STAT6021_Mod0Qs

Kip McCharen

7/13/2020

Topic 1: Sampling Distributions

1a: $\bar{x}$ (the sampling distribution of the sample mean) = $\boldsymbol{N}(\mu, \frac{\sigma}{\sqrt{n}})$

1b: the central limit theorem informs us that the sampling distribution of the sample mean, $\bar{x}$, can be well-approximated by a normal distribution.

2a: Probability of subcomponent length: one-tailed test.

```
roundme <- 4

value = 118
sample_mean = 116
ssd = 4.8
zscore = (value - sample_mean)/ssd
print(paste("zscore: ", round(zscore, roundme)))
```

```
## [1] "zscore:  0.4167"
```

```
curve <- pnorm(zscore)
print(paste("area under the curve: ", round(1-curve, roundme)))
```

```
## [1] "area under the curve:  0.3385"
```

There is ~33.85% probability that one selected subcomponent is longer than 118cm.

2b: 1 tailed test checking above sample mean length.

```
original_SD = 4.8
sample_size = 3
ssd = original_SD / sqrt(sample_size)
value = 118
sample_mean = 116
zscore = (value - sample_mean)/ssd
print(paste("zscore: ", round(zscore, roundme)))
```

```
## [1] "zscore:  0.7217"
```

```r
curve <- pnorm(zscore)
print(paste("area under the curve: ", round(1-curve, roundme)))
```

```
## [1] "area under the curve:  0.2352"
```

There is about a 23.52% chance that 3 randomly selected subcomponents have a mean length > 118cm.

Topic 2: Confidence Intervals

3: We construct confidence intervals to be as accurate as possible in quantifying the amount of uncertainty in an estimation of plausible values for an unknown value of interest.

4: If we increase the confidence level, the margin of error increases. The bigger they are, the harder they fall.

---

5 As the sample size increases, so too increases the denominator of the standard error of the sample mean, thus reducing the magnitude of the margin of error. In clearer language, with more samples we can assume our data will look more exactly like a normal curve and we can also assume our estimates will be more generally accurate (with less error).

---

6 Finding t-multipliers for μ CI

```r
t_mult <- function(alpha=0.05, sample_size) {
  percentile = (1-(alpha / 2))
  print(paste((percentile * 100), "th percentile value in t distribution", sep = ""))
  deg_free = sample_size - 1
  t_mult <- qt(percentile, deg_free)
  print(paste("t multiplier for curr alpha/percentile:  ", round(t_mult, roundme)))
  print("")
}
t_mult(0.06, 49)
```

```
## [1] "97th percentile value in t distribution"
## [1] "t multiplier for curr alpha/percentile:   1.9263"
## [1] ""
```

```r
t_mult(0.14, 82)
```

```
## [1] "93th percentile value in t distribution"
## [1] "t multiplier for curr alpha/percentile:   1.4904"
## [1] ""
```

```r
t_mult(0.26, 150)
```

```
## [1] "87th percentile value in t distribution"
## [1] "t multiplier for curr alpha/percentile:   1.1307"
## [1] ""
```

7a GPA sample of 100 students

```
conf_interval <- function(s_n, s_mean, s_SD, integer_confidence_pct) {
  alpha <- 1 - (integer_confidence_pct/100)
  deg_free <- s_n - 1
  multiplier <- (1 - (alpha / 2))#two tailed test
  print(paste((multiplier * 100), "th percentile value in normal distribution"))

  Z_score <- qnorm(multiplier, s_mean, s_SD)
  print(paste("z score:   ",round(Z_score,roundme)))

  t_mult <- qt(multiplier, deg_free)
  print(paste("t multiplier:   ",round(t_mult,roundme)))

  stand_error_of_s_mean = s_SD/(sqrt(s_n))
  print(paste("SE of estimate:   ", round(stand_error_of_s_mean,roundme) ))
              #standard error because estimating w sample mean instead of pop mean
              #estimate vs sample mean?
  margin_of_error <- t_mult * stand_error_of_s_mean

  print(paste("one sided interval / margin of error: ",round(margin_of_error,roundme) ))

  print(paste(integer_confidence_pct, "% confidence interval : ",
        round(s_mean - margin_of_error,roundme)," to ",round(s_mean + margin_of_error,roundme)))
}
conf_interval(100, 3.2, 0.2, 97)
```

```
## [1] "98.5 th percentile value in normal distribution"
## [1] "z score:    3.634"
## [1] "t multiplier:    2.2018"
## [1] "SE of estimate:    0.02"
## [1] "one sided interval / margin of error:  0.044"
## [1] "97 % confidence interval :  3.156  to  3.244"
```

7b The margin of error for above is 0.044. It tells us that we can be 98.5% confident that the true population mean GPA is 0.044 around the estimate.

7c Based on this confidence interval, no it is not reasonable to say that, it is outside of the very high % confidence interval.

Topic 3: Hypothesis Testing

8 The goal of conducting a hypothesis test is to test whether the evidence is statistically insignificant and demonstrates no effect or no difference in variables, or whether the evidence is statistically significant so we can reject that null hypothesis and accept the alternative possibility.

---

9 Hypothesis statements are always about the population parameter.

---

10. For each of the situations, state the appropriate null and alternative hypotheses, in symbols and in words. Sketch how you would find the p-value based on the calculated test statistic.

(a) David's car averages 29 miles per gallon on the highway. He just switched to a new motor oil that is advertised as increasing gas mileage. He wants to investigate if the advertisement is accurate.

(b) The diameter of a spindle in a small motor is supposed to be 4 millimeters. If the spindle is too small or too large, the motor will not function properly. The manufacturer wants to investigate whether the mean diameter is moved away from the target.

(c) The average time in traffic between 2 points of a congested highway used to be 2 hours. The government invested money to improve travel times by building extra lanes and overpasses. Citizens want to access if travel times have improved, on average.

10a $\boldsymbol{H_0}$