p-Values

April 1, 2020

1 p-Values

Throughout this course in many problems we were given an α from which to construct a confidence interval or Hypothesis test. This α was also chosen somewhat arbitrarily, depending on how risk-averse you were. An acceptable α might satisfy one person but not another. So instead of simply saying reject or not reject on a null hypothesis from a given data set, it can be desirable to state for which levels a hypothesis can be rejected. This is accomplished by p-values. Almost all scientific data is published with p-values.

DEFINITION 1.1. *The* ***p-value****, or* ***attainted significance level****, of a test statistic, is the smallest level α for which the null hypothesis should be rejected.*

Let's look at this is a very explicit example:

EXAMPLE 1.2. *You have a coin that you suspect is biased (meaning the probability of a heads is not .5). To test this hypothesis you flip the coin 10 times. If you get k heads what is the corresponding p-value?*

Solution: Let p be the probability the coin is heads, when flipped (Note: this p is different than the p in p-value, this is somewhat unfortunate notation, but rather standard, I guess you just get used to it).

Let X be the number of heads flipped and $\hat{p}=X/10$. Under $H_0\!:\!p=1/2$ we have the following probabilities for $\hat{p}$. (This is just the binomial distribution with $n=10$ and $p=1/2$.)

k	0	1	2	3	4	5	6	7	8	9	10
$\mathbb{P}(\hat{p}=k/10)$	.00098	.0098	.044	.12	.2	.25	.2	.12	.044	.0098	.00098

So if $p=.5$ the probability

$$\mathbb{P}(|\hat{p}-1/2|\geq 5/10)=\mathbb{P}(\hat{p}=10)+\mathbb{P}(\hat{p}=0)=2*.00098,$$

and the probability

$$\mathbb{P}(|\hat{p}-1/2|\geq 4/10)=\mathbb{P}(\hat{p}=10)+\mathbb{P}(\hat{p}=0)+\mathbb{P}(\hat{p}=9)+\mathbb{P}(\hat{p}=1)=2*.00098+2*.0098$$

and so on, we record this information in the following table:

k	5	6	7	8	9	10
$\mathbb{P}(\lvert\hat{p}-.5\rvert\geq(k-5)/10)$	1	.75	.34	.11	.021	.0020

(These numbers might seem slightly off due to rounding errors)

What this table means is that you flipped 0 or 10 heads, then you would reject $H_0 : p = 1/2$ in favor of $H_a : p \neq 1/2$, for any $\alpha > .002$, so your p-value is .002.

Similarly, if you flipped 7 heads, then you would be in the rejection region for any $\alpha > .34$. In other words the above table is the p-values if you flip k or $10-k$ heads.

If you want to report you data, you could say we reject the null hypothesis that $p = 1/2$ with a p-value and then state your number from the table. Then the reader could decide for themselves if that is a high enough level of significance for their purposes.

We now return to the general theory. If we want to do a one-sided hypothesis test, $H_0 : \theta = \theta_0$ versus $H_a : \theta > \theta_0$ using the test statistic $\hat{\theta}$ then we would have a rejection region of the form

$$RR = \{\hat{\theta} > k\}$$

for some k, depending on the distribution of $\hat{\theta}$ and α. Now that we are fixing not α, we compute the p-value as follows: For the observation $\hat{\theta} = t$ the p-value is:

$$p\text{-value} = \mathbb{P}(\hat{\theta} \geq t; \text{ when } H_0 \text{ is true})$$

Likewise if $H_a : \theta < \theta_0$ then

$$p\text{-value} = \mathbb{P}(\hat{\theta} \leq t; \text{ when } H_0 \text{ is true})$$

The case for two-sided hypothesis tests $H_a : \theta \neq \theta_0$, can be a little different, because there is not a unique way to choose 2-sided RR. However, when the distribution of the test statistic is symmetric around θ_0 (for example in the normal case with mean θ_0) then it is natural to choose a symmetric RR region. In this case our p-value is

$$p\text{-value} = \mathbb{P}(|\hat{\theta} - \theta_0| \geq t; \text{ when } H_0 \text{ is true})$$

for example with a normally distributed test statistic this quantity would be computed using $2z_{|\hat{\theta}-\theta_0|}$ multiplied by the appropriate standard deviation. Another example was done above with the coin flipping.

Let's look at a more typically example where we'll use the asymptotic normality of the test statistic to compute the p-value.

EXAMPLE 1.3. *In order for a certain drug to be sold it must be effective in* 60% *of the population. In a sample of 120 people, it is effective in* 65.8% *percent of the population. Find the p-value for this test. Would you claim the drug is effective at an* $\alpha = .1$ *level?*

Solution: Let p be the proportion of the population for which the drug is effective. (Note this p is different from the p in p-value).

The null hypothesis is $H_0 : p = .6$ and the alternative is $H_a : p > .6$ (so a one-sided test). The test statistic is $\hat{p}$, the number of people the drug is effective for divided by 120, in this case .658. Since we are reporting a p-value, there is no RR.

Under H_0, we assume $\hat{p}$ is approximately normal with mean .6 and standard deviation $\sqrt{\frac{.6(.4)}{120}}$. Our p-value is then the probability

$$\mathbb{P}(\hat{p} > .658)$$

which we standardize to

$$\mathbb{P}(\frac{\hat{p}-.6}{\sqrt{\frac{.6(.4)}{120}}} > \frac{.658-.6}{\sqrt{\frac{.6(.4)}{120}}}) \approx \mathbb{P}(Z > 1.3) = .0968$$

where Z is a standard normal.

So the p-value for this test is 9.6%. Since $9.6 < 10$, we accept the claim that the drug is effective at an $\alpha = .1$ level. On the other hand, we wouldn't at a, say, $\alpha = .05$ level.

In the future sections, we'll see how this works with other distributions, beyond the normal case.