Answers to Items from Problem Set 1

Item 1

Identify the scale of measurement most appropriate for each of the following variables. (Use A = *nominal*, B = *ordinal*, C = *interval*, D = *ratio.*)

a. ___response latency (i.e., the elapsed time between being exposed to a stimulus and responding to that stimulus)

D (ratio), there is a TRUE zero point.

b. ___motivation measured by scores on the *XYZ Motivation Inventory*

C (interval), the trait being measured (motivation) is continuous, but there is no TRUE zero point. Usually scores on tests are assumed to be interval.

c. ___political party affiliation (Democrat, Republican, Independent, Other)

A (nominal or categorical).

d. ___academic rank in high school

B (ordinal), ranking implies ordinarily.

e. ___scores on the *SAT* or *GRE*

C (interval), again, scores on tests are usually assumed to be interval.

f. ___grade point average

Usually treated as C (interval), but probably makes better sense to treat it as B (ordinal).

g. ___responses to a Likert-type item.

B (ordinal)

h. ____the sum (or average) of responses to Likert-type items.

C (interval)

Item 2. Identify the independent and dependent variables in each of the following experiments.

	Independent	Dependent
a. students are taught statistics either with or without a textbook and test scores of the two groups are subsequently compared	Type of Teaching	Test Scores
b. ratings of self-confidence are found to be correlated with high school	This is a correlational study, there are no independent or dependent variables	
c. fifteen minutes a day practicing shooting 3 point goals proved more effective than thirty minutes of practice twice a week	Type of practice	Number of 3-point goals

Item 3

Use the following sets of data to compute the values requested

Data (X): 9 12 13 14 15 16 16 17 18 20

Data (Y): 2 1 0 -1 2 2 -2 1 0 1

Answer

$\Sigma X = 150$
$\Sigma X^2 = 2340$
$(\Sigma X)^2 = 22500$
$\Sigma X + 2 = 152$
$\Sigma(X+2) = 170$
$\Sigma(X+2)^2 = 2980$
$\Sigma 2X = 300$
$\Sigma 2X + 2 = 302$
$\Sigma 2(X+2) = 340$
$\Sigma XY = 83$
$\Sigma XY + 2 = 85$
$\Sigma(X+2)(Y+2) = 435$

$\Sigma X/N = 15$

$\Sigma X^2/N = 234$
$\Sigma XY/N = 8.3$

$\Sigma X \Sigma Y = 900$
$\Sigma 6y = 36$
$5\Sigma y = 30$

Item 4

Using a table of areas under the normal curve, compute the following:

a. the proportion of cases below a z-score of -.72.
b. the proportion of cases below a z-score of 1.65.
c. the proportion of cases above a standard score of 2.0.
d. the proportion of cases above a standard score of -2.5.
e. the proportion of cases between z-scores of -1.5 and 1.5.
f. the proportion of cases between standard scores of 1.5 and 2.5.

To answer these, we need the **Table of Areas Under the Normal Curve** given in any introductory statistics textbook. Also, we need to remember that the table gives (1) the area (or **proportion**) of cases between the mean and the **absolute** value of a given standard score (**z-score**) and (2) the proportion of cases greater than the absolute value of a specified z-score.

a. In the table, the proportion of cases between the mean and a z-score of .72 is .2642. Since we are working, here, with a **negative** z-score we note, first, that the proportion of cases below the mean is .5. Since the proportion of cases between the mean and a z-score of -.72 is .2642, we subtract .2642 from .5, yielding .2358, the proportion **below** a z-score of -.72.

   A easier way to have obtained this answer is to use the area **beyond** the z-score. Of course, in this case, since z is negative, the area beyond gives the proportion **below** a z-score of -.72. Reading directly from the table, the area beyond a z-score of .72 is .2358. Hence the proportion of cases below a z-score of -.72 is .2348.

b. Here, we need to remember that the table only gives proportions for **half** the normal distribution. Looking up a z-score of 1.65 we find that the proportion between the mean and 1.65 is .4505. To this we must add the proportion **below** the mean (.5). Adding the two proportions gives .9505. Hence 95.05% of the cases in a normal distribution fall below a z-score of 1.65.

c. .0228

d. 9938 (*Hint: This is the sum of .4938 and .5)*

e. .8664 *(Hint: This is 2 x .4332)*

f. .0606 *(This can be obtained either by using the area between the mean and z,* [.4938 - .4332*] or the area beyond z, [*.0668 - .0062*].*

Item 5

Assume the mean on an IQ test is 100 with a standard deviation of 16. Using a table of areas under the normal curve, estimate each of the following:

To answer these, we need the formula for a standard (z) score. Recall that this is:

$$Z = \frac{(X - \overline{X})}{s}$$

Where

X = *the score,*

$\overline{X}$ = *the mean of the distribution, and*

s = *the sandard deviation.*

a. Find the percentage of individuals having an IQ of 130 or higher.

First compute the z-score

z = (130-100)/16 = 30/16 = 1.875 (*round to 1.88*).

Then look up the **proporton in tail** (Column C) for a z-score of 1.88. The proportion is .0301. Hence, about 3% have IQs of 130 or higher.

b. Find the proportion of individuals with an IQ of 90 or lower.

z = (90 - 100)/16 = -10/16 = -.625

According to the Unit Normal Curve Table (Appendix B), the proportion **in the tail to the left** (hence, below, Column C) a z-score of -.62 is .2676. Hence, a little less than 27% of the individuals have IQs lower than 90.

c. Find the percentage of individuals with IQs between 85 and 115.

The z-score for an IQ of 115 is .94 (*rounded from .9375*); the z-score for an IQ of 85 is -.94.

The area to the left (below) a z-score of .94 is .8264 (Column B).

The area to the left (below) a z-score of -94 is .1736 (Column C).

Hence the proportional area between -.94 and .94 is the *difference* between the two proportions (.8264 - .1736*)* or .6528.

d. Find the proportion of individuals with IQs between the first and third quartiles.

The first and third quartiles correspond the the 25th and 75th percentiles, respectively. Furthermore, percentiles are directly related to proportions (or areas) of a normal distribution: .25 (or 25%) of the distribution lies below the 25th percentile; .75 (or 75%) lies below the 75th percentile. Hence .5 lie between the 25th and 75th percentile.)

e. Find the percentage of individuals with IQs above 130.

An IQ score of 130 corresponds to a z-score of 1.875 (or 1.87).

Using Column C in the Unit Normal Curve Table we see that the proportion *above* a z of 1.87 is .0397.

Hence, approximately 4% of individuals have IQ scores above 130.

f. Find the IQ of individuals at the 85th percentile.

An individual at the 85th percentile has a *z*-score of 1.04 (Note: since the 85th percentile is *above the median* we use Column B and scan down until we find .8508. The corresponding to z-score is 1.04).

Now, solve the z-score equation for X:

$$X = Zs + \overline{X}$$

Then compute, X = 1.04(16) + 100 = 116.64, or, after rounding to the nearest whole score, an IQ of 117.

g. Find the IQ of individuals at the 35th percentile.

35% of the distribution lies **below** the 35th percentile. We can use the Table of Unit Normal Curve Table to determine that this corresponds to a z-score of -.38 (*using Column C*).

Next, using the formula above to solve for X, we obtain,

X = -.38(16) + 100 = 83.36, or an IQ of 94.

Item 6

A survey of 1,000 kindergarten children reveals that the average child watches 145 minutes of TV per week. The uncorrected *sample* standard deviation is 40 minutes. Construct the 95% confidence to estimate the mean number of minutes spent viewing TV per week in the *population* of kindergarteners.

We have the sample size (1000) and the sample standard deviation (40). To compute the confidence interval we need the **standard error of the mean (***SEM***)**:

$$\begin{aligned} SEM &= \frac{s}{\sqrt{(N-1)}} \\ &= \frac{40}{\sqrt{999}} \\ &= \frac{40}{31.607} \\ &= 1.266 \end{aligned}$$

The 95% confidence interval is then given by:

$$\begin{aligned} 95\%\, CI\!: \; & \bar{X} - 1.96(SEM) \leq \mu \leq \bar{X} + 1.96(SEM) \\ & \bar{X} - (1.96 \times 1.266) \leq \mu \leq \bar{X} + (1.96 \times 1.266) \\ & 145 - 2.48 \leq \mu \leq 145 + 2.48 \\ & 142.52 \leq \mu \leq 144.48 \end{aligned}$$

We conclude, 95% confidence, that the **true** mean number of minutes children spend watching TV per week is somewhere between 142.52 minutes and 144.48 minutes per week.

Item 7

a. Is it important that Schmalling first examine the *reliability* of her new instrument? Explain your reasoning.

Yes, it is important that she first establish the reliability the instrument. If the instrument is not reliable, then scores on the instrument reflect error (noise or random responding), in which case not valid inferences can be drawn.

b. What type of evidence for validity was Schmalling looking for? What results would provide evidence instrument is valid?

Schmalling was looking for construct (convergent-divergent) evidence of validity. She hoped for a reasonably strong correlation between scores on the new instrument and scores on the *Attitude Toward Mathematics* instrument (since the two instruments presumably assess similar characteristics) and a relatively weak correlation between scores on the new instrument and scores on the *Motivations for Reading* instrument (since attitudes and interests in reading is, presumably, independent of attitudes and interests in mathematics).

Item 8

The table below shows the number of students, classified by age and whether or not a parent had attended college, over a 10-year period, who attained an Associate's Degree from a local community college.

a. What is the probability of attaining an Associate's Degree for a student who is 28 years old?

b. What is the probability of attaining an Associate's Degree for a first-generation student who is 23 years old?

c. Of the six combinations (of age and parent attendance) which group has the highest probability of attaining an Associate's Degree?

Number of Students Attaining an Associate's Degree by Parent College Attendance Status and Age Category

	Age Category		
Status of at least one parent attended college	Less than 25 yrs old	25 to 35 yrs old	Greater than 35 yrs old
Parent attended college	1252	1215	632
No parent attended college	643	736	592

a. Probability of an Associate's Degree for a student 28 years old = 1951/5106 = .38.
b. Probability of attaining an Associate's Degree for a first-generation student who is 23 years old = 643/5106 = .13.
c. Highest probability of attaining an Associate's Degree = Less than 25 yrs old whose parents attended college: prob. = 1252/5106 = .256.

Item 9

A school district dietitian wants to construct a confidence interval to estimate the mean number of soft drinks high-school students consume daily in her district. She intends to survey the students. Assuming a *population* standard deviation of 1.8 soft drinks a day,:

a. how many students must she survey in order to obtain a 95% confidence interval that is .5 soft drinks wide?

b. how many students would she have to survey to obtain a 99% confidence interval that is .5 soft drinks wide?

A 95% confidence interval that is .5 wide is obtained when the interval,

$$\mu = \bar{X} \pm 1.96(SEM)$$

is .5 wide. This occurs when,

$$1.96(SEM)$$

is equal to .25 (half on either side of the mean). Hence, solving for the SEM gives

$$SEM = {}^{.25}\!/_{1.96} = .1276.$$

We use this value and the equation for SEM $\left(= \sqrt{\sigma^2/(n-1)}\right)$ to solve for N. We have,

$$SEM^2 = {}^{\sigma^2}\!/_{(N-1)}$$, or

$$.0163 = {}^{3.24}\!/_{(N-1)}$$. Solving for N

$$(N-1) = {}^{3.24}\!/_{.0163} = 198.77.$$

So that, when rounded up, *N* = 199.

A 99% confidence interval that is .5 wide is obtained when the interval,

$$\mu = \bar{X} \pm 2.58(SEM)$$

is .5 wide. This occurs when,

$$2.58(SEM)$$

is equal to .25 (half on either side of the mean). Hence, solving for the SEM gives

$$SEM = {}^{.25}\!/_{2.58} = .0969.$$

Using logic identical to that used for computing the *N* needed for 95% CI, we arrive at (*N*-1) = 344.681. Hence, we would need a sample size of 346.

Item 10

The distribution of SAT-V scores is assumed to be normal with μ = 500 and σ = 100.

a. What is the *probability* of someone having an SAT-V score *higher* than 650?

b. What is the *probability* of an SAT-V score *between* 550 and 650?

Probabilities are associated with areas under the normal curve. Think of the TOTAL area under the normal curve as being equal to 1. Then portions of this area give probabilities.

The distribution of SAT-V scores is normal with m = 500 and s = 100.

a. What is the *probability* of someone having an SAT-V score *higher* than 650?

First convert the SAT-V score of 650 to a *z*-score of 1.5.

Then, using Column C in the Unit Normal Curve Table, find that the proportion, .0668, lies *above* a *z*-score of 1.5. Hence, the probability of an SAT-V score above 650 in .0668 (or, after rounding, .07).

b. What is the *probability* of an SAT-V score *between* 550 and 650?

Again, using *z*-score transformations and the Unit Normal Curve Table, we find that a proportion of .9332 lies *below* a *z*-score of 1.5 and that a proportion of .6915 lies *below* a *z*-score of .5 (the *z*-score associated with an SAT-V score of 550).

The difference between these two proportionis (.9332 - .6915) is .2417. Since proportions can be read as probabilities, the probability of an SAT-V score *between* 550 and 650 is .24 (rounded).

Item 11

The correlation between SAT scores and first semester GPA in the general population is about .53. What do you suppose is the correlation between SAT and GPA at Harvard? Will it higher, lower, or about the same? Explain your reasoning.

The correlation will be lower (probably close to zero), because Harvard admits a highly selected group of freshmen. Most of them would be expected to have very high GPAs and to score very high on the SAT. Hence, Harvard's freshmen should exhibit a very narrow range (i.e., very little variance) of both SAT scores and GPSs.

The correlation between two variables, *x* and *y* is given by

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_x{}^2 \times s_y{}^2}}$$

where s_{xy} is the *covariance* between x and y, and $s^2{}_x$ and $s^2{}_y$ are, respectively the *variances* of x and y. Because of the highly restricted range of both x and y, the correlation between them is *attenuated.*

Item 12

The mean GRE score for a sample of 121 students who completed a GRE preparation course was 1030 with a corrected standard deviation of 175. The national mean GRE score is 1000. Is this difference statistically significant? Do you think the difference is practically significant? Explain your reasoning.

Compute a one-sample *t* test, where

t = (1030-1000)/SEM

SEM = $SD_{Corrected}$/sqrt(N)

= 175/sqrt(121)

= 175/11

= 15.9091

t = 30/15.9091

= 1.886

The critical value of t at α = .05 with 120 degrees of freedom is 1.98. Hence, the t is not statistically significant. Since t is not statistically significant, it cannot be practically significant.

13. Using the dataset below set up an SPSS database and run the following descriptive analyses: gender, ethnicity, and score. Also, create a graphic display (of your choice) that explains the data. Export all SPSS output into Word.

Be sure to label all values so the output is readable and clear. Data is listed in order for each subject (gender, ethnicity, grade).

1) 1, 3, 75

2) 1, 4, 98
3) 2, 6, 90
4) 1, 4, 85
5) 1, 6, 88
6) 2, 4, 90
7) 2, 1, 94
8) 2, 1, 83
9) 1, 1, 75
10) 1, 4, 70

Gender: 1=female; 2=male

Ethnicity: 1=African American; 2=American Indian; 3=Asian; 4=Caucasian; 5=Hispanic; 6=Multiracial

Answer:

Statistics

		gender	ethnicity	score
N	Valid	10	10	10
	Missing	0	0	0
Mean		1.40	3.40	84.80
Median		1.00	4.00	86.50
Mode		1	4	75[a]

a. Multiple modes exist. The smallest value is shown

gender

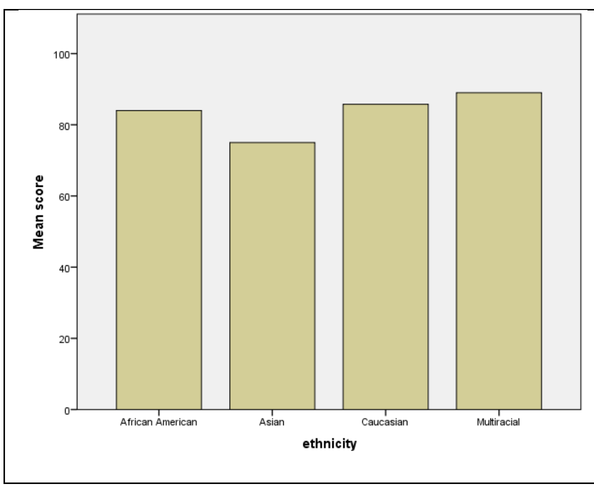
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	female	6	60.0	60.0	60.0
	male	4	40.0	40.0	100.0
	Total	10	100.0	100.0	

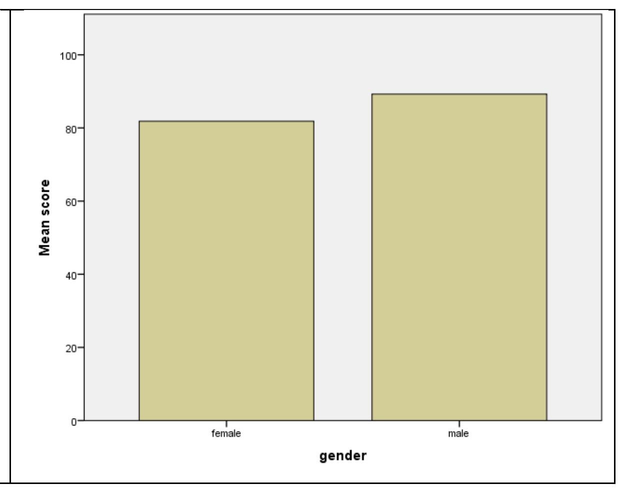
ethnicity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	African American	3	30.0	30.0	30.0
	Asian	1	10.0	10.0	40.0
	Caucasian	4	40.0	40.0	80.0
	Multiracial	2	20.0	20.0	100.0
	Total	10	100.0	100.0	

score

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	70	1	10.0	10.0	10.0
	75	2	20.0	20.0	30.0
	83	1	10.0	10.0	40.0
	85	1	10.0	10.0	50.0
	88	1	10.0	10.0	60.0
	90	2	20.0	20.0	80.0
	94	1	10.0	10.0	90.0
	98	1	10.0	10.0	100.0
	Total	10	100.0	100.0	





Item 14. Using the data set below, conduct an independent-samples t-test examining the differences between Cohort 18 (18) and Cohort 19 (19) on exam scores.

Is there a significant difference between the two groups? Produce SPSS output and write a p value statement to indicate your answer.

1) 99 18
2) 80 18
3) 65 18
4) 50 18
5) 88 18
6) 90 18
7) 80 18
8) 61 18
9) 77 18
10) 70 18
11) 90 19
12) 92 19
13) 85 19
14) 84 19
15) 71 19
16) 90 19
17) 60 19
18) 86 19
19) 83 19
20) 71 19

Answer:

The means of Cohort 18 and Cohort 19 did not differ significantly, p = 3.75.

Group Statistics

	Cohort	N	Mean	Std. Deviation	Std. Error Mean
ExamScores	Cohort 18	10	76.00	14.757	4.667
	Cohort 19	10	81.20	10.422	3.296

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
ExamScores	Equal variances assumed	1.076	.313	-.910	18
	Equal variances not assumed			-.910	16.189

Independent Samples Test

		t-test for Equality of Means		
		Sig. (2-tailed)	Mean Difference	Std. Error Difference
ExamScores	Equal variances assumed	.375	-5.200	5.713
	Equal variances not assumed	.376	-5.200	5.713

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
ExamScores	Equal variances assumed	-17.203	6.803
	Equal variances not assumed	-17.300	6.900

Item 15. Using the data set below, conduct a paired-samples t-test examining the differences between pre- and post-exam scores.

Is there a significant difference between the two groups? Produce SPSS output and write a p value statement to indicate your answer.

1) 9998
2) 8087
3) 6577
4) 5077
5) 8885
6) 9096
7) 8089
8) 6189
9) 7780
10) 70 90
11) 90 88
12) 92 98
13) 85 98
14) 84 98
15) 71 89
16) 90 98
17) 60 59
18) 86 89
19) 83 92
20) 71 90

Answer:

The means of Exam 1 and Exam 2 did differ significantly, $p < .05$, $p = .000$.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	ExamScores1	78.60	20	12.717	2.844
	ExamScores2	88.35	20	9.588	2.144

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	ExamScores1 & ExamScores2	20	.698	.001

Paired Samples Test

		Paired Differences			
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference
					Lower
Pair 1	ExamScores1 - ExamScores2	-9.750	9.136	2.043	-14.026

Paired Samples Test

		Paired Differences	t	df	Sig. (2-tailed)
		95% Confidence Interval of the Difference			
		Upper			
Pair 1	ExamScores1 - ExamScores2	-5.474	-4.773	19	.000