**CONCORDIA UNIVERSITY**
**FACULTY OF ENGINEERING AND COMPUTER SCIENCE**

**Probability and Statistics in Engineering**
(ENGR 371 Winter 2008)

## Instructions

- This is a closed book exam.
- Standard type calculator may be used.
- You may use the formula sheet which is attached.
- You may only write on the provided exam booklets.
- You may not separate any sheet from the exam paper.
- Cellular phones are not allowed.
- Read carefully all questions
- Show all the intermediate steps of your solution.
- Make reasonable assumptions if necessary.
- Please do not write in red (colour used for correction)
- Questions answered on the question paper will not be corrected.
- Everything which is not readable will not be corrected
- Available time is 3 hours

**Exam has two parts each 50% as follows:**

**Part I - Probability**: Five questions are provided. Choose four questions and answer them.

**Part II – Statistics:** Five questions are provided. Choose four questions and answer them.

Two exam booklets are provided. Use one of them specifically for “Probability” and the other one for “Statistics”. Write your name, ID number and subject of the exam (Probability or Statistics) on the cover page.

**Important note:**

If you answer five questions in part I or part II of the exam, question P5 or S5 will not be considered, respectively.

# Probability Questions

**P1:** A product is made by three different manufacturers. Manufacturers 1, 2 and 3 produce 20%, 35% and 45% of the product, respectively. It is known from the past experience that 1%, 1% and 2% of the products manufactured by each manufacturer, respectively, are defective.

a) If a finished product is randomly selected, what is the probability that it is defective?

b) If a finished product is randomly selected and is found to be defective, which manufacturer was most likely used?

**Part a: (2.5 marks)**

D: the event that the selected product is defective

M1: the event that the selected product is manufactured by manufacturer 1

M2: the event that the selected product is manufactured by manufacturer 2

M3: the event that the selected product is manufactured by manufacturer 3

Using total probability:

$$\begin{aligned} P(D) &= P(D \mid M1)P(M1) + P(D \mid M2)P(M2) + P(D \mid M3)P(M3) \\ &= 0.01 \times 0.20 + 0.01 \times 0.35 + 0.02 \times 0.45 = 0.0145 \end{aligned}$$

**Answer:** $P(D) = 0.0145$

**Part b: (2.5 marks)**

Using Bayes' theorem:

$$P(M1 \mid D) = \frac{PD \mid M1) \times P(M1)}{P(D)} = \frac{0.01 \times 0.20}{0.0145} = \frac{4}{29} = 0.1379$$

$$P(M2 \mid D) = \frac{PD \mid M2) \times P(M2)}{P(D)} = \frac{0.01 \times 0.35}{0.0145} = \frac{7}{29} = 0.2414$$

$$P(M3 \mid D) = 1 - P(M1 \mid D) - P(M1 \mid D) = 1 - \frac{4}{29} - \frac{7}{29} = \frac{18}{29} = 0.6207$$

$P(M3 \mid D) > P(M2 \mid D) > P(M1 \mid D)$ and therefore most likely manufacturer 3 has been used.

**Answer:** most likely manufacturer 3

**P2:** The percentage of people exposed to a bacteria who become ill is 20%. Assume that people are independent and 100 people are exposed to the bacteria.

a) Evaluate the probability that less than 5 people become ill.
b) Evaluate the probability that more than 60 people become ill.

**Part a: (2.5 marks)**

X is a random variable indicating the number of people becoming ill and has binomial distribution with mass function of $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $n$=100 and $p$=0.2

$$P(X < 5) = P(X \le 4) = \sum_{x=0}^{4} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \binom{100}{0} 0.2^0 (1-0.2)^{100} + \binom{100}{1} 0.2^1 (1-0.2)^{99} + \binom{100}{2} 0.2^2 (1-0.2)^{98} + \binom{100}{3} 0.2^3 (1-0.2)^{97} + \binom{100}{4} 0.2^4 (1-0.2)^{96}$$

$$P(X < 5) = 2.037\times10^{-10} + 5.092\times10^{-9} + 6.302\times10^{-8} + 5.146\times10^{-7} + 3.120\times10^{-6} = 3.7\times10^{-6}$$

**Answer:** $P(X < 5) = 3.7\times10^{-6}$

**Part b: (2.5 marks)**

Since X is a binomial random variable with $\mu = E(X) = np$ and $\sigma^2 = V(X) = np(1-p)$ then $Z = \frac{X-\mu}{\sigma}$ is approximately a standard normal random variable and

$P(X \le x) = P\left(Z \le \frac{x+0.5-\mu}{\sigma}\right)$. The approximation is good for $np > 5$ and $n(1-p) > 5$.

Since $np = 100\times0.2 = 20 > 5$ and $n(1-p) = 100\times0.8 = 80 > 5$, so we can use above approximation.

$\mu = np = 100\times0.2 = 20$ and $\sigma = \sqrt{np(1-p)} = \sqrt{100\times0.2\times(1-0.2)} = 4$

$$P(X > 60) = 1 - P(X \le 60) = 1 - P\left(Z \le \frac{60.5-20}{4}\right) = 1 - P(Z \le 10.125)$$

From table $P(Z \le 10.125) = 1$ and therefore $P(X > 60) = 1 - 1 = 0$

**Answer:** $P(X > 60) = 0$

**P3:** To evaluate the technical support from a computer manufacturer, the number of rings before a call is answered by a service representative is tracked. Historically, 70% of the calls are answered in two rings or less, 25% are answered in three or four rings, and the remaining calls require five rings or more. Suppose you call this manufacturer 10 times and assume the calls are independent.

a) What is the probability that eight calls are answered in two rings or less, one call is answered in three or four rings, and one call requires five rings or more?
b) What is the conditional distribution of the number of calls requiring five rings or more given that eight calls are answered in two rings or less.
c) Are the number of calls answered in two rings or less and the number of calls requiring five rings or more independent random variables? Why?

**Part a: (1.5 marks)**

Let X, Y, and Z denote the number of calls answered in two rings or less, three or four rings, and five rings or more, respectively. The joint mass function is extension of binomial with *p*=0.7, *q*=0.25, *r*=0.05, where n=10.

$$f_{XYZ}(x,y,z)=\frac{n!}{x!\,y!\,z!}p^x q^y r^z$$

$$P(X=8,Y=1,Z=1)=f_{XYZ}(8,1,1)=\frac{10!}{8!1!1!}0.7^8 0.25^1 0.05^1=0.0649$$

**Answer:** $P(X=8,Y=1,Z=1)=0.0649$

**Part b: (2 marks)**

$$f_{XZ}(x,z)=\frac{10!}{x!\,z!(10-x-z)!}0.7^x 0.25^{(10-x-z)}0.05^z \Rightarrow f_{XZ}(8,0)=0.1617,\ f_{XZ}(8,1)=0.0645,\ f_{XZ}(8,2)=0.0063$$

$$f_X(x)=\frac{10!}{x!(10-x)!}0.7^x 0.3^{(10-x)} \Rightarrow f_X(8)=0.233$$

$$f_{Z|8}(z)=\frac{f_{XZ}(8,z)}{f_X(8)}=\frac{\frac{10!}{8!\,z!(10-8-z)!}0.7^8 0.25^{(10-8-z)}0.05^z}{\frac{10!}{8!(10-8)!}0.7^8 0.3^{(10-8)}}=\frac{(10-8)!0.25^{(10-8-z)}0.05^z}{(10-8-z)!0.3^{(10-8)}}=\frac{2!0.25^{(2-z)}0.05^z}{(2-z)!0.3^2}$$

**Answer:** $f_{Z|8}(z)=\frac{2!0.25^{(2-z)}0.05^z}{0.3^2 z!(2-z)!}$ : $f_{Z|8}(0)=0.694$, $f_{Z|8}(1)=0.277$, $f_{Z|8}(0)=0.027$

**Part c: (1.5 marks)**

$f_Z(z)=\frac{10!}{z!(10-z)!}0.95^{(10-z)}0.05^z$ and knowing $f_{XZ}(x,z)$ & $f_X(x)$ from part b we can easily show that

**Answer:** $f_{XZ}(x,z)\neq f_X(x)f_Y(y)$ and therefore *X* and *Y* are dependent

**P4:** The permeability of a membrane used as a moisture barrier in a biological application depends on the thickness of three integrated layers. Layers 1, 2, and 3 are normally distributed with means of 0.5, 1, and 1.5 millimeters, respectively. The standard deviations of layer thickness are 0.1, 0.2, and 0.3, respectively. Also, the correlation between layers 1 and 2 is 0.7, between layers 2 and 3 is 0.5, and between layers 1 and 3 is 0.3.

a) Determine the mean and variance of the total thickness of the three layers.
b) What is the probability that the total thickness is less than 3.2 millimeters?

**Part a: (3 marks)**

$X_1$:Normal random variable indicating thickness of layer 1with mean 0.5 and standard deviation of 0.1

$X_2$ :Normal random variable indicating thickness of layer 2 with mean 1 and standard deviation of 0.2

$X_3$ :Normal random variable indicating thickness of layer 3 with mean 1.5 and standard deviation of 0.3

*T*: Random variable indicating total thickness. $T = X_1 + X_2 + X_3$

$\mu_T = E(T) = E(X_1) + E(X_2) + E(X_3) = 0.5 + 1 + 1.5 = 3\,\text{mm}$

**Answer**: $\mu_T = 3$ mm

$Cov(X_i, X_j) = \rho\sigma_{X_i}\sigma_{X_j}$, therefore $Cov(X_1, X_2) = 0.7 \times 0.1 \times 0.2 = 0.014$,
$Cov(X_1, X_3) = 0.3 \times 0.1 \times 0.3 = 0.009$, $Cov(X_2, X_3) = 0.5 \times 0.2 \times 0.3 = 0.030$

$$V(T) = V(X_1) + V(X_2) + V(X_3) + 2Cov(X_1, X_2) + 2Cov(X_1, X_3) + 2Cov(X_2, X_3)$$
$$V(T) = 0.1^2 + 0.2^2 + 0.3^2 + 2 \times 0.014 + 2 \times 0.009 + 2 \times 0.030 = 0.246$$

**Answer**: $\sigma_T^2 = 0.246\ mm^2$

**Part b: (2 marks)**

The total thickness *T* has normal distribution with mean of $\mu_T = 3\,\text{mm}$ and standard deviation of $\sigma_T = \sqrt{V(T)} = \sqrt{0.246} = 0.4959\,\text{mm}$ and therefore:

$P(X < 3.2) = P\left(Z < \dfrac{3.2 - 3}{0.4959}\right) = P(Z < 0.403)$ and from table III, $P(Z < 0.403) = 0.666$

**Answer**: $P(X < 3.2) = 0.655422$

**P5:** A popular clothing manufacturer receives Internet orders via two different routing systems. The time between orders for each routing system in a typical day is known to be exponentially distributed with mean of 3.2 minutes. Both systems operate independently.

a) What is the probability that no orders will be received in a 5 minute period?
b) What is the probability that both systems receive two orders between 10 and 15 minutes after the site is officially open for business.

**Part a: (2.5 marks)**

X: Time between two orders for routing system 1 with exponential density function

Y: Time between two orders for routing system 2 with exponential density function

Mean and parameter of the exponential density functions are $\mu = 3.2$ & $\lambda = \frac{1}{\mu} = \frac{1}{3.2}$.

$f_X(x) = \frac{1}{3.2} e^{-\frac{x}{3.2}}$, $f_Y(y) = \frac{1}{3.2} e^{-\frac{y}{3.2}}$ and since X and Y are independent $f_{XY}(x, y) = f_X(x) f_Y(y)$

$$P(X > 5, Y > 5) = \int_5^\infty \int_5^\infty f_{XY}(x, y)dxdy = \int_5^\infty f_Y(y) \int_5^\infty f_X(x)dxdy = \int_5^\infty \frac{1}{3.2} e^{-\frac{y}{3.2}} \int_5^\infty \frac{1}{3.2} e^{-\frac{x}{3.2}} dxdy$$

$$= \frac{1}{3.2 \times 3.2} \int_5^\infty e^{-\frac{y}{3.2}} \int_5^\infty e^{-\frac{x}{3.2}} dxdy = \frac{1}{3.2} \int_5^\infty e^{-\frac{y}{3.2}} e^{-\frac{5}{3.2}} dy = e^{-\frac{5}{3.2}} e^{-\frac{5}{3.2}} = 0.20961^2 = 0.0439$$

**Answer:** $P(X > 5, Y > 5) = 0.0439$

**Part b: (2.5 marks)**

Let X denote the number of orders in a 5-minute interval for routing system 1, then $X$ is poisson random variable with $\lambda = \frac{5}{3.2} = 1.5625$ and $f_X(x) = \frac{e^{-\lambda} \lambda^x}{x!}$

Let Y denote the number of orders in a 5-minute interval for routing system 2, then Y is poisson random variable with $\lambda = \frac{5}{3.2} = 1.5625$ and $f_Y(y) = \frac{e^{-\lambda} \lambda^y}{y!}$

Since they are independent: $f_{XY}(x, y) = f_X(x) f_Y(y)$

$$P(X = 2, Y = 2) = f_{XY}(2,2) = f_X(2) f_Y(2) = \frac{e^{-1.5625} (1.5625)^2}{2!} \times \frac{e^{-1.5625} (1.5625)^2}{2!} = 0.256^2$$

**Answer:** $P(X = 2, Y = 2) = 0.0655$

# Statistics Questions

## Select four questions:

**S1:** If X is a binomial random variable with parameters *p* (probability of a success) and *n* (number of Bernoulli trials) verify that

a) $\hat{P} = X / n$ is a unbiased estimator of *p*,

b) $\hat{Q} = \dfrac{X + \sqrt{n}/2}{n + \sqrt{n}}$ is a biased estimator of *p*. Evaluate the value of bias.

**Solution S1:**

a) $E\left(\hat{P}\right) = E(X/n) = \dfrac{E(X)}{n} = \dfrac{\mu}{n} = \dfrac{np}{n} = p$ and $E\left(\hat{P}\right) = p$ and therefore $\hat{P} = X/n$ is an unbiased estimator of *p*.

b) $E(\hat{Q}) = E\left(\dfrac{X + \sqrt{n}/2}{n + \sqrt{n}}\right) = \dfrac{E(X + \sqrt{n}/2)}{n + \sqrt{n}} = \dfrac{E(X) + \sqrt{n}/2}{n + \sqrt{n}} = \dfrac{\mu + \sqrt{n}/2}{n + \sqrt{n}} = \dfrac{np + \sqrt{n}/2}{n + \sqrt{n}}$

$$E(\hat{Q}) = \left(\frac{n}{n + \sqrt{n}}\right)p + \frac{\sqrt{n}/2}{n + \sqrt{n}}$$

$$\text{Bias is } = \left(\frac{n}{n + \sqrt{n}}\right)p + \frac{\sqrt{n}/2}{n + \sqrt{n}} - p = \left(\frac{n}{n + \sqrt{n}} - 1\right)p + \frac{\sqrt{n}/2}{n + \sqrt{n}}$$

**S2:** An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with a standard deviation of 40 hours.

a) If a sample of 30 bulbs has an average life of 780 hours, find a 95% confidence interval for the population mean of all bulbs produced by this firm.
b) If the population is not normal distributed, can you still compute the 95% confidence interval?
c) How large a sample is needed if we wish to be 96% confident that our sample mean will be within 10 hours of the true mean?

**Solution S2:**

a) This is a 2-sided 95% ($1 - \alpha = 0.95$) confidence interval problem for population mean ($\mu$) when the population standard deviation is known ($\sigma = 40$). We have a sample of size $n = 30$ with sample mean of $\bar{x} = 780$.

$$\bar{x} - z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{x} + z_{\alpha/2}\sigma/\sqrt{n}$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \Rightarrow z_{\alpha/2} = z_{0.025} = 1.96$$

$$780 - 1.96 \times 40/\sqrt{30} \le \mu \le 780 + 1.96 \times 40/\sqrt{30}$$

So, the 95% confidence interval is:

$$765.69 \le \mu \le 794.31$$

b) In many cases of practical interest, if $n \geq 30$ , a random sample of size *n* has normal distribution even if the population is not normally distributed. This is the case for this problem; since $n = 30$, we could find the 95% confidence interval very close to that of part a.

c) Confidence interval = $1 - \alpha = 0.96$, so $\alpha / 2 = 0.02$.
From normal table: $z_{\alpha/2} = z_{0.02} = 2.06$

$$u = \bar{x} - z_{\alpha/2}\sigma / \sqrt{n} \Rightarrow u - \bar{x} = z_{\alpha/2}\sigma / \sqrt{n} \Rightarrow 10 = 2.06 \times 40 / \sqrt{n} \Rightarrow n = 67.89$$

So the sample size should be $n = 68$.

**S3:** The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint:

| | | | | |
|---|---|---|---|---|
| 3.4 | 2.5 | 4.8 | 2.9 | 3.6 |
| 2.8 | 3.3 | 5.6 | 3.7 | 2.8 |
| 4.4 | 4.0 | 5.2 | 3.0 | 4.8 |

a) Assuming the measurements represent a random sample from a normal population, find the 95% confidence interval for the population mean.
b) Evaluate the P-value of a test based on the above observed samples such that the drying time is larger than 3.9 hours.

## Solution S3:

a) $$\bar{x} = \frac{\sum_{i=1}^{15} x_i}{15} = 3.786667$$

$$s^2 = \frac{\sum_{i=1}^{15} x_i^2 - \frac{\left(\sum_{i=1}^{15} x_i\right)^2}{15}}{14} = \frac{228.08 - \frac{56.8^2}{15}}{14} = 0.942667 \text{ and } s = 0.970910$$

The sample size is $n = 15$. Since $\bar{x} = 3.786667$ and $s = 0.970910$ are the mean and standard deviation of a random sample from a normal distribution with unknown variance, a $95\% = 100(1-\alpha)\%$ $(\alpha = 0.05)$ confidence interval on $\mu$ is given by

$$\bar{x} - t_{\alpha/2,n-1} s / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2,n-1} s / \sqrt{n}$$

and

$$3.786667 - t_{0.025,14} \times 0.970910 / \sqrt{15} \leq \mu \leq 3.786667 + t_{0.025,14} \times 0.970910 / \sqrt{15}$$

where $t_{\alpha/2,n-1} = t_{0.025,14}$ is the upper $100\alpha / 2 = 2.5$ percentage point of $t$ distribution with $n - 1 = 14$ degrees of freedom.

$t_{0.025,14} = 2.145$ and therefore, the 0.95% confidence interval is:

$$3.24894 \leq \mu \leq 4.32439$$

b) We use,
$H_0 : \mu = 3.9$ vs. $H_1 : \mu > 3.9$ (the drying time greater than 3.9 hr is not desired)

$$t_0 = \left| \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \right| = \left| \frac{3.786667 - 3.9}{0.970910 / \sqrt{15}} \right| = 0.45$$

| $t_{\alpha,14}$ | $\alpha$ |
|---|---|
| 0.258 | 0.40 |
| 0.45 | $\alpha_{0.45}$ |
| 0.692 | 0.25 |

By linear interpolation:

$\alpha_{0.45} = 0.33$

$p\text{-value} = 1 - \alpha_{0.45} = 1 - 0.33 = 0.67$

**S4:** A manufacturer has developed a new fishing line, which he claims has a mean breaking strength of 15 kilograms with a standard deviation of 0.5 kilogram. To test the hypothesis that $\mu = 15$ kilograms against the alternative that $\mu \neq 15$ kilograms, a random sample of 50 lines will be tested.

a) If the critical values are 14.9 and 15.1, find the probability of committing a type I error.
b) If the probability of type I error is 0.05, evaluate the probability of type II error for the alternative $\mu = 14.8$ kilograms.

**Solution 4:**

a) Probability of type I error = $\alpha = P(\overline{X} < 14.9 \quad when \quad \mu = 15) + P(\overline{X} > 15.1 \quad when \quad \mu = 15)$

$z$-values corresponding to the critical values 14.9 and 15.1 are

$$z_1 = \frac{14.9 - 15}{0.5/\sqrt{50}} = -1.41 \text{ and } z_2 = \frac{15.1 - 15}{0.5/\sqrt{50}} = 1.41$$

$$\alpha = P(Z < -1.41) + P(Z > 1.41) = 2 \times P(Z < -1.41) = 2 \times 0.079270$$

Probability of type I error = $\alpha = 0.15854$

b) If the probability of type I error is $\alpha = 0.05$, then $z_{0.025} = 1.96$ and the critical values are:

$$l = 15 - \frac{1.96(0.5)}{\sqrt{50}} = 14.861 \qquad \text{and} \qquad u = 15 + \frac{1.96(0.5)}{\sqrt{50}} = 15.139$$

Probability of type II error = $\beta = P(14.861 \leq \overline{X} \leq 15.139 \quad when \quad \mu = 14.8)$

$z$-values corresponding to the critical values 14.861 and 15.139 when $\mu = 14.8$ are

$$z_1 = \frac{14.861 - 14.8}{0.5/\sqrt{50}} = 0.86 \text{ and } z_2 = \frac{15.139 - 14.8}{0.5/\sqrt{50}} = 4.79$$

$$\beta = P(0.86 \leq Z \leq 4.79) = P(Z \leq 4.79) - P(Z \leq 0.86) = 1.000000 - 0.805106 = 0.194894$$

Probability of type II error = $\beta = 0.194894$

**S5:** The contents of a random sample of 10 containers are 10.2, 9.7, 10.1, 10.3, 10.1, 9.8, 9.9, 10.4, 10.3 and 9.8 litters.

a) Test the hypothesis that the average content of containers of a particular lubricant is 10 litters. Use a 0.01 of significance and assume that the distribution of the contents is normal. Formulate appropriate hypothesis test.

b) Test the hypothesis that $\sigma^2 = 0.03$ against the alternative that $\sigma^2 < 0.03$. Consider a type I error of 0.01.

**Solution S5:**

$n = 10$

$s = 0.2459$

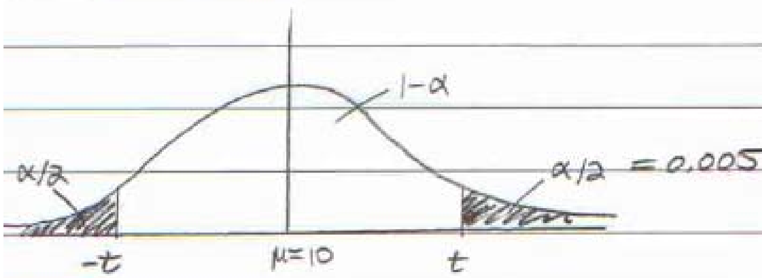
$\bar{x} = 10.06$ ✓

$H_0 : \mu = 10$ $\alpha = 0.01$ normally distributed

$H_1 : \mu \neq 10$ $\nu = 9$ t-distribution

✓ ✓ – $\sigma$ unknown

– Mean for $\mu$.




$\alpha = 0.01$

$\alpha/2 = 0.005$

$t_{0.005, 9} = 3.250$ ✓

$$10 - \frac{3.250(0.2459)}{\sqrt{10}} \leq \mu \leq 10 + \frac{3.250(0.2459)}{\sqrt{10}}$$

$$9.747 \leq \mu \leq 10.253$$

$\bar{x}$ is in the range so we fail to reject Ho ✓

**Part b)**

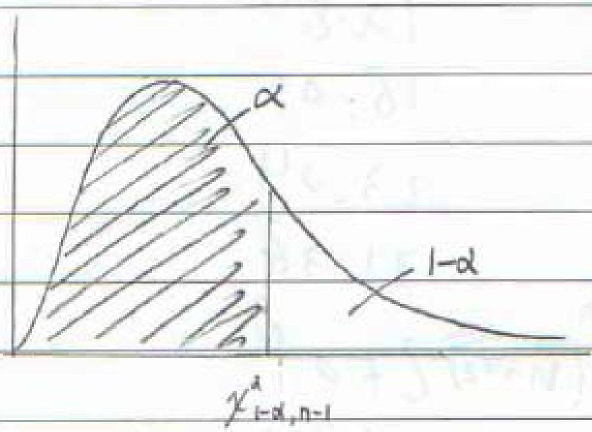
$H_0 : \sigma^2 = 0.03$ ✓ $\alpha = 0.01$

$H_1 : \sigma^2 < 0.03$ ✓ $\nu = 9$

$\chi^2$ distribution ✓

$s^2 = 0.06044$

$n = 10$




$\alpha = 0.01$

$1-\alpha = 0.99$

$\chi^2_{0.99,9} = 2.09$ ✓

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{9(0.06044)}{0.03} = 18.132 \checkmark$$

$\chi_0^2 > \chi^2_{0.99,9} \rightarrow$ we fail to reject $H_0$