Chapter 4.4 Basic tests concerning one parameter

Grethe Hystad

November 14, 2012

The p-value

Definition

The **p-value** associated with a test is the probability that we obtain a value of the test statistics that is at least as extreme (in the direction of the alternative) as the observed value of our test statistics assuming the null hypothesis is true.

- Instead of selecting the critical region in advance, we can report the p-value and consider all the values of α that would give the rejection of the null hypothesis.

The p-value

Definition

The **p-value** associated with a test is the probability that we obtain a value of the test statistics that is at least as extreme (in the direction of the alternative) as the observed value of our test statistics assuming the null hypothesis is true.

- Instead of selecting the critical region in advance, we can report the p-value and consider all the values of α that would give the rejection of the null hypothesis.
- If the p-value is less than the significance level, α, we reject H_0 and say that the result is **statistically significant** at the level α.

The p-value

Definition

The **p-value** associated with a test is the probability that we obtain a value of the test statistics that is at least as extreme (in the direction of the alternative) as the observed value of our test statistics assuming the null hypothesis is true.

- Instead of selecting the critical region in advance, we can report the p-value and consider all the values of α that would give the rejection of the null hypothesis.
- If the p-value is less than the significance level, α, we reject H_0 and say that the result is **statistically significant** at the level α.
- The smaller the p-value, the stronger the evidence we have against the null hypothesis, H_0.

Example

Lets go back to the last example in chapter 4.3 where we tested,

$$H_0 : \mu = 500 \quad \text{against} \quad H_1 : \mu > 500$$

with $n = 49$, $\overline{x} = 530$ and $\sigma = 100$. We rejected H_0 if $z = \frac{\overline{x} - 500}{\frac{100}{\sqrt{49}}} > 1.645$. We have $z = \frac{530 - 500}{\frac{100}{7}} \approx 2.1$. Since $Z = \frac{\overline{X} - 500}{\frac{100}{7}}$ is $N(0, 1)$ if H_0 is true, we have that the p-value$= P(Z \geq 2.1) = 0.0179 < 0.05$.

- If the p-value is less than α, we reject H_0 and accept H_1.

Example

Lets go back to the last example in chapter 4.3 where we tested,

$$H_0 : \mu = 500 \quad \text{against} \quad H_1 : \mu > 500$$

with $n = 49$, $\overline{x} = 530$ and $\sigma = 100$. We rejected H_0 if $z = \frac{\overline{x} - 500}{\frac{100}{\sqrt{49}}} > 1.645$. We have $z = \frac{530 - 500}{\frac{100}{7}} \approx 2.1$. Since $Z = \frac{\overline{X} - 500}{\frac{100}{7}}$ is $N(0, 1)$ if H_0 is true, we have that the p-value$= P(Z \geq 2.1) = 0.0179 < 0.05$.

- If the p-value is less than α, we reject H_0 and accept H_1.
- Since the p-value$= 0.0179 < 0.05$, we reject H_0 at the significance level $\alpha = 0.05$.

Example

Lets go back to the last example in chapter 4.3 where we tested,

$$H_0 : \mu = 500 \quad \text{against} \quad H_1 : \mu > 500$$

with $n = 49$, $\overline{x} = 530$ and $\sigma = 100$. We rejected H_0 if $z = \frac{\overline{x} - 500}{\frac{100}{\sqrt{49}}} > 1.645$. We have $z = \frac{530 - 500}{\frac{100}{7}} \approx 2.1$. Since $Z = \frac{\overline{X} - 500}{\frac{100}{7}}$ is $N(0, 1)$ if H_0 is true, we have that the p-value$= P(Z \geq 2.1) = 0.0179 < 0.05$.

- If the p-value is less than α, we reject H_0 and accept H_1.
- Since the p-value$= 0.0179 < 0.05$, we reject H_0 at the significance level $\alpha = 0.05$.
- However, we would **not** reject H_0 at the significance level $\alpha = 0.01$ since p-value$= 0.0179 > 0.01$.

Example

In the previous example, if we instead had tested

$$H_0 : \mu = 500 \quad \text{against} \quad H_1 : \mu \neq 500,$$

then the

$$\text{p-value} = P(|Z| > 2.1) = 2 \cdot 0.0228 = 0.0456 < 0.05$$

and we would reject H_0 at the 5% significance level.

If the underlying distribution is $N(\mu, \sigma^2)$, where σ^2 is known, the test statistics is given by

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}.$$

If σ is unknown, we replace σ with the sample standard deviation, s, to get the test statistics

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

which has the t-distribution with $n - 1$ degrees of freedoms.

Test of hypothesis

One sided composite hypothesis test about mean with unknown variance:

$$H_0 : \mu = \mu_0 \quad \textit{against} \quad H_1 : \mu > \mu_0.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$t \geq t_\alpha$$

which is equivalent to

$$\overline{x} \geq \mu_0 + t_\alpha \frac{s}{\sqrt{n}}.$$

Test of hypothesis

One sided composite hypothesis test about mean with unknown variance:

$$H_0 : \mu = \mu_0 \quad \textit{against} \quad H_1 : \mu < \mu_0.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$t \leq -t_\alpha$$

which is equivalent to

$$\overline{x} \leq \mu_0 - t_\alpha \frac{s}{\sqrt{n}}.$$

Test of hypothesis

Two sided composite hypothesis test about mean with unknown variance:

$$H_0 : \mu = \mu_0 \quad \textit{against} \quad H_1 : \mu \neq \mu_0.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$|t| \geq t_{\alpha/2}$$

which is equivalent to

$$|\overline{x} - \mu_0| \geq t_{\alpha/2} \frac{s}{\sqrt{n}}.$$

Lets return to the following example from chapter 4.2

Example

The Trial Urban District Assessment (TUDA) is a study sponsored by the government of student achievement in large urban school district. The math test-score is on a scale from 0 to 500. A "basic" math level is a score of 262, a "proficient" level is a score of 299 and a "advanced" level is a score of 333. In 2007, a random sample of 2000 *eighth*-graders from Los Angeles had an average math scale score of $\overline{x} = 257$ with a standard deviation of 49.19. (The study reports the standard error of the mean instead of the standard deviation.) Source: TUDA results for 2007 from the National Center for Education Statistics, at nces.ed.gov/nationsreportcard

(A) Determine if the mean scale score for all the eighth graders in LA was significantly lower than the basic level at the 1% level.
(B) Determine the approximate p-value of this test.

Solution

We test

$$H_0 : \mu = 262 \quad \textit{against} \quad H_1 : \mu < 262.$$

$>$ *qt(0.01,1999)*
[1] -2.328215

We have $t_{0.99}(1999) = -2.328$ *and*

$$t = \frac{\overline{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{257 - 262}{\frac{49.19}{\sqrt{2000}}} = -4.545 < -2.328.$$

(A) We reject the null hypothesis, H_0*, at the* 1% *significance level. Also recall from chapter 4.2 that* 262 *was not in the* 99% *confidence interval for* μ*. Thus, the mean scale score for the eighth graders in LA was significantly lower than the basic level at the* 1% *level*

Solution continue

Solution

(B)

```
> pt(-4.545,1999)
[1] 2.910936e-06
```

$$p\text{-}value = P(T < -4.545) = 2.910936e^{-06}.$$

Example

In 1879 Michelson determined the speed of light. He had $n = 100$ observations (suitably coded (km/sec, with 299000 subtracted).) Stigler reported the true speed of light to be 710.5. Michelson data can be found by typing

$>$ morley

in R.

The mean in Michelson's data is $\overline{x} = 852.4$ and the standard deviation is 79.01.

(A) Determine if Michelson's mean of the data is significant different than 710.5 at the 5% significance level.

(B) Find the approximate p-value of this test.

Solution

We want to test

$$H_0 : \mu = 710.5 \quad \textit{against} \quad H_1 : \mu \neq 710.5.$$

We have
> qt(0.975,99)
[1] 1.984217
so $t_{0.025}(99) = 1.984217$
and

$$t = \frac{|\overline{x} - \mu_0|}{\frac{s}{\sqrt{n}}} = \frac{852.4 - 710.5}{\frac{79.01055}{\sqrt{100}}} \approx 17.960 > 1.984.$$

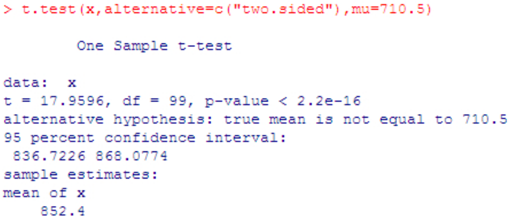
We reject the null hypothesis at the 5% *significance level. Thus Michelson's determinations of speed of light is significantly different than Stigler's value of speed of light.*

Solution continue

Solution

In the previous example, the p-value is

$$p\text{-}value = P(|T| > 17.960) \approx 2 \cdot 3.252e^{-33} = 6.504e^{-33}.$$



```
> t.test(x,alternative=c("two.sided"),mu=710.5)

        One Sample t-test

data:  x
t = 17.9596, df = 99, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 710.5
95 percent confidence interval:
 836.7226 868.0774
sample estimates:
mean of x
    852.4
```

Matched pair t-test

Let $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_n)$ be two random samples measured on the same subjects such that X and Y are **dependent**. For example X could be the old treatment given to a group of patients and Y could be the new treatment give to the same group of patients. We then want to measure the effect of Y compared to X. Let $W = X - Y$. The null hypothesis is then $H_0 : \mu_W = 0$ and the alternative hypothesis is either one-sided, ($\mu_W < 0$ or $\mu_W > 0$), or two sided, $\mu_W \neq 0$.

The t-statistics is

$$T = \frac{\overline{X} - \overline{Y}}{\frac{S_{X-Y}}{\sqrt{n}}} = \frac{\overline{W}}{\frac{S_W}{\sqrt{n}}}$$

which has a t-distribution with $n - 1$ degrees of freedom.

Example

In 1908 William Gosset under the pseudonym **Student** published the article "The Probable error of a mean". He seeded plots with two different types of seed, regular and kiln-dried, and reported the result. Each type of seed was planted in adjacent plot, giving 11 pairs of "split" plots. His article is maybe the most famous in statistical literature and is being read even today.

Here are his data given in the 1908 paper:

$X = 1903\ 1935\ 1910\ 2496\ 2108\ 1961\ 2060\ 1444\ 1612\ 1316\ 1511$

$Y = 2009\ 1915\ 2011\ 2463\ 2180\ 1925\ 2122\ 1482\ 1542\ 1443\ 1535$

X=REG=Corn yield (lbs/acre) from regular seed.

Y=KILN = Corn yield from kiln-dried seed.

Source: DASL at http://lib.stat.cmu.edu/DASL/

Is the mean corn yield from kiln-dried seed higher compared to the mean corn yield from regular seed at the 5% significance level?

Solution

Let $W = X - Y$. *We test the hypothesis:*

$$H_0 : \mu_W = 0 \quad \textit{against} \quad H_1 : \mu_W < 0.$$

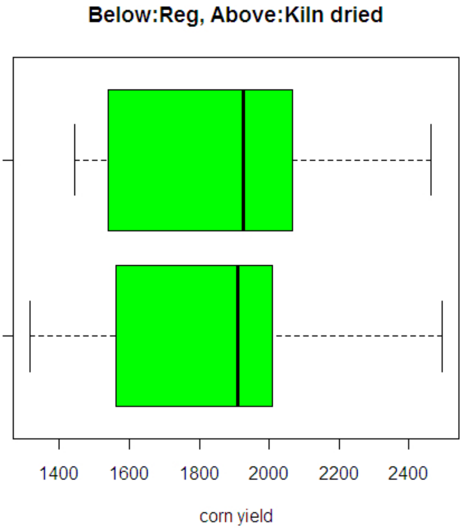
We have
$W = c(-100, 20, -101, 33, -72, 36, -62, -38, 70, -127, -24)$,
where $\overline{W} = -33.72727$ *and* $s_W = 66.17113$. *We have*
$>$ *qt(0.05,10)*
[1] *-1.812461*
so $t_{0.95}(10) = -1.812461$. *Then*

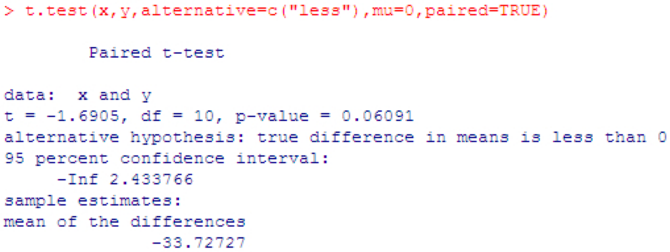
$$t = \frac{\overline{w} - 0}{\frac{s_W}{\sqrt{n}}} = -\frac{33.72727}{\frac{66.17113}{\sqrt{11}}} = -1.690476 > t_{0.95}(10).$$

Solution continue

Solution

The sample size, $n = 11$*, is small, but it can be shown that the differences in data is approximately normal. No outliers does appear in the data of differences. Thus, we accept* H_0 *at the* 5% *significance level. Therefore one should not treat the seeds by kiln drying them. However, this does not prove that the corn yields are equal, only that the data of differences does not deviate enough from zero for us to reject the null hypthesis. Maybe a higher sample size would have given a different result.*


Below:Reg, Above:Kiln dried
1400
1600
1800
2000
2200
2400
corn yield



```
> t.test(x,y,alternative=c("less"),mu=0,paired=TRUE)

        Paired t-test

data:  x and y
t = -1.6905, df = 10, p-value = 0.06091
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 2.433766
sample estimates:
mean of the differences
              -33.72727
```

Tests about the variance

Suppose again that the underlying distribution is $N(\mu, \sigma^2)$, but we now want to find a test about the variance, σ^2. The null hypothesis is $H_0 : \sigma^2 = \sigma_0^2$ and the test statistics is

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

which has the chi-square distribution with $r = n - 1$ degrees of freedom when H_0 is true.

Test of hypothesis

One sided composite hypothesis test about the variance:

$$H_0 : \sigma^2 = \sigma_0^2 \quad \textit{against} \quad H_1 : \sigma^2 > \sigma_0^2.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$s^2 \geq \frac{\sigma_0^2 \chi_\alpha^2(n-1)}{n-1}$$

which is equivalent to

$$\chi^2 \geq \chi_\alpha^2(n-1).$$

Test of hypothesis

One sided composite hypothesis test about the variance:

$$H_0 : \sigma^2 = \sigma_0^2 \quad \textit{against} \quad H_1 : \sigma^2 < \sigma_0^2.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$s^2 \leq \frac{\sigma_0^2 \chi^2_{1-\alpha}(n-1)}{n-1}$$

which is equivalent to

$$\chi^2 \leq \chi^2_{1-\alpha}(n-1).$$

Test of hypothesis

Two sided composite hypothesis test about the variance:

$$H_0 : \sigma^2 = \sigma_0^2 \quad \textit{against} \quad H_1 : \sigma^2 \neq \sigma_0^2.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$s^2 \leq \frac{\sigma_0^2 \chi^2_{1-\alpha/2}(n-1)}{n-1} \quad \text{or} \quad s^2 \geq \frac{\sigma_0^2 \chi^2_{\alpha/2}(n-1)}{n-1}$$

which is equivalent to

$$\chi^2 \leq \chi^2_{1-\alpha/2}(n-1) \quad \text{or} \quad \chi^2 \geq \chi^2_{\alpha/2}(n-1).$$

Example

Suppose $X_1, ..., X_{25}$ is a random sample of $n = 25$ i.i.d. $N(\mu, \sigma^2)$. We want to test

$$H_0 : \sigma^2 = 49 \quad \text{against} \quad H_1 : \sigma^2 < 49$$

at an $\alpha = 0.05$ significance level. Suppose the observed value of the sample variance was $s^2 = 40$. Do we accept or reject H_0?

Solution

Let $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = \frac{(24)(40)}{49} = 19.592$. *We have*

```
> qchisq(0.05, 24)
[1] 13.84843
```

so $\chi^2_{0.95}(24) = 13.84843$. *Since* $\chi^2 > \chi^2_{0.95}(24)$, *we accept* H_0.
Thus, the true σ^2 *is not significant smaller than* 49 *at the* 5% *level.*

Hypothesis test for one proportion

- Let Y be binomial, $b(n, p)$, where p is unknown.
- We want to test the hypothesis, $H_0 : p = p_0$ against $p > p_0$, $p < p_0$ or $p \neq p_0$.
- Let $\hat{p} = \frac{Y}{n}$ which is the estimator of p.
- Recall that $E(\frac{Y}{n}) = p$ and $Var(\frac{Y}{n}) = p(1-p)$.
- If H_0 is true, $E(\hat{p}) = E(\frac{Y}{n}) = p_0$.
- We estimate the standard deviation of $\hat{p}$ to be $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ which is called the standard error of $\hat{p}$.
- The hypothesis tests are summarized on the next three pages.

Test of hypothesis

One sided composite hypothesis test for one proportion:

$$H_0 : p = p_0 \quad \textit{against} \quad H_1 : p > p_0.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$z = \frac{y/n - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} \geq z_\alpha$$

Test of hypothesis

One sided composite hypothesis for one proportion:

$$H_0 : p = p_0 \quad against \quad H_1 : p < p_0.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$z = \frac{y/n - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} \leq -z_\alpha$$

Test of hypothesis

Two sided composite hypothesis for one proportion:

$$H_0 : p = p_0 \quad \textit{against} \quad H_1 : p \neq p_0.$$

We reject H_0 at the $(\alpha \times 100)\%$ significance level if

$$|z| = \frac{|y/n - p_0|}{\sqrt{\hat{p}(1-\hat{p})/n}} \geq z_{\alpha/2}$$

Test of hypothesis

Note: In testing $H_0 : p = p_0$ some statisticians use p_0 rather than $\hat{p}$ in the denominator of z. That is

$$z = \frac{y/n - p_0}{\sqrt{p_0(1 - p_0)/n}}.$$

The numerical results are about the same.

Example

The national Assessment of Educational Progress (NAEP) is a congressionally authorized project of the National Center for Education statistics (NCES) within the Institute of Education Science of the U.S. department of Education. NAEP has since 1969 measured achievement periodically in several subjects including mathematics. The NAEP assessment in mathematics in 2011 measured the student's knowledge of mathematics and their problem solving skill across five mathematics content areas: number properties and operations, measurement, geometry, data analysis, statistics, and probability, and algebra.
Source: TUDA results for 2011 from the National Center for Education Statistics, at
http://nces.ed.gov/nationsreportcard/pdf/main2011/2012458.pdf

Example continue

Example

"The Trial Urban District Assessment (TUDA) is designed to explore the feasibility of using National Assessment of Educational Progress (NAEP) to report on the performance of public school students at the district level." In 2011, between 1,000 and 2,700 fourth- and eighth-grade public school students from 21 urban districts participated in the NAEP TUDA in mathematics.
Source: http://nationsreportcard.gov/tuda.asp
In 2011, 792 out of a sample of 1200 students in San Diego scored above the Basic level in mathematics. In large public city, 63% scored above the Basic level in mathematics.
Source:TUDA results for 2011 from the NCES.
Is the proportion of students who scored above the Basic level in San Diego significantly larger than 0.63 at the 5% level?

Solution

$$H_0 : p = 0.63 \quad against \quad H_1 : p > 0.63.$$

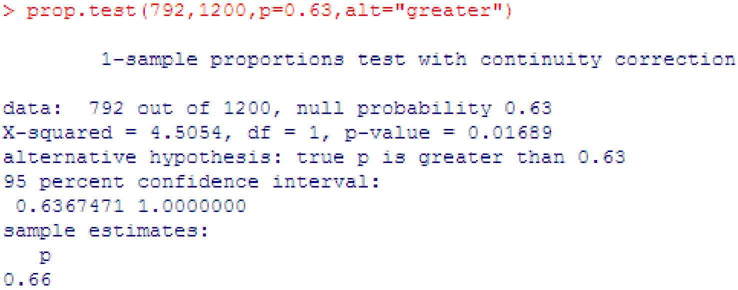
We have $z_{0.05} = 1.645$. *We have* $\frac{y}{n} = \hat{p} = \frac{792}{1200} = 0.66$ *and*

$$z = \frac{y/n - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} = \frac{0.66 - 0.63}{\sqrt{0.66(1-0.66)/1200}} \approx 2.1934 > 1.645.$$

Using p_0 *rather than* $\hat{p}$ *in the denominator, we obtain*

$$z = \frac{y/n - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.66 - 0.63}{\sqrt{0.63(1-0.63)/1200}} \approx 2.152 > 1.645.$$

Thus we reject H_0 *at the* 5% *significance level. Thus the proportion of students who scored above the Basic level in San Diego is larger than* 0.63 *at the* 5% *significance level.*



```
> prop.test(792,1200,p=0.63,alt="greater")

        1-sample proportions test with continuity correction

data:  792 out of 1200, null probability 0.63
X-squared = 4.5054, df = 1, p-value = 0.01689
alternative hypothesis: true p is greater than 0.63
95 percent confidence interval:
 0.6367471 1.0000000
sample estimates:
   p 
0.66 
```