ST 516: Foundations of Data Analytics

Mood's Test for Two Population Medians

Inference for Two Population Medians

Just like we did in the one-sample (one-population) case, we now consider inference for population *medians* instead of population *means*.

Now we are working in the setting where we have *independent* samples from two different populations of interest:

- $X_{11}, X_{12}, \ldots, X_{1n_1}$ are the observations from Population 1, which has (unknown) population median parameter M_1
- $X_{21}, X_{22}, \ldots, X_{2n_2}$ are the observations from Population 2, which has (unknown) population median parameter M_2

Note that we may have different sample sizes n_1 and n_2 in the two samples.

Our question of interest now is whether the population median parameters are equal: does $M_1 = M_2$?

Example

Suppose we are comparing real estate prices in two different cities: Seattle, WA vs. Orlando, FL.

- What would we learn by comparing *mean* prices, and when might that be the more relevant question?

Example

Suppose we are comparing real estate prices in two different cities: Seattle, WA vs. Orlando, FL.

- What would we learn by comparing *mean* prices, and when might that be the more relevant question?

  Comparing mean prices would tell us what we can *expect* to spend if we buy a randomly chosen piece of real estate. This might be the more relevant comparison if we are a developer and are going to buy multiple lots, and we are interested in whether our total cost would be higher or lower in Seattle or Orlando.

Example

Suppose we are comparing real estate prices in two different cities: Seattle, WA vs. Orlando, FL.

- What would we learn by comparing *medians*, and when might that be the more relevant question?

Example

Suppose we are comparing real estate prices in two different cities: Seattle, WA vs. Orlando, FL.

- What would we learn by comparing *medians*, and when might that be the more relevant question?

  Comparing median prices would tell us whether the 50th percentiles of real estate prices differ, and which city has the larger 50th percentile. It is important to note that this *is not* the same as telling us whether *most* lots of real estate are cheaper in one of the cities: it is possible that Seattle has the lower median, but there are more cheap lots in Orlando.

Example

We decide that we are interested in comparing the median price of real estate in these two cities because we want to answer "Is the price for which 50% of Seattle real estate lots are cheaper equal to the price for which 50% of Orlando real estate lots are cheaper?".

Therefore, we cannot use the t-test, as the t-test answers questions about population means. We want to answer whether $M_1 = M_2$ (that is, are the population medians equal), so we perform a hypothesis test of the following hypotheses:

- Null Hypothesis: $H_0 : M_1 = M_2$
- Alternative Hypothesis: $H_A : M_1 \neq M_2$

Mood's Test

We use the following steps to perform this test, which is called **Mood's Test**:

1. Find the median of the *combined* samples.
2. Calculate the proportion of observations from each sample *separately* that are less than or equal to the combined median.
3. Use the two-sample test of equality of binomial proportions to test whether the population proportions of observations less than the combined median are equal in the two populations.

These steps are explained in more detail on the next slide.

Mood's Test, More Detail

1. Find the median of the *combined* samples.

   That is, put the samples $X_{11}, X_{12}, \ldots, X_{1n_1}$ and $X_{21}, X_{22}, \ldots, X_{2n2}$ together as a single group, and find the middle value, m, of this combined group.

2. Find be the proportion of the first sample that are smaller than m:

$$\hat{\pi}_1 = \frac{\text{\# Observations in Sample 1 } \leq m}{n_1}$$

   Similarly, find the proportion of the second sample that are smaller than m:

$$\hat{\pi}_2 = \frac{\text{\# Observations in Sample 2 } \leq m}{n_2}$$

3. Use the two-sample binomial proportions test to test that $\pi_1 = \pi_2$.

Mood's Test Example

Suppose we obtain a random sample of real estate listings from Seattle, and another random sample of real estate listings from Orlando. Hypothetical data is displayed below, in units of $1000:

Seattle	133	145	222	232	290	344
$(n_1 = 12)$	356	382	464	532	783	1349
Orlando	240	278	383	408	418	431
$(n_2 = 10)$	472	484	509	583		

The median real estate prices in these two samples are

Seattle	Median = $350 Thousand
Orlando	Median = $424.5 Thousand

Mood's Test Example

We use Mood's Test to determine whether we would reject the null hypothesis that the two cities have the same median real estate prices:

- First, we find the median of the combined sample, shown below in sorted order (smallest to largest):

Seattle	133	145	222	232	240	278
and	290	344	356	382	$\boxed{383}$	$\boxed{408}$
Orlando	418	431	464	472	484	509
(combined)	532	583	783	1349		

The median (middle value) is half way between the 11th and 12th values (since there are $n_1 + n_2 = 22$ total observations). Therefore, the combined median is

$$m = \frac{383 + 408}{2} = 395.5$$

Mood's Test Example

- Next, we count up how many of the observations from *each sample separately* are less than or equal to the combined median $m = 395.5$:

	$\# \leq m$	$\# > m$
Seattle	8	4
Orlando	3	7

Mood's Test Example

- Then we compute the sample proportions of observations less than or equal to m in each sample:

	$\# \leq m$	$\# > m$	Total	Sample Proportion
Seattle	8	4	12	$\hat{\pi}_1 = \frac{8}{12} = 0.67$
Orlando	3	7	10	$\hat{\pi}_2 = \frac{3}{10} = 0.30$

Mood's Test Example

- Finally, we perform a two-sample test of binomial proportions to test that $\pi_1 = \pi_2$.

```r
x <- c(8, 3)
n <- c(12, 10)
prop.test(x, n, correct=F)
```

```
##    2-sample test for equality of proportions without
##    continuity correction
##
## data:  x out of n
## X-squared = 2.9333, df = 1, p-value = 0.08677
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.02295977  0.75629311
## sample estimates:
##    prop 1    prop 2
## 0.6666667 0.3000000
```