

# Introduction to R – Wednesday AM

1) Getting help / external resources

2) Basic statistics

- Probability distributions
- Descriptive stats
- Inference from single populations
- Inference from two populations
- Simple linear regression
- Exporting output
- Examples and exercises

# 5.1 Probability distributions

prefix	function
d	density
p	cumulative probabilities
q	quantiles
r	simulate

stem	distribution
binom	Binomial
pois	Poisson
norm	Normal
t	Student's t
chisq	$\chi^2$
f	F

## Normal distribution

dnorm(quantiles, mean, sd)  
pnorm(quantiles, mean, sd)  
qnorm(probabilities, mean, sd)  
rnorm(length, mean, sd)

## Binomial distribution

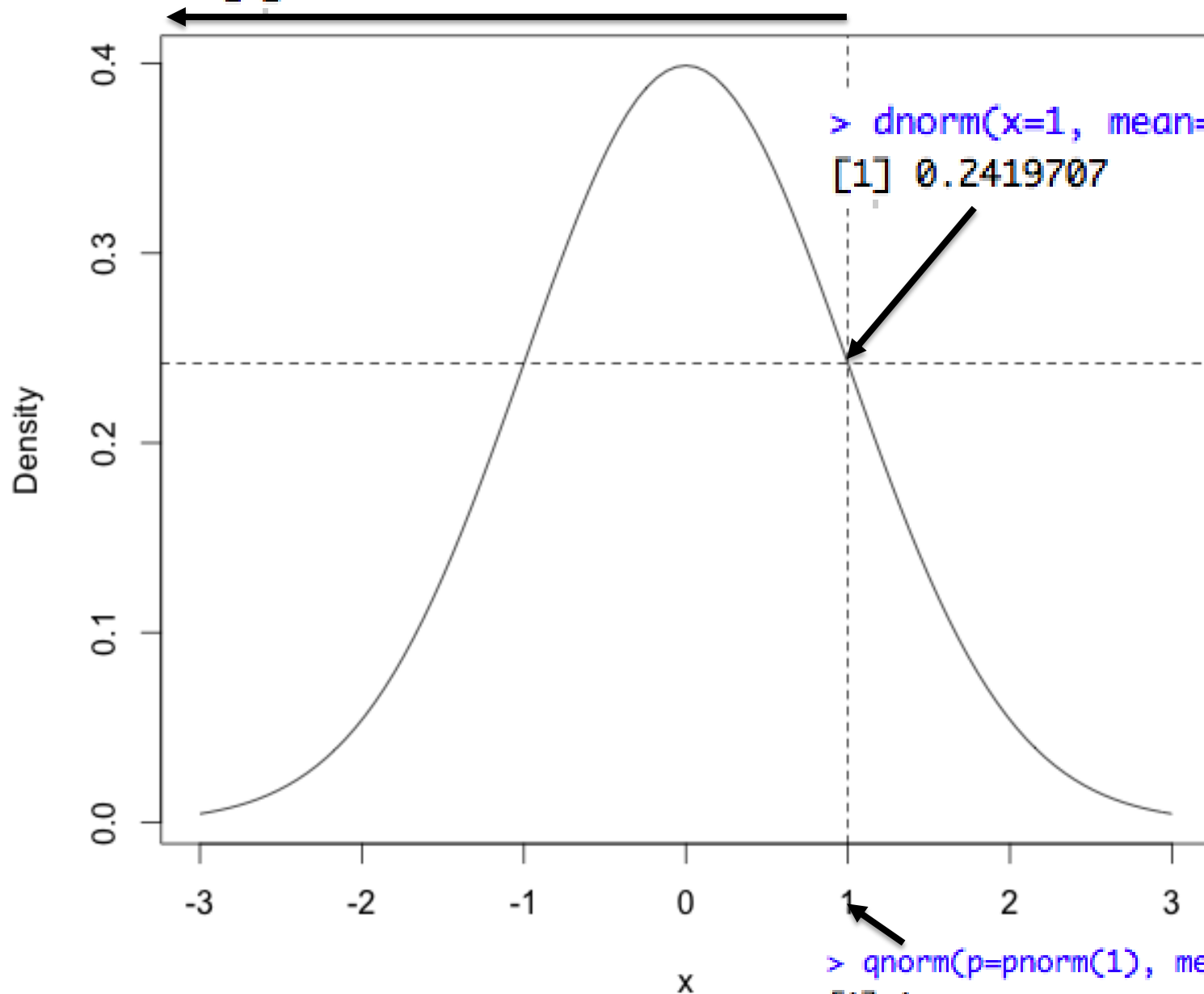
dbinom(quantiles, # trials, probability)  
pbinom(quantiles, # trials, probability)  
qbinom(probabilities, # trials, probability)  
rbinom(length, # trials, probability)

## F distribution

df(quantiles, num DF, denom DF)  
pf(quantiles, num DF, denom DF)  
qf(probabilities, num DF, denom DF)  
rf(length, num DF, denom DF)

```
> pnorm(q=1, mean=0, sd=1)
```

```
[1] 0.8413447
```



```
> dnorm(x=1, mean=0, sd=1)
```

```
[1] 0.2419707
```

```
> qnorm(p=pnorm(1), mean=0, sd=1)
```

```
[1] 1
```

## 5.2 Descriptive Statistics

- Several functions that work with single numeric vectors: `mean()`, `median()`, `var()`, `sd()`, `length()`, `IQR()`, `quantile()`
- `summary()` on a numeric vector returns mean, median, and quartiles
- `skewness()` and `kurtosis()` in the 'moments' library -> data not normally distributed, overdispersed, underdispersed
- NB! – NAs cause functions to return NA, need to use 'na.rm=T' as argument

## 5.4 Two Populations

- Notice three points with `t.test()` and `wilcox.test()`:
  1. Use the same function for both the one-sample and the two-sample cases:

```
## Default S3 method:  
t.test(x, y = NULL,  
       alternative = c("two.sided", "less", "greater"),  
       mu = 0, paired = FALSE, var.equal = FALSE,  
       conf.level = 0.95, ...)
```

1. Can specify 'X, Y' or use the formula interface ('Y ~ X') for two-sample tests
2. Provide the functions with one or two numeric vectors of equal or unequal length

# 5.5 Simple Linear Regression

- Use to estimate relationships between two variables and to test hypotheses regarding these relationships
- Two-step process (see example in manual):
  1. Fit the model:  
`model <- lm(Y ~ X, data=data)`
  2. Obtain information on the model fit:  
`summary(model)` # coefficients, R<sup>2</sup>, p-values  
`abline(model)` # adds prediction following `plot(X, Y)`  
`resid(model)` # returns model residuals  
`plot(model)` # returns diagnostic plots (normality, heteroscedasticity, leverage)

## 5.6 Exporting output

- Results printed to the console are lost once R is shut down, can use `sink()` to send these results to a file as they are produced

```
# write output to file, no more on screen  
sink('output.txt')
```

```
# stop writing to file  
sink()
```

```
# append to end of existing output file  
sink('output.txt', append=T)
```

# Introduction to R – Wednesday AM

1) Getting help / external resources

2) Basic statistics

- Probability distributions
- Descriptive stats
- Inference from single populations
- Inference from two populations
- Simple linear regression
- Exporting output
- **Examples and exercises**



