KEY

Part I. (20 points) A new shampoo is tested in a random sample of 81 people to see if it helps to recover hair loss. The recovery levels are determined on a scale of 1 to 10 ranging from smallest to largest recovery. The mean score is 7.5 and the standard deviation is 2.7 in the sample.

1. (12 points) In the advertisements, the company wants to state that ¾ of the hair-loss is recovered. This level, namely ¾ , corresponds to a score of 8. Does the evidence from the sample support the company's claim? Conduct a test of significance at $\alpha = 1\%$.

$H_0: \mu = 8$
$H_a: \mu < 8$

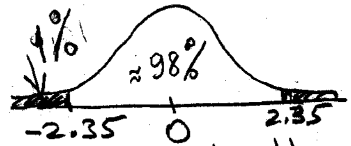
$$SE = \frac{SD}{\sqrt{n}} = \frac{2.7}{\sqrt{81}} = 0.3$$

$$z = \frac{7.5 - 8}{0.3} = -1.67$$

$$\text{P-value} = \frac{100\% - 90.11\%}{2} = 4.945\%$$

Since $4.945\% > 1\%$, we do not reject Ho.

(or z-critical = -2.35 since



and $-1.67 > -2.35 \Rightarrow$ do not reject Ho)

The evidence supports the company's claim.

2. (3 points) If a confidence interval was constructed for the mean recovery score, would this interval be for the sample or the population mean?

It would be for the population mean, which is to be estimated.

3. (5 points) If 50 different such samples are taken, and a 90% confidence interval is constructed from each sample, how many of those intervals do you expect to cover the true mean recovery score?

$50 \times 90\% = 45$ of them are expected to cover μ.

Part II. (20 points) A hybrid car is an environment-friendly car that uses both electricity and gasoline. Since it is new technology, only the following price data (in dollars) are available from a random sample of car dealers.

27000, 28400, 27500, 24800, 23300, 20200, 23400, 24600, 28700, 22200

1. (4 points) What is your estimate of the average price of hybrid cars in the market?

estimate of μ is $\bar{X} = \dfrac{27000 + 28400 + \cdots + 22200}{10}$

$= 25010$ dollars.

2. (8 points) Find a 90% confidence interval for the average price of hybrid cars in the market. (Hint: The standard deviation of the sample is 2829 dollars).

$n = 10$, small sample $\Rightarrow$ assume the price is approximately normal and hence use t-table.

$\bar{x} \mp t\,SE \Rightarrow 25010 \mp 1.83 \dfrac{2829}{\sqrt{10}}$

$d.f. = 9 \Rightarrow t_* = 1.83$ with 5%

$25010 \mp 1.83\,(894.6)$

$[23372.88, 26647.12]$

3. (8 points) The mean price for the ordinary cars (using only gasoline) is 22500 dollars. Test the claim that hybrid cars are more expensive than ordinary cars on the basis of the given data.

$H_0: \mu = 22500$

$H_a: \mu > 22500$

$t = \dfrac{25010 - 22500}{894.6}$

$= 2.8$

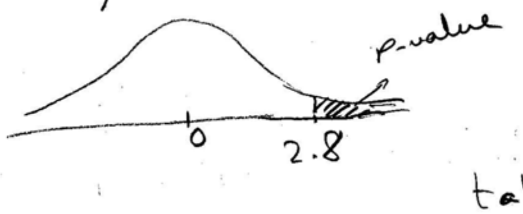


table $\Rightarrow$ p-value $\cong 1\%$

Since P-value $< 5\%$, reject H_0.

The mean price of hybrid cars is significantly higher than that of ordinary cars.

Part III. (20 points) A group of sociologists study the link between the level of masculinity and criminal behavior in men. They use a random sample of 19 men who were involved in violent events and another random sample of 35 men who avoided violent events. Each of the sampled men takes a Masculunity-Feminity Scale (MFS) test to determine his level of masculinity. MFS scores range from 0 to 56 points, with higher scores indicating a more masculine orientation.

Here is a MINITAB output for the MFS scores of the violent and non-violent groups:

Variable	N	Mean	SE Mean	StDev
Violent	19	41.6	1.95	8.48
Non-violent	35	34.0	1.64	9.68

1. (12 points) Perform a test of significance to decide if the violent group has a more masculine orientation than the non-violent group. Show all steps and state your conclusion in plain English.

$H_0: \mu_1 = \mu_2 \Rightarrow H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 > \mu_2 \qquad H_a: \mu_1 - \mu_2 > 0$

$SE_{2\text{-sample}} = \sqrt{(1.95)^2 + (1.64)^2} \cong 2.55$

$z = \frac{41.6 - 34.0 - 0}{2.55} = 2.98 \Rightarrow$ P-value $= \frac{100 - 99.73}{2}\% = 0.135\%$

Since $0.135\% < 5\%$, reject Ho.

The violent group has a more masculine orientation. The result is highly significant.

2. (4 points) Find a 95% CI for the mean MFS score of the men with violent criminal activity.

$n = 19$, small sample. Assuming normality, we use t-table. d.f. $= n-1 = 18 \Rightarrow t = 2.10$ with $\frac{\alpha}{2} = 2.5\%$

$\Rightarrow 41.6 \mp (2.10)\, 1.95$

$\Rightarrow [37.505, 45.695]$

3. (4 points) Find a 95% CI for the mean MFS score of the men who avoid violent events.

$n = 35$, large $\Rightarrow$ use z-table

$95\% \Rightarrow z = 2$

$\bar{x} \mp z\, SE \Rightarrow 34 \mp 2(1.64)$

$\Rightarrow [30.72, 37.28]$

Part IV. (20 points) A newspaper published a study of the "tip-of-the-tongue" phenomenon in elderly people. The researchers compared 40 people between 60-72 years of age with 50 people between 73-83 years of age. When the initial syllable of a missing word was given, the subject could either recall the whole word or not, in an experimental set up. Suppose 32 of the subjects could recall the word in the younger group, while 30 could recall it in the older group.

1. (16 points) Does the first (younger) group of elderly people have significantly higher recall rate than the second (older) group? Show all steps of a test of significance and state your conclusion in plain English.

$\hat{p}_1 = \frac{32}{40} = 80\%$ $\quad \hat{p}_2 = \frac{30}{50} = 60\%$

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 > 0$

$SE_1 = \sqrt{\frac{(0.80)(0.20)}{40}} = 6.3\%$

$SE_2 = \sqrt{\frac{(0.60)(0.40)}{50}} = 6.9\%$

$SE_{2\text{-sample}} = \sqrt{(0.063)^2 + (0.069)^2} = 0.093 = 9.3\%$

(OR $\sqrt{\frac{(0.80)(0.20)}{40} + \frac{(0.60)(0.40)}{50}} = 9.4\%$, difference is due to rounding

$z = \frac{80\% - 60\%}{9.3\%} = 2.15 \Rightarrow$ P-value $= \frac{100 - 96.84}{2}\% = 1.58\%$

Since P-value is less than 5%, reject Ho.

There is a significant difference between the two groups; the younger group has a significantly higher recall rat

2. (4 points) Give an estimate of the percentage of people who can remember well (in the context of the "tip-of-the-tong" experiment) among all elderly people in the population from 60 to 83 years of age.

$40 + 50 = 90$ in total.

$30 + 32 = 62$ recall

$\Rightarrow \hat{p} = \frac{62}{90} = 68.9\%$

Part V. (20 points) A four-year study of various brands of bottled water found that 25% of bottled water is just tap water packaged in a bottle. (brand=marka, tap=musluk)

Consider a sample of 10 bottles of water obtained recently from a specific brand name.

1. (10 points) What is the probability that at most 2 of the bottles has tap water? Assume that this brand is like any other brand in the market.

$p = 25\% = 0.25, \quad n = 10$

$$p(0) + p(1) + p(2) = \binom{10}{0}(0.25)^0(0.75)^{10} + \binom{10}{1}(0.25)^1(0.75)^9 + \binom{10}{2}(0.25)^2(0.75)^8$$

$$= 0.0563 + 0.188 + 0.282$$

$$= 0.526$$

2. (5 points) In fact, only 2 of the bottles contained tap water in the sample. What do you conclude about the considered brand: does it contain tap water by 25% as in general? Or more, or less?

Since the probability of observing such an event is not small by question 1 ($0.526 > 0.05$), we conclude that it contains tap water by 25%.

3. (5 points) Next, a random sample of 15 bottles from various brands is obtained to see if the percentage of tap water in the market has dropped down lately. It is found that 3 of them contain tap water. Fill in the blanks:

The null hypothesis is $H_0: p = 25\%$

The alternative hypothesis is $H_a: p < 25\%$

I cannot do a z-test because the sample size is small.

I cannot do a t-test because the number of bottles with tap water is binomially distributed, not normally distributed, although the sample size is small.

The estimate of the percentage of tap water in today's bottled water market is $3/15 = 20\%$.