AP Statistics: Guided Notes Chapter 11
11.1 Chi-square tests
Read 678–681

What is a one-way table? What is a chi-square test for goodness-of-fit?

What are the null and alternative hypotheses for a chi-square goodness-of-fit test?

How do you calculate the expected counts for a chi-square goodness-of-fit test? Should you round these to the nearest integer?

Don't round the expected counts.

What is the chi-square test statistic? Is it on the formula sheet? What does it measure?

Must use counts, not proportions!

HW page 692 (1–4)

Read 682–687

In a goodness-of-fit test, when does the chi-square test statistic follow a chi-square distribution? How do you calculate the degrees of freedom for a chi-square goodness-of-fit test?

Describe the shape, center, and spread of the chi-square distributions. How are these based on the degrees of freedom?

How do you calculate *p*-values using chi-square distributions?

What are the conditions for conducting a chi-square goodness-of-fit test?

Do following example before GOF test on calculator

Alternate Example: *Landline surveys*

According to the 2000 census, of all U.S. residents aged 20 and older, 19.1% are in their 20s, 21.5% are in their 30s, 21.1% are in their 40s, 15.5% are in their 50s, and 22.8% are 60 and older. The table below shows the age distribution for a sample of U.S. residents aged 20 and older. Members of the sample were chosen by randomly dialing landline telephone numbers. Do these data provide convincing evidence that the age distribution of people who answer landline telephone surveys is not the same as the age distribution of all U.S. residents?

Category	Count
20–29	141
30–39	186
40–49	224
50–59	211
60+	286
Total	1048

State: *We want to perform a test of the following hypotheses using* α *= 0.05:*
H_0*: The age distribution of people who answer landline telephone surveys is the same as the age distribution of all U.S. residents. <NOT ABOUT THE SAMPLE!>*
H_a*: The age distribution of people who answer landline telephone surveys is not the same as the age distribution of all U.S. residents.*

Plan: *If conditions are met, we will perform a chi-square goodness-of-fit test.*

- *Random: The data came from a random sample of U.S. residents who answer landline telephone surveys.*
- *Large Sample Size: The expected counts are 1048(0.191) = 200.2, 1048(0.215) = 225.3, 1048(0.211) = 221.1, 1048(0.155) = 162.4, 1048(0.228) = 238.9. All expected counts are at least 5.*
- *Independent: Because we are sampling without replacement, there must be at least 10(1048) = 10,480 U.S. residents who answer landline telephone surveys. This is reasonable to assume.*

Do:

- *Test statistic:* $\chi^2 = \dfrac{(141-200.2)^2}{200.2} + \cdots = 48.2$
- *P-value: Using 5 − 1 = 4 degrees of freedom, P-value =* χ^2 *`cdf(48.2,1000,4)` ≈ 0.*

Conclude: *Because the P-value is less than* α *= 0.05, we reject* H_0*. We have convincing evidence that the age distribution of people who answer landline telephone surveys is not the same as the age distribution of all U.S. residents.*

Read 687–690

Can you use your calculator to conduct a chi-square goodness-of-fit test?

When should you do a follow-up analysis? How do you do a follow-up analysis?

Alternate Example: *Birthdays and hockey*
In his book *Outliers,* Malcolm Gladwell suggests that a hockey player's birth month has a big influence on his chance to make it to the highest levels of the game. Specifically, since January 1 is the cutoff date for youth leagues in Canada (where many National Hockey League players come from), players born in January will be competing against players up to 12 months younger. The older players tend to be bigger, stronger, and more coordinated and hence get more playing time and more coaching and have a better chance of being successful. To see if birth date is related to success (judged by whether a player makes it into the NHL), a random sample of 80 NHL players from the 2009–2010 season was selected and their birthdays were recorded. Overall, 32 were born in the first quarter of the year, 20 in the second quarter, 16 in the third quarter, and 12 in the fourth quarter.
(a) Do these data provide convincing evidence that the birthdays of NHL players are not uniformly distributed throughout the year?
(b) If the results are significant, do a follow-up analysis.

HW page 692 (9, 11, 13, 17)