Other two sample comparisons

ST551 Lecture 25

Charlotte Wickham

2017-11-20

Charlotte to ask about Weds lab

So far

Our two sample comparisons have focused on means (or proportions)

What else could we compare?

- Medians
- Variances
- Whole distributions

Comparing medians: Mood's median test

Mood's median test

Setting: two indpendent samples

Y_i i.i.d sample of size n from popuation with c.d.f F_Y
X_i i.i.d sample of size m from popuation with c.d.f F_X

$m_Y = F_Y^{-1}(0.5) =$ median of population that Y is sampled from.
$m_X = F_X^{-1}(0.5) =$ median of population that X is sampled from.

Comparison of interest: Is m_Y the same as m_X?

Example

A study is performed to assess the effect of fish oil supplements on diastolic blood pressure

- 25 subjects are randomly assigned to receive fish oil ($n_Y = 12$) or regular vegetable oil ($n_X = 13$) for two weeks.
- Each subject's decrease in diastolic blood pressure over those two weeks is recorded (bigger numbers $=>$ better reduction in blood pressure)

Fish oil: *-2.2*, *-0.8*, *3.7*, *4.9*, *5*, *5.2*, *5.3*, *6*, *8*, *8*, *10.4* and *14*

Regular oil: *-6.4*, *-6.4*, *-5.9*, *-5.8*, *-5.3*, *-4.9*, *-4.4*, *0.2*, *2.1*, *2.5*, *2.5*, *6.1* and *8.9*

Question: Is the median blood pressure reduction the same for these two treatments?

Your turn

- If the null is true, $m_Y = m_X = m$, what is our best guess for the median m?
- If the null is true, what proportion of the sample from Y should be larger than m?
- If the null is true, what proportion of the sample from X should be larger than m?

Estimating the combined median

$$\hat{m}_Y = \hat{m}_X = \hat{m} = \text{median}(Y_1, Y_2, \ldots, Y_n, X_1, X_2, \ldots, X_m)$$

If the null is true, this estimate is an unbiased and consistent estimate of the common median, m.

We expect $P(Y_i > m) = P(X_i > m)$.

Mood's median test

Procedure:

1. Find the combined median $\hat{m}$.
2. Test the true proportion of Y's greater than $\hat{m}$ is equal to the true proprtion of X's greater than $\hat{m}$.
   - Z-test for proportions/Chi-square test or Fishers exact test

Example cont.

Combined sample:

```
##  [1] -6.4 -6.4 -5.9 -5.8 -5.3 -4.9 -4.4 -2.2 -0.8
## [10]  0.2  2.1  2.5  2.5  3.7  4.9  5.0  5.2  5.3
## [19]  6.0  6.1  8.0  8.0  8.9 10.4 14.0
```

Combined median, $\hat{m} = 2.5$

	Number $> \hat{m}$	Number $\leq \hat{m}$
Fish Oil	10	2
Regular Oil	2	11

Example cont.

$$
\begin{aligned}
Z &= \frac{\hat{p}_Y - \hat{p}_X}{\sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n}+\frac{1}{m}\right)}} \\
&= \frac{\frac{10}{12} - \frac{2}{13}}{\sqrt{\frac{12}{25}(1-\frac{12}{25})\left(\frac{1}{12}+\frac{1}{13}\right)}} \\
&= 3.4
\end{aligned}
$$

p-value $= 6.8 \times 10^{-4}$.

There is convincing evidence that the median BP reduction on fish oil is different to the median BP reduction on regular oil.

Wilcoxon Rank Sum test

Wilcoxon Rank Sum

Wilcoxon Rank Sum, a.k.a Mann-Whitney U-test

Often presented as a test for equality of medians, like Wilcoxon Signed Rank, **this isn't true without further assumptions**.

Wilcoxon Rank Sum Procedure

1. Combine the samples
2. Rank the observations in the combined sample from smallest (1) to largest ($n + m$). If there are ties, assign the average rank to the tied observations.
3. **Test statistic:** Sum of the ranks in the sample with the smaller sample size
4. **p-value:** either use Normal approximation, or via permutation

Intutition: if all the observations come from the same distribution, it would be unlikely for all the observations in the samller sample to have all the highest ranks (or lowest).

Example

Combined sample:

```
## Regular Oil Regular Oil Regular Oil Regular Oil
##        -6.4        -6.4        -5.9        -5.8
## Regular Oil Regular Oil Regular Oil    Fish Oil
##        -5.3        -4.9        -4.4        -2.2
##    Fish Oil Regular Oil Regular Oil Regular Oil
##        -0.8         0.2         2.1         2.5
## Regular Oil    Fish Oil    Fish Oil    Fish Oil
##         2.5         3.7         4.9         5.0
##    Fish Oil    Fish Oil    Fish Oil Regular Oil
##         5.2         5.3         6.0         6.1
##    Fish Oil    Fish Oil Regular Oil    Fish Oil
##         8.0         8.0         8.9        10.4
##    Fish Oil
##        14.0
```

```
## [1] 208
```

Problems

- Location-shift assumption

- Not location shift