# Week 6

Goal: In these assignments, practice interpreting and analyzing data from contingency tables. Furthermore, they evaluate the scope of inferences that can be made from observational studies, and practice statistical literacy by writing a letter to a friend.

## Part 0 – Extra Credit

A 2009 study by Baker et. al. examined PTSD by branch of service. This week we will use data they collected to answer the question: "Is there evidence that there is a difference in the proportion of military veterans diagnosed with PTSD by service branch?"

The study included 339 veterans from four branches: the army, navy, marines, and national guard.  Download the paper here: [Baker et. al. 2009](https://blackboard.gwu.edu/bbcswebdav/pid-11083926-dt-content-rid-64870399_2/xid-64870399_2). Read the paper and use it to answer the following questions. Refer to the section of the paper where the researchers discuss how they recruited their participants as well as the introduction.

(1a) What is the researchers' population of interest?

(1b) Did the researchers select participants by random sampling from their entire population of interest? Support your answer by explaining their sampling procedure, and why it does or does not fit the definition of a random sample.

(2a) What kind of study design the the researchers use?

(2b) Did the researchers use random allocation to assign study participants into branch of service?  Support your answer by explaining their design procedure, and why it does or does not fit the definition of a random allocation (a.k.a. randomization a.k.a. randomized experiment).

(3) Based on your answers to questions 1 and 2, can the researchers establish a causal claim? Can the researchers generalize their results to the entire population of interest?

**due Thursday**: Submit a single .docx or .pdf file with your completed work, including screenshots of StatKey, written out equations, and explanations as appropriate.

## Part 1

The following data comes from the study by Baker et al. (2009). The table below provides counts for the number of veterans diagnosed with PTSD by branch of service. Use the following table to complete each step. Submit your results following this same outline:

|  |  |  |  |
| --- | --- | --- | --- |
| Branch | PTSD | No PTSD | Total |
| Army | 34 | 35 | 69 |
| Marines | 48 | 65 | 113 |
| National Guard | 10 | 12 | 22 |
| Navy | 33 | 102 | 135 |
| Total | 125 | 214 | 339 |

(1) Calculate the risk of PTSD for each service branch. (Hint: Review the audio lectures for Week 6).

(2) When we calculate absolute risk difference (ARD) and relative risk (RR), we designate one group as the reference group. The reference group is what we compare to. Use the Army as the reference group and compare PTSD risks in each of the other branches to the Army. Calculate "navy - army", "usmc - army", and "ng - army" for the ARDs. For the RRs, it would be "navy / army", "usmc / army", and "ng / army".   (Hint: Review the audio lectures for Week 6).

(3) Inspect the risks, relative risks, absolute risk differences, and relative risk reductions. Do there appear to be any clinically significant differences?  (Hint: Review the Youtube Lecture Part 1 by Professor Rao). (Hint: don't forget to consider sampling variability - don't overinterpret these point estimates.)

(4) One explanation for the differences in risks is that sampling variability has made it look like there are differences even though there aren't. To test this hypothesis, conduct a chi-squared test for association to assess whether the risk of PTSD is related to military branch. Report and interpret the null hypothesis, your p-value, and your conclusion. Include a screenshot  of StatKey.

(5) If you find that there is a statistically significance difference, identify which branches seem to be different from the others. (Hint: Review the Youtube Lecture Part 2 by Professor Rao and the resource video by Dr. Melvin).

(6) We are particularly interested in comparing the risk of PTSD in the Navy to the Army. Estimate the difference in PTSD risk between the Army and Navy by creating an interval estimate. Use the confidence interval for a difference in proportions procedure (2\*SE method). Report and interpret the results. (Hint: review the resource video by Dr. Zimmerman).

**due Thursday:** Submit a single .docx or .pdf file with your completed work, including screenshots of StatKey, written out equations, and explanations as appropriate.

## Part 2

Suppose you receive the following letter from a friend:

Hey stranger! How ya doin'? Listen, I've been thinking of joining the Marines, and my parents are really worried. They read some study that examined    PTSD by branch of service. Apparently, the study interviewed vets and my parents are saying if I join the Marines I'll get PTSD -- They want me to join the Navy instead! I know you're in a stats class so I was hoping you could help me decipher the paper. Are they really saying that joining the Marines makes you get PTSD? [*Here's a link to the paper*](https://blackboard.gwu.edu/bbcswebdav/pid-11083934-dt-content-rid-71372244_2/xid-71372244_2).

Thanks for your help!

Your assignment is to review the highlighted portions of the attached paper and write a one paragraph response (i.e., 4-5 sentences) to your friend to answer their questions. Be sure to consider and address (a) how participants were selected for the study, (b) how participants were grouped, (c) estimates of risk and relative risk, (d) the effect of sampling variability, and (e) either confidence intervals or hypothesis tests. Use 'simple' language while attending to the important concepts we've been discussing in the class (i.e., sampling variability, random sampling, random allocation, estimation and inference). Address the fact that your friend's parents are making a causal claim - is this accurate from a statistical standpoint?

Post your answers to the discussion board by **Thursday**.

## Part 3

Respond to at least two other colleagues' posts. Comment on the *statistical* information they included in their summary, any holes in their justification, or their communication of statistical information. Ask clarifying questions. This is your chance to practice and improve your statistical communication and understanding.

Answer any questions that your classmates or instructors ask about your post.

due **Sunday.**