

APPENDIX FOR DAY I

Elaboration for p. 24:

Terminology:

- The *certain event* is the event "some possibility occurs."
 - For example, in rolling a die, the certain event is "One of 1, 2, 3, 4, 5, 6 comes up."
 - In considering the stock market, the certain event is "The Dow Jones either goes up or goes down or stays the same."
- Two events are called *mutually exclusive* if they cannot both occur simultaneously.
 - For example, the events "the die comes up 1" and "the die comes up 4" are mutually exclusive (assuming we are talking about the same toss of the same die).
- The *union* of events is the event that at least one of the events occurs.
 - For example, if E is the event "a 1 comes up on the die" and F is the event "an even number comes up on the die," then the union of E and F is the event "the number that comes up on the die is either 1 or even."

The ***axiomatic model of probability*** says that probability is a function (i.e., a rule; we'll call it P) that assigns a number to each event, and satisfies the three conditions (axioms; coherence conditions) below. (Just what constitutes events will depend on the situation where probability is being used.)

The three axioms of probability:

- I. 0 is the smallest allowable probability and 1 is the largest allowable probability (i.e., $0 \leq P(E) \leq 1$).
- II. $P(\text{certain event}) = 1$
- III. $P(\text{union of mutually exclusive events}) = \text{sum of } P \text{ of individual events}$

Example: If we have a fair die, the axioms of probability *require* that each number come up with probability $1/6$.

Proof: Since the die is fair, each number comes up with the same probability.

Since the outcomes "1 comes up," "2 comes up," ... "6 come up" are mutually exclusive and their union is the certain event, Axiom III says that

$$P(1 \text{ comes up}) + P(2 \text{ comes up}) + \dots + P(6 \text{ comes up})$$

= P(the certain event),

which is 1 (by Axiom II).

Since all six probabilities on the left are equal, that common probability must be 1/6

Common source of misunderstanding: Different uses of the word "risk."

Everyday meaning: risk = danger

Technical meanings: A number quantifying a danger.

1. Risk as a probability (*absolute risk*)

Example: "the risk that a U.S. resident dies from a heart attack is about 25%."

Note: Risk may be misunderstood if the reference category is not understood.

2. *Relative risk* (also called *risk ratio*).

- A method of comparing the risk for one group with the risk for another.
- E.g., one group might be people with a certain condition (or receiving a certain treatment) and the other group people without that condition (or not receiving the treatment). Or the groups might be men and women. Or smokers and non-smokers; etc.
- *Relative risk is the ratio of the risks for the two groups.*
- Three possible source of confusion:

i. What are the two groups?

ii. Which group's risk is in the numerator and which group's risk is in the denominator?

iii. A relative risk is difficult to interpret without knowing the absolute risks.

Example:

You are told that a certain treatment will reduce your risk of contracting a certain disease by 25%.

This means that the relative risk of those having the treatment compared to those who don't have the treatment is 0.75.

Scenario 1: The absolute risk for those not using the treatment is 40% (i.e., 4 out of 10).

- A risk reduction of 25% reduces the risk to 3 out of 10, giving absolute risk 30%.
- So 10% of people who use the treatment benefit from it.
- This is substantial, but not as substantial as “25% risk reduction” might sound.

Scenario 2: The absolute risk for those not using the treatment is 0.0004 % (i.e., 4 out of 1,000,000).

- A risk reduction of 25% reduces the risk to 3 out of 1,000,000.
- This is not a substantial reduction.

3. Risk = Probability X Consequences

Elaboration for p. 30: Conditional probabilities in medical testing

To figure out the positive predictive value if you know the sensitivity, you also need to know the *specificity*

$P(\text{tests negative} \mid \text{does not have the disease})$

and the *prevalence rate*

$P(\text{has the disease})$,

which itself might be a conditional probability – for example,

$P(\text{infected with HIV} \mid \text{intravenous drug user})$, or

$P(\text{infected with HIV} \mid \text{age over 80})$

For more detail, see the references given in the Notes at the webpage <http://www.ma.utexas.edu/users/mks/statmistakes/misundcond.html>

Elaboration for p. 37: Some other types of random samples

1. *Stratified random sample*:

- The population is first classified into groups (called *strata*) with similar characteristics.
- Then a simple random sample is chosen from each stratum separately.
- These simple random samples are combined to form the overall sample.

Examples of characteristics on which strata might be based include: gender, state, school district, county, age.

Reasons to use a stratified rather than simple random sample include:

- The researchers may be interested in studying results by strata as well as overall. Stratified sampling can help ensure that there are enough observations within each stratum to be able to make meaningful inferences by strata.
- Statistical techniques can be chosen taking the strata into account to allow stronger conclusions to be drawn.
- Practical considerations may make it impossible to take a simple random sample.

2. *One-stage cluster sample:*

- The population is also divided into groups, called *clusters*.
- Instead of sampling within each cluster, a simple random sample of *clusters* is selected.
- The overall sample consists of all individuals in the clusters that constitute this simple random sample of clusters.

Example: If the purpose of the study is to find the average hourly wage of convenience store employees in a city, the researcher might randomly select a sample of convenience stores in the city and find the hourly wages of all employees in each of the stores in the sample.

Note: The results from cluster samples are not as reliable as the results of simple random samples or stratified samples, so they should only be used if practical considerations do not allow a better sample scheme. For example, in the convenience store example, it may be practically speaking impossible to draw up a list of all convenience store employees in the city, but it would be much less difficult to draw up a list of all the convenience stores in the city.

Elaboration for p. 57: Asking questions

Devising good questions is much more complicated and subtle than it may initially appear; whole books have been written about wording questions; entire courses are devoted to survey design; there are labs devoted to testing out questions for surveys. See <http://www.ma.utexas.edu/users/mks/statmistakes/wordingquestions.html> and the links and references therein for more examples and resources.

Suggestions for researchers using questionnaires:

- Educate yourself about the problems involved in designing good survey questions.
- Test out your tentative questions on subjects *similar to those you plan to survey*.
- Then modify questions and re-test as needed.

- When reporting results of a survey, be sure to provide access to the exact questions asked, so others can verify whether or not the questions are likely to be ambiguous or influential.
- In reporting results, discuss any questions that turned out to be problematical.
 - Take the problems into account in interpreting analysis of the data.

Suggestions for reviewers, editors, etc.:

- Have the researchers done the above?

Suggestions for consumers of research:

- Be cautious in interpreting the results of a survey.
- In particular, try to find the exact questions asked and check them over for possible ambiguity or other problems in wording.
- If the authors of the survey are not willing to reveal the questions, be doubly cautious in making interpretations.

Suggestions for teachers:

- Even though you cannot include thorough coverage of the topic of wording in a general statistics course, be sure to
 - Mention it.
 - Give some examples
 - Ask some questions on it on exams, and
 - Have students pay attention to question wording if they are designing or carrying out a study.

Elaboration for p. 60: Mean vs median for skewed distributions

- *Example:* The mean of the ten numbers 1, 1, 1, 2, 2, 3, 5, 8, 12, 17 is $55/10 = 5.2$.
- Seven of the ten numbers are less than the mean, with only three of the ten numbers greater than the mean.
- A better measure of the center for this distribution would be the *median*, which in this case is $(2+3)/2 = 2.5$.
- Five of the numbers are less than the median 2.5, and five are greater.
- Notice that in this example, the mean (5.2) is *greater* than the median (2.5).
- This is common for a distribution that is *skewed to the right* (that is, bunched up toward the left and with a "tail" stretching toward the right).
- Similarly, a distribution that is *skewed to the left* (bunched up toward the right with a "tail" stretching toward the left) typically has a mean *smaller* than its median.
- See von Hippel 2005 for discussion of exceptions.

Elaboration for p. 69: Unusual Events

More examples of unusual events, where a measure of center is not appropriate:

3. Traffic safety interventions typically are aimed at high-speed situations.
 - So the average speed is not as useful as, say, the 85th percentile of speed.

4. If two medications for lowering blood pressure have been compared in a well-designed, carefully carried out randomized clinical trial, and the average drop in blood pressure for Drug A is more than that for Drug B, we cannot conclude just from this information alone that Drug A is better than Drug B. We also need to consider the incidence of undesirable side effects.
 - One might be that for some patients, Drug A lowers blood pressure to dangerously low levels.
 - Or it might be the case that for some patients, Drug A actually increases blood pressure.
 - Thus in this situation, we need to consider extreme events in *both* directions.
 - Note that this is another example where variability is important.

5. Unusual events such as earthquakes and extreme behavior of the stock market can have large effects, so are important to consider.

Many techniques have been developed for studying unusual events.

- However, these techniques are not usually mentioned in introductory courses in statistics.
- And, like other statistical techniques, they are not "one-size-fits-all."
- See <http://www.ma.utexas.edu/users/mks/statmistakes/unusualevents.html> for some references.

REFERENCES

Agresti (2010) *Analysis of Ordinal Categorical Data*, Wiley

Eddington, Eugene S., *Randomization Tests*, 1995, Marcel Dekker.

Ernst, Michael D., "Permutation methods: A basis for exact inference", *Statistical Science*, 2004, v.19, no.4, 676-685.

Gigerenzer, Gerd, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M. Schwartz, Steven Woloshin (2007), "Helping Doctors and Patients Make Sense of Health Statistics," *Psychological Science in the Public Interest*, vol. 8, No. 2, pp. 53 - 96.

Download from

http://www.psychologicalscience.org/journals/index.cfm?journal=pspi&content=pspi/8_2

Good, P. (2005) *Introduction to Statistics Through Resampling Methods and R/S-PLUS*. Wiley.

Good, P. (2005) *Introduction to Statistics Through Resampling Methods and Microsoft Office Excel*. Wiley.

Jobling, MA and Tyler-Smith C (2003), The human Y chromosome: an evolutionary marker comes of age, *Nature Review Genetics*, 4(8):598-612

Koenker, Roger, Quantile Regression website,
<http://www.econ.uiuc.edu/~roger/research/rq/rq.html> Includes references to books, articles, and software.

Limpert, E. and Stahel, W. (1998) Life is Lognormal!,
<http://www.inf.ethz.ch/personal/gut/lognormal/brochure.html>

Moore, Thomas (2010), Using baboon “mothering” behavior to teach permutation tests, Cause Webinar, <http://www.causeweb.org/webinar/teaching/2010-09/>.

Video and power-point slides. A gentle introduction to permutation tests.

Pablos-Mendez, a. et al (1998), Run-in periods in randomized trials: Implications for the application of results in clinical practice, *JAMA* 279(3),
<http://mres.gmu.edu/pmwiki/uploads/Main/RunInTrialsRCT>

Senn, Stephen and Steven Julious (2009), Measurement in clinical trials: A neglected issue for statisticians, *Statistics in Medicine* **28**:3189-3209.

Von Hippel, Paul (2005), “Mean, Median, and Skew: Correcting a Textbook Rule,” *Journal of Statistics Education*, v. 13 no. 2,
www.amstat.org/publications/jse/v13n2/vonhippel.html

Wainer, Howard (2011) The first step toward wisdom, *Chance* vol 24, no. 2,
<http://chance.amstat.org/2011/05/visualrev/>