# Q1 (10 points)

# A pharmaceutical manufacturer forms tablets by compressing a granular material that contains

# the active ingredient and fillers. The force in kilograms (kg) applied to the tablets varies

# a bit, with the N(11.6, 0.2) distribution. The process specifications call for applying a force

# between 11.2 and 12.0 kg.

mean <- 11.6

sd <- 0.2

x\_min <- 11.2

x\_max <- 12.0

#

# (a) What percent of tablets are subject to a force that meets the specifications?

100 \* (pnorm(x\_max, mean=mean, sd=sd) - pnorm(x\_min, mean=mean, sd=sd))

# 95.44997

# (b) The manufacturer needs that at least 99 percent of the tablets meet the specification. He can

# adjust the process so that standard deviation is at specific rates while the mean remains

# the same mu=11.6 kg.

# What should be standard deviation of the forces so that 99 percent of tablets meet the

# specification?

# Describe your solution:

# Confidence level is 99%, so z = 2.576

# x\_bar = 11.6

#

# x\_bar + z\*(sd/sqrt(n)) = 12.0

# 11.6 + 2.576\*(sd/sqrt(1)) = 12.0

# sd = (12.0-11.6)/2.576

# sd = 0.1552795

#

# x\_bar - z\*(sd/sqrt(n)) = 11.2

# 11.6 - 2.576\*(sd/sqrt(1)) = 11.2

# sd = (11.6-11.2)/2.576

# sd = 0.1552795

# test sd by substituting into pnorm functions

sd <- 0.1552795

100 \* (pnorm(x\_max, mean=mean, sd=sd) - pnorm(x\_min, mean=mean, sd=sd))

# 99.00049

# Q2 (5 points)

# The scores of adults on an IQ test are approximately Normal with mean 100 and standard

# deviation 15.

# What percent of adults have IQ score of 130 or higher?

mean <- 100

sd <- 15

x <- 130

100 \* (1 - pnorm(x, mean=mean, sd=sd))

# 2.275013 (B)

# Q3 (5 points)

# What happens to the variability of the sampling distribution of the mean as the number of

# observations units decreases?

# A. The variability increases

# B. The variability decreases

# C. The variability is independent of the sample size so there is no change

# Central Limit Theorem states that the variance of the sample mean decreases with an increase

# sample size, and the variance of the mean increases with a decrease in sample size.

# (A)

# Q4 (10 points)

# In a study of exercise, a large group of male runners walk on a treadmill for 10 minutes.

# Their heart rates in beats per minute at the end vary from runner to runner according to the

# N(104, 12.6) distribution

# The heart rates for male nonrunners after the same exercise have the N(120, 15) distribution.

# (a) What percent of the runners have heart rates above 130?

x <- 130

mean <- 104

sd <- 12.6

100 \* (1 - pnorm(x, mean=mean, sd=sd))

# 1.953295

# (b) What percent of the nonrunners have heart rates above 130?

mean <- 120

sd <- 15

100 \* (1 - pnorm(x, mean=mean, sd=sd))

# 25.24925

# Q5 (10 points)

# The 95% confidence interval for the population mean was calculated based on a random sample

# with sample size of n=31 as (50 to 60). What is the outcome of a 2-sided alpha = 0.05 test

# based on the null hypothesis of H0: mu = 55.

# A. We would reject the null hypothesis

# B. We would accept the null hypothesis

# C. We would fail to reject the null hypothesis

# D. Not enough information to decide

# H0: mu = 55

# alpha = 0.05

# 2-sided critical value for 5% (z): 1.960

# x\_bar + z\*(sigma/sqrt(n)) = 60

# x\_bar + 1.960\*(sigma/sqrt(31)) = 60

#

# x\_bar - z\*(sigma/sqrt(n)) = 50

# x\_bar - 1.960\*(sigma/sqrt(31)) = 50

#

# 60 - 1.960\*(sigma/sqrt(31)) = 50 + 1.960\*(sigma/sqrt(31))

# 60\*sqrt(31) - 1.960\*sigma = 50\*sqrt(31) + 1.960\*sigma

# 10\*sqrt(31) = 3.920\*sigma

# sigma = 14.203480517423525

# x\_bar - z\*(sigma/sqrt(n)) = 50

# x\_bar - 1.960\*(14.203480517423525/sqrt(31)) = 50

# x\_bar = 50 + 1.960\*(14.203480517423525/sqrt(31))

# x\_bar = 55

# (B) Accept the null hypothesis

# Q6 (5 points)

# A two sample t-test was conducted to determine if there was a difference in the underlying

# mean BMI between smokers and non-smokers. The t-statistic was calculated to be 1.7 with 20

# degrees of freedom. What is the associated p-value?

# A. 0.025

# B. Between 0.025 and 0.05

# C. 0.052

# D. Between 0.05 and 0.10

# E. 0.10

# according to table, for df=20 and t-statistic 1.724718, p = 0.05

# and t-statistic 1.325341, p = 0.10, so the answer is D

# Q7 (10 points)

# Do college students who have volunteered for community service work differ from those who

# have not?

# A study obtained data from 57 students who had done service work and 18 who had not.

#

# --------------------------------------------------

# Group | Condition | n | x\_bar | S |

# --------------------------------------------------

# 1 | Service | 57 | 104.12 | 14.60 |

# 2 | No Service | 18 | 97.82 | 14.20 |

# --------------------------------------------------

#

# Is there strong evidence (alpha = 0.05 Significance level) that students who have engaged in

# community service are on the average more attached to their friends?

# H0: students who have engaged in community service are on average more attached to their friends

# in other words, t > 0 in test below

x\_bar\_s <- 104.12

s\_s <- 14.60

n\_s <- 57

x\_bar\_ns <- 97.82

s\_ns <- 14.20

n\_ns <- 18

t <- (x\_bar\_s - x\_bar\_ns)/sqrt((s\_s^2/n\_s) + (s\_ns^2/n\_ns))

# t = 1.629814. Proving that students who have engaged in community service are on average more

# attached to their friends.

# Q8 (20 points)

# We have data on the lean body mass and resting metabolic rate for 10 women who are subjects

# in a study on dieting. Lean body mass, given in kilograms, is a person's weight leaving out all

# fat. Metabolic rate, in calories burned per 24 hours, is the rate at which the body consumes

# energy.

#

# ---------------------------------------------------------------------------------------

# Mass | 35.2 | 54.6 | 48.5 | 42.0 | 50.6 | 42.0 | 40.3 | 33.1 | 42.4 | 34.5 |

# Rate | 991 | 1455 | 1395 | 1418 | 1502 | 1246 | 1189 | 913 | 1124 | 1052 |

# ---------------------------------------------------------------------------------------

#

mass <- c(35.2,54.6,48.5,42.0,50.6,42.0,40.3,33.1,42.4,34.5)

rate <- c(991,1455,1395,1418,1502,1246,1189,913,1124,1052)

# (a) Find the least-squares regression line for predicting metabolic rate from body mass.

lm(rate~mass)

# Beta\_0 = 115.80

# Beta\_1 = 26.29

beta\_0 <- 115.80

beta\_1 <- 26.29

# equation for the least-squares regression line for predicting metabolic rate from body mass

# is: y = 115.80 + 26.29\*x

#

# (b) Another woman has lean body mass 48 kilograms. What is her predicted metabolic rate?

x <- 48

beta\_0 + beta\_1\*x

# 1377.72

#

# (c) What percentage of the variability in the metabolic rate can be explained by the body

# mass?

# index,xi,yi,y\_hat\_i

# 1,35.2,991

# 2,54.6,1455

# 3,48.5,1395

# 4,42.0,1418

# 5,50.6,1502

# 6,42.0,1246

# 7,40.3,1189

# 8,33.1,913

# 9,42.4,1124

# 10,34.5,1052

# sample mean,42.32,1228.5

# sample sd,7.139063119,208.0658283

#

# xi,yi,y\_hat\_i,y\_hat\_i-y\_bar,(y\_hat\_i-y\_bar)^2,y\_i-y\_hat\_i,(y\_i-y\_hat\_i)^2

# 35.2,991,1041.208,-187.292,35078.29326,-50.208,2520.843264

# 54.6,1455,1551.234,322.734,104157.2348,-96.234,9260.982756

# 48.5,1395,1390.865,162.365,26362.39323,4.135,17.098225

# 42,1418,1219.98,-8.52,72.5904,198.02,39211.9204

# 50.6,1502,1446.074,217.574,47338.44548,55.926,3127.717476

# 42,1246,1219.98,-8.52,72.5904,26.02,677.0404

# 40.3,1189,1175.287,-53.213,2831.623369,13.713,188.046369

# 33.1,913,985.999,-242.501,58806.735,-72.999,5328.854001

# 42.4,1124,1230.496,1.996,3.984016,-106.496,11341.39802

# 34.5,1052,1022.805,-205.695,42310.43303,29.195,852.348025

# ,,,,317034.3229,,72526.24893

# ---------------------------------------------------------------------------------

# | SS | df | MS |

# ---------------------------------------------------------------------------------

# Regression | 317034.3229 | k=1 | 317034.3229/1=317034.3229 |

# Residual | 72526.24893 | n-k-1=10-1-1=8 | 72526.24893/8=9065.78111625 |

# ---------------------------------------------------------------------------------

# Total | 389560.5719 | | |

# ---------------------------------------------------------------------------------

# R^2 = reg\_ss/total\_ss

317034.3229/389560.5719

# 0.8138255. So, 81.38255% of the variability in the metabolic rate can be explained

# by the body mass.

# Q9 (5 points)

# A least-squares simple linear regression model was fit predicting husbands' age based the age

# of the wife from a sample of 12 couples.

# What percentage of the variability in the husbands' age can be explained by the age of the

# wife?

#

# -----------------------------------------------------------

# | SS | df | MS | F-statistic |

# -----------------------------------------------------------

# Regression | 10 | 1 | 10 | 5 |

# Residual | 20 | 10 | 2 | |

# -----------------------------------------------------------

# Total | 30 | | | |

# -----------------------------------------------------------

# reg\_df = k = 1

# res\_df = n - k - 1 = 10

# res\_ms = res\_ss/res\_df = 2 = res\_ss/10 => res\_ss = 2\*10 = 20

# reg\_ss + res\_ss = total\_ss = 30 = reg\_ss + 20 => reg\_ss = 30-20 = 10

# reg\_ms = reg\_ss/reg\_df = 10/1 = 10

# f-statistic = reg\_ms/res\_ms = 10/2 = 5

# R^2 = reg\_ss/total\_ss = 10/30 = 0.3333333

# 33.333% of the variability in the husbands' age can be explained by the age of the wife.

# Q10 (20 points)

# A research studied the correlation between physical characteristics of sisters and brothers.

# Here are data on the heights (in inches) of 11 adult pairs.

# -----------------------------------------------------------------------------

# Brother | 71 | 69 | 66 | 67 | 70 | 71 | 70 | 73 | 72 | 65 | 66 |

# Sister | 69 | 63 | 65 | 63 | 65 | 62 | 65 | 64 | 66 | 59 | 62 |

# -----------------------------------------------------------------------------

# (a) Find the correlation and the equation of the least-squares line for predicting sister's height

# from brother's height.

brother <- c(71,69,66,67,70,71,70,73,72,65,66)

sister <- c(69,63,65,63,65,62,65,64,66,59,62)

r <- cor(brother,sister)

# r = 0.5596833

lm(sister~brother)

# Beta\_0 = 26.8653

# Beta\_1 = 0.5362

beta\_0 <- 26.8653

beta\_1 <- 0.5362

# equation for the least-squares regression line for predicting sister's height from brother's

# is: y = 26.8653 + 0.5362\*x

#

# (b) Carlos is 72 inches tall. Predict the height of his sister.

x <- 72

y <- 26.8653 + 0.5362\*x

# y = 65.4717 inches

#

# (c) Based on the scatterplot and the correlation r, do you expect your prediction to be very

# accurate? Why?

plot(brother, sister)

# scatterplot shows a positive relationship and a positive slope. with r = 0.5596833, which is

# much closer to 1 than -1, this confirms the positive slope. Though, because the correlation is

# not closer to the value of 1, I would claim that the prediction is somewhat accurate. There is

# a significant trend, so there is definitely a degree of accuracy to the prediction. However,

# I would hesitate to go as far as to claim it is very accurate.

#

# (d) Is there evidence of a significant linear association between physical characteristics of

# sisters and brothers (alpha = 0.05 level)?

# 1

# H0:ρ=0 (there is no linear association)

# H1:ρ≠0 (there is a linear association)

# alpha=0.05

# 2

# t = r\*sqrt((n-2)/(1-r^2))

# 3

# Determine the appropriate value from the t-distribution table with n−2=11−2=9

# degrees of freedom and associated with a right hand tail probability of alpha/2=0.025

# Using the table, t=2.262157

n <- length(brother)

df <- n-2

qt(0.975, df=df)

# Decision Rule: Reject H0 if t≥2.262157 or if t≤−2.262157 (|t|≥2.262157).

# Otherwise, do not reject H0

# 4

# 5

t = r\*sqrt((df)/(1-r^2))

# from t-distribution table, t=2.026109 for df=9 and alpha/2 = 0.025

# |t| = 2.026109 < 2.262157

# Therefore, we do not reject the null hypothesis and there is not enough significant linear

# association between physical characteristics of sisters and brothers at alpha level 0.05.