ST 314 R Functions

CDFs, PDFs, and Percentiles of Named Distributions

Distribution	Function	Function Values	What does it do?
Binomial	`dbinom(x,n,p)`	x = value of interest n = number of trials p = probability of success	This is the probability mass function for a binomial distribution. $P(X = x)$
Binomial	`pbinom(x,n,p)`	x = value of interest n = number of trials p = probability of success	This is the cumulative distribution function for a binomial distribution. $P(X \leq x)$
Poisson	`dpois(x,lambda)`	x = value of interest lambda = rate parameter	This is the probability mass function for a Poisson distribution. $P(X = x)$
Poisson	`ppois(x,lambda)`	x = value of interest lambda = rate parameter	This is the cumulative distribution function for a Poisson distribution. $P(X \leq x)$

Uniform	`punif(x,a,b)`	x = value of interest a = lower bound b = upper bound	This is the cumulative distribution function for a Uniform distribution. $P(X \leq x)$
Uniform	`qunif(p,a,b)`	p = proportion that falls below x a = lower bound b = upper bound	This is the inverse cumulative distribution function. Finds percentiles for the Uniform distribution. That is, finds x_p for the expression $P(X \leq x_p) = p$
Gamma	`pgamma(x,shape,rate)`	x = value of interest shape = α rate = $1/\beta$	This is the cumulative distribution function for a Gamma distribution. $P(X \leq x)$
Gamma	`qgamma(p,shape,rate)`	p = proportion that falls below x shape = α rate = $1/\beta$	This is the inverse cumulative distribution function. Finds percentiles for the Gamma distribution. That is, finds x_p for the expression $P(X \leq x_p) = p$
Exponential	`pexp(x,lambda)`	x = value of interest lambda = λ	This is the cumulative distribution function for an Exponential distribution. $P(X \leq x)$
Exponential	`qexp(p,lambda)`	p = proportion that falls below x lambda = λ	This is the inverse cumulative distribution function. Finds percentiles for the

			Exponential distribution. That is, finds x_p for the expression $P(X \leq x_p) = p$
Normal	`pnorm(x,mu,sigma)`	x = value of interest mu = mean sigma = standard deviation	This is the cumulative distribution function for a Normal distribution. $P(X \leq x)$
Normal	`qnorm(p,mu,sigma)`	p = proportion that falls below x mu = mean sigma = standard deviation	This is the inverse cumulative distribution function. Finds percentiles for the Normal distribution. That is, finds x_p for the expression $P(X \leq x_p) = p$
t	`pt(x,df)`	x = value of interest df = degrees of freedom	This is the cumulative distribution function for a *t* distribution. $P(X \leq x)$
t	`qt(p,df)`	p = proportion that falls below x df = degrees of freedom	This is the inverse cumulative distribution function. Finds percentiles for the *t* distribution. That is, finds x_p for the expression $P(X \leq x_p) = p$
F	`pf(x,v1,v2)`	x = value of interest v1 = numerator degrees of freedom	This is the cumulative distribution function for an *F* distribution.

		v2 = denominator degrees of freedom	$P(X \leq x)$
F	`qf(p,v1,v2)`	p = proportion that falls below x v1 = numerator degrees of freedom v2 = denominator degrees of freedom	This is the inverse cumulative distribution function. Finds percentiles for the *F* distribution. That is, finds x_p for the expression $P(X \leq x_p) = p$

Organizing Data

- Create a vector of values for a single variable:
  - Use the `c()` function to combine values into a single vector.
  - Example: `variable <- c(12, 15, 13, 11, 10, 14)`
- Create a data frame from a vector(s):
  - Use the `data.frame()` function.
  - Within the function, define the name you want a column to have in quotes and then set equal to the vector you want to use for that variable.
  - Example: `dataframe <- data.frame("values" = variable)`.
  - If you want your data frame to include more than on column, include new variables in the same way as above, separated by a comma.
    - Example: `dataframe <- data.frame("values" = variable, "var2" = variable2)`
- Referencing a specific column within a data frame:
  - To reference a column, first define the data frame you want to use, then specify the column name. The data frame name and column should be separated by `$`
  - Example: `dataframe$values`
- Check the type of an object stored in R:
  - Use the `str()` function. str is short for structure.

- Possible outputs may include: `num` - a vector of numeric values, `chr` - a vector where store values are characters, `Factor` - a vector where store values are defined categories, `data.frame` - object is defined as a data frame
- Examples: `str(variable)` or `str(dataframe)`

Summarizing Data

Function	What does it do?	Example (using `variable` and `dataframe` from above)
`mean()`	Calculates the mean from a vector of values.	`mean(variable)`
`median()`	Calculates the median from a vector of values.	`median(variable)`
`sd()`	Calculates the standard deviation from a vector of values.	`sd(variable)`
`summary()`	Calculates the five-number summary (min, Q1, median, Q3, max) and the mean from a vector of values.	`summary(variable)`
`length()`	Calculates the number of values stored in a vector.	`length(variable)`
`nrow()`	Calculates the number of rows in a data frame.	`nrow(dataframe)`