FISH 102

Statistical distributions in R – Gaussian (Normal) Distribution

`rnorm` generates a random sample from a Normal distribution. eg `rnorm(10, 0, 1)` samples 10 points from a Normal distribution with mean ($\mu = 0$) and standard deviation $\sigma = 1$.

- Randomly sample data from a Normal distribution and plot it using a histogram. eg. `hist(rnorm(100, 0, 1))` samples 100 points with $\mu = 0$ and $\sigma = 1$ and plots a histogram.
- Using `help(rnorm)` change the sample size. Then change μ and σ. What happens to the distribution as these parameters change?
- Calculate the mean of the random sample for a range of sample sizes. What happens to the mean as the sample size increases?
- An alternative plot is a density plot. eg. `plot(density(rnorm(100,0,1)))`. Sample data using `rnorm` and store as an R object, then plot the same data using both methods. You can see more than one plot at once using `par(mfrow=c(m,n))`, where m is the number of rows of plots and n the number of columns of plots.
- Calculate $z_{1-\alpha/2}$ for $\alpha = 0.01, 0.05$. (eg. `qnorm(1-0.05/2)` is 2–sided with $\alpha = 0.05$).
- Given a dataset $x = (2, 6, 8, 3, 5, 6, 9, 3, 2, 5, 7, 8)$, calculate the mean, standard deviation and confidence interval with $\alpha = 0.05$, then with $\alpha = 0.01$.
- If $\mu \approx 50$ and $\sigma = 10$ what sample size is required to ensure (with $\alpha = 0.05$) a confidence interval width ≤ 0.5?

STATS 150

R — loops and functions

`source`

Remember: scripts can be run in R from a text file using `source`.

- Write some commands in a text editor eg

```
x <- 1:10
y <- x*x
plot(x,y)
print(c(mean(x), mean(y)))
```

- and save the file eg test.r
- In R, `source("test.r")`. If R is not running in the same directory as the text file you will also need to provide the path.

Loops – `for`

It is often useful to be able to repeat the same operation and one way to do this is using a for–loop.

- In a text file save this code:

```
for(i in 1:5){
  print(i*i)
}
```

- Run the script in R.
- A loop can be used to compare samples from a Normal distribution. In the example below σ varies.

```
x <- seq(0.5,2.5, 0.5)
t <- 0
plot(density(rnorm(100, 0, 1)), ylim=c(0, 0.1), type="n")
for(i in x) {
  t <- t+1
  lines(density(rnorm(100, 0, i)), col=t)
}
```

- What does `type="n"` do?
- Compare samples from Normal distributions where you repeat the same sampling procedure. What happens as you decrease/increase n?
- Try the examples in your lecture notes.

Loops – `while`

It is also possible to write loops with the `while` command.

- Try this:

```
sumi <- 0
i <- 0
while(sumi < 10){
  i <- i+1
  sumi <- sumi+i
  print(c(sumi, i))
}
```

- a `while` loop repeats the commands between the {}'s until the while statement is met.

The `if` statement

eg

- eg

```
x <- 1:10
if(length(x) > 5){
  print(mean(x))
}
```

- can also to `if` *cond* `else` eg

```
x <- 1:3
if(length(x) >= 5){
  print(mean(x))
 } else {
  print("Error: x < 5")
}
```

Writing a function in R

- Create the following function:

```
mymean <- function(data){
 m <- sum(data)/length(data)
 return(m)
}
```

- Create a small dataset `x`.
- Compare `mymean(x)` with `mean(x)`