Hao-4

Bruce Hao

9/25/2016

```
library(dplyr)
library(ggplot2)
library(ggthemes)

# setup normal plots
xv = seq(-4, 4, 0.01)
yv = dnorm(xv, 0, 1)
df = data.frame(xv, yv)
p = ggplot(df, aes(x=xv, y=yv)) + geom_line()
```

4.4

a) Point estimate of average = 171.1; point estimate of median = 170.3
b) Point estimate of standard deviation = 9.4; point estimate of IQR = 177.8 - 163.8 = 14
c) A 180cm person is 0.95 sd taller than the mean and thus taller than about 83% of the population of active individuals. A 155cm person is 1.7 sd shorter than the mean and thus taller than only about 4% of the population.
d) No, some variation across samples should be expected.
e) Standard error = 9.4 / sqrt(507) = 0.42.

4.14

a) False, the confidence interval represents our confidence that the average spending for all American adults fall within the interval.
b) False, the sample is probably large enough that the skew can be ignored when determining the confidence interval.
c) True.
d) True.
e) True.
f) False, since the denominator of the standard error calculation is the square root of the number of observations within the sample, we would need a 9x bigger sample to decrease the standard error 3x (and thus tighten the interval by a third).
g) True.

4.24

a) Yes, 36 students likely represents <10% of the population in a large city thus satisfying the independence requirement; the data do not appear particularly skewed, and there are >30 observations.

b)

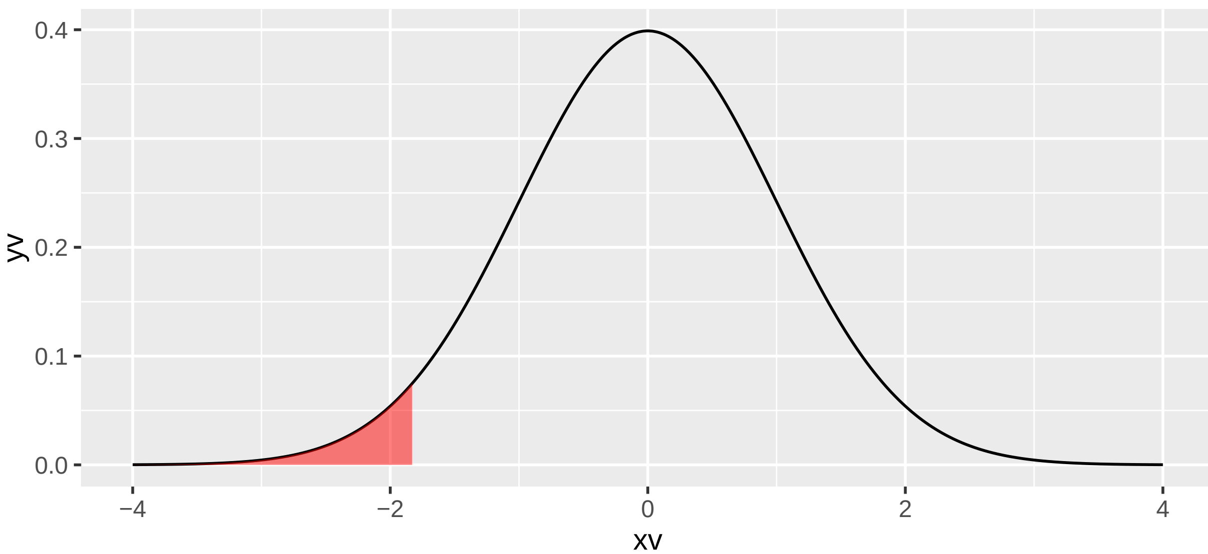
- H_0: $mu_{(gifted\ children)} = 32$
- H_A: $mu_{(gifted\ children)} < 32$

Observed difference = (30.69 - 32) / (4.31 / sqrt(36)) = -1.82. This represents a p-value of 0.034, which is less than the significance level of 0.10; as such, we reject H_0 and accept H_A that gifted children learn to count to 10 more quickly than children in the general population do on average.

```
shade = df[df$xv < (-1.82), ]
pnorm(-1.82)
```

```
## [1] 0.0343795
```

```
p + geom_ribbon(data = shade, aes(ymax=yv, ymin=0), fill = 'red', alpha = 0.5)
```



d) 90% CI = 30.69 +/- 1.64 * 4.31 / sqrt(36) = (29.51, 31.87)

e) Yes, this agrees with the hypothesis test. Our CI suggests that we are 90% confident that on average gifted children learn to count to 10 between 29.51 and 31.87 months of age. This interval excludes the 32 months it takes the children in the general population on average to learn the same. As such, the difference in speed is statistically significant at the 0.10 significance level.

4.26

a)

- H_0: $mu_{(\text{gifted mother IQ})} = 100$
- H_A: $mu_{(\text{gifted mother IQ})} > 100$

Observed difference = (118.2 - 100) / (6.5 / sqrt(36)) = 16.8. This represents a p-value of basically 0.00, which is less than the significance level of 0.10; as such, we reject H_0 and accept H_A that the mothers of gifted children have statistically significant higher IQs than do mothers in general.

b) 90% CI = 118.2 +/- 1.64 * 6.5 / sqrt(36) = (116.42, 119.98)

c) Yes, this agrees with the hypothesis test. Our CI suggests that we are 90% confident that on average mothers of gifted children have IQs between 116.42 and 119.98. This interval excludes the 100, the average IQ for mothers in general. As such, the difference in IQ is statistically significant at the 0.10 significance level.

4.34

The term ‘sampling distribution’ of the mean describes the distribution of averages of various samples of the population. With enough samples, the distribution of the sample averages should be approximately normal and centered on the population average. As the number of samples increases, in addition to the distribution becoming more and more normal, the variation of the distribution also decreases.

4.40

a) 6.7%

b) Since the population is normally distributed, n<30 is not a concern; thus, the mean lifespan of 15 light bulbs should be normally distributed, centered on the population mean of 9000 with an sd equal to the standard error of 1000 / sqrt(15) = 258.20.

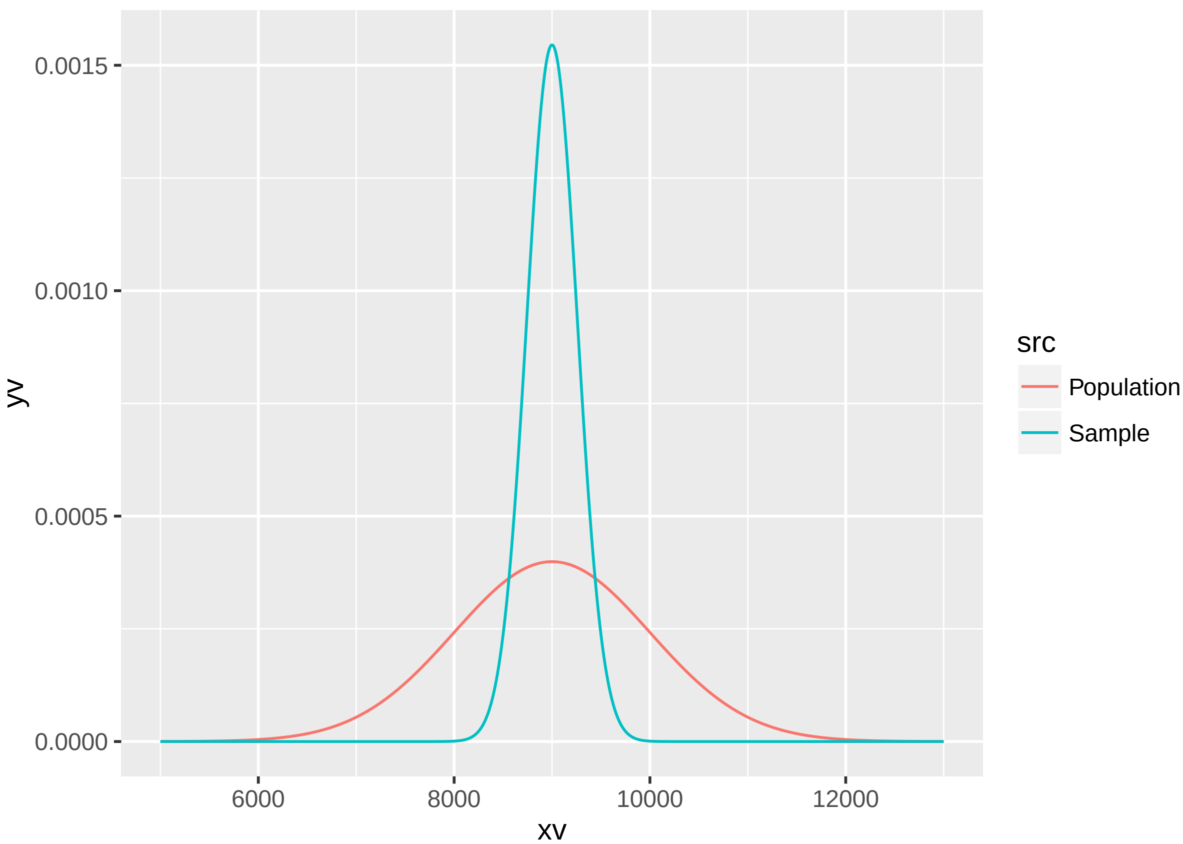
c) Basically zero. A mean life span of 10500 hours is almost 6 standard errors away from the population mean.

d)

```
xv = seq(9000 - 4*1000, 9000 + 4*1000, 1)
yv = dnorm(xv, 9000, 1000)
dfPop = data.frame(src = 'Population', xv, yv)
xv = seq(9000 - 4*1000, 9000 + 4*1000, 1)
yv = dnorm(xv, 9000, 1000 / sqrt(15))
dfSamp = data.frame(src = 'Sample', xv, yv)
df = bind_rows(dfPop, dfSamp)
```

```
## Warning in bind_rows_(x, .id): Unequal factor levels: coercing to character
```

```
p = ggplot(df, aes(x=xv, y=yv, color=src)) + geom_line()
p
```



e) We could estimate the probabilities for part a because the sample size is large; however, we could not do the same for part c because the sample size is too small.

4.48

All else equal, the higher the sample size, the smaller the standard error, the higher the z-scores and t-statistics, and the lower the p-values.