Hypothesis Tests about the Population Mean Using the Normal Distribution

Note: don't refer back to this sheet when doing problems in the future.

As we will see, the procedure we are about to use can't be applied in real problems!!

Next class, we will discuss our actual hypothesis test procedure that will be used from then on. We're doing this to develop understanding of the method and ideas.

Example 1

In the population of all babies born in December of 1998, the standard deviation of the gestation time was about 2.6 weeks. Suppose I take a carefully selected, representative random sample of 100 babies from this population and obtain a sample mean of 39.22 weeks. A density plot of my sample data shows its distribution to be unimodal, approximately symmetric, and have no outliers. Conduct a hypothesis test of the claim that the mean gestation time is 40 weeks, at the $\alpha = 0.05$ significance level.

Step 1: Identify population parameter of interest

- **Population parameter**:

Step 2: State hypotheses

- **Null Hypothesis** (H_0):

- **Alternative Hypothesis** (H_A):

Step 3: Identify the sample statistic and the sampling distribution of the sample statistic, assuming H_0 is true. Check all necessary conditions.

Check conditions:

- Were there any forms of bias in our sampling procedure?
- Was a quantitative variable recorded for each observational unit?
- Is the mean a reasonable summary of the center, and is the sample size large enough that the Central Limit Theorem applies?
- Are different observational units in our sample independent?

State the sampling distribution, assuming H_0 is true:

Step 4: Calculate the p-value for the test

The p-value is the probability of getting a test statistic at least as extreme as what we observed, assuming H_0 is true.

In this example, that means "the z-score is at least as big as what we got based on our sample".

```
pnorm(-3) + pnorm(3, lower.tail = FALSE)
```

```
## [1] 0.002699796
```

. . . or. . .

```
2 * pnorm(-3)
```

```
## [1] 0.002699796
```

. . . or. . .

```
2 * pnorm(3, lower.tail = FALSE)
```

```
## [1] 0.002699796
```

Step 5: Draw a conclusion

What is the strength of evidence against the null hypothesis provided by the data?

Make a decision using a significance level of $\alpha = 0.05$.

Example 2:

Suppose I take a random sample of 64 houses in South Hadley, MA, and calculate that the mean square footage of the houses in my sample is 1525 square feet. A histogram of the sample data shows the distribution of values in the sample to be unimodal and approximately symmetric, and there are no outliers. According to town records, the standard deviation of the square footage of all houses in the town is 100 square feet. Conduct a hypothesis test to see whether the average square footage of houses in South Hadley is different from the national average of 1480 square feet reported by the Census Bureau.

Step 1: Identify population parameter of interest

- **Population parameter**:

Step 2: State hypotheses

- **Null Hypothesis** (H_0):

- **Alternative Hypothesis** (H_A):

Step 3: Identify the sample statistic and the sampling distribution of the sample statistic, assuming H_0 is true. Check all necessary conditions.

Check conditions:

- Were there any forms of bias in our sampling procedure?
- Was a quantitative variable recorded for each observational unit?
- Is the mean a reasonable summary of the center, and is the sample size large enough that the Central Limit Theorem applies?
- Are different observational units in our sample independent?

State the sampling distribution, assuming H_0 is true:

Step 4: Calculate the p-value for the test

Use space in lab 12 on Gryd to do this calculation.

Step 5: Draw a conclusion, using a significance level of $\alpha = 0.05$.