Attention: This is a Pre-Statistics self-diagnostic test. Its main goal is to help you get acquainted with the typical thought processes used in Statistics. The questions (and required answers) tackled in our course will be more sophisticated.

Show your complete calculations and explain your reasonings. Do all of this clearly and in an organized way.

1. How do you represent mathematically the expression "x is at least 1"?
   (a) $x < 1$      (b) $x \leq 1$      (c) $x > 1$      (d) $x \geq 1$

   Answer: (c). Another way to say the same thing would be 'x can be 1 or more'.

2. How do you represent mathematically the expression "x is at most 1"?
   (a) $x < 1$      (b) $x \leq 1$      (c) $x \geq 1$      (d) $x > 1$

   Answer: (b). Another way to make the same statement would be to say 'x can be 1 or less'.

3. a=1, b=-2, c=2. Add 'a' and '-c', divide the result by 'b' and multiply the result by 2. What’s the result?
   (a) $\frac{1}{2}$      (b) $-1$      (c) $-\frac{1}{4}$      (d) 1

   Answer: (d). Careful with the signs.

4. Which country will be more crowded: one with average population density of 120 persons/mi$^2$ or the one with 200 persons/mi$^2$.

   Answer: The one with 200 persons/mi$^2$. In one mile we have on average 200 persons, against 120 persons in the other country.

5. A retail electronics company promotes the following slogan: "No other digital video recorder retailer can beat our prices". Say whether the following sentences are equivalent to their slogan, justify your answer.
   (a) All competitor’s prices are higher     (b) All competitors have either equal prices or higher
   (c) All competitor’s prices are lower     (d) All competitors have either equal prices or lower

   Answer: (b), notice this statement is basically admitting you could find similar deals with the competition!! Was that your feeling when you first read the slogan? If it wasn’t their marketing strategy worked!

6. An organization plans a campaign to make people aware of tuberculosis and promote a new state-funded project for free immunization. They decide to put up most of their posters in gas stations all over the state. Will they be able to reach pretty much everyone? Will there be anyone left out? Will this matter?

   Answer: No. Even though a sizeable population has cars and therefore will sooner or later stop by a gas station, there is part of the population that doesn’t own a car, in particular there’s a sub-group (homeless people) that doesn’t on a car and is actually particularly vulnerable to the disease. If your first thought was that this setup for the campaign was good, you probably fell prey to thinking mostly about your own socio-economical group (a mistake very easy to make and very pervasive in our society). Thinking carefully before you state your answer is a plus in this class, answering questions hastily will most likely produce low results.
7. What’s the difference between the expression ”Wear a jacket and an umbrella” and ”Wear either a jacket or an umbrella”?

**Answer:** In the first sentence the person wears both items, in the second the person wears one or the other but not both simultaneously. In this question we expect most students to go the right idea. The goal of this question is to see if you can express your ideas clearly. Where you able to convey the correct answer in a single, clear sentence?

8. Spam filters are programs that help keep spam mail away from your email accounts. Sometimes they don’t work well. What kind of mistakes do they do?

(a) Only one: refuse to deliver mail that is genuine
(b) Only one: deliver spam mail to your inbox
(c) Mainly two: refuse to deliver mail that is genuine and deliver spam mail to your inbox
(d) they never make mistakes

**Answer:** (c). We can see this from our own experience, but also we can just think about how spam filters work, they are basically automatic classifiers; this means they can misclassify an email in two ways: either by mistaking a real email for spam or vice-versa.

9. Will an AIDS test suffer from the same error(s)? Say explicitly what would it (they) consist of in this new context?

**Answer:** Yes. AIDS tests are also automatic classifiers ("This subject is either infected or not"), and is also subject to saying someone is infected when in reality it is not, or the other way around, it can classify someone as not being infected when that person is, in fact, infected.

10. (Refers to question about spam filters). Can you say which of the errors mentioned in question 8 will happen the most and which would happen fewer times, or do you think they don’t happen at all, or they happen roughly the same amount of time.

**Answer:** The error that typically occurs the fewer times is classifying genuine email as spam, classifying spam mail as genuine occurs far more times. Again this can be seen from our own experience.

11. (Refers to question about aids tests). Can you say which of the errors mentioned in question 9 will happen the most and which would happen fewer times, or do you think they don’t happen at all, or they happen roughly the same amount of times.

**Answer:** We don’t hear very much about these, but we can safely say that classifying an infected person as non-infected is the most rare of the occurrences.

12. (Refers to question about spam filters). Which of the errors in question 8 is more serious from the point of view of the user?

**Answer:** Classifying a genuine e-mail as spam, since we don’t usually have the patience to check the spam-mail folder and an important email can be lost. Having spam mail being classified as genuine is annoying (granted), but certainly less serious of an event.

13. (Refers to question about aids tests). Which is the error (mentioned in question 9) we would be especially interested in minimizing?

**Answer:** Classifying an infected person as not being infected. Right? (Think about the consequences of that). In fact, tests for contagious diseases are especially designed to have the lowest possible error of this kind.

14. A group of 5 students is asked to independently make one measurement of the height of a door with a measuring tape. Which list of numbers (in inches) is the most likely to be the true one? Say what’s wrong with each of the lists you reject.
(a) (70, 70, 70, 70, 70)  (b) (69, 71, 70, 78, 71)  (c) (69, 71, 70, 69, 71)  (d) (65, 61, 78, 81, 65)

**Answer:** The correct list should be (c). (a) is very unlikely to happen because we expect each person to measure the length slightly different from one another; (b) although we expect some variability in the measurements, the fourth is really unrealistic; (d) the variability on this one is also unrealistic. The goal of this question is to get you to start thinking about variability of measurements.

15. These are the enrollment figures for a school over a period of 10 years: 2200, 2000, 2050, 2100, 2300, 2150, 2300, 2050, 2100, 2000. Do you think that the evidence supports the idea that enrollment is decreasing? Someone says that the school should set up a package costing US$2 million to counter what is perceived as a drop in enrollment. Do you think this person is justified? If you disagree, state your reasons.

**Answer:** Looking at enrollment throughout the years we see some variability. The enrollment in the 10th year is different from the 9th year, yes, but with a magnitude quite typically of the overall variability. We don’t think the evidence is clear to support a real drop in enrollment and maybe we should hold off spending a small fortune to boost the enrollment. This question is also designed to get you to start thinking about variability and it is intrinsically connected with uncertainty, and how to make decisions inspite of that uncertainty.

16. Let’s say we are both looking out the window and see a black bird. Consider the following reasoning: *All crows are black. That bird is black, therefore it is a crow.* Is the reasoning correct? If it isn’t say what’s wrong with it (state your reasons clearly and succinctly).

**Answer:** The first two statements are true, but the conclusion *this bird is a crow* does not follow from them. The conclusion would only follow if all black birds are black (which is not true). This question deals with your capability of reading a question carefully, extracting information from it, and without getting confused, to be able to deliver a clear answer. Where you able to do it? Where you able to answer this question in one or two short sentences?

17. In the last few months I have been measuring the time bus number 20 takes to ride from the Metro Center up to the Science Hill bus stop. These are the measurements (in minutes): 15, 20, 19, 18, 16, 17, 21. Let’s say you’re waiting for that same bus now at the Metro Center,

a) What is the best estimate you can make for the length of the trip?

**Answer:** Around 18 minutes. I exclude the most extreme values as being less usual and see that the trips all go around 18 (17 or 19 would also be good estimates). This question deals again with your ability to make decisions inspite of the uncertainty that arises from variability.

b) Do you expect the actual trip to take the exact amount of time you predict or do you expect a difference?

**Answer:** No, I expect the actual trip to last a different amount of time most of the times.

c) Can you state an approximate magnitude you expect for the difference mentioned in part (b)? (Note: you can think in terms of a full statement like this: ”I expect this trip to take ___ minutes, give or take ___ minutes”, this question deals with the second blank).

**Answer:** Looking a the typical variability I observed I’d say the trip should last 18 minutes give or take 3 minutes. Values of 4 or 5 would also be reasonable estimates.

18. When designing surveys or questionnaires one has to be careful about the words one uses in the questions, since some questions have some emotional charge associated with them that can bias the answers. One example (which will most likely be covered in class) is about the use of the word ‘forbid’. In one study the following question was asked of American adults ”Do you think the US should forbid public speeches against democracy?”. 21.4% of the respondents
answered ‘yes’. Another study (run in pretty much the same conditions as the previous one) asked the question “Do you think the US should allow public speeches against democracy?”. In this study 47% answered ‘no’.

(a) To make the point that the word ‘forbid’ provokes a different reaction that the word ‘allow’, we compare the percentage of people who said ‘yes’ to the first question to the percentage of people who said ‘no’ to the second question. Is this the correct comparison? Should we compare the percentage who say ‘yes’ to both questions? Justify both answers carefully.

**Answer:** The comparison is correct because the two questions ask opposite things. If we are quantifying support for measures to defend democracy then we should be looking for who answers ‘yes’ to the first question and who answers ‘no’ to the second measure. Again, this is a question that is not supposed to measure your level of knowledge or intelligence, we would be much more interested in knowing how you analyze the data given in the question and if you are able to express your reasoning clearly and succinctly.

(b) So, what’s the final conclusion? Does using the word ‘forbid’ increase or decrease the support for measures to defend democracy? By how much?

**Answer:** Using the word ‘forbid’ decreases support for measures to defend democracy by 25%. Use of that word seems to bias the answers against what is being asked.