

1. (a) variables: width of the last upper molar is numerical-continuous; observational unit: specimen; sample size: 36.
- (b) variables: birth weight is numerical-continuous, date of birth is numerical-discrete, mother's race is categorical-nominal; observational unit: baby; sample size: 65.
- (c) variables: height is numerical-continuous, weight is numerical-continuous; observational unit: child; sample size: 37.
- (d) variables: blood type is categorical-nominal, cholesterol level is numerical-continuous; observational unit: person who donated blood; sample size: 129.
- (e) variables: number of leaves is numerical-discrete; observational unit is plant; sample size is 25.
- (f) variables: number of seizures during an eight-week period is numerical-discrete; observational unit is patient; sample size is 20.

2. (a) *Describe the sampling units (objects that are sampled) in this sample.*

Children in motor vehicle crashes from the PCPS database.

*What is the sample size?*

11859 children.

*What is the primary response variable associated with each sampling unit?*

Was the child injured?

*What are some other variables that are measured?*

Some (not all) other variables measured include: "driver age, gender, and restraint use; driver relationship to the child occupant; occurrence of driver injury; child age and gender; child-restraint use and seat position; vehicle type, model year, and weight; crash characteristics, including direction of initial impact and rollover occurrence, single or multivehicle crash, and severity on the basis of reported intrusion and/or whether the vehicle was towed from the scene." (pp. 290-291).

- (b) *Provide a description of a population for which it is reasonable to generalize the results of this study.*

Children in motor vehicle crashes in the United States (and other countries with similar driving conditions) of vehicles driven by parents or grandparents with similar insurance coverage and socioeconomic status as those in the study.

- (c) *Find and quote a passage in the Pediatrics that summarizes the results of the study that compares the risk of injury to children when grandparents or parents drive.*

"Children driven by grandparents comprised 9.5% of the sample but resulted in only 6.6% of the total injuries. Injuries were reported for 1302 children, for an overall injury rate of 1.02 (95% CI: 0.90–1.17) per 100 child occupants. These represented 161 weighted injuries (0.70% injury rate) with grandparent drivers and 2293 injuries (1.05% injury rate) with parent drivers. Although nearly all children were reported to have been restrained, children in crashes with grandparent drivers used optimal restraint slightly less often. Despite this, children in grandparent-driven crashes were at one-half the risk of injuries as those in parent-driven crashes (OR: 0.50 [95% CI: 0.33–0.75]) after adjustment." (p. 289).

*Is the USA Today summary an accurate report of the findings of the Pediatrics article? Consider the process by which data was sampled in your answer.*

One can take issue with the following statement from the summary "Kids may be safest in cars driven by grandma or grandpa instead of mom or dad, according to study results that even made the researchers do a double-take." The study can only show that *if* a child gets into an accident, then they are less likely to be injured if a grandparent is driving than if a parent is driving. However, to make the quoted assertion, we also need to know how likely kids are to get into car crashes when grandparents drive versus when parents drive.

Later in the summary, the USA Today author does mention that the study did not look at who had more crashes, but then the author immediately says that the study found that "children's risk for injury was 50% lower when riding with grandparents than with parents". For the reasons mentioned in my previous

paragraph, this is true only with a special definition of “risk for injury”, namely “risk for injury, given that a crash occurs”.

- (d) *Is it accurate to conclude on the basis of this research that children are less likely to be injured when driven by grandparents than when driven by parents?*

Not really, because of the reason given in my previous answer.

*If not, how can the statement be modified to provide a more accurate interpretation of the results?*

We can say that *if* a child gets into an accident, then they are less likely to be injured if a grandparent is driving than if a parent is driving.

- (e) *How might you collect data if wanted to address the question of whether or not grandparents drove more cautiously when they had grandchildren in the car than when they did not?*

We could use insurance records (like in this study) to find out how many accidents in a given time period involve motor vehicles driven by grandparents with no grandchildren in the car, versus how many such accidents involve motor vehicles driven by grandparents with grandchildren in the car. Here we assume that “drives more cautiously” means “gets in fewer accidents”.

3. (a) *Provide a citation for the article.*

Milk Production in Hereford Cows I. Means and Correlations  
O. W. Robison, M. K. M. Yusuff and E. U. Dillard  
J ANIM SCI 1978, 47:131-136.

- (b) *Describe the sample of individuals (units) for which data is collected. Describe how this sample was selected.*

The sample of individuals was 528 cows. This sample was selected from two herds of Hereford cattle located at Raleigh and Plymouth, NC.

- (c) *Report one scientific question that your article addresses.*

How does the weaning weight affect milk yield?

- (d) *What results does the article include regarding the scientific question?*

Weaning weight is positively correlated with milk yield. Also, about 40% of variability in milk yield is accounted for by variation in milk yield.

- (e) *Select one graph from the article. What variables are graphed on each axis? What does the graph illustrate?*

The horizontal axis shows age of cow, in years. How vertical axis shows milk yield, in kilograms. The lower three lines show the milk yield of cows, measured at three different observation times. One of the upper lines shows the sum of the bottom three lines, and the other line shows a prediction made by the authors’ model. The point of the plot is to provide validation of the authors’ model as being reasonable, since the upper two lines are generally close to each other.

- (f) *Identify two variables that are measured by the authors, one a response variable and one an explanatory variable, if applicable.*

V1: weaning weight, V2: total milk yield.

*What are the units of measurement?*

V1: kilograms, V2: kilograms.

*Are the variables summarized with statistics? If so, what are the values of the statistics?*

V1: N/A, V2: the total milk yield has mean 16.24 and standard deviation 2.71.

*Classify each variable as either categorical or numerical, as either experimental or observational, and as either explanatory or response.*

V1: numerical, V2: numerical.

V1: experimental, V2: observational.

V1: explanatory, V2: response.

- (g) *Identify a method of statistical analysis used by the authors and report a given result.*

Least squares regression. Their regression model predicted that the herds’ combined yield would be .72 of the original yield in 1975.