Ethics

The purpose of this section is to describe what is proper conduct for a statistician in accordance with the standards of the profession. Most of these notes are developed from my own experiences and the discussions given in

- The ASA's *Ethical Guidelines for Statistical Practice* that is available through a link at `https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practi` `aspx`. These guidelines were developed by the ASA's Committee on Professional Ethics in 2016. There was a JSM 2016 session on these guidelines as well.
- The ISI's *Declaration of Professional Ethics* that is available through a link at `https://www.isi-web.org/index.php/activities/professional-ethics/isi-declaration`. These guidelines were developed by the ISI's Professional Ethics Committee in 2010.
- Section 10.4 of Hahn and Doganaksoy (2011).

Some of the content in these notes should not be surprising (e.g., do not fabricate data), but other content may be new (e.g., do not plagiarize yourself). At the very least, it is good to be reminded what is and is not proper conduct for a statistician.

Plagiarism

Below are quotes from the introductions of two different papers which are meant to describe what is group (pooled) testing:

- Ebert et al. (*Annals of the Entomological Society of America*, 2010, p. 827, `http://dx.doi.org/10.1603/AN09158`)

> In practice, pooling is a simple process. If 30,000 mosquitoes are collected from the field, they could be tested one at a time for a viral pathogen. If each test takes 10 min and costs US$15, then this project will take 5,000 h and cost US$450,000. A shorter approach would be to smash 10 mosquitoes together and test this pooled sample. This approach would take 500 h and cost US$45,000. Even greater savings are achieved with larger pool sizes. However, there is an obvious problem. If two individuals are infected and pool size

- Montesinos-Lopez et al. (*PLoS ONE*, 2012, p. 2, `http://dx.doi.org/10.1371/journal.pone.0032250`)

> In practice, pooling is a simple process; for example, if 40,000 plants are collected from the field, they could be tested one at a time for detecting unwanted transgenic plants (AP). If each test takes 15 minutes and costs US$12, then this project will take 10,000 hours and cost US$480,000. A shorter approach would be to smash 10 plants together and test this pooled sample [15]. This approach would take 1000 hours and cost US$48,000. Even greater savings are achieved with larger pool sizes. However, because the maximum likelihood estimator (MLE) of p under binomial [16] and negative binomial [9,10] group testing is biased

What is wrong with the second paper? Note that the numbered references in the last line should not influence your answer.

The key ways to avoiding plagiarism are:

- Use quotation marks or offset text from the main text any content which is a direct quote taken from someone's work; include a citation to this work.
- Paraphrase someone's work and include a citation to this work.

Could one simply fix the excerpt from the second paper by adding a citation and quotation marks in the appropriate locations? Yes, but the amount of quotations would likely cause the journal's

editor to reject the paper because it does not represent the authors own work.

Below is my description of what group testing is in the same context (transgenic plants, cost/time included) as Montesinos-Lopez et al. (2012) without plagiarism of Ebert et al. (2010):

> Testing for transgenic plants is a relatively simple process. For example, suppose there are 40,000 plants from a field that need to be test for unwanted transgenic plants. One could simply test each plant one by one. If each test costs US$12 and takes 15 minutes, the total cost and testing time would be US$480,000 and 10,000 hours, respectively. In most cases, one would find this to be unsatisfactory. Instead, one could pool together 10 plants (say, grind them up to form an amalgamation) at a time and perform tests on each pool formed. Completing this process for all 40,000 plants would reduce the total cost and time by one-tenth while still allowing for the computation of a maximum likelihood estimator for p. Of course, using larger pools would lead to even larger reductions in costs and time.

Overall, the process of group testing is well known, so no references are needed for it. If one wanted to be safe, a reference to the earilest paper on group testing could be given. For example, the previous paragraph could be changed to

> ... to be unsatisfactory. Instead, one could use a process known as pooled testing (Dorfman 1943) that would involve compositing together 10 plants at a time and performing tests on each pool formed. Completing this process for all 40,000 plants ...

Because group testing is my main research area, I very often need to explain the basics of group testing in my writing. Below are a few examples of how I introduce group testing.

1. Bilder and Tebbs (*Statistics in Medicine*, 2012): I had to quickly describe group testing due to smaller than normal page limits given by the journal.

1. Introduction

Pooled testing, also known as group testing, is a procedure where individual specimens (e.g., urine or blood) are combined into a pooled specimen to test for a binary response (e.g., positive or negative HIV status). In the most widely used form of pooled testing known as 'Dorfman testing' [1], pools that test negative have all individuals within them declared negative. Pools that test positive indicate that at least one individual within each pool is positive, and individual retesting of each specimen is subsequently used to decode the positives from the negatives. The strong appeal of pooled testing is that it can significantly reduce the number of tests and associated costs when the prevalence for a disease is small. This has led to the application of pooled testing in a wide variety of infectious disease screening settings, such as blood donation screening by the American Red Cross [2, 3], chlamydia and gonorrhea opportunistic testing in medical clinics [4], influenza surveillance through blood donations [5], and West Nile virus surveillance in mosquitoes [6].

2. Zhang, Bilder, and Tebbs (*Biometrical Journal*, 2013): This is another quick description.

1 Introduction

Pooling specimens to screen a population for infectious diseases has a long history dating back to Dorfman's (1943) proposal to screen American soldiers for syphilis during World War II. Today, testing individuals in pools through group testing (also known as "pooled testing") has been successfully adopted in many additional areas, including entomology (Gu et al., 2004), veterinary medicine (Muñoz-Zanzi et al., 2000), DNA screening (Berger et al., 2000), and drug discovery (Kainkaryam and Woolf, 2009). When compared to testing specimens individually, group testing can provide considerable savings in time and costs when the overall prevalence of the disease (or some other binary characteristic of interest) is low. This makes the use of group testing particularly desirable in applications where there are limitations in resources.

3. Bilder, Tebbs, and Chen (Journal of the American Statistical Association, 2010): This is a longer description.

1. INTRODUCTION

Chlamydia and gonorrhea are the two most prevalent bacteria-based sexually transmitted diseases in the United States, and $4 billion is spent annually on these infections (Infertility Prevention Project, Region VII 2003). Infected persons are often asymptomatic, resulting in individuals being left untreated and others becoming infected unknowingly. Both diseases can lead to severe consequences including infertility and a higher susceptibility to HIV infection (Kacena et al. 1998a, 1998b). To address this public health problem, the Infertility Prevention Project (IPP), funded by the Centers for Disease Control and Prevention, has been implemented nationwide. Its purpose is to screen and to provide treatment for chlamydia and gonorrhea in higher risk populations while monitoring disease prevalence. We will focus on the screening aspect in this paper, specifically its implementation in Nebraska. Screening is especially important in Nebraska because chlamydia and gonorrhea infections have been characterized as being at epidemic levels (Zagurski 2006). Currently, more than 30,000 tests are completed annually in Nebraska, and all testing is performed on *individual* specimens.

Given the large number of tests in Nebraska and the associated cost, it is important to find ways to reduce the amount of testing needed without screening fewer individuals. In similar situations where a large number of individuals are screened for infectious diseases, it has become standard practice to perform screening tests on *pools* or *groups* of individual specimens (e.g., blood, urine, etc.). Group testing, also known as pooled testing, was introduced by Dorfman (1943) to screen

World War II soldiers for syphilis. Since this seminal work, the usefulness of pooling has been demonstrated in blood donation screening (Stramer et al. 2004), in screening individuals for drug use (Gastwirth and Johnson 1994), in preventing the potential spread of bioterrorist agents (Schmidt et al. 2005), and in other applications including genetics, plant pathology, veterinary, and drug discovery (Gastwirth 2000; Tebbs and Bilder 2004; Peck 2006; Remlinger et al. 2006). In general, group testing for case identification involves testing individuals first in groups. Positive groups are then "decoded" through algorithmic procedures to identify positive individuals. Perhaps due to its simplicity, Dorfman's (1943) original procedure, where each individual in a positive pool is retested, is the most widely used. However, many other retesting strategies have been proposed; see Hughes-Oliver (2006) for a review.

4. NIH grant R01AI121351: I tried to make my introduction stand out to the grant reviewer.

Specific Aims

A patient anxiously awaits a disease diagnosis in an examination room. Chlamydia, gonorrhea, or even HIV are possibilities. "Positive or negative," the patient wonders aloud as minutes seemingly feel like hours. Finally, the doctor enters and reports the diagnosis as ...

This situation plays out every day in hospitals and clinics throughout the United States. Patients expect their disease diagnosis to be accurate and timely while still at a reasonable cost. In many cases, these expectations are met using an innovative procedure known as *group testing* (also known as *pooled testing*). Rather than testing specimens one by one, group testing works by pooling specimens, such as blood or urine, from separate individuals to form a single specimen. Individuals within negative testing groups (pools) are declared negative. Individuals within positive testing groups are retested in some predetermined manner to distinguish the positive individuals from the negative ones. Accuracy, timeliness, and cost effectiveness are achieved as long as the overall disease prevalence in the population is small. Examples where group testing is used include

- Chlamydia and gonorrhea testing at laboratories across the United States as part of federally sponsored sexually transmitted disease (STD) assessment and prevention programs; see Lewis et al. (2012) and Centers for Disease Control and Prevention (2015)
- HIV, hepatitis B, and hepatitis C screening of blood donations; see Schmidt et al. (2010), Stramer et al. (2011), and O'Brien et al. (2012).

You may be wondering

> Why can't you use the same description each time since you wrote each of them?

The reason is because you would be plagiarizing yourself! A BIG reason why this is plagiarism is because most publishing companies own the copyright to all papers published in their journals (authors are required to give up the copyright). Outside of the

copyright issues or even writing papers for journals, there is still an expectation in academics that the work is original and unique for a particular setting.

When would it be o.k. then to "re-use" previous items that you have written:

- Publishing in a journal your work from a dissertation or thesis
- Publishing in a journal your work that was already given in a conference proceedings, although make sure to check the rules of the proceedings; note that it is o.k. to use re-use work published in the JSM proceedings

The Office of Graduate Studies provides a good discussion of plagiarism at `https://www.unl.edu/gradstudies/current/integrity#plagiarism`. Below is a screen capture of an example that they provide (pay attention to what is acceptable paraphrasing).

Examples of proper use of others' words and ideas

To illustrate an example of plagiarism, as well as proper ways to use the words and ideas of someone else, we present a short original passage, followed by examples of a plagiarized paraphrase and an acceptable paraphrase.

ORIGINAL TEXT	Dengue virus infections in humans can be subclinical or can cause illnesses ranging from a mild, flulike syndrome with rash and some hemorrhagic manifestations (dengue fever [DF]) to a severe and sometimes fatal disease, with coagulopathy, capillary leakage, and hypovolemic shock (dengue hemorrhagic fever [DHF]).	This is the original text from page 1 of "Dengue Fever in Humanized NOD/SCID Mice" by D.A. Bente, et al. in the *Journal of Virology*, November 2005.
UNACCEPTABLE PARAPHRASE	Dengue virus infections in humans can range in intensity from subclinical manifestations, to a mild flulike illness with a rash and some hemorrhaging (dengue fever [DF]) to a severe and sometimes fatal disease with blood clotting defects, leaking capillaries, and hypovolemic shock (dengue hemorrhagic fever [DHF]).	This is considered **plagiarism** because the writer has: • only changed around a few words and phrases • failed to cite a source for any of the facts or ideas
ACCEPTABLE PARAPHRASE	Dengue virus infections affect humans in a variety of ways. In some, the disease doesn't show up at all; others may have a rash and some minor bleeding, while still others may experience severe bleeding, shock, and even death (Bente et al., 2005).	This is **acceptable paraphrasing** because the writer: • accurately relays the information in the original • uses her own words • lets her reader know the source of her information.
ACCEPTABLE PARAPHRASE WITH QUOTATION	In humans, dengue virus infections can range from mild to severe, from a flu-like syndrome "to a severe and sometimes fatal disease, with coagulopathy, capillary leakage, and hypovolemic shock" (Bente, et al., 2005, p.1).	This is **acceptable paraphrasing** because the writer: • gives credit for the ideas in this passage • indicates which parts are taken directly from the source by putting them in quotation marks and citing the page number.

Final notes:

- What is the acceptable style to use when citing a paper? There

is not one standard style in Statistics. Below are some examples,

  - In text: Bilder (2009) shows how group testing could be used to detect Cylons on the TV show *Battlestar Galactica*.
  - Parenthetical: Group testing has been described in a number of settings, including Cylon detection on the TV show *Battlestar Galactica* (Bilder 2009).

  In contexts where a bibliography is not given (like these notes!), it can be helpful to include the journal name as Bilder (*Chance*, 2009). The use of commas within the parentheses in the last example and the parenthetical example is journal dependent. If you are writing something where there is not a comma standard, pick one and be consistent in your writing! Overall, there are very few statistics journals that use numerical citations like shown in my previous paper excerpt from *Statistics in Medicine*.

- UNL Writing Center: `https://www.unl.edu/writing`

Integrity of data and methods

The data observed is the data that needs to be analyzed! This may seem like common sense, but pressure to obtain results, unusual observations, and different subsets of the data may make it more difficult.

Falsifying data

The most egregious cases of violating this code of conduct involve falsifying data to achieve a desired outcome. One very likely example of where this occurred was with Anil Potti at Duke University a few years ago. Potti developed a personalized cancer treatment method and published his research in top

medical journals. Unfortunately, it was eventually discovered that most likely the data which supported Potti's conclusions was falsified. See `https://www.youtube.com/watch?v=W5sZTNPMQRM` for a discussion about this from the TV show *60 Minutes*.

Unusual and influential observations

Observations may not conform to an otherwise well fitting statistical model or they may unduly influence the model leading to potentially different conclusions. How to handle these situations can be difficult. Here are some potential solutions:

- Develop a different model; for example, a model which is robust to influential observations, where reasoning for it is given in a corresponding report or paper
- Remove the observations from the data but detail the consequences in the corresponding report or paper
- Leave the observations in but detail the consequences in the corresponding report or paper

I encountered a situation like this when doing the work for Bilder and Loughin (*Chance*, 1998). This paper developed a logistic regression model to estimate the probability of success for placekicks in football. During the model building process, I found two observations which were very influential and then narrowed down this influence to an interaction term for the distance of the placekick and the type of placekick (PAT or field goal).

These two observations represented very unusual situations for football–non-20 yard PATs (at this time in football, all PATs were 20 yards unless there was a penalty on the original placekick attempt). To solve the problem, I decided to remove all non-20 yard PATs from the data set (not just those that were influential). I reported in the paper that the removal occurred and the population of inference was subsequently reduced by not including this

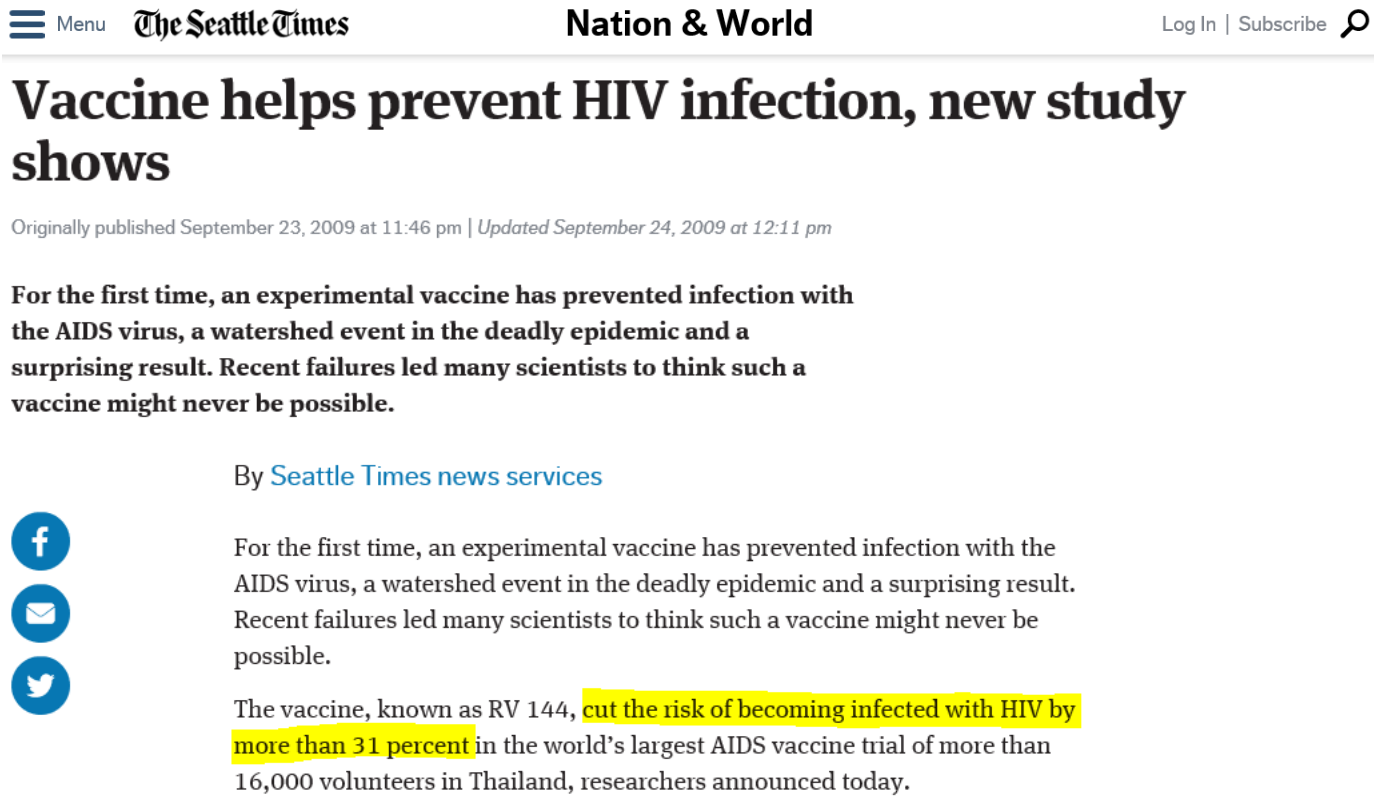
situation. There are other justifiable possible solutions too, and these are detailed in STAT 875.

Tell the whole story

When presenting an analysis, it is always important to "tell the whole story" rather than only talk about what benefits you and your colleagues. Bilder and Loughin (2014, p. 55) discuss a situation where there are questions whether researchers presented all of the pertinent information (at least at first). Below is the background regarding it:

> On September 24, 2009, news reports hailed the findings from an HIV vaccine clinical trial as being the first time that a vaccine worked. These news reports often made front-page headlines in newspapers and lead stories on television newscasts:

- *The Seattle Times*:



Menu The Seattle Times **Nation & World** Log In | Subscribe

Vaccine helps prevent HIV infection, new study shows

Originally published September 23, 2009 at 11:46 pm | *Updated September 24, 2009 at 12:11 pm*

For the first time, an experimental vaccine has prevented infection with the AIDS virus, a watershed event in the deadly epidemic and a surprising result. Recent failures led many scientists to think such a vaccine might never be possible.

By Seattle Times news services

For the first time, an experimental vaccine has prevented infection with the AIDS virus, a watershed event in the deadly epidemic and a surprising result. Recent failures led many scientists to think such a vaccine might never be possible.

The vaccine, known as RV 144, cut the risk of becoming infected with HIV by more than 31 percent in the world's largest AIDS vaccine trial of more than 16,000 volunteers in Thailand, researchers announced today.

- *The PBS News Hour*: `http://www.pbs.org/newshour/bb/health-july-dec09-hiv_09-24`

The clinical trial was performed in Thailand. Study participants were given the vaccine or a placebo a number of times over a period of months, and then subsequently tracked to determine if the vaccine was effective in preventing HIV infection. At the time of original news reports, results from the study were not published yet but would be subsequently a few weeks later in Rerks-Ngram et al. (*New England Journal of Medicine*, 2009). The released conclusions were partially based on a hypothesis test performed on the *modified intent-to-treat data* that gave a p-value of 0.04 (low values indicate evidence to reject a null hypothesis of "no effect").

When the paper was published on October 20, 2009, the paper also included two additional analyses on versions of the data referred to as the *intent-to-treat* and the *per-protocol*. In summary,

- 16,402 participants were in the intent-to-treat data; these are individuals who were in the study but perhaps did not complete the entire vaccine regimen; p-value = 0.08
- 16,395 participants were in the modified intent-to-treat data, where seven individuals were removed from the intent-to-treat data because they were later found to be infected with HIV prior to beginning the vaccine regimen; p-value = 0.04
- 12,542 participants were in the per-protocol data; these are individuals who completed the entire vaccine regimen; p-value = 0.16

These new results again were publicized by the media, but not with the same types of headlines as before. A *Los Angeles Times* article (Maugh, 2009) said the following:

> A secondary analysis of data from the Thai AIDS vaccine trial—announced last month to much acclaim—suggests that the vaccine might provide some protection against the virus, but that the results are not statistically significant. In short, they could have come about merely by chance.

Given the results here, this leads to a number of questions:

- Why were the modified intent-to-treat data results released first?
- Are the results *really* different for the modified intent-to-treat and the intent-to-treat data?
- Why didn't the media make as big of a deal about the results on October 20 as on September 24?

Overall, there are a number of issues here without clear answers. Hopefully, this example will make you think about these issues.

Additional comments

1. Acknowledge assumptions inherent to a statistical method
2. Discuss potentially important items that are not accounted for in an analysis
3. Account for a larger chance of an inference error as more are made (e.g., control a familywise error rate)
4. Protect the privacy of the data
5. Recognize that results should not be made to conform to prior beliefs of a subject-matter researcher
6. Use up-to-date methods

Reproducibility of research

Research needs to be reproducible! Any research findings found by yourself should be able to be found by others as well if they employ your methods. Without reproducibility, the research has no meaning and can not be applied in practice. What causes studies to not be reproducible? Below are some statistical reasons:

- Poor understanding of statistics by non-statisticians

- Integrity of the data and methods not upheld
- Lack of explanation

With respect to statistics research, improvements are being made to make our research reproducibile. Below are some important recent improvements:

- R is the "lingua franca" for Statistics. The software has revolutionized how we communicate with others because its free and there are easy ways to disseminate code (e.g., packages posted on CRAN or GitHub). When I was a student, a statistical journal paper would rarely have a computer program associated with it that would allow readers to immediately try out the research. Now, it can be difficult to get published without one. Most statistical research will have the computations programmed in R. When computations take a long time in R, at least part of the programming will be done in a faster language, like C++.
- Some journals, including *Biometrics*, *Biometrical Journal*, *Biostatistics*, and *Journal of the American Statistical Association*, require a submission of programs, data, and additional results (i.e., web appendix) that were used in a paper. At least for me, it is still surprising that until recently no journals had the program and data requirement. In fact, even referees could not easily check any computations. *Biometrical Journal* is one of the best with respect to reproducibility because they have an editor that actually examines these additional submissions (they likely were the first statistics journal to have one). Below are requirements given by *Biometrical Journal* on their website:

Reproducible Research

Biometrical Journal aims to increase the practical impact of methodological research published in the journal by making computational tools used in an article available to readers. Authors are therefore encouraged to submit data sets used as examples, source code of data analyses presented, sources of simulation studies and implementations of new methodology, *etc.*, as Supporting Information for online publication. Open-source statistical software environments or programming languages with freely available interpreters/ compilers are preferred. The submitted material will be reviewed by the Reproducible Research Editor after acceptance of the manuscript. It is mandatory that results reported in the manuscript coincide with results produced by the software code submitted.

Files for Reproducible Research need to be uploaded as one single ZIP file together with the manuscript, using the file designation "Data and Software" from the dropdown box.

Please note that it is mandatory that the submission follows the "Guidelines for Code and Data Submission" on the Journal's homepage. For further details on Biometrical Journal's reproducible research policy, please consult the editorial on reproducible research published in *Biometrical Journal* 51(4), 553–555.

Their *Guidelines for Code and Data Submission* is a good read. Even if you do not ever publish papers, there are a number of good practices described in that document (e.g., "The code must be well documented") to use for your own academic and post-academic work.

For Zhang, Bilder, and Tebbs (*Biometrical Journal*, 2013), we

- Created a 20 page web appendix that provided additional results
- Provided an R program that can fit the proposed models to a data set
- Included additional functions for the `binGroup` package in R
- Gave a simulated data set similar to the data used to illustrate the proposed statistical methods (my data source did

not allow me to share the data)

Surprisingly, the journal did not post this information to their website!

Nature discusses this issue in a September 2016 article at `http://www.nature.com/news/why-scientists-must-share-their-research-code-1.20504?WT.mc_id=TWT_NatureNews`. This article focuses on the *Journal of the American Statistical Association*.

Even if a journal does not have a requirement for reproducibility, one should still make available a web page that contains additional information like a program. My group testing research website at `www.chrisbilder.com/grouptesting` provides some examples.

- Document creation software (like LaTeX/LyX) allow for one to embed R code used for a statistical analysis *within* a document. When the document is created, the code is run and the output is put into the document. These dynamically created document tools include the use of the `knitr` package from R. Advances in R markdown and bookdown now provide alternative ways to do this as well.

Final comments

1. Check your work and strive to minimize the chances for programming errors
2. Include the major competing methods in comparisons made for statistical research
3. Share data

Additional resources

- UNL
  - Student Code of Conduct: https://studentconduct.unl.edu/student-code-conduct
  - Graduate Studies and academic integrity - https://www.unl.edu/gradstudies/current/integrity
  - Ethics Center - https://ethics.unl.edu
  - Responsible conduct of research training - https://research.unl.edu/researchcompliance/responsible-con
- ASA's Committee on Professional Ethics has a set of case studies - https://community.amstat.org/communities/community-home?CommunityKey=b06cd782-6770-4feb-94b1-c
- Ethics and Statistics column in Chance Magazine - http://chance.amstat.org/category/columns/ethics-and-statist
- He, X. (2013). "Ethics in Publishing," *IMS Bulletin* 42(3), 4. http://bulletin.imstat.org/2013/04/ethics-in-publishi
- Ioannidis, J. (2005). "Why Most Published Research Findings Are False," *PLoS Medicine* 2(8), e124. http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020124
- ISI's international work to provide help on ethics cases - http://www.isi-web.org/index.php/activities/professional-et/isi-statements-letters
- National Institutes of Health
  - Resnik, D. (2015). "What is ethics in research and why is it important?" National Institute of Environmental Health Sciences website, http://www.niehs.nih.gov/research/resources/bioethics/whatis

  - NIH bioethics research https://osp.od.nih.gov/clinical-r bioethics-research/

- North Carolina State University Department of Statistics - PhD course requirements include an ethics course, https://statistics.sciences.ncsu.edu/graduate/phd-programs
- P-values:
  - Nuzzo, R. (2014). "Scientific method: Statistical errors," *Nature* 506, 150-152.
  - Wassertsein, R. and Lazar, N. (2016). "The ASA's Statement on p-Values: Context, Process, and Purpose," *American Statistician* 70(2), 129-133.