

Data Mining: PSY 392

Greg Francis, PhD
Department of Psychological Sciences
Final Exam

Name _____

Total points on the exam is 100. The exam will count as 20% of your class grade. Answers can be written, typed, or mixed. Do organize your answers so that it is obvious which problem they are for. To get partial credit, you need to show your work, which can be based on calculations or on values from calculators. Explain how you get your answer.

The exam is take home. It is due on Wednesday, December 11 by 5:00 pm. You can hand in a hard copy or email Garrett an electronic copy.

You can use the textbook, lecture notes, and look up information online. You can talk with the professor and the TA. Do not work with other students on the problems. If we determine that you worked with other students in the class or had someone else answer the questions for you, you will receive a score of 0 on the exam and you will be reported to the Dean of Students.

When you do use an on-line calculator or program, print a screen shot of the calculator to show your work. Label each figure to indicate which problem it goes with.

(1) (15 points) Using the *p*-hacker app at

<https://shinyapps.org/apps/p-hacker/>

enter the “Use seed” value of 3539 at the bottom left of the page. Keep everything else with the default values.

- a) Using the “Do an expert subgroup analysis” feature, what significant result do you find?

- b) Continue *p*-hacking until you get a total of three significant test results for the sample (they can be across different subgroups). Explain what you did to get this pattern of results (screen shots would be helpful). Write a short paragraph describing your findings as if you wanted to deceive readers into believing that there are legitimate scientific findings here (be creative).

(2) (15 points) Use the STATCHECK app at

<http://statcheck.io/index.php>

to evaluate the paper by Lazareva et al. (2005), which is available for download on the class web site.

- a) Summarize the results of STATCHECK. What does the analysis say about the paper?

- b) For each “Decision inconsistency” discuss what might have happened for there to be an inconsistency.

- c) Evaluate how well STATCHECK does, itself. Does it find and check all the statistics in the paper? Give an example where it misses a statistic. Why do you think it has a problem with these cases?

(3) (20 points) Does humor vary across countries? In the STATLAB sense of humor experiment, subjects rate jokes on a scale from 0–10, with bigger numbers meaning a joke was considered to be funnier. At the end of the experiment, each subject produces an average joke rating, which measures the sense of humor for that subject. The experiment has been run by subjects from three different countries: Canada, United Kingdom (UK), and United States of America (USA). Data for 186 subjects is available in an Excel spreadsheet called “SenseOfHumor.xlsx” at the class web site.

- a) Run an independent means ANOVA to test for differences in sense of humor across these countries. Use the on-line calculator at

<https://introstatsonline.com/chapters/calculators/OneWayANOVA.shtml>

- b) Run three contrasts to test all possible pairs of means against each other. Summarize your results. (It can be helpful to provide screenshots of the calculator.)

(4) (15 points) A study by Gao, McCarthy and Scholl (2010) studied the “Wolfpack effect”, where animacy (pointing toward a target) seems to imply intentional tracking behavior. For a demonstration, see the first two videos at

<http://perception.yale.edu/Brian/demos/Animacy-Wolfpack.html>

In Experiment 1, subjects were asked to determine whether there existed an element on the display, called the “wolf,” that actually tracked another element, called the “sheep.” The expectation was that it would be harder to detect the presence/absence of the wolf when other elements (arrows) pointed toward the sheep element (the Wolfpack effect). Indeed, *t*-tests found that performance detecting the presence/absence of the wolf was worse for the Wolfpack condition compared to three other conditions (arrows pointing perpendicular, arrows pointing in a direction that matches how they move, and the elements are discs).

The table below shows the mean percent correct (identifying whether a wolf tracked the sheep, or not) for the different conditions.

	Wolfpack	Perpendicular	Match	Discs
Mean	62	72	75	72

Using $n = 12$ observers, Gao *et al.* (2010) ran a within-subjects experiment (each subject ran each condition) and found significant differences between the Wolfpack condition and each of the other conditions. (For each condition $s \approx 9$ and $r \approx 0.5$.) Suppose you want to repeat the experiment using a between-subjects design (each subject runs only one condition). Using the means and standard deviation from Gao *et al.* (2010), run a power analysis for an independent ANOVA and three contrasts that compare the Wolfpack condition to each of the other conditions. Use the on-line app at

<https://introstatsonline.com/chapters/calculators/OneWayANOVAPower.shtml>

- What power would the new experiment have to get all four tests to produce a significant result if you used the same sample size $n_1 = n_2 = n_3 = 12$ as Gao *et al.*?
- What sample size is required to have 90% power for the ANOVA and all of the contrasts? Show a screenshot of the calculator.
- Why is the required sample size larger than what was used by Gao *et al.* (2010)?

(5) (15 points) In a paper about superstition, Damish, Stoberock and Mussweiler (2010) reported that people performed better on a variety of tasks if they were given an object with “good luck.”

- a) In experiment 1, 28 participants were randomly assigned to a superstition-activated condition or to a control condition. In the superstition-activated condition, subjects were given a golf ball and told that it is a “lucky ball.” In the control condition, subjects were told that the ball has been “used by everyone.” With the given ball, each subject then performed 10 golf putts, and the number of successful putts was counted. The average success rate was reported as $\bar{X} = 6.42$ for the lucky ball condition and as $\bar{X} = 4.75$ for the control condition. The description of the experiment does not indicate how many subjects were assigned to each condition. Use the GRIM test to figure out how many subjects must have been assigned to each condition. [Hint: Use the Mini-GRIM test Excel file that can found on the class web site. Play with different values of n_1 and compute n_2 . Both means must satisfy the GRIM test with sample sizes that sum so that $n_1 + n_2 = 28$.]
- b) In experiment 4 of the same article, subjects brought a “lucky charm” with them to the lab. They performed an anagram (word scramble) task. Some of the subjects had their lucky charm with them while doing the anagram task and other subjects did not have their lucky charm with them while doing the task. Performance was the number of correctly identified words in the anagram task. The mean number of words identified was higher for subjects with their lucky charm, $\bar{X} = 45.84$ compared to subjects without their lucky charm, $\bar{X} = 30.56$. There were a total of 29 participants, but the article did not report how many subjects were in each condition. Use the GRIM test to either figure out how many subjects must have been assigned to each condition or determine that the means are impossible given the sample sizes. Justify your conclusion.

(6) (20 points) A paper by Xu, Schwarz & Wyer, Jr. (2015) reported that hunger not only motivates people to consume food but also motivates people to acquire nonfood objects. The table below reports the statistical tests used to support their conclusion (no reported tests went against the conclusion).

	Test	p -value
Study 1	$t(67) = 2.01$	0.05
Study 2	$F(1, 75) = 4.55$	0.04
Study 3	$t(86) = 2.28$	0.03
Study 4	$F(1, 59) = 6.24$	0.02
Study 5	$t(77) = 2.11$	0.04

- a) Run a Test for Excess Success on the reported results. Interpret the results of the test. [Hint: Use the ComputePower.R code provided on the class web site.]
- b) Use the on-line app at
<http://shinyapps.org/apps/p-checker/>
to analyze the same statistics using the p -curve analysis.
- c) Do the two analyses give a similar interpretation of the validity of the results reported by Xu *et al.*?