MA4413 2013

Week 9 Tutorial

November 6, 2013

1 Accuracy Precision and Recall

	Null hypothesis (H_0) true	Null hypothesis (H_0) false
Reject null hypothesis	**Type I error** False positive	**Correct outcome** True positive
Fail to reject null hypothesis	**Correct outcome** True negative	**Type II error** False negative

In the context of a binary classification prediction procedure

	Predicted Negative	Predicted Positive
Observed Negative	True Negative	False Positive
Observed Positive	False Negative	True Positive

Accuracy, Precision and recall are defined as

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

The F measure is computed as

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Questions

	Predicted Negative	Predicted Positive
Negative Cases	TN: 9,700	FP: 165
Positive Cases	FN: 35	TP: 100

With reference to the table above, compute each of the following appraisal metrics.

a. Accuracy

b. Precision

c. Recall

d. F measure

2 Computer Output

Statistical Procedures were performed using a statistical programming language called R. A brief description of each procedure is provided. For each procedures, identify the null and alternative hypotheses, the $p-$value, and your conclusion for this test.

- If the $p-$value is less than 0.05 : reject the null hypothesis.
- If the $p-$value is greater than 0.05 : fail to reject the null hypothesis.

Test 1. Single Sample Test for Proportions

- Sample size $(n) = 500$
- Number of successes $(x) = 280$
- Expected value under null hypothesis (Usually π, but here as p)

```
> prop.test(x=280,n=500,p=0.60)

   1-sample proportions test with continuity correction

data:  280 out of 500, null probability 0.6
X-squared = 3.1688, df = 1, p-value = 0.07506
alternative hypothesis: true p is not equal to 0.6
95 percent confidence interval:
 0.5151941 0.6038700
sample estimates:
   p
0.56
>
```

Test 2. F Test for equality of variance

- In this procedure, we determine whether or not two ***populations*** have the same variance.
- The assumption of equal variance of two populations underpins several inference procedures. This assumption is tested by comparing the variance of samples taken from both populations.
- The null and alternative hypotheses are as follows:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

```
> var.test(X,Y)

        F test to compare two variances

data:  X and Y
F = 2.5122, num df = 9, denom df = 9, p-value = 0.1862
alternative hypothesis: true ratio of variances
  is not equal to 1
95 percent confidence interval:
  0.6239986 10.1141624
sample estimates:
ratio of variances
          2.512215
```

Test 3. Shapiro Wilk's Test for Normality

- We will often be required to determine whether or not a data set is normally distributed. This assumption underpins many statistical models.
- The null hypothesis is that the data set is normally distributed.
- The alternative hypothesis is that the data set is not normally distributed.
- One procedure for testing these hypotheses is the Shapiro-Wilk test, implemented in `R` using the command `shapiro.test()`.

```
> shapiro.test(X)

        Shapiro-Wilk normality test

data:  X
W = 0.9849, p-value = 0.1012
```

Test 4. Grubbs' Test for Determining an Outlier

The Grubbs' test is used to determine if there are any outliers in a data set.

There is no agreed formal definition for an outlier. The definition of outlier used for this procedure is a value that unusually distant from the rest of the values (For the sake of clarity , we shall call this type of outlier a **Grubbs' Outlier**). Consider the following data set: is the lowest value 4.01 an outlier?

```
6.98 8.49 7.97 6.64
8.80 8.48 5.94 6.94
6.89 7.47 7.32 4.01
```

```
> grubbs.test(x, two.sided=T)
Grubbs test for one outlier
data: x
G = 2.4093, U = 0.4243, p-value = 0.05069
alternative hypothesis: lowest value 4.01 is an outlier
```