Question 1: (15 points) The owner of an appliance store produced the following joint probability distribution for the number of refrigerators and stoves sold daily in her store.

	Refrigerators			
Stoves	**0**	**1**	**2**	
0	0.08	0.14	0.12	0.34
1	0.09	0.17	0.13	0.39
2	0.05	0.18	0.04	0.27
	0.22	0.49	0.29	

a) In a given day, what is the probability that at least one stove is sold in this store?

$p(1)+p(2) = 0.39+0.27=0.66$

b) What is the approximate probability that at least 100 stoves are sold in six months? Assume that each month has 20 working days. $20\times 6=120$ days

$P(X_1+\dots+X_{120}\geq 100)=?$ $\sim N(120\mu, \sqrt{120}\,\sigma)$

$\sim N(111.6, 8.52)$

$\mu=(0.39)1+(0.27)2=0.93$

$\sigma^2 = 1^2(0.39)+2^2(0.27)-(0.93)^2=0.6051$ $\Rightarrow \sigma \simeq 0.778$

$\Rightarrow P\left(Z\geq \frac{100-0.5-111.6}{8.52}\right)=P(Z\geq -1.42)$

$=0.5+0.4222=0.9222$

c) Find the correlation coefficient between the stoves and refrigerator sales. Interpret the result: weak, medium or strong correlation?

$E(X,Y)= 0.17\,(1)(1)+0.13(1)(2)+(0.18)(2)(1)+0.04(2)(2)$

$=0.17+0.26+0.36+0.16=0.95$

$E(X)= 0.49+2(0.29)=1.07$ $E(Y)=0.39+(2)(0.27)=0.91$

$Cov(X,Y)= 0.95-(1.07)(0.91)=-0.0237$

$Var(X)= E(X^2)-(E(X))^2 = 0.49+(4)(0.29)-\{1.07\}^2=1.65-1.14\simeq 0.50$

$Var(Y)= E(Y^2)-(E(Y))^2 = 0.39+(4)(0.27)-\{0.91\}^2=1.47-0.82\simeq 0.64$

$\frac{-0.0237}{\sqrt{(0.50)(0.64)}}=\frac{-0.0237}{0.56}$

$=-0.042$

weak correlation

d) Find the conditional distribution of the number of refrigerators sold given that one stove is sold in a given day.

$P(R|S=1)=\frac{P(R\cap S=1)}{P(S=1)}\Rightarrow$

$P(R=0|S=1)=\frac{0.09}{0.39}\simeq 0.23$

$P(R=1|S=1)=\frac{0.17}{0.39}\simeq 0.43$

$P(R=2|S=1)=\frac{0.13}{0.39}\simeq 0.33$

Question 2: (20 points) It is recently reported in a newspaper that 4% of all mining workers consists of child labor. An investigator wants to take a random sample and test this claim.

a) The investigator wants to estimate the percentage of child labor with an error margin of 0.02 and 95% level of confidence. How large should the sample be? Use the newspaper's estimate in your calculations.

$$1.96\sqrt{\frac{(0.04)(0.96)}{n}} \le 0.02$$

$$\Rightarrow \sqrt{n} \ge 1.96\,\frac{\sqrt{(0.04)(0.96)}}{0.02} \Rightarrow n \ge 368.\ldots$$

$$n = 369$$

b) Test the claim of the newspaper at $\alpha = 0.05$, if 19 are classified as child labor in a random sample of 370 mining workers.

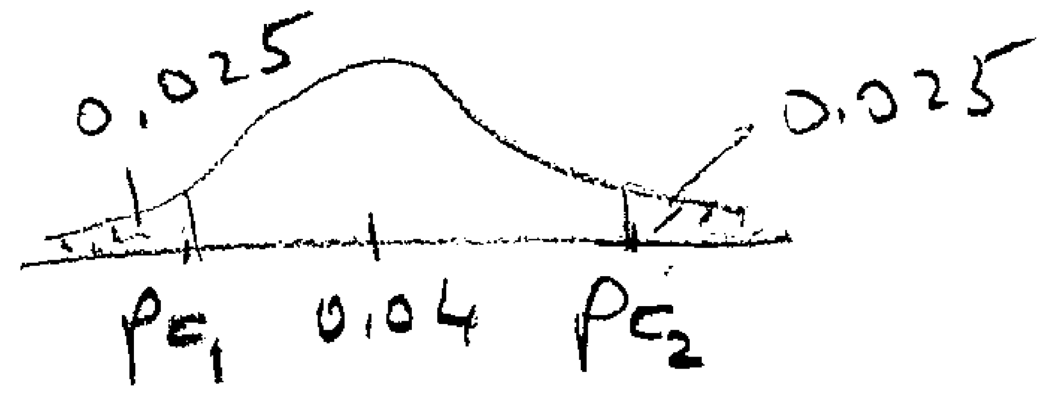
$H_0: p = 0.04$

$H_a: p \neq 0.04$

$$z^* = \frac{\frac{19}{370} - 0.04}{\sqrt{\frac{(0.04)(0.96)}{370}}} \cong 1.114$$

Since $1.114 < 1.96$, we fail to reject Ho.
So, the percentage of child labor can be considered as 4%.

c) Find the Type II error (β) if the true percentage of child labor is 4.8%.



$P_{critical}$

$$\frac{P_c - 0.04}{\sqrt{\frac{(0.04)(0.96)}{370}}} = \mp 1.96 \Rightarrow P_{c_1} = 0.02,\quad P_{c_2} = 0.06$$

$$\beta = P\left(\frac{0.02 - 0.048}{\sqrt{\frac{(0.048)(0.952)}{370}}} \le Z \le \frac{0.06 - 0.048}{\sqrt{\frac{(0.048)(0.952)}{370}}}\right) = P(-2.52 \le Z \le 1.08)$$

$$= 0.8599 - (1 - 0.9941) = 0.854$$

d) Assume the true percentage is 4%. What is the probability that there are at least two children among 50 randomly chosen mining workers?

$$P(X \ge 2) = 1 - \binom{50}{0}(0.04)^0(0.96)^{50} - \binom{50}{1}(0.04)(0.96)^{49} = 0.5996$$

or

$$P(X \ge 2) \cong P\left(Z > \frac{2 - 0.5 - (50)(0.4)}{\sqrt{50(0.4)(0.96)}}\right) = P(Z > -0.36) = 0.6406$$

or since $(50)(0.4) = 2$,

$$P(x \ge 2) \cong 1 - e^{-2}\frac{2^0}{0!} - e^{-2}\frac{2^1}{1!} = 0.594$$

Question 3: (15 points) Researchers from University of Michigan and Princeton University tested the placebo effect as follows: Each volunteer was put inside a magnetic resonance imaging (MRI) machine for two consecutive sessions. During the first session electric shocks were applied to their arms and the blood oxygen level-dependent (BOLD) signal (a measure related to neural activity in the brain) was recorded during the pain. The second session was identical to the first, except that, prior to applying the electric shocks, the researchers put a cream on the volunteer's arms. The volunteers were informed that the cream would block the pain experience, when in fact, it was just a regular skin lotion (that is, a placebo). If the placebo is effective in reducing the pain experience, the BOLD measurements should be higher, on average, in the first MRI session than the second.

The differences between the BOLD measurements in the first and second sessions had an average of 0.21 and a standard deviation of 0.47 for 24 volunteers.

a) Perform a test of hypothesis to test if the placebo is effective in reducing the pain experience. Report the P-value.

$H_0: \mu_D = 0$

$H_a: \mu_D \neq 0$

$$t^* = \frac{\bar{d}}{s_D/\sqrt{n}} = \frac{0.21-0}{0.47/\sqrt{24}} = 2.189$$

d.f. = 23

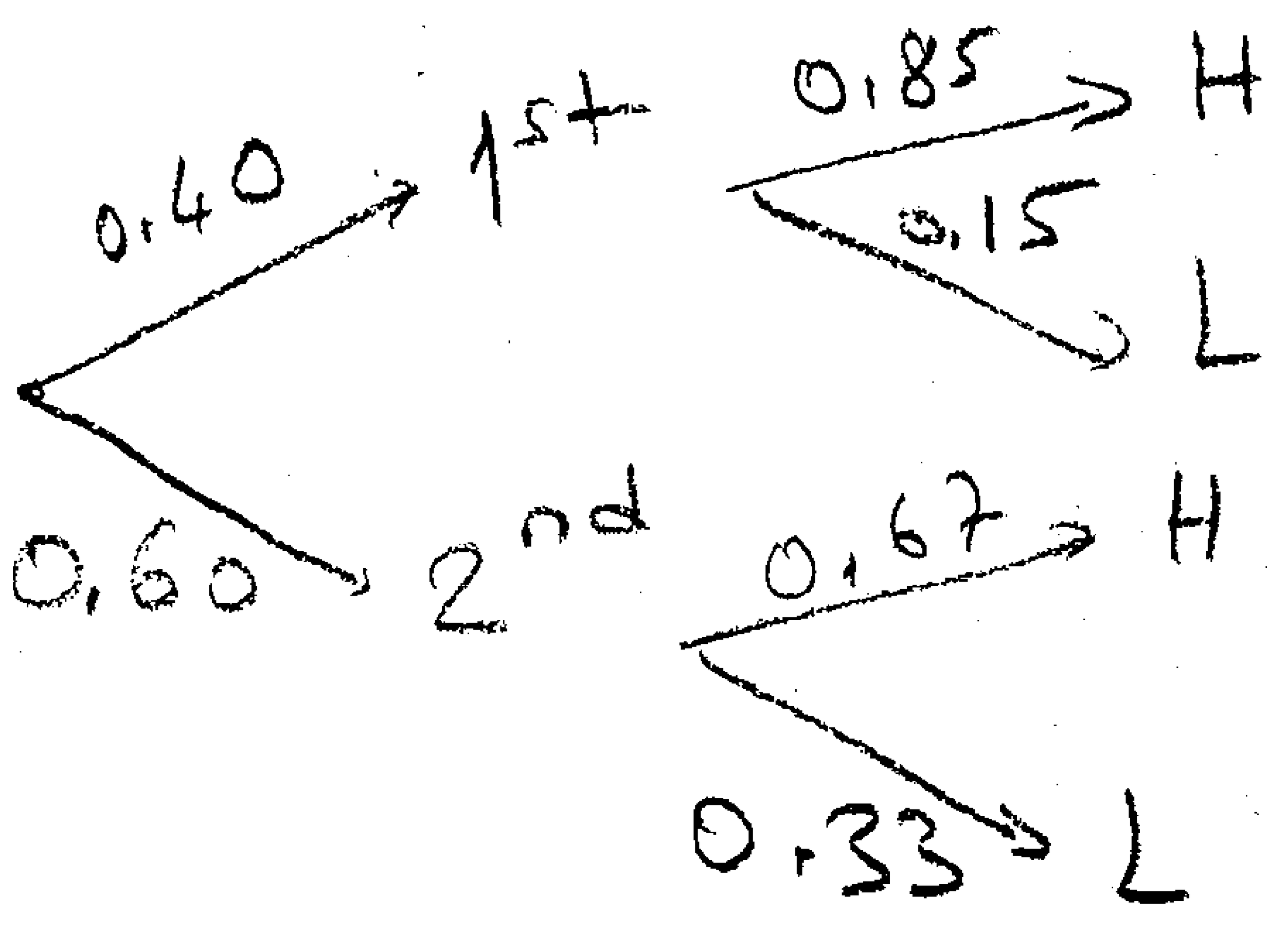
$t_{0.025,23} = 2.0687$ $\quad t_{0.01,23} = 2.5$

$\Rightarrow 0.01 < \frac{P\text{-value}}{2} < 0.025 \Rightarrow 0.02 < P\text{-value} < 0.05$

Small p-value, so reject H_0.

Yes, the placebo is significantly effective in reducing the pain.

b) In addition, the researchers have estimated the following probabilities in a different experiment. 40% of the volunteers are put in the first session, and 60% are put in the second session with placebo. The probability of high level of pain in the first session is 0.85, whereas in the second session this probability is 0.67. If a patient reports high level of pain in the end of the experiment, what is the probability that s/he was in the first session?



$$P(1^{st} \mid H) = \frac{(0.40)(0.85)}{(0.40)(0.85)+(0.6)(0.67)}$$

$$\cong 0.458$$

Question 4: (20 points) A researcher suggests that male nurses earn more than female nurses. A survey of independent samples of 15 male and 15 female nurses produced the following summary statistics for their yearly incomes (TL) given as MINITAB (statistics software) output. Here, SE Mean stands for standard error of the mean.

Variable	N	Mean	StDev	SE Mean
MALE	15	23800	300	77.46
FEMALE	15	23650	250	64.55

a) Is there enough evidence to support the claim of the researcher? Assume that the income distributions are normally distributed and the population standard deviations are equal for male and female nurses. Report the P-value.

$H_0: \mu_1 = \mu_2$ Assume $\sigma_1 = \sigma_2$ $d.f. = n_1 + n_2 - 2 = 28$

$H_a: \mu_1 > \mu_2$ $\Rightarrow S_p^2 = \frac{(14)(300)^2 + 14(250)^2}{28} = 76250$

$$t^* = \frac{23800 - 23650}{\sqrt{76250\left(\frac{1}{15} + \frac{1}{15}\right)}} \cong 1.49$$

Since $\underbrace{1.313}_{t_{0.1}} < 1.49 < \underbrace{1.701}_{t_{0.05}}$, $0.05 < \text{P-value} < 0.1$

Large P-value $\Rightarrow$ Do not reject H_0.
There is not sufficient evidence to support the claim that male nurses earn more.

b) Using the sample statistics given in MINITAB output to estimate the expectation μ and variance σ^2 of the income distributions for male and female nurses, respectively, estimate the expectation and variance of the family income of a married couple of nurses by assuming that their incomes are independent. $\hat{\mu}_1 = 23800$ $\hat{\mu}_2 = 23650$ $\hat{\sigma}_1 = 300$ $\hat{\sigma}_2 = 250$

$E(X_1 + X_2) = 23800 + 23650 = 47450$ TL
$V(X_1 + X_2) = 300^2 + 250^2 = 152500$ TL

c) Take the parameter estimates in part b) to be true values for a married couple of nurses. Find the probability that the average household income of a random sample of 15 ~~nurses~~ couples is at most 47200 TL? $W \sim N(47450, \sqrt{152500})$

$$P(\overline{W} \leq 47200) = P\left(Z \leq \frac{47200 - 47450}{\sqrt{\frac{152500}{15}}}\right)$$

$$\cong P(Z \leq -2.48)$$

$$= 1 - 0.9934 = 0.0066$$

Question 5: (20 points) The time it takes for a certain pain reliever to reduce symptoms is assumed to be normally distributed. The pharmaceutical company advertises that their medication reduces pain in less than half an hour.

a) In a random sample of 35 patients, the mean and the standard deviation of the time for reducing the pain are found to be 28.3 minutes and 5.1 minutes, respectively. Is there sufficient evidence that the mean time to reduce the pain for this pain reliever is less than half an hour at 5% level of significance?

$H_0: \mu = 30$ $\quad \alpha = 0.05 \quad \Rightarrow t_{0.05,34} = 1.6909$

$H_a: \mu < 30$ $\quad n = 35$

$$t^* = \frac{28.3 - 30}{5.1/\sqrt{35}} = -1.97$$

Since $-1.97 < -1.6909$, we reject H_0.

Yes, there is sufficient evidence that it reduces the pain in less than 30 min.

b) Suppose the true mean to reduce the pain for this pain reliever is 30 minutes and the true standard deviation is 4 minutes. What is the probability that it will take the medication between 35 and 40 minutes to begin to work?

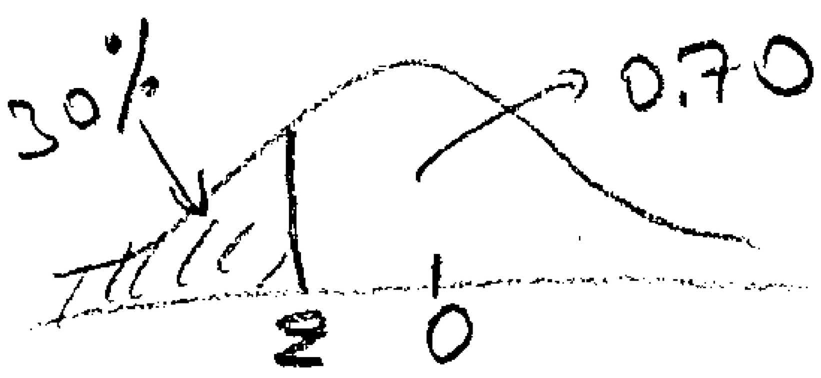
$$P(35 < X < 40) = P\left(\frac{35-30}{4} < Z < \frac{40-30}{4}\right)$$

$$= P(1.25 < Z < 2.5)$$

$$= 0.9938 - 0.8944$$

$$= 0.0994$$

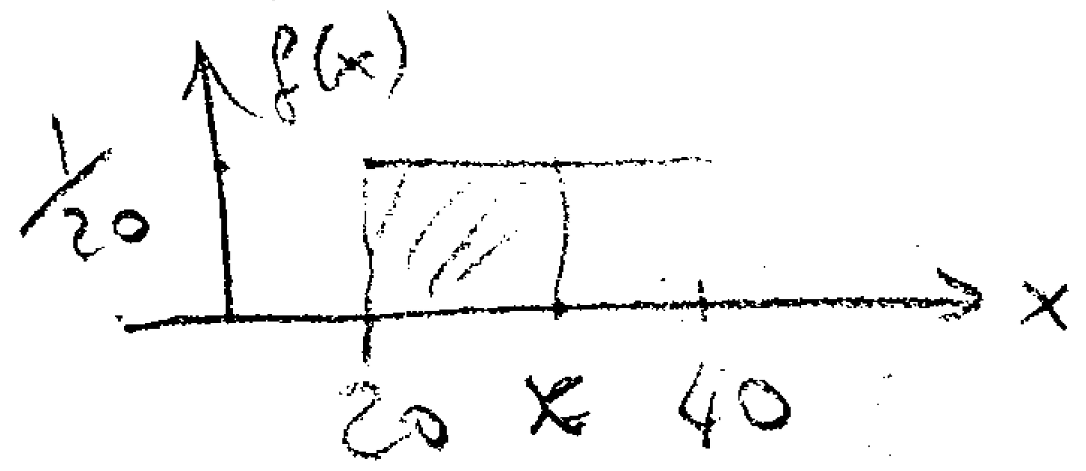
c) Find the 30^{th} percentile of the distribution of the time it takes for this pain reliever to reduce the symptoms, assuming the parameter values given in part b).



$$\Rightarrow z = -0.52 = \frac{x - 30}{4}$$

$$\Rightarrow x = 27.92 \text{ minutes}$$

d) Find the cumulative distribution function of the time it takes for this pain reliever to reduce the symptoms by assuming that the distribution is not normal and actually it is Unif(20,40).



$$\Rightarrow F(x) = \begin{cases} \frac{x-20}{20} & \text{if } 20 < x < 40 \\ 0 & \text{if } x \le 20 \\ 1 & \text{if } x \ge 40 \end{cases}$$

Question 6: (20 points) The relationship between the age (in years) and price (in thousand $) of a certain brand of automobiles is investigated. The following observations are obtained.

	x Age	y Price	x^2	y^2	xy
	5	16	25	256	80
	12	6	144	36	72
	9	8	81	64	72
	15	4	225	16	60
Sum	41	34	475	372	284

a) Find the regression equation for the given data set.

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{284 - \frac{(41)(34)}{4}}{475 - \frac{(41)^2}{4}} = \frac{-64.5}{54.75} = -1.18$$

$$\hat{\beta}_0 = \frac{34}{4} - (-1.18)\frac{41}{4} \cong 20.6$$

$$\hat{y} = -1.18x + 20.6$$

b) Estimate the price when the age is 7.

$$\hat{y} = -1.18(7) + 20.6 = 12.34 \Rightarrow 12340\$$$

c) Interpret the slope in the context of the problem.

The price of the automobile decreases by 1180 $ for every year that it depreciates.

d) Test if the true slope is 0 or not at the $\alpha = 0.01$ level of significance.

$H_0: \beta = 0$

$H_a: \beta \neq 0$

$t_{0.005,2} = 9.9248$ $\quad n=4 \quad n-2=2 \quad 0.01/2 = 0.005$

$$s^2 = \frac{1}{2}\left[372 - (20.6)(34) - (-1.18)(284)\right]$$

$$\Rightarrow s^2 = 3.36 \Rightarrow s = 1.833$$

$$t^* = \frac{-1.18 - 0}{1.833/\sqrt{54.75}} = -4.76$$

Since $-4.76 \nless -9.9248$ do not reject H_0.

So, the slope is not significantly different from 0.