

# STA 6384, Report 1.10

Carson Slater     *Baylor University*

**Problem: Work problem 1.12, p. 30 of Agresti.**

**1.12** A researcher routinely tests using a nominal  $P(\text{type I error}) = 0.05$ , rejecting  $H_0$  if the  $P$ -value  $\leq 0.05$ . An exact test using test statistic  $T$  has null distribution  $P(T = 0) = 0.30$ ,  $P(T = 1) = 0.62$ , and  $P(T = 2) = 0.08$ , where a higher  $T$  provides more evidence against the null.

**a.** With the usual  $p$ -value, show that the actual  $P(\text{type I error}) = 0$ .

To solve part (a), we must determine the actual probability of committing a Type I error. A Type I error occurs when we reject the null hypothesis,  $H_0$ , when it is in fact true. The given decision rule is to reject  $H_0$  if the calculated  $p$ -value is less than or equal to the nominal significance level,  $\alpha = 0.05$ .

The test statistic  $T$  is discrete and has the following probability distribution under the null hypothesis:

- $P(T = 0) = 0.30$
- $P(T = 1) = 0.62$
- $P(T = 2) = 0.08$

The problem states that a higher value of  $T$  provides more evidence against  $H_0$ , indicating a right-tailed test. For a right-tailed test with a discrete test statistic, the usual  $p$ -value corresponding to an observed value  $t_{obs}$  is defined as  $P(T \geq t_{obs})$ .

We now calculate the  $p$ -value for each possible outcome of the test statistic  $T$ :

**1. If the observed outcome is  $T = 2$ :** The  $p$ -value is calculated as:

$$P(T \geq 2) = P(T = 2) = 0.08$$

**2. If the observed outcome is  $T = 1$ :** The  $p$ -value is calculated as:

$$P(T \geq 1) = P(T = 1) + P(T = 2) = 0.62 + 0.08 = 0.70$$

**3. If the observed outcome is  $T = 0$ :** The  $p$ -value is calculated as:

$$P(T \geq 0) = P(T = 0) + P(T = 1) + P(T = 2) = 0.30 + 0.62 + 0.08 = 1.00$$

Next, we apply the researcher's decision rule ( $p$ -value  $\leq 0.05$ ) to each possible outcome:

- For  $T = 2$ , the  $p$ -value is 0.08. Since  $0.08 > 0.05$ , we fail to reject  $H_0$ .

- For  $T = 1$ , the  $p$ -value is 0.70. Since  $0.70 > 0.05$ , we fail to reject  $H_0$ .
- For  $T = 0$ , the  $p$ -value is 1.00. Since  $1.00 > 0.05$ , we fail to reject  $H_0$ .

In every possible case, the calculated  $p$ -value is greater than 0.05. Therefore, the null hypothesis is never rejected. The actual probability of a Type I error is the sum of the probabilities of the outcomes in the rejection region. Since the rejection region is the empty set, its probability is 0.

Thus, the actual  $P(\text{type I error}) = 0$ .

**b.** With the mid  $p$ -value, show that the actual  $P(\text{type I error}) = 0.08$ .

In this part, we are asked to show that using the mid- $p$ -value, the actual probability of a Type I error is 0.08. The mid- $p$ -value for a discrete test statistic in a right-tailed test is defined as:

$$\text{mid-}p\text{-value} = \frac{1}{2}P(T = t_{\text{obs}}) + P(T > t_{\text{obs}})$$

where  $t_{\text{obs}}$  is the observed value of the test statistic  $T$ .

We use the same null distribution as in part (a):

- $P(T = 0) = 0.30$
- $P(T = 1) = 0.62$
- $P(T = 2) = 0.08$

The researcher rejects the null hypothesis  $H_0$  if the  $p$ -value is less than or equal to 0.05. We calculate the mid- $p$ -value for each possible outcome of  $T$ :

**1. If the observed outcome is  $T = 2$ :** The mid- $p$ -value is:

$$\frac{1}{2}P(T = 2) + P(T > 2) = \frac{1}{2}(0.08) + 0 = 0.04$$

**2. If the observed outcome is  $T = 1$ :** The mid- $p$ -value is:

$$\frac{1}{2}P(T = 1) + P(T > 1) = \frac{1}{2}(0.62) + P(T = 2) = 0.31 + 0.08 = 0.39$$

**3. If the observed outcome is  $T = 0$ :** The mid- $p$ -value is:

$$\frac{1}{2}P(T = 0) + P(T > 0) = \frac{1}{2}(0.30) + (P(T = 1) + P(T = 2)) = 0.15 + (0.62 + 0.08) = 0.15 + 0.70 = 0.85$$

Now, we apply the decision rule (reject if  $p$ -value  $\leq 0.05$ ) to these mid- $p$ -values:

- For  $T = 2$ , the mid- $p$ -value is 0.04. Since  $0.04 \leq 0.05$ , we **reject**  $H_0$ .

- For  $T = 1$ , the mid- $p$ -value is 0.39. Since  $0.39 > 0.05$ , we fail to reject  $H_0$ .
- For  $T = 0$ , the mid- $p$ -value is 0.85. Since  $0.85 > 0.05$ , we fail to reject  $H_0$ .

The actual probability of a Type I error is the probability of the set of outcomes for which we reject  $H_0$ . In this case, we only reject  $H_0$  when the outcome is  $T = 2$ . Therefore, the actual probability of a Type I error is the probability of this outcome occurring under the null hypothesis.

$$P(\text{type I error}) = P(\text{Reject } H_0) = P(T = 2) = 0.08$$

This confirms that with the mid- $p$ -value, the actual  $P(\text{type I error}) = 0.08$ .

c. Find  $P(\text{type I error})$  in parts (a) and (b) when  $P(T = 0) = 0.30$ ,  $P(T = 1) = 0.66$ ,  $P(T = 2) = 0.04$ . Note that the test with mid  $p$ -value can be conservative or liberal. The exact test with ordinary  $p$ -value cannot be liberal.

In this part, we are asked to find the probability of a Type I error for the methods in parts (a) and (b) using a new null distribution for the test statistic  $T$ :

- $P(T = 0) = 0.30$
- $P(T = 1) = 0.66$
- $P(T = 2) = 0.04$

The nominal significance level remains  $\alpha = 0.05$ .

*Part (a) Revisited: Usual  $p$ -value*

We first calculate the usual  $p$ -value,  $P(T \geq t_{\text{obs}})$ , for each possible outcome.

1. **If  $T = 2$ :** The  $p$ -value is  $P(T \geq 2) = P(T = 2) = 0.04$ .
2. **If  $T = 1$ :** The  $p$ -value is  $P(T \geq 1) = P(T = 1) + P(T = 2) = 0.66 + 0.04 = 0.70$ .
3. **If  $T = 0$ :** The  $p$ -value is  $P(T \geq 0) = P(T = 0) + P(T = 1) + P(T = 2) = 0.30 + 0.66 + 0.04 = 1.00$ .

We reject  $H_0$  if the  $p$ -value  $\leq 0.05$ .

- For  $T = 2$ :  $0.04 \leq 0.05$ , so we **reject**  $H_0$ .
- For  $T = 1$ :  $0.70 > 0.05$ , so we fail to reject  $H_0$ .
- For  $T = 0$ :  $1.00 > 0.05$ , so we fail to reject  $H_0$ .

The rejection region consists only of the outcome  $T = 2$ . The probability of a Type I error is the probability of this outcome.

$$P(\text{type I error}) = P(T = 2) = 0.04$$

With the new distribution, the test using the usual  $p$ -value has an actual Type I error rate of 0.04. This is less than the nominal  $\alpha = 0.05$ , so the test is conservative.

*Part (b) Revisited: Mid- $p$ -value*

Next, we calculate the mid- $p$ -value,  $\frac{1}{2}P(T = t_{\text{obs}}) + P(T > t_{\text{obs}})$ , for each outcome.

1. **If  $T = 2$ :** The mid- $p$ -value is  $\frac{1}{2}P(T = 2) + P(T > 2) = \frac{1}{2}(0.04) + 0 = 0.02$ .
2. **If  $T = 1$ :** The mid- $p$ -value is  $\frac{1}{2}P(T = 1) + P(T > 1) = \frac{1}{2}(0.66) + P(T = 2) = 0.33 + 0.04 = 0.37$ .
3. **If  $T = 0$ :** The mid- $p$ -value is  $\frac{1}{2}P(T = 0) + P(T > 0) = \frac{1}{2}(0.30) + (P(T = 1) + P(T = 2)) = 0.15 + (0.66 + 0.04) = 0.85$ .

We reject  $H_0$  if the mid- $p$ -value  $\leq 0.05$ .

- For  $T = 2$ :  $0.02 \leq 0.05$ , so we **reject**  $H_0$ .
- For  $T = 1$ :  $0.37 > 0.05$ , so we fail to reject  $H_0$ .
- For  $T = 0$ :  $0.85 > 0.05$ , so we fail to reject  $H_0$ .

The rejection region is again just the outcome  $T = 2$ . The probability of a Type I error is:

$$P(\text{type I error}) = P(T = 2) = 0.04$$

In this case, the mid- $p$ -value test is also conservative, as  $0.04 < 0.05$ . This, in combination with the result from part (b) where the error rate was 0.08 (liberal), demonstrates the note that the test with the mid- $p$ -value can be either conservative or liberal. The exact test with the ordinary  $p$ -value cannot be liberal because the rejection rule requires the tail probability  $P(T \geq t_{\text{obs}})$  to be  $\leq \alpha$ . The actual error rate is the sum of  $P(T = t)$  for outcomes  $t$  in the rejection region, which by definition cannot exceed  $\alpha$ .

**d.** In part (a), a randomized-decision test generates a uniform random variable  $U$  from  $[0, 1]$  and rejects  $H_0$  if both  $T = 2$  and  $U \leq \frac{5}{8}$ . Show the actual  $P(\text{type I error}) = 0.05$ . Is this a sensible test?

This part of the problem introduces a randomized-decision test. We revert to the original null distribution from part (a):

- $P(T = 0) = 0.30$
- $P(T = 1) = 0.62$
- $P(T = 2) = 0.08$

The new rejection rule is: Reject  $H_0$  if both the observed test statistic is  $T = 2$  and a generated uniform random variable  $U$  from  $[0, 1]$  satisfies  $U \leq \frac{5}{8}$ .

### Calculating the Probability of a Type I Error

A Type I error occurs when we reject  $H_0$  when it is true. The probability of this error is the probability of the rejection event, calculated under the null distribution.

$$P(\text{type I error}) = P\left(T = 2 \text{ and } U \leq \frac{5}{8}\right)$$

The outcome of the test statistic  $T$  and the generated uniform random variable  $U$  are independent events. Therefore, we can multiply their probabilities:

$$P(\text{type I error}) = P(T = 2) \times P\left(U \leq \frac{5}{8}\right)$$

From the null distribution, we have  $P(T = 2) = 0.08$ . For a uniform random variable  $U$  on the interval  $[0, 1]$ , the probability  $P(U \leq k)$  is equal to  $k$  for any  $k \in [0, 1]$ . Thus,  $P(U \leq \frac{5}{8}) = \frac{5}{8}$ .

Substituting these values into the equation gives:

$$P(\text{type I error}) = 0.08 \times \frac{5}{8} = \frac{8}{100} \times \frac{5}{8} = \frac{5}{100} = 0.05$$

This shows that the actual probability of a Type I error for this randomized test is exactly 0.05.

### Is this a sensible test?

Whether this test is “sensible” is a point of statistical debate.

**Arguments for the test being sensible:** From a purely theoretical standpoint, the test is sensible because it achieves the desired nominal significance level of  $\alpha = 0.05$  exactly. In part (a), using the standard  $P$ -value, we found the actual error rate was 0, making the test overly conservative. The smallest attainable non-zero significance level with a non-randomized test would be  $P(T \geq 2) = 0.08$ . This randomized procedure provides a mechanism to create a rejection region with a precise probability, overcoming the limitations of the discrete test statistic.

**Arguments against the test being sensible:** From a practical and philosophical standpoint, the test can be seen as not sensible. The decision to reject a scientific hypothesis depends not only on the observed data (the value of  $T$ ) but also on the outcome of an auxiliary, unrelated random event (the value of  $U$ ). This means two researchers with the exact same data ( $T = 2$ ) could arrive at opposite conclusions based solely on the flip of a coin (or in this case, the output of a random number generator). This introduction of arbitrary randomness into the final decision-making process is unsettling to many practitioners and is a primary reason why such randomized tests are very rarely used in practice.

**Conclusion:** While the test is a valid statistical tool to achieve a desired Type I error rate exactly, it is not generally considered sensible for scientific practice due to its reliance on an external, arbitrary random device to make a final conclusion.