To Do

Read Sections 5.1 - 5.2 (Hypothesis Testing)

Do End-of-Chapter Problems 1 - 8.

Assignment 3 due Friday November 11.

See detailed information posted on Learn regarding material covered by Midterm Test 2 (4:40 - 6:10 on Tuesday November 15).

Last Class

(1) Null hypothesis, Alternative Hypothesis

(2) Test statistic or Discrepancy Measure

(3) Steps of a Test of Hypothesis

(4) Interpretation of a p-value

Steps of a Statistical Test of Hypothesis

(1) Assume that the null hypothesis H_0 will be tested using data Y.

(2) Adopt a test statistic or discrepancy measure $D(Y)$ for which, large values of D are less consistent with H_0. Let $d = D(y)$ be the corresponding observed value of D.

(3) Calculate

$$p\text{-value} = \mathrm{P}(D \geq d \text{ assuming } H_0 \text{ is true})$$
$$= \mathrm{P}(D \geq d;\, H_0)$$

(4) Draw a conclusion based on the p-value.

Guidelines for Interpreting the *p*-value

These are only guidelines for this course.

The interpretation of a *p*-value must always be made in the context of a given study.

p-value	Interpretation
$p > 0.1$	There is no evidence against H_0 based on the data.
$0.05 < p \leq 0.1$	There is some evidence against H_0 based on the data.
$0.01 < p \leq 0.05$	There is evidence against H_0 based on the data.
$0.001 < p \leq 0.01$	There is strong evidence against H_0 based on the data.
$p \leq 0.001$	There is very strong evidence against H_0 based on the data.

Interpreting the *p*-value

See the *p*-value bears:

http://www.youtube.com/watch?v=ax0tDcFkPic&feature=related

ESP Experiment: $n = 100$

Suppose we did the ESP experiment for $n = 100$ trials and Student answered correctly 60 times.

The test statistic would now be $D = |Y - 50|$ and the observed value is $d = |60 - 50| = 10$.

ESP Experiment: $n = 100$

p-value = P($D \geq 10$; assuming H_0 is true)

= P($D \geq 10$; H_0)

= P($|Y - 50| \geq 10$) where $Y \backsim$ Binomial(100,0.5)

$$= P\left(\frac{|Y-50|}{\sqrt{100(0.5)(0.5)}} \leq \frac{10}{\sqrt{100(0.5)(0.5)}} \right) \text{ (no continuity correction used)}$$

$\approx$ P($|Z| \geq 2$)   where $Z \backsim N(0,1)$

= 2[1 - P($Z \leq 2$) - 1] = 2(1 - 0.97725)

= 0.04550

What would we conclude now about Student's ESP ability?

Today's Lecture

(1) Testing H_0: $\mu = \mu_0$ when σ is unknown for G(μ,σ) model.

(2) Statistical significance versus practical significance.

(3) Relationship between tests of hypothesis and confidence intervals

Tests of hypotheses for the parameters in a G(μ,σ) model

The G(μ,σ) has two parameters μ and σ.

Today we look at testing H_0: $\mu = \mu_0$ when σ is unknown.

Next class we look at testing H_0: $\sigma^2 = \sigma_0^2$ when μ is unknown.

Testing H_0: $\mu = \mu_0$ when σ is unknown

Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from a G(μ,σ) distribution.

There is a close relationship between the pivotal quantities we used to find confidence intervals and test statistics for testing hypotheses.

Recall the pivotal quantity:

$$\frac{\bar{Y} - \mu}{S / \sqrt{n}} \sim t(n-1)$$

Testing H_0: $\mu = \mu_0$ when σ is unknown

To test H_0: $\mu = \mu_0$ we use the test statistic

$$D = \frac{\left|\overline{Y} - \mu_0\right|}{S / \sqrt{n}}$$

Why does this test statistic make sense?

Testing H_0: $\mu = \mu_0$ when σ is unknown

To test H_0: $\mu = \mu_0$ we use the test statistic

$$D = \frac{\left|\overline{Y} - \mu_0\right|}{S / \sqrt{n}}$$

Why does this test statistic make sense?

$E\left(\overline{Y}\right) = \mu_0$ **if H_0: $\mu = \mu_0$ is true.**

Testing H_0: $\mu = \mu_0$ when σ is unknown

Let

$$d = \frac{\left|\bar{y} - \mu_0\right|}{s / \sqrt{n}}$$

be the observed value of *D* for an experiment which has been conducted.

$$p - value = P\left(D \geq d;\, H_0 \text{ is true}\right)$$

$$= P\left(\frac{|\bar{Y} - \mu_0|}{S / \sqrt{n}} \geq \frac{|\bar{y} - \mu_0|}{s / \sqrt{n}}\right)$$

$$= P\left(|T| \geq d\right) \text{ where } T \sim t(n-1)$$

$$= 2\left[1 - P\left(T \leq d\right)\right]$$

Example: Bias in a Measurement System

An inexpensive weight scale is tested by taking ten weighings of a known 1 kg weight.

The measurements were:

1.026	0.998	1.017	1.045	0.978
1.004	1.018	0.965	1.010	1.000

Assume $Y_i \backsim G(\mu,\sigma)$, $i = 1,2,\ldots,10$ where Y_i = ith measurement and μ represents the mean measurement in repeated weighings of the 1 kg weight using this scale.

The hypothesis of interest is H_0: $\mu = 1$. (Why?)

Example: Bias in a Measurement System

For these data

$$\bar{y} = 1.0061,\ \ \mu_0 = 1,\ \ s = 0.0230,\ \ n = 10$$

$$d = \frac{|\,\bar{y} - \mu_0\,|}{s/\sqrt{n}} = \frac{|1.0061 - 1|}{0.0230/\sqrt{10}} = 0.839$$

$$p - value = 2\left[1 - P\left(T \le 0.839\right)\right]\ \ T \sim t\left(9\right)$$

$$= 2\left(1 - 0.7884\right) \approx 0.42$$

Since the p-value ≈ 0.42 then based on the observed data there is no evidence against H_0: $\mu = 1$. There is no evidence that the scale is over or under weighing.

Example: Bias in a Measurement System

For a different set of inexpensive weigh scales the observed data were:

1.011	0.966	0.965	0.999	0.988
0.987	0.956	0.969	0.980	0.988

Example: Bias in a Measurement System

For these data

$$\bar{y} = 0.981,\ \ \mu_0 = 1,\ \ s = 0.0170,\ \ n = 10$$

$$d = \frac{|\,\bar{y} - \mu_0\,|}{s/\sqrt{n}} = \frac{|\,0.981 - 1\,|}{0.0170/\sqrt{10}} = 3.534$$

$$p - value = 2\left[1 - P\left(T \le 3.534\right)\right]\ \ T \sim t\left(9\right)$$

$$= 0.0064$$

Based on the observed data there is no evidence against H_0: $\mu = 1$, that is, there is strong evidence that the scale is over or under weighing. The observed data strongly suggest that the second scale is biased.

Example: Bias in a Measurement System

Although there is strong evidence against H_0 for the second scale, the degree of bias in its measurements is not necessarily large enough to be of practical concern.

In fact, a 95% confidence interval for the mean μ is given by

$$\bar{y} \pm 2.2622 s/\sqrt{10} = 0.981 \pm 0.012$$

$$[0.969, 0.993]$$

where $P(T \leq 2.2622) = 0.975$ and $T \sim t(9)$.

Example: Bias in a Measurement System

Evidently the second scale consistently understates the weight but the bias in measuring the 1 kg weight is likely fairly small (about 1% - 3%).

Is this bias of practical significance?

Statistical Significance versus Practical Significance

Although we might be able to find evidence against a given hypothesis, this does not mean that the difference is of practical significance.

For example a person who is willing to toss a particular coin one million times can almost certainly find evidence against H_0: P(heads) = 0.5.

Statistical Significance versus Practical Significance

Similarly, if we collect large amounts of financial data, it is quite easy to find evidence against H_0: stock index returns are Normally distributed.

Nevertheless for smaller amounts of data, the Normality assumption is usually made and considered useful.

p-values and Confidence Intervals

If the evidence against H_0 is statistically significant, the size of the *p*-value DOES NOT imply how "wrong" H_0 is.

A confidence interval however does indicate the magnitude and direction of the departure from H_0.

If strong evidence against H_0 is found in a particular direction then this might suggest conducting further experiments to investigate this evidence.

Relationship Between Tests of Hypothesis and Confidence Intervals

Suppose we test H_0: $\mu = \mu_0$ for G(μ,σ) data. Then

$$p-value \geq 0.05$$

$$\text{iff } P\left(\frac{|\bar{Y}-\mu_0|}{S/\sqrt{n}} \geq \frac{|\bar{y}-\mu_0|}{s/\sqrt{n}}; H_0 \text{ is true}\right) \geq 0.05$$

$$\text{iff } P\left(|T| \geq \frac{|\bar{y}-\mu_0|}{s/\sqrt{n}}\right) \geq 0.05 \text{ where } T \sim t(n-1)$$

$$\text{iff } P\left(|T| \leq \frac{|\bar{y}-\mu_0|}{s/\sqrt{n}}\right) \leq 0.95$$

$$\text{iff } \frac{|\bar{y}-\mu_0|}{s/\sqrt{n}} \leq a \text{ where } P(|T| \leq a) = 0.95$$

$$\text{iff } \mu_0 \in \left[\bar{y} - as/\sqrt{n}, \bar{y} + as/\sqrt{n}\right]$$

Which is a 95% confidence interval for μ.

Relationship Between Tests of Hypothesis and Confidence Intervals

In other words:

the p-value for testing H_0: $\mu = \mu_0$ is greater than or equal to 0.05

if and only if

the value $\mu = \mu_0$ is inside a 95% confidence interval for μ

(assuming we use the same pivotal quantity).

Relationship Between Tests of Hypothesis and Confidence Intervals

More generally, suppose we use the same pivotal quantity to construct a confidence interval for a parameter θ and a test of the hypothesis H_0: $\theta = \theta_0$.

The parameter value $\theta = \theta_0$ is inside a 100q% confidence interval for θ if and only if the p-value for testing H_0: $\theta = \theta_0$ is greater than 1- q.

Relationship Between Tests of Hypothesis and Confidence Intervals

For the weigh scale example a 95% confidence interval for the mean μ for the second scale was [0.969, 0.993].

Since $\mu = 1$ is not in this interval we know that the *p*-value for testing H_0: $\mu = 1$ would be less than 0.05.

In fact we showed the *p*-value equals 0.0064 which is indeed less than 0.05.