14] Want $E_i \geq 5$, $i=1,\dots,k$. Combine neighboring classes if necessary, unless the Exam problem says to use the given classes.

ex] <u>HW6 #4</u> combines the 3, 4+ classes into a 3+ class, so use the 3+ class.

15] $$Q = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j} = \sum_{j=1}^{k} \frac{O_j^2 - 2O_jE_j + E_j^2}{E_j}$$

$$= \sum_{j=1}^{k} \frac{O_j^2}{E_j} - 2\underbrace{\sum_{j=1}^{k} O_j}_{n} + \underbrace{\sum_{j=1}^{k} E_j}_{n} = \left(\sum_{j=1}^{k} \frac{O_j^2}{E_j}\right) - n$$

16] The df $k-r-1$ occurs if the n claims are divided into k groups since then $\sum_{i=1}^{k} n_i = n$.

If claims are just observed in k groups (so n is a RV or groups are independent), df $= k-r$. } rather rare ex} Test Poisson dist for the following table except data must be grouped so each group has $E_i \geq 5$.

# claims	
0	410
1	75
2	10
3	5
4+	0
total	500

	O_i	E_i
0	410	401.2594
1	75	88.2771
2+	15	10.4635
		500

$$\hat{\lambda} = \frac{75(1) + 10(2) + 5(3)}{500} = 0.2200 = \bar{x}, \ r=1$$

$$500\,\hat{p}_0 = 500\,e^{-0.22} = 401.2594$$

$$500\,\hat{p}_1 = 500\,e^{-0.22}(0.22) = 88.2771$$

$$500\,\hat{p}_2 = 500\,e^{-0.22}(0.22)^2/2 = 9.7105$$

$$500\,\hat{p}_3 = 500\,e^{-0.22}(0.22)^3/6 = 0.7121$$

$$500\,\hat{p}_{4+} = \quad = 0.0409$$

$$500 \qquad \text{so sum} = 500$$

$$p_k = P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

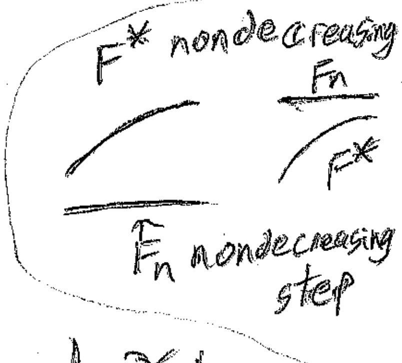
df	α .05 / .95	.025 / .975
1	3.841	5.024

$Q = 4.1541$

$df = 3-1-1 = 1$

reject H_0 at 5% level but not at 2.5% level

17] p448 Kolmogorov Smirnov test statistic: Let $X_1, \ldots, X_n$ be iid or possibly truncated and censored. Let d be the truncation point ($d = 0$ no truncation), and let U be the censoring point ($U = \infty$ no censoring). Then the test statistic is



$$D = \max_{d \le x \le U} \left| F_n(x) - \underbrace{F^*(x)}_{\text{fitted cdf}} \right|.$$

The max occurs at an observed x_i where $F_n(x)$ jumps, and is $\max_{x_i} \left(|F_n(x_i) - F^*(x_i)|, |F_n(x_i-) - F^*(x_i)| \right)$.

18) *critical values

α	0.1	0.05	0.01
	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

19) know 4 step Kolmogorov Smirnov test:

i) Ho fitted distribution is good HA not Ho

ii) D

iii) reject Ho if D > critical value
Otherwise fail to reject Ho.

(v) nontechnical conclusion

reject Ho: fitted dist is not good

fail to reject Ho; fitted dist is good

(or there is not enough evidence to conclude that the fitted dist is not good.

20] *Variant: Find smallest significance level for which Ho is rejected.

	α	0.1	0.05	0.01
eg $n=100$	critical value	0.122	0.136	0.163

if $D = 0.14$ then reject Ho at 0.05 = 5% significance but not at 0.01 = 1% significance.

$$F_n(x) = \frac{\# x_i \le x}{n} \qquad F_n(x-) = \frac{\# x_i < x}{n}$$

21]

ordered

x_1	$1/n$	0
x_2	$2/n$	$1/n$
x_3	$3/n$	$2/n$
x_4	$4/n$	$3/n$
$\vdots$	$\vdots$	$\vdots$
x_{n-1}	$(n-1)/n$	$(n-2)/n$
x_n	n/n	$(n-1)/n$

assuming no ties

ex) C40 Payments

29, 64, 90, 135, 182 are assumed to follow an exponential distribution. Estimate the mean with the method of moments and do the 4 step Kolmogorov–Smirnov test.

soln $\hat{\theta}_{MME} = \hat{\theta}_{MLE} = \bar{x} = \frac{\sum x_i}{n} = \frac{500}{5} = 100$

ordered x_i	$\frac{1}{n} \# x_i \le x = F_n(x)$	$\frac{1}{n} \# x_i < x = F_n(x_i-)$	$1-e^{-x/100} = F(x)$	$\max(\lvert F_n(x)-F(x)\rvert, \lvert F_n(x-)-F(x)\rvert)$	
29	1/5	0	0.2517	0.2517	$= .2517-0$ $> .2517-.2$
64	2/5	1/5	0.4727	0.2727	$= D$
90	3/5	2/5	0.5934	0.1934	
135	4/5	3/5	0.7407	0.1407	
182	5/5	4/5	0.8378	0.1602	$= 1-.8378$ $> .8378-.8$

i) H_0 the fitted model is good H_A not H_0

ii) $D = 0.2727$ use $\alpha = .05$ if α is not given

iii) critical value $= \frac{1.36}{\sqrt{n}} = \frac{1.36}{\sqrt{5}} = 0.6082 > D$

$D < .6082 =$ crit. value

iv) fail to reject H_0

(v) not enough evidence to conclude that the fitted distribution is bad.

22) p450 Suppose data is truncated at d and censored at u. The Anderson Darling test statistic

$$A^2 = -nF^*(u) + n\left[\sum_{j=0}^{k}\left[S_n(y_j)\right]^2\left[\ln\left(S^*(y_j)\right) - \ln\left(S^*(y_{j+1})\right)\right]\right] +$$

$$n\left[\underbrace{\sum_{j=1}^{k}}_{F_n^*(0)=0}\left[F_n(y_j)\right]^2\left[\ln\left(F^*(y_{j+1})\right) - \ln\left(F^*(y_j)\right)\right]\right].$$

$S^* = 1-F^*$, $S_n = 1-F_n$

23) The Anderson Darling test is a competitor of the χ^2 test and Kolmogorov Smirnov test, but computations rarely appear on exams.

24) <u>Know for multiple choice questions</u>

	Kolmogorov Smirnov	Anderson Darling	χ^2
i)	indiv data	indiv data	indiv or grouped data
ii)	continuous fits	continuous fits	continuous or discrete fits
iii)	lower the critical value if $u<\infty$	lower the critical value if $u<\infty$	no adjustment of the critical value if $u<\infty$
iv)	lower the critical value if $r>0$	lower the critical value if $r>0$	$df = k-r-1$ adjusts for $r\geq 0$
v)	critical value $\downarrow$ as $n\uparrow$	critical value free of n	critical value free of sample size n
vi)	no discretization ← continuous fits	no discretization	discretization with grouped data
vii)	uniform weight on all parts of distribution	higher weight on tails of dist	higher weight on intervals with lower prob (eg right tail has low prob)

25] *Likelihood Ratio Test LRT

H_0 data came from population with distribution A (pdf or pmf f_A)

H_1 data came from population with distribution B (pdf or pmf f_B)

* where model A is a special case of model B.

26] Let Θ_0 be the parameter space for H_0 and let Θ_1 be parameter space for H_1.

Let $\hat{\underline{\theta}}_0$ be the MLE where $\hat{\underline{\theta}}_0 \in \overset{A}{\Theta_0}$.

Let $\hat{\underline{\theta}}_1$ be the MLE where $\hat{\underline{\theta}}_1 \in \overset{B}{\Theta_1}$.

Let $L_0 = L(\hat{\underline{\theta}}_0)$ and $L_1 = L(\hat{\underline{\theta}}_1)$.

The LRT test statistic is $T = -2\ln\left(\frac{L_0}{L_1}\right)$

$$= 2\ln\left(\frac{L_1}{L_0}\right) = 2\left[\ln(L_1) - \ln(L_0)\right].$$

df = d = # "free" parameters in B − # "free" parameters in A
$= d_B - d_A$.

Reject H_0 if $T > \chi^2_d$ critical value.

So $\alpha = P(T > \chi^2_{d,\underbrace{1-\alpha}_{P \text{ on } \chi^2 \text{ table}}}) = P(\chi^2_d > \chi^2_{d,1-\alpha})$.

45.5

A "free" parameter is not specified, so must be estimated using MLE.

27] Need " $df \geq 1$. So model B needs to have more free parameters than model A. This makes sense because A is a special case of B for the χ^2 approx to hold.

i) A $EXP(\theta) \sim$ Weibull $(\tau = 1, \theta)$ vs B Weibull(τ, θ)

ii) A $EXP(\theta) \sim$ Gamma$(\alpha = 1, \theta)$ vs B Gamma(α, θ)

iii) A Geometric $(\beta) = NB(r = 1, \beta)$
vs B Negative Binomial (r, β)

iv) eg A Pareto $(\alpha = 2, \beta)$
vs B Pareto (α, β)

one of 2 parameters estimated for A

ex] H_0 dist is gamma with mean μ_0
H_1 dist is gamma(α, β)

under H_0 $\widehat{EX} = \hat{\alpha}\hat{\beta} = \mu_0$

$\hat{\alpha} = \frac{\mu_0}{\hat{\beta}}$

so $d_A = 1$, $d_B = 2$.
reject $H_0 \Rightarrow$ use dist given in H_1 (model B)

28] Since model A is a special case of model B,

$\ln L_B \geq \ln L_A$.

For ex, $A \sim EXP(\theta) = \text{weibull}(\tau=1, \theta)$.

$$\ln(L_A) = \ln f_{W(\tau=1, \theta=\bar{x})}(\underline{x}) \leq \ln f_{W(\tau=\hat{\tau}, \theta=\hat{\theta})}(\underline{x}) = \ln(L_B)$$

Since $(\hat{\tau}, \hat{\theta})$ maximizes $\ln f_{W(\tau,\theta)}(\underline{x}) = \ln(L(\tau,\theta))$.

a model selection method

29] Suppose model A is not a subset of model B or there are models

$A_1, A_2, \ldots, A_K$.

	# of parameters	(minimum $-$log likelihood) or maximal log likelihood
eg		-321.32
	1	-319.93
	2	-319.12
	3	-318.12
	4	

Select for every number of parameters,
could have 3 models with 3 parameters

the model with the highest log likelihood (so lowest − loglikelihood). 46.5

Suppose $\alpha = .05$ is the significance level.

In order to prefer the best 2 parameter model over the best 1 parameter model, need $2(\ln L_2 - \ln L_1) \geq \chi^2_{1,.95} = 3.841$.

If the best 2 parameter model is not good, need $2(\ln L_3 - \ln L_1) \geq \chi^2_{2,.95} = 5.991$

and so on.

If the 2 parameter model is preferred, then start over comparing 3, 4, ... parameter models to the 2 parameter model

eg need $2(\ln(L_3) - \ln(L_2)) \geq \chi^2_{1,.95} = 3.841$

If the 3 parameter model is not good,
need $2(\ln(L_4) - \ln(L_2)) \geq \chi^2_{2,.95} = 5.991$

and so on. ex] For models with up to 5 parameters, if L_1 is preferred to L_2 and L_3, but L_4 is preferred to L_1 and L_5, then the 4 parameter model is best.

a) For the table below 29) with $\alpha = .05$, M404 47

$2\,(\ln L_2 - \ln L_1) = 2(-319.93 + 321.32) =$

$2.78 < \chi^2_{1,.95} = 3.841$ prefer L_1 to L_2

$2\,(\ln L_3 - \ln L_1) = 2\,(-319.12 + 321.32) =$

$4.4 < \chi^2_{2,.95} = 5.991$ prefer L_1 to L_3

$2\,(\ln L_4 - \ln L_1) = 2\,(-318.12 + 321.32) =$

$6.4 < \chi^2_{3,.95} = 7.815$ L_1 preferred over L_4

So the 1 parameter model is preferred.

b) If $\alpha = 0.1$, the thresholds are 2.706, 4.605 and 6.251. Dividing by 2, we get 1.351, 2.3025, 3.1255, which are the thresholds required to add additional parameters. (df 1, 2, 3) The 2 parameter model increases the loglikelihood by 1.39 (321.32 − 319.93), meeting the threshold: it is preferred over the 1 parameter model. The 3 and 4 parameter models have loglikelihoods 0.81 and 1.81 higher than (319.93 − 319.12, 319.93 − 318.12; df 1, 2) the 2 parameter model, and do not pass the thresholds (1.351 and 2.3025). 47.5

So the 2 parameter model is selected.
(L2 is preferred to L1 and L3 and L4.)

ex)

model	maximal log likelihood	# parameters
Burr	−75.8	3
Exponential	−79.5	1
Inverse Pareto	−77.7	2 ← worse than paralogistic
Paralogistic	−77.2	2

given

For which level of confidence is the Paralogistic model preferred by the LRT? (confidence $p = 1 - \underset{\text{level}}{\alpha}$).

soln) To prefer paralogistic to Exponential,

$$2\,(-77.2 + 79.5) = 4.6 > \chi^2_{1,p}$$

df \ p	.95	.975
1	3.841	5.024

So prefer Paralogistic (2 param) to Exponential at 95% confidence. To prefer Paralogistic to Burr (3 param), need

$$2\,(-75.8 + 77.2) = 2.8 < \chi^2_{1,p}$$

(Burr (3 param) is preferred if $2.8 > \chi^2_{1,p}$)

df \ p	.9	.95
1	2.706	3.841

So 95%