EPI 203: *Epidemiologic Methods*
Problem Set 7: Understanding Measurement: Aspects of Reproducibility and Validity

Due: November 2, 2021 at 1:30 pm
Possible points: 30

NOTE: This Excel file already has data entered that you will need to answer the questions.
- Calculations are most easily performed using "functions" & graphing capacity of a spreadsheet (e.g., MS Excel or Google Sheets).
- Regarding spreadsheet functions for standard deviation, choose the function based on a sample with (n-1) in the denominator.
- A goal of this assignment is to develop some basic facility with spreadsheets such as MS Excel if you do not already have it.
- Type your text answers into the available text boxes to display your responses.
- Please be sure to review your final version to make sure all your responses are visible.

NAME: _____

1. You are asked to be a consultant for the development of a new at-home non-invasive needleless blood sugar (glucose) monitoring device. The company has presented you with some reproducibility data for the device. They have sampled 30 participants and performed 3 replicate measurements on each patient, one replicate right after the other, which are shown below. (Clinical note: The normal range for blood glucose in non-diabetics is about 70 to 110 mg/dl; differences less than 10 mg/dl are not considered clinically significant. In diabetic patients, blood glucose levels can be as high as 500 to 700 mg/dl.)

Participant	rep#1	rep#2	rep#3
1	95	96	91
2	142	145	139
3	168	171	167
4	123	126	123
5	110	113	107
6	115	116	113
7	122	125	118
8	148	148	147
9	105	108	102
10	155	159	154
11	168	169	163
12	188	188	187
13	118	120	117
14	114	118	113
15	134	137	130
16	139	143	137
17	149	150	147
18	130	133	125
19	128	129	125
20	124	129	120
21	130	131	130
22	140	141	136
23	142	144	141
24	137	141	135
25	141	145	140
26	105	107	104
27	180	182	180
28	178	182	177
29	108	108	107
30	110	113	107

a. Determine the within-person standard deviation for each participant. (2 pts)

b. Plot the within-person standard deviation versus the mean for each person. Be sure to label the axes.
Insert plot below. Comment on any relationship between the within-person standard deviation and the mean. (2 pts)

Insert plot here:

Comment here:

c. Determine the "common" (or mean) within-person standard deviation. Explain or show calculation. (1 pt)

d. The difference between any individual single measurement of glucose on a participant and the value that would be obtained if you could take the mean of numerous replicates on the participant (i.e., the participant's true value as measured by the device) would be expected to be less than what value for 95% of all individual single measurements? (Note: Assume that any requirements for a normal distribution are met.) (2 pts)

e. The difference between any two measurements on the same individual would be expected to be less than what value for 95% of all pairs of measurements? (2 pts) (Again, assume any requirement for normal distribution is met.)

f. Is the coefficient of variation a useful measure of reproducibility here? How about the geometric within-person standard deviation? Explain your answers. (2 pts)

g. The following Stata output used the above reproducibility data to calculate the ICC of the device. State the ICC. State in words what the ICC means. (1 pt) (Note: In this analysis, although 3 replicates were available for reproducibility testing, the device in practice is defined as producing just one determination. In other words, the final answer for each patient is obtained from just one run of the device.)

```
. loneway Glucose Subject

                  One-way Analysis of Variance for Glucose:

                                             Number of obs =        90

    Source                SS         df      MS            F     Prob > F
------------------------------------------------------------------------------
Between Subject         50978.056     29     1757.864    271.37     0.0000
Within Subject            388.66667     60     6.4777778
------------------------------------------------------------------------------
Total                           51366.722     89    577.15418

         Intraclass
         correlation     Asy.
         coefficient     S.E.       [95% Conf. Interval]
         ---------------------------------------------------
            0.98903     0.00348       0.98220     0.99585
```

2. These same 30 participants also had blood drawn, with conventional phlebotomy methods, during the session when the measurements with the new device were taken. The blood glucose level taken from the conventional phlebotomy was measured at the UCSF Hospital Clinical Laboratory — with one assay — as shown below:

Participant	UCSF glucose
1	96.0
2	139.0
3	168.0
4	126.0
5	112.0
6	116.0
7	121.0
8	149.0
9	106.0
10	157.0
11	166.0
12	188.0
13	117.0
14	116.0
15	134.0
16	140.0
17	149.0
18	129.0
19	128.0
20	124.0
21	133.0
22	139.0
23	142.0
24	137.0
25	143.0
26	104.0
27	181.0
28	180.0
29	108.0
30	110.0

a. Examine the difference between the UCSF Clinical Lab value (consider it the gold standard) and the new device value with a Bland-Altman plot. Insert plot below. (1 pt) Does the difference between the new method and the gold standard appear constant over the range of values? (1 pt) (Note: Use the mean of the three replicates of the new device as the "final" result for the new device.)

Insert plot here:

Comment here:

b. Determine the numerical bias associated with the new device as compared to the gold standard. (Note: Again, use the mean of the three replicates of the new device as the "final" result for the new device.) (1 pt)

c. In comparison with the UCSF blood glucose measurement, determine the 95% limits of agreement for the new device. Note: Again use the mean of the three replicates of the new device as the "final" result for the new device. Also, assume normal distributions. (1 pt)

d. Does this new device (operationalized as using the average of 3 replicates as the final value) thus far appear promising for use in research? Explain your answer. Does the new device thus far appear promising for use in individual patient measurement and management (i.e., clinical practice)? Explain your answer. If you could collect additional data, what would you want to know? (3 pts)

e. What if the bias (the entity referred to in part b) was -40 units? Without any major physical re-engineering of the device, what else could be done to make it clinically useful? (1 pt)

f. What if you wanted to use this new device to measure blood glucose in a new research study where a new drug is being investigated for toxicity in already well managed patients with diabetes (blood glucose below 200). Blood glucose is one of the outcomes. While each replicate of glucose measurement is not that expensive, money for the study is limited.
Would you take one measurement per person or would you take 1 or 2 additional replicates? Explain your answer. (1 pt)

3. Examine the following data from Costantini et al. *J. Immunological Methods.* 278:145-55, 2003.

Cryopreservation (i.e., freezing) is the most commonly used procedure to store peripheral blood lymphocytes for prolonged periods of time. To determine the effect of prolonged cryopreservation on the immunophenotype of peripheral blood lymphocytes, we performed a comparative analysis of fresh blood versus cryopreserved blood in a group of 19 healthy individuals. Each individual had fresh cells tested prior to freezing and then had their cells thawed after 2 months of being frozen at -80° C. Cells were tested for the proportion of lymphocytes which were CD4+ and CD8+.

As shown in the table, the authors conclude that no significant differences were observed following cryopreservation in the mean proportions of the two major lymphoid subpopulations, CD4+ and CD8+ cells (p=0.08 and 0.27, respectively).

	Fresh cells	Thawed cells after cryopreservation	p value (Student's T test)
CD4+ cells	44% +/- 6.4*	40% +/- 7.5	0.08
CD8+ cells	27% +/- 7.8	30% +/- 8.9	0.27

*values are expressed as mean +/- standard deviation in 19 individuals

The authors concluded that cryopreservation of cells was equivalent to the testing of fresh cells for the enumeration of CD4+ and CD8+ cells. Do you agree with their conclusion? (1 pt) What other analysis would you perform if you wanted to know if cryopreserved cells could be substituted for fresh cells in research regarding CD4+ and CD8+ cells? (1 pt)

4. Consider the following abstract:

Objective

To determine the ICC of a self-report questionnaire (the Adolescent Sedentary Activities Questionnaire; ASAQ) which assesses total time spent per day in sedentary activities, among school-aged children.

Methods

Two-hundred and fifty school students aged 11–15 years from four primary and four high schools in Australia completed the questionnaire under the same conditions on two separate days, 2 weeks apart.

Results

Intraclass correlation coefficient (ICC) for time total spent in sedentary behavior were ≥ 0.90 for all students, except for Grade 6 boys (in which the ICC = 0.57, 95% CI: 0.25 to 0.76). Reproducibility was generally higher on week days compared with weekend days.

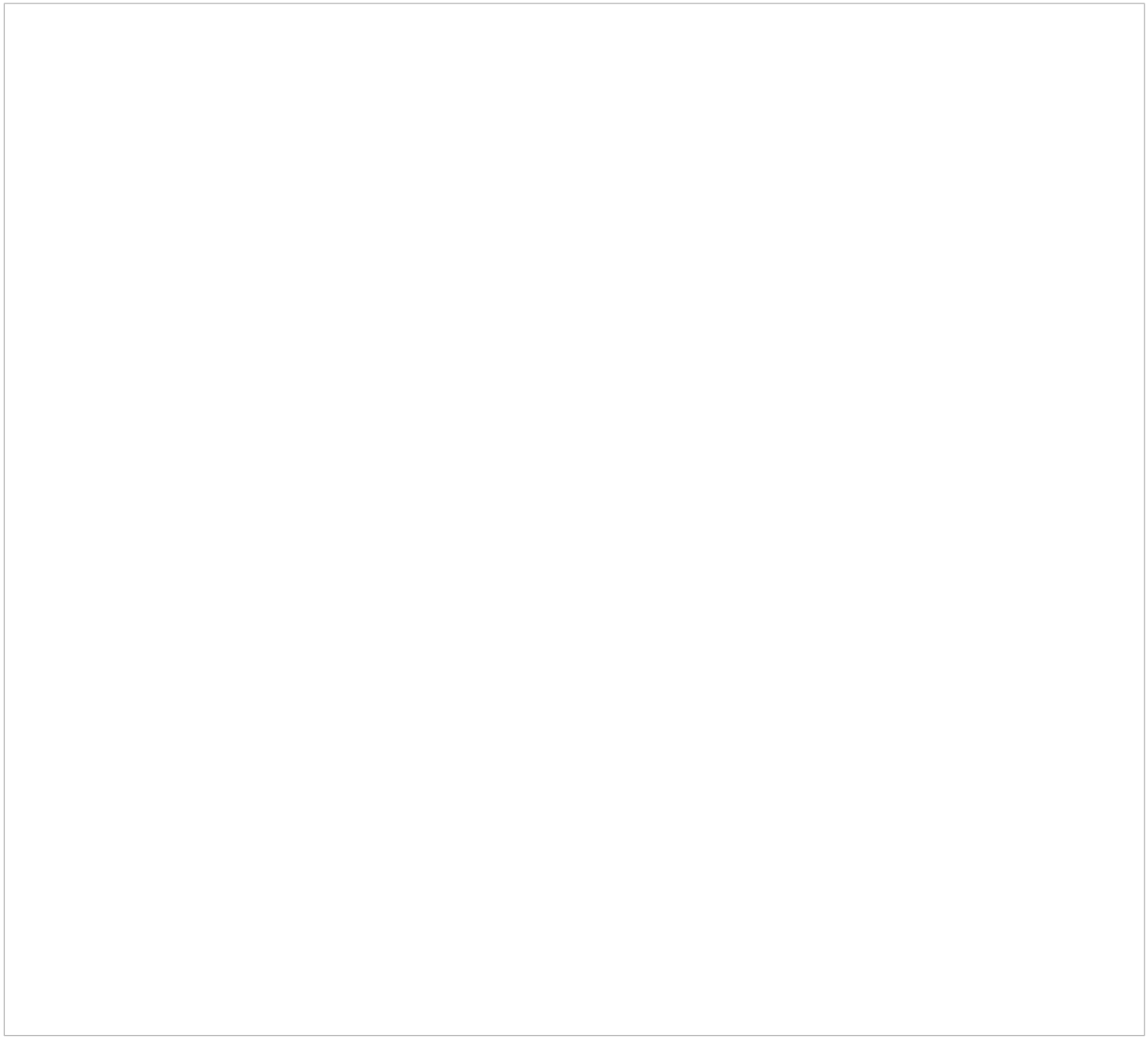
a. Describe what the intraclass correlation coefficient for Grade 6 boys means. (1 pt)
What concerns would you have, if any, about any inferences derived from the ASAQ for Grade 6 boys if you used the ASAQ as an exposure variable, measured on one occasion, in a research study about Grade 6 boys in Australia? (1 pt)

b. Name two reasons why the ICC in the Grade 6 boys might be so low? (1 pt)

c. You would like to use this same ASAQ measurement tool in a research study you are doing among 15-year old students in San Francisco. Is there any other information you would like to know before you use it "as is" (i.e., without any modifications)? (1 pt)

5. Assume that you work in a general primary care clinic, one of 5 such clinics in a linked network in San Francisco. Patients in the network are free to choose among the 5 different primary care clinics. Your supervisor has become concerned with "patient satisfaction" and has asked you to come up with a short questionnaire to administer to patients at your clinic. You create the following questions: "Are you satisfied with the sensitivity of your primary care provider?"; "Are you satisfied with the nursing staff?"; "Are you satisfied with the reception desk staff?"; and "Are you satisfied with the transportation options to the clinic?". (Response options include: very satisfied, somewhat satisfied, somewhat dissatisfied, and very dissatisfied.)

Unfortunately, there are no concurrent gold standards for "patient satisfaction" to which to compare your measurement questions. Describe how you will perform content validation, criterion validation, and construct validation of your questionnaire using the definitions provided in Chapter 10 (Validity), pages 227 to 240, of the Streiner et al. textbook on health measurement scales. (3 pts)



6. FOR DISCUSSION IN SECTION ONLY
In the Journal Club article by Lane et al. (*Arth Rheum* 56:3319, 2007), the authors refer to "the coefficients of variation for Dkk-1" in the Methods section. Describe what the CV means and how it helps you to determine whether or not this measurement of DKK-1 is useful for research.

7. FOR DISCUSSION IN SECTION ONLY
Consider one or two of the main measurements in your research. It can be an outcome variable, exposure variable, or something else. What is known about the reproducibility of the measurement? What is known about the validity of the measurement? Can you identify any feasible actions which can improve either reproducibility and/or validity?