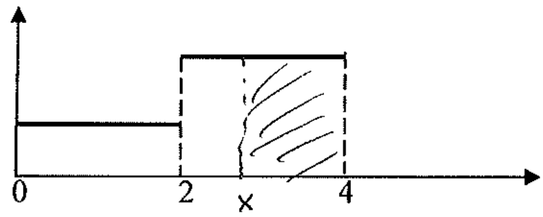
Spring 2013 Final Exam Solutions

Question 1. (20 points) The probability density function $f(x)$ of a random variable X is given as follows:

$f(x) = \frac{1}{6}$ for $0 < x < 2$, $f(x) = \frac{1}{3}$ for $2 < x < 4$, and $f(x) = 0$ elsewhere.

Hint: The sketch is below.



The expected value of X is $\frac{7}{3}$, and the variance of X is 1.22

(a) Find the median

$$P(0 < X < 2) = 2\left(\frac{1}{6}\right) = \frac{1}{3}$$

$$P(2 < X < x) = (x-2)\cdot\frac{1}{3} = \frac{x-2}{3} \quad \text{where } x \text{ : median}$$

$$\Rightarrow \frac{1}{3} + \frac{x-2}{3} = \frac{1}{2} \Rightarrow \frac{x-1}{3} = \frac{1}{2} \Rightarrow x = \frac{5}{2} = 2.5$$

(b) Find the median of the sum of 40 random variables, each of which has the same probability density function given in part (a)

CLT $\Rightarrow X_1 + \cdots + X_{40} \sim N(40\mu, \sqrt{40}\,\sigma)$ approximately.

Mean = median for normal random variables.

So, median $= 40\mu = 40\left(\frac{7}{3}\right) \cong 93.33$

(c) Find the probability that the 40 random variables in part (b) have a sum of more than 100.

$$\sqrt{40}\,\sigma = \sqrt{40}\sqrt{1.22} \cong 6.986$$

$$P(X_1 + \cdots + X_{40} > 100) \cong P\left(Z > \frac{100 - 93.33}{6.986}\right)$$

$$= P(Z > 0.95)$$

$$= 0.5 - 0.3289$$

$$= 0.1711$$

Question 2. (30 points) Cacık is a popular dish in Törki. Sometimes garlic is added, and sometimes it is not. It is believed that 32% of all Törks like garlic in their cacık. Dr. Çokmeraklı, who is investigating this important gastronomical issue, asks a random sample of 74 about their preference, and 33 say that they like garlic in their cacık.

(a) Construct a 98% Confidence Interval for the proportion of Törks who like garlic in their cacık.

$\hat{p} = \frac{33}{74} \cong 0.45$ $\quad 0.45 \mp 2.33 \sqrt{\frac{(0.45)(0.55)}{74}}$

$z_{0.01} = 2.33 \quad \Rightarrow [0.32, 0.58]$

(b) Is the proportion of Törks who like garlic in their cacık more than 32%? Answer this question using the P-value.

$H_0: p = 0.32$

$H_a: p > 0.32$

$z^* = \frac{0.45 - 0.32}{\sqrt{\frac{(0.32)(0.68)}{74}}} \cong 2.40$

P-value $= P(Z > 2.40) = 0.5 - 0.4918$
$= 0.0082 \Rightarrow$ Since P-value < 0.05 reject Ho.

Yes, the proportion is significantly more than 32%.

(c) In Griys, cacık is called Tsatsiki, because it's hard for them to say 'cacık'. In a survey made in Griys, 23 out of 85 randomly sampled Griyks declare that they like garlic in their Tsatsiki. Is the proportion of people who like garlic in this dish lower in Griys than in Törki? Test with significance 0.01.

$\hat{p}_1 = 0.45 \quad \hat{p}_2 = \frac{23}{85} = 0.27$

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 > 0$

$\bar{p} = \frac{33+23}{74+85} \cong 0.35$

$z^* = \frac{0.45 - 0.27}{\sqrt{(0.35)(0.65)\left(\frac{1}{74} + \frac{1}{85}\right)}} \cong 2.37$

$z_{0.01} = 2.33$ Since $z^* > z_{0.01}$, reject Ho.

Yes, the proportion is significantly lower in Griys.

(d) In a binomial table for n=74 and p=0.32, the numbers given for k=32, 33, and 34, are 0.9844, 0.9916, and 0.9957, respectively. What is the probability that 33 out of a random sample of 74 people like garlic in cacık, if the population proportion really is 32%? (You must use the numbers in the binomial table to find the solution)

$P(X=33) = P(X \leq 33) - P(X \leq 32)$
$= 0.9916 - 0.9844 = 0.0072$

(e) The normal approximation to the binomial distribution is good if the interval $[\mu - 3\sigma, \mu + 3\sigma]$ is included in the interval $[0,n]$, where μ and σ are the mean and standard deviation of the binomial distribution. Is this the case for the binomial distribution in part (d)?

$\mu = np = 74(0.32) = 23.68$

$\sigma = \sqrt{npq} = \sqrt{74(0.32)(0.68)} = 4.01$

$\Rightarrow [11.65, 35.71]$ yes, included in $[0, 74]$.

Question 3. (20 points) A new billing system is to be established in a consulting firm. The manager thinks that the new system is cost effective only if the mean monthly account is more than \$170. Using the new billing system, a random sample of monthly accounts of 400 customers is drawn, for which the sample mean is \$178 and the standard deviation is \$65.

a) Can the manager conclude that the new system is cost effective at $\alpha = 0.05$?

$H_0: \mu = 170$

$H_a: \mu > 170$

$$z^* = \frac{178-170}{65/\sqrt{400}} \cong 2.46 \qquad z_{0.05} = 1.65$$

Since $z^* > z_{0.05}$, reject H_0.

Yes, the new system is cost effective.

b) The old billing system of the firm had previously produced a mean monthly account of \$167 in a random sample of 300 customers, where the standard deviation was \$82. Is there a significant difference between the old and the new system? Find the P-value.

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

$$z^* = \frac{167-178}{\sqrt{\frac{65^2}{400} + \frac{82^2}{300}}} = -1.92$$

$P(Z < -1.92) = 0,5 - 0,4726 = 0,0274$

p-value $= 2(0.0274) = 0.0548$

Since p-value > 0.05, do not reject H_0.

No, there is no significant difference between the old and the new system.

Question 4. (12 points) In a gas station, the daily demand for regular gasoline (95 oktan benzin) is normally distributed with a mean of 4000 liters and a standard deviation of 400 liters.

a) The z-score of demand for a given day is 1.3. How many liters of gasoline is the demand for that day?

$$z = \frac{x-\mu}{\sigma} \Rightarrow 1.3 = \frac{x-4000}{400}$$

$$\Rightarrow x = 4520 \text{ liters}$$

b) Today, there is 4500 liters in storage. What is the probability that the gasoline in storage is sufficient for today's demand?

$$P(X < 4500) = P\left(Z < \frac{4500-4000}{400}\right)$$

$$= P(Z < 1.25)$$

$$= 0.5 + 0.3944$$

$$= 0.8944$$

per day 30 days

c) Find the probability that the average demand for the month of June is less than 3900 liters.

$$P(\overline{X} < 3900) = P\left(Z < \frac{3900-4000}{400/\sqrt{30}}\right)$$

$$= P(Z < -1.37)$$

$$= 0.5 - 0.4147$$

$$= 0.0853$$

Question 5. (18 points) For investigating the eventual success of family-run businesses when the boss's son or daughter takes over, a random sample of 6 businesses is considered. The operating income before the change of the management is subtracted from the operating income after the change, to yield the following data (in ten thousand dollars):

-1.95, 0.56, 1.44, 1.5, -0.32, 0.57

a) Test if the difference is less than 0.40 at level of significance 0.0~~5~~25. Also report the P-value.

$$\bar{X} = \frac{-1.95+0.56+\dots+0.57}{6} = 0.3$$

$$s^2 = \frac{(-1.95-0.3)^2+\dots+(0.57-0.3)^2}{5} = 1.647$$

$\Rightarrow S = 1.28$

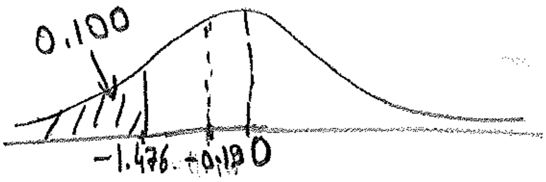
$H_0: \mu_d = 0.40$

$H_a: \mu_d < 0.40$

$$t^* = \frac{0.3-0.4}{1.28/\sqrt{6}} = -0.19$$

$t_{0.025,5} = 2.571$

Since $|t^*| < t_{0.025}$, we do not reject $H_0 \Rightarrow$ no significant difference



P-value $= P(t < -0.19) > 0.100$

since $t_{0.100} = -1.476 < -0.19$

b) A related research question for family-run businesses is that "does the business do better if the new boss is son or daughter of the owner (population 1), or if an outsider person is made chief executive officer, CEO, (population 2)"? Find a 99% confidence interval for the difference in the mean operating incomes if $\bar{x}_1 - \bar{x}_2 = -1.25$, $s_1^2 = 3.79$, $s_2^2 = 8.03$, $n_1 = 25$, $n_2 = 17$.

d.f. $n_1 + n_2 - 2 = 40 \Rightarrow t_{0.005,40} = 2.704$

$$S_p^2 = \frac{(3.79)(24)+(8.03)(16)}{40} = 5.486$$

$$-1.25 \pm 2.704\sqrt{(5.486)\left(\frac{1}{25}+\frac{1}{17}\right)}$$

$[-3.24, 0.74]$

Question 6. **(10 points)**) The operating incomes of 15 businesses before and after a change of management are obtained. Let X: income before the change, and Y: income after the change (in hundred thousand dollars).

Accordingly, the summary statistics are as follows: $\sum x_i^2 = 1672$ $\sum y_i^2 = 1764$ $\sum x_i y_i = 1691$

$\bar{x} = 9.73$ $\bar{y} = 9.87$

$\sum x_i = n\bar{x}$ $\sum y_i = n\bar{y}$ $n = 15$

a) Find the regression line.

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{1691 - \frac{(15)(9.73)\cancel{(15)}(9.87)}{\cancel{15}}}{1672 - \frac{(15)^{\cancel{2}}(9.73)^2}{\cancel{15}}}$$

$$= \frac{250.4735}{251.9065} \cong 0.99$$

$$\hat{\beta}_0 = 9.87 - 0.99(9.73) = 0.2373$$

$$\hat{y} = 0.99x + 0.2373$$

b) Find the correlation coefficient and interpret.

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}}$$

$$SS_{yy} = 1764 - \frac{(15)^{\cancel{2}}(9.87)^2}{\cancel{15}} = 302.7465$$

$$= \frac{250.4735}{\sqrt{(251.9065)(302.7465)}}$$

$$\cong 0.91$$

r is high, which implies a strong positive relationship between incomes before and after change.