Math 211: Statistics for Sciences

January 5, 2013

Fall 2012 Final Exam

Closed book & notes; only a two-sided and handwritten A4 formula sheet and a calculator allowed; 2 hours 15 minutes. No questions accepted!

Instructions: There are six pages (one cover and five pages with questions) in this exam. Please inspect the exam and make sure you have all 6 pages. You may only use your calculator and your formula sheet. Do all your work on these pages. If you use the back of a page, make sure to indicate that. **You may not exchange any kind of material with another student.**

Remember: ***You must show your work to get proper credit.***

Academic Honesty Code: Koç University Academic Honesty Code stipulates that "copying from others or providing answers or information, written or oral, to others is cheating." By taking this exam, you are assuming full responsibility for observing the Academic Honesty Code.

NAME and SURNAME: KEY

SIGNATURE: ____________________

1	/15
2	/20
3	/15
4	/15
5	/15
6	/20
Total:	/100

Question 1 (15 Points): USA Today/Gallup polling company (anket şirketi) in October 2008 located a random sample of adults in the US who said they were cutting household spending (eve ait harcama) as a result of the economic crisis.

(a) What is the required sample size to estimate at 90% confidence the actual proportion of people who were cutting household spending in the population within 1%? Show your work.

$d = 0.01$, 90% confidence $\Rightarrow \alpha = 0.10 \Rightarrow \frac{\alpha}{2} = 0.05 \Rightarrow z_{0.05} = 1.645$

$$n \geq \left(\frac{z_{\alpha/2}}{2d}\right)^2 = \left(\frac{1.645}{2(0.01)}\right)^2 = (82.25)^2$$

$$= 6765.46 \nearrow \underline{\underline{6766}}$$

(b) The polling company located a random sample of 524 adults in the US who said they were cutting household spending as a result of the economic crisis. Of those, 29% said they were cutting back on cable TV and cell phone service. Find a 95% confidence interval for the proportion of this population who were cutting back on cable TV and cell phone service. Show your work.

$n = 524 \qquad \hat{p} = 0.29$

n is large so 95% CI on p is $\hat{p} \mp (z_{\alpha/2}) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

$95\% \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025 \Rightarrow z_{0.025} = 1.96$

$$\Rightarrow 0.29 \mp (1.96) \sqrt{\frac{0.29(1-0.29)}{524}}$$

$$= 0.29 \mp (1.96)(0.0198)$$

$$= 0.29 \mp 0.0389$$

$$= \underline{\underline{(0.2511, 0.3289)}}$$

Generated by CamScanner from intsig.com

Question 2 (20 points):

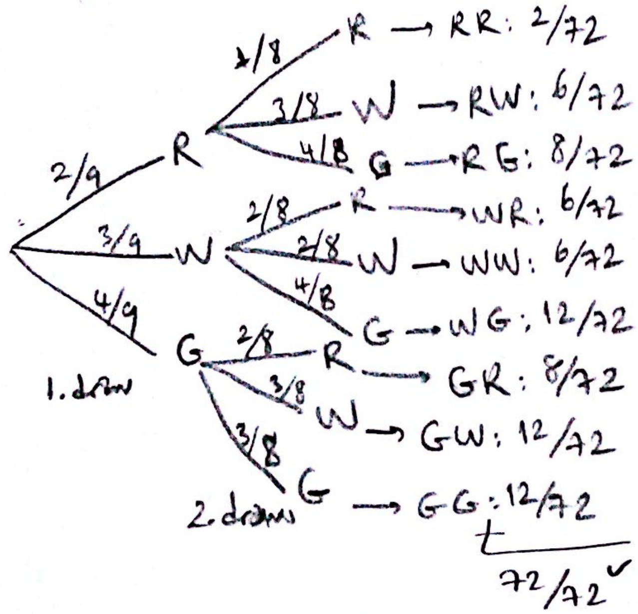
(a) A data set contains 78 numbers, and its stem-and-leaf plot is given in the right. Find the mode in this data set, and also find the relative frequency of the data value, 62.

1	5 9
2	4 5 7 9
3	5 6 6 8 8 8 9
4	2 4 4 5 5 5 7 7 9 9 9
5	1 3 3 3 3 3 5 5 5 6 6 7 7 7 8 8 8 9
6	2 2 2 5 5 5 5 5 5 6 6 6 8 8 8
7	1 2 2 3 3 4 5 5 7 8 9 9
8	8 8 8 9
9	5 6 7 7 7

The mode is 65 (which is the most frequent value, occurring 6 times).

Relative frequency of 62 is $\frac{3}{78} = \frac{1}{26}$.

(b) A box contains 2 red, 3 white, and 4 green balls. Two balls are drawn without replacement. Let events E and F be defined as E= "balls are of the same color" and F = "second ball is red". Find $P(E)$ and $P(F)$, the probabilities of the events. Are events E and F independent?



E = "RR or WW or GG"

$\Rightarrow P(E) = P(RR) + P(WW) + P(GG)$

$= \frac{2}{72} + \frac{6}{72} + \frac{12}{72} = \frac{20}{72} = \frac{5}{18}$

F = "RR or WR or GR"

$\Rightarrow P(F) = P(RR) + P(WR) + P(GR)$

$= \frac{2}{72} + \frac{6}{72} + \frac{8}{72} = \frac{16}{72} = \frac{2}{9}$

$E \cap F$ = "same color AND 2nd is R"

= "RR" $\Rightarrow P(E \cap F) = 2/72 = 1/36$

But $P(E)P(F) = \frac{5}{18} \times \frac{2}{9} = \frac{5}{81}$ so $P(E \cap F) \neq P(E)P(F)$

$\Rightarrow$ E & F are not independent.

(c) Two people are selected in order from ten married couples. How many choices are possible?

$_{20}P_2 = 20 \times 19 = 380$

(d) In the selection problem in part (c), what is the probability that two people selected are a married couple? There are 10 ways to select 1 couple from 10 couples and for each couple, two outcomes, (H, W), (W, H), result in a married couple. Hence $10 \times 2 = 20$ different outcomes that result in a married couple's selection.

husband wife

So prob is $20/380 = 1/19$

Question 3 (15 points): Although bratwurst (mangalda kızarmış sosis) is a very popular fast-food item in Germany, it is not so popular among Turks living in Berlin. But Salim Dönerci thinks that more than 10% of the Turks eat bratwurst regularly. Mr. Dönerci asked 40 Turks picked at random, and 9 confirmed that they eat bratwurst regularly.

(a) What is the p-value of the test, and is Mr. Dönerci's claim that "more than 10% of the Turks eat bratwurst regularly" correct according to the statistical evidence at 0.05 level of significance?

$n=40, \ X=9$ $\quad H_0: p=0.10$ vs $H_a: p>0.10$

$\hat{p}=\frac{9}{40}=0.225$, since n is large, will use a z-test.

$$Z_{calc}=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}=\frac{0.225-0.10}{\sqrt{\frac{0.10(0.90)}{40}}}=\frac{0.125}{0.0477}=2.64$$

$$\text{p-value}=P(Z>2.64)=1-0.9959=\underline{0.0041}$$

so, since p-value $<\alpha=0.05$, we reject Ho and conclude that there is sufficient evidence to support his claim.

(b) If actually 10% of the Turks in Berlin eat bratwurst regularly, what is the probability that between 3 and 6 (both inclusive) out of 40 randomly selected Turks in Berlin eat bratwurst regularly?

Let $X=$ # of Turks out of 40 who eat bratwurst regularly.

Then $X \sim BIN(n=40, p=0.10)$, $np=40(0.10)=4$ $\quad np(1-p)=40(0.10)(0.9)=3.6$

Since n is large $X \xrightarrow{approx} N(np, np(1-p))$
$=N(4, 3.6)$

$$\text{so } P(3\le X\le 6)=P(2.5\le X\le 6.5)$$
$$=P\left(\frac{2.5-4}{\sqrt{3.6}}\le Z\le\frac{6.5-4}{\sqrt{3.6}}\right)$$
$$=P(-0.79\le Z\le 1.32)$$
$$=0.9066-0.2148$$
$$=\underline{0.6918}$$

Question 4 (15 points): On Friday nights, long lines of hungry people wait in front of the fast-food stands (tezgah) in Berlin. Hence Salim Dönerci opened a döner stand in Berlin, and he thinks that his customers wait less than those of Siegfried's, his main competitor (rakip). Mr. Dönerci observes that the average waiting time of 14 people who bought a bratwurst at Siegfried's is 8.4 minutes with a standard deviation of 2.1 minutes. For 12 people who bought döner from Mr. Dönerci, the average waiting time was 6.8 minutes with a standard deviation of 1.8 minutes. Do people wait less at Salim's stand? Test at $\alpha = 0.05$. (You may assume that population variances are equal).

X: Siegfried's: $\bar{X} = 8.4$, $s_x = 2.1$, $n = 14$

Y: Salim's: $\bar{Y} = 6.8$, $s_y = 1.8$, $m = 12$

$H_0: \mu_x = \mu_y$ vs. $H_a: \mu_x > \mu_y$

$\Rightarrow \mu_x - \mu_y = 0$ $\Rightarrow \mu_x - \mu_y > 0$

$n=14$ & $m=12$ are both small, so use a t-test.

Also, $\sigma_x^2 = \sigma_y^2 = \sigma^2$, so we estimate the pooled variance

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} = \frac{(14-1)(2.1)^2 + (12-1)(1.8)^2}{14+12-2} =$$

$$= \frac{92.97}{24} = 3.874 \Rightarrow s_p = 1.968 \text{ and } df = 24$$

So $$T = \frac{(\bar{X}-\bar{Y}) - (\mu_x - \mu_y)}{s_p\sqrt{\frac{1}{n}+\frac{1}{m}}} = \frac{(8.4-6.8)-0}{1.968\sqrt{\frac{1}{14}+\frac{1}{12}}}$$

$$= \frac{1.6}{0.774} = 2.07$$

D.R: Reject Ho, if $T > t_{\alpha, df} = t_{0.05, 24} = 1.7109$

So, since $2.07 > 1.7109$, we reject Ho and conclude that people wait less at Salim's.

Question 5 (15 points):

(a) Bratwurst lengths in Bayern have a mean of 22.4 cm with a standard deviation of 2.4 cm. What is the probability that 52 of them have a combined length of more than 12 meters? (Assume that bratwurst lengths are independent).

Let X_i be the length of bratwurst i, so $\mu_x = 22.4$ & $\sigma_x = 2.4$

and let $Y = \sum_{i=1}^{52} X_i$, since $n = 52 > 30$, by CLT, we have

$$Y \xrightarrow{\text{approx}} N(\mu_y, \sigma_y^2) \quad \text{where } \mu_y = 52\mu_x = 52(22.4) = 1164.8\text{ cm}$$

and by independence, $\sigma_y^2 = 52(\sigma_x^2) = 52(2.4)^2 = 299.52$

So $P(Y > 12\text{ m}) = P(Y > 1200\text{ cm})$

$$= P\left(\frac{Y - 1164.8}{\sqrt{299.52}} > \frac{1200 - 1164.8}{\sqrt{299.52}}\right) = P\left(Z > \frac{35.2}{17.3066}\right)$$

$$= P(Z > 2.03) = 1 - 0.9788 = \underline{0.0212}$$

(b) If the bratwurst lengths in Bayern are distributed normally, what is the approximate probability that more than 5 out of 250 have a length of more than 28 cm? (The mean and standard deviation are given in part (a)).

$$X \sim N(22.4, (2.4)^2), \text{ so } P(X > 28) = P\left(Z > \frac{28 - 22.4}{2.4}\right)$$

$$= P\left(Z > \frac{5.6}{2.4}\right) = P(Z > 2.33)$$

$$= 1 - 0.9901 \cong 0.01$$

let Y = # of bratwursts longer than 28 cm among 250

So $Y \sim BIN(n = 250, p = 0.01)$

Since n is large, $Y \xrightarrow{\text{approx}} N(np, np(1-p))$

$$= N(2.5, 2.475)$$

$np = 250(0.01) = 2.5$

$np(1-p) = 250(0.01)(0.99) = 2.475$

so $P(Y > 5) = P(Y \geq 5.5) = P\left(\frac{Y - 2.5}{\sqrt{2.475}} \geq \frac{5.5 - 2.5}{\sqrt{2.475}}\right)$

$$= P\left(Z \geq \frac{3}{1.5732}\right) = P(Z \geq 1.91) = 1 - 0.9719$$

$$= \underline{0.0281}$$

Question 6 (20 points):

(a) According to a recent test, there was sufficient statistical evidence at $\alpha = 0.01$ to conclude that the weights of the bratwursts in Bayern were heavier than those in Berlin. The sample sizes in both states (eyalet) were 50, and the sample average for Berlin was 193 gr. with a sample standard deviation of 10.4 gr. The sample standard deviation for Bayern was 12.1 gr. The sample average for Bayern was erased accidentally. At least what weight is the erased sample average?

X: Bayern: $\bar{X} = ?$, $s_x = 12.1$, $n = 50$ — n, m are both large, so we use a

Y: Berlin: $\bar{Y} = 193$, $s_y = 10.4$, $m = 50$ — z-test.

$H_0: \mu_x = \mu_y \Rightarrow \mu_x - \mu_y = 0$ vs $H_a: \mu_x > \mu_y \Rightarrow \mu_x - \mu_y > 0$

D.R: Reject H_0, if $Z_{calc} > Z_{0.01} = 2.33$

$$Z_{calc} = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}} = \frac{(\bar{x}-193)-0}{\sqrt{\frac{12.3^2}{50}+\frac{10.4^2}{50}}} = \frac{\bar{x}-193}{2.2564}$$

$$\Rightarrow \frac{\bar{x}-193}{2.2564} > 2.33 \Rightarrow \bar{x} > 193 + (2.33)(2.2564)$$

$$= 193 + 5.2574$$

$$= \underline{198.26 \text{ gr}}$$

(b) Construct a 95% confidence interval for the mean (ortalama) weight of bratwursts in Berlin.

$$\bar{y} \mp (z_{0.025}) \frac{s_y}{\sqrt{m}} = 193 \mp (1.96)\left(\frac{10.4}{\sqrt{50}}\right)$$

$$= 193 \mp 2.883$$

$$= (190.12, 195.88)$$

(c) Test the claim that mean weight of bratwursts in Berlin are less than 195 gr. Use $\alpha = 0.05$.

$H_0: \mu = 195$ vs $H_a: \mu < 195$ — $n = 50$ is large

$$Z_{calc} = \frac{\bar{y}-\mu_0}{s_y/\sqrt{m}} = \frac{193-195}{10.4/\sqrt{50}} = \frac{-2}{1.471} = -1.36$$

$$\text{p-value} = P(Z < -1.36) = 0.0869.$$

Since p-value $\not< 0.05$, we fail to reject H_0 and conclude that there is not sufficient evidence to indicate that mean bratwurst weight in Berlin is less than 195 gr.