Problem Set #3

Josh Klier 4/3/19

Question #1

plot(newhamp$pObama, newhamp$Dean, main= "Proportion of New Hampshire voter preference between Dean and Obama and hand counted vs machine ballots", cex.main=.5, xlab="proportion Obama", ylab="proportion Dean", col=my\_cols, xlim=c(0,1), ylim=c(0,1))

my\_cols <- ifelse(newhamp$votesys=="H", "Blue", "Red")

legend("topleft", legend = c("Hand counted", "Machine counted"), col = c("Blue", "Red"), lty = 1, cex = 1)



Question #2

when the df is large, the t-distribution approximates the normal distribution.

t.values <- seq(-4,4,.1)

plot(x=t.values, y=dt(t.values,20), type="l", col="red", ylim = c(0,.4), xlab = "t distribution", ylab = "frequency", main = "differing dfs")

lines(x=t.values, y=dt(t.values,3), type = "l", col="blue", lty=2)

lines(x=t.values, y=dt(t.values,1), type= "l", col="green", lty=3)

lines(t.values, dnorm(t.values), type="l", col="Black", lty=4)

legend("topright", title="df", legend = c("20","3","1","dnorm"), lty = c(1,2,3,4), col=c("red","blue","green","black"), cex=.5)

My plot shows that as the number of dfs goes up, the graph gets closer to resembling a normal distribution.

#df= sample size- 1. so as df goes up, the sample size goes up and vice versa. So as the sample size goes up the graph more closely resembles a normal distribution which will affect the population parameters.



Question #3

A)

#null hypothesis = average voting age among american voters equals 50

#alternative hypothesis = average voting age among american voters is different than 50

B)

standard.error <- sd(voteincome$age, na.rm = TRUE)/sqrt(length(voteincome$age[!is.na(voteincome$age)]))

standard.error

# standard error = .4511027

avgvote <- voteincome$age[voteincome$vote=="1"]

voteincome[voteincome$vote=="1", "age"]

# I am taking the data from american voters. Meaning they had to vote. Meaning vote=1.

t.test(avgvote, mu=50)

mean(avgvote)

49.6516

zscore <- ((50 - 49.6516)/(standard.error))

zscore\*2

#1.544659, corresponds with a p value of,

#p value 0.122429

C)

p value is greater than alpha so we cant reject the null.

#the p value is not less than significance level of .05, therefore the null hypothesis that the mean=10 cant be rejected. We fail to reject the null.

D)

95% confidence interval

49.6516 + 1.96\*(.4511027)

49.6516 - 1.96\*(.4511027)

(48.76744, 50.53576)

E)

In C we say that we cant reject the null because P>alpha.

In D we are showing this by calculating the 95% confidence interval

the fact that the 95% confidence interval contains 50, means that we can't reject the null because the null is that the avg is 50.

Question #4

A)

sqrt(pi0(1-pi0)/n). one sample t test

t\*= ybar-mu null/(s/sqrt(n))

given the small sample size we will use a t statistic. Assume normal distribution

B)

were need to assume that the null hypothesis is true.

C)

((9.5-10)/(1.2/sqrt(16)))

t\*=-1.66667

df=n-1=15, p-value= 1.753, one tailed t table. One tailed because alternative hype is Ha<10.

P>alpha, we fail to reject the null.

D)

We are already assuming that we know the population standard deviation because of normal distribution. So use the same test. Also we are using a sampling distribution so we are using the standard error of the sampling distribution.

E)

the sample standard deviation equals the population standard deviation.

F)

test statistic -1.66667, se = 1.2/sqrt(16), df=15 pvalue=1.753

(1.2/sqrt(16))

#se = .3

((9.5-10)/(1.2/sqrt(16)))

# t statistic = -1.666667

G)

they're the same

Question #5

A)

binomial distribution for population distribtion. The type of sampling distribution is normal because of the sample size.

B)

341/698

.488

C)

assume .488=pi0

sqrt(.488\*(1-.488)/698)

.018 standard error

D)

95% confidence interval

z.score <- qnorm(1-.05/2)

z.score

1.96

.488 + 1.96\*(.018)

.488 - 1.96\*(.018)

(.45272, .52328)

Question #6

A)

face‐to‐face voter mobilization was effective in stimulating voter turnout across a wide spectrum of local elections.

B)

people who received face‐to‐face contact from a coalition of nonpartisan student and community organizations, encouraging them to vote.

C)

voter turnout rates

D)

they had treatment and control groups. Since the treatment group had higher turnout rate than the control group, they can make a causal statement.

Question #7

A)

null hypothesis population mean male= population mean female. Ha = hypo pop mean male doesnt equal hypo pop mean female

se=sqrt((s1)^2/n1+(s2)^2/n2)

sqrt((2.34)^2/1117+(2.22)^2/870)

se= .1027954

t stat = estimate of paramater - null hypo of parameter / se of estimate

((2.86-2.99)-0)/.1027954

t stat = -1.264648

p value = .103

p>alpha, can't reject the null.

B)

yes because we can't reject the null. Meaning the the difference in means could be zero.

this is shown by the 95% confidence interval.

(2.86-2.99)+ 1.96\*(.102)

(2.86-2.99)- 1.96\*(.102)

(-.32992, .06992)

C)

yes it's approximately normal because of the sample size of 1117 and 870.

This affects the validity of the inferences because we are assuming taht our data is normally distributed.

Question #8

A)

se

sqrt((2.34)^2/11+(2.22)^2/16)

se = .8976674

t stat

(2.86-2.99-0)/.8976674

t stat = -.1448198

11+16= 27 = N

N-2=25 = DF

P=2.060

P>.05, can't reject the null.