**STAT400 Final.**

**1.** *Good memory* company wants to send advetisements to senior citizens of Happy Town. They send their booklets to all customers of *SuperSlim* company. They base their strategy in the following data. It is known that 25% of adults in that town age aged between 20 and 39, 40% are aged between 40 and 59 and 35% are 60 or older. Among people aged 20-39, 28% are obese, among people aged 40-59, 34% are obese and among people aged 60 or older 42% are obese.

(a) Find a probability that a randomly chosen citizen of Happy Town is aged 60 or older and is obese.

(b) Find a probability that a randomly chosen citizen of Happy Town is obese.

(c) Given that a citizen of the Happy Town is obese find the probability that they are aged 60 or older.

**Solution.** Let $Y$ denote the event that the citizen is aged between 20 and 39, $M$ denote the event that the citizen is aged between 40 and 59, $S$ denote the event that the citizen is aged 60 or older and $O$ denote the event that a citizen is obese.

(a) $P(SO) = P(S)P(O|S) = 0.35 * 0.42 = 0.147$.

(b) $P(O) = P(SO) + P(MO) + P(YO) = 0.35 * 0.42 + 0.40 * 0.34 + 0.25 * 0.28 = 0.353$.

(c) $P(S|O) = \frac{P(SO)}{P(O)} = \frac{147}{353} \approx 0.42$.

**2.** Let $X$ have density equal to $\frac{x}{2}$ if $0 \leq x \leq 2$ (and equal to 0 otherwise).

(a) Compute $EX$.

(b) Compute $EX^2$.

(c) Compute $VX$.

(d) Compute $\mathrm{Cov}(X, X^2)$.

(e) Find the cumulative distribution function of $X$;

(f) Find 25th percentile of the distribution of $X$.

**Solution.**

$$(a) \quad EX = \int_0^2 \frac{x}{2} * x dx = \int_0^2 \frac{x^2}{2} dx = \frac{x^3}{6}\Big|_0^2 = \frac{4}{3}.$$

$$(b) \quad EX^2 = \int_0^2 \frac{x}{2} * x^2 dx = \int_0^2 \frac{x^3}{2} dx = \frac{x^4}{8}\Big|_0^2 = 2.$$

$$(c) \quad VX = EX^2 - (EX)^2 = 2 - \frac{16}{9} = \frac{2}{9}.$$

(d) $\mathrm{Cov}(X, X^2) = E(X * X^2) - (EX)(EX^2) = E(X^3) - (EX)(EX^2)$.

$$EX^3 = \int_0^2 \frac{x}{2} * x^3 dx = \int_0^2 \frac{x^4}{2} dx = \frac{x^5}{10}\Big|_0^2 = \frac{16}{5}.$$

Thus $\mathrm{Cov}(X, X^2) = \frac{16}{5} - 2 * \frac{4}{3} = \frac{8}{15}$.

(e) For $0 \le x \le 2$ $F(x) = \int_0^x \frac{s}{2} ds = \frac{x^2}{4}$. Also $F(x) = 0$ if $x \le 0$ and $F(x) = 1$ if $x \ge 2$.

(f) From part (e) the percentile satisfies $\frac{x_{0.25}^2}{4} = 0.25$ so $x_{0.25} = 1$.

**3.** Error in filling a 30-lb bag of mulch has distribution $X$ with density $\frac{3x^2}{2}$ if $-1 \le x \le 1$ (and equal to 0 otherwise). Let $S$ be the combined error of weighting 1500 bags. Thus $S = X_1 + X_2 + \ldots X_{1500}$ where $X_j$ are independent and have the same distribution as $X$.

(a) Compute $ES$.

(b) Compute $VS$.

(c) Find approximately $P(S > 25)$.

**Solution.**

$$(a) \quad EX = \int_{-1}^1 \frac{3x^2}{2} * x dx = \int_{-1}^1 \frac{3x^3}{2} dx = \frac{3x^4}{8}\Big|_{-1}^1 = 0.$$

Thus $ES = 1500 * EX = 0$.

$$(b) \quad VX = EX^2 = \int_{-1}^1 \frac{3x^2}{2} * x^2 dx = \int_{-1}^1 \frac{3x^4}{2} dx = \frac{3x^4}{10}\Big|_{-1}^1 = \frac{3}{5}.$$

Thus $ES = 1500 * VX = 900$. Accordingly $\sigma_S = \sqrt{900} = 30$.

(c) By Central Limit Theorem $S \approx N(0, 900) = 0 + 30Z$ where $Z$ is standard normal. Hence $P(S > 25) \approx P(30Z > 25) \approx P(Z > 0.83) = 1 - P(Z < 0.83) \approx 1 - 0.80 = 0.20$.

**4.** The EPA sets an airborne limit of 5 parts per million (ppm) on vinyl chloride, a colorless gas used to make plastics, adhesives, and other chemicals. It is both a carcinogen and a mutagen (New Jersey Department of Health, Hazardous Substance Fact Sheet, 2009). A major plastic manufacture, attempting to control the amount of vinyl chloride its workers are exposed to, has given instructions to halt production if the mean amount of vinyl chloride in the air exceeds 3.0 ppm. A random sample of 50 air specimens produced the following statistics: $\bar{x} = 3.1$ and $s = 0.5$.

(a) Construct 99% confidence interval for the average level of vinyl chloride.

(b) Find 99% of lower confidence bound for the average level of vinyl chloride.

(c) Do the observations above provide sufficient evidence to halt the production at confidence level 99%?

(d) For the hypothesis test descibed in part (c) explain in words the meaning of type I and type II errors.

**Solution.** (a) $z_{0.005} = 2.58$. Hence the confidence interval takes form

$$\bar{x} \pm z_{0.005}\frac{s}{\sqrt{n}} = 3.1 \pm \frac{2.58 * 0.5}{\sqrt{50}} = 3.1 \pm 0.18 = [2.92, 3.28].$$

(b) $z_{0.01} = 2.33$ hence the lower bound is $\mu > 3.1 - \frac{2.33*0.5}{\sqrt{50}} = 3.1 - 0.16 = 2.94$.

(c) We have to test $H_0 = \{\mu = 3\}$ versus $H_a = \{\mu > 3\}$. Since 3 is above the lower confidence bound of 2.94 there is not sufficient evidence that the level of vinyl chloride is above 3ppm so we will not recommend to halt the production.

(d) Type I error is if we halt production process when in fact the level of vynil chloride is 3 ppm of below. Type II error is if we do not halt the production while the level of vynil chloride is above 3 ppm. Type II error is less serious since even if the actual level of vynil chloride was 3.28 it would still be much lower than the safe level of 5 pmm. This explains our choice of null and alternative hypothesis in part (c).

**5.** A sample of 9 radon detectors of a certain type was selected and each was exposed to 100pCi/L of radon. The test resulted in $\bar{x} = 101.0$ and $s = 1.2$. Suppose that results of radon measurement have normal distribution.

(a) Construct 95% confidence interval for population mean;

(b) Do the observations above provide sufficient evidence that the population mean differs from 100pCi/L using $\alpha = 0.05$?

(c) If you buy a radon detector of the type described above and expose it to 100pCi/L of radon, give a 95% prediction interval for the resulting reading.

**Solution.** (a) The CI has form

$$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} = 101 \pm \frac{2.306 * 1.2}{\sqrt{9}} = [100.08, 101.92].$$

(b) We have to test $H_0 = \{\mu = 100\}$ vs $H_a = \{\mu \neq 100]$. Since 100 is not inside our confidence interval there is sufficent evidence that the detectors are inaccurate.

(c) The prediction interval has form

$$\bar{x} \pm st_{\alpha/2,n-1}\sqrt{1 + \frac{1}{n}} = 101 \pm \frac{2.306 * 1.2\sqrt{10}}{\sqrt{9}} = [98.08, 103.92].$$