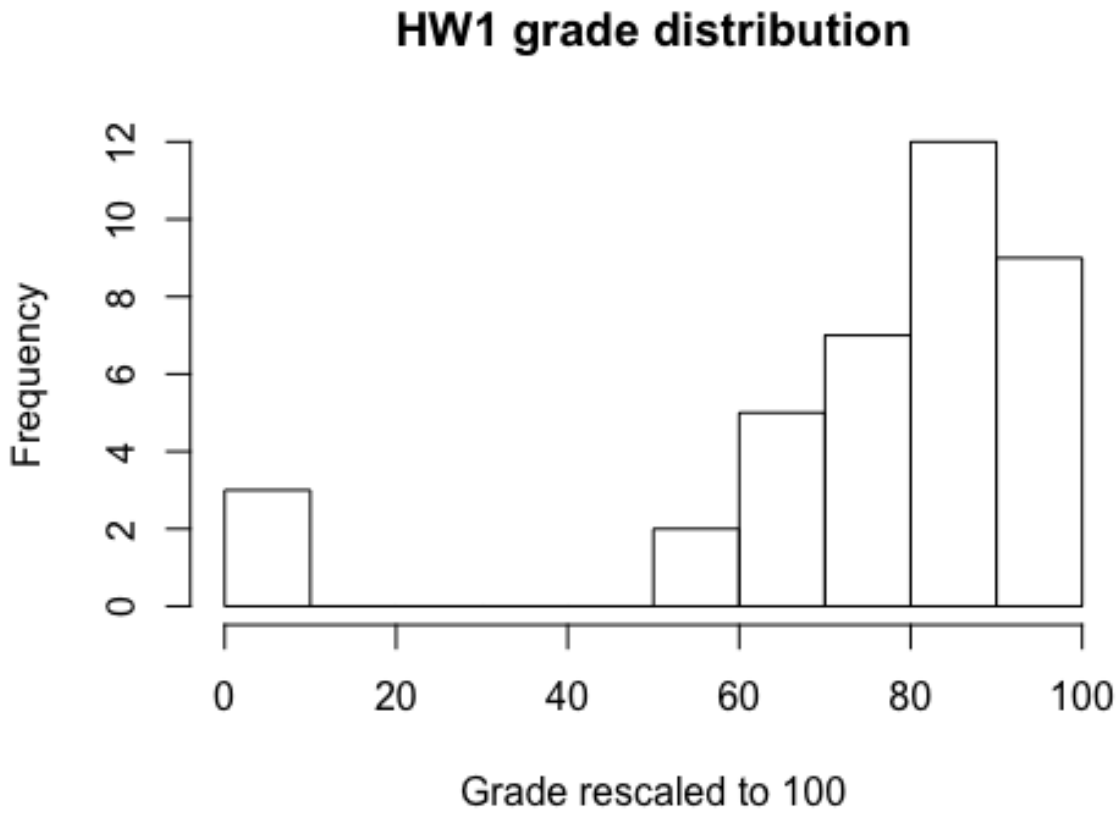
Statistics

Week 3 Recitation

ESD, SUTD

Term 5, 2017

Homework 1 Revisit



Note: Points deducted from your submissions are based on a 40-point marking scheme.
The final scores on eDimension are rescaled to 100 points.

Homework 1 Revisit

Question 7. We wish to conduct a survey on a sensitive issue, namely to find the proportion x of people who regularly chew gum. n people are randomly chosen. Each is asked to toss a biased coin. If the coins shows H(which occurs with probability q), the person is asked to answer QA; otherwise s/he is to answer QB.

Question A: Do you regularly chew gum?

Question B: Is your birthday within 2 weeks of Christmas?

Suppose proportion p of the responses are Yes.

- $P(answer\ Yes) =$
  $P(answer\ Yes|QA) \times P(QA) + P(answer\ Yes|QB) \times P(QB)$
- The calculated x is an *Estimation* of a probability, thus CAN be out of the range of [0,1]

Homework 1 Revisit

Question 8. A population consists of N = 4 numbers: 1, 2, 3, 4. Each number has equal chance of being selected.

(b) A random sample of size n of 2 is selected with replacement from the population. Find the sampling distribution for the sample variance.

Homework 1 Revisit

Question 9. Suppose that the enemy has tanks numbered 0, 1, 2, . . . ,N. You observe n of the tank with replacement at random and note down their numbers. Using the sample mean of these numbers, find an unbiased estimator for the total number of tanks.

- By symmetry,
  $E(\bar{X}) = E(X_i) = \frac{1}{N+1}(0 + 1 + 2 + ... + N) = \frac{N}{2}$
- Total number of tanks is $N + 1$
- If prove your answer with simulation, make sure to have large enough sample size.

Homework 1 Revisit

Talk to me after this recitation / email me if you have any questions about the homework grading.

For future homework submission:

- Put all your write-up in a single document (pdf or word document)
- Copies of handwritten solutions are OK, but make sure to scan or take photos under good lighting condition
- DO NOT submit separated image files of copied handwritten solutions
- Document your working process; explain your answers
- Attatch separate documents (such as excel spread sheet) to show your working process if necessary

Simulation in R - Distribution Functions

Standard distributions are built in with R: Normal, Uniform, Binomial, Exponential, Gamma, Poisson, Student t, Chi-squared, etc.

Probability distribution functions usually have four functions associate with them. The functions are prefixed with a

- `d` for density
- `p` for cumulative distribution
- `q` for quantile function
- `r` for random number generation

Examples:

- Normal distribution: `dnorm()`, `pnorm()`, `qnorm()`, `rnorm()`
- Uniform distribution: `dunif()`, `punif()`, `qunif()`, `runif()`
- Exponential distribution: `dexp()`, `pexp()`, `qexp()`, `rexp()`

Simulation in R - Distribution Functions

Normal distribution:

- `dnorm(x,mean=0,sd=1,log=FALSE)`
- `pnorm(q,mean=0,sd=1,lower.tail=TRUE,log.p=FALSE)`
- `qnorm(p,mean=0,sd=1,lower.tail=TRUE,log.p=FALSE)`

`pnorm(q)` $= \Phi($`q`$)$, `qnorm(p)` $= \Phi^{-1}($`p`$)$,
where Φ is the CDF of standard normal distribution.

If `lower.tail` is `TRUE` (default), probabilities are $P[X \leq x]$
otherwise, $P[X > x]$.

If `log, log.p` is `TRUE`, probabilities p are given as $log(p)$.

Simulation in R (Exercise)- Distribution Functions

Exercise:

- Plot the PDF and CDF of a normal distribution with mean = 1 and sd = 2

Functions to use

`seq(from, to, by)` # to generate a sequence of numbers
`dnorm()` and `pnorm()`
`plot()`

Simulation in R - Generating Random Numbers

Generating random numbers according to Normal distribution, given mean and standard deviation:

- `rnorm(n,mean=0,sd=1)`

Pseudo-random:
Random numbers generated in R are actually deterministic. They are generated by using mathematical formulas or precalculated lists, and are periodic (with very long periods).

To control the randomness:

- `set.seed()`

Randomly drawing samples from a given population, allowing you to sample from an arbitrary distribution:

- `sample(x, size, replace = FALSE, prob = NULL)`

Simulation in R (Exercise) - HW1 Q9 German Tank

Question 9. Suppose that the enemy has tanks numbered 0, 1, 2, . . . ,N. You observe n of the tank with replacement at random and note down their numbers. You want to test if $2\bar{X}$ is an unbiased estimator for the total number of tanks by simulation.

Assume $N = 250$, and you observed $n = 5$ tanks. Use R to:

- Draw 5 random samples with replacement from 0, 1, 2, ... 250
- Calculate $2\bar{X}$ of the samples
- Record the value
- Repeat above steps 500 times

Questions:

- What is the mean of your estimator, and the mean's 95% CI?
- Can you reject the hypothesis that $2\bar{X}$ is an unbiased estimator of the total number of tanks?