WORKSHEET: t

Problem 1. In each of the following, suppose T is a random variable with a t distribution with the indicated number of degrees of freedom. Use the R functions `pt` and `qt` to calculate the indicated quantity.

(a) df $= 5$. Calculate $P(T \leq 1)$.

(b) df $= 10$. Calculate $P(T \leq 1)$.

(c) df $= 15$. Calculate $P(T \leq 1)$.

(d) df $= 15$. Calculate $P(T \geq 1)$.

(e) df $= 15$. Calculate $P(0.5 \leq T \leq 1)$.

(f) df $= 15$. Calculate the value of the observation that is larger than exactly 80% of observations.

(g) df $= 15$. Calculate the value of the observation that is less than exactly 80% of observations.

(h) df $= 15$. Calculate the number t^* such that 60% of observations are within t^* standard deviations of the mean.

Problem 2. A 90% confidence interval for a population mean is (65, 77). The population distribution is approximately normal and the population standard deviation is unknown. This confidence interval is based on a simple random sample of 25 observations. Calculate the sample mean, the margin of error, and the sample standard deviation.

Problem 3. The standard deviation of SAT scores for students at a particular Ivy League college is 250 points. Two statistics students, Raina and Luke, want to estimate the average SAT score of students at this college as part of a class project. They want their margin of error to be no more than 25 points.

(a) Raina wants to use a 90% confidence interval. How large a sample should she collect?

(b) Luke wants to use a 99% confidence interval. Without calculating the actual sample size, determine whether his sample should be larger or smaller than Raina's, and explain your reasoning.

(c) Calculate the minimum required sample size for Luke.

Problem 4. Researchers interested in lead exposure due to car exhaust sampled the blood of 52 police officers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these officers had an average lead concentration of 124.32 g/l and a SD of 37.74 g/l. A previous study of individuals from a nearby suburb, with no history of exposure, found an average blood level concentration of 35 g/l. Does this data provide convincing evidence that the blood samples from these police officers have a higher lead concentration than the group with no history of exposure? (Be sure to state your hypotheses clearly, explicitly state and check all conditions necessary for inference, and state your conclusion in the context of the data.)

Problem 5. An investigator collects test data about reading scores and writing scores for a simple random sample of 200 students. The observed average value of reading test score minus writing test score is -0.545, and the standard deviation is 8.887. Does this data provide convincing evidence of a difference between the average scores on the reading and writing tests? (Be sure to state your hypotheses clearly, explicitly state and check all conditions necessary for inference, and state your conclusion in the context of the data.)

Problem 6. An investigator is interested in understanding the relationship, if any, between the analytical skills of young gifted children (`score`) and the following variables:

- `fatheriq`: father's IQ
- `motheriq`: mother's IQ
- `speak`: age in months when the child first said 'mummy' or 'daddy'
- `count`: age in months when the child first counted to 10 successfully
- `read`: average number of hours per week the child's mother or father reads to the child
- `edutv`: average number of hours per week the child watched an educational program on TV during the past three months
- `cartoons`: average number of hours per week the child watched cartoons on TV during the past three months

The analytical skills are evaluated using a standard testing procedure, and the score on this test is used as the response variable. This data was collected as a simple random sample from schools in a large city on a set of 36 children who were identified as gifted children soon after they reached the age of four.

Load this data set into R using the following command:

```
gifted <- read.csv("https://sagrawalx.github.io/teaching/sp20-b6_ma117/gifted.csv")
```

Then conduct hypothesis tests to answer the following questions. Make sure to clearly state your hypotheses, check the relevant conditions, and state your conclusion in the context of the data.

(a) Are IQ scores of mothers and fathers of young gifted children equal?

(b) Do gifted children spend as much time watching educational programs on TV as they do cartoons?