University of New Mexico

Hypothesis Testing-1 (Fall 2018)

BIOM 505: Biostatistical Methods I (by Fares Qeadan)

Hypothesis Testing about one Population Proportion (one sample proportion z-test) [using sample statistics]:

Assumptions of this test:

- We are sampling less than 10% of the total population.
- The sample size is sufficiently large such that $np_0 \geq 10$ and $n(1-p_0) \geq 10$.
- Sample observations should be random.
- Sample individuals are independent.

(1) **In a 1999-2000 longitudinal study of youth baseball, researchers found that 77 of 248 young pitchers complained of elbow pain after pitching. A spokesman for the league asserts that the proportion of young pitchers who will develop elbow pain is less than 37%. Test this claim at the 5% significance level.**

(a) The significance level α is:

(b) Give the claim as a mathematical statement:

(c) The null and alternative hypotheses are:

(d) The decision rule (about H_0) is:

(e) Compute the test-statistic, critical value, and p-value

```
%MACRO percentzt(n,x,p0);
data _null_;
  file print;
  p0 = &p0.; n = &n.; x = &x.;
  phat = x/n;
  Z = (phat - p0)/sqrt(p0 * (1-p0)/n);
  ptwosided = 2*(1 - probnorm(abs(Z)));
  prightsided = 1 - probnorm(Z);
  pleftsided = probnorm(Z);
  put '===== Test for one sample proportion =====';
  put 'n = ' n  '  x =' x  '  p.hat=' phat;
  put 'p0 = ' p0;
  put 'Z = ' Z;
  put 'Pr > |Z|: ' ptwosided pvalue.;
  put 'Pr > Z:  ' prightsided pvalue.;
  put 'Pr < Z:  ' pleftsided pvalue.;
run;
%MEND percentzt;
%percentzt(248,77,0.37);


===== Test for one sample proportion =====
n = 248   x =77   p.hat=0.310483871
p0 = 0.37
Z = -1.941285364
Pr > |Z|: 0.0522
Pr > Z:  0.9739
Pr < Z:  0.0261
```

(f) Decision:

(g) Conclusion:

(h) State the error you might have made in the decision above and identify it as Type I or Type II?

(2) **In a 1999-2000 longitudinal study of youth baseball, researchers found that 77 of 248 young pitchers complained of elbow pain after pitching. A spokesman for the league asserts that the proportion of young pitchers who will develop elbow pain is different than 36%. Test this claim at the 5% significance level.**

(a) The significance level α is:

(b) Give the claim as a mathematical statement:

(c) The null and alternative hypotheses are:

(d) The decision rule (about H_0) is:

(e) Compute the test-statistic, critical values and p-value:

```
%percentzt(248,77,0.37);


===== Test for one sample proportion =====
n = 248   x =77   p.hat=0.310483871
p0 = 0.37
Z = -1.941285364
```

```
Pr > |Z|: 0.0522
Pr > Z:  0.9739
Pr < Z:  0.0261
```

(f) Decision:

(g) Conclusion:

(h) State the error you might have made in the decision above and identify it as Type I or Type II?

(i) Construct the 95% confidence interval about the true proportion of young pitchers who will develop elbow pain?

(j) Is your conclusion in (h) consistent with your findings in (j)?

Hypothesis Testing about one Population Mean (one sample t-test)[using sample statistics]

Assumptions of this test:

- The population from which the sample has been drawn should be roughly normal and minor departures from normality do not affect this test. In fact, this assumption is not too important with large samples. We will be studying formal statistical methods for testing this assumption (For example the Shapiro-Wilk test).
- The underlying population standard deviation is unknown.
- Sample observations should be random.
- Sample individuals are independent.

(3) **The claim is made that the average number of days of sick leave taken by women at UNM is higher than 8.4 days (the overall average for all employees). In order to investigate the validity of this claim, a randomly selected records of 100 women were examined. The mean number of days sick leave for this sample was 8.5 with a standard deviation of 3.4 days. Conduct a test of hypothesis about this mean using a level of significance of 10% while assuming that number of days of sick leave taken by women at UNM is roughly normally distributed?**

(a) The significance level α is:

(b) Give the claim as a mathematical statement:

(c) The null and alternative hypotheses are:

(d) The decision rule (about H_0) is:

(e) Compute the test-statistic, critical values and p-value

```
%MACRO ttest1(n,xbar,s,mu0);
data _null_;
  file print;
  n = &n.;  xbar = &xbar.; s = &s.;  mu0 = &mu0.;
  t = (xbar - mu0)/(s / sqrt(n));
  ptwosided = 2*(1 - cdf('t', abs(t),n-1));
  prightsided =1 - cdf('t', t,n-1);
  pleftsided =cdf('t', t,n-1);
  put '===== Test for one sample mean =====';
  put 'n = ' n  '  xbar =' xbar '  s =' s  ;
  put 'mu0 = ' mu0;
   put 't = ' t;
  put 'Pr > |t|: ' ptwosided pvalue.;
  put 'Pr > t:  ' prightsided pvalue.;
  put 'Pr < t:  ' pleftsided pvalue.;
run;
%MEND ttest1;
%ttest1(100,8.5,3.4,8.4);
```

```
===== Test for one sample mean =====
n = 100   xbar =8.5   s =3.4
mu0 = 8.4
t = 0.2941176471
Pr > |t|: 0.7693
Pr > t:  0.3846
Pr < t:  0.6154
```

(f) Decision:

(g) Conclusion:

(h) State the error you might have made in the decision above and identify it as Type I or Type II?

Hypothesis testing about one population proportion/mean in STATA [using sample data]

(4) **Consider the Diabetes and obesity, cardiovascular risk factors data set we have used in Lab 1 (link: http://www.mathalpha.com/lab1/diabetesfall17.sas7bdat) to answer the following questions:**

(a) Test the claim that the true proportion of diabetes among African Americans in Virginia is higher than 12%?

SOLUTION:
```
proc freq data=biom505.diabetesfall17;
table diab/binomial(level='1' p=0.12);
run;
```

(b) Test the claim that the true mean HDL among African Americans in Virginia is less than 52?

SOLUTION:
```
proc ttest data=biom505.diabetesfall17  SIDES= L alpha=0.05 H0 = 52;
var hdl;
run;
```

(c) Test the claim that the true mean Age of African Americans in Virginia is significantly different than 46?

SOLUTION:
```
proc ttest data=biom505.diabetesfall17  SIDES= 2 alpha=0.05 H0 = 46;
var age;
run;
```