

6B Hypothesis testing

This is the last set of exercises. There are no more home problems.

Class problems

6B1 (Nail factory) A factory is producing nails, with the aim that the average length should be 10.00 cm. Because of imprecisions in the manufacturing process, each produced nail has a slightly random length, according to a normal distribution with an unknown mean μ and an unknown standard deviation σ . For quality control, a sample of 120 nails were inspected. In the sample, the average was 10.08 cm and the sample standard deviation was 0.40 cm.

Because of the large sample size, we work as if the standard deviation of the nail-generating process is $\sigma = 0.40$ (i.e. equals the sample standard deviation). This is Ross's "known variance" situation (§7.3 and 8.3.1).

- (a) Compute a 95% confidence interval for the unknown mean.
- (b) Compute a 99% confidence interval for the unknown mean.
- (c) Calculate the p -value for a hypothesis test with the null hypothesis $\mu = 10.00$, against the alternative hypothesis that $\mu \neq 10.00$.
- (d) Using the p -value from (b), report whether the null hypothesis is accepted or rejected at significance level $\alpha = 0.05$.
- (e) Using the p -value from (b), report whether the null hypothesis is accepted or rejected at significance level $\alpha = 0.01$.
- (f) (Optional) Redo (a)–(e), but now take into account the uncertainty about σ (see Ross's "unknown variance" situation §7.3.1 and 8.3.2). That is, instead of using the standard normal distribution, use Student's t distribution with parameter $n - 1 = 119$. Compare the numerical results to the previous ones.

The CDF and its inverse for the t distribution can be computed in R with `pt` and `qt`, and in Matlab with `tcdf` and `tinv`.

Solution.

- (a) At 95% confidence level, the interval is

$$m(\vec{x}) \pm z \frac{\sigma}{\sqrt{n}} = 10.08 \pm 1.96 \frac{0.40}{\sqrt{120}} \approx 10.08 \pm 0.07 = [10.01, 10.15],$$

where $z = 1.96$ is the number such that a standard normal distributed random variable has probability 0.025 of exceeding z , that is, $\Phi(1.96) = 0.975$, and also probability 0.025 of being smaller than $-z$, that is, $\Phi(-1.96) = 0.025$. (Recall exercise 5A.) Note that the confidence interval does not contain 10.00.

- (b) At 99% confidence level, we apply $z = 2.58$ because $\Phi(2.58) \approx 0.995$. Now we have the interval

$$m(\vec{x}) \pm z \frac{\sigma}{\sqrt{n}} = 10.08 \pm 2.58 \frac{0.40}{\sqrt{120}} \approx \mathbf{10.08 \pm 0.09} = \mathbf{[9.99, 10.17]}.$$

Note that the confidence interval contains 10.00.

- (c) The null hypothesis is $H_0 : \mu = 10.00$, and the alternative hypothesis is $H_1 : \mu \neq 10.00$. We first compute the test statistic

$$t(\vec{x}) = \frac{m(\vec{x}) - \mu}{\sigma/\sqrt{n}} = \frac{10.08 - 10.00}{0.40/\sqrt{120}} \approx 2.19.$$

If the null hypothesis is true, then the test statistic $t(\vec{X})$ follows the standard normal distribution. The p -value is the probability (if the null is true) that $|t(\vec{X})| \geq 2.19$, that is, the probability that we would observe a test statistic “as far out” as we actually observed, or even further.

Because our test statistic is positive, we compute first the *right* tail probability

$$P(t(\vec{X}) \geq 2.19) = 1 - \Phi(2.19) \approx 0.0142.$$

Then (by symmetry) the left and right tails together have probability

$$P(|t(\vec{X})| \geq 2.19) = 2 \cdot (1 - \Phi(2.19)) \approx \mathbf{0.0285}.$$

This is the p -value.

- (d) Because the p -value is $0.0285 < 0.05$, we **reject** the null hypothesis. That is, at significance level $\alpha = 0.05$, we consider the observed sample mean 10.08 to be “too far out” from 10.00, or “too surprising”, to be compatible with the null hypothesis.
- (e) Because the p -value is $0.0285 > 0.01$, we **accept** the null hypothesis. That is, at significance level $\alpha = 0.01$, we consider the observed sample mean 10.08 to be “not far enough” from 10.00, or “not surprising enough”, to reject the null hypothesis.
- (f) For the confidence intervals, we now use the quantiles 0.975 and 0.995 of the t distribution, instead of the standard normal distribution. With R commands `qt(.975, 119)` and `qt(.995, 119)` we get coefficients 1.98 and 2.62, that is, our confidence intervals are about 1...2% wider than in the “known variance” situation.

Similarly, in hypothesis testing, we compute the tail probabilities using the $t(119)$ distribution. Now the right tail has probability $1 - \text{pt}(2.19, 119) \approx 0.0152$, the left tail has probability $\text{pt}(-2.19, 119) \approx 0.0152$ (note that the t distribution is also symmetric about the origin), so the p -value is $\approx 2 \cdot 0.0152 \approx 0.0305$. It is slightly bigger than in (c), that is, our observation is not *quite* as “surprising” but the difference is not big. We still reject the null hypothesis at significance level $\alpha = 0.05$, and accept it at level $\alpha = 0.01$.

The difference between known and unknown variance scenarios becomes large when the sample is small, something like $n < 30$.

6B2 (Peas) Gregor Mendel (1822–1884) laid the foundation for modern genetics by studying how certain features of pea plants are inherited. Some of his experiments led to the theory that when pea plants are cross-bred in a particular way, then 75% of the offspring plants produce yellow peas, and 25% produce green peas.

In order to test this theory, a geneticist cross-breeds 80 plants using the same procedure, and observes that 56 of them produce yellow peas. Is this observation compatible with Mendel's theory? Formulate and perform a (two-sided) hypothesis test. Let the null hypothesis be that each plant has probability $\theta = 0.75$ of producing yellow peas (independently of the other plants). The alternative hypothesis is that $\theta \neq 0.75$. Apply the significance level 5%.

You can use either the exact binomial test (which exists in two versions; see Lecture 6A and Ross 8.6), or the normal approximation.

Note that in a binary model, one might commonly use “ p ” to denote the probability parameter (for a plant to produce yellow peas), but here we use θ to avoid confusion with the p -value, which is a probability associated with the hypothesis test.

Solution. The null hypothesis and the alternative hypothesis are:

$$H_0 : \theta = 0.75, \quad H_1 : \theta \neq 0.75.$$

For the exact binomial test, we can use the *number of yellow-producing plants* X as the test statistic. We have observed $X = 56$.

If the null hypothesis is true, then each plant (independently) produces yellow plants with probability $\theta = 0.75$, so the test statistic X has the binomial distribution with parameters $n = 80$ and $\theta = 0.75$. Then the expected number of yellow-pea plants is $E(X) = n\theta = 60$.

In a two-sided test, we will have *critical regions* on both sides of $E(X)$; both too low and too high values of X lead to rejection.

There are two slightly different methods using the exact binomial distribution, and a third using the normal approximation.

Method 1. Equal distances from mean. This is the method in lecture slides. First measure how far X is from $E(X)$. Here we observe $X - E(X) = 56 - 60 = -4$, so the observed value is 4 units left (smaller) from $E(X)$. Now to obtain the p -value, calculate the probability of X deviating from $E(X)$ *by at least 4 units* to either direction. We do this by looking at the binomial distribution.

$$\begin{aligned} P(|X - E(X)| \geq 4) &= P(X \leq 56 \text{ or } X \geq 64) \\ &= 1 - P(57 \leq X \leq 63) && \text{(complement rule)} \\ &= 1 - \sum_{k=57}^{63} f(k), \end{aligned}$$

where

$$f(k) = \binom{80}{k} 0.75^k 0.25^{80-k}$$

is the density function of $\text{Bin}(80, 0.75)$, that is the distribution of X if the null hypothesis is true. If we do not want to compute the seven densities manually, we can use the R command `sum(dbinom(57:63, 80, 0.75))` which gives 0.634.

So the p -value (= sum of tail probabilities) is $1 - 0.634 = \mathbf{0.366}$. Because $p > 0.05$, we **accept** the null hypothesis (did not find enough evidence against Mendel's theory).

Method 2. Divide tail probabilities equally (half and half). This is the method in Ross section 8.6. Calculate separately the probabilities (if null is true) that $X \geq x$ and that $X \leq x$, where $x = 56$ is the observed value. Reject the null hypothesis if either of these probabilities is $\alpha/2 = 0.025$ or greater.

If the null hypothesis is true, then $X \sim \text{Bin}(80, 0.75)$. Then $P(X \leq 56) = \sum_{k=0}^{56} f(k) = F(56)$, where f and F are the density and CDF of the binomial distribution. The R command `pbinom(56, 80, 0.75)` gives 0.182.

Similarly, $P(X \geq 56) = 1 - P(X < 56) = 1 - F(55) \approx 0.876$, using the R command `pbinom(55, 80, 0.75)`. Note that we were careful about whether 56 is included or not in the interval, because X is a discrete random variable.

Because neither of the numbers 0.182 or 0.876 is below $\alpha/2 = 0.025$, we **accept** the null hypothesis at significance level $\alpha = 0.05$.

Method 3. Approximate the binomial with a normal distribution. We approximate that if the null hypothesis is true, then X follows the *normal* distribution with mean $\mu = n\theta = 60$ and standard deviation $\sigma = \sqrt{n\theta(1-\theta)} \approx 3.87$. Both values were taken from the corresponding values of $\text{Bin}(80, 0.75)$. Then the *normalized* test statistic $Z = (X - \mu)/\sigma = (X - 60)/3.87$ should follow the standard normal distribution.

We observed $x = 56$, which corresponds to the normalized test statistic $(x - 60)/3.87 \approx -1.033$. This is negative, so the left tail probability is $P(Z \leq -1.033) = \Phi(-1.03) \approx 0.1508$, and the (by symmetry) the corresponding right tail probability is $P(Z \geq 1.033) \approx 0.1508$.

So in total, the probability of observing a value that is "at least as far out" as our observed value $x = 56$ (normalized $z = -1.03$) is

$$2 \cdot \Phi(-1.033) \approx \mathbf{0.3016}.$$

This is our p -value, and because it is greater than 0.05, we **accept** the null hypothesis.

As you see, in a binary situation, there are several slightly different procedures for performing a hypothesis test. They all are based on the same underlying idea (reject the hypothesis if the test statistic is too far out). Because the binomial distribution is not symmetric, methods 1 and 2 are slightly different. And method 3 is an approximation. In many cases, especially if n is large, the results are not *too* different.