Psychology 205, Revelle, Fall 2014
Research Methods in Psychology
Mid-Term

Name: _____

1. (2 points) What is the primary advantage of using the median instead of the mean as a measure of central tendency?

It is less affected by outliers.

2. (2 points) Why is counterbalancing important in a within-subjects experiment?

Ensuring that conditions are independent of order and of each other. This allows us to determine effect of each variable independently of the other variables. If conditions are related to order or to each other, we are unable to determine which variable is having an effect. Short answer: order effects.

3. (6 points) Define reliability and compare it to validity. Give an example of when a measure could be valid but not reliable.

2 points: Reliability is the consistency or dependability of a measurement technique. ["Getting the same result" was not accepted; it was too vague in that it did not specify the conditions (e.g., the same phenomenon) in which the same result was achieved.]
2 points: Validity is the extent to which a measurement procedure actually measures what it is intended to measure.
2 points: Example (from class) is a weight scale that gives a different result every time the same person stands on it repeatedly. Another example: a scale that actually measures hunger but has poor test-retest reliability. [Other examples were accepted.]

4. (4 points) A consumer research company wants to compare the "coverage" of two competing cell phone networks throughout Illinois. To do so fairly, they have decided that they will only compare survey data taken from customers who are all using the same cell phone model - one that is functional on both networks and has been newly released in the last 3 months. They also plan to focus their analysis only on actual phone calls (not text messaging or other functionality). Assuming that all technological variables are controlled for, will this provide an accurate comparison of network coverage for the state? Explain.

Example answer: Customers who are using a newly released cell phone may not be representative of the state's population of cell phone users – may be more likely to live in (or outside of) urban areas, may be early adapters who have usage patterns

dramatically different from the general population, may include a higher percentage of users who have recently switched networks based on coverage, etc.

In addition, the limited focus on phone calls is a large confound as many users rely on their phone for texts and other non-voice functions exclusively. Missing data about coverage for these services is likely quite important.

[Other well-argued points were accepted.]

5. (6 points) Name or describe the three criteria that separate science from pseudoscience.

2 points: Systematic empiricism
2 points: public verification
2 points: solvability
[You can find the definition of each of these in the book.]

6. (6 points) Compare *a priori* and *post hoc* theories. State which is preferred and why.

2 points: A priori theories are theories that may be informed by past research, but are created before the data are collected or examined.
2 points: Post hoc theories are made after the fact, e.g. after analyses have already been run and examined.
2 points: A priori theories are preferred because post hoc theories are contaminated by the particular result of the study. It is easy to retrospectively explain any results with a post hoc explanation.

7. (4 points) Explain why positive findings do not prove theories, and null findings do not disprove theories.

2 points: It is invalid to prove the antecedent of an argument (the theory) by affirming the consequent (the hypothesis). Many hypotheses could support the data you just found.
2 points: Although it is logically possible to disprove a theory, practically it is impossible due to error-- poor measuring techniques, biased sample of participants, etc.

8. (2 points) Cohen's d, correlation, and odds ratio are all ways of measuring the strength of the relationship between two or more variables. What is the general term that encompasses all these measurements?

Effect Size

9. (6 points) Describe what an interaction effect is. Draw an example of an interaction, using a line graph (and make up the variables). Make sure to label your graph.

2 points: An interaction between two or more IVs is present if the effect of one IV is different under one level of another IV than it is under another level of that IV. [Many students got this wrong because they said something to the effect of "when two IVs have a relationship with each other." This answer is incorrect because it is not specific enough. Two IVs can have a relationship (e.g., a positive correlation) without interacting.]
2 points: Basic interaction graph drawn
2 points: Graph properly labeled

10. (6 points) There are four different levels/scales of measurement. List them, ordering them from the least to the most informative. Circle the scale of measurement that reflects this type of ordering.

Nominal, Ordinal, Interval, Ratio
(one point each for names, one point for ordering, one point for circling Ordinal)

11. (4 points) What is a confound and how is it different from error?

2 points: Confound- when a variable besides the independent variable differs between groups and thus affects the dependent variable. It is systematic and could conceivably invalidate a study.
2 points: Error is unsystematic differences in participants, measurements, etc. It does not invalidate research, although you do want to minimize it. It can be considered as "static noise."

12. (6 points) Briefly describe deductive reasoning and inductive reasoning. In the process of iterative research, how do they inform one another?

2 points: Deduction- reasoning from a general proposition to a specific implication of that proposition; In research, hypotheses are often deduced from theories.
2 points: Induction- reasoning from specific instances to a general proposition about those instances; for example, hypotheses are sometimes induces from observed facts.
2 points: In research, deductive reasoning determines the questions you ask by formulating theories into more narrow hypotheses, then research is performed, and data are analyzed. Inductive reasoning helps you make sense of results and draw new theories and hypotheses that are further refined by deductive reasoning, starting a new iteration of research.

13. (8 points) Briefly describe each type of research: descriptive, correlational, experimental, and quasi-experimental.

2 points: Descriptive- describes phenomena, usually using descriptive statistics (frequency, count, mean, standard deviation, etc.), without looking at complex interactions.
2 points: Correlational- uses more advanced statistics (correlations, regression, ANOVA, etc.) to determine the relationship of two or more variables; does not attempt to manipulate variables.
2 points: Experimental- A standard experiment manipulates at least one independent variable to see its effect on a dependent variable. Experiments are highly controlled.
2 point: Quasi-experimental: Like experimental, but not as controlled; often is used in "natural" settings where it's difficult to control all variables.

14. (2 points) Data = _____ + _____

1 point: Model
1 point: Error/Residual

15. (8 points) A researcher interested in language and intelligence found that Northwestern students who had taken Latin in high school got better grades during their freshman year than Northwestern classmates who had taken French or Spanish. The researcher concluded that Latin helps students do better in college.

a) What kind of research is this?

2 points: Correlational

b) List three weak points in the researcher's study and/or conclusion

2 points: Correlation does not imply causation
2 points: There are any number of confounds that could account for this relationship, for example maybe students who took Latin were more driven
2 points: The study is not long term enough--better grades freshman year does not mean better grades for the next three years
[Other well-argued points can supplement these three]

16. (8 points) Two labs independently run a study on the test-retest reliability of a self-report scale that is supposed to measure hunger. One lab finds the scale to have a high reliability (.80), while the other finds the scale to have low reliability (.20). Assuming no error in recording the data, give a thoughtful explanation as to why the two labs could have arrived at such different answers. (It's possible there is more than one explanation. Give one). Also, make one suggestion on how to structure both studies to increase the likelihood of similar results.

4 points: thoughtful explanation. Example: Time of day could have an effect on outcome. One lab tested subjects during the similar time of day, while the other tested before and after lunch. Inter-rater reliability would also be acceptable. [Other well-reasoned answers were accepted.]
4 points: meaningful suggestion. Example: To get similar results, they should test and retest at the same time of day. Also, Bill mentioned spreading out participants by time of day would increase reliability (but of course you need to keep t1 and t2 for each participant the same).

17. (6 points) Disgruntled parents claim that a high school psychology exam is biased against male students because they consistently score lower on the test than female students. As a researcher, describe how you would determine whether there is systematic bias on the test or whether there is a true performance difference.

6 points: Example: look at each gender's correct rate on each question to see if there are certain questions that have big gender differences. If so, examine those questions to see if they are somehow biased in subject-matter against males. Summary of another acceptable answer: give the test to a completely psychology-naïve sample; if you still see score differences by gender, you probably have a biased test since none of the sample has any prior psychology knowledge.
[Students often missed this question because they failed to address the biased nature of the TEST. In order to get full credit, students needed to describe a procedure that would evaluate the test itself. Many students gave answers equalizing study time or giving the test to other samples, but this would not help if females were naturally more gifted at learning psychology. The psychology-naïve example was the only particular case that would work without examining the test, since it eliminates the issue of females being more gifted at learning psychology.]

18. (12 points) A psychologist was interested in developing a test that would predict the success of prospective lawyers. She selected a random sample of lawyers in a fashion website's list of "Top 1,000 American Lawyers," under the assumption that they would be successful lawyers. She then contacted them by means of a mail questionnaire that contained several hundred questions. The results from the questionnaires that were returned were analyzed and a profile of successful lawyers was compiled. The questionnaire was given to a group of prospective law students, and those students whose scores were significantly divergent from the successful lawyer group were advised not to pursue a law career.

a) Identify three weak points in this design.

2 points: Missing value problem (not all returned)
2 points: Need to know how the ones from successful differ from the less successful.
2 points: In addition, we don't know characteristics of law students related to success – for instance, successful lawyers are older than law students, does age lead to success
(Other well-argued issues can supplement these three)

b) Give three suggestions on how to redesign this study.

2 points: Do a prospective, longitudinal study, starting when participants are law students and tracking future success
2 points: Try to find unsuccessful lawyers as well
2 points: Try to find out characteristics of these lawyers when they were students
(Other well-argued suggestions can supplement these three)

19. (6 points) An employee at a certain cafeteria wanted to show his boss that the policy of serving "exotic" juices with breakfast was a wasteful practice. He had a theory that people just bought the exotic juice to impress their friends and that they did not really like the taste. In order to gather evidence, the man recorded the flavor of the juice contained in the glasses of those patrons who failed to finish their drinks. At the end of the morning he had counted a total of 50 less than completely empty glasses of juice. Of these, 40 were of the "exotic" variety and only 10 were of the non-exotic variety (e.g., good old orange juice). Because 80 percent of the unfinished juices were of the exotic variety, the employee concluded he had solid evidence of his theory. Provide one possible, realistic explanation as to why the employee is wrong about his conclusion.

6 points: Maybe there were more people who ordered exotic juice to begin with, therefore the chance that people did not finish exotic juice was higher than people who did not finish orange juice. Or there is a subject variable: people who liked orange juice were more frugal and were more likely to finish their drinks, than people who ordered exotic juice. [Many other possible explanations.]

20. (10 points) A memory researcher wants to test the idea that classical music improves memory, while heavy metal hinders it. She asks subjects for their musical preferences and accordingly assigns them to three different groups, classical music, jazz, and heavy metal, each of size 30. She gives to each group the same list of 100 words, which subjects have to memorize for 10 minutes. Then she exposes the subjects to their corresponding kind of music for another 10 minutes, followed by recall test. Here is what she finds in the test results:

	Mean Word Recall	SD
Classical	49	4.2
Jazz	40	3.5
Heavy Metal	32	4.0

She concludes that classical music improves memories and should be used in education to facilitate learning.

a) What are the independent variable and the dependent variable?

1 point: IV: type of music
1 point: DV: mean word recall/memory

b) What type of research is this?

2 points: Quasi-experiment / experiment / Between-subjects / applied

c) List one probable confound.

2 points: Self-assignment, heavy-metal listeners may have lower memory capacity.
[Many other reasonable acceptable answers.]

d) Propose two changes in the design of this experiment that would lead to more credible results.

2 points: Random assignment
2 points: Control group
[Many other reasonable answers can supplement those two.]