

Psychology 205, Fall 2015
Research Methods in Psychology
Midterm #2 Key

Name: _____

Short Answer (15 questions):

1. Name and define the four traditional scales of measurement discussed in class. (4 pts)

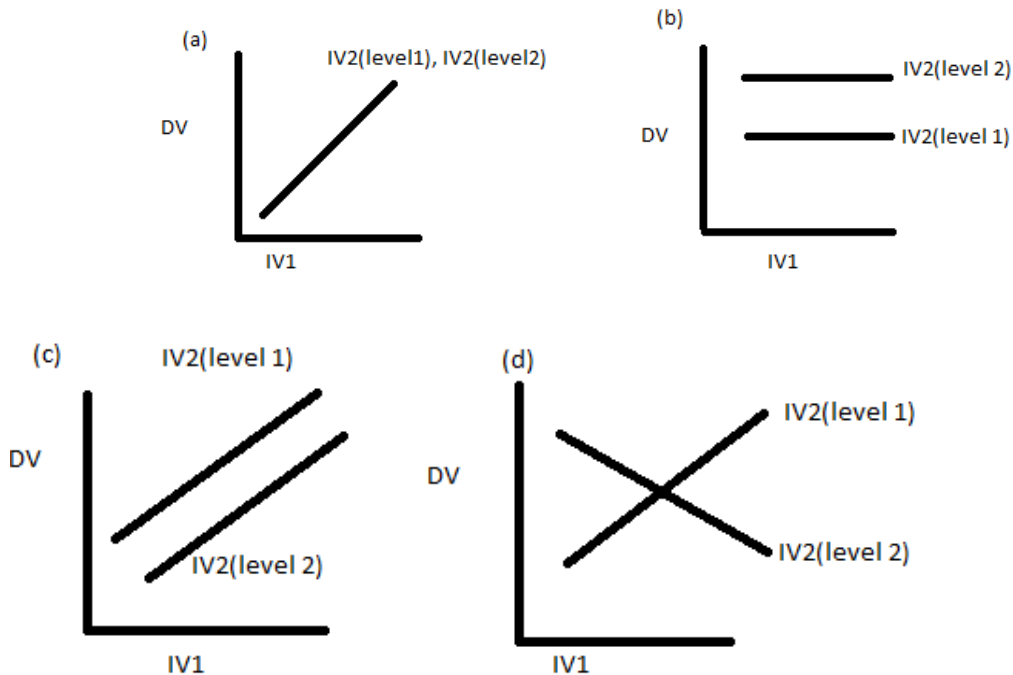
Nominal scales describe categories (e.g., X, Y, W, V)

Ordinal scales rank-order data (now you can say, e.g., that $X > Y > W > V$)

Interval scales allow us to compare the magnitudes of intervals, presupposing equal differences between each unit on the scale (so, e.g., we can now say that $X - Y > W - V$)

Ratio scales are interval scales that ALSO have a true zero point (we can say e.g., $X / Y > W / V$)

2. Imagine that you have two independent variables (IV1 and IV2) that you're using to predict some dependent variable (DV). Draw the following effects using the axes below: (a) A main effect of IV1, but no effect of IV2. (b) A main effect of IV2, but no effect of IV1. (c) A main effect of IV1 and a main effect of IV2. (d) An interaction between IV1 and IV2. (4 pts)



4 pts if you got all four drawings right (with or without labels)

2.5 pts if you got 2 drawings right and mislabeled the graphs

1.5 pts if you got 0 drawings right but put two lines on graphs b-d

1 pt if you got 0 drawings right and only put one line on each graph

3. What is the relationship between what statisticians call “alpha” and “beta” and Type I and Type II errors? What is the formula for power in terms of alpha, beta, or both? (3 pts)

Alpha = the probability of making a Type I error

Beta = the probability of making a Type II error

Power = 1 - Beta

4. In question 2, you drew some examples of additive and interactive relationships in multivariate research designs. What are the three types of relationship that were discussed in class for *univariate* research designs? Draw an example of each on the axes provided. (4 pts)

Linear

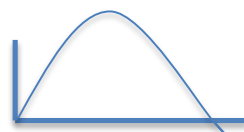


Monotonic



(with or without inflection was fine here)

Non-monotonic
(or quadratic)



5. In class, we discussed three types of deceptive scientific misconduct. What were these, and what did they involve? (4pts)

(1 point for each of the three on the list; 1/3 of a point each for correct definitions)

Fabrication: Making up data or results

Falsification: changing or misreporting data or results

Plagiarism: using the ideas or words of another person without giving appropriate credit

6. What is epidemiology, and what kind of design does epidemiological research typically employ? Discuss at least one threat to the validity of most epidemiological research. Does using random assignment in such a study help much? (4 pts)

1 point: Epidemiology = the branch of medicine that deals with the incidence, distribution, and possible control of diseases and other factors relating to health

1 point: Most epidemiological research uses observational/correlational designs

1 point: Mention one of the following subject variables: health, compliance/conscientiousness, or prescribing effect

1 point: Using random assignment gets rid of one potential threat to validity, but it doesn't negate the effect of subject variables like compliance.

7. (2pts) What are the benefits of using a correlation instead of a regression?

[1] The units of a correlation are standardized, unlike a regression.
 [2] A correlation is not as biased an estimate as a regression.
 [3] A correlation is prediction agnostic, whereas a regression implies that X predicts Y.
 [4] A correlation is also a ratio of data/model, unlike a regression.

CIRCLE ONE OF THE BELOW.

- a. 1 and 3
- b. 2 and 4
- c. 3 and 4
- d. 1, 2, and 3
- e. 1, 3, and 4
- f. 2, 3, and 4

8. What were Bill's three "ethical guidelines" that were meant to summarize all of research ethics? (2 pts)

Be honest.
 Be fair.
 Be accurate.

9. Name and define the four types of validity (that a scale can have) that were discussed in class. (4pts)

Face validity: does the scale *look like* it measures what it's supposed to measure?

Concurrent validity: does the scale correlate with a *current* criterion measure (other individual differences that are measured at the same time as the scale is administered)?

Predictive validity: does it correlate with a *future* criterion measure (other individual differences that are measured after time has been allowed to elapse following the initial scale administration)?

Construct validity: General type of validity that describes whether the measure correlates as it should with other measures (i.e. it correlates with other measures that claim to measure the same thing, convergent validity; and doesn't correlate with other measures that claim to measure different things). Okay (but not necessary) to discuss incremental validity here, too.

10. Give two examples of “trait” subject variables and two examples of “state” subject variables. Which kinds of experimental designs are most affected by state subject variables, and what method is normally used to control for the latter? (6 pts)

2 points for any two of the following “trait” variables: ability, age, prior practice/skill, any specific trait (e.g. anxiety, extraversion, etc.), other intelligent responses

2 points for any two of the following “state” variables: prior practice/order/learning, interest, fatigue, any specific state (e.g. anxiety, arousal, mood, etc.), other intelligent responses

1 point: Within-subjects designs frequently fall prey to state effects

1 point: Counterbalancing helps control for state effects

11. List 5 of the 8 elements that an informed consent form must include. (5 pts)

Give 5 of the following 8 (paraphrasing is fine):

1] Purpose of the research, expected duration, and procedures;

2] Notification of participants’ right to decline to participate and to withdraw from the research once participation has begun;

3] Consequences of declining or withdrawing;

4] Reasonably foreseeable factors that may be expected to influence subjects’ willingness to participate (e.g., risks, discomfort, other adverse effects)

5] Any prospective research benefits;

6] Limits of confidentiality;

7] Incentives for participation;

8] Whom to contact for questions about the research and research participants’ rights.

12. What is the advantage of looking at subject variable interactions rather than main effects? (2 pts)

Apparent main effects of individual differences can be hard to interpret because it’s difficult to tell whether the effect is due to the individual difference of interest or to other individual differences. Adding another IV may allow you to disconfirm the original finding, thus eliminating certain alternative explanations (e.g., either the individual difference of interest or other individual differences).

13. How is Item Response Theory conceptually different from Classical Test Theory? (Hint: what factors does it take into account?) (5 pts)

IRT takes into account “item difficulty” and “ability of the test-taker” and is better than Classical Test Theory (4 points). *One additional point for mentioning either greater efficiency of IRT, avoiding floor and ceiling effects, or issues it poses for anxious test-takers.*

14. A psychologist finds in her sample that a parent’s IQ predicts their child’s IQ, with a regression coefficient of 0.70 that has a $p < 0.05$. The psychologist concludes that because the slope is significant, there’s at least a 95% chance that a parent’s IQ really does predict their child’s IQ by the specified amount (0.70). Describe what is wrong with the psychologist’s conclusion about the meaning of the p value. (4 pts)

The p value tells you the likelihood (in this case, a 5% chance) that you would get your regression slope (in this case, 0.70) by chance; it does not tell you the likelihood that the regression slope actually predicts the outcome variable by the specified amount.

15. Name and define (or give a good example of) four of the five types of investigator effects discussed in class (4 pts)

Four of the five below:

Paradigm effects: Investigators may fail to see effects that don’t fit a current way of thinking

Design effects: Designs fail to pit theory against theory and/or seek disconfirmation

Loose procedure effects: Research procedure is vague and poorly specified

Data analysis effects: P-hacking, HARKing

Fudging effects: Scientific fraud, scientific misconduct

Long Answer (5 questions):

16. (10 pts) A researcher is interested in the effects of ambiguity and number of bystanders on helping behavior. Subjects fill out a questionnaire in a room with zero or two other people who appear to be other subjects but aren't. The experimenter distributes the questionnaire and then goes into the room next door. After 5 minutes, there is a loud crash, possibly caused by the experimenter falling. For half of the subjects, the experimenter unambiguously calls out that he has fallen, is hurt, and needs help. For the remaining subjects, the situation is more ambiguous—the experimenter says nothing after the apparent fall. There are 20 subjects tested in each condition and the number who go to help the experimenter are as follows:

0 bystanders; ambiguous	14
2 bystanders; ambiguous	8
0 bystanders; unambiguous	18
2 bystanders; unambiguous	18

Identify the independent variables and the levels of each, as well as the dependent variable. (2 pts)

IVs: number of bystanders, ambiguity or not; DV: number who go to help

If you put something ambiguous like "helping behavior" for DV, I subtracted ½ pt

If you said DV was "whether or not the subject helps" or "proportion of subjects who help" rather than the total number who helped, I subtracted ¼ pt

If you only gave IVs or only DV, then you only got 1 pt for this part

Place the data into the correct cells of a factorial matrix. (2 pts)

	Ambiguous	Unambiguous	Row Means
0 bystanders	14	18	16
2 bystanders	8	18	13
Column Means	11	18	

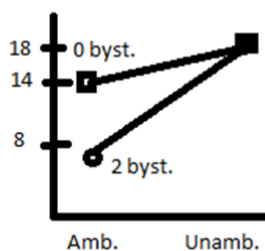
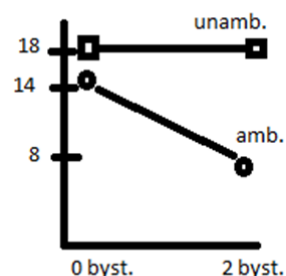
2 pts if you made this kind of matrix or if you put sum totals rather than column or row means

1 pt if you had no column or row sums or means

.5 pt for unclear or unlabeled table

Draw a graph of the results. (2 pts)

Either of these was acceptable (see notes at top of next page, too):



I took off one point if you failed to label 1 or both axes. I also took off one point if you graphed the row and column means rather than the DV numbers themselves, or if you made a bar graph.

I took off .5 points if your graphs or axes were in some way 'weird' (crossing without showing on the axes where the crossing occurred, e.g., or assuming 'ambiguity' was continuous)

Determine if main effects and/or interactions exist (Note: assume that a difference of 2 between any of the row or column means is a significant difference and that nonparallel = interaction.) (2 pts)

Two main effects, one interaction.

-1/2 point for each of the effects left off

Give a verbal description of each of the effects that occurs. (2 pts)

There are two main effects. More helping is offered when there are zero bystanders than when there are two bystanders ($16 > 13$). More helping is offered when the situation is unambiguous than when it is ambiguous ($18 > 11$).

There is one interaction. When the situation is not ambiguous, helping is the same and fairly high regardless of the number of bystanders ($18 = 18$), but when the situation is ambiguous, helping is more likely to occur with fewer bystanders ($14 > 18$).

2pts: if you described all three of these effects

1.5 pts: if you only described two effects

1 pt: if you only described one effect

17. An investigator was interested in the long-term effects of child sexual abuse. In a demographics questionnaire, participants were asked for their gender, ethnicity, age, and similar information. Questionnaires measuring history of sexual abuse, trait self-esteem, feelings of subjective well-being, substance abuse, sexual adjustment, anxiety, depression, interpersonal sensitivity, psychotic symptoms, hostility, and suicidal ideation were also included in the packet. All questionnaires were randomly ordered.

Experimenters were stationed at Norris, the library, and three busy walkways throughout campus. They asked every third person to fill out a packet of questionnaires and note the proportion of acceptances to all individuals asked. About 80% of students asked actually do fill out the questionnaires. Of that 80%, 70% complete the entire packet. All participants read and signed an informed consent form if they agreed to participate. They then filled out the questionnaires individually with either a black pen or pencil. Participants took 15-25 minutes to complete the packet. Upon completion, participants received a debriefing form in which the purpose of the study was explained.

After the data is entered and analyzed, the investigator finds that the effect of child sexual abuse on the other measures is very, very small. From this evidence, the investigator says that child sexual abuse does not cause long-term, intense, and pervasive harm.

a. (2 pts) What was the construct of interest?

--child sexual abuse

b. (2 pts) What is the design of the study?

--within subjects design (I also accepted "between subjects;" original key answer may be wrong)
This question was a bit unclear, so I gave 1.5 points for a wide range of other responses.

c. (2 pts) Is there anything wrong with this study?

--maybe sensitive nature of materials, attrition, subject variables

d. (2 pts) Was the researcher justified in making his final conclusion?

--overgeneralization; college folks are different from the whole population of sexual abuse victims (i.e., they are better adjusted and able to continue on with life)

18. For each of the following, write a sentence or short phrase describing at least one possible threat to internal validity that the study described might face. (2 pts each; 8 pts total)

(i) A college dean is upset about the so-called attrition rate—the percentage of freshmen who return to the college as sophomores. Historically, the rate has been around 75%, but in the academic year just begun, only 60% of last year's freshmen return. The dean puts a tutoring program into effect and then claims credit for its effectiveness when the following year's attrition rate is 71%. (2 pts)

Regression, History (1 pt: other reasons like no control group, no random assignment)

(ii) An educational psychologist develops a program to improve the reading skills of disadvantaged youth. She administers a reading test to 50 children, then uses their scores to form two groups with the same average score. The first group gets the reading program, but the second group does not. Both groups' reading ability is tested again when the program ends. (2 pts)

Regression (1.5 pt: history, maturation, selection bias; 1 pt: nonequivalent groups, no random assignment)

(iii) Two nearby colleges agree to cooperate in evaluating a new computerized instructional system. College A gets the program, and college B doesn't. Midway through the study, college B announces that it has filed for bankruptcy (however, it remains open and continues to collect data). One year later, computer literacy is higher at college A (2 pts).

Selection (no pre-test), history; 1 pt if you gave a vague response like 'non-equivalent groups'

(iv) Clinical psychologists at an institution for the criminally insane wanted to see whether playing certain low-frequency sounds (outside the normal range of hearing) over the loudspeakers would have a calming effect on patients. A number of aggressive behaviors were targeted and the frequency with which they occurred was measured on a monthly basis for three months. From the fourth through the sixth month, the sounds issued constantly from the loudspeakers; behavior frequency was assessed monthly. Between the seventh and twelfth month, the procedure described above was repeated. (2 pts)

Test sensitization, time-of-year effects, history, etc.

19. A pair of psychology graduate students had a bet on who could get published first in *The Journal of Personality and Social Psychology*. The first graduate student studied affective priming;

therefore, she ran a study in which she recorded participants' reaction times as they responded by pressing different computer keys in response to positively-valenced, negatively-valenced, and neutral words. Unfortunately, her study found no effect of emotional valence type on reaction time (RT); the mean RT was 500 ms for positive, neutral, and negative words.

The second graduate student was more interested in personality; he decided that the key to finding a significant effect would be to study the role of personality in determining reaction time. He administered the Neuroticism scale of the NEO PI-R to his participants, then recorded their reaction times in response to negatively valenced words. When he divided his participants into high, average, and low neuroticism groups, he found a significant main effect of neuroticism: high neurotics took an average of 400 ms to respond to the words, average neurotics took 500 ms to respond, and low neurotics took an average of 600 ms to respond. The second graduate student rejoiced, feeling sure that he would win the bet. His graduate adviser wasn't so sure.

What might be going on here? Could either of these studies be made better? How? (8 pts)

Neither Graduate Student 1 nor Graduate Student 2 has considered the possibility that a person \times situation interaction might be relevant to their variables of interest. If the two students joined forces to look at the effect of personality and word valence on reaction time, then they might get far more interesting results; for instance, hypothetically,

	<i>Positive Valence</i>	<i>Neutral Valence</i>	<i>Negative Valence</i>
<i>High Neurotic</i>	<i>600 ms</i>	<i>500 ms</i>	<i>400 ms</i>
<i>Average Neurotic</i>	<i>500 ms</i>	<i>500 ms</i>	<i>500 ms</i>
<i>Low Neurotic</i>	<i>400 ms</i>	<i>500 ms</i>	<i>600 ms</i>

Note that there is no main effect of valence, nor is there a main effect of personality when all word valences are considered together. However, there is an interaction between personality and valence.

8 pts: For a thoughtful answer that included consideration of a person \times condition interaction (among other acceptable means of improving the students' studies)

7 pts: For suggesting that the two studies could be 'combined' (among other acceptable means of improving the students' studies), even if no mention of interaction is made.

6 pts: For noticing that only negative words were used in the second study

5 pts: Other acceptable discussions of sources of error and potential improvement (with possibly a few errors in the details)

20. A number of studies suggest that people don't like to have their personal space invaded. A group of researchers wished to study this phenomenon more thoroughly by examining the physiological manifestations of personal space intrusions. One such physiological indicator concerns urination; under stress, people find it more difficult to begin urination or will stop urinating faster. The researchers predicted that in a men's room with three urinals, subjects standing at urinal 1 would experience greater personal space intrusion (therefore stress, therefore the physiological effects of stress on urination) if another person entered the room and stood at urinal 2 (next to urinal 1)

rather than at urinal 3 (one removed from 1). Rather than simply wait for these events to occur naturally, the researchers used 'confederates' who followed a subject into the men's room. If the subject was alone and standing at 1, the confederate stood either at 2 or 3 (or, in a control condition, didn't enter), as determined randomly. Subjects were unaware that they were participating in the study. The amount of time it took for the subject to begin to urinate and the total time spent urinating were recorded by a researcher hidden in one of the stalls using a periscope, stopwatches, and acute hearing.

What ethical issues are raised by this study (give at least two)? How could you make the study better? (9 pts)

No informed consent, invasion of privacy. A variety of improvements were acceptable.

9 pts for informed consent, privacy, and good discussion and improvements

8 pts if you got one or the other of informed consent and privacy, and were making an effort at getting at the one you didn't mention explicitly; discussions and improvements generally good

7 pts if you failed to mention one or the other of informed consent and privacy; discussions and improvements otherwise good.