



통계적 품질관리[1]



1. 데이터 정리법 / 4-02
2. 확률분포 / 4-16
3. 관리도 / 4-23
4. 기출 · 예상문제 및 착안점 / 4-49



2. 확률분포 [품기2회]

2.1 이산형 확률분포

2.1.1 이산형 확률분포의 종류

* 이산확률분포(discrete probability distribution)는 수를 셀 수 있는 경우에 나타나는 계수치의 분포로서 다음과 같은 것들이 있다.

- ① 베르누이분포(Bernoulli distribution) ② 이항분포(binomial distribution)
③ 초기하분포(hypergeometric distribution) ④ 포아송분포(Poisson distribution)
⑤ 다항분포(multinomial distribution)

* 이들 분포들 중 ②~④분포들을 중심으로 차례대로 살펴보기로 한다.

2.1.2 이산형 확률분포의 확률밀도함수

(1) 이항분포의 확률밀도함수 [경지1회] [기지1회] [품기8회]

* 성공률이 P 인 베르누이 시행이 n 번 반복 시행되었을 때, 확률변수 X 를 “ n 번 시행에서의 성공횟수”라고 정의한다. 이때 확률변수 X 의 확률분포를 시행횟수 n 과 성공률 P 를 갖는 이항분포 $B(n, P)$ 라 한다. 확률변수 X 가 취할 수 있는 값은 $0, 1, 2, \dots, n$ 이다.

* 통계적 품질관리의 경우로서 부적합품률이 P 인 관리상태에 있는 공정(이는 무한모집단)으로부터 크기 n 인 표본을 취하거나, 또는 부적합품률 P 를 갖는 로트(이는 유한모집단임)로부터 크기 n 인 표본을 복원추출(sampling with replacement : 한 개를 추출하여 측정 한 다음 곧 그것을 모집단에 되돌려 넣고 다시 다음의 추출단위를 추출하여 가는 방법)할 때에, 표본 중에 발견되는 부적합품수 X 는 이항확률변수가 된다.

* 이항분포의 확률밀도함수 p.d.f.는 다음과 같다.

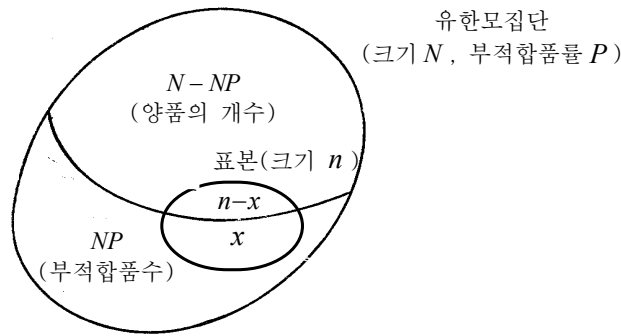
$$P_r(X=x) = p(x) = {}_nC_x P^x (1-P)^{n-x}, \quad x=0, 1, 2, \dots, n \quad (4.14)$$

* 일반적으로 속성 A (부적합이라는 것도 하나의 속성임)를 갖는 비율이 P 이고, 속성 A 를 갖지 않는 비율이 $1-P$ 인 무한모집단 또는 복원추출되는 유한모집단에서 크기 n 의 표본을 추출할 때 n 개 중에서 속성 A 를 갖는 개수 X 는 이항분포 $B(n, P)$ 에 따른다.

(2) 초기하분포의 확률밀도함수 [품기2회]

* 모부적합품률 P 이고 크기 N 인 로트(유한모집단)에서 표본(크기 n)을 채취할 때 부적합품수 X 의 확률분포는 복원추출을 하는 경우로서 이항분포를 따름을 이미 앞에서 보았다.

* 한편 복원추출이 아니고 비복원추출(sampling without replacement)인 경우 부적합품수 X 는 초기하분포(hypergeometric distribution)를 하게 된다.



[그림 4.8] 초기하분포의 샘플링 모델

* [그림 4.8]은 초기하분포의 샘플링 모델이고, 초기하분포는 다음과 같은 p.d.f를 가진다.

$$P_r(X=x) = p(x) = \frac{NP C_x \cdot N-NP C_{n-x}}{N C_n}, \quad x=0,1,2,\dots, \min(n, NP) \quad (4.15)$$

여기서, N : 로트 크기, P : 부적합품률, NP : 로트내 부적합품 수

$N-NP$: 로트내의 양품 수, n : 표본 크기

x : 표본 중의 부적합품 수, $n-x$: 표본 중의 양품 수

(3) 포아송분포의 확률밀도함수 [기지1회] [품기2회]

* 이항분포의 특수한 경우로서, 예를 들면 공장에서 일정기간 동안 발생한 사고건수, 직물의 일정한 단위면적 내에 있는 흠집의 수, 단위시간내의 걸려 오는 전화 수, 강판 등의 일정 면적당 흠의 수 등이 포아송분포(Poisson distribution)를 하는 예가 된다.

* 이항분포의 식 (4.14)에서 n 과 P 를 결합하여 “ $nP = m = \text{일정}$ ”하게 놓고, P 를 매우 작게, n 을 충분히 크게 취하면 식 (4.14)는 더욱 간단한 다음의 포아송분포로 된다.

* 포아송분포의 확률밀도함수 p.d.f.는 다음 식과 같다.

$$P(X=x) = p(x) = \frac{e^{-m} m^x}{x!}, \quad x=0,1,2,\dots, \quad m>0 \quad (4.16)$$

2.2 연속형 확률분포

2.2.1 연속형 확률분포의 종류

* 연속확률분포(continuous probability distribution)는 계량치의 분포라고도 말하며, 이것은 무게, 시간, 길이 등과 같은 양을 측정하는 경우에 나타난다.

* 연속확률분포로 다음과 같은 분포들이 많이 사용된다.

- ① 정규분포(normal distribution) ② χ^2 분포(χ^2 -distribution)
- ③ t 분포(t -distribution) ④ F 분포(F -distribution)
- ⑤ 지수분포(exponential distribution) ⑥ 와이블분포(Weibull distribution)
- ⑦ 균등(일양)분포(uniform distribution)

* 여기에서는 이들 분포 중 주로 쓰이는 ①~④의 분포들을 중심으로 살펴보고, ⑤~⑥은 신뢰성에 주로 활용되는 분포로서 제6장 신뢰성공학에서 다루기로 한다.

2.2.2 연속형 확률분포 내용

(1) 정규분포 [품기1회]

(가) 정규분포의 의미

* 정규분포(normal distribution)는 수학자 K. F. Gauss(1777~1855)가 발견한 것으로서, 확률변수 X 의 기대치가 μ 이고 표준편차가 σ 일 때 그 확률밀도함수(p.d.f.)가

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (-\infty < x < \infty, \sigma > 0) \quad (4.17)$$

이 되는 확률변수 X 는 정규분포 $N(\mu, \sigma^2)$ 에 따른다고 하였다.

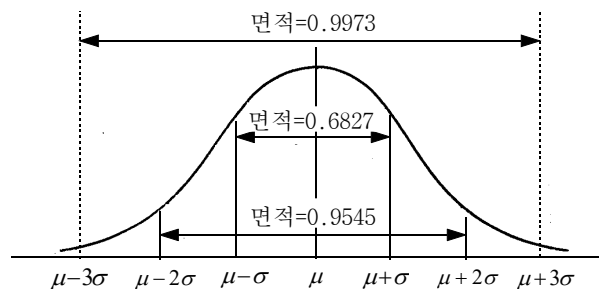
여기서 e 는 자연대수의 밑(base)으로 2.7183...인 무리수이다.

* 이 분포의 p.d.f.를 그래프로 그려보면 [그림 4.9]와 같이 평균(기대치) μ 에 대해 좌우대칭이고, 산포의 크기는 σ 에 의하여 결정되며, 대표적인 범위의 면적은 다음과 같다.

$$P[\mu - \sigma \leq X \leq \mu + \sigma] = 0.6827 \text{ (68.27\%)}$$

$$P[\mu - 2\sigma \leq X \leq \mu + 2\sigma] = 0.9545 \text{ (95.45\%)}$$

$$P[\mu - 3\sigma \leq X \leq \mu + 3\sigma] = 0.9973 \text{ (99.73\%)}$$



[그림 4.9] 정규분포의 p.d.f.의 그래프

* 확률변수 X 가 $N(\mu, \sigma^2)$ 에 따를 때, 다음의 변수변환

$$U = \frac{X - \mu}{\sigma} \quad (4.18)$$

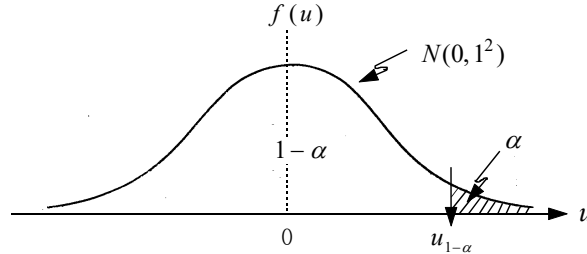
을 하면 확률변수 U 는 $\mu = 0, \sigma^2 = 1$ 이 되는 표준정규분포에 따르게 되며, $N(0, 1^2)$ 로 나타낸다. 정규분포의 p.d.f.는 다음과 같은 식이 된다.

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad (4.19)$$

* 부록의 정규분포표에는 $u_{1-\alpha}$ 에서 상측확률 α 를 구하는 표 등이 제시되어 있다.

$$P(U \geq u_{1-\alpha}) = \int_{u_{1-\alpha}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \quad (4.20)$$

* [그림 4.10]에 $u_{1-\alpha}$ 의 위치가 표시되어 있다.



[그림 4.10] 표준정규분포 $N(0, 1^2)$

* 확률변수 X 가 $N(\mu, \sigma^2)$ 에 따를 때 확률 $P[a \leq X \leq b]$ 은 식 (4.20)을 사용하여 다음과 같이 변수변환시켜 표준정규분포를 만든 후 부표를 사용하여 확률을 구할 수 있다. 즉,

$$\begin{aligned} P[a \leq X \leq b] &= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} \leq U \leq \frac{b-\mu}{\sigma}\right) \\ &= 1 - \left[P\left\{U < \frac{a-\mu}{\sigma}\right\} + P\left\{U > \frac{b-\mu}{\sigma}\right\} \right] \end{aligned} \quad (4.21)$$

예제 4.3 $N(3 \text{ cm}, 0.25\text{cm})$ 일 때 X 가 2.8~3.5cm내에 들어갈 확률은? (기사 기출)

해설

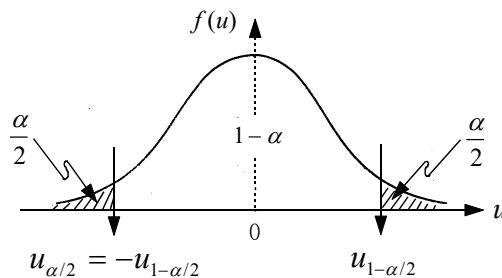
$\mu = 3, \sigma^2 = 0.25$ 이므로

$$\begin{aligned} P(a \leq X \leq b) &= P(2.8 \leq X \leq 3.5) = P\left(\frac{2.8-3}{0.5} \leq U \leq \frac{3.5-3}{0.5}\right) = P(-0.4 \leq U \leq 1.0) \\ &= 1 - [P(U < -0.4) + P(U > 1.0)] = 1 - (0.3446 + 0.1587) = 0.4967 (49.67\%) \end{aligned}$$

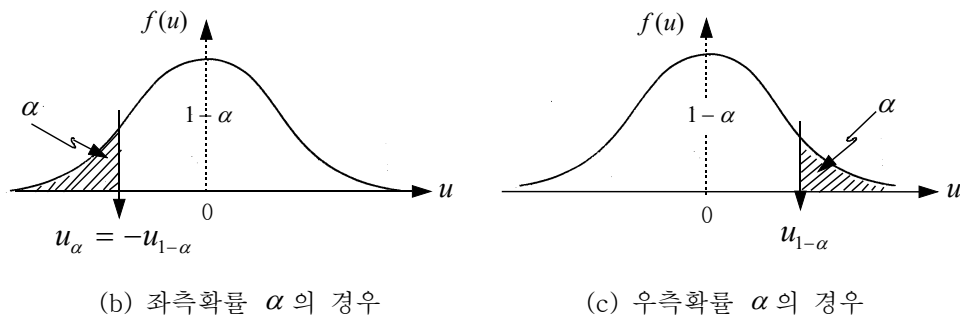
(나) 표준화정규분포의 확률밀도

* 분위점 u_α 는 $P_r(U \leq u_\alpha) = \alpha$, $P_r(U > u_\alpha) = 1 - \alpha$, $P_r(u_{\alpha/2} < U < u_{1-\alpha/2}) = 1 - \alpha$ 가 성립하고,

$u_\alpha = -u_{1-\alpha}$, $u_{\alpha/2} = -u_{1-\alpha/2}$ 가 성립한다.



(a) 양측확률 α 의 경우



[그림 4.11] 표준화정규분포의 확률밀도

<표 4.7> 분위점과 확률 관계

유의수준 α	양쪽의 경우	한쪽의 경우
$\alpha=0.10$ 일 때	$u_{1-\alpha/2} = u_{0.95} = 1.645$	$u_{1-\alpha} = u_{0.90} = 1.282$
$\alpha=0.05$ 일 때	$u_{1-\alpha/2} = u_{0.975} = 1.960$	$u_{1-\alpha} = u_{0.95} = 1.645$
$\alpha=0.01$ 일 때	$u_{1-\alpha/2} = u_{0.995} = 2.576$	$u_{1-\alpha} = u_{0.99} = 2.326$

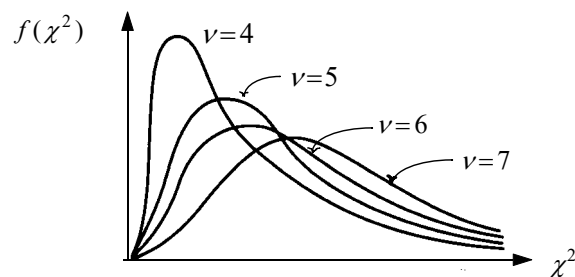
(2) χ^2 (카이제곱)분포

(가) χ^2 분포의 의미

* $N(\mu, \sigma^2)$ 인 정규모집단으로부터 크기 n 의 샘플을 샘플링하여, 그 데이터로부터 편차제곱합(변동) S 를 구했을 때, S 를 모분산 σ^2 으로 나누어서

$$\chi^2 = \frac{S}{\sigma^2} \quad (4.22)$$

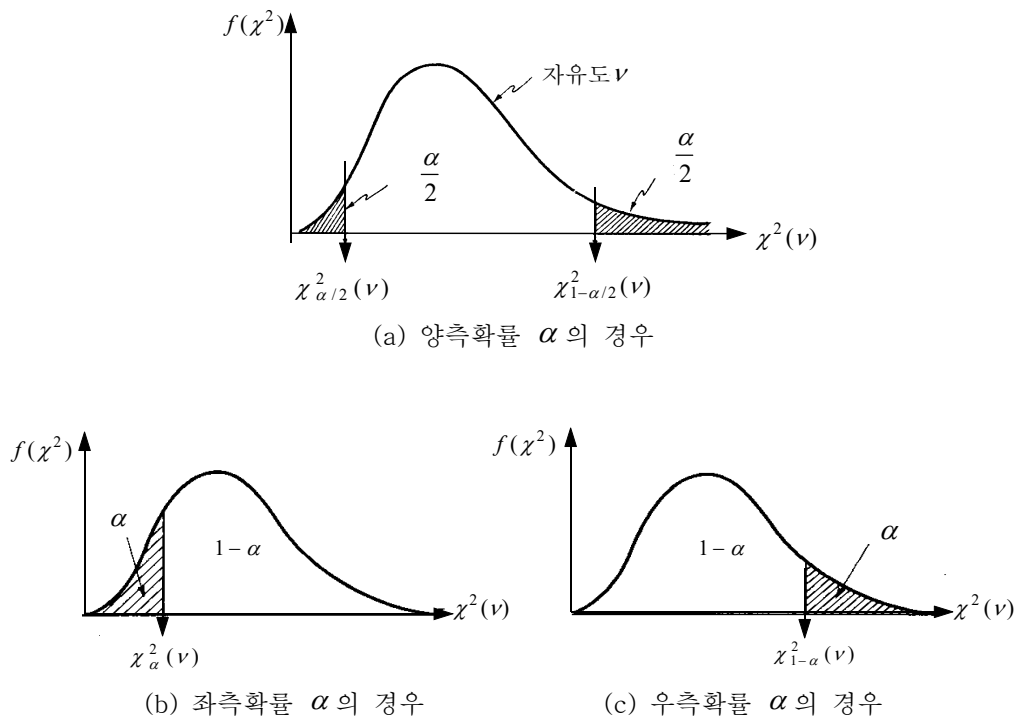
이라는 새로운 통계량을 구하면, 이 χ^2 의 값은 [그림 4.12]와 같은 자유도 $\nu = n-1$ 의 χ^2 분포를 한다고 알려져 있다.

[그림 4.12] χ^2 분포

(나) χ^2 분포의 확률밀도

* $\chi^2(\nu)$ 가 자유도 ν 인 χ^2 분포에 따르는 확률변수이면, $\chi^2_\alpha(\nu)$ 는 $P_r\{\chi^2(\nu) \leq \chi^2_\alpha(\nu)\} = \alpha$ 를 만족하는 점이며, 다음 식이 성립한다.

$$P_r\{\chi^2(\nu) > \chi^2_\alpha(\nu)\} = 1 - \alpha, \quad P_r\{\chi^2_{\alpha/2}(\nu) < \chi^2(\nu) < \chi^2_{1-\alpha/2}(\nu)\} = 1 - \alpha$$

[그림 4.13] 자유도 $\nu = n - 1$ 인 χ^2 분포의 확률밀도

(3) t 분포

(가) t 분포의 의미

- * $N(\mu, \sigma^2)$ 인 정규모집단으로부터 랜덤하게 샘플링한 크기 n 개의 표본(시료)의 평균 $\bar{X}$ 의 분포는 $N(\mu, \sigma^2/n)$ 의 정규분포로 되며, 이를 표준화하여

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

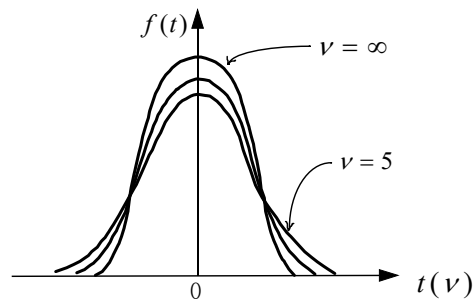
으로 놓으면, 확률변수 U 는 표준정규분포 $N(0, 1^2)$ 에 따르게 됨을 앞에서 설명하였다.

- * 한편, σ 대신에 그 추정값($\hat{\sigma}$)인 시료표준편차 $\hat{\sigma} = s = \sqrt{V}$ 를 대입하여

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad (4.23)$$

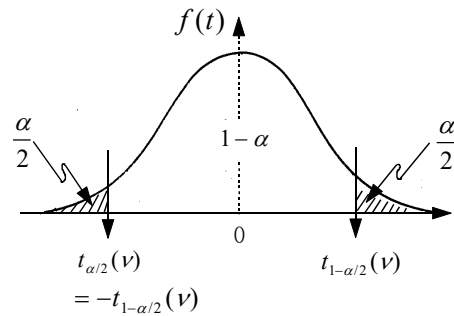
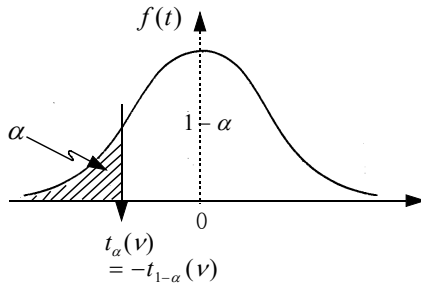
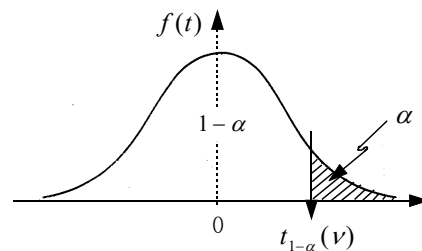
라는 새로운 확률변수 T 를 만들 때, 이 확률변수 T 는 정규분포로 되지 않고, 자유도 $\nu = n - 1$ 의 t 분포를 하는 것으로 알려져 있다.

- * t 분포는 자유도에 따라서 형태가 달라지나 어느 경우를 막론하고 좌우대칭이 된다.
- * t 분포는 자유도가 커짐에 따라서 정규분포에 접근하며, 자유도가 ∞ 일 때 $\sqrt{V} = \sigma$ 가 되므로 정규분포에 일치하게 된다.

[그림 4.14] t 분포(나) t 분포의 확률밀도

* 자유도 ν 인 t 분포를 따르는 확률변수가 $t(\nu)$ 일 때, 분위점 $t_\alpha(\nu)$ 는 $P_r[t(\nu) \leq t_\alpha(\nu)] = \alpha$ 를 만족하며, 다음이 성립한다.

$$P_r[t_{\alpha/2}(\nu) < t(\nu) < t_{1-\alpha/2}(\nu)] = 1 - \alpha, \quad P_r[t(\nu) > t_\alpha(\nu)] = 1 - \alpha$$

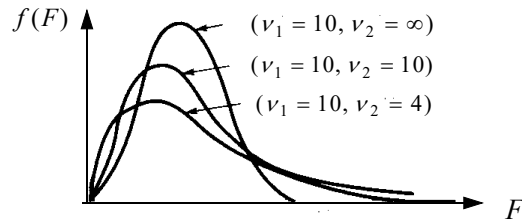
(a) 양측확률 α 의 경우(b) 좌측확률 α 의 경우(c) 우측확률 α 의 경우[그림 4.15] 자유도 $\nu=n-1$ 인 t 분포의 확률밀도(4) F 분포 [품기1회](가) F 분포의 의미

* 분산이 동일한 2개의 정규모집단으로부터 각각 랜덤하게 샘플링한 크기 n_1, n_2 의 2조의 표본에서 불편분산이 $V_1 = \frac{S_1}{n_1 - 1}$, $V_2 = \frac{S_2}{n_2 - 1}$ 이라면, V_1 과 V_2 의 비

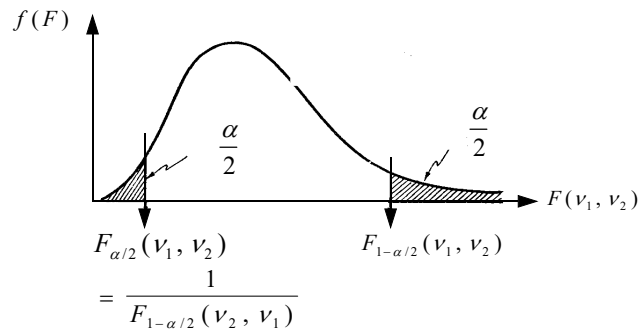
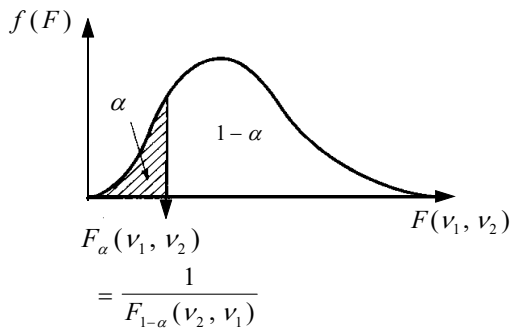
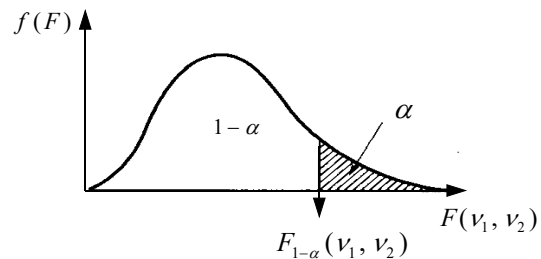
$$F = \frac{V_1}{V_2} \quad (4.24)$$

은 자유도 $\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$ 의 F 분포를 한다.

* F 분포는 [그림 4.16]에서와 같이 오른쪽으로 비뚤어진 형을 이루고 있으므로 ν_1, ν_2 의 조합에 의하여 분포 모양이 달라진다.

[그림 4.16] F 분포

(나) F 분포의 확률밀도

(a) 양측확률 α 의 경우(b) 좌측확률 α 의 경우(c) 우측확률 α 의 경우[그림 4.17] 자유도 ν_1, ν_2 인 F 분포의 확률밀도

3. 관리도

* 관리도(control chart)는 공정관리에 있어서의 우연원인에 의한 변동과 이상원인에 의한 원인을 구분하여 공정을 관리하기 위하여 고안된 것으로서, 제조공정의 관리 및 통계적 해석을 위해 많이 활용되는 기법이다.

* 관리도는 기존의 KS A 3201(관리도)가 KS Q ISO 8258(슈하트 관리도)를 거쳐 KS Q ISO 7870-2(관리도-제2부: 슈하트 관리도)로 명칭이 변경되고 용어 및 표현방법이 변경되었다.



통계적 품질관리[2]

-
1. 규격, 공차 및 공정능력 / 5-02
 2. 통계적 검정 및 추정 / 5-14
 3. 상관분석 및 회귀분석 / 5-47
 4. 샘플링 및 오차 / 5-73
 5. 샘플링검사 / 5-79
 6. 기출 · 예상문제 및 착안점 / 5-137
-

3. 상관분석 및 회귀분석

3.1 상관·회귀분석 개요

3.1.1 상관·회귀의 개념

- * 금속의 성분과 항장력 또는 항장력과 연신률, 소둔온도와 금속의 경도 등과 같이 만약 한쪽의 측정치가 변하면 다른 측정치가 이것에 따라 변화하는 관계가 있을 때 이들 사이에는 “상관이 있다”고 말하며, 상관관계를 해석하는 방법을 상관분석이라 한다.
- * 상관분석의 방법에는 확률변수가 X, Y 와 같이 2개 있을 경우의 해석법을 단상관분석, 3개 이상의 확률변수에 대한 관계를 해석하는 방법을 중상관분석이라 한다.
- * 상관분석은 X, Y 다 같이 랜덤하게 변하는 확률변수로서 2변량 정규분포를 하는 것이 전제로 되어 있다.
- * 한편, X 가 확률변수가 아닐 경우, 예를 들면 온도를 $100^{\circ}\text{C}, 110^{\circ}\text{C}, 120^{\circ}\text{C}, \dots$ 와 같이 지정하여 변화시키고, 이때의 반응생성물의 수확량을 측정하여, 이 온도와 수확량과의 관계를 문제로 삼을 경우에는, X 가 랜덤하게 변하는 확률변수가 아니므로 상관분석을 적용하는 것은 적절하지 못하다. 이와 같이 X 의 값을 알고 있는 확정수일 때 Y 와의 관계는 회귀분석으로 취급한다.

3.1.2 상관·회귀의 목적

- * 품질관리에서는 다음의 목적 때문에 상관·회귀가 사용된다.
 - ① 특성과 특성간의 상관관계, ② 대용관리특성의 발견, ③ 특성과 요인간의 관계
 - ④ 증별하여 상관을 취함으로써 교호작용 발견 등
- * 상관과 회귀의 중국 목적은 변수 x 로부터 변수 y 를 추정하는데 있다.
- * 상관·회귀의 공통점과 차이점을 살펴보면, 공통점은 변수들간의 관계를 규명하는데 쓰이는 것이며, 차이점은 회귀는 통계적 모형을 구축하고 그 모형으로부터 미래값을 예측하는데 활용하는 데 있다.

3.2 상관분석 [정지1회] [품기4회]

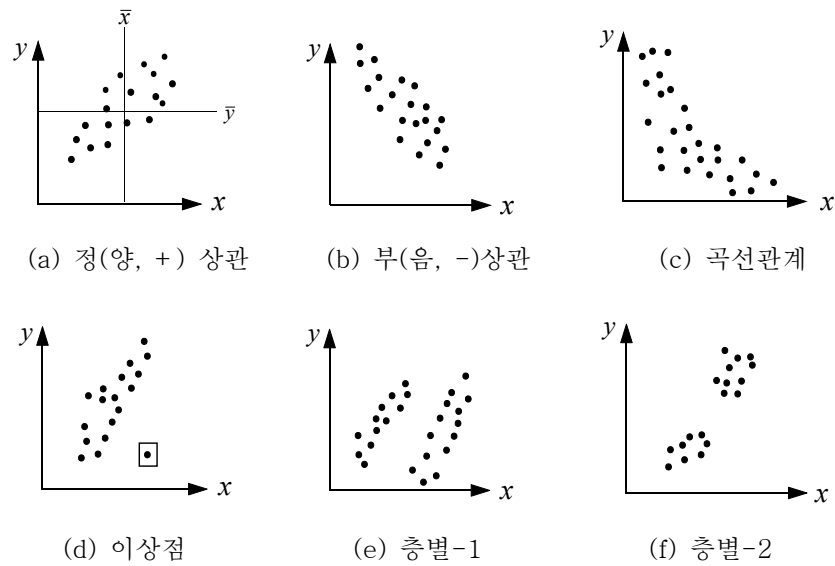
3.2.1 상관분석의 순서

- * 상관분석은 다음의 순서에 의해 이루어진다.
 - ① 산점도(산포도라고도 함)를 그린다. ② 상관계수(r)를 계산한다.
 - ③ 상관계수의 유의성에 대한 검정 및 추정을 한다.
 - ④ 유의이면 x 에 대한 y 의 회귀선을 구한다.
 - ⑤ 결정계수(기여율, r^2)을 구한다.

3.2.2 상관분석을 위한 산점도의 작성

(1) 상관·회귀의 기초로서의 산점도

- * 2변량 사이의 상관관계를 알기 위하여 2개의 변량 x , y 를 그래프에 타점하여 2개의 변량을 각자 세로축과 가로축에 잡아 측정치를 타점하여 [그림 5.13]과 같이 작성한 것을 산점도(scatter diagram)라고 한다.
- * 산점도 작성 방법에 대해서는 데이터정리법에서의 7가지 QC기초수법 중 산점도 항목을 참조하도록 한다.



[그림 5.13] 산점도

(2) 산점도를 그린 후 검토사항

- * 산점도를 그린 후에 검토해야 할 사항은 다음과 같은 것들이다.
 - ① 점들이 산재해 있는 모양으로부터 x 와 y 사이에 관계가 있을 듯한가에 대해 검토한다.
또한 양(positive) 상관인가 음(negative) 상관인가를 검토함[그림 5.13의 (a)와 (b)].
 - ② x 와 y 가 직선관계인가 곡선관계인가를 검토한다.
 - * [그림 5.13]의 (c)처럼 곡선관계인 경우에는 상관계수를 구하는 것은 의미가 없고, 이와 같은 경우는 곡선회귀분석을 적용해야 한다. 회귀선이 2차식이면 $r=0$ 이 된다.
 - ③ 이상한 데이터가 있는지를 확인한다.
 - * 산점도에서 이상점의 발생은 데이터의 수집시 다른 모집단의 시료가 혼입되었거나, 측정이나 계산이 잘못되었거나, 데이터의 기입 등에 착오가 있는 것 등이다.
 - * 이와 같은 이상점이 발견되면 원인을 규명하여 수정하여 주도록 한다.
 - ④ 점들이 뚜렷하게 두 개 또는 그 이상으로 층별되는 경우가 있는가 검토한다.
 - * [그림 5.13]의 (e)처럼 두 개의 층으로 나누어지는 경우에는 두 개의 모집단으로부터 얻어진 두 개의 시료가 섞여 있는 것이므로 두 개의 서로 다른 시료로 취급해야 한다.

3.2.3 표본상관계수 r 을 구함 [정지1회] [품기3회]

* 상관계수는 두 변수간의 선형(직선) 관계가 존재하는지 알아보는 계수이다. 회귀분석과 유사하지만 인과관계에 대한 분석은 아니다.

* 두 확률변수 X, Y 에 대하여 n 개의 데이터 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 이 얻어졌을 때 그 표본(또는 시료)상관계수(sample correlation coefficient)는 다음과 같이 정의된다.

$$r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} \quad (5.61)$$

$$\text{여기서, } S_{(xx)} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$S_{(yy)} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$S_{(xy)} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$$

* 점들이 직선에 모여 있으면 상관계수는 커진다. 상관계수의 부호가 양이면 한 변수의 값이 커질수록(작아질수록) 다른 변수의 값도 커짐(작아짐)을 의미한다.

* 표본의 크기가 커지면 상관계수의 값이 커지며, 상관계수의 값이 얼마 이상이어야 함은 기준은 없으므로, 가설검정에 의한 유의확률(P value)을 계산하도록 한다.

* 상관계수에 대해 다음 사항을 특히 주의하도록 한다.

- ① 상관계수는 두 변수간의 선형관계를 알아보는 것이다. 2차관계의 상관계수는 0이다.
- ② 상관계수는 점들이 직선에 모여 있는 정도를 나타내는 지표이지 직선의 기울기 크기를 나타내는 것이 아니다. (완전 일직선이면 기울기에 관계없이 $r=1$ 이 됨)

* 표본상관계수가 갖는 몇 가지 성질을 들어보면 다음과 같다.

- ① r 의 범위는 $-1 \leq r \leq 1$ 이 된다.
- ② r 의 값은 두 확률변수 X, Y 간의 선형관계(linear relationship)를 나타내는 척도로서 [그림 5.13]의 (a)의 경우는 $0 < r < 1$ 이고 (b)의 경우는 $-1 < r < 0$ 이 된다.
 $r = \pm 1$ 인 경우는 모든 점이 기울기가 영이 아닌 일직선상에 놓이게 됨을 의미한다.
- ③ 만약 x 가 $ax + b$ 로 바뀌고, y 가 $cy + d$ 로 바뀌어도 a 와 c 의 부호가 같으면 r 의 값에는 변함이 없다. 따라서 계산의 편의상 아래와 같이 수치변환(단, $h > 0, g > 0$ 임)

$$X = (x - x_0) \times h \quad Y = (y - y_0) \times g$$

을 하여도 x, y 간의 상관계수는 X, Y 의 상관계수와 동일하다. 즉,

$$r = \frac{S_{(xy)}}{\sqrt{S_{(xx)} \cdot S_{(yy)}}} = \frac{\frac{1}{hg} S_{(XY)}}{\sqrt{\frac{1}{h^2} S_{(XX)} \cdot \frac{1}{g^2} S_{(YY)}}} = \frac{S_{(XY)}}{\sqrt{S_{(XX)} \cdot S_{(YY)}}}$$

예제 5.22 두 변수 x, y 에 대하여 다음 6개의 데이터가 있다. 표본상관계수(r)를 구하라.

x	1	2	3	5	6	7
y	3	4	6	8	9	12

해설

표본상관계수를 구하기 위해 다음 [표 1]과 같은 계산을 위한 보조표를 만든다.

[표 1] 표본상관계수를 구하기 위한 보조표

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	3	-3	-4	9	16	12
2	4	-2	-3	4	9	6
3	6	-1	-1	1	1	1
5	8	1	1	1	1	1
6	9	2	2	4	4	4
7	12	3	3	9	25	15
$\bar{x} = 4$	$\bar{y} = 7$			$S_{(xx)} = 28$	$S_{(yy)} = 56$	$S_{(xy)} = 39$

$$r = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}} = \frac{39}{\sqrt{28 \times 56}} = 0.985$$

3.2.4 상관계수 유의성 검정 및 추정

(1) 상관계수 유의성 검정 및 추정의 개념

(가) 모상관계수

* 두 확률변수 X, Y 의 기대치와 분산을 각각 μ_x, μ_y 및 σ_x^2, σ_y^2 이라고 하고, 다음과 같은 기대치를 구하면, $E[(X - \mu_x)(Y - \mu_y)]$ 와 같은 기대치를 X, Y 의 공분산(covariance)이라고 하며, $Cov(X, Y)$ 혹은 σ_{xy} 로써 표시한다.

$$\begin{aligned} E[(X - \mu_x)(Y - \mu_y)] &= E(XY - \mu_y X - \mu_x Y + \mu_x \mu_y) \\ &= E(XY) - \mu_x \mu_y \quad (\because E(X) = \mu_x, E(Y) = \mu_y) \end{aligned}$$

* 만약 X, Y 가 서로 독립이면 다음 식과 같이 되므로 공분산은 영(0)이 된다.

$$E(XY) = E(X) \cdot E(Y) = \mu_x \cdot \mu_y$$

* 여기서 공분산의 추정치는 $\hat{\sigma}_{xy} = V_{xy}$ 로 표기되기도 하며, $V_{xy} = S_{(xy)} / (n-1)$ 로 계산된다.

* 두 확률변수 X, Y 간의 상관계수(correlation coefficient)는 $Corr(X, Y)$ 또는 간단히 ρ 로 나타내며, 다음 식으로 정의된다.

$$\rho = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sqrt{E(X - \mu_x)^2} \cdot \sqrt{E(Y - \mu_y)^2}} = \frac{Cov(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y} \quad (5.62)$$

상기 ρ 의 식에서 X, Y 가 서로 독립이면 $Cov(X, Y) = 0$ 이므로 $\rho = 0$ 이 된다.

상기 식에서 $\sigma_x, \sigma_y, \sigma_{xy}$ 는 다음과 같은 의미이다.

$$\sigma_x^2 = E(X - \mu_x)^2, \sigma_y^2 = E(Y - \mu_y)^2, \sigma_{xy} = E(X - \mu_x)(Y - \mu_y)$$

(나) 상관계수의 분포와 Z 변환

* 시료상관계수 r 은 통계량으로서 분포를 가지며, r 의 기대치와 분산은 다음 식과 같다.

$$\left. \begin{aligned} E(r) &= \rho \\ V(r) &= \frac{(1 - \rho^2)^2}{n} \end{aligned} \right\} \quad (5.63)$$

* r 의 분포는 매우 복잡해서 ρ 의 신뢰한계를 r 에 의해 직접 계산하는 것은 곤란하므로, 시료의 크기가 상당히 크고($n \geq 25$), ρ 의 절대치가 그다지 크지 않을 때 다음과 같이 r 의 Z 로의 Z 변환을 함(Z 변환표에서 Z 대신 z' 로 표기되기도 함). R. A. Fisher에 의하면

$$Z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \quad [= \tanh^{-1}(r)] \quad (5.64)$$

로 정의되는 Z 값은 근사적으로 정규분포에 따르고, 기대치 $E(Z)$ 와 분산 $V(Z)$ 는

$$E(Z) = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) + \frac{\rho}{2(n-1)} \approx \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \quad (5.65)$$

$$V(Z) = \frac{1}{n-3} \quad (5.66)$$

임이 알려져 있다. 이 변환을 Z 변환이라고 부른다.

* 따라서 표준화된 확률변수인 통계량, 즉

$$U = \frac{Z - E(Z)}{\sqrt{V(Z)}} = \sqrt{n-3} \cdot \left\{ \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) - \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \right\} \quad (5.67)$$

는 대략적으로 표준화 정규분포인 $N(0, 1^2)$ 에 따르게 된다.

(2) 상관계수의 유의성 검정

(가) 검정순서

* 상관계수 ρ 에 대한 유의성 검정은 다음의 순서로 실시된다.

① 가설 설정. 귀무가설 H_0 및 대립가설 H_1 을 정한다.

* 귀무가설 H_0 는 내리고자 하는 판단결과에 반대가 되도록 하는 가설이고, 대립가설 H_1 은 내리고자 하는 결론에 일치가 되는, 즉 H_0 에 반(反)하는 가설이다.

② 유의수준 α 를 정함 : 유의수준 α 란 제1종과오를 일으킬 확률(즉, 위험률)로서, H_0 가 채택되어야 하는 데도 불구하고 기각될 과오(예를 들면, 공정이 변하지 않았는데도 변했다고 판단할 과오)를 뜻한다.

③ 검정통계량의 값($|t_0|$ 혹은 $|r_0|$) 계산 : $t_0 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ 혹은 $r_0 = \frac{S_{(xy)}}{\sqrt{S_{(xx)}S_{(yy)}}}$

④ 기각역 설정 : $|t_0| > t_{1-\alpha/2}(n-2)$ 혹은 $|r_0| > r_{1-\alpha/2}(n-2)$ 이면 H_0 기각

⑤ 판정 : 이들의 검정순서에 따른 구체적인 검정방법은 다음 항에서 기술하기로 한다.

(나) 검정 방법

1) 상관계수의 검정 방법 ($\rho = 0$ 일 때)

적용	H_0	H_1	검정통계량	기각역
t 표	$\rho = 0$	$\rho \neq 0$	$t_0 = r\sqrt{n-2} / \sqrt{1-r^2}$	$ t_0 > t_{1-\alpha/2}(n-2)$
r 표	$\rho = 0$	$\rho \neq 0$	$r_0 = S_{(xy)} / \sqrt{S_{(xx)} \cdot S_{(yy)}}$	$ r_0 > r_{1-\alpha/2}(n-2)$
비고	$r = S_{(xy)} / \sqrt{S_{(xx)}S_{(yy)}}$, r 분포표의 값은 부록의 <부표 11>을 참조하도록 한다.			

* 이는 모상관계수 ρ 가 0인가 아닌가, 즉 두 데이터간의 상관관계 유무의 검정으로서, 이 경우는 다음 2가지의 성질에 의거 검정이 가능하다.

① $t = r\sqrt{n-2} / \sqrt{1-r^2}$ 로 정의되는 t 가 자유도 $\nu = n-2$ 의 t 분포

② $r = S_{(xy)} / \sqrt{S_{(xx)}S_{(yy)}}$ 로 정의되는 r 이 자유도 $\nu = n-2$ 의 r 분포

* 이러한 상관계수의 유의성 검정으로 H_0 가 기각되면, 회귀선을 $\hat{y} - \bar{y} = b(x - \bar{x})$ 로 구할 수 있게 된다.

[참조] 수치변환에 의한 상관계수의 계산 방법

상관계수(r)을 구하기 위한 수치변환의 경우는 다음과 같은 방법으로 구한다.

$X = (x - x_0) \times h$, $Y = (y - y_0) \times g$ 로 수치변환하면, 변동(편차제곱합)은

$$S_{(XX)} = \sum X^2 - (\sum X)^2 / n, \quad S_{(YY)} = \sum Y^2 - (\sum Y)^2 / n$$

$$S_{(XY)} = \sum XY - (\sum X)(\sum Y) / n$$

이고, 이들을 원데이터의 값으로 환원시키기 위해 다음과 같이 계산한다.

$$S_{(xx)} = S_{(XX)} \times \frac{1}{h^2}, \quad S_{(yy)} = S_{(YY)} \times \frac{1}{g^2}, \quad S_{(xy)} = S_{(XY)} \times \frac{1}{hg}$$

예제 5.23 어떤 제품의 특성치 X, Y 의 상관계수를 32개 시료에서 구한 결과 0.674를 얻었다. 상관관계가 없다는 귀무가설은 있다는 대립가설에 대하여 유의수준 $\alpha = 0.01$ 로 검정하시오. (단, t 검정을 하시오) [기사 기출]

[해설]

☞ 상관계수의 유의성 검정을 t 검정에 의거 행한다.

① 가설설정 : $H_0 : \rho = 0$, $H_1 : \rho \neq 0$ ② 유의수준 : $\alpha = 0.01$ 로 한다.

③ 검정통계량의 값(t_0) 계산 : $t_0 = r\sqrt{n-2} / \sqrt{1-r^2} = 0.674\sqrt{32-2} / \sqrt{1-(0.674)^2} = 4.997$

④ 기각역 설정 : $|t_0| > t_{1-\alpha/2}(n-2) = t_{0.995}(30) = 2.750$ 이면 H_0 기각

⑤ 판정 : $|t_0| = 4.997 > t_{0.995}(30) = 2.750$ 이므로 위험률(유의수준) 1%로 H_0 를 기각한다.

즉, 상관계수는 고도로 유의하다(상관관계가 큼).

2) 상관계수의 유의성 검정 ($\rho = \rho_0$, 즉 $\rho \neq 0$ 일 때)

H_0	H_1	검정통계량	기각역
$\rho = \rho_0$	$\rho \neq \rho_0$	$U_0 = \sqrt{n-3} \cdot \left\{ \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2} \ln\left(\frac{1+\rho_0}{1-\rho_0}\right) \right\}$	$ U_0 > u_{1-\alpha/2}$

* 종전의 상관계수(모상관계수) ρ 가 얼마라고 주어져 있는데 공정변화 후 ρ 가 달라졌는가에 대한 검정으로서, 이 경우

$$Z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right), \quad E(Z) = \frac{1}{2} \ln\left(\frac{1+\rho_0}{1-\rho_0}\right) \quad (5.68)$$

를 사용한 다음의 검정통계량을 사용하여 검정한다.

$$U_0 = \frac{Z - E(Z)}{\sqrt{V(Z)}} = \sqrt{n-3} \cdot \left\{ \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2} \ln\left(\frac{1+\rho_0}{1-\rho_0}\right) \right\} \quad (5.69)$$

예제 5.24 어느 공장에서 종래에 생산한 탄소강의 시험편의 지름과 항장력 사이의 상관계수는 $\rho_0 = 0.749$ 였다. 최초 재료의 일부가 변경되었다. 그로 인해 상관계수가 달라지지 않았는가를 조사하기 위해 크기 64인 시험편에 대하여 그 지름과 항장력을 측정하여 상관계수를 계산하였더니 $r = 0.818$ 이었다. 재료가 바뀜으로써 모상관계수가 달라졌다고 할 수 있겠는가?

유의수준 5%로 검정하시오. [기사 기출]

해설 [품기1회]

☞ 상관의 유의성 검정($\rho = \rho_0$, 즉 $\rho \neq 0$ 일 때)으로서

① 가설 설정 : $H_0 : \rho = 0.749(\rho_0)$, $H_1 : \rho \neq 0.749$ ② 유의수준 : $\alpha = 0.05$

③ 검정통계량의 값(U_0) 계산

$$\begin{aligned} U_0 &= \sqrt{n-3} \cdot \left\{ \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) - \frac{1}{2} \ln\left(\frac{1+\rho_0}{1-\rho_0}\right) \right\} \\ &= \sqrt{64-3} \cdot \left\{ \frac{1}{2} \ln\left(\frac{1+0.818}{1-0.818}\right) - \frac{1}{2} \ln\left(\frac{1+0.749}{1-0.749}\right) \right\} = 1.406 \end{aligned}$$

여기서, $n = 64$, $r = 0.818$

④ 기각역 설정 : $|U_0| > u_{1-\alpha/2} = u_{0.975} = 1.960$ 이면 H_0 기각

⑤ 판정 : $|U_0| = 1.406 < u_{0.975} = 1.960$ 이 되므로 H_0 를 기각할 수 없다.

즉, 모상관계수가 달라졌다고 할 수 없다.

(3) 상관계수에 관한 추정

(가) 두 확률변수 X 와 Y 간의 모상관계수 ρ 의 추정

* 두 확률변수 X, Y 간의 모상관계수 ρ 는 다음 식과 같다.

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{V(X) \cdot V(Y)}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

여기서, $\text{Cov}(X, Y) = \sigma_{xy}$, $V(X) = \sigma_x^2$, $V(Y) = \sigma_y^2$ 의 관계이다.

* 이들의 시료에 의한 추정치는 각각

$$\hat{\sigma}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{S_{(xy)}}{n-1}$$

$$\hat{\sigma}_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{S_{(xx)}}{n-1} \quad \hat{\sigma}_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{S_{(yy)}}{n-1}$$

를 사용할 수 있으므로, ρ 의 점추정치는 다음과 같이 추정된다.

$$\hat{\rho} = \frac{\frac{S_{(xy)}}{n-1}}{\sqrt{\frac{S_{(xx)}}{n-1} \cdot \frac{S_{(yy)}}{n-1}}} = \frac{S_{(xy)}}{\sqrt{S_{(xx)} S_{(yy)}}} = r \quad (5.70)$$

(나) 모상관계수에 관한 신뢰구간의 추정 (대략 $n \geq 25$ 일 때)

* 상관계수의 분포에서 $E(Z)$ 값은 다음 식으로 정의가 됨을 앞에서 이미 보았다.

$$E(Z) = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$$

* Z 의 $100(1-\alpha)\%$ 신뢰구간은 다음 식으로 된다.

$$\left. \begin{matrix} Z_U \\ Z_L \end{matrix} \right\} = Z \pm u_{1-\alpha/2} \sqrt{\frac{1}{n-3}} \quad (5.71)$$

구간상한 Z_U 와 구간하한 Z_L 에 대응하는 ρ_U, ρ_L 의 값이 모상관계수의 구간추정값이다.

* Z 값으로부터 ρ 를 추정하는 방법은 다음의 방법들이 있다.

1) 방법 1 [$r = \tanh(Z)$ 이용]

* $Z = \tanh^{-1}(r)$ 이므로 $r = \tanh(Z)$ 이 된다. 따라서 $\hat{\rho} \approx r = \tanh(Z)$ 가 된다.

$$Z_U \rightarrow \rho_U \text{ 변환 : } \hat{\rho}_U \approx r_U = \tanh(Z_U)$$

$$Z_L \rightarrow \rho_L \text{ 변환 : } \hat{\rho}_L \approx r_L = \tanh(Z_L)$$

* 따라서 ρ 의 구간추정은 $\hat{\rho}_L \leq \hat{\rho} \leq \hat{\rho}_U$ 로 구해진다.

2) 방법 2 [$r = (e^{2Z} - 1)/(e^{2Z} + 1)$ 의 공식 사용]

* ρ 의 $100(1-\alpha)\%$ 신뢰구간은 다음의 식으로 계산되는 것이 증명되어져 있다.

$$\hat{\rho}_L \approx r_L = \frac{e^{2Z_L} - 1}{e^{2Z_L} + 1} \quad (5.72)$$

$$\hat{\rho}_U \approx r_U = \frac{e^{2Z_U} - 1}{e^{2Z_U} + 1} \quad (5.73)$$

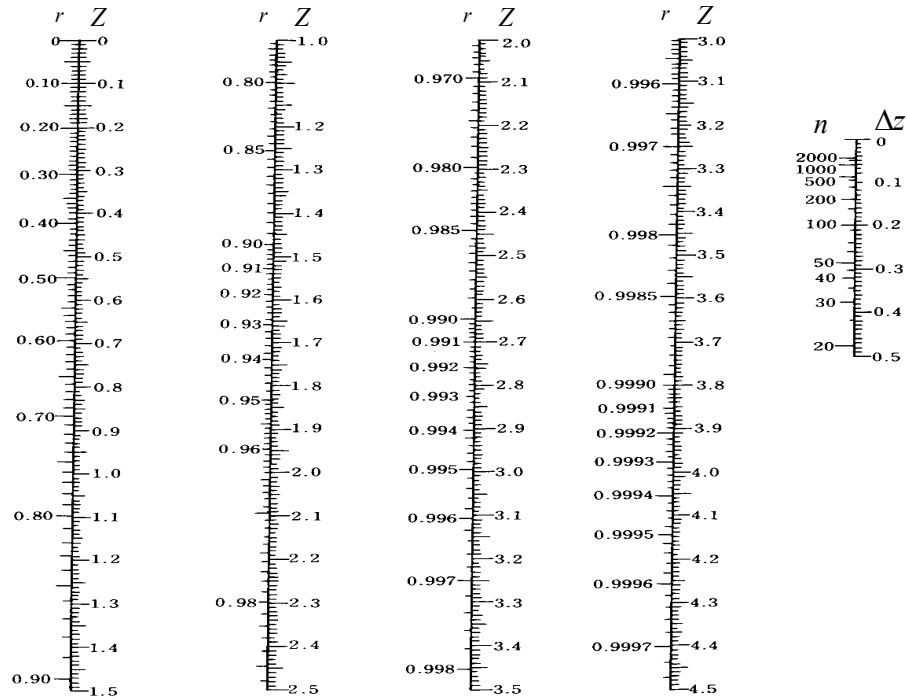
* 따라서 ρ 의 구간추정은 $\hat{\rho}_L \leq \hat{\rho} \leq \hat{\rho}_U$ 로 구해진다.

3) 방법 3 (Z 변환표에 의한 경우)

▶(순서 1) 시료상관계수 r 의 값을 Z 변환도표 상에서 찾는다.

▶(순서 2) 시료의 크기 n 에 대응하는 ΔZ 의 폭을 구한다.

▶(순서 3) 앞에서 Z 변환도표 위에서 잡은 r 의 값에서 상하에 ΔZ 의 폭을 주었을 때 얻어지는 상하 두 점의 r 의 값이 신뢰구간의 하한 ρ_L 과 상한 ρ_U 의 추정값이 된다.



$$Z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = \tanh^{-1}(r) \quad r = \tanh(Z) \quad \Delta Z = \frac{1.96}{\sqrt{n-3}}$$

[그림 5.14] Z 변환도표

3.2.5 x 에 대한 y 의 회귀선 [품기1회]

* x 에 대한 y 의 회귀선은 다음의 식으로 구한다.

$$\hat{y} - \bar{y} = b(x - \bar{x}) \quad (5.74)$$

* 상기 식에서 회귀계수 b 는 다음 식으로 구한다.

$$b = \frac{S_{(xy)}}{S_{(xx)}} \quad (5.75)$$

* 상기 b 의 식에서 $S_{(xx)}$ 와 $S_{(xy)}$ 는 다음의 식에서 구해진 값을 구하여 대입시킴.

$$S_{(xx)} = \sum x^2 - \frac{(\sum x)^2}{n}, \quad S_{(xy)} = \sum xy - \frac{\sum x \sum y}{n}$$

[참조] y 에 대한 x 의 회귀선을 구할 경우 (이는 특수한 경우임) [품기1회]

$$x - \bar{x} = b(y - \bar{y}) \quad \text{여기서, 회귀계수 } b = S_{(xy)} / S_{(yy)}$$

[참조] 수치변환에 의한 회귀식의 계산의 경우

* 위의 회귀선에서의 x 의 평균 $\bar{x}$ 와 y 의 평균 $\bar{y}$, 회귀계수 b 의 값을 구할 때 계산의 단순화를 위해 $X = (x - x_0) \times h$, $Y = (y - y_0) \times g$ 에 의해 수치변환한 값에 의한 경우에는 다음의 식으로 계산한 값을 식 (5.74)에 대입시킨다.

$$\bar{x} = x_0 + \frac{\sum X}{n} \times \frac{1}{h}, \quad \bar{y} = y_0 + \frac{\sum Y}{n} \times \frac{1}{g}$$

3.2.6 결정계수(기여율) R^2 [기지1회] [품기2회]

(1) 결정계수의 정의 [품기1회]

* 상관계수(r)에 의거하여 설명력을 보는 결정계수(기여율) R^2 은 다음 식에 의하여 구한다.

$$R^2 = (r)^2 = \left(\frac{S_{(xy)}}{\sqrt{S_{(xx)} \cdot S_{(yy)}}} \right)^2 \quad (5.76)$$

* 결정계수(기여율)의 의미는 총변동(SST ; Sum of Squared Total) 중에서 기여율에 해당하는 수치가 회귀에 의하여 설명되는 변동(SSR ; Sum of Squared Regression)을 의미하고, 나머지 값은 잔차변동(SSE ; Sum of Squared Error)을 나타낸다. $R^2 > 0.65$ 이면 회귀식이 자료를 설명하기에 좋다고 판정한다.

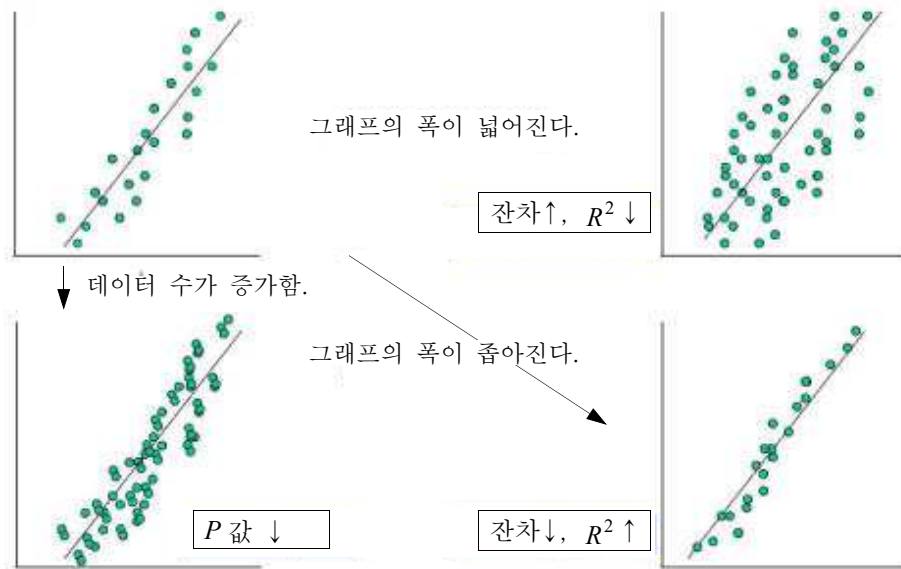
$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (5.77)$$

결정계수는 두 변수간의 선형관계 정도가 높으면(관측치들이 가까이 모여 있는 것을 의미함) 결정계수는 1에 가까워 진다. 특히 단순회귀 모형에서 상관계수는 $r = \pm \sqrt{R^2}$ 이 성립한다.

- * 유의하지 않는 설명변수라도 모형에 삽입되면 R^2 값은 올라가므로 모형의 유의성 비교에는 R^2 이 아닌 설명변수의 개수에 의해 조정된 수정(adjusted) 결정계수 $R^2_{(adj)}$ 가 사용된다.
- * $R^2_{(adj)}$ 은 독립변수(인자)의 수와 데이터의 수를 고려한 결정계수이고, $R^2_{(adj)} < R^2$ 이며, 다중회귀에 특히 유효하다. R^2 과 $R^2_{(adj)}$ 의 차이는 작을수록 좋다.

(2) 산점도의 폭, 잔차 등과 결정계수의 관계

- ① 상관관계↑ → 산점도 폭이 좁음 → 결정계수 R^2 ↑
 - ② 잔차↑ → 상관관계↓ → 결정계수 R^2 ↓
- * 데이터의 수, 산점도 폭, P값, 잔차, R^2 등의 관계를 정리하면 [그림 5.15]와 같다.



[그림 5.15] 데이터의 수, 산점도 폭, P값, 잔차, R^2 등의 관계

(3) 결정계수와 P값과의 관계

- * P value(P값)는 통계적인 의미를 보는 것으로서, 유의확률을 의미한다. Minitab 통계프로그램 등으로 구해진 P value가 5% 이하이면 귀무가설이 정말 틀렸다고 결론을 내린다. 즉 $p < 0.05$ 이면, H_1 , 즉 선형회귀모형이 존재함을 의미한다. P value는 작은 것이 좋고, $R^2_{(adj)}$ 는 큰 것이 좋다.
- * 결정계수 $R^2_{(adj)}$ 와 P 값과의 관계를 정리하면 <표 5.7>과 같다.

<표 5.7> 결정계수 R^2 과 p 값과의 관계

	$R^2_{(adj)} \uparrow$	$R^2_{(adj)} \downarrow$
$p \downarrow$	통계적 의미가 있고, 설명력이 크다. 이상적인 모형 → 중요한 인자를 찾았다.	통계적 의미가 있고, 설명력이 작다. 데이터의 수가 많을 경우에 발생 → 이상값을 찾는다. 다른 중요한 인자를 찾는다.
$p \uparrow$	통계적 의미가 없고, 설명력이 크다. 데이터의 수가 작을 때나 이상값이 존재할 경우에 발생 → 데이터의 수를 늘린다. 이상값을 찾는다.	통계적 의미가 없고, 설명력이 작다. 선형적인 관계가 없다. → 비선형 회귀분석 다른 중요한 인자를 찾는다.

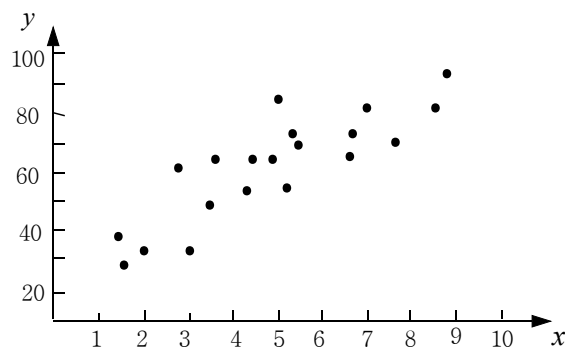
예제 5.25 화학반응공정에서 전해시에 첨가물의 양(x g)과 수량(y %)의 관계 데이터를 20개 취하였다. x 와 y 의 관계를 산점도를 그린 후 상관의 유의성 검정을 $\rho=0$ 인 경우에 대해 실시하고, 이를 이용하여 $\rho \neq 0$ 인 경우, 즉 종래의 상관계수(모상관계수)는 $\rho_0 = 0.75$ 일 때, 이번에 (공정에 약간 변화가 있는 후) 주어진 표와 같이 새로 측정한 데이터에 의해 계산한 상관계수에 의해 모상관계수가 달라졌다고 할 수 있는가에 대한 검정과 모상관계수의 추정을 실시하고, 상관계수의 유의성검정 결과가 유의인 경우 x 에 대한 y 의 회귀선과 기여율을 구하여라.

x	y	x	y	x	y	x	y
1.3	40	3.4	50	5.0	87	6.8	75
1.5	30	3.5	66	5.1	57	7.0	83
2.0	35	4.2	56	5.3	75	7.6	72
2.8	62	4.3	66	5.5	71	8.5	83
3.0	35	4.9	66	6.6	68	8.8	94

해설 [품기 2023 2회차]

(1) 산점도의 작성

* [그림 1]의 산점도를 보면 플로팅된 점이 이종(異種)의 모집단에 속해 있다고 생각되는 점은 없고, 전체적으로 양(+)상관으로 볼 수 있다.



[그림 1] 산점도

제 6 장

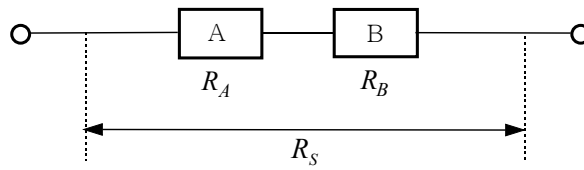
신뢰성공학

1. 신뢰성의 기초 / 6-02
 2. 고장률과 고장확률밀도함수 / 6-07
 3. 신뢰성추정 / 6-12
 4. 신뢰성 샘플링검사 / 6-16
 5. 시스템의 신뢰도 / 6-19
 6. 스트레스 · 강도 모델 및 안전계수 / 6-25
 7. 보전성과 가동성 / 6-27
 8. 고장해석기법 FMEA · FTA / 6-32
 9. 신뢰성 설계기술 / 6-42
 10. 기출 · 예상문제 및 착안점 / 6-48
-

5. 시스템의 신뢰도 [공기3회] [기지1회] [품기4회]

5.1 직렬결합모델의 신뢰도 [경지1회] [품기6회]

- * 조립품이나 컴포넌트·시스템 등을 구성하고 있는 여러 개의 소자나 부품 중 어느 하나라도 고장이 나게 되면 시스템 전체가 기능을 상실하게 되도록 소자나 부품이 결합된 것을 직렬결합모델이라고 한다.
- * [그림 6.3]과 같이 A, B 2개의 부품이 직렬결합모델로 되어 있으면 이 기기가 제대로 기능을 발휘하기 위해서는 A와 B 2개의 부품이 모두 정상작동하여야 한다.



[그림 6.3] 직렬결합모델

- * 직렬결합모델의 전체 시스템이 작동하는 전체 신뢰도 R_S 는 다음과 같다.

$$R_S = P_r(A \text{ AND } B) = P_r(A \cap B)$$

- * 만일 A, B가 서로 독립사상이면

$$R_S = P_r(A) \cdot P_r(B) = R_A \cdot R_B \quad (6.56)$$

여기서, $P_r(A)$ 는 부품 A의 신뢰도 R_A , $P_r(B)$ 는 부품 B의 신뢰도 R_B 를 의미한다.

일반적으로 n 개 부품이 직렬결합시의 시스템의 신뢰도 R_S 는 다음 식과 같이 된다.

$$R_S = R_1 \cdot R_2 \cdots R_n = \prod_{i=1}^n R_i = e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)t} \quad (6.57)$$

- * 전체의 고장률 λ_S 는 각 부품의 고장밀도함수가 지수분포에 따르는 경우

$$R_S = e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)t} = e^{-\lambda_S \cdot t}$$

의 관계로부터 λ_S 는 다음의 식으로 구해진다.

$$\lambda_S = \sum_{i=1}^n \lambda_i = \lambda_1 + \lambda_2 + \cdots + \lambda_n \quad (6.58)$$

- * 한편, 시스템의 신뢰성을 구할 때 중요한 것은 t 시간 후의 잔존확률인 신뢰성 $R(t)$ 로서, 다음의 근사식에 의해 $R(t)$ 의 값을 구하면 편리하다.

$$R(t) = e^{-\lambda t} = e^{-\frac{t}{\text{MTBF}}} \doteq 1 - \frac{t}{\text{MTBF}} \quad (6.59)$$

여기서, t = 사용시간, $\lambda = 1/\text{MTBF}$

* 또한, 고장률 λ_i 인 부품이 여러 개 직렬연결된 시스템의 MTBF는 다음과 같이 구한다.

$$\text{MTBF}_S = \frac{1}{\lambda_S} \quad (6.60)$$

$$\text{여기서, } \lambda_S = \sum_{i=1}^n \lambda_i = \lambda_1 + \lambda_2 + \dots + \lambda_n$$

예제 6.8 각 부품의 평균고장률이 0.002(/시간)인 부품 7개가 동시에 모두 작동해야만 기능을 발휘하는 기기가 있다. 이 기기의 평균수명을 구하라.

해설

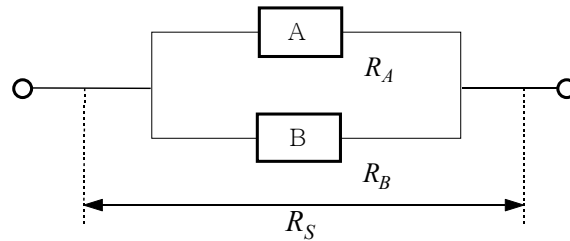
$$\text{기기의 평균수명 } \text{MTBF}_S (\text{혹은 } \theta_S) \text{는 } \text{MTBF}_S = \frac{1}{\lambda_S} = \frac{1}{0.014} = 71.4 \text{ (시간)}$$

여기서, 부품 7개의 직렬결합모델이므로 $\lambda_S = \lambda_1 + \lambda_2 + \dots + \lambda_7 = 7 \times 0.002 = 0.014 (/시간)$

5.2 병렬결합모델의 신뢰도 [경지1회] [품기6회]

* [그림 6.4]는 부품을 여분으로 한 개 더 부가시켜서 부품 2개 중 어느 한 개만 작동하면 전체가 기능을 발휘할 수 있도록 결합한 것이다.

* 이와 같은 설계를 병렬설계라 하며, 용장(冗長)설계, redundancy설계, 여유설계, 과잉설계 등으로도 불린다. 병렬설계를 하면 전체의 신뢰도를 크게 증대시킬 수 있다.



[그림 6.4] 병렬결합모델

* 병렬결합모델의 전체 시스템이 작동하는 전체 신뢰도 R_S 는 다음과 같다.

$$\begin{aligned} R_S &= P_r(A \text{ OR } B) = P_r(A \cup B) \\ &= P_r(A) + P_r(B) - P_r(A) \cdot P_r(B) \\ &= R_A + R_B - R_A \cdot R_B \end{aligned} \quad (6.61)$$

* 또한, $R(t) + F(t) = 1$ 의 관계에서 다음 식으로 될 수 있다.

$$\begin{aligned} R_S &= (1 - F_A) + (1 - F_B) - (1 - F_A)(1 - F_B) \\ &= 1 - F_A F_B = 1 - (1 - R_A)(1 - R_B) \end{aligned} \quad (6.62)$$

* 일반적으로 n 개의 부품이 병렬결합시 시스템 전체의 신뢰도 R_S 는 다음과 같다.

$$R_S = 1 - \prod_{i=1}^n F_i = 1 - \prod_{i=1}^n (1 - R_i) \quad (6.63)$$

* 병렬결합모델의 시스템 MTBF는 식 (6.61)에서

$$R(t) = e^{-\lambda t}, R_1(t) = e^{-\lambda_1 t}, R_2(t) = e^{-\lambda_2 t}$$

의 지수분포를 따를 때, 이것으로 대치하고 양변을 적분하여 MTBF를 구한다.

$$\int_0^{\infty} R(t) dt = \int_0^{\infty} R_1(t) dt + \int_0^{\infty} R_2(t) dt - \int_0^{\infty} R_1(t) R_2(t) dt$$

$$\int_0^{\infty} e^{-\lambda t} dt = \int_0^{\infty} e^{-\lambda_1 t} dt + \int_0^{\infty} e^{-\lambda_2 t} dt - \int_0^{\infty} e^{-(\lambda_1 + \lambda_2)t} dt$$

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2}$$

* 이에 의거 병렬결합시스템의 $MTBF_S$ 는 다음과 같이 구한다.

$$MTBF_S = \frac{1}{\lambda_S} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2} - \frac{1}{\lambda_1 + \lambda_2} \quad (6.64)$$

* 상기 식에서 $\lambda_1 = \lambda_2 = \lambda_0$ 라면 $MTBF_S$ 는 다음과 같이 주어진다.

$$MTBF_S = \frac{3}{2} \cdot \frac{1}{\lambda_0} \quad (6.65)$$

* 일반적으로 $\lambda_1 = \lambda_2 = \dots = \lambda_n = \lambda_0$ 인 n 개 구성부품의 병렬결합시스템의 $MTBF_S$ 는 다음 식으로 구해진다.

$$MTBF_S = \frac{1}{\lambda_0} + \frac{1}{2\lambda_0} + \dots + \frac{1}{n\lambda_0} \quad (6.66)$$

* 또는, 평균수명 $\theta_0 = 1/\lambda_0$ 이라면 시스템의 $MTBF_S$ 는 다음과 같다.

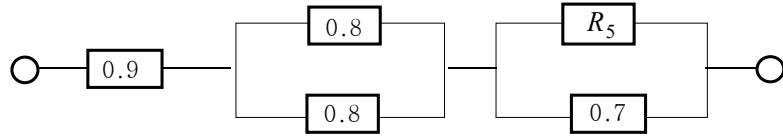
$$MTBF_S = \sum_{i=1}^n \frac{\theta_0}{i} \quad (6.67)$$

* 그리고, 시스템의 고장률 λ_S 는 다음 식으로 구해진다.

$$\lambda_S = \frac{1}{MTBF_S} = \frac{1}{\frac{1}{\lambda_0} + \frac{1}{2\lambda_0} + \dots + \frac{1}{n\lambda_0}} = \frac{1}{\sum_{i=1}^n \frac{\theta_0}{i}} \quad (6.68)$$

여기서, $\theta_0 = 1/\lambda_0$ 인 경우이다.

예제 6.9 다음과 같이 구성된 시스템이 있다. 전체 신뢰도 $R_S=0.85$ 로 하고자 할 때 R_5 의 신뢰도는 얼마인가? [기사 기출]



해설 [품기5회]

시스템의 신뢰도 : $R_S = R_1 \times R_{S1} \times R_{S2} = 0.9 \times 0.96 \times (0.7 + 0.3R_5) = 0.85$

이를 풀면 $R_5 = 0.946$

여기서, 신뢰도 0.8의 부품 2개로 병렬결합된 부분의 신뢰도 :

$$R_{S1} = 1 - (1 - 0.8)^2 = 0.96$$

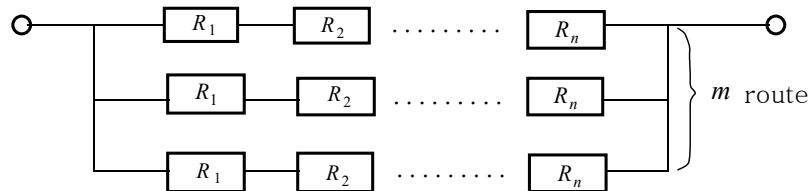
신뢰도 R_5 , 0.7의 부품 2개로 병렬결합된 부분의 신뢰도 :

$$R_{S2} = 1 - (1 - R_5)(1 - 0.7) = 0.7 + 0.3R_5$$

5.3 특수결합 모델의 시스템 신뢰도 [품기3회]

5.3.1 m route 시스템 신뢰도

* 병렬결합모델의 특수한 경우인 m route 시스템 신뢰도는 다음 그림과 같은 구조로 된다.



[그림 6.5] m route의 서브시스템 결합

* 직렬인 경우의 서브시스템에서의 신뢰성은 다음과 같다.

$$R_{SS} = R_1 \cdot R_2 \cdots R_n = \prod_{i=1}^n R_i \quad (6.69)$$

* 병렬인 경우의 m route에서의 신뢰성은 다음과 같다.

$$R_S = 1 - (1 - R_{SS})^m = 1 - (1 - \prod_{i=1}^n R_i)^m \quad (6.70)$$

여기서, $R_1 = R_2 = \cdots = R_n = R$ 일 경우에는 다음 식으로 된다.

$$R_S = 1 - (1 - R^n)^m \quad (6.71)$$

예제 6.10 $\lambda = 0.002$ (/시간)인 부품 3개가 병렬결합되어 있다. 전체 평균고장률 λ_s 는 얼마인가?

해설

$$\text{MTBF}_s = \frac{1}{\lambda_s} = \frac{1}{\lambda} + \frac{1}{2\lambda} + \frac{1}{3\lambda} = \frac{11}{6\lambda} \rightarrow \lambda_s = \frac{6\lambda}{11} = \frac{6 \times 0.002}{11} = 1.09 \times 10^{-3} \text{ (/시간)}$$

5.3.2 n 중 k (k out of n) 시스템 신뢰도 [품기1회]

* n 개 중 k 개만 작동하면($1 \leq k \leq n$) 시스템이 작동하는 경우 각 구성품의 신뢰도를 R 이라 하면 시스템의 신뢰도 R_s 는 다음과 같다.

$$R_s = \sum_{i=k}^n \binom{n}{i} R^i (1-R)^{n-i} \quad (6.72)$$

* 만일 $R(t) = e^{-\lambda t} = e^{-t/\theta}$ 의 지수분포에 따른다면 시스템의 MTBF_s 는 다음과 같이 된다.

$$\text{MTBF}_s (= \theta_s) = \sum_{i=k}^n \frac{\theta_0}{i} = \text{MTBF}_0 \left(\frac{1}{k} + \dots + \frac{1}{n} \right) \quad (6.73)$$

예제 6.11 신뢰도가 0.99인 미사일 4개가 설치된 미사일발사 시스템이 있음. 그런데 4개의 미사일 중 3개만 작동하면 이 미사일발사 시스템은 임무수행이 가능함. 이 4 중 3개 미사일 시스템의 신뢰도를 구하여라.

해설

$n = 4, k = 3, R = 0.99$ 이므로, 이 4 중 3개 미사일 시스템의 신뢰도 R_M 은 다음과 같다.

$$\begin{aligned} R_M &= \sum_{i=3}^4 \binom{4}{i} (0.99)^i (1-0.99)^{4-i} = \binom{4}{3} (0.99)^3 (0.01)^1 + \binom{4}{4} (0.99)^4 (0.01)^0 \\ &= (4)(0.970)(0.01) + (1)(0.960) = 0.0388 + 0.960 = 0.9988 \end{aligned}$$

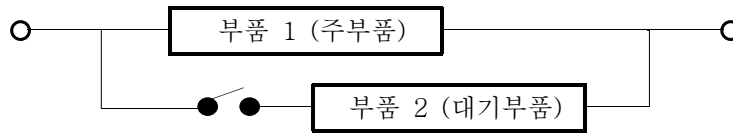
$$\text{여기서, } \binom{4}{3} = {}_4C_3 = \frac{4!}{3!(4-3)!} = \frac{4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 1} = 4$$

5.3.3 대기결합구조의 시스템 신뢰도

(1) 대기결합구조의 의미

* 여분의 부품이 처음부터 병렬로 연결되어 있지 않고 처음에는 주부품이 그의 기능을 수행하다가 이것이 고장나면 여분의 부품인 대기부품이 그의 기능을 이어받아 계속 수행하도록 결합되어 있는 것을 말한다.

* 그리고 대기결합구조로 시스템을 구성하면 신뢰도가 증가하게 되는 성질을 대기리던던시(stand-by redundancy)라고 한다.



[그림 6.6] 대기결합구조

(2) 대기결합구조의 시스템 신뢰도

* 대기결합구조의 시스템 신뢰도에 대해 알려진 결과로서의 식만을 간단히 보도록 한다.

- ① 주부품인 부품 1의 신뢰도 R_1 과 대기부품인 부품 2의 신뢰도 R_2 가 각각 평균고장률이 λ_1 과 λ_2 인 지수분포에 따르고, 이 시스템의 임무기간(mission time)이 T 라 할 때

$$R_S = e^{-\lambda_1 T} + \int_0^T \lambda_1 e^{-\lambda_1 t} \cdot e^{-\lambda_2 (T-t)} dt = \frac{1}{\lambda_1 - \lambda_2} [\lambda_1 e^{-\lambda_2 T} - \lambda_2 e^{-\lambda_1 T}] \quad (6.74)$$

- ② $\lambda_1 = \lambda_2 = \lambda$ 인 경우는 주부품 한 개일 경우보다 MTBF_S는 2배, 신뢰도 R_S 는 $(1 + \lambda T)$ 배가 증가한다. 즉,

$$\textcircled{1} R_S = e^{-\lambda T} + \int_0^T \lambda e^{-\lambda t} \cdot e^{-\lambda (T-t)} dt = e^{-\lambda T} (1 + \lambda T) \quad (6.75)$$

$$\textcircled{2} \text{ 전환스위치 신뢰도 } R_{SW} \text{를 고려할 경우 } R_S = e^{-\lambda T} (1 + R_{SW} \cdot \lambda T) \quad (6.76)$$

$$\textcircled{3} \text{ MTBF}_S = \int_0^T e^{-\lambda t} (1 + \lambda t) dt = \frac{2}{\lambda} \quad (6.77)$$

- ③ 일반적으로 $\lambda_1 = \lambda_2 = \dots = \lambda_n = \lambda$ 인 n 개의 부품이 대기결합구조로 구성된 경우

$$R_S = e^{-\lambda t} \left[1 + \lambda t + \frac{(\lambda t)^2}{2!} + \dots + \frac{(\lambda t)^{n-1}}{(n-1)!} \right] \quad (6.78)$$

$$\text{MTBF}_S = \left(\frac{1}{\lambda} + \frac{1}{\lambda} + \dots \right)_{n\text{개}} = \frac{n}{\lambda} \quad (6.79)$$

예제 6.12 평균고장률이 0.001이고 $T=24$ 일 때 신뢰도가 0.9763인 주레이더가 그림과 같이 대기결합구조로 결합되어 있다. 이 대기결합 레이더 시스템의 신뢰도를 구하라.



해설

* $\lambda = 0.001$, $T=24$ 이므로 대기결합 레이더 시스템의 신뢰도 R_{S1} 은 다음과 같다.

$$R_{S1} = e^{-\lambda T} (1 + \lambda T) = e^{-0.001 \times 24} (1 + 0.001 \times 24) = 0.9763(1 + 0.0024) = 0.9997$$