LECTURE: SAMPLING DISTRIBUTIONS (II)

1. Normally Distributed Populations (continued)

Recall:

Let $Y_1, \ldots, Y_n$ be iid then the **sample mean** is

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

Note that $\bar{Y}$ is a random variable. In general, the distribution of $\overline{Y}$ is unknown, except for an important special case:

Recall:

If all the $Y_i \sim N(\mu, \sigma^2)$ then $\overline{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Example 1: (continued)

A ball bearing machine produces ball bearings whose diameters are normally distributed with mean μ mm and standard deviation $\sigma = 0.1$ mm

(b) How many ball bearings should we sample if we want the sample mean to be within 0.02 mm of μ with probability 0.95?

Since $\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ by the 68-95-99 rule, we want $\overline{Y}$ to lie within 2 standard deviations of the population mean μ

Hence we want the standard deviation of $\overline{Y}$ to be $\frac{0.02}{2} = 0.01$

But since the standard deviation of $\overline{Y}$ is $\frac{\sigma}{\sqrt{n}} = \frac{0.1}{\sqrt{n}}$ we get

$$\begin{aligned} \frac{0.1}{\sqrt{n}} &= 0.01 \\ \sqrt{n} &= \frac{0.1}{0.01} = 10 \\ n &= 100 \end{aligned}$$

(c) How many ball bearings should we sample if we wish the sample mean to be within 0.02 mm of μ with probability 0.98?

Here we actually have to do the math. We want:

$$P(|\bar{Y} - \mu| \leq 0.02) = P(-0.02 \leq (\bar{Y} - \mu) \leq 0.02) = 0.98$$

Converting to the standard normal, we get

$$\begin{aligned} 0.98 &= P\left(\frac{-0.02}{\sigma/\sqrt{n}} \leq \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.02}{\sigma/\sqrt{n}}\right) \\ &= P\left(\frac{-0.02}{0.1/\sqrt{n}} \leq Z \leq \frac{0.02}{0.1/\sqrt{n}}\right) \\ &= P(-0.2\sqrt{n} \leq Z \leq 0.2\sqrt{n}) = 0.98 \end{aligned}$$

By complements and symmetry, we need

$$P(Z \leq -0.2\sqrt{n}) = 0.01$$

Looking at the $Z-$table, we find $P(Z \leq -2.32) = 0.102$ and so

$$\begin{aligned} -0.2\sqrt{n} &= -2.32 \\ \sqrt{n} &= 11.6 \\ n &= 134.56 \end{aligned}$$

Answer: $n = 135$

In other words, we need a sample size of 135 or more to have 98% confidence that our sample mean is within 0.02 of the true population mean.

Note: Ok to choose $P(Z \leq -2.33) = 0.0099$ instead.

2. Chi-Square distribution

We will now briefly discuss two other useful distributions, the chi-square and t distribution. Computations with these distributions is done via tables, and not by using densities.

The chi-square distribution is the distribution of the sum of *squares* of normal random variables. This is useful to make inferences about the variance of a population, since variance involves squares of random variables.

Definition:

Let $Y_1, \dots, Y_n$ be iid $N(\mu, \sigma^2)$

Then $Z_i = (Y_i - \mu)/\sigma$ are iid $N(0,1)$ and

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma}\right)^2$$

Has a **chi-square distribution** with n degrees of freedom (df)

To use the chi-square distribution, we use a chi-square table, which is provided on the course website. Here is how to use it:

Example 2:

Suppose Y has a chi-square distribution with 10 df

(a) Find y such that $P(Y > y) = 0.995$

Look at the 10 df row and the $\chi^2_{0.995}$ column. The entry is precisely y such that $P(Y > y) = 0.995$

Therefore $P(Y > 2.156) = 0.995$ and $y = 2.156$

(b) Find y such that $P(Y \leq y) = 0.9$

By the complement rule, we want $P(Y > y) = 1 - 0.9 = 0.1$

Looking at the $\chi^2_{0.1}$ column, we have $P(Y > 15.987) = 0.1$

Therefore $y = 15.987$

Note: The columns in the table are *extreme* probabilities, i.e. very high or very low. Those are the probabilities we are usually interested in. For other values, you need to use a statistical package such as R.

Example 3:

Let $Z_1, \ldots, Z_6$ be iid $N(0,1)$. Find a value a such that

$$P\left(\sum_{i=1}^{6} Z_i^2 \leq a\right) = 0.95$$

$$\text{We want } P\left(\sum_{i=1}^{6} Z_i^2 > a\right) = 1 - 0.95 = 0.05$$

Since $\sum_{i=1}^{6} Z_i^2$ has a chi-square distribution with 6 df, looking at the chi-square table in the 6 df row and $\chi^2_{.050}$ column, we get $a = 12.592$

Question: How do we use this practically?

Recall:

The **sample variance** is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Then we can say the following about the distribution of S^2:

Fact:

Let $Y_1, \ldots, Y_n$ be iid $N(\mu, \sigma^2)$ then

$$\left(\frac{n-1}{\sigma^2}\right) S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Has a chi-square distribution with $n-1$ deg of freedom

Note: The $n-1$ is not a typo. This is because $\overline{Y}$ depends on $Y_1, \cdots, Y_n$ so the sum above really just depends on $n-1$ independent random variables and not n