Introductory Statistics Lectures

Review for Test II

Random variables, probability densities, confidence intervals, hypothesis testing

ANTHONY TANBAKUCHI
DEPARTMENT OF MATHEMATICS
PIMA COMMUNITY COLLEGE

(Compile date: Tue May 19 14:51:46 2009)

Contents

1 Review for Test II

1.1 Key concepts

This review is not fully inclusive.

Be able to differentiate:
- simple random sample, random sample
- qualitative variable, quantitative variable
- discrete variable, continuous variable
- parameter, statistic
- biased statistic, unbiased statistic
- sampling error, non-sampling error
- population distribution, sampling distribution
- distribution function, density function, cumulative density function, inverse cumulative density function
- point estimate, confidence interval

Be able to answer questions such as:
- What is the easy way to find "the probability of at least one"?
- What is a random variable?
- What is the binomial distribution used for? What are the requirements? What does it look like? How do you find probability with it? What is its mean and standard deviation?

1

- What is the normal distribution used for? What does it look like? How do you find probability with it?
- What is a z score. What is μ_z and σ_z equal to? What is the standard normal distribution?
- What does the Central Limit Theorem state? What are the requirements? Why is it useful?
- What does a sampling distribution represent?
- If you increase sample size n, would you expect the variance in the sampling distribution to increase or decrease?
- What do confidence intervals represent? Why are they useful?
- What is hypothesis testing?

In terms of hypothesis testing:
- Know all eight steps.
- Know the requirements for the tests.
- What is H_0 and H_a?
- What do we assume is true?
- Do we use the sampling distribution or population distribution to find the p-value?
- What does the p-value represent?
- How do you find the p-value if you have the test statistic?
- What are the two types of errors? What do they represent.
- What is power? Is it better to have higher or lower power?
- If you reject H_0, what is the probability you made the wrong decision?
- Why do we say a hypothesis test does not prove a hypothesis? How does proof and statistical evidence differ?

1.2   Examples

Given the following density function on the left and it's corresponding CDF for the χ^2 distribution, answer the following questions.

Chi−Squared Density

Chi−Squared CDF
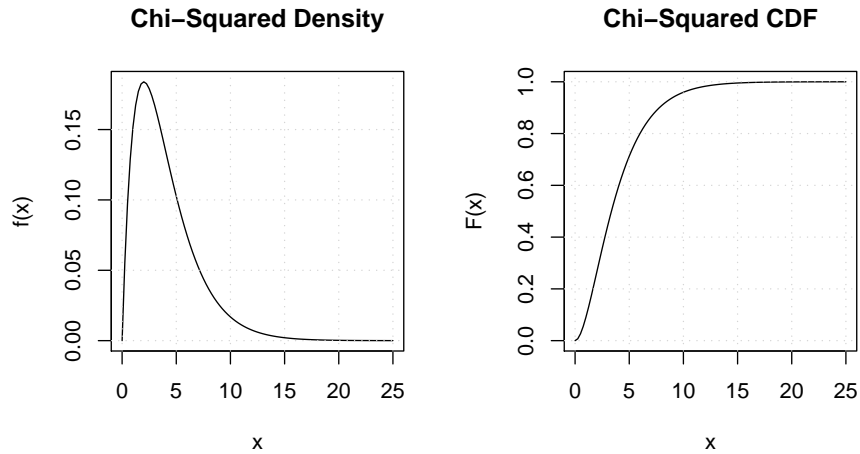
Question 1. Find $P(x > 5)$

Question 2. Find P_{25}

Question 3. What percent of data lies within ± 1.5 standard deviations on a normal distribution? (Check: 0.866)

Question 4. What is the probability that a student who randomly guesses on a 10 choice true/false exam will get at least 1 correct? (Check: 0.999)

Question 5. What is the probability that a student who randomly guesses on a 50 choice true/false exam pass the exam ($70\% = 35$ or more correct)? (Check: 0.0036 using normal approx. Using exact: 0.0033)

If a researcher is conduction a 1-sample proportion hypothesis test with the hypothesis $H_a : p > 0.7$. The study finds $x = 78$ and $n = 100$.

Question 6. What is the test statistic? (Check:1.75)

Question 7. What is the p-value? (Check:0.0404)

Question 8. What would the p-value have been if $H_a : p \neq 0.7$

A manufacturer of paper used for packaging requires a minimum strength of 20 lb/in^2. A quality control inspector randomly samples 35 pieces of paper from the previous hour's production and tests them in a machine the measures the force at which the paper breaks. The standard deviation σ of the strength measurements, computed over many sample, is 2 lb/in^2.

Question 9. What is the probability distribution of the sample mean strength?

Question 10. What is the expected average variation for $\bar{x}$? (Check: 0.338 lb/in^2)

Question 11. If one piece of paper is tested, what is the probability that its strength is at least 21 lb/in^2? (Assume $\mu = 20$ lb/in^2, and the individual values have a normal distribution.) (Check: 0.309)

Question 12. If 35 pieces of paper are tested, what is the probability that their mean strength is at least 21 lb/in^2? (Assume $\mu = 20$ lb/in^2) (Check: 0.00155)

Question 13. The mean strength of the paper from the sample is 18.9 lb/in^2. Based on the sample data, construct a 98% confidence interval for the true mean strength. (Check: $z_{\alpha/2} = 2.33$, $E = 0.786$ lb/in^2)

Question 14. The mean strength of the paper from the sample is 18.9 lb/in^2. Conduct a hypothesis test at the 1% significance level to check the quality

control inspector's concern that the strength is too low. (Check: $z = -3.25$, p-value= 0.000569)

Question 15. The manufacturer changed the process to increase the strength. If the quality control engineer wants to estimate the new strength to within 0.25 lb/in², what sample size should be used? (Check: $n = 246$)

The manufacturer changed the process to increase the strength. However, your boss is a real cheapskate, and he thinks your recommended sample size is too expensive! A new sample of 5 pieces of paper is measured (in lb/in²):

20.4, 22.1, 23.3, 25.6, 23.2

Question 16. Since the process is different, assume σ unknown. Test the hypothesis $H_a : \mu > 20$ using the sample data. Does the process appear to be ok?

A researcher who is trying to determine the proportion of people who support increasing the tax on gas guzzlers. Use the output below to answer the following question.

```
R:  prop.test(87, 200, p = 0.5, alternative = "less")
        1-sample proportions test with continuity correction

data:    87 out of 200, null probability 0.5
X-squared = 3.125, df = 1, p-value = 0.03855
alternative hypothesis: true p is less than 0.5
95 percent confidence interval:
 0.00000 0.49565
sample estimates:
    p
0.435
```

Question 17. What type of hypothesis test is being conducted?

Question 18. What was the study size and number of successes?

Question 19. What is are the null and alternative hypothesis?

Question 20. What is the formal decision (assume $\alpha = 0.05$)

Question 21. What is the conclusion?

Question 22. What is the best point estimate for p?