Chapter 8 Lecture/Tutorial

1. Definitions

Please give examples of sample statistics and population parameters. Are sample statistics random variables?

Solution:

Sample statistics:

- Sample mean ($\bar{X}$),
- Sample variance (S^2),
- Sample proportion ($\hat{P}$), etc.

All of them are random variables as their values are likely to differ from sample to sample.

Population parameters:

- Population mean (μ),
- Population variance (σ^2),
- Population proportion (p), etc.

2. Definitions

How are the following statistics defined? Which distributions do they follow?

(a) Z-statistic, T-statistic, χ^2-statistic for a single sample

(b) F-statistic for two samples.

(c) When we deal with t, f or χ^2 , is it always necessary that the population is normal?

(d) When we deal with Z-statistic, is it always necessary that the population is normal?

Solution:

(a) First, when $\bar{X}$ follows normal distribution, the random variable Z is defined in the following manner so that Z follows the standard normal distribution. This greatly facilitates probability calculations.

$$Z = \frac{\bar{X} - mean\ of\ \bar{X}}{standard\ deviation\ of\, \bar{X}}$$

We can show that the mean of all possible sample means = population mean, and the standard deviation of sample means is population mean divided by the square root of sample size. Then,

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

[Note: In chapter 9, we will define Z-statistic for sample proportion in the analogous manner. Sample proportion can be viewed as mean when population has the characteristics of a binomial random variable.].

Second, the random variable T (T-statistic) is defined in the following manner

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

This variable is defined to deal with situations when we know that $\bar{X}$ follows normal distribution, but do not know sigma which prevents the use of Z. The random variable T (T-statistic) follows a distribution called T-distribution. For each n value, there is one T distribution. Therefore, we specify a quantity called the degree of freedom Nu = (n-1) to specify a particular T-distribution. T-distribution is symmetric with mean value of zero (See Fig 8.8).

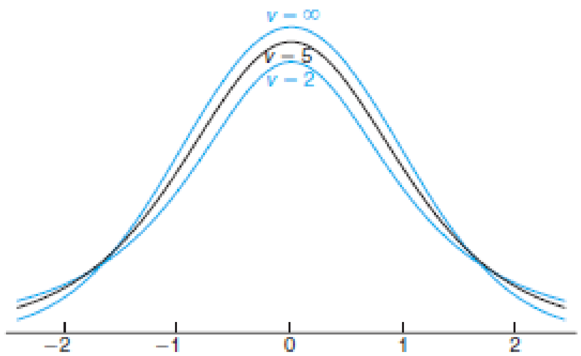


Figure 8.8: The t-distribution curves for $v = 2, 5$, and ∞.

Third, chi-square, χ^2 is defined as the ratio of variance of sample variance to population variance multiplied with (n-1), the associated degree of freedom.

$$\chi 2 = (n-1)\frac{S^2}{\sigma^2}$$

Chi-square can be considered as the square of a random variable X.

(b) Finally, the random variable F is defined for two samples. If the two samples are collected from the same population, then F is defined as the variance of two samples (and is expected to be close to 1).

$$\mathrm{F} = \frac{S_1^2}{S_2^2}$$

In general, F is defined in the following manner.

$$\mathrm{F} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

F-statistic follows F-distribution. F-distribution is not symmetric and depends on the two degrees of freedom (d.f.) (see the two different pdfs shown in Fig. 8.11). The first degree of freedom ($Nu_1 = n_1$-1) arises from the first sample size, and the second degree of freedom ($Nu_2 = n_2$-1) arises from the second sample size.

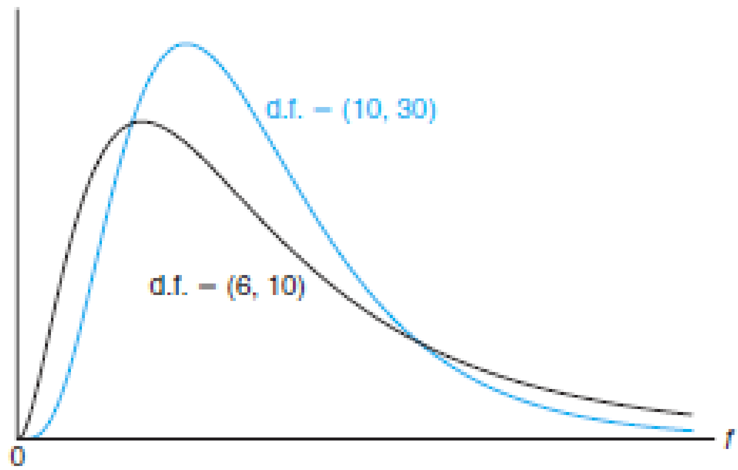


Figure 8.11: Typical F-distributions.

(c) Yes, sample must be drawn from normal population to use T, F, chi-square distributions.
(d) Population need not be always normal for the application of Z-statistic
Case I: If population is normal, then sample mean ($\bar{X}$) is always normal with mean = μ and sample standard deviation = $\sigma/\sqrt{n}$ for any value of n. We can then use Z-statistic.
Case II: However, when population is not normal, we know, from central limit theory, that for large sample size (greater than or equal to 30), sample mean ($\bar{X}$) can be assumed normal and Z-statistic can be used.

3. ***Understanding Notations***: What do f_α represent?
Solution:
f_α represents a specific f-value on the horizontal f-axis that divides the f-axis in such a way that the right portion of the f-axis contain an area equal to alpha above it (see the shaded region in Fig 8.12 below). The left portion of the f-axis then contains the remaining area of 1-alpha.

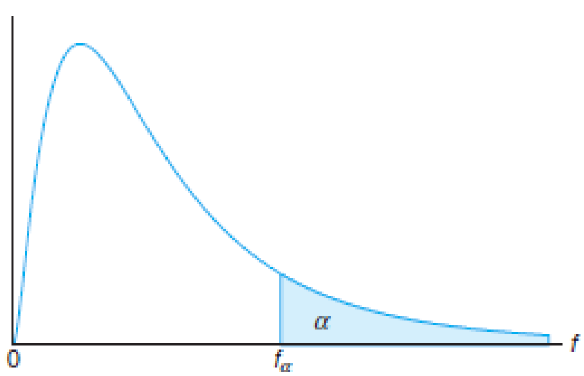


Figure 8.12: Illustration of the f_α for the F-distribution.

4. ***Notations***: What do t_α , $t_{1-\alpha}$, $-t_\alpha$ represent (consider df, the degree of freedom, remains the same)?

Solution:

(i) t_α represents a specific t-value on the horizontal t-axis that divides the t-axis in such a way that the right portion of the t-axis contain an area of alpha above it. The left portion of the t-axis then contains the remaining area of 1-alpha (See fig 8.9 below).

(ii) $t_{1-\alpha}$ represents a specific t-value on the horizontal t-axis that divides the t-axis in such a way that the right portion of the t-axis contain an area of (1-alpha) above it. The left portion of the t-axis then contains the remaining area of alpha.

(iii) $-t_\alpha$ is the negative of t_α . Due to symmetry of t-distribution, $t_{1-\alpha} = -t_\alpha$. See Figure 8.9 below. The points t_α and $-t_\alpha$ are located equidistant to the right and to the left respectively, from the mean value of zero. The area to the right of t_α is alpha. On the other hand, since the area to the left of $-t_\alpha$ is equal to alpha, the area to the right of $-t_\alpha$ is 1-alpha. Consequently, $-t_\alpha$ = $t_{1-\alpha}$ = t-value which has area (1-alpha) to its right.

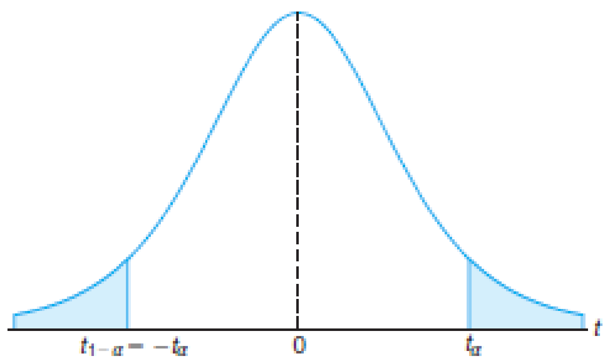


Figure 8.9: Symmetry property (about 0) of the t-distribution.

5. ***Notations***: What do $\chi2_\alpha$, $\chi2_{1-\alpha}$ represent (for same df)?

Solution:

(i) χ^2_α represents a specific chi-square value on the horizontal χ^2-axis that divides the χ^2-axis in such a way that the right portion of the axis contain an area of alpha above it. The left portion of the axis then contains the remaining area of 1-alpha

(ii) $\chi^2_{1-\alpha}$ represents a specific χ^2-value on the horizontal χ^2-axis that divides the axis in such a way that the right portion of the axis contain an area of (1-alpha) above it. The left portion of the axis then contains the remaining area of alpha.

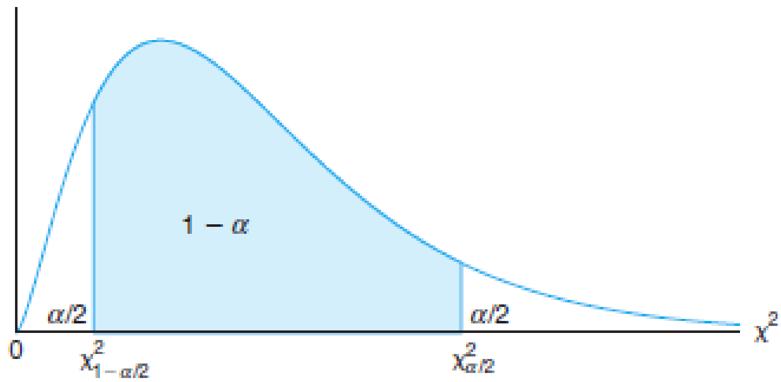
6. ***Notations***: What do $\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}$ represent (for same df)?

Solution:

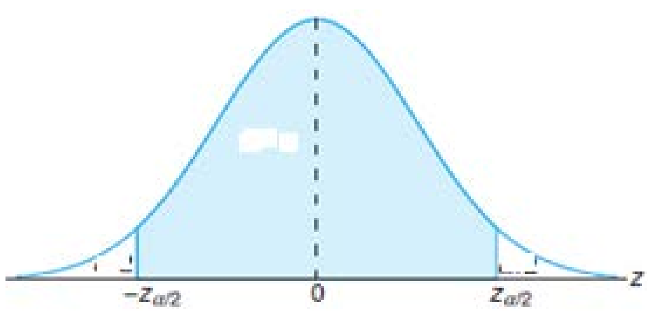
(i) $\chi^2_{\alpha/2}$ represents a specific chi-square value on the horizontal χ^2-axis that divides the χ^2-axis in such a way that the right portion of the axis contain an area of alpha/2 above it. The left portion of the axis then contains the remaining area of (1-alpha/2)

(ii) $\chi^2_{1-\alpha/2}$ represents a specific χ^2-value on the horizontal χ^2-axis that divides the axis in such a way that the right portion of the axis contain an area of (1-alpha/2) above it. The left portion of the axis then contains the remaining area of alpha/2.

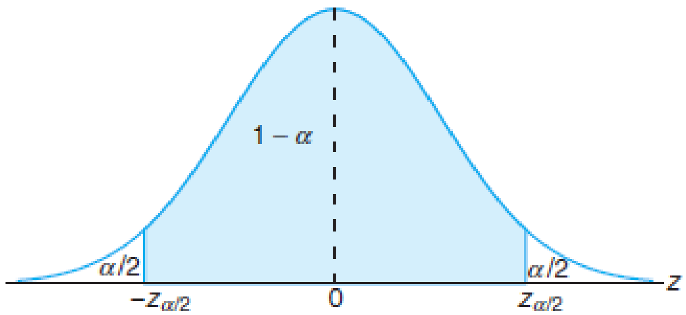
$\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ values are shown in the horizontal axis of the figure below. Note that the area to the right of $\chi^2_{1-\alpha/2}$ is 1-alpha +alpha/2 = 1-alpha/2.



7. ***Notations:*** In the figure below, label the areas contained above the three portions of the horizontal axis
(i) above $z > z_{\alpha/2}$ (ii) above $z < -z_{\alpha/2}$ and (iii) above $-z_{\alpha/2} < z < z_{\alpha/2}$.



Solution:



8. ***Probability statements:*** *Sketch the following probabilities [These probability statements will be used a lot in chapter 9]*

   (a) $P\left(-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$

   (b) $P\left(-t_{\frac{\alpha}{2}} < t < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$

   (c) $P\left({\chi^2}_{1-\alpha/2} < \chi^2 < {\chi^2}_{\alpha/2}\right) = 1 - \alpha$

   (d) $P\left(f_{1-\alpha/2} < f < f_{\alpha/2}\right) = 1 - \alpha$

Solution: The areas corresponding to each probabilities are shown below. The shaded region in each of the figure has an area of 1-alpha.

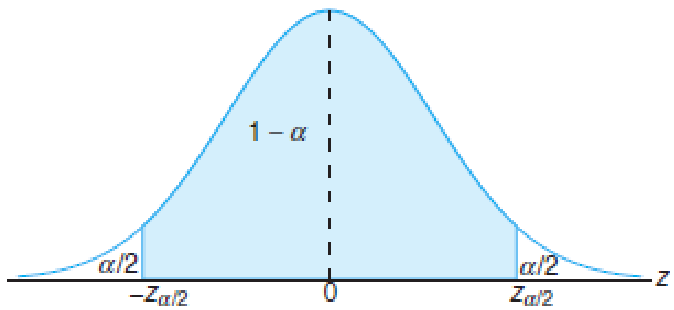


Figure 9.2: $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$.

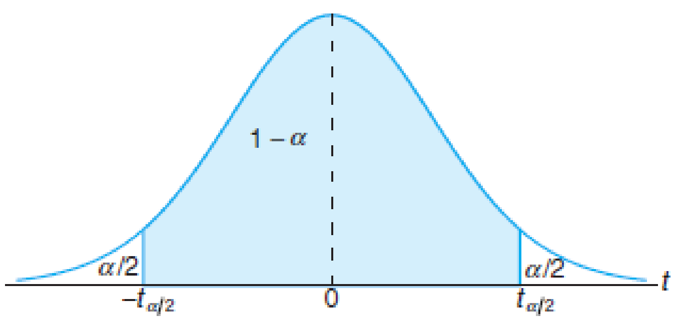


Figure 9.5: $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$.

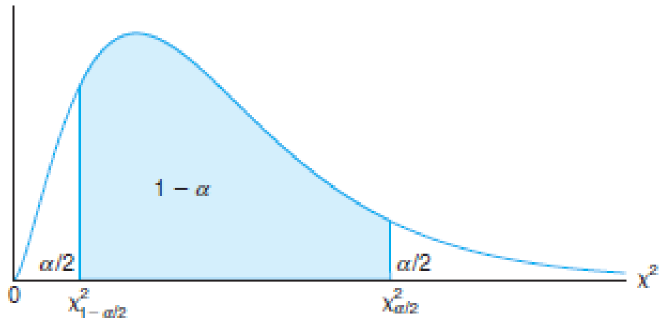


Figure 9.7: $P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha$.

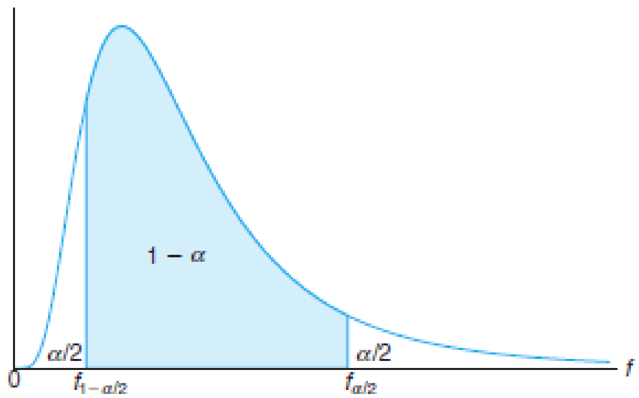


Figure 9.8: $P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha$.

9. ***An example to show the transition from Chapter 6 to Chapter 8***

An industrial sewing machine uses ball bearings that are targeted to have a diameter within specified limits. Because of the variation in the manufacturing process, some of the ball bearings have diameter that fall outside the specified limits. Consider that the diameter of the ball bearings is normally distributed with a mean of 0.503 inch and a standard deviation of 0.004 inch.

(a) What is the probability of obtaining **a particular ball bearing** having a diameter between 0.503 and 0.507 inch?

(b) What is the probability of obtaining **a particular sample of four ball bearing**s whose average diameter is between 0.503 and 0.507 inch?

Solution:

(a) We studied such problems in Chapter 6. Let X = diameter of a particular ball bearing chosen. Since the population follows normal distribution, X follows normal distribution.

$P(0.503 < X < 0.507) = P\left(\frac{0.503-0.503}{0.004} < Z < \frac{0.507-0.503}{0.004}\right)$

$= P(0 < Z < 1) = P(Z < 1) - P(0 < Z) = 0.8413 - 0.5 = 0.3413 = 34.13\%\ (Table\ A.6)$

We can say that, 34.13% of all ball bearings will have diameter in the given range.

(b) To solve this problem, we need concepts from Chapter 8. Here the random variable is the mean of a random sample of n=4 ball bearings. In Chapter 8, we learned that this random variable (sample mean) also follow normal distribution with the same population mean of 0.503 inch, but with a standard deviation of 0.004/sqrt(4) =0.002 inch

$$P(0.503 < \bar{X} < 0.507) = P\left(\frac{0.503 - 0.503}{0.002} < Z < \frac{0.507 - 0.503}{0.002}\right)$$
$$= P(0 < Z < 2) = P(Z < 2) - P(0 < Z) = 0.9772 - 0.5 = 0.4772 = 47.72\% \; (Table\ A.6)$$

We can say that, 47.72% of all ball bearings will have diameter in the given range. The probability has increased (compared to part (a)) because sample mean with n = 4 has less variability (0.002 inch in part (b) compared to 0.004 in part (a)). Note also that the random variable in part (a) may be considered as a sample mean with n = 1.

10. ***Definition:*** Define Z-statistic for the difference of means of two independent samples drawn from two populations.

Solution:
If independent samples of size n_1 and n_2 are drawn at random from two populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively, then the sampling distribution of the differences of means, $\bar{X}_1 - \bar{X}_2$, is approximately normally distributed with mean and variance given by

$$Mean\ of\ (\bar{X}_1 - \bar{X}_2) = \mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$
$$Variance\ of\ (\bar{X}_1 - \bar{X}_2) = \sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

The Z-statistic is then defined in the following manner.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - mean\ of\ (\bar{X}_1 - \bar{X}_2)}{standard\ deviation\ of (\bar{X}_1 - \bar{X}_2)}$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Note: An important theorem regarding f-statistic is the following. Please note that the order of the two d.f. (nu_1 and nu_2) is important when using this theorem.

Theorem 8.7: Writing $f_\alpha(v_1, v_2)$ for f_α with v_1 and v_2 degrees of freedom, we obtain

$$f_{1-\alpha}(v_1, v_2) = \frac{1}{f_\alpha(v_2, v_1)}.$$

Thus, the f-value with 6 and 10 degrees of freedom, leaving an area of 0.95 to the right, is

$$f_{0.95}(6, 10) = \frac{1}{f_{0.05}(10, 6)} = \frac{1}{4.06} = 0.246.$$

[Please solve the example problems in chapter 8 from the textbook]