

Rewrite of script confidence intervals.R

We have a conceptual mistake in the last lab which I want to clear up. In the lab write up, it says that for n_{10} , n_{30} , and n_{100} , the CI's are (-.0052, .5452), (.1111, .4289), and (.1830, .3570), respectively. THESE ARE NOT CI's. but rather are 2 standard deviations either side of the center of the sampling distribution of the population (of $p = .27$) for $n = 10, 30$, and 100 , respectively. This formula for the CI is stated correctly on page 6, and will be each sample proportion (\hat{p}) + or -

$1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where $1.96 = z$ for 95% CI. So, we must change the script given to you (which captured p much better than 95% of the time, called `confidence_intervals.R` to that shown below.

```
# Confidence Intervals

# initiate constants
n.samples <- 100
n <- 100
p1 <- .28

# create samples and histogram of phats
distrib1 <- matrix(rbinom(n.samples*n, size=1,
  prob=p1), ncol=n)
sample1 <- rowMeans(distrib1[,1:n])
hist(sample1)

# compute std. dev. of histogram (2 std errors)
summary(sample1)
sd(sample1)
diff <- 1.96 * sqrt(p1*(1-p1)/n)
lower <- p1 - diff
upper <- p1 + diff
print("95% CI is")
lower ; upper
hist(sample1)
abline(v=lower,lwd=4)
abline(v=upper,lwd=4)

# count proportion of CI's not capturing p
diff1 <- 1.96 * sqrt(sample1 * (1 - sample1)/n)
lower1 <- sample1 - diff1
upper1 <- sample1 + diff1
test1 <- which(lower1>p1 | upper1<p1)
length(test1)
cat("proportion of ci's not capturing",
  p1, "are", length(test1)/n.samples, "\n")
cat("sample interval numbers not capturing p are",
  which(lower1>p1 | upper1< p1), "\n")
```

```

# plot ci's

x1 <- 1:n.samples
y1 <- lower1
y2 <- upper1
x <- c(x1,x1)
y <- c(y1,y2)
plot(x,y)
abline(p1,0)
lines(x1,y1)
lines(x1,y2)

```

This script differs from the original one offered in the lab by the following features:

- [line7] starting with `distrib1 <- matrix`, has `ncol=n` now, instead of 10. This was causing all runs of $n=10$ to come out the same!
- We created a `lower1`, `upper1`, and `diff1`, now, corresponding to the \hat{p} values of each sample obtained, instead of using the `lower`, `upper`, and `diff` values determined from the p value = .28
- We now check that the `upper1` limit of each little confidence interval with the $p = .28$, AS WELL AS check the `lower1` limit, to make sure we straddle the .28 with our sample CI. If not, we count the interval as missing the true parameter $p = .28$.

Use this script to replace the incorrect one given earlier, and remember that \hat{p} 's are used in CI determinations, NOT parameter p values, for the test!! And, thank you Brian S. for finding this error!!! I have posted the correct script on the web site.