Chapter 22 & 23

What Is a Test of Significance?

Use and Abuse of Statistical Inference

What is a hypothesis?

- Testable
- Falsifiable

Using Data to Make Decisions

- **Examining Confidence Intervals.**
- **Hypothesis Tests:**
  - **Is the sample data statistically significant, or could it have happened by chance?**

Steps for Testing Hypotheses

- Determine the **null** hypothesis and the **alternative** hypothesis.
- Collect **data** and summarize with a single number called a **test statistic**.
- Determine how **unlikely** test statistic would be ***if null hypothesis were true***.
- Make a **decision**.

7

Parental Discipline

Brown, C. S., (1994) "To spank or not to spank." *USA Weekend*, April 22-24, pp. 4-7.

What are parents' attitudes and practices on discipline?

18

Parental Discipline

- Nationwide random telephone survey of 1,250 adults.
  - 474 respondents had children under 18 living at home
  - results on behavior based on the smaller sample
- reported margin of error
  - 3% for the full sample
  - 5% for the smaller sample

19

Parental Discipline

"The 1994 survey marks the first time a majority of parents reported *not* having physically disciplined their children in the previous year. Figures over the past six years show a steady decline in physical punishment, from a peak of 64 percent in 1988."

- The 1994 sample proportion who ***did not*** spank or hit was 51%!
- *Is this evidence that a majority of the population did not spank or hit?*

20

The Null Hypothesis: H_0

- population parameter equals some value
- status quo
- no relationship
- no change
- no difference in two groups
- etc.
- **When performing a hypothesis test, we assume that the null hypothesis is true until we have sufficient evidence against it**

21

The Alternative Hypothesis: H_a

- population parameter differs from some value
- not status quo
- relationship exists
- a change occurred
- two groups are different
- etc.

22

The Hypotheses for Proportions

- Null: H_0: $p = p_0$
- One-sided alternatives

  H_a: $p > p_0$

  H_a: $p < p_0$
- Two-sided alternative

  H_a: $p \neq p_0$

23

The Hypotheses

- Null: The proportion of parents who physically disciplined their children in the previous year is the same as the proportion [p] of parents who *did not* physically discipline their children. [H_0: $p = .5$]
- Alt: A majority of parents *did not* physically discipline their children in the previous year. [H_a: $p > .5$]

24

Sampling Distribution for Proportions

If numerous simple random samples of size *n* are taken, the sample proportions $(\hat{p})$ from the various samples will have an ***approximately normal*** distribution with ***mean*** equal to *p* (the population proportion) and ***standard deviation*** equal to

$$\sqrt{\frac{p(1-p)}{n}}$$

Since we assume the null hypothesis is true, we replace *p* with p_0 to complete the test.

25

Test Statistic for Proportions

To determine if the observed proportion $\hat{p}$ is unlikely to have occurred under the assumption that H_0 is true, we must first convert the observed value to a standardized score:

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

26

Test Statistic

Based on the sample:

- n=474 (large, so proportions follow normal distribution)
- no physical discipline: 51%
  - $\hat{p} = 0.51$
  - standard error of $\hat{p}$: $\sqrt{.50(1-.50)/474} = 0.023$
  (where ***.50*** is p_0 from the null hypothesis)
- standardized score (test statistic)
  z = (0.51 - 0.50) / 0.023 = 0.43

27

P-value

- *The* ***<u>P-value</u>*** *is the probability of observing data this extreme or more so in a sample of this size, assuming that the null hypothesis is true.*
- A small *P*-value indicates that the observed data (or relationship) is unlikely to have occurred if the null hypothesis were actually true
  - The *P*-value tends to be small when there is evidence in the data against the null hypothesis
  - The *P*-value is **NOT** the probability that the null hypothesis is true

28

P-value for Testing Proportions

- **H_a: $p > p_0$**
- When the alternative hypothesis includes a greater than "**>**" symbol, the *P*-value is the probability of getting a value as large or larger than the observed test statistic (z) value.
  - **The area in the right tail of the bell curve**

29

P-value for Testing Proportions

- **H_a: $p < p_0$**
- When the alternative hypothesis includes a less than "**<**" symbol, the *P*-value is the probability of getting a value as small or smaller than the observed test statistic (z) value.
  - **The area in the left tail of the bell curve** (the same as the percentile value)

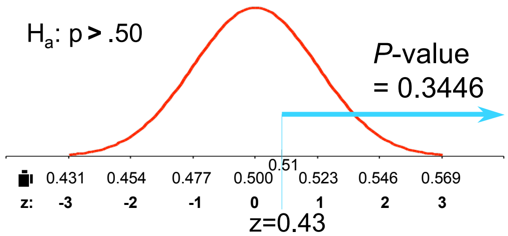
30

P-value for Testing Proportions

- **H_a: $p \neq p_0$**
- When the alternative hypothesis includes a not equal to "≠" symbol, the *P*-value is twice as large as a one-sided test (the sign of z could go either way, we just believe there is a difference).
  - **The area in both tails of the bell curve**
  - **double** the area in one tail (symmetry)

31

P-value



From Table B, z=0.4 is the 65.54th percentile.

32

Decision

- If we think the ***P-value is too low*** to believe the observed test statistic is obtained by chance only, then we would reject chance (***reject the null hypothesis***) and conclude that a statistically significant relationship exists (accept the alternative hypothesis).
- Otherwise, we fail to reject chance and ***do not reject the null hypothesis*** of no relationship (result not statistically significant).

33

Typical Cut-off for the *P*-value

- Commonly, P-values less than 0.05 are considered to be small enough to reject chance (reject the null hypothesis).
- Some researchers use 0.10 or 0.01 as the cut-off instead of 0.05.
- This "cut-off" value is typically referred to as the significance level (alpha) of the test.

34

Decision

- Since the *P*-value (.3446) is *not* small, we cannot reject chance as the reason for the difference between the observed proportion (0.51) and the (null) hypothesized proportion (0.50).
- We do not find the result to be statistically significant.
- We fail to reject the null hypothesis. It is plausible that there was not a majority (over 50%) of parents who *refrained* from using physical discipline.

35

The Two Types of Errors in Testing

	True State of Nature	
Decision Made	**Innocent, Healthy Null Hypothesis**	**Guilty, Diseased Alternative Hypothesis**
Not guilty Healthy Don't reject null hypothesis	Correct ☺	Undeserved freedom False negative Type 2 error
Guilty Diseased Accept alternative hypothesis	Undeserved punishment False positive Type 1 error	Correct ☺

- **Type 1 error** can only be made if the null hypothesis is actually true.
- **Type 2 error** can only be made if the alternative hypothesis is actually true.

40

Decision Errors: Type I

- If we *decide* there is a relationship in the population (reject null hypothesis)
  - This is an incorrect decision only if the null hypothesis is true.
  - The probability of this incorrect decision is equal to the cut-off (α) for the *P*-value.
- If the *null hypothesis is true* and the cut-off is 0.05
  - There really is no relationship and the extremity of the test statistic is due to chance.
  - About 5% of all samples from this population will lead us to wrongly reject chance.

41

Decision Errors: Type II

- If we *decide* not to reject chance and thus allow for the plausibility of the null hypothesis
  - This is an incorrect decision only if the alternative hypothesis is true.
  - The probability of this incorrect decision depends on
    - the magnitude of the true relationship,
    - the sample size,
    - the cut-off for the *P*-value.

42

Power of a Test

- This is the probability that the sample we collect will lead us to reject the null hypothesis when the alternative hypothesis is true.
- The power is larger for larger departures of the alternative hypothesis from the null hypothesis (magnitude of difference).
- The power may be increased by increasing the sample size.

43

Test for means

One of the conclusions made by researchers from a study comparing the amount of bacteria in carpeted and uncarpeted rooms was, “The average difference [in mean bacteria colonies per cubic foot] was 3.48 colonies (95% CI: -2.72, 9.68; *P-value* = 0.29).”

What are the null and alternative hypotheses being tested here?

Is there a statistically significant difference between the means of the two groups?

45

The Hypotheses for a Single Mean

- Null: H_0: $\mu = \mu_0$
- One-sided alternatives
  - H_a: $\mu > \mu_0$
  - H_a: $\mu < \mu_0$
- Two-sided alternative
  - H_a: $\mu \neq \mu_0$

47

The Hypotheses for a Difference in Two Means

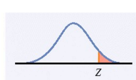
- Null: H_0: $\mu_{diff} = \mu_{diff,0}$ (usually = 0)
- One-sided alternatives
  - H_a: $\mu_{diff} > \mu_{diff,0}$
  - H_a: $\mu_{diff} < \mu_{diff,0}$
- Two-sided alternative
  - H_a: $\mu_{diff} \neq \mu_{diff,0}$

48

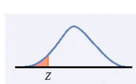
P-value in one-sided and two-sided tests

One-sided (one-tailed) test

H_a: $\mu > \mu_0$ is $P(Z \geq z)$

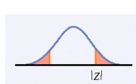


H_a: $\mu < \mu_0$ is $P(Z \leq z)$



Two-sided (two-tailed) test

H_a: $\mu \neq \mu_0$ is $2P(Z \geq |z|)$



49

Warnings about Reports on Hypothesis Tests: *Data Origins*

For any statistical analysis to be valid, the data must come from proper samples. Complex formulas and techniques cannot fix bad (biased) data. In addition, be sure to use an analysis that is appropriate for the type of data collected.

53

Warnings about Reports on Hypothesis Tests: *P-value or C.I.?*

P-values provide information as to whether findings are more than just good luck, but *P*-values alone may be misleading or leave out valuable information (as seen later in this chapter). Confidence intervals provide both the estimated values of important parameters and how uncertain the estimates are.

54

Warnings about Reports on Hypothesis Tests: *Significance*

If the word *significant* is used to try to convince you that there is an important effect or relationship, determine if the word is being used in the usual sense or in the statistical sense only.

55

Warnings about Reports on Hypothesis Tests: *Large Sample*

If a study is based on a very large sample size, relationships found to be statistically significant *may not* have much practical importance.

56

Warnings about Reports on Hypothesis Tests: *Small Sample*

If you read "no difference" or "no relationship" has been found in a study, try to determine the sample size used. Unless the sample size was large, remember that it could be that there is indeed an important relationship in the population, but that not enough data were collected to detect it. In other words, the test could have had very *low power*.

57

Warnings about Reports on Hypothesis Tests: *1- or 2- Sided*

Try to determine whether the test was one-sided or two-sided. If a test is one-sided, and details are not reported, you could be misled into thinking there was no difference, when in fact there was one in the direction opposite to that hypothesized.

58

Warnings about Reports on Hypothesis Tests: *Only Significant are Reported?*

Sometimes researchers will perform a multitude of tests, and the reports will focus on those that achieved statistical significance. Remember that if nothing interesting is happening and all of the null hypotheses tested are true, then [about] 1 in 20 (.05) tests should achieve statistical significance just by chance. Beware of reports where it is evident that many tests were conducted, but where results of only one or two are presented as "significant."

59

Key concepts

- Steps of Hypothesis Testing
- P-values and Statistical Significance
- Decision Errors
- Statistically significant vs practically important
- Large/Small Samples and Statistical Significance
- Multiple Tests and Statistical Significance

74