Gaussian Normal Distribution

Dustin Ma

1/21/2020

Load packages

```
library(dplyr)
library(ggplot2)
```

Gaussian Normal Distribution

```
dnorm(x = 1.15)
```

```
## [1] 0.2059363
```

```
dnorm(x = 1.15, mean = 0, sd = 1) # of mean and std
```

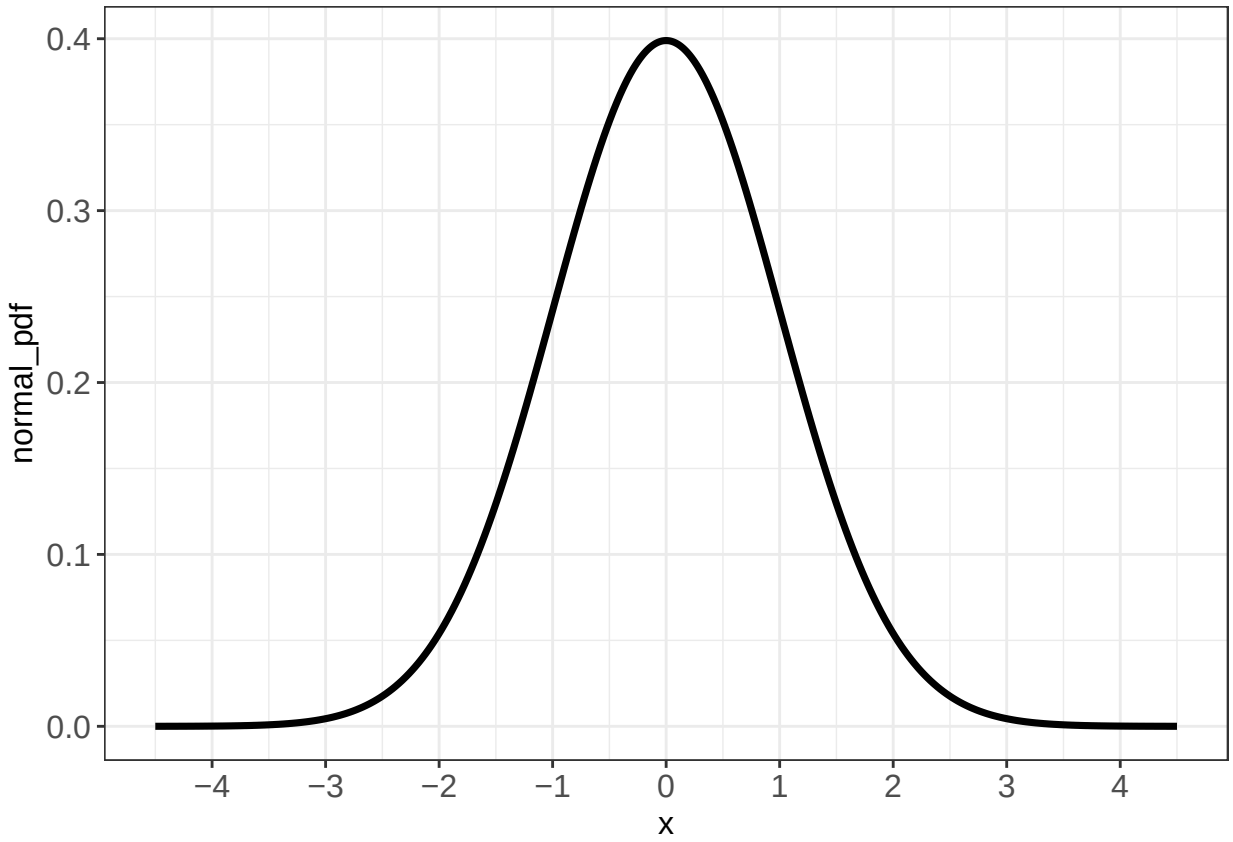
```
## [1] 0.2059363
```

```
dnorm(x = -2:2, mean = 0, sd = 1)
```

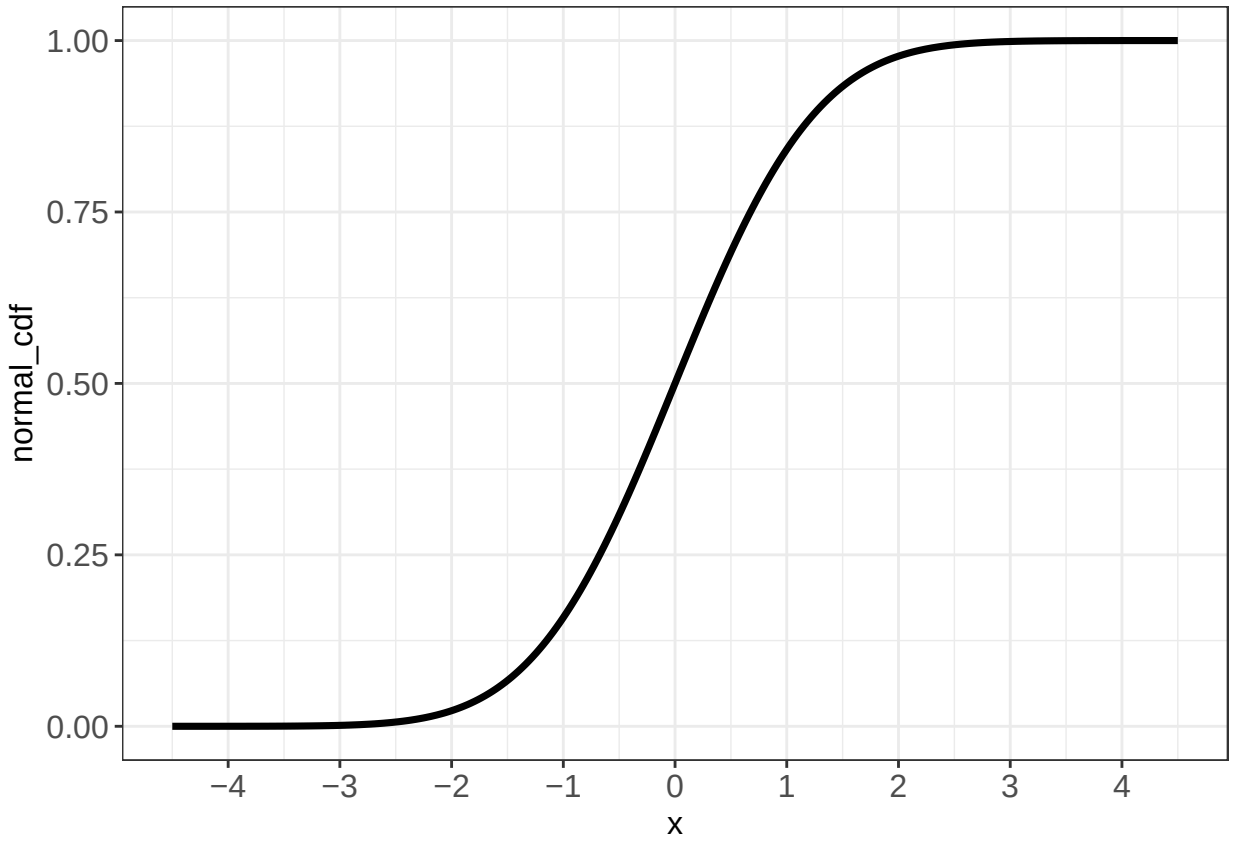
```
## [1] 0.05399097 0.24197072 0.39894228 0.24197072 0.05399097
```

```
df_gauss <- tibble::tibble(
  x = seq(-4.5, 4.5, length.out = 1001)
) %>%
  mutate(normal_pdf = dnorm(x))
```

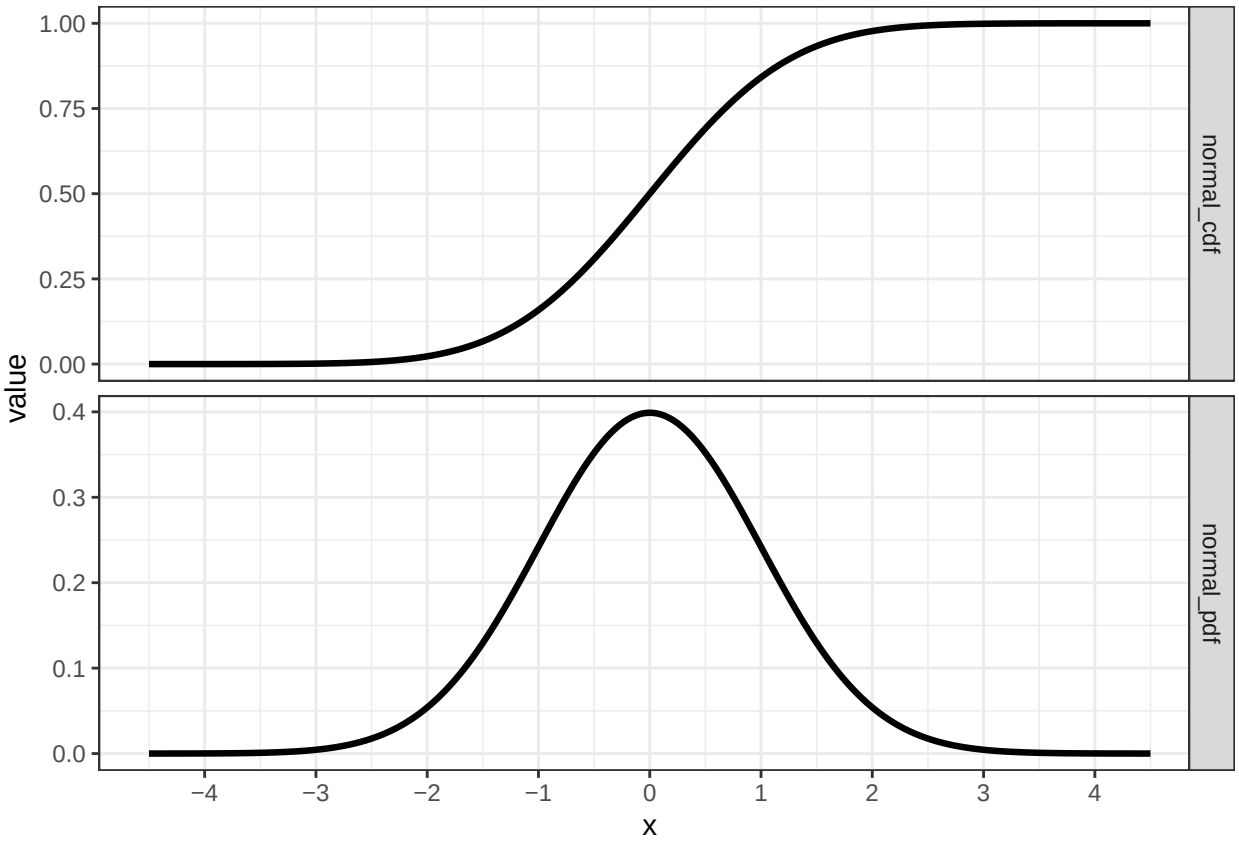
```
df_gauss %>%
  ggplot(mapping = aes(x = x, y = normal_pdf)) +
  geom_line(mapping = aes(group = 1), size = 1.25) +
  scale_x_continuous(breaks = -4:4) +
  theme_bw() +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 12))
```




```
df_gauss %>%
  mutate(normal_cdf = pnorm(q = x)) %>%
  ggplot(mapping = aes(x = x, y = normal_cdf)) +
  geom_line(mapping = aes(group = 1), size = 1.25) +
  scale_x_continuous(breaks = -4:4) +
  theme_bw() +
  theme(axis.text = element_text(size = 12),
        axis.title = element_text(size = 12))
```




```
df_gauss %>%
  mutate(normal_cdf = pnorm(q = x)) %>%
  tibble::rowid_to_column() %>%
  tidyr::gather(key = "key", value = "value", -rowid, -x) %>%
  ggplot(mapping = aes(x = x, y = value)) +
  geom_line(mapping = aes(group = key), size = 1.15) +
  facet_grid(key ~ ., scales = "free_y") +
  scale_x_continuous(breaks = -4:4) +
  theme_bw()
```

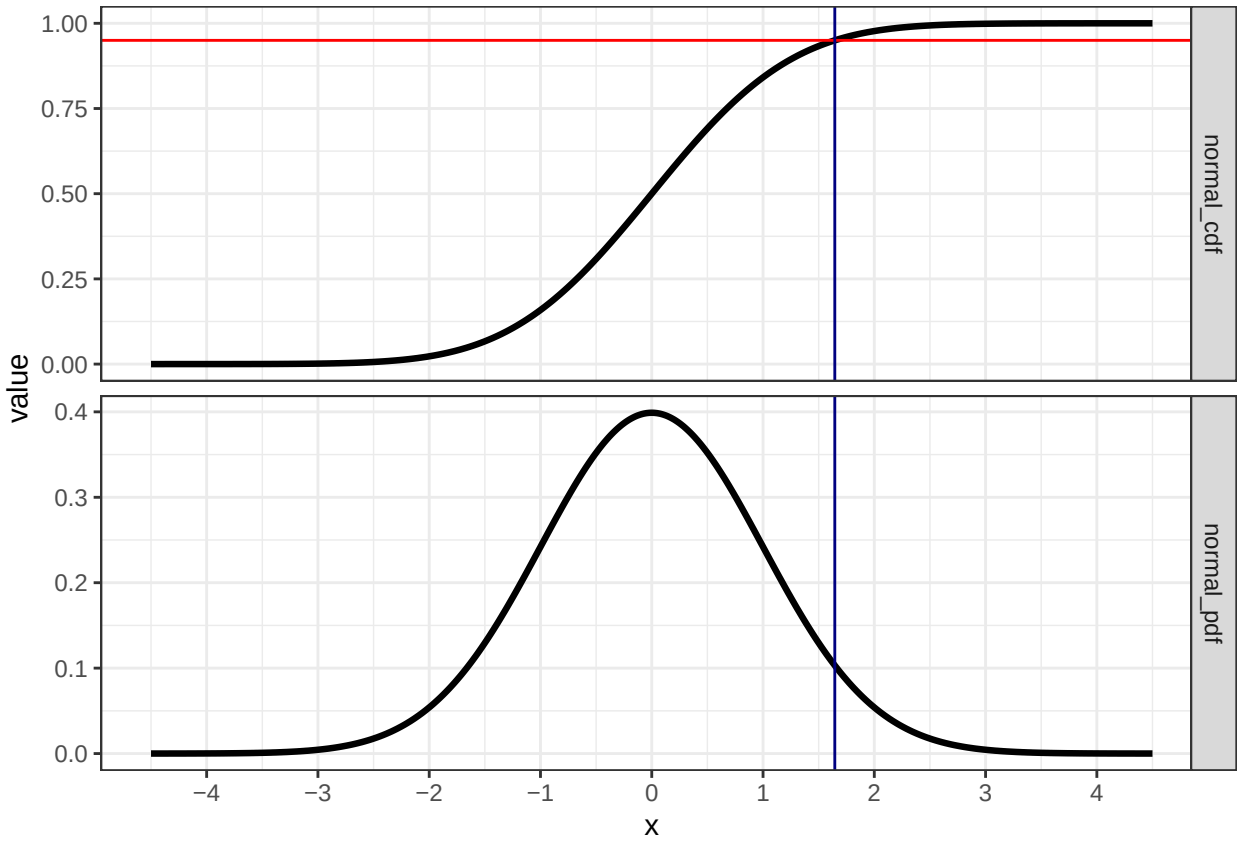


Quantiles

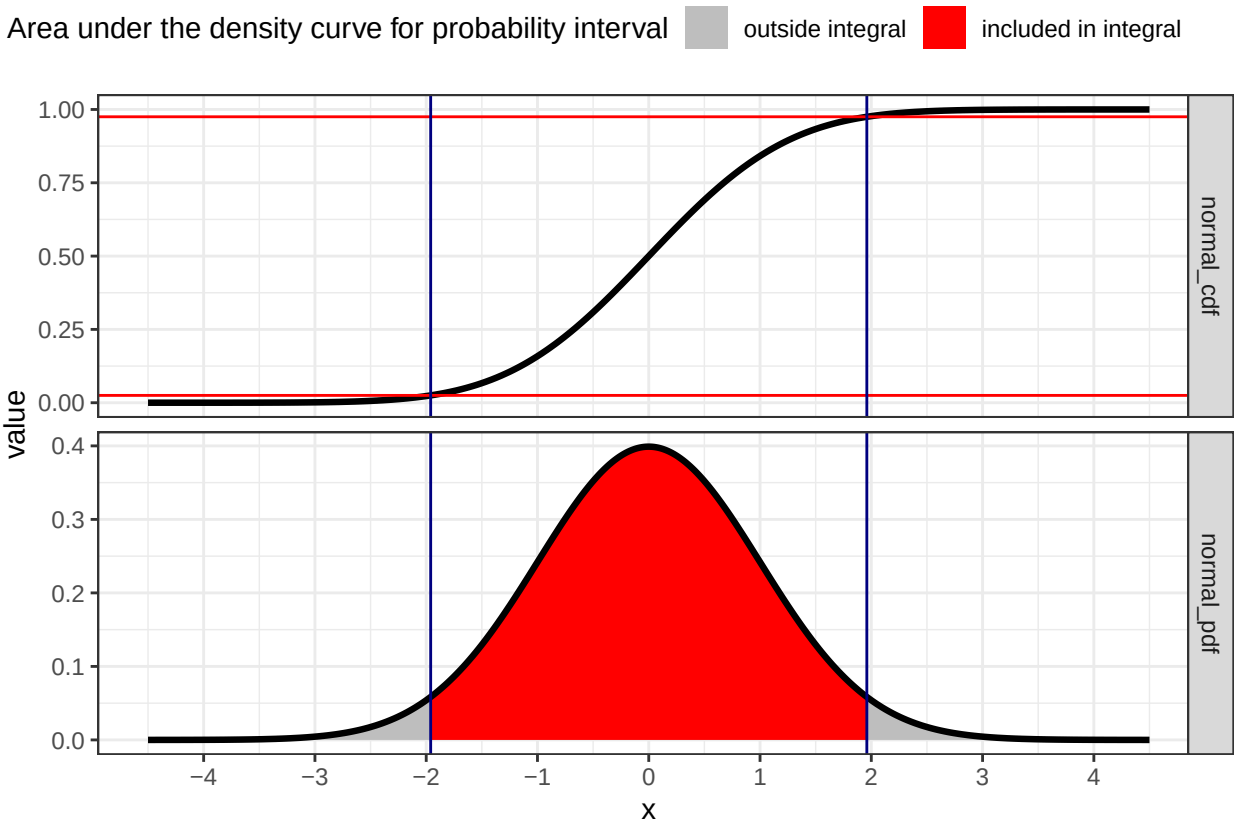
```
qnorm(p = 0.5) # value of the 50th quantile
```

```
## [1] 0
```

```
quant = 0.95
df_gauss %>%
  mutate(normal_cdf = pnorm(q = x)) %>%
  tibble::rowid_to_column() %>%
  tidyr::gather(key = "key", value = "value", -rowid, -x) %>%
  ggplot(mapping = aes(x = x, y = value)) +
  geom_line(mapping = aes(group = key), size = 1.15) +
  geom_hline(data = data.frame(prob_val = quant, key = c("normal_cdf")),
             mapping = aes(yintercept = prob_val),
             color = "red") +
  geom_vline(xintercept = qnorm(quant), color = "navyblue") +
  facet_grid(key ~ ., scales = "free_y") +
  scale_x_continuous(breaks = -4:4) +
  theme_bw()
```




```
df_gauss %>%
  mutate(normal_cdf = pnorm(q = x)) %>%
  tibble::rowid_to_column() %>%
  tidyr::gather(key = "key", value = "value", -rowid, -x) %>%
  mutate(show_area = ifelse(key == "normal_cdf",
                            NA,
                            between(x, qnorm(0.025), qnorm(0.975)))) %>%
  ggplot(mapping = aes(x = x, y = value)) +
  geom_area(mapping = aes(fill = show_area)) +
  geom_line(mapping = aes(group = key), size = 1.15) +
  geom_hline(data = data.frame(prob_val = c(0.025, 0.975), key = c("normal_cdf")),
             mapping = aes(yintercept = prob_val),
             color = "red") +
  geom_vline(xintercept = qnorm(c(0.025, 0.975)), color = "navyblue") +
  facet_grid(key ~ ., scales = "free_y") +
  scale_fill_manual("Area under the density curve for probability interval",
                    values = c("TRUE" = "red",
                               "FALSE" = "grey",
                               "NA" = NA),
                    labels = c("TRUE" = "included in integral",
                               "FALSE" = "outside integral",
                               "NA" = NA)) +
  scale_x_continuous(breaks = -4:4) +
  theme_bw() +
  theme(legend.position = "top")
```


Area under the density curve for probability interval
outside integral
included in integral
1.00
0.75
0.50
0.25
0.00
normal_cdf
value
0.4
0.3
0.2
0.1
0.0
normal_pdf
−4
−3
−2
−1
0
1
2
3
4
x