

Stats Support 6: Mid-term review

Practice questions

Will be taken up in Stats Support on Fri February 28th – at 1130 in W170 and 230 in LH101. Remember you do not need to complete the questions to attend this portion of stats support.

**note these review questions are similar in size, content, style and coverage to your midterm exam questions. We will use most of our time this week specifically reviewing and strategizing for the exam.*

1. We have measured anxiety levels in a random sample of 356 university students. Our anxiety scale ranged between 0 (no anxiety) and 25 (highest anxiety). We measured a mean anxiety score of 11.8 and a standard deviation of 5 in our sample. Assuming the distribution is normal, would someone who scored a 6 be estimated to be in the bottom 10% of all cases? (5)
What about if the standard deviation was 2? (5)

2. Here is data from a community centre looking at the number of hours per week put in by the most active volunteers. The hours per week for the top 10 volunteers are: 22, 14, 13, 17, 21, 11, 14, 20, 14, 29. Calculate the appropriate measure(s) of central tendency and dispersion and report. If possible, include discussion of potential skewness. (10)

3. Below is raw data from a study of Canadian municipalities looking at current public transit expansion plans. For each city we have information on what type of transit expansion is being planned next. Use this information to complete the questions/computations in this section.

1	add buses
2	bus rapid transit
3	light rail
4	bus rapid transit
5	subway line
6	add buses
7	subway line
8	bus rapid transit
9	light rail
10	light rail
11	light rail
12	add buses

What are the cases in this data? What is the variable in this data? At what level of measurement is this data? (2)

Create a frequency table of this data. Include percentage and if appropriate, cumulative percentage columns. (4)

Compute appropriate measure(s) of central tendency and dispersion and report. (4)

4. We are given information on a retraining program for displaced manufacturers. We have assigned each a score out of 10 to represent their success at retraining in the program. The workers had a mean success score of 4.91. We also have a subsample of workers ($n=22$) that have gone through a newly added requirement in the program. Their mean success score was 5.64 with a standard deviation of 0.81. Has the addition of the new requirement resulted in better success scores? (10)

5. We surveyed a random sample of 142 world class athletes at the IAAF world championships. We asked them to rank on a scale of 1 (low) to 7 (high) how prohibitive funding concerns are to their participation in competition on this level. 35% of all athletes rated funding concerns at the highest possible level (7 out of 7).
 In what range can we be 90% sure that the true percentage for all athletes in this competition is likely to lie. (5)
 What would the size of the interval be if we wanted to be 99% sure instead? (5)

6. Below we have 'livability scores' for 4 countries across two different years. We also have information on the number of survey respondents in each country that ranked the country in the top half vs bottom half of all countries. Use this information to answer the questions below:

Country	livability score		Is my country better or worse than average?	
	1998	2018	better	worse
country 1	72	78	765	142
country 2	41	61	327	345
country 3	33	55	439	230
country 4	57	52	288	689
country 5	66	65	452	321

Which country has had the greatest change in scores between 1998 and 2018? (5)
 What is the ratio of respondents who think their country is better than average to those that think it is worse for each country? Which country has the lowest number of respondents who think it is worse for every respondent who thinks its better? (5)