

Homework 1 solution

Joshua Loftus

Random variables

R's four functions for each distributions

Question 1: Normal random variables

```
mu <- sqrt(2)
sigma2 <- 9
# Expected value  $E[Z] = \mu$ 
mu

## [1] 1.414
# Variance  $\text{Var}(Z) = \sigma^2$ 
sigma2

## [1] 9
# Distribution function:  $\text{Prob}(Z = 1)$ 
# [note: for continuous random variables this is only approximate]
dnorm(1, mean = mu, sd = sqrt(sigma2))

## [1] 0.1317
# Cumulative distribution function:  $\text{Prob}(Z \leq 1.5)$ 
pnorm(1.5, mean = mu, sd = sqrt(sigma2))

## [1] 0.5114
# Upper tail instead:  $\text{Prob}(Z > 1.5)$ 
1- pnorm(1.5, mean = mu, sd = sqrt(sigma2))

## [1] 0.4886
# Quantile: what is the specific value  $z$ 
# such that  $P(Z \leq z) = .95$ ?
qnorm(.95, mean = mu, sd = sqrt(sigma2))

## [1] 6.349
# Random sampling: generate a sample of 8 observations of  $Z$ 
rnorm(8, mean = mu, sd = sqrt(sigma2))

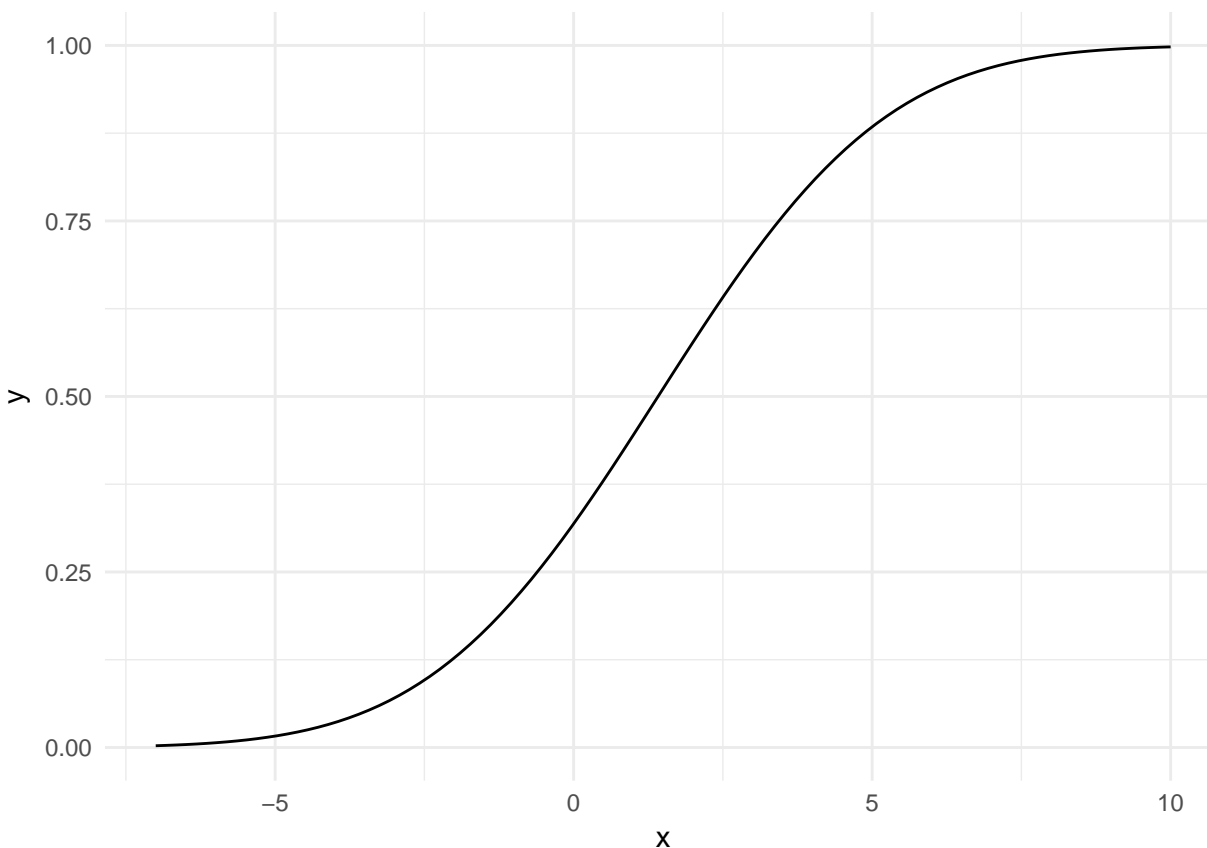
## [1] 5.1924 2.5742 0.2388 0.2200 3.9790 0.2039 0.2968 -0.1689
```

Plotting distribution functions

Normal cumulative distribution function

Plotting the normal cdf:

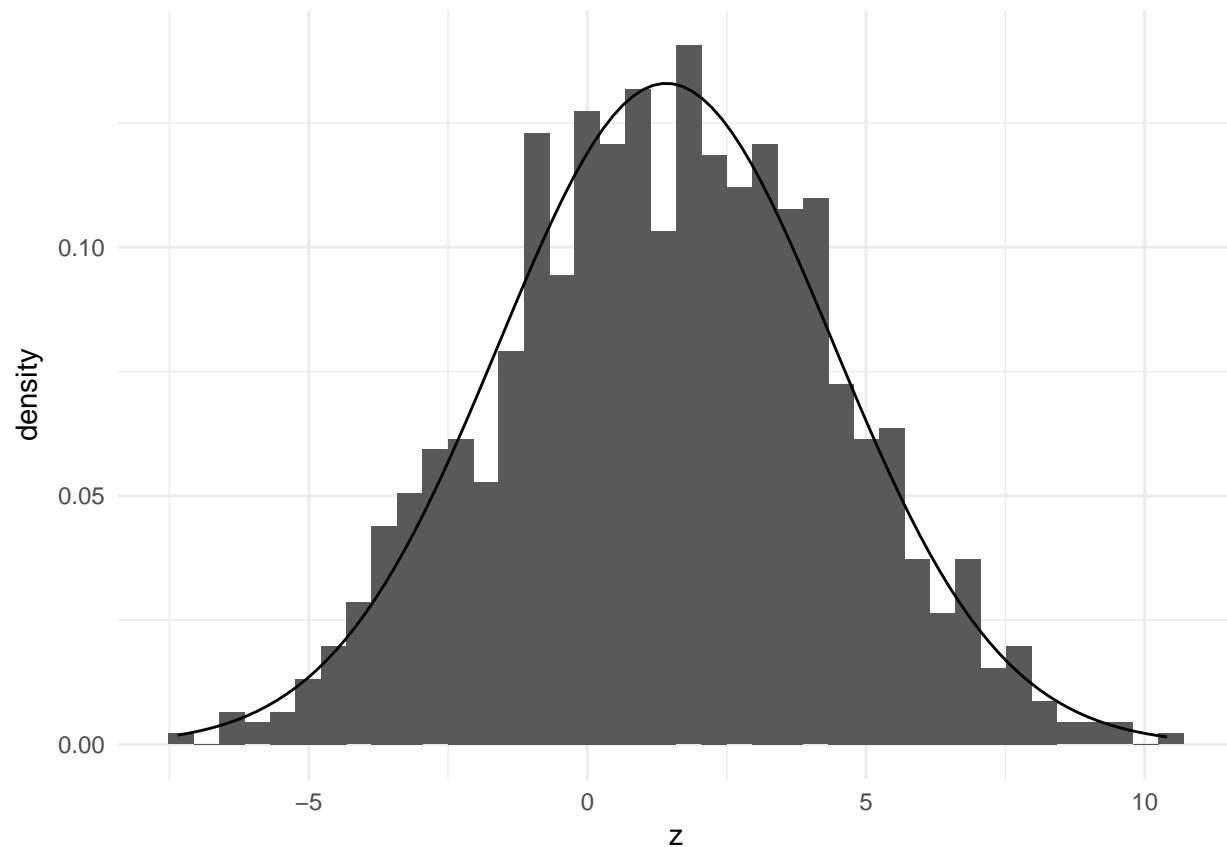
```
# copy and paste the code above, change dnorm to pnorm -- that's it!
range <- data.frame(x = c(-7, 10))
ggplot(range, aes(x)) +
  stat_function(fun = pnorm, args = list(mean = mu, sd = sqrt(sigma2))) +
  theme_minimal()
```



Question 2

Create a histogram plot like the one above, but with a larger sample size.

```
# Copy and paste the code above
# Change sample_size to 1000 and nbins to 40
sample_size <- 1000
nbins <- 40
Z_sample <- data.frame(z = rnorm(sample_size, mean = mu, sd = sqrt(sigma2)))
ggplot(Z_sample, aes(z)) + geom_histogram(aes(y = ..density..), bins = nbins) +
  stat_function(fun = dnorm, args = list(mean = mu, sd = sqrt(sigma2))) +
  theme_minimal()
```



Real data

GDP and life expectancy in gapminder

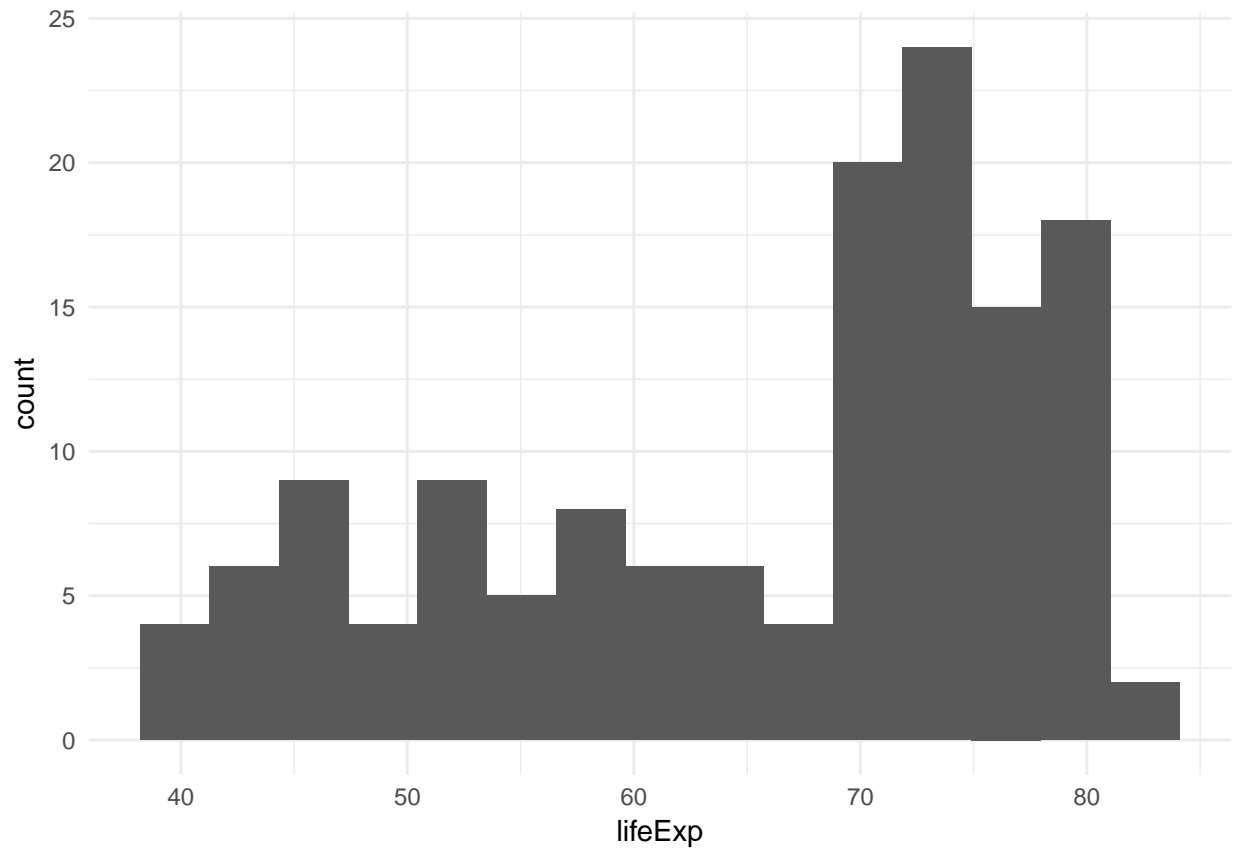
Let's use data from the year 2002

```
world <- filter(gapminder, year == 2002,
  continent != "Oceania")
```

Problem 3

Histogram of lifeExp for the world.

```
# paste here, change gdpPercap to lifeExp
ggplot(world, aes(lifeExp)) +
  geom_histogram(bins = 15) +
  theme_minimal()
```



Density plots of lifeExp for each continent.

```
# paste here, same as above  
ggplot(world, aes(lifeExp)) +  
  geom_density() +  
  facet_wrap(~continent, scales = "free") +  
  theme_minimal()
```

