Hw 1-Ex: 3, 8, 11, 12, 22, 32, 36

Julia Andrus

3a) Calculate a point estimate of the mean value of coating thickness, and state which estimator you used.

Sample of observations:

0.83	0.88	0.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

Using Sample mean to find the point estimate of mean value:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{n}\frac{\sum_{i=1}^{n} x_i}{n}$$

So,
$x_1 + x_2 + ...x_{16} = 0.83 + 0.88 + 0.88 + ...1.83$=21.57
n=16

Answer: $\frac{1}{n}\frac{\sum_{i=1}^{n} x_i}{n} = \frac{1}{16}(21.57) = 1.3481$

3b) Calculate a point estimate of the median of the coat ing thickness distribution, and state which estimator you used.

Since there are 16 ordered observations, 1.31 and 1.48 are our two values. By using the Sample Median, we can get our point estimate of the median:

Answer: $\bar{x} = \frac{1.31+1.48}{2} = 1.395$

3c) Calculate a point estimate of the value that separates the largest 10% of all values in the thickness distribution from the remaining 90%, and state which estimator you used. (Hint: Express what you are trying to estimate in terms of μ and σ.)

Finding Sample Variance:

$$s^2 = \frac{1}{n-1}\frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{1}{n-1}\sum x_i^2 - \frac{1}{n}(\sum x_i)^2$$

0.83	0.88	0.88	1.04	1.09	1.12	1.29	1.31
1.48	1.49	1.59	1.62	1.65	1.71	1.76	1.83

x_i^2**:**

0.6889	0.7744	0.7744	1.0816	1.1881	1.2544	1.6641	1.7161
2.1904	2.2201	2.5281	2.6244	2.7225	2.9241	3.0976	3.3489

Now we can find s^2:

$s^2 = \frac{1}{n-1}\frac{\sum(x_i-\bar{x})^2}{n-1} = \frac{1}{n-1}\sum x_i^2 - \frac{1}{n}(\sum x_i)^2$
$s^2 = \frac{1}{n-1}(0.6889 + 0.7744 + ...3.3489 - \frac{1}{16}(0.83 + 0.88 + ...1.83)^2)$
$s^2 = \frac{1}{n-1}(30.7981 - \frac{1}{16}(21.57)^2)$
$s^2 = \frac{1}{n-1}(1.719)$
$s^2 = \frac{1}{16-1}(1.719) = 0.1146$

Finding the sample Standard Deviation: $s = \sqrt{s^2} = \sqrt{0.1146} = 0.3385$

We now have to use $\mu + z_1 -_\alpha \sigma$ to find the 90th percentile:

$\mu = \bar{x} = 1.3481$ $\alpha = 0.1(10\%)$ $z_1 -_\alpha = 1.28$ $\sigma = s = 0.3385$
$\mu + z_1 -_\alpha \sigma = 1.3481 + 1.28(0.3385) = 1.7814$

Answer: the 90th percentile is 1.7814

3d) Estimate P(X , 1.5), i.e., the proportion of all thick ness values less than 1.5.

$$P(X < 1.5) = P(\frac{X - \bar{x}}{s} < \frac{1.5 - 1.3481}{0.3385}) = P(Z < 0.45) \quad (1)$$

We can use the appendix or use R to find our answer:

Answer: $P(Z < 0.45) = 0.6737$

3e) What is the estimated standard error of the estimator that you used in part (b)?

The estimated standard error of the estimator:

Answer: $\bar{X} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{0.3385}{\sqrt{16}} = 0.0846$

8a) In a random sample of 80 components of a certain type, 12 are found to be defective. Give a point estimate of the proportion of all such components that are not defective.

Since there are 12 that are defective from the 80 components, 68 are not defective (80-12). So the point estimate of components that are not defective is:

Answer: $\hat{p} = \frac{68}{80} = 0.85$

8b) A system is to be constructed by randomly selecting two of these components and connecting them in series, as shown here.The series connection implies that the system will function if and only if neither component is defective (i.e., both components work properly). Estimate the proportion of all such systems that work properly. [Hint: If p denotes the probability that a component works properly, how can P(system works) be expressed in terms of p?]

Since both components work properly, then P(system works)=p^2
So, $\hat{p^2} = (0.85)^2 = 0.723$

Answer: $\hat{p}^2 = (0.85)^2 = 0.723$

11a) Show that $(X_1/n_1) - (X_2/n_2)$ is an unbiased estimator for $p_1 - p_2$. [Hint: $E(X_i) = n_i p_i$ for i=1,2.]

Since we are given the hint of $E(X_i) = n_i p_i$ for i=1,2, then,

$$E(X_1) = n_1 p_1$$
$$E(X_2) = n_2 p_2$$

So,

$$E(\frac{X_1}{n_1}) - E(\frac{X_2}{n_2})$$

$$\frac{n_1 p_1}{n_1} - \frac{n_2 p_2}{n_2}$$

$$p_1 - p_2$$

Since $E(\frac{X_1}{n_1}) - E(\frac{X_2}{n_2}) = p_1 - p_2$, then $\frac{X_1}{n_1}) - \frac{X_2}{n_2}$ is an unbiased estimate of $p_1 - p_2$.

11b) What is the standard error of the estimator in part (a)?

We need to use the variance of the binomial distribution:

$$V(X_1) = \sigma_1^2 = n_1 p_1 q_1 = n_1 p_1 (1 - p_1)$$
$$V(X_2) = \sigma_2^2 = n_2 p_2 q_2 = n_2 p_2 (1 - p_2)$$

Finding the variance of $\frac{X_1}{n_1} - \frac{X_2}{n_2}$:

$V(\frac{X_1}{n_1} - \frac{X_2}{n_2}) = \frac{1}{n_1^2} V(X_1) + \frac{1}{n_2^2} V(X_2) = \frac{1}{n_1^2} n_1 p_1 (1 - p_1) + \frac{1}{n_2^2} n_2 p_2 (1 - p_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$

Answer: Standard error of the estimator:

$$(\frac{X_1}{n_1} - \frac{X_2}{n_2}) = \sqrt{V(\frac{X_1}{n_1} - \frac{X_2}{n_2})} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

11c) How would you use the observed values x_1 and x_2 to estimate the standard error of your estimator?

We can let $\hat{p_1} = \frac{x_1}{n_1}$ and $\hat{p_2} = \frac{x_2}{n_2}$, then the standard estimator is

Answer:

$$(\frac{X_1}{n_1} - \frac{X_2}{n_2}) = \sqrt{V(\frac{X_1}{n_1} - \frac{X_2}{n_2})} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{\frac{x_1}{n_1}(1 - \frac{x_1}{n_1})}{n_1} + \frac{\frac{x_2}{n_2}(1 - \frac{x_2}{n_2})}{n_2}}$$

11d) If $n_1 = n_2 = 200$, $x_1 = 127$, and $x_2 = 176$, use the estimator of part (a) to obtain an estimate of $p_1 - p_2$.

$n_1 = 200$
$n_2 = 200$
$x_1 = 127$
$x_2 = 176$

Since $(\frac{X_1}{n_1}) - (\frac{X_2}{n_2})$ is an estimator of $p_1 - p_2$, then:

$$\hat{p_1} - \hat{p_2} = (\frac{X_1}{n_1}) - (\frac{X_2}{n_2}) = \frac{127}{200} - \frac{176}{200} = -0.245$$

Answers: $\hat{p_1} - \hat{p_2} = -0.245$

11e) Use the result of part (c) and the data of part (d) to estimate the standard error of the estimator.

Using the result from c:

$(\frac{X_1}{n_1} - \frac{X_2}{n_2}) = \sqrt{V(\frac{X_1}{n_1} - \frac{X_2}{n_2})} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} = \sqrt{\frac{\frac{x_1}{n_1}(1-\frac{x_1}{n_1})}{n_1} + \frac{\frac{x_2}{n_2}(1-\frac{x_2}{n_2})}{n_2}}$

$$= (\frac{X_1}{n_1} - \frac{X_2}{n_2}) = \sqrt{\frac{\frac{127}{200}(1 - \frac{127}{200})}{200} + \frac{\frac{176}{200}(1 - \frac{176}{200})}{200}}$$

$$= 0.0411$$

Answer: Standard error=0.0411

12) Certain type of fertilizer has an expected yield per arce of $\mu_1$with variance σ^2, where as the expected yield for a second type of fertilizer is μ_2 with the same variance σ^2. Sample variances pf yields based on sample sizes n_1, n_2 of the two fertilizers: S_1^2, S_2^2. Show the estimator is unbiased of σ^2:

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

Now solving for unbiased estimator:

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

Multiply:

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

To help us understand what $E(S_1)^2 \& E(S_2)^2$, we have to use a long proof from the book (Chp. 6.1; pg 253):

$$E(S^2) = \frac{1}{n-1}[\sum X_i^2 - \frac{(\sum X_i)^2}{n}]$$

$$E(S^2) = \frac{1}{n-1}[\sum E(X_i^2) - \frac{1}{n}E(\sum X_i)^2]$$

.

.

.

$$= \sigma^2$$

With this proof, we can now start solving:

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

$$= \frac{n_1-1}{n_1+n_2-2}\sigma^2 + \frac{n_2-1}{n_1+n_2-2}\sigma^2$$

Combine:

$$= \frac{n_1+n_2-2}{n_1+n_2-2}\sigma^2$$

Cancel:

$$= \sigma^2$$

Answer: So the estimators S_1^2, S_2^2 are unbiased estimators of the variance.

22a) Use the method of moments to obtain an estimator of θ, and then compute the estimate for this data.

$$f(x;\theta) = \begin{cases} (\theta+1)x^{\theta} & 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

Doing the integral:

$$\bar{X} = \int_0^1 x(\theta+1)x^{\theta}dx = (\theta+1) * \frac{x^{\theta}+2}{\theta+2}|_0^1$$
$$= \frac{\theta+1}{\theta+2}$$

Now we can find $\hat{\theta}$:

$\bar{X} = \frac{\theta+1}{\theta+2}$
$\bar{X} = \frac{\hat{\theta}+1+(1-1)}{\hat{\theta}+2}$
$\bar{X} = \frac{\hat{\theta}+2}{\hat{\theta}+2} - \frac{1}{\hat{\theta}+2}$
$\bar{X} - 1 = -\frac{1}{\hat{\theta}+2}$
$\hat{\theta} + 2 = -\frac{1}{\bar{X}-1}$
$\hat{\theta} = -\frac{1}{1-\bar{X}} - 2$

Now we have to find $\bar{X}$:

$$\bar{X} = \frac{X_1 + X_2 + ...X_{10}}{10}$$
$$= \frac{0.92 + 0.79 + ... + 0.88}{10}$$
$$= 0.8$$

Finally, finding the estimate $\hat{\theta}$:

$\hat{\theta} = -\frac{1}{1-\bar{X}} - 2$
$= -\frac{1}{1-0.8} - 2$
Answer: $\hat{\theta} = 3$

22b) Obtain the maximum likelihood estimator of θ, and then compute the estimate for the given data.

The likelihood function for this would be:

$$\begin{aligned} f(x_1, x_2, ...x_n; \theta) &= (\theta+1)x_1^{\theta} * (\theta+1)x_2^{\theta} * (\theta+1)x_n^{\theta} \\ &= (\theta+1)^2 * (x_1, x_2, ...x_n)^{\theta} \end{aligned}$$

To get the maximum likelihood estimator we have to include the log likelihood function:

$$\ln f(x_1, x_2, ...x_n; \theta) = \ln[(\theta+1)^n * (x_1 * x_2 * ...x_n)^{\theta}]$$

$$= n * \ln(\theta+1) + \theta * \sum_{i=1}^{n} \ln x_i$$

If we took the derivative of the log likelihood function respect to $\theta$and equating it to 0 which the maximum likelihood estimator is obtained:

$$\begin{aligned} \frac{d}{d\theta} f(x_1, x_2, ...x_n; \theta) &= \frac{d}{d\theta}[2 * \ln(\theta+1) + \theta * \sum_{i=1}^{n} \ln x_i] \\ &= n * \frac{1}{(\theta+1)} + \sum_{i=1}^{n} \ln x_i \end{aligned}$$

Now solve for $\hat{\theta}$:

$$\begin{aligned} n * \frac{1}{(\hat{\theta}+1)} + \sum_{i=1}^{n} \ln x_i &= 0 \\ n * \frac{1}{(\hat{\theta}+1)} &= -\sum_{i=1}^{n} \ln x_i \\ \frac{1}{(\hat{\theta}+1)} &= \frac{-\sum_{i=1}^{n} \ln x_i}{n} \\ \hat{\theta}+1 &= \frac{n}{-\sum_{i=1}^{n} \ln x_i} \\ \hat{\theta} &= -\frac{n}{\sum_{i=1}^{n} \ln x_i} - 1 \end{aligned}$$

Now with $x_1, x_2....., x_{10}$ and using the calculator, then

$$\hat{\theta} = -\frac{10}{-2.4295} - 1$$

$$= 3.12$$

Answer: $\hat{\theta} = 3.12$

32a) Let $X_1, ...X_n$ be a random sample from a uniform distribution on $[0, \theta]$. Then the mle of θ is $\hat{\theta} = Y = \max(X_i)$. Use the fact that $Y \leq y$ iff each $X_i \leq y$ to the derivative the cdf of Y. Then show that the pdf of $Y = \max(X_i)$ is:

$$f_y(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n} & 0 \leq y \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

cdf:

$$F_Y(y) = P(Y \leq y) = P(max(X_i) \leq y$$

all the X_i are small, if max is small:

$$= P(X_1 \leq y * ....X_n \leq y)$$

they are independent:

$$= P(X_1 \leq y) * ....(X_n \leq y)$$

cdf of uniform distribution on $[0, \theta]$:

$$= (\frac{y}{0})^n, 0 \leq y \leq \theta$$

Now taking the derivative:

$$\frac{d}{df_Y} = \begin{cases} \frac{ny^{n-1}}{\theta^n} & 0 \leq y \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

32b) Use the result of part (a) to show that mle is biased but that (n + 1)max$(X_i)/n$ is unbiased.

For the estimator ($E(y) = \theta$) to be unbiased, then:

$$E(Y) = \int_0^\theta y * \frac{ny^n - 1}{\theta^n} dy$$
$$= \frac{n}{\theta^n} \frac{y^n + 1}{n+1} |_0^\theta$$
$$= \frac{n}{n+1} \theta$$

As you can see that $E(Y) \neq \theta$, and $\therefore$ the estimator is biased.

To show that (n + 1)max$(X_i)/n$ is unbiased:

Let $\tilde{Y} = \frac{n+1}{n}$, then

$$E(\tilde{Y}) = E\frac{(n+1)}{n}$$

Using result from (a):

$$= E(Y)\frac{n+1}{n}$$
$$= \theta \frac{n}{n+1} \frac{n+1}{n}$$

Cancel:

$$= \theta$$

This shows that $\tilde{Y}$ is unbiased.

36) Compute both the corresponding point estimate and s for the data of Ex 6.2.

The Data from Ex 6.2:

24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88

To find S (Standard Deviation), we need to find S^2 (Sample Variance) first:

$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

$\bar{X} = X_1 + X_2 + ...X_{20} = 555.86$
$X_i^2 = 24.46^2 + 5.61^2 + ...30.88^2 = 15489.62$

Now that we found the little parts we need to solve this problem, we can find our answer:

$$S^2 = \frac{\sum(X_i - \bar{X})^2}{n-1}$$
$$= \frac{1}{n-1}(\sum x_i^2 - \frac{1}{n}(\sum x_i)^2)$$
$$= \frac{1}{20-1}(15489.62 - \frac{1}{20}(555.86)^2)$$
$$= \frac{1}{20-1}(40.6034)$$
$$S^2 = 2.137$$

Now we can find Standard Deviation (S):

$$S = \sqrt{(S^2)}$$
$$= \sqrt{(2.137)}$$
$$= 1.462$$

So S=1.462

Now to find the estimate of standard deviation, we have to find median:

$\cancel{24.46}$	$\cancel{25.61}$	$\cancel{26.25}$	$\cancel{26.42}$	$\cancel{26.66}$	$\cancel{27.15}$	$\cancel{27.31}$	$\cancel{27.54}$	$\cancel{27.74}$	**27.94**
27.98	$\cancel{28.04}$	$\cancel{28.28}$	$\cancel{28.49}$	$\cancel{28.50}$	$\cancel{28.87}$	$\cancel{29.11}$	$\cancel{29.13}$	$\cancel{29.50}$	$\cancel{30.88}$

Median:

$$\frac{27.94 + 27.98}{2} = 27.96$$

Now we have too subtract our sample by the median and absolute the sample after that:

3.5	2.35	1.71	1.54	1.3	0.81	0.65	0.42	0.22	0.02
0.02	0.08	0.32	0.53	0.54	0.91	1.15	1.17	1.54	2.92

Now that we have this, we can now divide by .6745 (given in the problem) and put them in order:

$$3.5/.6745, 2.35/.6745......2.92/.6745$$

.

.

.

.

0.03	0.03	0.12	0.33	0.47	0.62	0.79	0.80	0.96	1.20
1.35	1.70	1.73	1.93	2.28	2.28	2.54	3.48	4.33	5.19

Now finding the median for this data:

~~0.03~~	~~0.03~~	~~0.12~~	~~0.33~~	~~0.47~~	~~0.62~~	~~0.79~~	~~0.80~~	~~0.96~~	**1.20**
1.35	~~1.70~~	~~1.73~~	~~1.93~~	~~2.28~~	~~2.28~~	~~2.54~~	~~3.48~~	~~4.33~~	~~5.19~~

Because there are two number for the median, we have to add the two values and divid it by 2:

$$\frac{1.20 + 1.35}{2} = 1.275$$

With everything being solved:

$$\hat{\sigma} = 1.275$$

Answers: $S = 1.462$ and $\hat{\sigma} = 1.275$