Fall 2011, Final Exam

MATH 211

Question 1: (20 Points) Consider the experiment of tossing a fair coin three times. Let X denote the number of heads on the last flip (=0 if not heads, =1 if heads), and let Y denote the total number of heads on the three flips.

a) (5 points) Find $p_{X,Y}(x,y)$ for relevant values of x, y. Show the probabilities on the bivariate table:

x \ y	0	1	2	3	
0	1/8	2/8	1/8	0	1/2
1	0	1/8	2/8	1/8	1/2

b) (6 points) Find Cov(X, Y).

$$\text{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}X\,\mathbb{E}Y \quad , \quad \mathbb{E}X = 1/2 \quad , \mathbb{E}Y = 3/2$$

$$\mathbb{E}(XY) = 1\cdot 1\left(\tfrac{1}{8}\right) + 1\cdot 2\cdot\left(\tfrac{2}{8}\right) + 1\cdot 3\left(\tfrac{1}{8}\right) = 1.$$

$$\Rightarrow \text{Cov}(X,Y) = 1 - \tfrac{1}{2}\cdot\tfrac{3}{2} = \boxed{1/4}$$

c) (2 points) What is the probability that the total number of heads on the three flips is 2, given that the last flip is heads?

$$\mathbb{P}\{Y=2 \mid X=1\} = \mathbb{P}\{Y=2 \text{ and } X=1\} / \mathbb{P}\{X=1\} = \frac{2/8}{1/2}$$

$$= \boxed{1/2}$$

d) (7 points) If this experiment is repeated 50 times, what is the approximate probability that heads will appear on the third flip in more than 20 of these experiments?

$$\mathbb{P}\{\text{success}\} = \mathbb{P}\{X=1\} = 1/2$$

$$\mathbb{P}\{\underbrace{X_1 + \dots X_{50}}_{W} \geq 21\} = \mathbb{P}\left\{\frac{W-np}{\sqrt{np(1-p)}} \geq \frac{20.5-(50)(\frac{1}{2})}{\sqrt{50(\frac{1}{2})(\frac{1}{2})}}\right\}$$

$$= \mathbb{P}\left\{Z \geq \frac{-4.5}{\sqrt{12.5}}\right\} = \boxed{0.898}$$

Question 2: (20 Points) Let Y denote the percentage change in Dow Jones Industrial Average for a year (at the end of the year, compared to its beginning). Let Y_i denote the percentage change in year , $i = 1,2, \dots$, independently from each other.

a) (5 points) If Y is assumed to have a uniform distribution on [-5,15], find the cumulative distribution function for the maximum of Y_1 and Y_2 . Assume they are identically distributed.

$$f_Y(y) = \frac{1}{20}, \quad -5 \le y \le 15 \qquad Y_1 \perp\!\!\!\perp Y_2$$

$$\mathbb{P}\{\max(Y_1, Y_2) \le y\} = \mathbb{P}\{Y_1 \le y \text{ and } Y_2 \le y\} = \mathbb{P}\{Y_1 \le y\}\,\mathbb{P}\{Y_2 \le y\} = \left(\mathbb{P}\{Y \le y\}\right)^2$$

$$= \left(\int_{-5}^{y} \frac{1}{20}\, dy'\right)^2 = \left(\frac{1}{20}(y+5)\right)^2 = \boxed{\begin{array}{ll} \frac{(y+5)^2}{400} & \text{for } -5 \le y \le 15 \\ 0 \text{ for} & y \le -5 \\ 1 \text{ for} & y \ge 15 \end{array}}$$

b) (7 points) If Y is assumed to have a uniform distribution on [-5,15], find $P(Y_1 + \cdots + Y_{30} > 20)$ approximately. Assume all Y_i are identically distributed.

$$\mathbb{E}Y = 5, \quad \text{Var } Y = \frac{1}{12}(20)^2, \quad \sigma = \frac{20}{\sqrt{12}} \approx 5.77$$

$$= \mathbb{P}\left\{\frac{Y_1 + \cdots + Y_{30} - (30)(5)}{\sqrt{30} \cdot 5.77} > \frac{20 - (30)(5)}{\sqrt{30} \cdot 5.77}\right\} \cong \mathbb{P}\{Z > -4.11\}$$

$$\cong \boxed{1}$$

c) (8 points) Suppose i. i. d. observations $Y_1, \dots, Y_{30}$ are obtained and it is found that $\sum Y_i = 195.3$ Construct a 95% confidence interval for μ_Y assuming Var(Y)= 32.

$$\left(\bar{y} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \ \bar{y} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

$$\bar{y} = \frac{195.3}{30}, \quad \sigma = \sqrt{32}, \quad z_{0.025} = 1.96$$

$$6.51 \pm \frac{1.96\sqrt{32}}{\sqrt{30}} \rightarrow [4.486, 8.534]$$

Question 3: (20 Points) To predict the trends in stock market as indicated by Y, the percentage change in Dow Jones Industrial Average for a year, the following data were used:

Year	1991	1992	---	---	---	---	2007	2008
x_i	-2.6	-0.1					-0.5	-1.5
y_i	24.8	1.6					7.2	-31.4

where x denotes the percentage change for first week in Dow Jones Industrial Average for the first week in January. Some useful statistics are $\sum x_i = -0.9$, $\sum y_i = 160.2$, $\sum x_i^2 = 92.6$, $\sum y_i^2 = 6437.7$ $\sum x_i y_i = 221.4$

a) (8 points) Fit the regression line $y = \beta_0 + \beta_1 x$. $n = 18$

$$\beta_1 = \frac{18 \cdot 221.4 - (-0.9)(160.2)}{18(92.6) - (-0.9)^2} = 2.48$$

$$\beta_0 = \bar{y} - b\bar{x} = \frac{160.2 - (2.48)(-0.9)}{18} = 9.024$$

$$\boxed{y = 9.024 + 2.48x}$$

b) (1 point) What is your assumption for the distribution of Y while fitting the regression line?

normal

c) (11 points) Do the data indicate that change in January can be used to predict the change in a given year? Answer by conducting an appropriate hypothesis test for β_1 at $\alpha = 0.01$ and also by quantifying the strength of linear relationship between x and y.

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

$$t = \frac{\hat{\beta}_1 - 0}{s / \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} = \frac{2.48}{16.67 / 9.62} = 1.43$$

$$s = \left[\frac{1}{\underbrace{n-2}_{16}} \cdot \left(\sum_{i=1}^{18} y_i^2 - \beta_0 \sum_{i=1}^{18} y_i - \beta_1 \sum_{i=1}^{18} x_i y_i\right)\right]^{1/2} = 16.67$$

$$\sqrt{\sum (x_i - \bar{x})^2} = \sqrt{\sum x_i^2 - \left(\frac{1}{n}\right)\left(\sum x_i\right)^2} = 9.62$$

$t_{\alpha/2, n-2} = t_{0.005, 16} = 2.92 > 1.43 \Rightarrow$ do not reject H_0

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

$= 0.3368$

$\Rightarrow r^2 = 0.1135$

Question 4: (13 Points) For a sample of seven automobiles of the same brand, the mileage at which the original from brake pads (=fren balataları) were worn (aşınmış) to 10% of their original thickness was measured, as well as the same mileage for rear brake pads. The results are given in the following table where the numbers denote mileage in thousands of miles (e.g. 32.8 refers to 32800 miles).

Front	32.8	26.6	35.6	36.4	29.2	40.9	34.8
Rear	41.2	35.2	46.1	46.0	39.9	51.7	47.1
d	-8.4	-8.6	-10.5	-9.6	-10.7	-10.8	-12.3
$\lvert d-\bar{d}\rvert$	1.7	1.5	0.4	0.5	0.6	0.7	2.2

$\bar{d} = -10.1$

$S_D^2 = \boxed{11.24/6 = 1.87}$

$\sqrt{S_D^2} = \boxed{1.37}$

The sample means are 33.8 for front, and 43.9 for rear tires.

a) (8 points) Can you conclude that the mean lifetime of front brake pads is less than that of rear brake pads for this brand of automobiles?

$H_0: d = 0$

$H_1: d < 0$

$\bar{d} = 33.8 - 43.9 = -10.1$

$df = 6$

$\alpha = 0.05$

$$t^* = \frac{-10.1}{\boxed{1.37}/\sqrt{7}} = -19.5$$

$t_{0.05,6} = -1.94$

$t^* < t_{0.05,6} \Rightarrow$ reject Ho, yes you can conclude that mean lifetime of brake pads is less than that of rear brake pads.

b) (5 points) Can you conclude that the mean lifetime of rear brake pads exceeds that of front brake pads by more than 10000 miles for this brand of automobiles?

$H_0: d = -10$

$H_1: d < -10$

$$t^* = \frac{-10.1 + 10}{1.37/\sqrt{7}} = -0.19$$

$t^* > -1.94 \Rightarrow$ do not reject Ho, no you can't conclude that mean difference is greater than 10000 miles.

Question 5: (13 Points) Eight independent measurements were taken of the dissolution rate of a certain chemical at a temperature of c, and eight independent measurements were taken at 10°C. The results are as follows

0°C	2.28	1.66	2.56	2.64	1.92	3.09	3.09	2.48
10°C	4.63	4.56	4.42	4.79	4.26	4.37	4.44	4.21

The mean and standard deviation are 2.47 and 0.51, respectively for 0°C. They are 4.46and 0.19 for 10°C

a) (8 points) Find a 99% confidence interval for the difference of the dissolution rate of this chemical at 0°C and 10°C.

$$\left(\bar{x} - \bar{y} - t_{\alpha/2,\, n+m-2} \cdot S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \;,\; \bar{x} - \bar{y} + t_{\alpha/2,\, n+m-2} \cdot S_p \sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

$\alpha/2 = 0.005$

$n+m-2 = 14$

$\bar{x} - \bar{y} = 2.47 - 4.46 = -1.99$

$t_{0.005,14} = 2.9768$

$$S_p^2 = \frac{7(0.51)^2 + 7(0.19)^2}{14} = 0.1481$$

$S_p = 0.385$

CI

$\boxed{[-2.56, -1.42]}$

b) (2 points) What were your assumptions about
-the distribution
-and the variances
of the dissolution rates, in your solution approach to part a) ?

normal distr.

$\sigma_X = \sigma_Y$

c) (3 points) Can you conclude that there is a significant difference in the dissolution rates of the chemical at the two different temperatures?

Yes, since $0 \notin [-2.56, -1.42]$

Question 6: (14 Points) A manufacturing company purchases resistors labeled as 100ohm from two different vendors. The quality specification for such resistors is that their actual resistance must be close enough to their labeled value (here 100ohms). In a sample of 180 resistors from vendor A, 150 of them met the specification. In a sample of 270 resistors purchased from vendor B, 233 of them met the specification.

Currently, vendor A is the supplier of the company. If the data demonstrate convincingly that a greater proportion of the resistors from vendor B meet the specification, a change will be made.

Should a change be made? Answer by performing a hypotheses test and report the *P*-value.

$$A \rightarrow P_X = \frac{150}{180} = 0.83 \qquad H_0 : P_X = P_Y$$

$$B \rightarrow P_Y = \frac{233}{270} = 0.86 \qquad H_1 : P_X < P_Y$$

$$P_e = \frac{150 + 233}{450} = 0.85 \qquad z = \frac{P_X - P_Y}{\sqrt{P_e(1-P_e)\left(\frac{1}{\underset{180}{n}}+\frac{1}{\underset{270}{m}}\right)}} = -0.87$$

$$\text{p-value} = \mathbb{P}\{Z \le -0.87\} = 0.1922$$

p-value is high $\Rightarrow$ don't reject H_0

a change shouldn't be made, b/c data doesn't demonstrate that the proportion for B is significantly greater than for A.