Discussion #5

Review

1. The notion "mean" in three contexts:
   (i) Sample mean is a certain number. After you get a set of samples, $\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ is calculated as the sample mean;
   (ii) Population mean μ, a mean of entire population, usually unknown. We use sample mean $\bar{x}$ to estimate it;
   (iii) Random variable $\bar{X}$: Suppose you decide to get a sample of size n from the population. Before your experiment, you know you will get n random variables from the population, and their average $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is still a random variable. When you get different samples, $\bar{X}$ may change. Any certain sample mean is a realization of the random variable $\bar{X}$.

2. Steps for a significant test:

   (a) **Formulate the null(H_0) and alternative(H_A) hypothesis.**

      i. H_0 is position that we wish to support unless there is strong evidence against it – standard, known from before, established value.

      ii. H_A is challenging assertion or new idea. Usually two-sided is prefered unless there is a strong reason to use one-sided.

   (b) **Decide the parameter we want to test and the distribution for the data.**

      i. Normal distribution

         A. if σ^2 is known, interested in μ, then use

         $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

         where μ is the hypothesized value in H_0. Z is exactly N(0,1).

         B. if σ unknown, then use

         $$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

         where μ is the hypothesized value in H_0. T is the T-distribution with degree of freedom $n - 1$.

      ii. Unknown any distribution, interested in μ, n is large ($n \geq 30$),

         A. if σ known, then use

         $$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

         where μ is the hypothesized value in H_0. Z is approximately N(0,1)

         B. if σ unknown, then use

         $$Z = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

         where μ is the hypothesized value in H_0. Z is approximately N(0,1) (double approximation)

      iii. Binomial data, interested in p

         A. if n is large($np \geq 5, nq \geq 5$), then use

         $$Z = \frac{X_{NA} - \mu_X}{\sigma_X} \quad \text{where} \quad \mu_X = np \text{ and } \sigma_X^2 = npq$$

         where p is the hypothesized value in H_0. Z is approximately N(0,1)

         B. if n is small, use binomial formula directly. see pp146-147.

(c) **Calculate p-value.**
p-value is the probability of observing an event as extreme or more extreme than what we observed. This is calculated assuming that H_0 is true and using H_A to determine what kinds of data constitute "extreme" data.

(d) **Interpretation**
small p-value $\longrightarrow$ evidence against H_0
large p-value $\longrightarrow$ evidence for H_0
In hypothesis testing, if we use 5% significance level
p-value $\leq 0.05 \longrightarrow$ reject H_0
p-value $> 0.05 \longrightarrow$ accept H_0
You can also see pp140 for the guidelines of interpretation.

Practice Problem

1. Consider taking a random sample from a $N(\mu, \sigma^2)$ distribution with $\sigma = 6$. Consider testing hypothesis H_0: $\mu = 35$ versus H_A: $\mu \neq 35$.

   (a) Suppose that a random sample of size 9 is taken and that $\bar{x} = 32$. What is the p-value for your hypothesis test? Are the results significant at 5%?

   (b) Suppose that a random sample of size 49 is taken and that $\bar{x} = 32$. What is the p-value for your hypothesis test? Are the results significant at 5%?

   (c) Suppose that $\sigma = 3$ instead of $\sigma = 6$. Suppose that a random sample of size 9 is taken and that $\bar{x} = 32$. What is the p-value for your hypothesis test? Are the results significant at 5%?

2. A particular strain of bacteria is used for nitrogen fixation on a certain variety of alfafa. The nitrogen fixation is known to follow (approximately) a normal distribution. A scientist claims the mean amount of nitrogen fixed in a plant is 26.7 mg. Data are available on a random sample of 12 plants:

   23.9, 26.2, 27.9, 22.2, 24.4, 25.8, 25.6, 29.1, 26.6, 26.0, 24.9,23.3

   State symbolically the null and alternative hypothesis. Find the p-value for the test of the claim. Are the results significant at $\alpha = 10\%$? at $\alpha = 5\%$? at $\alpha = 1\%$?

3. A five-year-old census recorded that 20% of the families in a large community lived below the poverty level. To determine if this percentage has changed, a random sample of 400 families is studied and 70 are found to be living below the poverty level. Does this finding indicate that the current percentage of families earning incomes below the poverty level has changed from what it was five years ago?

4. My cat Felix likes to hunt mice. I claim that each day his probability for hunting success (catching some mice) is 0.6. Assume that his hunting success is independent from day to day. I observe him carefully for 10 days dand found that he has hunting success 9 days. Perform a test of my claim.