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

1.1.1 Type I Error Rate

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

a. Let's denote the test statistic t with the given formula $t = X_1 + Y_1 + X_2 + Y_2 + X_3 + Y_3$.

b. 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.

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

c. 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 $t = 1$.

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

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

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

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

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

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

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

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 $H_a : p = \frac{3}{4}$.

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

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.

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

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?*

For the results of a t-test on some sample to be reliable, then each draw from the sample must come from an underlying random variable that is **metric** in scale, meaning they are continuous and numeric. As the preference statements in this survey are each measured on a 5-point Likert-scale, we know that this assumption is violated.

Running a paired t-test on Likert-scale data means calculating the test statistic $t = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$, which gives us the difference from the mean in units of standard deviation. These units lose their meaning when calculated using Likert-scale data; it doesn't make sense to subtract or take a mean of these values. This is because data from Likert-scale have **non-standard intervals**, are **categorical instead of continuous**, and have **incomparable magnitudes** as values are subjective to each customer in the sample (a "very good" from one person can mean something different than another person's "very good"). Therefore, our interpretation of the test statistic would be invalid and so would be any conclusions we draw from it.

1.2.2 Remedial Measure

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

Since Likert-scale data contains paired values that are non-metric, it is better to use a **hypothesis of comparisons under the Wilcoxon Rank-Sum Test**. If we want to compare Likert-scale responses about customers' sentiment about our regular website (X) and our mobile website (Y), we would want to test the null hypothesis that $H_0 : P(X < Y) = P(X > Y)$, meaning equal likelihood. If the evidence suggests that we can reject this null hypothesis, then we can provide a useful conclusion to our organization about how our customers view these websites differently.

The assumptions required under this test are that the data must follow an ordinal scale (which Likert-scale does) and that each pair is drawn from the same distribution (which the organization can accomplish through random sampling).

1.3 Test Assumptions

1.3.1 World Happiness

Assumptions: In order to run a paired t-test on some sample $\{X_i\}_{i=1}^n$ and receive reliable results, there are 3 assumptions that must be met.

1. Each X_i comes from an underlying random variable that is **metric** in scale, meaning they are continuous and numeric.
2. $\{X_i\}_{i=1}^n$ is an **independent and identically distributed** random sample.
3. $\{X_i\}_{i=1}^n$ is **normally distributed**, or $\{X_i\}_{i=1}^n \sim N(\mu_x, \sigma_x^2$
   - With sufficiently large n (e.g., surveying at least 30 customers), this assumption is met asymptotically since the Central Limit Theorem will cause a normal distribution of the mean.

Violations:

```
2 + 2
```

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

1.3.2 Legislators

```r
2 + 2
```

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

1.3.3 Wine and Health

```r
2 + 2
```

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

1.3.4 Attitudes Toward the Religious

```r
2 + 2
```

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