※ 통계학 요약(시험 대비 위주로 작성)

* 확률의 공리적 정리 -> 0≤P(A)≤1 , $P(\overline{A})=1-P(A)$, $P(A \sqcup B)=P(A)+P(B)-P(A \sqcap B)$
* 통계적 독립을 표현하는 식

$$P(X \sqcup Y)=P(X)+P(Y) \,/\, P(X \sqcap Y)=P(X)P(Y)$$
$$P(Y|X)=P(Y) \,/\, P(X|Y)=P(X)$$
$$f_{xy}=f_x f_y \quad / f_{x|y}=f_x \,/ f_{y|x}=f_y$$

* 이항분포의 기대값, 분산 - 기대값 : $E(x)=np\,(n=\text{횟수}, p=\text{특정확률값})$ / 분산 : $Var(x)=np(1-p)$
* 변동계수(분산계수) $CV=\frac{\sigma}{\mu(=E(R))}$
* 결합확률분포의 공분산 $Cov(X,Y)=E(XY)-E(X)E(Y)$ <- 결합 확률표 주어졌을 때 E(XY)는 표 가운에의 결합확률을 의미, E(X)와 E(Y)는 각각의 한계확률(가장자리)을 의미

* 확률분포

	의미, 특성	검증 방법
이항분포	이산확률변수의 확률분포	
정규분포	연속활률변수의 대표적 확률분포, 모집단의 분포	$Z=\frac{X-\mu}{\sigma/\sqrt{N}}$
T 분포	표본의 평균에 대한 분포 정규분포보다 fat tail한 leptokurtic 분포	$T=\frac{\overline{X}-\mu}{S/\sqrt{N}}$ (S: 표본의 표준편차σ, N: 표본의 개수, $\overline{X}$: 관측결과의 평균)
X^2 분포	표본의 분산에 대한 분포	$\chi^2=(n-1)\times\frac{S^2}{\sigma^2}$ (자유도조정 : $n-1$)
F 분포	X^2 분포와 유사	

* 추정량(Estimator)의 특성

구분	특성	내용
소표본(30개 기준)	불편성(unbiased)	* $E(\hat{\theta})=\theta$ 추정치(θ)는 실제치($\hat{\theta}$)와 같아야, 효율성의 전제조건
	효율성(efficiency)	* 불편성을 갖는 추정량 중 가장 작은 분산을 갖는 추정량
	평균제곱오차(MSE)	* $MSE(\hat{\theta})=Var(\hat{\theta})+b(\theta)^2$
	BLUE	* 모든 <u>선형</u>불편추정량 중 최소분산을 갖는 추정량
대표본	점근적 불편성	* asymptotic 데이터량이 증가하면..
	일치성(일관성)	
	점근적 정규성	* 대수의 법칙(LLN)
	체비셰프 부등호	* $1-\frac{1}{k^2}$(k : 시그마, 평균에서 몇 시그마 떨어져 있는가)
체크사항		* 불편성 조건은 소표본과 대표본 모두 공통사항이다. * 대표본에만 있는 특성은 점근적 정규성, 일치성. 단, 넓게 해석하면 대표본의 일치성은 소표본의 효율성과 동일한 것으로 해석되므로, 대표본에만 유일한 특성은 점근적 정규성이다.

* 모수 추정

	의미	검정방법, 처리방법 등
자기상관	오차항(잔차)간의 상관 관계	* D-W검정 등. * D-W검정. $d = 2(1-\rho)$ (d=2->로우=0, d=0->로우=1, d=4->로우=-1) * 처리방법 -> 차분 $Y_t - \rho Y_{t-1}$=Y 새로운 Y값
다중공선성	독립변수간 높은 상관관계로 종속변수에 영향	* 다중공선성 의심 t(샘플) 값이 작은데 R^2값이 높은 경우 설명변수간 상관성 높은 경우, 회귀계수 값이 모형 설정에 매우 민감한 경우 * 처리방법 표본 크기(데이터) 증가/ 특정 설명변수 제거/ 모형외 정보 사용
이분산	독립변수 값 변화 -> 종속변수의 분산이 동일한 값을 취하지 않고 변하는 것.	* 원인 시계열 분포가 fat tail의 leptokurtic 한 분포. *검정방법 -> Goldfeld-Quandt 검정/ Breusch-Pagan 검정 / White 검정
회귀분석	두 독립변수 간 인과관계 함수관계 규명. 단순, 다중회귀분석	* 적합도 측정 TSS(총오차σ_i^2)=RSS(회귀식으로 설명된 오차 $\beta_i^2\sigma_m^2$)+ SSE(회귀식의 남은오차 $Var(\epsilon_i)$) * 결정계수 $R^2 = \frac{RSS}{TSS} = \frac{\beta_i^2\sigma_m^2}{\sigma_i^2} = \rho^2$ (값이 높을수록 설명력↑)
시계열 분석	한 변수의 추세변동, 주기적변동, 계절적 변동을 분해, 외생적 요인 제거	* 외삽법: 외생적 요인 제거위한 필터 적용하여 이전의 상태로. * 평활법: 불규칙 변동 등 제거위해 평활한(평균) 시계열 도출 * ARMA(Box-Jenkins) AR- 과거시차변수 $Y_t = C + \phi Y_{t-1} + \epsilon_t$ MA- 과거의 잔차 $Y_t = C + \epsilon_t - \theta_1 \epsilon_{t-1}$
ARCH, GARCH	이분산, 자기상관이 동시에 발생하는 경우, 두 문제를 동시에 처리하는 방법	*ARCH : 꼬리가 두터운 금융시계열을 모형화 하는데 유용 * GARCH(1,1) 의 비조건부 분산 $\sigma^2 = \frac{\alpha_0}{(1-\alpha_1-\beta_1)}$ * p=q=1 경우 GARCH의 약안전성조건 $\alpha_1 + \beta_1 < 1$ * GARCH(1,1) $\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 \epsilon_{t-1}^2$ 변동성의 지속 $\lambda = (\alpha_1 + \beta_1)$ $\lambda \fallingdotseq 1$일수록 변동성지속 / $\lambda < 1$으면 약안전성조건충족