Slides 17 Notes

John Hancock

April 1, 2018

Contents

1 References and License

In this document we are recording notes on reading material in [1].

Please see the references section for detailed citation information.

2 Significance Level

Significance level is area under the curve in the rejection region. So if a test has a high significance level, it means it has a large rejection region. In turn, it means, it would be easy to fail the test - it's easy to reject the null hypothesis in favor of the alternative because the rejection region is so big.

3 Board Question on Significance Testing

There is a board question on null hypothesis significance testing in [1]. This is our attempt to solve it. We are given a null hypothesis H_0 that some data follows a normal distribution $N\left(5, 10^2\right)$.
Orloff and Bloom also give us an alternative hypothesis that the data follows a normal distribution $N\left(\mu, 10^2\right)$ where $\mu \neq 5$.
They give a test statistic z which is equal to the standardized mean of the data $\bar{x}$.
Finally they give a significance level of $\alpha = 0.05$. We pause to remember that the significance level is the area under the probability density function where the x axis is in the rejection region.

Orloff and Bloom give us that the test statistic z is the standardized sample mean $\bar{x}$. Therefore, given the null hypothesis H_0 above

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{64}}}. \tag{1}$$

We us the values Orloff and Bloom give for the symbols in the equation above to calculate the value

$$z = \frac{6.25 - 5}{\frac{10}{8}}, \tag{2}$$

The equation above simplifies to

$$z = \frac{\frac{5}{4}}{\frac{5}{4}}. \tag{3}$$

Hence, $z = 1$.
We use R's qnorm function to find $z_0.025 \approx 1.96$. z is less than 1.96, so our test statistic does not fall in the rejection region. We use the pnorm function of R [2],to find that the p-value for our test statistic is 0.32. This is larger than the significance level of 0.05 - it means that if the null hypothesis is true, then there is a 0.32 probability of our observing the test statistic.
It is interesting to note that the R code for calcuating the p-value is

```
2*pnorm(1, lower.tail=FALSE)
```

We are asking R for the area under the probability distribution function of the normal distribution with mean 0 and standard deviation 1 for values of a normally distributed random variable x where the value of x is greater than 1. We multiply that area by 2 because our alternative hypothesis is that our random variable has a mean value not equal to something.
Our mean value could be greater than 5, or it could be less than 5. This means that our rejection region must encompass positive and negative numbers.

References

[1] Jeremy Orloff and Jonathan Bloom. *Frequentist Statistics and Hypothesis testing 18.05 Spring 2014*. Available at `https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/class-slides/MIT18_05S14_class17_slides.pdf` (Spring 2014).

[2] Chi Yaho. *Two-Tailed Test of Population Proportion.* Available at `http://www.r-tutor.com/elementary-statistics/hypothesis-testing/two-tailed-test-population-proportion` (2009-2018).