Jessica Bonin

Analysis of Environmental Data

Lab 4 Report

October 10, 2021

Worked with Juliana Berube, Julia Vineyard, Andrew Gordon

1. Show the code you used to create your vectors

lab.mean = 10.4

lab.sd = 2.4

norm\_17 = rnorm(n = 17, mean = lab.mean, sd = lab.sd)

norm\_30 = rnorm(n = 30, mean = lab.mean, sd = lab.sd)

norm\_300 = rnorm(n = 300, mean = lab.mean, sd = lab.sd)

norm\_3000 = rnorm(n = 3000, mean = lab.mean, sd = lab.sd)

1. Include the R code you used to create your figure. Your answer should include code that builds the figure as well as saves it to a file.

require(here)

image\_file = "lab\_04\_hist\_01.png"

png(

here("images", image\_file),

width = 1500, height = 1600,

res = 180)

par(mfrow = c(2, 2))

hist(norm\_17, xlab = "Random Data", ylab = "Frequency", main = "17 Data Points")

hist(norm\_30, xlab = "Random Data", ylab = "Frequency", main = "30 Data Points")

hist(norm\_300, xlab = "Random Data", ylab = "Frequency", main = "300 Data Points")

hist(norm\_3000, xlab = "Random Data", ylab = "Frequency", main = "3000 Data Points")

dev.off()

1. Upload your lab\_04\_hist\_01.png file to Moodle. Make sure you double check the image size and resolution requirements.

Done

1. Qualitatively describe the differences among the histograms.

With the increase of data points, there is an increase in number of bars on the histogram. In addition, as the number of data points increases, the shape of the histogram becomes more “bell shaped” or normally distributed.

1. Explain why the shapes of the histograms are different.

The more data points that are run, the closer the distribution to the normal curve.

1. What are the parameters and their values for the standard Normal distributions?

* Mean = 0
* SD = 1
* Total area under the curve = 1

1. Include the R code you used to create your figure. Your answer should include code that

builds the figure as well as saves it to a file.

image\_file = "norm\_1.png"

png(here("images", image\_file))

x = seq(0, 20, length.out = 1000)

y = dnorm(x, mean = 10.4, sd = 2.4, log = FALSE)

plot(x, y, main = "Normal PDF: mean = 10.4, sd = 2.4", type = "l", xlim = c(0, 20))

abline(h = 0)

dev.off()

1. Upload norm\_1.svg.

Done

1. Show the R code you used to create one of the random datasets in your figure.

set.seed(97)

n\_pts = 100

x\_min = 1

x\_max = 80

x = runif(n = n\_pts, min = x\_min, max = x\_max)

dat = data.frame(x = x, y\_observed = runif(n\_pts))

1. Upload an image file of your figure.

Done

1. Show the R code you used to create one of the random datasets in your figure.

set.seed(97)

n\_pts = 25

x\_min = 1

x\_max = 200

x = runif(n = n\_pts, min = x\_min, max = x\_max)

dat = data.frame(x = x, y\_observed = runif(n\_pts))

guess\_x = 50

guess\_y = 0.3

guess\_slope = .003

plot(y\_observed ~ x, data = dat, pch = 8, main = "Scatterplot: Measuring Error", xlab = "x-value", ylab = "Observed y-value")

curve(line\_point\_slope(x, guess\_x, guess\_y, guess\_slope), add = T)

1. Upload an image file of your figure.

Done

1. Paste the R code you used to create the columns of predicted values and residuals.

line\_point\_slope(dat$x, guess\_x, guess\_y, guess\_slope)

dat = data.frame(x = x, y\_observed = runif(n\_pts), y\_predicted = line\_point\_slope(dat$x, guess\_x, guess\_y, guess\_slope))

dat = data.frame(x = x, y\_observed = rnorm(n\_pts),

y\_predicted = line\_point\_slope(dat$x, guess\_x, guess\_y, guess\_slope),

resids = dat$y\_observed-dat$y\_predicted)

1. In your report, include the two following figures

* A histogram of the model’s residuals
* A scatterplot of your model’s predicted values (on the x-axis) and residuals (on the y-axis)Chart, histogram

  Description automatically generated