | Num | Lab test |
| 93 | Urine_PH | Num | Lab test |
| 94 | Protien | Num | Lab test |
| 95 | Ketones | Num | Lab test |
| 96 | Bilirubin_Total | Num | Lab test |
| 97 | Alpha_glucosidase_inhibitor | Num | 1=primary med is alpha glucosidase inhibitor 0=otherwise |
| 98 | Amylin_Mimetic | Num | 1=primary med is amylin mimetic 0=otherwise |
| 99 | Biguanide | Num | 1=primary med is biguanide 0=otherwise |
| 100 | DPP_4_Inhibitor | Num | 1=primary med is DPP 4 inhibitor 0=otherwise |
| 101 | Incretin_Mimetic | Num | 1=primary med is incretin mimetic 0=otherwise |
| 102 | Meglitinide | Num | 1=primary med is meglitinide 0=otherwise |
| 103 | Sulfonylurea | Num | 1=primary med is sulfonylurea 0=otherwise |
| 104 | Thiazolidinedione | Num | 1=primary med is thiazolidinedione 0=otherwise |
| 105 | ABDOMINAL_PAIN | Num | 1=adverse event is abdominal pain 0=otherwise |
| 106 | DIZZINESS | Num | 1=adverse event is dizziness 0=otherwise |

| Variable Number | Variable | Type | Variable Description |
|-----------------|---------------------|------|--|
| 107 | HALLUCINATIONS | Num | 1=adverse event is hallucinations 0=otherwise |
| 108 | HEADACHE | Num | 1=adverse event is headache 0=otherwise |
| 109 | ITCHING | Num | 1=adverse event is itching 0=otherwise |
| 110 | NAUSEA | Num | 1=adverse event is nausea 0=otherwise |
| 111 | PALPITATIONS | Num | 1=adverse event is palpitations 0=otherwise |
| 112 | SKIN_RASH | Num | 1=adverse event is skin rash 0=otherwise |
| 113 | TINNITUS | Num | 1=adverse event is tinnitus 0=otherwise |
| 114 | VOMITING | Num | 1=adverse event is vomiting 0=otherwise |
| 115 | MILD | Num | Severity of adverse event: 1=Mild, 0=Moderate, Severe |
| 116 | MODERATE | Num | Severity of adverse event: 1=Moderate, 0=Mild, Severe |
| 117 | Chest_Pain | Num | omit |
| 118 | State | Char | State |
| 119 | City | Char | City |
| 120 | State_Longitude | Num | State Longitude |
| 121 | State_Latitude | Num | State Latitude |
| 122 | City_Longitude | Num | City Longitude |
| 123 | City_Latitude | Num | City Latitude |
| 124 | CONTROLLED_DIABETIC | Num | if hemoglobin A1C lt 7 then CONTROLLED_DIABETIC = 1; if hemoglobin A1C ge 7 then CONTROLLED_DIABETIC = 0; |
| 125 | INCHES | Num | Height in inches |



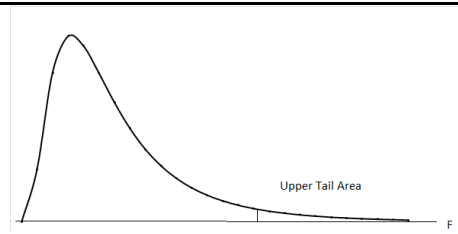
Appendix D: t-Distribution Upper Tail Area

| df | Upper-Tail Area | | | | |
|----|-----------------|-------|--------|--------|--------|
| | 0.10 | 0.05 | 0.025 | 0.01 | 0.01 |
| 1 | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 |
| 2 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 |
| 3 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 |
| 4 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 |
| 5 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 |
| 6 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 |
| 7 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 |
| 8 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 |
| 9 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 |
| 10 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 |
| 11 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 |
| 12 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 |
| 13 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 |
| 14 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 |
| 15 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 |
| 16 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 |
| 17 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 |
| 18 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 |
| 19 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 |
| 20 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 |
| 21 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 |
| 22 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 |
| 23 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 |
| 24 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 |
| 25 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 |
| 26 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 |
| 27 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 |
| 28 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 |
| 29 | 1.311 | 1.699 | 2.045 | 2.462 | 2.756 |
| 30 | 1.310 | 1.697 | 2.042 | 2.457 | 2.750 |



Appendix E: Chi-Square Distribution Upper Tail Area

| df | Upper-Tail Area | | | | | | | | | |
|----|-----------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 0.995 | 0.990 | 0.975 | 0.950 | 0.900 | 0.100 | 0.050 | 0.025 | 0.010 | 0.005 |
| 1 | 0.000 | 0.000 | 0.001 | 0.004 | 0.016 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | 0.010 | 0.020 | 0.051 | 0.103 | 0.211 | 4.605 | 5.991 | 7.378 | 9.210 | 10.597 |
| 3 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.064 | 7.779 | 9.488 | 11.143 | 13.277 | 14.860 |
| 5 | 0.412 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.070 | 12.833 | 15.086 | 16.750 |
| 6 | 0.676 | 0.872 | 1.237 | 1.635 | 2.204 | 10.645 | 12.592 | 14.449 | 16.812 | 18.548 |
| 7 | 0.989 | 1.239 | 1.690 | 2.167 | 2.833 | 12.017 | 14.067 | 16.013 | 18.475 | 20.278 |
| 8 | 1.344 | 1.646 | 2.180 | 2.733 | 3.490 | 13.362 | 15.507 | 17.535 | 20.090 | 21.955 |
| 9 | 1.735 | 2.088 | 2.700 | 3.325 | 4.168 | 14.684 | 16.919 | 19.023 | 21.666 | 23.589 |
| 10 | 2.156 | 2.558 | 3.247 | 3.940 | 4.865 | 15.987 | 18.307 | 20.483 | 23.209 | 25.188 |
| 11 | 2.603 | 3.053 | 3.816 | 4.575 | 5.578 | 17.275 | 19.675 | 21.920 | 24.725 | 26.757 |
| 12 | 3.074 | 3.571 | 4.404 | 5.226 | 6.304 | 18.549 | 21.026 | 23.337 | 26.217 | 28.300 |
| 13 | 3.565 | 4.107 | 5.009 | 5.892 | 7.042 | 19.812 | 22.362 | 24.736 | 27.688 | 29.819 |
| 14 | 4.075 | 4.660 | 5.629 | 6.571 | 7.790 | 21.064 | 23.685 | 26.119 | 29.141 | 31.319 |
| 15 | 4.601 | 5.229 | 6.262 | 7.261 | 8.547 | 22.307 | 24.996 | 27.488 | 30.578 | 32.801 |
| 16 | 5.142 | 5.812 | 6.908 | 7.962 | 9.312 | 23.542 | 26.296 | 28.845 | 32.000 | 34.267 |
| 17 | 5.697 | 6.408 | 7.564 | 8.672 | 10.085 | 24.769 | 27.587 | 30.191 | 33.409 | 35.718 |
| 18 | 6.265 | 7.015 | 8.231 | 9.390 | 10.865 | 25.989 | 28.869 | 31.526 | 34.805 | 37.156 |
| 19 | 6.844 | 7.633 | 8.907 | 10.117 | 11.651 | 27.204 | 30.144 | 32.852 | 36.191 | 38.582 |
| 20 | 7.434 | 8.260 | 9.591 | 10.851 | 12.443 | 28.412 | 31.410 | 34.170 | 37.566 | 39.997 |
| 21 | 8.034 | 8.897 | 10.283 | 11.591 | 13.240 | 29.615 | 32.671 | 35.479 | 38.932 | 41.401 |
| 22 | 8.643 | 9.542 | 10.982 | 12.338 | 14.041 | 30.813 | 33.924 | 36.781 | 40.289 | 42.796 |
| 23 | 9.260 | 10.196 | 11.689 | 13.091 | 14.848 | 32.007 | 35.172 | 38.076 | 41.638 | 44.181 |
| 24 | 9.886 | 10.856 | 12.401 | 13.848 | 15.659 | 33.196 | 36.415 | 39.364 | 42.980 | 45.559 |
| 25 | 10.520 | 11.524 | 13.120 | 14.611 | 16.473 | 34.382 | 37.652 | 40.646 | 44.314 | 46.928 |
| 26 | 11.160 | 12.198 | 13.844 | 15.379 | 17.292 | 35.563 | 38.885 | 41.923 | 45.642 | 48.290 |
| 27 | 11.808 | 12.879 | 14.573 | 16.151 | 18.114 | 36.741 | 40.113 | 43.195 | 46.963 | 49.645 |
| 28 | 12.461 | 13.565 | 15.308 | 16.928 | 18.939 | 37.916 | 41.337 | 44.461 | 48.278 | 50.993 |
| 29 | 13.121 | 14.256 | 16.047 | 17.708 | 19.768 | 39.087 | 42.557 | 45.722 | 49.588 | 52.336 |
| 30 | 13.787 | 14.953 | 16.791 | 18.493 | 20.599 | 40.256 | 43.773 | 46.979 | 50.892 | 53.672 |



Appendix F: F Distribution Upper Tail Area

| denominator | | numerator degrees of freedom | | | | | | | | | |
|--------------------|----------|------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| degrees of freedom | α | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 0.10 | 39.86 | 49.50 | 53.59 | 55.83 | 57.24 | 58.20 | 58.91 | 59.44 | 59.86 | 60.19 |
| | 0.05 | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 | 238.88 | 240.54 | 241.88 |
| | 0.01 | 4052.18 | 4999.50 | 5403.35 | 5624.58 | 5763.65 | 5858.99 | 5928.36 | 5981.07 | 6022.47 | 6055.85 |
| 2 | 0.10 | 8.53 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.35 | 9.37 | 9.38 | 9.39 |
| | 0.05 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 |
| | 0.01 | 98.50 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 | 99.39 | 99.40 |
| 3 | 0.10 | 5.54 | 5.46 | 5.39 | 5.34 | 5.31 | 5.28 | 5.27 | 5.25 | 5.24 | 5.23 |
| | 0.05 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 |
| | 0.01 | 34.12 | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.35 | 27.23 |
| 4 | 0.10 | 4.54 | 4.32 | 4.19 | 4.11 | 4.05 | 4.01 | 3.98 | 3.95 | 3.94 | 3.92 |
| | 0.05 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 |
| | 0.01 | 21.20 | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.55 |
| 5 | 0.10 | 4.06 | 3.78 | 3.62 | 3.52 | 3.45 | 3.40 | 3.37 | 3.34 | 3.32 | 3.30 |
| | 0.05 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 |
| | 0.01 | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 | 10.16 | 10.05 |
| 6 | 0.10 | 3.78 | 3.46 | 3.29 | 3.18 | 3.11 | 3.05 | 3.01 | 2.98 | 2.96 | 2.94 |
| | 0.05 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 |
| | 0.01 | 13.75 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 |
| 7 | 0.10 | 3.59 | 3.26 | 3.07 | 2.96 | 2.88 | 2.83 | 2.78 | 2.75 | 2.72 | 2.70 |
| | 0.05 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 |
| | 0.01 | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 | 6.62 |
| 8 | 0.10 | 3.46 | 3.11 | 2.92 | 2.81 | 2.73 | 2.67 | 2.62 | 2.59 | 2.56 | 2.54 |
| | 0.05 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 |
| | 0.01 | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 |
| 9 | 0.10 | 3.36 | 3.01 | 2.81 | 2.69 | 2.61 | 2.55 | 2.51 | 2.47 | 2.44 | 2.42 |
| | 0.05 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 |
| | 0.01 | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 |
| 10 | 0.10 | 3.29 | 2.92 | 2.73 | 2.61 | 2.52 | 2.46 | 2.41 | 2.38 | 2.35 | 2.32 |
| | 0.05 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 |
| | 0.01 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 |
| 11 | 0.10 | 3.23 | 2.86 | 2.66 | 2.54 | 2.45 | 2.39 | 2.34 | 2.30 | 2.27 | 2.25 |
| | 0.05 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 |
| | 0.01 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 |
| 12 | 0.10 | 3.18 | 2.81 | 2.61 | 2.48 | 2.39 | 2.33 | 2.28 | 2.24 | 2.21 | 2.19 |
| | 0.05 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 |
| | 0.01 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 |
| 13 | 0.10 | 3.14 | 2.76 | 2.56 | 2.43 | 2.35 | 2.28 | 2.23 | 2.20 | 2.16 | 2.14 |
| | 0.05 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 |
| | 0.01 | 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 | 4.10 |
| 14 | 0.10 | 3.10 | 2.73 | 2.52 | 2.39 | 2.31 | 2.24 | 2.19 | 2.15 | 2.12 | 2.10 |
| | 0.05 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 |
| | 0.01 | 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 |

NOTE: Additional F-values can be found in Excel Using F.INV.RT(alpha,numerator df,denominator df)

APPENDIX G: QUIZ ANSWER KEYS

| CHAPTER | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 |
|---|----|----|----|----|----|----|----|----|----|-----|
| Ch 2: Summarizing Your Data with Descriptive Statistics | B | C | A | D | D | B | A | C | C | D |
| Ch 3: Data Visualization | C | B | D | A | B | C | A | D | E | C |
| Ch 4: The Normal Distribution and Intro to Inferential Statistics | D | C | B | B | A | D | C | B | A | D |
| Ch 5: Analysis of Categorical Variables | C | B | C | D | E | A | B | A | F | E |
| Ch 6: Two-Sample t-Test | B | A | A | D | B | D | D | C | A | C |
| Ch 7: Analysis of Variance (ANOVA) | D | D | B | A | C | B | A | B | A | B |
| Ch 8: Preparing the Input Variables for Prediction | B | C | D | B | A | C | D | B | A | C |
| Ch 9: Linear Regression Analysis | D | C | A | D | E | B | C | A | B | B |
| Ch 10: Logistic Regression Analysis | B | D | A | C | B | D | B | B | C | A |
| Ch 11: Measure of Model Performance | B | B | D | A | C | C | A | A | D | C |

APPENDIX H: How to save the accompanying data sets on your desktop/laptop

There are 15 SAS data sets provided with this book. All programs referenced in this book are also provided. Each SAS program reads a SAS data set using a LIBNAME statement. In order for your SAS programs to run with no modification, create the following folders on your C: drive, and save the data set in the appropriate folders:

| FOLDER | SAS Data Set |
|---------------|---|
| C:\SASBA\AMES | ameshousing ames300 ames300miss amesreg300 ames70 ames30 amesnew alt40 |
| C:\SASBA\HC | diabetics diab200 diab25f |
| C:\SASBA\DATA | all sunglasses cas revenue |