

DATA 606 Homework 4

Ilya Kats

March 19, 2017

```
library('DATA606')
```

4.4 Heights of adults.

- Average height is 171.1. Median is 170.3.
- Standard deviation is 9.4. IQR is $Q3 - Q1 = 177.8 - 163.8 = 14$.
- $Z_{180} = \frac{x-\mu}{\sigma} = \frac{180-171.1}{9.4} \approx 0.95$ and $Z_{155} = \frac{155-171.1}{9.4} \approx -1.71$ The height of 180 cm is within 1 standard deviation from the mean, so it is not considered unusual. The height of 155 cm is within 2 standard deviations from the mean, and although it is more uncommon than 180 cm, it is still not unusually short.
- With another sample I would not expect the same point estimates. They approximate population values, but vary between samples.
- Variability of the estimate can be measured using standard error, $SE = \frac{\sigma}{\sqrt{n}} = \frac{9.4}{\sqrt{507}} \approx 0.417$.

4.14 Thanksgiving spending, Part I.

- FALSE.** Inference is measured on the population parameter. The point estimate of this sample (or any other point estimate) is always within the confidence interval.
- FALSE.** The sample is sufficiently large ($n = 436$) to account for the skew.
- FALSE.** Confidence interval for the mean of a sample is not about other sample means.
- TRUE.** This is similar to point a. The population parameter (average spending of all American adults) is being estimated by the point estimate and the confidence interval.
- TRUE.** With a 90% confidence interval we do not need such a wide interval to catch the values, so the interval would be narrower.
- FALSE.** In order to decrease the margin of error by 3, we need to increase the sample by $3^2 = 9$ (since n is under the square root in the margin of error formula).
- TRUE.** The margin of error is half the confidence interval or $\frac{89.11-80.31}{2} = 4.4$.

4.24 Gifted children, Part I.

- The sample is random and 36 children of a large city is certainly under 10% of the population. The sample size is over 30. There isn't appear to be any strong skew in the population. Based on this information the conditions for inference are satisfied.

b.

$$H_0 : \mu = 32 \text{ and } H_A : \mu < 32$$

$$SE = \frac{4.31}{\sqrt{36}} \approx 0.72$$

$$Z = \frac{30.69-32}{0.72} = -1.819444$$

Since $p - value = 0.0344219 < 0.10 = \alpha$, we **reject the null hypothesis** H_0 in favor of H_A .

- c. If the null hypothesis is true, then the probability of observing a sample mean lower than 30.69 for a sample of 36 children is only 0.0344 (p-value).
- d. The 90% confidence interval is $30.69 \pm 1.65 * SE = 30.69 \pm 1.188$ or (29.502, 31.878).
- e. Results from the hypothesis test and the confidence interval seem to agree. We are 90% confident that the average age at which gifted children first count to 10 is between 29.5 and 31.9 months. This is lower than the average age for all children at 32 months.

4.26 Gifted children, Part II.

a.

$$H_0 : \mu = 100 \text{ and } H_A : \mu \neq 100$$

$$SE = \frac{6.5}{\sqrt{36}} \approx 1.08$$

$$Z = \frac{118.2 - 100}{1.08} = 16.85185$$

With a Z value over 16, the p-value, even with a two-sided test, is close to 0. If the null hypothesis is true, then the probability of observing a sample mean as different as in our sample is lower than the significance level, α , of 0.10. We reject the null hypothesis H_0 in favor of H_A .

- b. The 90% confidence interval is $118.2 \pm 1.65 * SE = 118.2 \pm 1.782$ or (116.418, 119.982).
- c. Results from the hypothesis test and the confidence interval seem to agree. We are 90% confident that the average IQ of mothers of gifted children is between 116.4 and 120. This is significantly above population average of 100.

4.34 CLT.

The sampling distribution of the mean is the distribution of sample means of multiple samples. Per the Central Limit Theorem, it can be approximated by a normal model. As sample size increases the normal approximation becomes better and the spread of the sampling distribution of the mean becomes narrower.

4.40 CFLBs.

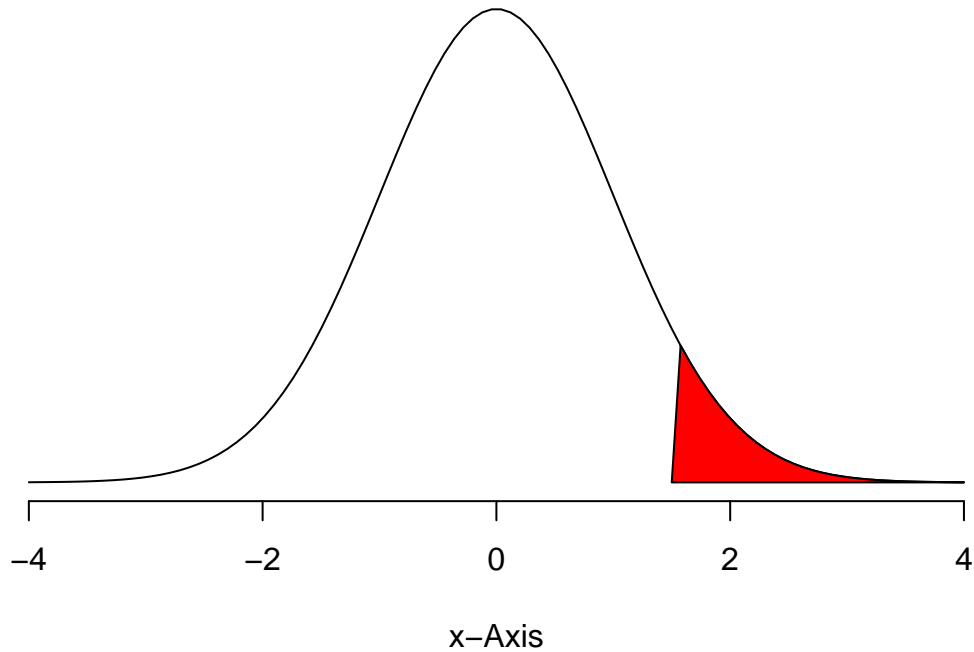
a.

$$Z = \frac{x - \mu}{\sigma} = \frac{10500 - 9000}{1000} = 1.5$$

```
normalPlot(mean = 0, sd = 1, bounds=c(1.5,4), tails = FALSE)
```

Normal Distribution

$$P(1.5 < x < 4) = 0.0668$$



Probability of $x > 10500$ is $1-\text{pnorm}(1.5)$ or 0.0668072.

b. Assuming light bulbs are selected at random, the distribution of the mean lifespan of 15 light bulbs is nearly normal with distribution $N(\mu, \frac{\sigma}{\sqrt{n}})$ or $N(9000, 258.1989)$.

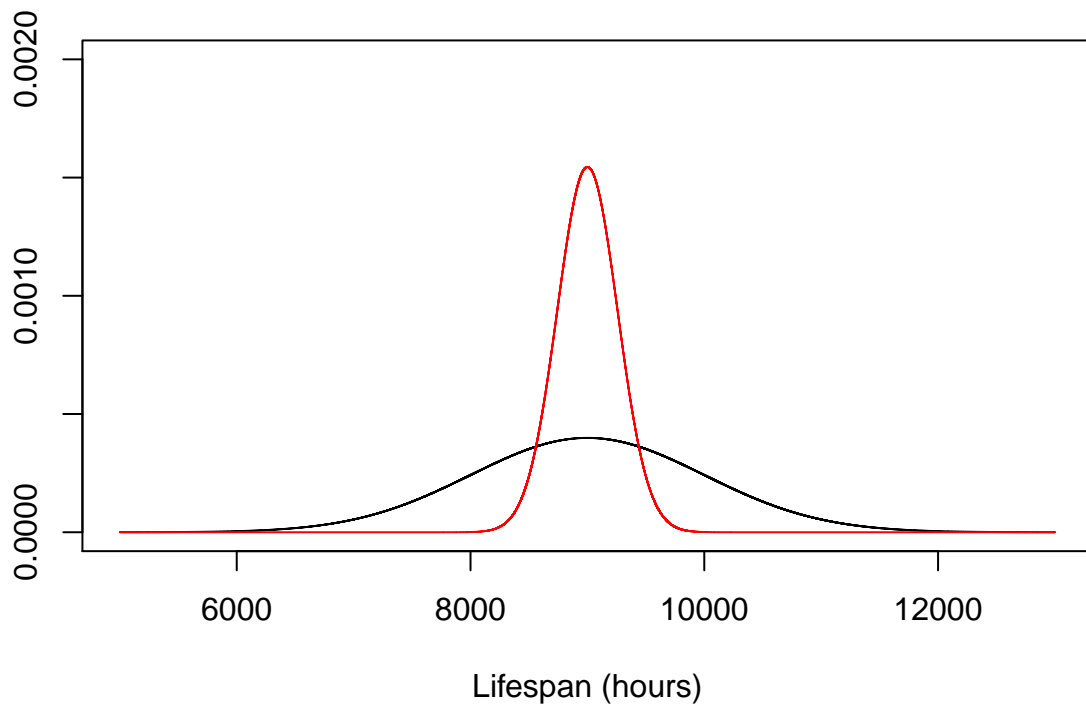
c.

$$Z = \frac{10500 - 9000}{258.1989} \approx 5.81$$

The probability is $1-\text{pnorm}(5.81)$ or ≈ 0 .

d. Population distribution is black, sampling distribution is red.

```
s <- seq(5000,13000,0.01)
plot(s, dnorm(s,9000, 1000), type="l", ylim = c(0,0.002), ylab = "", xlab = "Lifespan (hours)")
lines(s, dnorm(s,9000, 258.1989), col="red")
```



- e. We could not estimate part a without a nearly normal distribution. We could not estimate part c since the sample size is not sufficient to yield a nearly normal sampling distribution to account for a skewed distribution.

4.48 Same observation, different sample size.

$$Z = \frac{\text{point estimate} - \text{null value}}{SE_{\text{point estimate}}}, \text{ where } SE_{\text{point estimate}} = \frac{s}{\sqrt{n}}$$

If n is increased from 50 to 500, then SE will decrease and alternatively Z will increase in case of positive Z and decrease in case of negative Z . As Z is changed, p -value will decrease.