VATSAV SETHUPATHI

MATH-DATA 363 FINAL EXAM

⑧ (a) $H_0: p_{WB} = 0.55$, $p_O = 0.16$, $p_{CA} = 0.12$, $p_m = 0.06$, $p_L = 0.06$ & $p_a = 0.05$

H_1: Atleast one of the p_i values is not equal to the given values, where i is the species of trees

(b)

tree species	WB	O	CA	M	L	A
proport of forest	110	32	24	12	12	10

(c) $\chi^2 = \frac{(127-110)^2}{110} + \frac{(38-32)^2}{32} + \frac{(10-24)^2}{24} + \frac{(9-12)^2}{12} + \frac{(10-12)^2}{12}$

$+ \frac{(10-6)^2}{10}$

$= 14.6023$ $\quad$ [DF = 6 − 1 = 5]

(d) pvalue $= \chi^2 cdf(14.6023, \infty, 5)$ [Calculator]

$= 0.012204$

Since our p-value < 0.05, we can reject the null hypothesis at the 5% significance level.

This means that there were tree species ~~where~~ ~~whose~~ that march tits preferred at a different probability than what was predicted.

(7) (a) $H_0 : \mu_s \geq 78,000$

$H_1 : \mu_s < 78,000$

Where μ_s is the mean salary of the division.

(b) $\bar{x} = 71500$ $\quad s = 18000$

$t = \cancel{\frac{\bar{x} - \mu_s}{18000/\sqrt{21}}}$ $\quad \frac{\bar{x} - \mu_s}{s/\sqrt{n}}$

$= \frac{715000 - 78000}{18000/\sqrt{21}} = -1.6548$

(c) ~~p = tcdf(-∞, 0.0548~~

$p = \text{tcdf}(-\infty, -1.6548, 20)$ [Calculator]

$= 0.05678$

This means that there is a 5.678% probability That this salary is not significantly lower than the mean

(d) When we change the no. of observations to 31, we get a t score of -2.01 which gives us a p-value of 0.0267. This means we have more evidence to reject the null hypothesis.

This is because by ~~decreasing~~ increasing the no. of observations, we are reducing the spread of the data. Due to the ~~larger~~ spread, the critical region increases, meaning it is easier to reject H_0.

smaller

Here x = Rank & y - MCAT Score

① (a) The std of 4.21 is for the mean MCAT Score since its spread is quite small. The Scores remain roughly around 511, whereas rank ranges from 0 all the way to 67, both of which are far from the mean.

(b) $r = \frac{cov(x,y)}{S_x S_y} = -0.5135$

This just means that the variables are negatively associated somewhat strongly. ~~We get no~~ ~~This association is negative.~~ It tells us nothing about cause & effect.

(c) $\hat{\beta} = \frac{cov(x,y)}{S_x^2} = -0.1034$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \Rightarrow 511.33 + (0.1034)25.89$
$\Rightarrow 514.007$

Eq $\Rightarrow$ $\hat{y} = 514.007 + 0.1034 x_i$

(d) Washington $\Rightarrow$ $\hat{y} = 514.007 + 0.1034(12)$
$= 515.248$

Residual $= 507 - 515.248$
$= -8.248$

This means that the avg score at Washington is -8.248 points less than our prediction.

Date.......
Page.......

(5) (a) $L(\lambda|t) = (\lambda^2)^8 \left[t_1 e^{-\lambda t_1} \cdot t_2 e^{-\lambda t_2} \cdots t_8 e^{-\lambda t_8} \right]$

$= (\lambda^{16}) \left[t_1 t_2 t_3 t_4 t_5 t_6 t_7 t_8 \, e^{-\lambda(t_1 + t_2 + \cdots t_8)} \right]$

(b) $\ln L(\lambda|t) = 16 \ln(\lambda) + \ln(t_1) + \cdots \ln(t_8)$
$- \lambda(t_1 + t_2 + \cdots t_8)$

$$\frac{\partial \ln L}{\partial \lambda} = \frac{16}{\lambda} - (t_1 + t_2 + t_3 + \cdots + t_8)$$

But $\frac{\partial \ln L}{\partial \lambda} = 0$

$\Rightarrow \frac{16}{\lambda} - (t_1 + t_2 + \cdots + t_8) = 0$ $\Rightarrow \hat{\lambda} = \frac{16}{t_1 + t_2 + \cdots + t_8}$

$\frac{\partial^2}{\partial \lambda} (\ln L) = \frac{-16}{\lambda^2}$ which is -ve, so this is a maximum

(c) $\hat{\lambda} = \frac{16}{0.68 + 12.00 + \cdots + 1.72} = 0.3468$

(d) $P(\text{Wait} \leq 10) = F_T(10 | \hat{\lambda})$
$= 1 - (1 + (0.3468)10)e^{-0.3468 \cdot 10}$
$= \boxed{0.861}$

Date....../....../......

Page..................

(6) (a) $H_0: p_{\cancel{M}W} - p_M \leq 0$

$H_1: p_W - p_M > 0$

Where p_W is ~~popula~~ proportion of women & p_M is proportion of men agreeing with the statement.

(b) $\hat{p}_M = 91/914 = 0.09956$ $\quad \hat{p}_W = 135/921 = 0.14658$ $\quad \hat{p} = 221/1835 = 0.12316$

$$z = \frac{0.14658 - 0.09956}{\sqrt{0.12316(1-0.12316)\left(\frac{1}{914} + \frac{1}{921}\right)}}$$

$= 3.06459$

(c) pvalue = normcdf(3.06459, ∞, 0, 1) [Calc]

$= 0.00109$

This means that we can expect to see this diff in proportion 0.109% of the time ~~[illegible]~~ if the null hypothesis is true. Therefore, we can reject it at the 1% level.

(d) $C.I = 0.14658 \pm z^*_{0.975} \sqrt{\frac{0.14658(1-0.14658)}{921}}$

1.96 from calculator

$= (0.123737, 0.169423)$

Date....../....../......

Page..........

SAME

③ (a) $F_x^{-1}(x|\theta) = (x)^{1/\theta}$ | $F_x(x|\theta) = 1/2 = x^{\theta}$

$F_x^{-1}(1/2|\theta) = (1/2)^{1/\theta}$ | $x = (1/2)^{1/\theta}$

which is the median

(b) density $= \dfrac{\partial F_x(x|\theta)}{\partial \theta} \Rightarrow \dfrac{\partial x^{\theta}}{\partial \theta} = \theta x^{\theta-1}$

(c) $EX = \int_0^1 f_x(x|\theta)\cdot x \, dx = \theta \int_0^1 x^{\theta} dx$

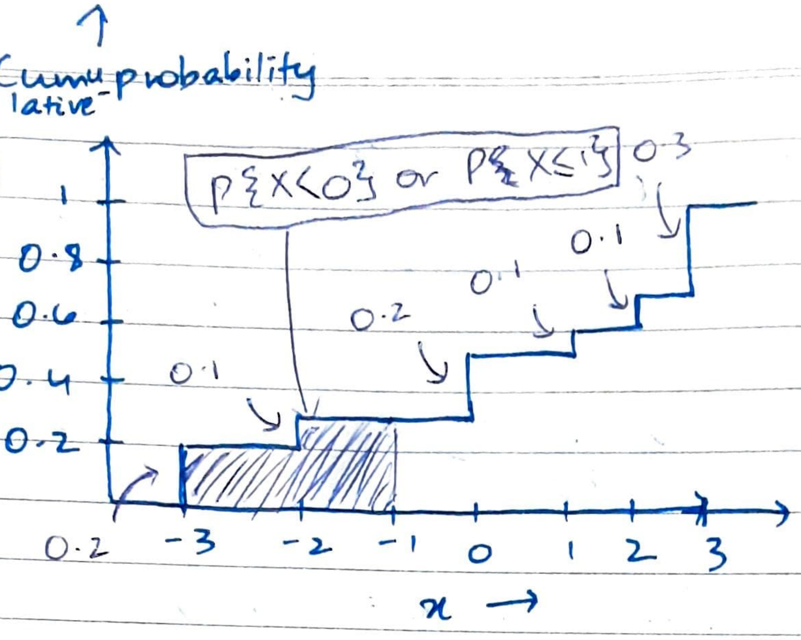
$= \theta \left[\dfrac{x^{\theta+1}}{\theta+1}\right]_0^1 \Rightarrow \dfrac{\theta\cdot(1)^{\theta+1}}{\theta+1} = \dfrac{\theta}{\theta+1} = \mu$

(d) $\bar{x} = \dfrac{\theta}{\theta+1}$ $\Rightarrow \bar{x}\theta + \bar{x} = \theta$

$\Rightarrow \bar{x} = \theta - \bar{x}\theta = \theta(1-\bar{x})$

$$\boxed{\hat{\theta} = \frac{\bar{x}}{1-\bar{x}}}$$

② (a)



(b) $EX = \sum f_X(x)\, x$

$\Rightarrow (-3)(0.2)+(-2)(0.1)+(-1)(0)+0(0.2)+(1)(0.1)+(2)(0.1)+(3)(0.3)$

$= 0.4$

(c) $Var(X) = EX^2 - (EX)^2$

$EX^2 = (-3)^2(0.2) + (-2)^2(0.1) + 1^2(0.1) + (2)^2(0.1) + (3)^2(0.3)$

$= 5.4$

$Var(X) = 5.4 - (0.4)^2 = 5.24$

(d) $E[5-3X] = E[5] - 3E[X] = 5 - 3(0.4)$

$= 3.8$

$Var[5-3X] = Var[5] - Var[3X]$

$= 5 - 3^2\, Var[X]$

$= 5 - (9)(5.24)$

$= -42.16$

$P(A|B) = \frac{P(A \cap B)}{P(B)}$

$P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C) + P(A|C^c)P(C^c)}$

Date....../....../......

Page..............

	Mean with σ known	Mean with σ unknown
1 sample ⇒	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ ‖ $\bar{x} + z^* \frac{\sigma}{\sqrt{n}}$	1 sam ⇒ $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ ‖ $\bar{x} + t^* \frac{s}{\sqrt{n}}$
2 sample	$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	2 sam ⇒ $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
	$\bar{x}_1 - \bar{x}_2 \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Regression

$\hat{\beta} = \frac{cov(x,y)}{Var(x)} = \frac{r s_x s_y}{s_x^2}$

$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$

$r = \frac{cov(x,y)}{s_x s_y}$

$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}^2}$ $df = (n-1) \times (m-1)$

Matched pairs:

$d = x_1 - x_2$ $\bar{d} = \bar{x}_1 - \bar{x}_2$ $sd = sd(x_1 - x_2)$

Proportions (Only z)

1 sample: $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ CI: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n}}$

2 sample: $z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} - \frac{1}{n_2}\right)}}$ [Here, $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$]

ANOVA Conf = $\bar{y} \pm t^* \frac{S_{res}}{\sqrt{1/n + 1/n}}$

Calc commands ⇒ normcdf, invcdf

CI: $\hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$\boxed{* = (1-\alpha)/2}$

Conf → $\bar{y}_j \pm t_{\frac{(1+\alpha)}{2}, n-q} \frac{S_{resid}}{\sqrt{n_j}}$

$F = \frac{SS_{between}/(q-1)}{SS_{resid}/(n-q)}$

↖ within

$SS_{resid} = \sum s_x^2 (n-1) \Rightarrow SSa$

$SS_{betw} = \sum n(\bar{y} - \bar{\bar{y}}) \Rightarrow SSE$