2028: Basic Statistical Methods
Solutions - Homework 3

1. (a) i.

$$X \sim N(7000, 600^2)$$

$$P(X < 5000) = P(Z < \frac{5000 - 7000}{600}) = P(Z < -3.33) = 0.00043$$

ii.

$$P(X > x) = 0.95 \Leftrightarrow P(Z > \frac{x - 7000}{600}) = 0.95$$

$$\frac{x - 7000}{600} = -1.64$$

$$x = 6016$$

iii.

$$P(X > 7000) = P(Z > \frac{7000 - 7000}{600}) = P(Z > 0) = 0.5$$

$$P(\text{three lasers operating after 7000 hours}) = (0.5)^3 = \frac{1}{8}$$

(b) $\chi^2_{0.05,10} = 18.31, \chi^2_{0.025,15} = 27.49, \chi^2_{0.01,12} = 26.22, \chi^2_{0.95,20} = 10.85, \chi^2_{0.99,18} = 7.01, \chi^2_{0.995,16} = 5.14, \chi^2_{0.005,25} = 46.93$

$\chi^2_{0.1,8} = 13.36, \chi^2_{0.5,8} = 7.34 \Rightarrow \chi^2_{0.30,8} = 7.34 + \frac{13.36-7.34}{0.1-0.5}(0.3 - 0.5) = 10.35$

(c) $t_{0.025,15} = 2.131, t_{0.05,10} = 1.812, t_{0.1,20} = 1.325, t_{0.005,25} = 2.787, t_{0.001,30} = 3.385$

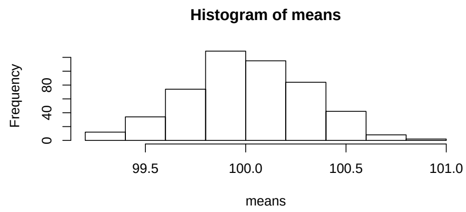
$t_{0.0025,3} = 7.453, t_{0.001,3} = 10.213 \Rightarrow t_{0.002,3} = 7.453 + \frac{10.213-7.453}{0.001-0.0025}(0.002 - 0.0025) = 8.373$

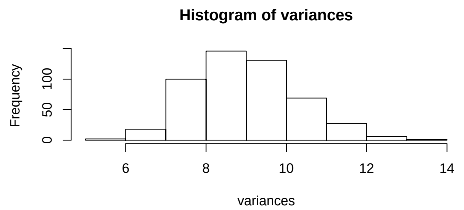
(d) $f_{0.25,7,15} = 1.47, f_{0.10,10,12} = 2.19, f_{0.01,20,10} = 4.41, f_{0.75,7,15} = \frac{1}{f_{0.25,15,7}} = \frac{1}{1.68} = 0.595, f_{0.90,10,12} = \frac{1}{f_{0.1,12,10}} = \frac{1}{2.28} = 0.439, f_{0.99,20,10} = \frac{1}{f_{0.01,10,20}} = \frac{1}{3.37} = 0.297$

$f_{0.10,2,3} = 5.46, f_{0.25,2,3} = 2.28 \Rightarrow f_{0.15,2,3} = 2.28 + \frac{5.46-2.28}{0.1-0.25}(0.15 - 0.25) = 4.4$

2. The five numerical summary for the sample from the sampling distribution of the sample mean estimator is $Q1 = 99.81$, $mean = 100.00$, $median = 100.00$, $Q3 = 100.90$ and $variance = 0.089$. The five numerical summary for the sample from the sampling distribution of the sample variance estimator is $Q1 = 5.382$, $mean = 8.950$, $median = 8.865$, $Q3 = 9.804$ and $variance = 1.662$. Note that this values will slightly differ since they are based on sample not the theoretical values.

   The histograms for the sample for $\widehat{\mu}$ and the sample of $\widehat{\sigma}^2$ are in the following figure.





The sampling distribution of $\widehat{\mu}$ is $N(\mu, \sigma^2/n)$ where $\mu = 100$, $\sigma^2 = 9$ and $n = 100$. The histogram for $\widehat{\mu}$ is approximately normal centered at 100. The estimated variance of the sample from $\widehat{\mu}$ is $variance = 0.089$ approximately equal to $\sigma^2/n = .09$.

The sampling distribution of $\widehat{\sigma}^2$ is given by

$$\frac{\widehat{\sigma}^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$$

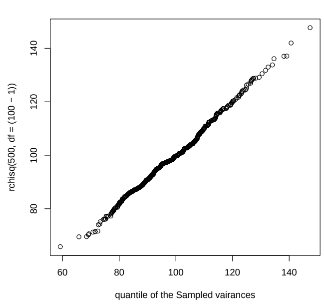
In order to compare the sampling distribution approximated by the 500 samples for $\widehat{\sigma}^2$ we will multiply the vector `variances` with $n-1$ and divided by σ^2:

```
distvar= variances*(n-1)/9
```

We compare the sample variance χ^2_{n-1} with the distribution approximated by `distvar`:

```
qqplot(distvar, rchisq(500, df = (100-1)))
```

The quantile-quantile plot shows a good approximation of the sample from $\widehat{\sigma}^2$ distribution with the theoretical sampling distribution. The quantile plot is in the figure below.



3. (a) $\bar{X}_B - \bar{X}_A \sim N(\mu_B - \mu_A, \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B})$
   If $\mu_A = \mu_B, then \bar{X}_B - \bar{X}_A \sim N(0, 2 * \frac{16^2}{25})$

$P(\bar{X}_B - \bar{X}_A > 3.5)$
$= P(\frac{\bar{X}_B - \bar{X}_A - 0}{4.52} > \frac{3.5-0}{4.52})$
$= P(Z > 0.7743)$
$= 1 - P(Z \leq 0.7743) = 1 - 0.7794 = 0.22$
So when the means are the same there is still a relatively high chance that the difference in the sample means is greater than 3.5. So we do not decide to adopt type B because we do not have conclusive evidence that $\mu_B \geq \mu_A$.

(b) p is the probability that a chip is defective
$X_i \sim$ Bernoulli(p), $\sum_{i=1}^n X_i \sim$ Binomial(n,p)
$E(\widehat{p}) = E(\frac{\sum_{i=1}^n X}{n}) = \frac{np}{n} = p$
$Var(\widehat{p}) = Var(\frac{\sum_{i=1}^n X}{n}) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$
  i. $n\widehat{p} \sim$ Binomial(50,p), $E(\widehat{p}) = p, Var(\widehat{p}) = \frac{p(1-p)}{50}$
  ii. $n\widehat{p} \sim$ Binomial(80,p), $E(\widehat{p}) = p, Var(\widehat{p}) = \frac{p(1-p)}{80}$
  iii. $n\widehat{p} \sim$ Binomial(100,p), $E(\widehat{p}) = p, Var(\widehat{p}) = \frac{p(1-p)}{100}$

(c) i. Find a Method of Moments (MOM) Estimator.
For the MOM estimator, we equate the first moment of the distribution with the sample variance. For this, we first need to find the first moment:

$$E(X) = \int_0^1 x(\theta+1)x^\theta \, dx = \left. \frac{(\theta+1)x^{\theta+1}}{\theta+2} \right|_0^1 = \frac{\theta+1}{\theta+2}$$

We therefore find the MOM by equating

$$\bar{X} = \frac{\theta+1}{\theta+2}$$

and solve this equation with respect to θ to obtain the MOM estimator

$$\widehat{\theta} = \frac{1-2\bar{X}}{\bar{X}-1}.$$

ii. Find the Maximum Likelihood Estimator (MLE).
The likelihood function is

$$L(\theta) = \prod_{i=1}^n (\theta+1)X_i^\theta = (\theta+1)^n (\prod_{i=1}^n X_i)^\theta$$

and the log-likelihood function is

$$l(\theta) = n\log(\theta+1) + \theta \sum_{i=1}^n \log(X_i)$$

with the first order derivative

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{n}{\theta+1} + \sum_{i=1}^n \log(X_i).$$

To find the MLE we equate $\frac{\partial l(\theta)}{\partial \theta} = 0$ and solve with respect to θ and obtain

$$\widehat{\theta} = \frac{n}{\sum_{i=1}^{n} \log(1/X_i)} - 1$$

this is the maximum since the second order derivative is

$$l''(\theta) = -\frac{n}{(\theta+1)^2} < 0.$$

(d) i. Find an unbiased estimator for $\mu_1 - \mu_2$ and find its standard error.
Since $\bar{X}_1$ is unbiased estimator from μ_1 and $\bar{X}_2$ is unbiased estimator for μ_2, it suggests that an unbiased estimator for $\mu_1 - \mu_2$ is $\bar{X}_1 - \bar{X}_2$:

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2.$$

Its standard error is

$$\sqrt{V(\bar{X}_1 - \bar{X}_2)} = \sqrt{V(\bar{X}_1) + V(\bar{X}_2)} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}.$$

Note that we can write the variance of the two sample means as the sum of their variances because we assume that the two samples of data are independent.

ii. Find the bias of the estimator $\bar{X}_1^2 - \bar{X}_2^2$ for the parameter $\mu_1^2 - \mu_2^2$. What happens to the bias as the sample sizes of n_1 and n_2 increase to ∞?
We find the from

$$E(\bar{X}_1^2 - \bar{X}_2^2) = E(\bar{X}_1^2) - E(\bar{X}_2^2) = (V(X_1) + E(X_1)^2) - (V(X_2) + E(X_2)^2) = (\mu_1^2 - \mu_2^2) + (\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2})$$

Therefore the bias is

$$E(\bar{X}_1^2 - \bar{X}_2^2) - (\mu_1^2 - \mu_2^2) = \frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}$$

which converges to 0 as $n_1 \to \infty$ and $n_2 \to \infty$.

iii. Assume that both populations have the same variance; that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Show that

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator of σ^2.

$$\begin{aligned} E(S_p^2) &= E\left(\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}\right) = \frac{1}{n_1 + n_2 - 2}\left[(n_1 - 1)E(S_1^2) + (n_2 - 1)E(S_2^2)\right] \\ &= \frac{1}{n_1 + n_2 - 2}\left[(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2\right] = \frac{n_1 + n_2 - 2}{n_1 + n_2 - 2}\sigma^2 = \sigma^2 \end{aligned}$$

(e) i. Bias($\widehat{\mu}_1$)=0 (unbiased),Bias($\widehat{\mu}_2$)= $-\frac{13}{60}\mu$ (biased), Bias($\widehat{\mu}_3$)=$2 - \mu/4$.
ii. Var($\widehat{\mu}_1$)=4.444, Var($\widehat{\mu}_2$)=2.682, Var($\widehat{\mu}_3$)=2.889 ($\widehat{\mu}_2$ has smallest variance).
iii. MSE($\widehat{\mu}_1$)=4.444, MSE($\widehat{\mu}_2$)= 2.682+0.0469μ^2 (when $\mu = 3$, MSE=3.104), MSE($\widehat{\mu}_3$)= 2.889 + $(2 - \mu/4)^2$ (when $\mu = 3$, MSE=4.452). ($\widehat{\mu}_2$ has smallest MSE).