Introduction to Biostatistics PHP 2500
Midterm Exam
Thursday, October 25, 2007

(Closed Book)

There are 32 `questions', each worth 3.125 points.

Write out your solutions and circle your final answer. Show all your work on these pages. If you need more space write on the back of the page.

You have 1.5 hours to complete this exam.

Name: Solutions

1. Below are the infant mortality rates per 1,000 live births for 24 countries in 2006. Make the calculations needed for a box plot, and enter your results in the appropriate blanks.

	Angola	184.4	Iran	38.1	
(d) –	Afghanistan	157.4	India	34.6	
	Mali	105.7	Indonesia	32.1	
	Zambia	100.7	Egypt	30.1	
	Nigeria	95.5	Algeria	28.8	
	Congo	86.6	Syria	27.7	– (c)
(b) –	Rwanda	85.3	Brazil	27.6	
	Congo	83.3	China	22.1	
	Uganda	67.2	Mexico	19.6	
	Haiti	63.8	Qatar	17.5	
	Sudan	59.6	Argentina	14.3	
	Togo	59.1	France	4.2	– (e)

(a) median = 48.6

$$\frac{12^{th} + 13^{th}}{2} = \frac{59.1 + 38.1}{2} = 48.6$$

(b) upper hinge = 85.3

$$3\left(\frac{n+1}{4}\right)^{th} obs = 18.75\ obs \Rightarrow 18^{th}$$

(c) lower hinge = 27.7

$$\left(\frac{n+1}{4}\right)^{th} obs = 6.25 \Rightarrow 6^{th}$$

$1.5\,IQR = 1.5(85.3 - 27.7) = 86.4$

upper fence $85.3 + 86.4 = 171.7$

lower fence $27.6 - 86.4 = -58.8$

(d) upper whisker = 157.4

(closest to fence w/o going over)

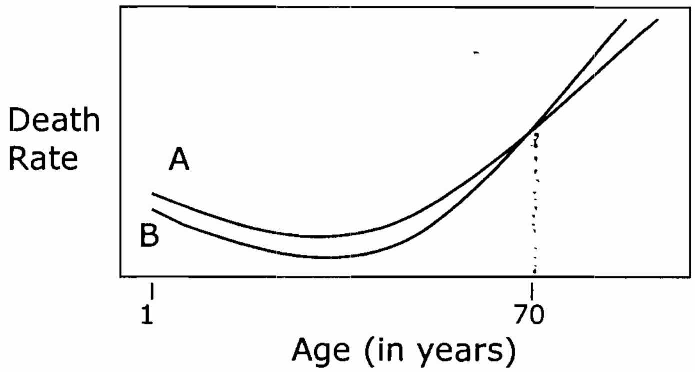
(e) lower whisker = 4.2

(closest to fence w/o going over)

(f) Are there any outliers in this data set? If so, identify those countries.

Yes, Angola.

2. The two lines in the graph below represent the age-specific death rates in two populations, A and B.



cross @ 70

(Check **one** box per question)

(a) If the age distributions in the two populations are the <u>same</u>,

☐ The crude rates for A and B will be equal
☐ A's crude rate will be greater than B's
☐ B's crude rate will be greater than A's
☒ Either one might be greater, depending on the (common) age distribution

(b) If the age distributions in the two populations are <u>different</u>,

☐ The crude rates for A and B will be equal
☐ A's crude rate will be greater or equal to B's
☐ B's crude rate will be greater or equal to A's
☒ Either one might be greater, depending on the two age distributions

(c) If the age-adjusted rates are calculated using a standard population where the age distribution consisted only of people less than 70 years old,

☐ The adjusted rates for A and B will be equal
☒ A's adjusted rate will be greater than B's
☐ B's adjusted rate will be greater than A's
☐ Either adjusted rate might be greater, depending on the age distribution of the standard population

3. The following table displays data on work related injuries by the level of highest education in ~~1749~~ 1747 people.

Education	Injured	Not Injured	
None	58	670	728
High School	17	456	473
Bachelors	5	324	329
Graduate	5	212	217
	85	1662	1747

(a) What is the probability of being injured and uneducated in this population?

$$P(I \text{ and None}) = \frac{58}{1747} = \boxed{0.0332}$$

(b) What is the probability of ~~being~~ not being injured?

$$P(\text{NOT } I) = \frac{1662}{1747} = \boxed{0.951}$$

(c) For individuals with graduate level education, what is the probability of being injured at work?

$$P(I \mid G) = \frac{5}{217} \quad \left[= \frac{5/1747}{217/1747} = \frac{P(I \cap G)}{P(G)} \right]$$

$$\boxed{= 0.023}$$

(d) For individual with at most a high school education, what is the probability of being injured?

	I	$\bar{I}$	
HS≤	75	1126	1201
HS>	10	536	
			1747

$$P(I \mid HS\le) = \frac{75}{1201} \quad \left[= \frac{58+17}{58+17+670+456} \right]$$

$$= \boxed{0.0624}$$

(e) Are the events {Injured} and {at most a high school education} Independent? (Justify your answer.)

IS $P(I \mid HS\le) = P(I)$

$$0.0624 \overset{?}{\neq} \frac{85}{1747} = 0.0489$$

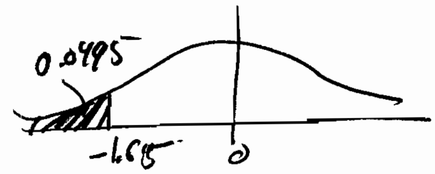
So, NO not indp!

4. After a natural disaster, the estimated distribution of clean-up costs for an American family is normally distributed with a mean of \$965 and a standard deviation of \$100.

(a) What is the probability that a randomly selected family will have to pay less than \$800 to clean-up after a natural disaster?

$X \sim N(965, 100^2)$

$$P(X < 800) = P\left(Z < \frac{800-965}{100}\right)$$
$$= P(Z < -1.65)$$
$$\boxed{= 0.0495}$$



(b) Find the top 14th percentile for a single family's clean-up costs.

$$P(X > ?) = 0.14$$
$$P\left(Z > \frac{?-\mu}{\sigma}\right) = 0.14$$
$$P\left(Z > \frac{?-965}{100}\right) = 0.14$$

But $P(Z > 1.08) = 0.14$

so $1.08 = \frac{?-965}{100}$

$$? = 965 + 1.08(100)$$
$$? = 1073$$
$$\boxed{\$1073}$$

(c) What is the distribution of the average clean-up cost for 49 families (give the family, mean and variance)?

$X_1, \dots, X_{49} \sim N(965, 100^2)$

so $\boxed{\bar{X}_{49} \sim N\left[965, \left(\frac{100}{7}\right)^2\right]}$

bc $E[\bar{X}_{49}] = \mu = 965$

$$VAR[\bar{X}_{49}] = \frac{VAR(X)}{49} = \frac{100^2}{49^2}$$
$$= 14.29^2$$
$$= 204.08$$

Problem 4 (continued)

(d) In a sample of 49 families, what is probability that their average clean-up cost will be between \$945 and \$985?

$$P(945 < \bar{x}_9 < 985)$$

$$P\left(\frac{945-965}{100/7} < Z < \frac{985-965}{100/7}\right)$$

$$P(-1.4 < Z < 1.4) = 1 - 2 \cdot P(Z > 1.4)$$

$$= 1 - 2 \cdot 0.0808$$

$$= \boxed{0.8384}$$

(e) Calculate the interval that captures the middle 80% of the clean-up cost for an American family.

$$P(?_L < X < ?_U) = .8$$

But $P(-1.28 < Z < 1.28) = .8$

$$P\left(\frac{?_L-\mu}{\sigma} < Z < \frac{?_U-\mu}{\sigma}\right) = .8$$

So $\frac{?_L-\mu}{\sigma} = -1.28$, $\frac{?_U-\mu}{\sigma} = 1.28$ $\Rightarrow$ $?_L = 965 - 1.28(100) = 837$, $?_U = 965 + 1.28(100) = 1093$

$$\boxed{[837, 1093]}$$

(f) If two independent families work together to clean-up, they get a \$200 'teamwork' rebate from the city (\$100 for each family). What is the distribution of the net clean-up cost when the two independent families work together (give the family, mean and variance)?

$X_1, X_2 \sim N(965, 100^2)$

$$\text{NetCost} = X_1 + X_2 - 200$$

$$E[NC] = \mu + \mu - 200 = 2 \cdot 965 - 200 = 1730$$

$$\text{Var}[NC] = \text{Var}(X_1) + \text{Var}(X_2) \quad \text{b/c indep}$$

$$= 100^2 + 100^2$$

$$= 2(100^2)$$

$$\boxed{NC \sim N[1730, 2(100^2)]}$$

20,000

5. Peripheral neuropathy (numbness in hands and feet) is a rare side effect of chemotherapy that occurs at a rate of 0.15 episodes per week. For very sick patients, multiple episodes per week are common. You are considering undergoing a cycle of chemotherapy that would last six weeks.

(a) How many episodes of peripheral neuropathy would you expect to have in a single cycle of chemotherapy?

$\lambda_{1week} = 0.15$ $X = \#$ PN Episodes in a 6 week cycle

$\lambda_{6weeks} = 0.15(6) = .9$ $X \sim Poiss(\lambda = .9)$ $E[X] = .9$

$\boxed{.9 \text{ Episodes per cycle}}$

(b) What is the standard deviation of the number of peripheral neuropathy episodes? (in a single cycle)

$SD(X) = \sqrt{var(X)} = \sqrt{.9}$

$\boxed{= 0.9487}$

(c) What is the probability that you will experience at least one episode of peripheral neuropathy in a single six week cycle?

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

$P(X \geq 1) = 1 - P(X=0)$

$= 1 - \frac{e^{-0.9}\,0.9^0}{0!}$

$= 1 - e^{-0.9}$

$= 1 - 0.4066 = \boxed{0.5934}$

Problem 5 (continued)

(d) Suppose on the very first day of chemotherapy you experience an episode of peripheral neuropathy. What is the chance that you will experience more episodes of peripheral neuropathy? (IN a single cycle)

$$P(X>1|X\geq 1) = P(X\geq 2|X\geq 1)$$

$$= \frac{P(X\geq 2)}{P(X\geq 1)}$$

$P(X\geq 2) = 1 - P(X=0) - P(X=1)$

$$= \frac{1-(e^{-0.9})-(e^{-0.9}0.9)}{1-e^{-0.9}} = \frac{0.2275}{0.5934} = \boxed{0.3833}$$

Later your doctor explains that you will have to undergo three cycles of Chemotherapy. However, after each cycle you will have a resting period so that the effects are not cumulative (i.e., cycles are independent).

(e) What is the probability that you will undergo three cycles of chemotherapy without having an episode of peripheral neuropathy?

$P(\text{No Episode of PN}) = P(X=0) = e^{-0.9} = 0.4066$

$Y = \#\text{cycles w/o PN} \Rightarrow Y \sim Bin(3, \theta)$ $\theta = e^{-0.9}$

$$P(Y=3) = \binom{3}{3}\theta^3(1-\theta)^0 = (.4066)^3$$

$$\boxed{= 0.0672}$$

≈ 0.0536

Problem 5 (continued)

(f) What is the probability that you have no more than one episode of peripheral neuropathy in the first and third cycles, but more than one episode during the second cycle?

$X \sim Poiss(0.9)$ $\quad P(X_1 \le 1 \text{ and } X_2 \ge 1 \text{ and } X_3 \le 1) = P(X \le 1)^2 P(X > 1)$

$$= [P(X=0) + P(X=1)]^2 [1 - P(X=0) - P(X=1)]$$

$$= \left[e^{-0.9} + \frac{e^{-0.9} 0.9^1}{1!}\right]^2 \left[1 - e^{-0.9} + e^{-0.9}(0.9)\right]$$

$$= [0.7725]^2 [0.2275]$$

$$= (0.5967)(0.2275)$$

$$= \boxed{0.1357}$$

(g) What is the probability that you will have a single episode of peripheral neuropathy in only one of the three cycles of chemotherapy?

Y = # NP Episodes in a cycle

$Y \sim Bin(3, \theta)$ $\quad \theta = P(Y=1) = \frac{e^{-0.9} 0.9^1}{1!} = 0.3659$

$$P(Y=1) = \binom{3}{1} \theta^1 (1-\theta)^2$$

$$= 3\theta(1-\theta)^2$$

$$= 3(0.3659)(1-0.3659)^2$$

$$= \boxed{0.4414}$$

6. Magnetic Resonance Spectroscopy of the prostate (MRS) is a new diagnostic test for men that is not always performed. At Hopkins hospital, radiologists perform MRS 92% of the time; at Harvard Hospital radiologists perform MRS 78% of the time; and at Dartmouth Hospital radiologists perform MRS 65% of the time. However, 70% of men go to Hopkins, while only 20% of men go to Harvard and 10% to Dartmouth.

Make Table Below

$P(Hop) = .7$; $P(Har) = .2$; $P(Dar) = .1$

$P(MRS \mid Hop) = .92$; $P(MRS \mid Har) = .78$; $P(MRS \mid Dar) = .65$

(a) How often is MRS preformed at these Hospitals?

$$P(MRS) = P(MRS \mid Hop)\,P(Hop) + P(MRS \mid Har)\,P(Har) + P(MRS \mid Dar)\,P(Dar)$$

$$= .92(.7) + .78(.2) + .65(.1)$$

$$= 0.865$$

or from table below $\frac{865}{1000}$

(b) What is the chance that the hospital was Harvard if an MRS was preformed?

$$P(Har \mid MRS) = \frac{P(MRS \mid Har)\,P(Har)}{P(MRS)}$$ ← Bayes

$$= \frac{.78(.2)}{.92(.7) + .78(.2) + .65(.1)}$$

$$= \frac{0.156}{0.865} = \boxed{0.1803}$$

From table or $\frac{156}{865} = 0.1803$

	Hop	Har	Dar	
MRS	.92(.7)·1000 644	.78(.2)·1000 156	.65(.1)·1000 65	865
$\overline{MRS}$	56	144	35	225
	700	200	100	1000

*Circled items are derived
not circled are given by problem

Problem 6 (continued)

(c) What is the probability that two independent radiologists from Hopkins each perform MRS on the same patient?

$$P(MRS_1 \text{ and } MRS_2 \mid \text{Hopkins}) = [P(MRS \mid Hop)]^2 \quad \text{b/c indep}$$

$$= [.92]^2$$

$$\boxed{= 0.8464}$$

(d) How often do two independent radiologists from the same hospital each perform MRS on the same patient?

$$P(MRS_1 \text{ and } MRS_2) = P(MRS_1 \text{ and } MRS_2 \mid Hop)\,P(Hop)$$

$$+ P(MRS_1 \text{ and } MRS_2 \mid HaR)\,P(HaR) + P(MRS_1 \text{ and } MRS_2 \mid DaR)\,P(DaR)$$

$$= [.92]^2(.7) + [.78]^2(.2) + [.65]^2(.1)$$

$$\boxed{= 0.7564} \quad \text{OR From new Table } \frac{756.4}{1000}$$

OR make new Table

	Hop	HaR	DaR	
	$.92^2(.7)\cdot 1000$	$.78^2(.2)\cdot 1000$	$.65^2(.1)\cdot 1000$	
MRS_1 MRS_2	592.4	121.7	42.3	756.4
NOT MRS_1 MRS_2	107.6	78.3	57.7	243.6
	700	200	100	1000

Problem 6 (continued)

(e) What is the probability that the hospital was Hopkins if you know that two independent radiologists each preformed MRS on the same patient?

$$P(Hop \mid MRS_1 \text{ and } MRS_2) = \frac{P(MRS_1 \text{ and } MRS_2 \mid Hop)\, P(Hop)}{P(MRS_1 \text{ and } MRS_2)}$$

Bayes

$$= \frac{[.92]^2(.7)}{[.92]^2(.7) + [.78]^2(.2) + [.65]^2(.1)}$$

$$= \frac{0.5924}{0.7564}$$

$$= \boxed{0.7832}$$

OR $\frac{592.4}{756.4}$ FROM New Table