Spring 2018 Math 211
Final Exam Solutions

Question 1: (20 Points) Consider two random variables X and Y. Note that we assume different distributions for them in the following parts a), b) and c).

a) (5 points) Find Cor(X,Y) (correlation) if their joint probability function $p(x,y)$ is given in the following table.

		y		
		0	1	2
x	1	0.1	0.2	0.1
	2	0.1	0.2	0.3

$Cov(X,Y) = E[XY] - E[X]E[Y]$

$E[X] = 1\cdot(0.4) + 2\cdot(0.6) = 1.6$

$E[Y] = 1\cdot(0.4) + 2\cdot(0.4) + 0\cdot(0.2) = 1.2$

$E[XY] = 1\cdot1\cdot(0.2) + 1\cdot2\cdot(0.1) + 2\cdot2\cdot(0.3) + 2\cdot1(0.2) = 2$

$Cov(X,Y) = 2 - (1.6)(1.2) = 0.08$

$E[X^2] = 1^2(0.4) + 2^2(0.6) = 2.8$

$E[Y^2] = 1^2(0.4) + 2^2(0.4) = 2.0$

$Var[X] = E[X^2] - E[X]^2 = 0.24$

$Var[Y] = E[Y^2] - E[Y]^2 = 0.56$

$Cor[X,Y] = \dfrac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}} = 0.22$

b) (7 points) Find c if their joint density $f(x,y)$ is given by $f(x,y) = c(x^2+y)$ for $0 \le y \le 1-x^2$, and $f(x,y) = 0$ otherwise.

Since f is pdf, we have

$$\int_{-1}^{1}\int_{0}^{1-x^2} c(x^2+y)\,dy\,dx = 1$$

$$\int_{-1}^{1}\int_{0}^{1-x^2} c(x^2+y)\,dy\,dx = 2\cdot\int_0^1 c\left(x^2y + \frac{y^2}{2}\right)\Bigg|_0^{y=1-x^2} dx = c\int_0^1 1 - x^4\,dx$$

$$= c\cdot\left(x - \frac{x^5}{5}\Bigg|_{x=0}^{x=1}\right) = \frac{4c}{5} = 1 \Rightarrow c = \frac{5}{4}$$

c) (8 points) If X has uniform distribution on the interval $[0,\vartheta]$, show that the estimator

$$\hat{\vartheta} = \frac{n+1}{n}\max(X_1, \ldots, X_n)$$

is unbiased.

$F_X(x) = \dfrac{x}{\vartheta}$ for $0 \le x \le \vartheta$

$$F_{max}(x) = P(\max(X_1 \ldots X_n) \le x) = P(X_1 \le x, X_2 \le x \ldots, X_n \le x)$$

$$= P(X_1 \le x) \cdots P(X_n \le x) = (F_X(x))^n = \frac{x^n}{\vartheta^n}$$

$$f_{max} = \frac{dF_X}{dx} = \frac{n\,x^{n-1}}{\vartheta^n}$$

$$E\left[\frac{n+1}{n}\cdot\max(X_1 \ldots X_n)\right] = \frac{n+1}{n}\int_0^{\vartheta} \frac{x\cdot n\cdot x^{n-1}}{\vartheta^n}\,dx = \frac{n+1}{n}\cdot\left(\frac{n}{(n+1)\vartheta^n}\,x^{n+1}\Bigg|_0^{\vartheta}\right)$$

$= \vartheta$. So $\hat{\vartheta}$ is unbiased

Question 2: (15 Points) Answer the following questions.

a) (7 points) A random sample of size 35 is obtained from the distribution of Y in Question 1 a). What is the probability that the mean of the sample is smaller than 1.3?

Is the probability you calculate exact or approximate? Why?

$$P(\bar{Y} < 1.3) = P\left(\frac{\bar{Y} - E[\bar{Y}]}{\sigma_{\bar{Y}}} \leq \frac{1.3 - E[\bar{Y}]}{\sigma_{\bar{Y}}}\right)$$

$$E[\bar{Y}] = E[Y] = 1.2$$

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{35}} = \frac{\sqrt{0.56}}{\sqrt{35}} = 0.126$$

$$\text{so } P(\bar{Y} < 1.3) = P\left(Z \leq \frac{0.1}{0.126}\right) = P(Z \leq 0.79)$$

$$= 0.7852$$

Since Y does not have normal distribution, we used central limit theorem. Hence it is approximate.

b) (8 points) The 75th percentile of a normally distributed random variable is 4. Find its standard deviation if its mean is 2.

What is the probability that the mean of a sample of 10 observations of this random variable is between 1.8 and 3.5?

Is the probability that you calculate exact or approximate? Why?

$$\mu = 2$$

$$P(X < 4) = 0.75 \Rightarrow$$

$$P\left(Z < \frac{4-\mu}{\sigma_x}\right) = 0.75 = P(Z < 0.67)$$

$$\Rightarrow \frac{4-2}{\sigma_x} = 0.67 \Rightarrow \sigma_x = 2.99$$

$$P(1.8 \leq \bar{X} \leq 3.5) = P\left(\frac{(1.8)-2}{2.99/\sqrt{10}} \leq Z \leq \frac{(3.5)-2}{2.99/\sqrt{10}}\right)$$

It is exact since X is normal distribution.

$$= P(-0.211 \leq Z \leq 1.58)$$

$$= P(Z \leq 1.58) - P(Z \leq -0.21)$$

$$= (0.9429) - (0.4168)$$

$$= 0.5263$$

Question 3: (20 Points) Suppose that on the average, 60% of the graduating seniors at a certain university have at least one parent attend the graduation ceremony.

a) (5 points) In a sample of 650 graduating seniors, what is the approximate probability that at most 400 of them will have at least one parent present in the ceremony?

$n = 650$
$p = 0.6$

$$P(X \leq 400) = P\left(Z \leq \frac{400 + 0.5 - (650)(0.6)}{\sqrt{650 \cdot (0.6)(0.4)}}\right)$$

$$= P(Z \leq 0.84) = 0.7995$$

b) (9 points) In a random sample of 650 graduating seniors, it is observed that exactly 397 students had at least one parent attend the ceremony. Conduct a test of hypothesis to test the claim that more than 60% of the graduating seniors have at least one parent attend the graduation ceremony in general, at $\alpha = 0.01$.

$H_0: p = 0.6$ vs $H_1: p > 0.6$, $k = 397, n = 650, \alpha = 0.01$

$$Z = \frac{k - n \cdot P_0}{\sqrt{n \cdot P_0 \cdot (1 - P_0)}} = 0.56$$

$$Z_{0.01} = 2.32$$

Since $0.56 < 2.32$

Fail to reject H_0.

c) (6 points) Find the Type II error of the test in part b), if the true percentage of graduating seniors who have at least one parent attend the graduation ceremony in general is 64%.

$$P(\text{fail to reject } H_0 \mid p = 0.64)$$

$$= P\left(\frac{k - n \cdot (P_0)}{\sqrt{n \cdot P_0 (1 - P_0)}} < 2.32 \,\middle|\, p = 0.64\right)$$

$$= P(k < 419.1 \mid p = 0.64)$$

$$= P\left(Z < \frac{419.1 - (650)(0.64)}{\sqrt{650 \cdot (0.64)(0.36)}}\right)$$

$$= P(Z < 0.25) = 0.5987$$

Question 4: (15 Points) Two different types of chemical solution are evaluated for polishing the contact lenses to be used in the human eye following cataract surgery. 300 lenses were polished using the first solution, and of those, 253 had no polishing-induced defects. Another 300 lenses were polished using the second solution, and 196 lenses were satisfactory in the end.

Is there any reason to believe that the two polishing solutions differ in performance?

Report the P-value.

$n = 300$ $\quad x = 253$ $\quad \alpha = 0.05$

$m = 300$ $\quad y = 196$ $\quad z_{\alpha/2} = 1.96$

$$Pe = \frac{x+y}{n+m} = \frac{253+196}{600} = 0.748$$

$H_0: P_x = P_y$ vs $H_1 = P_x \neq P_y$

$$z = \frac{\frac{x}{n} - \frac{y}{m}}{\sqrt{\frac{Pe(1-Pe)}{n} + \frac{Pe(1-Pe)}{m}}} = \frac{\frac{253}{300} - \frac{196}{300}}{\sqrt{\frac{(0.75)(0.25)}{150}}} = 5.37$$

$z > z_{\alpha/2}$ so reject H_0. Solutions differ in performance

$$P\text{-value} = 2 \cdot P(z < -5.37) \cong 0$$

Question 5: (15 Points) Samples of pottery from the Roman era as found in two different regions in Turkey are studied for their aluminum oxide content. The first sample $x_1, \ldots, x_5$ had an average aluminum oxide percentage of $\bar{x} = 18.2$ with $\sum(x_i - \bar{x})^2 = 12.6$. The second sample contained 6 observations, and had a mean of 17.3 and variance 6.2 .

Suppose that the variances of the aluminum content of the pottery in the two regions are different.

a) (10 points) Are the means of the aluminum content of the pottery in the two regions significantly different?

$n = 5$, $\bar{x} = 18.2$, $s_x^2 = 12.6$ $\qquad \alpha = 0.05$

$m = 6$, $\bar{y} = 17.3$, $s_y^2 = 6.2$ $\qquad H_0: \mu_x = \mu_y$ vs $H_1: \mu_x \neq \mu_y$

$$w = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}} = \frac{18.2 - 17.3}{\sqrt{\frac{12.6}{5} + \frac{6.2}{6}}} = 0.477$$

$$\hat{\theta} = \frac{s_x^2}{s_y^2} = \frac{12.6}{6.2} = 2.03$$

$$\nu = \frac{\left(\hat{\theta} + \frac{n}{m}\right)^2}{\frac{\hat{\theta}^2}{n-1} + \frac{1}{m-1} \cdot \left(\frac{n}{m}\right)^2} = 7.01 \simeq 7$$

$t_{0.025, 7} = 2.365$ Since $-2.365 < 0.47 < 2.365$

Fail to reject Ho. Means are not significantly dif.

b) (5 points) Construct a 99% confidence interval for the mean difference of the aluminum content of the pottery in the two regions.

$\alpha = 0.01$

Interval is

$$\left((\bar{x} - \bar{y}) \mp t_{\alpha/2, \nu} \cdot \sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}} \right)$$

$$= \left((18.2 - 17.3) \mp (3.50) \left(\sqrt{\frac{12.6}{5} + \frac{6.2}{6}} \right) \right)$$

$$= (-5.70, 7.50)$$

Question 6: (25 Points) When a small amount x of an insulin preparation is injected into a rabbit, the percentage decrease Y in blood sugar is observed. The following data are obtained.

x	0.6	1.0	1.7	2.2	2.8
Y	8	3	5	10	9

a) (8 points) Find the regression equation for this data set.

$$\beta_1 = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$\left.\begin{aligned} \sum x_i y_i &= 63.5 \\ \sum x_i &= 8.3 \\ \sum y_i &= 35 \\ \sum x_i^2 &= 16.93 \end{aligned}\right\} \quad \beta_1 = \frac{5\cdot(63.5) - (8.3)\cdot 35}{5\cdot(16.93) - (8.3)^2} = 1.71$$

$$\beta_0 = \bar{y} - \beta_1 \cdot \bar{x} = 4.16$$

b) (3 points) Interpret the slope by filling in the blanks:

If the amount of insulin injection is increased by 1 unit, then it is expected that ________

the blood sugar will decrease by 1.71 %

c) (8 points) Perform a test of hypotheses to decide if the slope is equal to 10 or not.

$H_0: \hat{\beta}_1 = 2$ vs $H_1: \hat{\beta}_1 \neq 2$, $n = 5$, $\alpha = 0.05$

$t_{\alpha/2, n-2} = t_{0.025, 3} = 3.1825$

$$s^2 = \frac{1}{n-2}\left(\sum y_i^2 - \beta_0 \sum y_i - \beta_1 \sum x_i y_i\right) = 8.27 \Rightarrow s = 2.86$$

$$t = \frac{\beta_1 - \beta_0}{s/\sqrt{\sum(x_i - \bar{x})^2}} = \frac{1.71 - 2}{2.86/\sqrt{3.152}} = -0.32.$$

Since $-0.32 > -3.18$
Fail to reject H_0.
Spole is not significantly diffent from 2.

d) (6 points) Compute R^2 and comment on the linear fit between x and Y.

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2}\ \sqrt{n \sum y_i^2 - (\sum y_i)^2}} = 0.52$$

$r^2 = 0.27$

27% of the variation Y is explained by the linear regration X.