Review of Probability Distributions (Z, t, χ^2, F)

Solutions - Gujarati Appendix C

C.1

(a) The number of independent observations available to compute an estimate, e.g., the sample mean or the sample variance.

(b) The probability distribution of an estimator.

(c) The (computable version or estimator of) the standard deviation of an estimator (commonly of the mean).

Note: Compare the standard deviation (sd) and standard error (se) of the mean estimator $\bar{X}$:

$$\mathrm{sd}(\bar{X}) = \frac{\sigma}{\sqrt{n}}, \quad \mathrm{se}(\bar{X}) = \frac{S}{\sqrt{n}}.$$

Also see **C.19**, **C.26**.

C.2

(a) True, $\mathrm{P}(Z > 1) = 0.5000 = 0.3413 = 0.1587 \approx 0.16$.

(b) True, $\mathrm{P}(1 \leq Z \leq 1.5) = 0.4332 = 0.3413 = 0.0919 \approx 0.09$.

(c) True, $\mathrm{P}(Z > 2.5) = 0.5000 - 0.4938 = 0.0062$.

C.3

(a) $X \sim N(8, 16/n)$.

(b) The variance of X depends on the sample size.

(c) Since $X \sim N(8, 16/25)$, the probability that $Z \leq -2.5 = 0.0062$.

C.4

Although both are symmetrical, the t distribution is flatter than the normal distribution. But as the degrees of freedom increase, the t distribution approximates the normal distribution.

C.5

(a) 0.10; (b) 0.10; (c) 0.20; (d) No.

C.6

True.

C.8

In large samples, the distribution of the sample mean of a r.v. can be approximated by a normal distribution regardless of the original population (i.e., PDF) from which the sample was drawn.

C.9

The chi-square (χ^2) distribution can be used to determine the probabilities for the sampling distribution of the sample variance S^2. In other words, a probability statement about a chi-square variable can be easily expressed into an equivalent probability statement about S^2. The F distribution can be used to find out if the variances of two normal populations are the same.

Note: For testing H_0: $\mu_1 = \mu_2$, we use t test; for testing $H_0 : \sigma_1^2 = \sigma_2^2$, we use F test.

C.10

(a) $Z = (1 - 1.5)/0.12 = -4.17$. The probability of obtaining a Z value equal to or less than -4.17 is extremely small.

(b) $Z_1 = (0.8 - 1.5)/0.12 = -5.8333$; $Z_2 = (1.3 - 1.5)/0.12 = -1.6667$. Therefore, $P(-5.8333 \leq Z \leq -1.6667)$ is very small. Note that the probability

$$P(Z_i \in (\text{mean} \pm 1.96\sigma)) = 0.95$$

is true for a normally distributed random variable. Given the mean of \$1.5 million and σ of \$0.12 million, the probability that a profit figure will be between 1.26 and 1.74 million is about 95%. Therefore, the probability that the profits will be between \$0.8 and \$1.3 million must be small indeed.

C.11

Since $P(Z \geq 1.28)$ is about 0.10, we obtain $1.28 = (X - 1.5)/0.12$, which gives $X =$ \$1.6536 million as the required figure.

C.12

From the preceding exercise, we know that $P(Z \geq 1.28) = 0.10$. Therefore, $1.28 = (80 - 75)/\sigma$, which gives $\sigma = 3.9063$.

C.13

(a) $Z = (6 - 6.5)/0.8 = -0.625 \approx -0.63$. Therefore, $P(Z \leq -0.63) = 0.2643$. Thus, approximately, 264 tubes will contain less than 6 ounces of toothpaste.

(b) The cost of the refill will be \$52.8 (= \$0.20 × 264).

(c) $Z = (7 - 6.5)/0.8 = 0.625 \approx 0.63$. The probability of $Z \geq 0.63$ is also ≈ 0.2643. Therefore, the profits lost will be \$13.2 (= \$0.05 × 264).

C.14

(a) $(X+Y) \sim N(25, 11)$.

(b) $(X-Y) \sim N(-5, 11)$.

(c) $3X \sim N(30, 27)$.

(d) $(4X+5Y) \sim N(115, 248)$.

C.15

In answering this question, note that if $W = aX + bY$,

$$\begin{aligned} \mathrm{E}(W) &= a\mu_X + b\mu_Y, \\ \mathrm{Var}(W) &= a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\rho\sigma_X\sigma_Y. \end{aligned}$$

(a) $(X+Y) \sim N(25, 16.88)$. (Note: $\sigma_X = 1.73$ and $\sigma_Y = 2.83$.)

(b) $(X-Y) \sim N(-5, 5.12)$.

(c) $3X \sim N(30, 27)$.

(d) $(4X+5Y) \sim N(115, 365.58)$, approximately.

C.16

Let $W = \frac{1}{2}(X) + \frac{1}{2}(Y)$. In this example,

$$\begin{aligned} \mathrm{E}(W) &= \tfrac{1}{2}(15) + \tfrac{1}{2}(8) = 11.5, \\ \mathrm{Var}(W) &= (\tfrac{1}{4})(25) + (\tfrac{1}{4})(4) + 2(\tfrac{1}{2})(\tfrac{1}{2})(-0.4)(5)(2) = 5.25. \end{aligned}$$

Therefore, $W \sim N(11.5, 5.25)$. The variance, hence the risk, of this portfolio is smaller than that of security X but greater than that of security Y. It is true that if you invest in security X, the expected return is higher than the portfolio return, but so is the risk. On the other hand, if you invest in security Y, the risk is smaller than that of the portfolio but so is the rate of return. Of course, you do not have to invest equally in the two securities.

C.17

If it is assumed that the SAT scores are normally distributed with mean and variance given in Example C.12, it can be shown that:

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

In the present example, we have: $\chi^2 = 9(142/102.07) = 12.521$, which is a chi-square variable with 9 d.f. From the χ^2 table, the probability of obtaining a chi-square of as much as 12.521 or greater is somewhere between 25% and 10%; the exact p-value being 18.55% (from a software package).

C.18

(a) We want $\mathrm{P}[(S^2/\sigma^2) > X] = 0.10$. That is,

$$\mathrm{P}\left[(n-1)\frac{S^2}{\sigma^2} > (n-1)X\right] = 0.10.$$

From the χ^2 table, we find that for 9 d.f., $9X = 14.6837$ or $X = 1.6315$. That is, the probability is 10 percent that S^2 will be more than 63% of the population variance.

(b) Following the same logic, it can be seen that:

$$\mathrm{P}\left[(n-1)X \leq (n-1)\frac{S^2}{\sigma^2} \leq (n-1)Y\right] = 0.95.$$

Using the χ^2 table, we find the X and Y values as 0.3000 and 2.1136, respectively.

Note: For 9 d.f., $\mathrm{P}(\chi^2 > 2.70039) = 0.975$ and $\mathrm{P}(\chi^2 > 19.0228) = 0.025$.

C.19

(a) Sample mean $\bar{X} = 15.9880$ ounces; sample variance $S^2 = 0.0158$ (ounces squared), sample standard deviation $S = 0.1257$.

(b) Calculate

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{15.988 - 16}{0.1257/10} = -0.3019.$$

For 9 d.f., the probability of obtaining a t value of -0.3019 or smaller is greater than 0.25 (one-tailed), the p-value being 0.3848. The t distribution is used here because the true variance (σ^2) is unknown.

Note: If σ^2 is known, then Z (standard normal) distribution should be used.

C.20

Use the F distribution. Assuming both samples are independent and come from the normal populations and that the two population variances are the same, it can be shown that:

$$F = \frac{S_1^2}{S_2^2} \sim F_{(m-1,n-1)}.$$

In this example, $F = 9/7.2 = 1.25$. The probability of obtaining an F value of 1.25 or greater is 0.2371.

C.24.

Recall that the following relationship between the F and the χ^2 distribution holds as the degrees of freedom in the denominator increases indefinitely ($n \to \infty$):

$$m \cdot F_{(m,n)} = \chi^2_{(m)}$$

where m are numerator d.f. From the statistical tables, we find that, at the 5% level, $\chi^2_{(10)} = 18.3070$. Now at the 5% level, the F values for $F_{(10,10)}, F_{(10,20)}$, and $F_{(10,60)}$ are, 2.98, 2.35, and 1.99, respectively. If we multiply the preceding values by 10, we obtain, 29.8, 23.5, and 19.9, which shows that as the denominator d.f. increase, the approximation becomes more accurate.

C.25.

Let $X \sim N(\mu, \sigma^2)$. Since $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$, therefore

$$\mathrm{E}(\bar{X}) = \frac{1}{n}[\mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)] = \frac{1}{n}[\mu + \cdots + \mu] = \frac{1}{n}[n\mu] = \mu,$$

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}[\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)] = \frac{1}{n^2}[\sigma^2 + \cdots + \sigma^2] = \frac{1}{n^2}[n\sigma^2] = \frac{\sigma^2}{n},$$

because $X_1, \cdots, X_n$ are *i.i.d.*

C.26

We have

$$\mathrm{E}(Z) = \mathrm{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\mathrm{E}(X-\mu) = 0,$$

because σ is a constant and the fact that $\mathrm{E}(X-\mu) = \mathrm{E}(X) - \mathrm{E}(\mu) = \mu - \mu = 0$.

Moreover,

$$\mathrm{Var}(Z) = \mathrm{E}[Z - \mathrm{E}(Z)]^2 = \mathrm{E}(Z^2)$$

since $\mathrm{E}(Z) = 0$. Now:

$$\mathrm{E}(Z^2) = \mathrm{E}\left[\frac{X-\mu}{\sigma}\right]^2 = \frac{1}{\sigma^2}\mathrm{E}[(X-\mu)^2] = \frac{1}{\sigma^2}\sigma^2 = 1.$$

Note: We usually standardize a r.v. by defining

$$\text{standardized r.v.} = \frac{\text{r.v.} - \text{mean}}{\text{sd}}$$

which has zero mean and unit variance. For a normally distributed r.v., the standardized r.v. follows $N(0,1)$. Applying this to the sample mean $\bar{X}$ of $X_1, \cdots, X_n$ in **C.25**, we have

$$\text{standardized r.v.} = \frac{\text{r.v.} - \text{mean}}{\text{sd}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}.$$

But when sd $(= \sigma/\sqrt{n})$ is unavailable and we have to use se $(= S/\sqrt{n})$ instead, then this is known as a t statistic which follows a t distribution:

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)},$$

as seen in **C.19**.