R Notebook

Question 1

Using the "Western Ontario and McMaster University index (WOMAC)" questionnaire, 76 women were classified with severe hip pain. The WOMAC mean function score was 70.7 with standard deviation of 14.6, using the above values, we wish to know if we may conclude that the mean function score for a population of similar women subjects with sever hip pain is less than 75 . Let $\alpha = 0.01$

Answer

We construct the hypothesis accordingly:
H_0: Average of women with severe hip pain <75 H_a: Average of women with severe hip pain $\geq$ 75

Our required confidence level is 99% or 0.01

Our sample space >30 so we can leverage the Central Limit Theorem and assume sample mean approximates population standard mean.

We can calculate z-Score test statistic accordingly:

```r
x = 75
m = 70.7
n = 76
s = 14.6

z_stat <- (x - m) / (s/sqrt(n))
z_stat
```

```
## [1] 2.567571
```

This is a high z-score and does not lend credence to H_0. However, we will calculate p-value to validate whether this z-score could occur just by chance.

```r
pnorm(z_stat) + pnorm(z_stat, lower.tail = FALSE)
```

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

This high p value implies that there is a good probability this z_score occurs by chance. We can compare p-stat with α which is our confidence level of 0.01. The p-value is much greater than our alpha so we cannot reject H_0. However, we can assert that H_0 is not significant at a 99% confidence level.

Question 2

A bottle filling machine is set to fill bottles with soft drink to a volume of 500 ml. The actual volume is known to follow a normal distribution. The manufacturer believes the machine is under-filling bottles. A sample of 20 bottles is taken and the volume of liquid inside is measured. The volumes were:
536.08, 454.51, 471.38, 512.01, 494.48, 528.63, 493.64, 485.03, 473.88, 521.59, 502.85, 538.08, 465.68, 495.03, 475.32, 529.41, 518.13, 464.32, 449.08, 469.29

Calculate the sample mean and standard deviation Use a one-sample t-test to determine whether the bottles are being consistently under filled using a significance level of 0.02

Answer

Calculate sample mean and standard deviation:

```
volumes <- c(536.08, 454.51, 471.38, 512.01, 494.48, 528.63, 493.64, 485.03, 473.88, 521.59, 502.85,
             538.08, 465.68, 495.03, 475.32, 529.41, 518.13, 464.32, 449.08, 469.29)
mean(volumes)
```

```
## [1] 493.921
```

```
sd(volumes)
```

```
## [1] 28.21998
```

Our sample mean is 493.921 ml and standard deviation of the sample is 28.22.

One-sample t-test to determine under filling. Our hypothesis states that we believe there is under filling occurring. Therefore, it is stated as follows:

H_0: $\mu < 500$ ml H_a $\mu = 500$ ml

We can validate the hypothesis accordingly.

```
## calculate z score test statistic
## renamed variables for safety while knitting
xbar <- 493.21
mus <- 500
st <- 28.22
nn <- 20

z <- (xbar - mus) / (st/sqrt(nn))
z
```

```
## [1] -1.076038
```

Using critical value (region) approach we establish critical value:

```
alpha01 <- 0.02
qnorm(1-alpha01)
```

```
## [1] 2.053749
```

```
qnorm(alpha01)
```

```
## [1] -2.053749
```

Based on these results we would accept H_0 as the z-score (standard deviation) is less than the standard deviation cutoff point for an alpha of 0.02 on a normal distribution. We can further validate via p-value comparison:

```r
pnorm(z, lower.tail = FALSE)
```

```
## [1] 0.859045
```

```r
1 - pnorm(z)
```

```
## [1] 0.859045
```

Since the p-value is substantially higher than our alpha of 0.02, we can confirm that H_0 cannot be rejected. μ is 98% likely to be $< \mu_a$

Question 3

The mean sale price of the 49 Ebay auctions for a used android tablet was $43.98 with a standard deviation of 4.14. Amazon sells the same item for 46.99. Set up an appropriate (one-sided!) hypothesis test to check the claim that Ebay buyers pay less during auctions.
The skeptic would say the average is the same on Ebay, and we are interested in showing the average price is lower on Ebay.
Does this provide sufficient evidence to reject the null hypothesis? Use a significance level of $\alpha = 0.01$

Answer

We can test in the following way using the hypothesis:

H_0: $\mu < 46.99$ H_a $\mu = 46.99$

```r
## establish variables
xbar1 <- 43.98
mu2 <- 46.99
sd2 <- 4.14
n2 <- 49

z_stat2 <- (xbar1 - mu2) / (sd2/sqrt(n2))
z_stat2
```

```
## [1] -5.089372
```

```r
## normal distribution sd of alpha 0.01
alpha02 <- 0.01
qnorm(1-alpha02)
```

```
## [1] 2.326348
```

```r
qnorm(alpha02)
```

```
## [1] -2.326348
```

```r
pnorm(z_stat2)
```

```
## [1] 1.796256e-07
```

We can reject the null hypothesis based on evaluating the test statistic against standard deviations for normal distributino of 0.01 and verified by finding a p-value much lower than our alpha of 0.01 which all suggests that more significance may be attached to the alternative hypothesis that $46.99 is much closer to the true mean of costs on Ebay.

Question 4

Assuming that the data in mtcars follows the normal distribution, find the 95% confidence interval estimate of the difference between the mean gas mileage of manual and automatic transmissions. (Hint: two sample t-test)

Answer

We have had enough of manually calculating things at this stage :) , ergo…

```r
## collect data
auto <- subset(mtcars, am == 0, select = c(am, mpg))
manual <- subset(mtcars, am == 1, select = c(am, mpg))
```

```r
mpgtest <- t.test(manual$mpg, auto$mpg, mu = 0, var.equal = FALSE, conf.level = .95)
mpgtest
```

```
##
##  Welch Two Sample t-test
##
## data:  manual$mpg and auto$mpg
## t = 3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

So, within a 95% confidence level, manual cars obtain between 3.2 and 11.3 additional miles per gallon compared to automatic transmissions. The mean mpg for manual cars is 24.4 and for automatic cars 17.1. The alternate hypothesis is proved and the true difference in means is not equal to zero, i.e, they are not the same.
A drastic range but then again, we have not differentiated between engine size, horsepower and other variables.