

Part 1: Foundational Exercises

Lab 1: Hypothesis Testing - W203 Section 8

Team Anglerfish: Aruna Bisht, Jing Lange, Chase Madson, Maria Manna

Contents

1.1 Professional Magic	2
1.1.1 Type I Error Rate	2
1.1.2 Power	3
1.2 Wrong Test, Right Data	4
1.2.1 Violating t-test Assumptions	4
1.2.2 Remedial Measures	4
1.3 Test Assumptions	4
1.3.1 World Happiness	4
1.3.2 Legislators	4
1.3.3 Wine and Health	5
1.3.4 Attitudes Toward the Religious	5

1.1 Professional Magic

1.1.1 Type I Error Rate

Question: What is the type I error rate of the test?

- Let's denote the test statistic t with the given formula $t = X_1 + Y_1 + X_2 + Y_2 + X_3 + Y_3$.
- The type I error rate α is the probability that we incorrectly reject the null hypothesis $H_0 : p = \frac{1}{2}$. In our test, we reject the null hypothesis when the test statistic is either zero or six.

$$\begin{aligned}\alpha &= Pr(H_0 \text{ is rejected when it is true}) \\ &= Pr(t \in \{0, 6\} | p = \frac{1}{2})\end{aligned}$$

- We can rearrange $t = 0$ as 3 consecutive flips of $(X_i, Y_i) = (0, 0)$, and similarly $t = 1$ as $(X_i, Y_i) = (1, 1)$. We note that there is $\binom{6}{0} = 1$ way of observing $t = 0$ and $\binom{6}{6} = 1$ way of observing $t = 1$.

$$\begin{aligned}&= \binom{6}{0} Pr([(X_1, Y_1) = (0, 0)] \cap [(X_2, Y_2) = (0, 0)] \cap [(X_3, Y_3) = (0, 0)] | p = \frac{1}{2}) \\ &+ \binom{6}{6} Pr([(X_1, Y_1) = (1, 1)] \cap [(X_2, Y_2) = (1, 1)] \cap [(X_3, Y_3) = (1, 1)] | p = \frac{1}{2})\end{aligned}$$

- Given that each flip of the pair is independent of all other flips of the pair, we can restate as the product of probabilities.

$$\begin{aligned}&= Pr([(X_1, Y_1) = (0, 0)] | p = \frac{1}{2}) \times Pr([(X_2, Y_2) = (0, 0)] | p = \frac{1}{2}) \times Pr([(X_3, Y_3) = (0, 0)] | p = \frac{1}{2}) \\ &+ Pr([(X_1, Y_1) = (1, 1)] | p = \frac{1}{2}) \times Pr([(X_2, Y_2) = (1, 1)] | p = \frac{1}{2}) \times Pr([(X_3, Y_3) = (1, 1)] | p = \frac{1}{2})\end{aligned}$$

- We express this in terms of the joint distribution functions we are given.

$$= (f_{X,Y|p=\frac{1}{2}}(0, 0))^3 + (f_{X,Y|p=\frac{1}{2}}(1, 1))^3$$

- When $p = \frac{1}{2}$, then $f_{X_i, Y_i}(0, 0) = f_{X_i, Y_i}(1, 1) = \frac{1}{2} = \frac{1}{4}$, and thus we plug that probability into the above.

$$\begin{aligned}&= \left(\frac{1}{4}\right)^3 + \left(\frac{1}{4}\right)^3 \\ &= \frac{1}{32} = 0.03125\end{aligned}$$

1.1.2 Power

Question: What is the power of your test for the alternative hypothesis that $p = \frac{3}{4}$

- a. The power of this test is $1 - \beta$, where β is the type II error rate (i.e., the probability that we fail to reject H_0 when it is false). Specifically, we want to calculate the power under the alternative hypothesis $H_a : p = \frac{3}{4}$.

$$\begin{aligned}
 \text{Power} &= 1 - \beta \\
 &= 1 - \Pr(\text{Fail to reject } H_0 \text{ when } H_a \text{ is true instead}) \\
 &= 1 - \Pr(t \notin \{0, 6\} \mid H_a : p = \frac{3}{4}) \\
 &= 1 - [1 - \Pr(t \in \{0, 6\} \mid H_a : p = \frac{3}{4})] \\
 &= \Pr(t \in \{0, 6\} \mid H_a : p = \frac{3}{4})
 \end{aligned}$$

- b. We repeat our steps c, d, and e from our solution to 1.1.1 above to reach the joint probability density function when $p = \frac{3}{4}$.

$$= (f_{X,Y|p=\frac{3}{4}}(0,0))^3 + (f_{X,Y|p=\frac{3}{4}}(1,1))^3$$

- c. When $p = \frac{3}{4}$, then $f_{X_i,Y_i}(0,0) = f_{X_i,Y_i}(1,1) = \frac{\frac{3}{4}}{2} = \frac{3}{8}$, and thus we plug that probability into the above.

$$\begin{aligned}
 &= \left(\frac{3}{8}\right)^3 + \left(\frac{3}{8}\right)^3 \\
 &= \frac{27}{256} = \mathbf{0.10546875}
 \end{aligned}$$

1.2 Wrong Test, Right Data

Imagine that your organization surveys a set of customers to see how much they like your regular website, and how much they like your mobile website. Suppose that both of these preference statements are measured on 5-point Likert scales.

1.2.1 Violating t-test Assumptions

Question: If you were to run a paired t-test using this data, what consequences would the violation of the metric scale assumption have for your interpretation of the test results?

To begin, the first assumption for a paired t-test is that the data must be metric. As the performance statements in this survey are measured using Likert-scales, we know that this assumption is violated. Data from Likert-scales have non-standard intervals, are not continuous, and the magnitudes for the values can be different for each value in the sample (a “Very good” from one person can mean something different than another person’s “very good”). Therefore, without even looking at the other assumptions, we already know that the paired t-test is inappropriate for this data.

Ordinal data can only be used to do order statistics, not to calculate values such as μ , σ , ε , or a t-statistic. The consequence of running the suggested paired t-test using ordinal data is that all findings and interpretations would be invalid.

The other main assumptions for a paired t-test are as follows:

- There is a natural dependency between observations
- I.I.D. data across samples ((X_i, Y_i) is drawn from the same distribution, independently of all other pairs)
- Sample size of at least 30, $n \geq 30$, and not too much skew so that CLT can be applied

1.2.2 Remedial Measures

Question: What would you propose to do to remedy this problem?

The data contains paired values but is not metric, which would encourage a Sign Test to be used to evaluate this data. If the test could not be changed, the data could be converted into a 0/1 binary data set. Binary data is metric and thus a paired t-test would then be appropriate.

1.3 Test Assumptions

1.3.1 World Happiness

2 + 2

[1] 4

1.3.2 Legislators

```
2 + 2
```

```
## [1] 4
```

1.3.3 Wine and Health

```
2 + 2
```

```
## [1] 4
```

1.3.4 Attitudes Toward the Religious

```
2 + 2
```

```
## [1] 4
```