Activity 2

EMH

2025-10-30

1

This problem relies on the following set of variables: ethnicity, education level, race, age, sex, BMI, AHI, ESS, MMSE, adherence (for grouping), ODSI score at baseline, and ADCS-MCI caregiver score at 12 months.

A

Create a "Table 1: Demographic and Clinical Characteristics (n=174)" using the following notes: Column headers should include "Characteristic", along with "Total", "Non-adherent", "Adherent", and "Effect Size". The last column should provide a p-value for appropriate statistical comparisons of distributions across adherent groups for each characteristic. Include sample sizes in parentheses in the column headers for the adherent and non-adherent groups. Populate the cells with the appropriate measures of central tendency and variability and label appropriately. Include 95% confidence intervals where appropriate and label appropriately. Use footnotes to illustrate the statistical tests used for comparisons.

B

Choose a single characteristic with a significant p-value when comparing between adherence and non-adherent groups, and describe in 2-3 sentences what this means in plain English. ### C Choose a single characteristic with a non-significant p-value when comparing between adherence and non-adherent groups, and describe in 2-3 sentences what this means in plain English.

2

In the Observation and Interview Based Diurnal Sleepiness Inventory (ODSI), each of the three items is rated on a seven-point Likert scale. The first item examines sleepiness during basic activities of daily living; the second item relates to falling asleep during periods of inactivity; and the third item asks about hours of daytime sleep. The total ODSI score ranges from 0 (no somnolence) to 24 (excessive somnolence), so a lower score is better. Test the null hypothesis that there is no difference in change from baseline to 6 months for ODSI for adherent versus non-adherent participants. Write out each step of the hypothesis test and clearly interpret your results in plain English in 2-3 sentences.

3

Based on the ODSI, a cutoff score of 6 or higher is used to identify older adults with excessive daytime sleepiness. Test the null hypothesis that the probability of excessive daytime sleepiness is the same at baseline versus at 6 months. Write out each step of the hypothesis test and clearly interpret your results in plain English in 2-3 sentences.

4

The Mini Mental State Examination (MMSE) is a tool that can be used to systematically assess mental status. It is an 11-question measure that tests five areas of cognitive function: orientation, registration, attention and calculation, recall, and language. The maximum score is 30, where higher scores are better. A score of 23 or lower is indicative of cognitive impairment.

A

Test the null hypothesis that the underlying population representing all study participants is not cognitively impaired, assuming a known population standard deviation equal to 2.0. Interpret the results in 2-3 sentences for a non-statistical audience.

```
## [1] 23.72847
```

```
## [1] 9.167268e-125
```

```
##
##  One Sample t-test
##
## data:  CPAPAdherence_Data$mmse
## t = 26.668, df = 173, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 24
## 95 percent confidence interval:
##  27.37461      Inf
## sample estimates:
## mean of x
##   27.5977
```

The average MMSE score in this scenario is about 27.6 which is above the cutoff of 24 that indicates cognitive impairment.This means that it is unlikely to have occurred by chance. Participants in this study are not cognitively impaired on average.

B

Test the null hypothesis that the underlying population representing all study participants is not cognitively impaired, assuming the population standard deviation is unknown. Interpret the results in 2-3 sentences for a non-statistical audience.

```
##
##  One Sample t-test
##
## data:  CPAPAdherence_Data$mmse
## t = 26.668, df = 173, p-value < 2.2e-16
## alternative hypothesis: true mean is greater than 24
## 95 percent confidence interval:
##  27.37461      Inf
## sample estimates:
## mean of x
##   27.5977
```

The average MMSE score in this scenario is about 27.6 which is above the cutoff of 24 that indicates cognitive impairment.This means that it is unlikely to have occurred by chance. Participants in this study are not cognitively impaired on average.

5

This problem relies on the variable representing average daily CPAP use. A. Compute and report the mean and standard deviation for the variable representing average daily CPAP use. B. Assuming the estimates you computed represent population parameters, estimate and report the proportion of participants with average daily CPAP use less than 3 hours. C. In 2-3 sentences, explain the meaning of this probability in plain language suitable for a non-statistical audience.

```
## [1] 5.150766
```

```
## [1] 2.504029
```

```
## [1] 20.68966
```

About 20% of participants use their CPAP machines for less than 3 hours a night.

6

Compute and report the mean and standard deviation for BMI.Assuming the estimates you computed represent population parameters, estimate and report the proportion of participants who are considered obese (BMI 30+). In 2-3 sentences, explain the meaning of this probability in plain language suitable for a non-statistical audience.

```
## [1] 42.18415
```

```
## [1] 7.206295
```

```
## [1] 95.97701
```

Summary About 96% of participants have a body mass index of 30 or higher. This classifies 96% or 96 people out of every group of 100 people as obese.
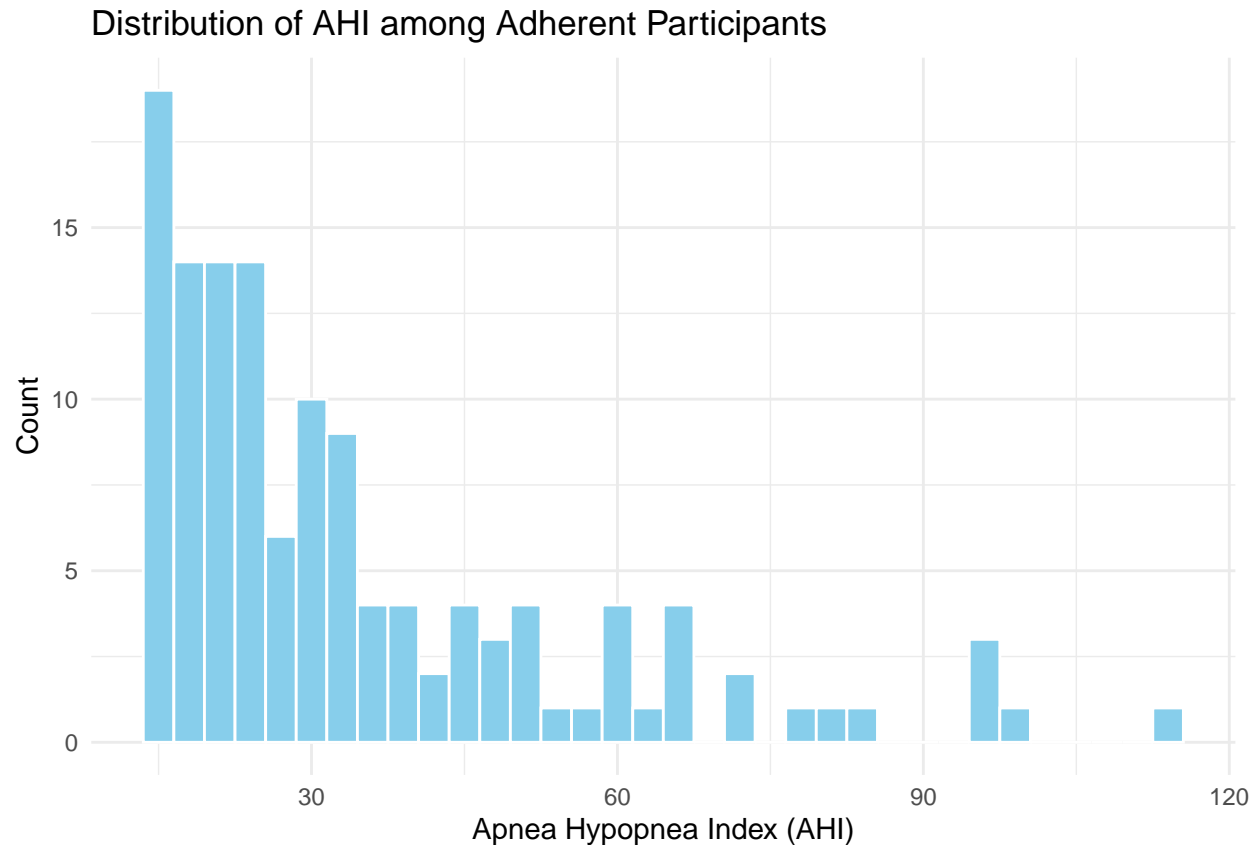
7

This problem relies on the Apnea Hypopnea Index (AHI) variable. Use the SamplingScript.R Download SamplingScript.Rfor drawing random samples.

A. Among those who are adherent to CPAP, describe the distribution of the variable AHI with appropriate statistics and data visualizations. Then, summarize your findings in 2-3 sentences in plain language suitable for a non-statistical audience.

```
##
##     adherent non-adherent
##          128           46
```

```
## # A tibble: 1 x 7
##   mean_AHI median_AHI sd_AHI min_AHI max_AHI Q1_AHI Q3_AHI
##      <dbl>      <dbl>  <dbl>   <dbl>   <dbl>  <dbl>  <dbl>
## 1     34.5       26.8   21.2      15    113.   19.5   43.6
```
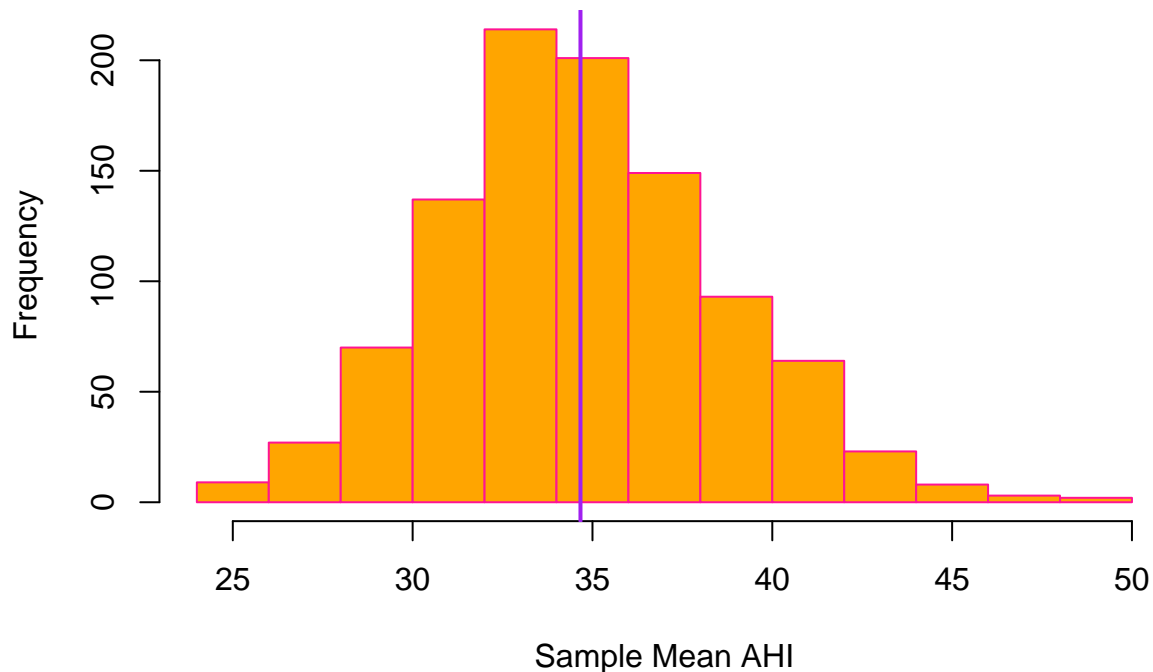
Distribution of AHI among Adherent Participants

Summary : Among people who use their CPAP machines regularly, most have moderate sleep apnea, but a few still have severe sleep apnea. This means that even with CPAP treatment, some individuals continue to experience serious sleep apnea episodes.

B

Among those who are adherent to CPAP, describe the sampling distribution based on samples of size 30 from the variable AHI based on theory with appropriate statistics and data visualizations. Then, summarize findings in 2-3 sentences in plain language suitable for a non-statistical audience.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   24.23   32.11   34.51   34.67   37.07   48.12
```

Sampling Distribution of Mean AHI (n = 30)
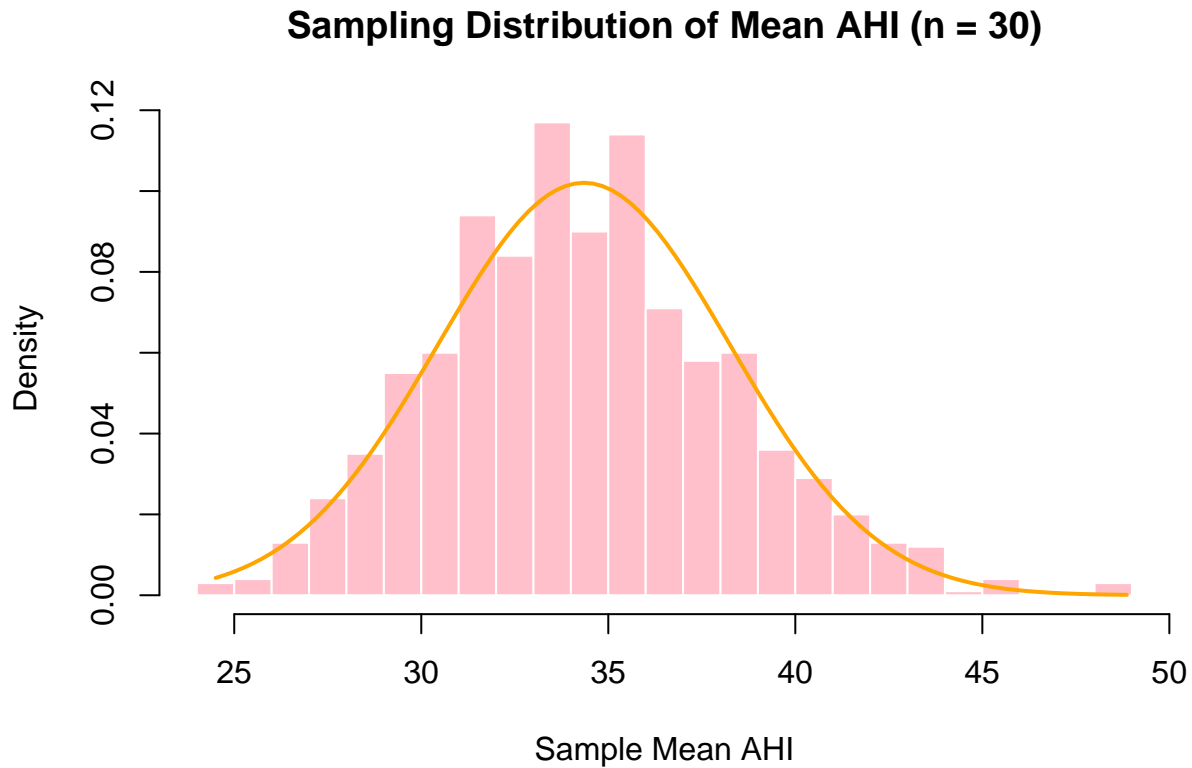


```
## [1] 3.898865
```

```
## [1] 3.870953
```

Summary: When we look at averages from radom groups of 30 people who use CPAP machines, most averages end up between 32 and 37 events per hour, with an overall average of about 34.5. This shows that the group as a whole is more stable than individual measurements. This gives a picture of the typical sleep apnea severity in people who use their CPAP machines as instructed

C. Randomly draw 1,000 samples of size 30 (with replacement) from the 128 AHI scores of adherent participants. For each sample, compute the sample mean AHI. Use these 1,000 sample means to:

  i. Calculate and report the mean and standard deviation of this sampling distribution.
  ii. Plot the sampling distribution of the sample means using a histogram.
  iii. Overlay a normal curve on your histogram of sample means.

```
## [1] 34.34837
```

```
## [1] 3.90928
```

Sampling Distribution of Mean AHI (n = 30)

iv. In 1-2 sentences, describe how your sampling distribution compares to a normal distribution. Summarize the shape and spread of the sampling distribution in plain language.
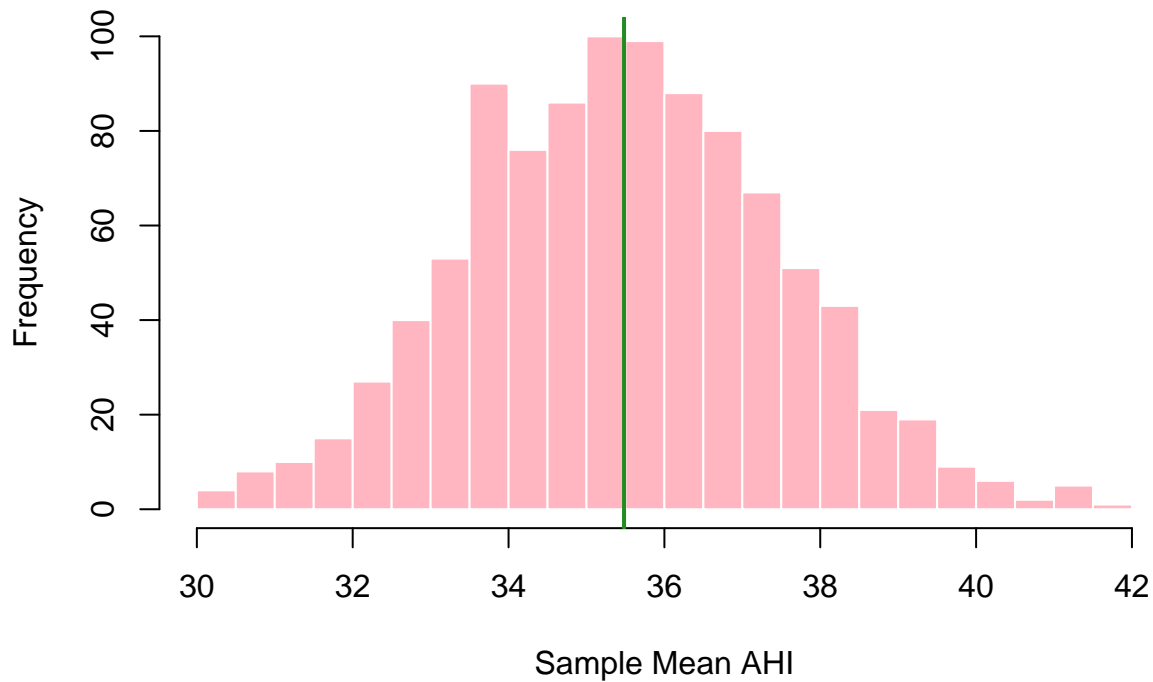
The sampling distribution is pretty normally distributed. It is fairly symmetrical on both sides. In plain language this means the the averages of groups containing 30 groups are consistent

v. In 2-3 sentences, describe how this distribution compares to your response from part (b) in plain language. The distribution in part c is very similar to what was gathered from part b. Both have very similar means and are roughly symmetrical.

D. Among those who are non-adherent to CPAP, describe the sampling distribution based on samples of size 100 from the variable AHI based on theory with appropriate statistics and data visualizations. Then summarize your findings in 2-3 sentences in plain language suitable for a non-statistical audience.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   30.01   34.03   35.45   35.48   36.88   41.63
```

Sampling Distribution of Mean AHI (n = 100)
Non–adherent Participants

```
## [1] 1.996526
```
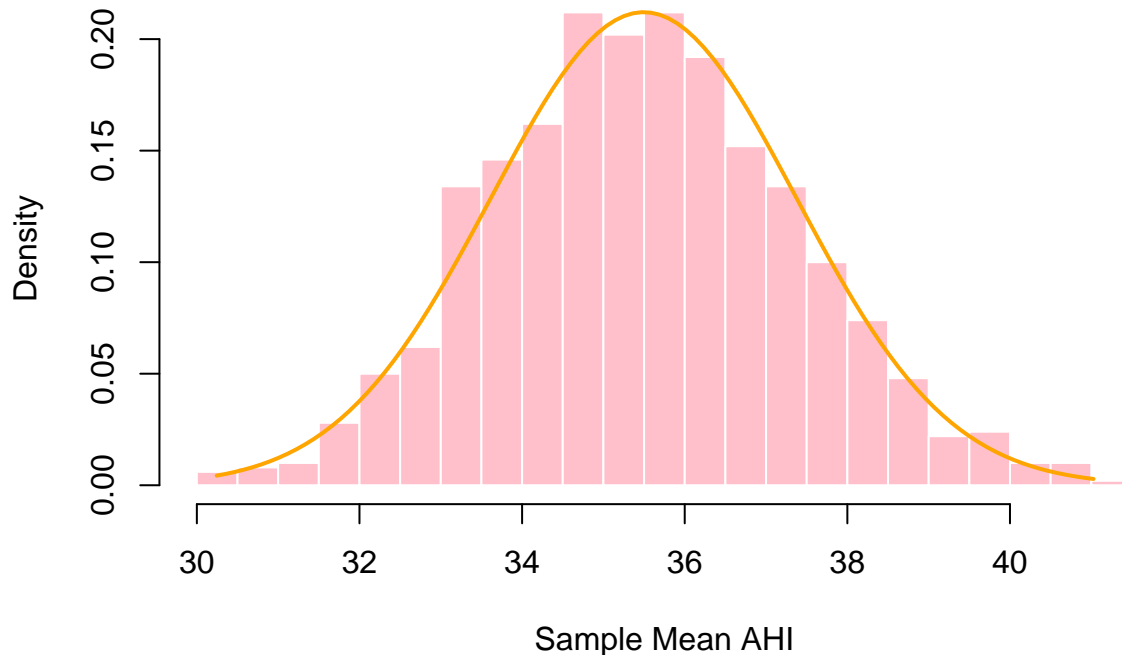
```
## [1] 1.990511
```

Summary : When random groups of 100 non-adherent CPAP users are analyzed, the average AHI for each group would usually be within about 2 points of the overall population average. The distribution of these group averages is roughly normal meaning extreme averages are not probable.

E. Randomly draw 1,000 samples of size 100 (with replacement) from the 46 AHI scores of non-adherent participants. For each sample, compute the sample mean AHI. Use these 1,000 sample means to: i. Calculate and report the mean and standard deviation of this sampling distribution. ii. Plot the sampling distribution of the sample means using a histogram. iii. Overlay a normal curve on your histogram of sample means.

```
## [1] 35.49504
```

```
## [1] 1.881024
```

Distribution of Mean AHI (Non–Adherent, n = 100)

iv. In 1-2 sentences, describe how your sampling distribution compares to a normal distribution. Summarize the shape and spread of the sampling distribution in plain language. The sampling distribution of the mean AHI is farily normal and symmetrical in shape.Individual AHI scores vary.

v. In 2-3 sentences, describe how this distribution compares to your response from part (d) in plain language. The sampling distribution of the mean AHI is much more consistent than the individual AHI scores. Individual scores for non-adherent participants ranged widely, but the sample means cluster tightly around 35.5, and had a much smaller standard deviation of 1.88. While individual scores are spread out widely the averages from large groups are consistent and can form patterns.

F Compare the sampling distributions among adherent versus non-adherent participants. How do the means and spreads of the two sampling distributions compare? Why do you think they are the same or different? Summarize your findings in 3-5 sentences in plain language suitable for a non-statistical audience.

The sampling distributions among adherent versus non-adherent participants are both roughly normal and centered around their means. The spread of the sample means is smaller for non-adherent participants than adherent participants because the non-adherent group had a larger sample size for each extrapulated mean. These differences demonstrate that averaging over larger samples produces more consistent and predictable estimates.