Quiz 3 Review

171:161 Spring 2014

Distribution Summaries

- Binomial: $\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$
  - Used to model “k” successes out of “n” binary trials
  - Has a mean of $n * p$ (number of trials times probability of success)
  - Standard Deviation of $\sqrt{n * p * (1-p)}$
  - Note that when using the normal approximation the following formula is often used:

$$SE = \sqrt{\frac{p_0 * (1-p_0)}{n}}$$

Normal Distribution and Central Limit Theorem

- Central limit theorem tells us that the distribution of sample means will be Normal with a mean equal to the population mean and a standard error equal to the population standard deviation divided by the square root of n
- In practice we do not know the population standard deviation, so t-tests are more appropriate for inference
- Because of this, the Normal Distribution is most often used in estimating probabilities/approximating the binomial and not for hypothesis testing.

Student’s t-distribution

- Bell shaped, symmetric and centered around zero.
- Heavier tails than the normal distribution.
- Has 1 parameter (degrees of freedom)
- As degrees of freedom increase the tails become thinner and the distribution looks more normal.
- When sample size is small, the t-distribution performs much better than the normal distribution.

How do I know which distribution to use??

Binomial problems will typically ask for the probability of k successes out of n trials or an expected number of successes. For example:

- “What is the probability of exactly 4 out of 10 individuals having a positive test?”
- “If 8 cards are randomly drawn (with replacement) how many do you expect to be spades?”
- “What are the chances that at least one confidence interval out of 20 does not contain the true mean?”

Normal distribution problems will focus on a single probability. For example:

- “What is the probability an infant weighs less than 3000 grams?”
- “What is the probability that a woman is over 5’6’’?”
- “What probability that a sample of 10 students has an average IQ over 110?”

Example #1:

Height for US males is normally distributed with a mean of 70 inches and standard deviation of 3 inches.

A) What is the probability that randomly selected male is over 6’6’’?

B) What is the probability that at least 1 male out of a group of 5 is over 6’6’’?

C) What is the probability that a sample of 5 males has an average height of over 6’6’’?

A) $Z = (76 – 70)/3 = 2$ $P(Z > 2) = 0.023$

B) Binomial … $P = 1 – \text{Prob}(\text{Zero males} > 6'6'') = 1 – (1 - 0.023)^5 = 0.11$

C) $Z = (76 – 70)/(3/\text{sqrt}(5)) = 4.47$ $P(Z > 4.47) = .000004$

Example #2:

Dobson et al. [1976], 36 patients with a confirmed diagnosis of phenylketonuria (PKU) were identified and placed on dietary therapy before reaching 121 days of age. The children were tested for IQ (Stanford-Binet test) between the ages of 4 and 6; subsequently, their normal siblings of closest age were also tested with the Stanford-Binet. The following are the first 15 pairs listed in the paper.

Pair	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
IQ of PKU case	89	98	116	67	128	81	96	116	110	90	76	71	100	108	74
IQ of sibling	77	110	94	91	122	94	121	114	88	91	99	93	104	102	82

	PKU	Sibling
Mean	94.66667	98.80000
Standard Deviation	18.54980	13.36413

Part A) Construct a 90% confidence interval for the mean IQ of PKU siblings.

Part B) The average IQ is 100, conduct a hypothesis test to determine whether PKU cases have an IQ different from normal.

Part A) 90% CI: $\bar{x} \pm t_{.95} * SE$

$$94.667 \pm 1.761 \; * (\frac{18.550}{\sqrt{15}})$$

(82.2, 103.1)

Part B) $T_{obs} = \frac{94.667 - 100}{\frac{18}{\sqrt{15}}} = -1.15$

For alpha = 0.05 this t-statistic does not indicate significance, thus we conclude there is no evidence of a difference in IQ from normal.

Example #3

The population of sonar operators has a mean identification rate of 82 targets out of 100. Recently a psychologist developed a new sonar training system that tested on 15 trainees. The group of trainees had a mean id rate of 85.7 and a standard deviation of 5.1. Conduct a hypothesis test to evaluate the effectiveness of the new training scheme.

$$H_0: ID\ Rate = 82$$
$$H_A: ID\ Rate \neq 82$$

$$t_{obs} = \frac{85.7 - 82}{\frac{5.1}{\sqrt{15}}} = 2.8$$

With 14 degrees of freedom a t-statistic of 2.8 will have a two-sided p-value less than 0.02

Thus we can conclude that the trainee group scored significantly higher than average and the training was effective at improving identification rates.

Example #4: Conceptual Questions:

- How does the width of a 90% CI compare to the width of a 95% CI?
- When a population is skewed why can we still use the Normal Distribution to make inferences about the sample mean?
- How does the t-distribution with 10 degrees of freedom compare to the t-distribution with 20 degrees of freedom? How to each compare to the Normal Distribution?
- Why does the t-distribution used instead of the Normal Distribution in most circumstances?

- The 95% CI is wider
- The mean is calculated as a sum of many observations, central limit theorem thus says it will be normally distributed if there are sufficient observations included.
- The t-distribution with fewer degrees of freedom will have thicker tails, both of these t-distributions will have thicker tails than the normal distribution.
- We are adding variability to our inference by using the sample standard deviation as an estimate of the population standard deviation.