L10 Group Assignment

Hypothesis Test - Height Study

In a study of women's heights, we obtain the following measurements:

```
heights = [1.66 1.69 1.50 1.80 1.68 1.64 1.65 1.70 1.72 1.67 1.69 ...
    1.68 1.20 1.90 1.2]
```

```
heights =

    1.6600    1.6900    1.5000    1.8000    1.6800    1.6400    1.6500    1.7000
```

In the population, the heights of woman is assumed to be Gaussian (normal) distributed with a mean of 1.68m and a standard deviation of 0.2m. Assume that you only test on the mean.

1. Estimate the mean of the population sample.

```
mu = 1.68; sigma = 0.2;

mu_hat = mean(heights)
```

```
mu_hat = 1.6253
```

2. Formulate the NULL hypothesis to test wether the sample has the sample mean as the same mean as the rest of the population.

```
mu_hat1 = mu1;
```

Sample mean is the same mean as the rest of the population.

3. Formulate the alternative hypothesis to the NULL hypothesis.

```
mu_hat1 ~= mu1;
```

4. Calculate the test statistics z.

```
n = length(heights);
z = (mu_hat - mu)/(sigma/sqrt(n))
```

```
z = -1.0586
```

5. Find the p-value based on a Gaussian pdf.

Since mu_hat is below mu $2 * \Pr(Z < z)$ is the p value.

```
p = 2*min(normcdf(z))
```

```
p = 0.2898
```

6. With a significance level of α = 0.05, can we reject the NULL hypothesis?

Since $p > \alpha$ the NULL hypothesis can not be rejected.

7. If we changes the significance level to α = 0.1, what would that imply?

There will still be $p > \alpha$ so it doesn't change anything.

8. Now repeat the experiment 100 times: Make a Matlab program, where 30 samples are drawn from a Gaussian distribution with a mean of 1.68m and a standard deviation of 0.2m

```
rep = 100; % number of repetitions
n = 30; % number of samples
p = zeros(1,rep);
for i = 1:rep
    heights = sigma*randn(1,n)+mu; % 30 samples created
    mu_hat = mean(heights);
    z = (mu_hat - mu)/(sigma/sqrt(n));
    p(i) = 2 * min(normcdf(z,0,1),1 - normcdf(z,0,1));
end
```

.

- With a significance level of $\alpha = 0.05$, how often do we falsely reject the NULL hypothesis?

Every time $p < \alpha$ the NULL hypothesis is falsely rejected.

```
alpha = 0.05;

fn = sum(p < 0.05)
```

```
fn = 3
```

- Now draw samples from a distribution with a mean of 1.78m and a standard deviation of 0.2m. How often do we falsely fail to reject the NULL hypothesis?

Every time $p > \alpha$ the NULL hypothesis is a false positives (fail to reject).

```
mu_new = 1.78;

for i = 1:rep
    heights = sigma*randn(1,n)+mu_new; % 30 samples created
    mu_hat = mean(heights);
    z = (mu_hat - mu)/(sigma/sqrt(n));
    p(i) = 2 * min(normcdf(z,0,1),1 - normcdf(z,0,1));
end

fp = sum(p > 0.05)
```

```
fp = 18
```

```
Prfp = fp/rep % Probability of false positive
```

```
Prfp = 0.1800
```