Stat 252.01 Winter 2006
Assignment #1 Solutions

**Printed Lecture Notes**

**(1.1)** If the population mean is $\mu$, then $\mathbb{E}(Y_i) = \mu$, $i = 1, \ldots, n$. Hence,

$$\mathbb{E}\left(\overline{Y}\right) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(Y_i) = \frac{n \cdot \mu}{n} = \mu$$

so that $\overline{Y}$ is an unbiased estimator of $\mu$. In order to show that $S^2$ is an unbiased estimator of $\sigma^2$, we begin by expanding $(Y_i - \overline{Y})^2 = Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2$. This gives

$$\begin{aligned} S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - 2\overline{Y}\sum_{i=1}^{n} Y_i + \overline{Y}^2\sum_{i=1}^{n} 1\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - 2n\overline{Y}^2 + n\overline{Y}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}^2\right) \end{aligned}$$

where we used the fact that

$$\sum_{i=1}^{n} Y_i = n\overline{Y}.$$

If the population variance is $\sigma^2$, then $\mathrm{Var}(Y_i) = \sigma^2$, $i = 1, \ldots, n$, so that $\mathbb{E}(Y_i^2) = \mathrm{Var}(Y_i) + (\mathbb{E}(Y_i))^2 = \sigma^2 + \mu^2$. Hence,

$$\begin{aligned} \mathbb{E}(S^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}(Y_i^2) - n\mathbb{E}(\overline{Y}^2)\right) &= \frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\mathbb{E}(\overline{Y}^2)\right) \\ &= \frac{1}{n-1}\left(n(\sigma^2 + \mu^2) - n\mathbb{E}(\overline{Y}^2)\right) \\ &= \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \mathbb{E}(\overline{Y}^2)\right). \end{aligned}$$

However, we still must compute $\mathbb{E}(\overline{Y}^2)$. As above, $\mathbb{E}(\overline{Y}^2) = \mathrm{Var}(\overline{Y}) + (\mathbb{E}(\overline{Y}))^2 = \mathrm{Var}(\overline{Y}) + \mu^2$ which leaves us with $\mathrm{Var}(\overline{Y})$ to compute. It is common to assume that the data were collected independently of each other; that is, if $i \neq j$, then $\mathrm{Cov}(Y_i, Y_j) = 0$. Therefore, from Theorem 5.12 (that's Stat 251 material)

$$\begin{aligned} \mathrm{Var}(\overline{Y}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{Var}(Y_i) + 2\sum\sum_{1 \le i < j \le n}\mathrm{Cov}(Y_i, Y_j)\right) &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Y_i) \\ &= \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}. \end{aligned}$$

Finally, we conclude that $\mathbb{E}(\overline{Y}^2) = \sigma^2/n + \mu^2$ so

$$\mathbb{E}(S^2) = \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \mathbb{E}(\overline{Y}^2)\right) = \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \left(\frac{\sigma^2}{n} + \mu^2\right)\right) = \frac{n}{n-1} \cdot \frac{(n-1)\sigma^2}{n} = \sigma^2$$

meaning that $S^2$ is an unbiased estimator of $\sigma^2$ as required.

**(4.3)** It is a simple matter to compute:

- $\mathbb{E}(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = 1 \cdot p + 0 \cdot (1 - p) = p$;
- $\mathbb{E}(X^2) = 1^2 \cdot P(X = 1) + 0^2 \cdot P(X = 0) = 1^2 \cdot p + 0^2 \cdot (1 - p) = p$;
- $\mathbb{E}(e^{\theta X}) = e^{\theta \cdot 1} P(X = 1) + e^{\theta \cdot 0} P(X = 0) = e^{\theta} \cdot p + 1 \cdot (1 - p) = 1 - p(1 - e^{\theta})$.

**(4.4)** In order to solve this problem, we will need to compute several integrals. Since the density function for any random variable integrates to 1, we have

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = 1.$$

After substituting $u = y^2/2$, and carefully handling the infinite limits of integrations, we find

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy = 0.$$

Finally, using parts with $u = y$, $dv = ye^{-y^2/2}dy$, and carefully handling the infinite limits of integration,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = 1.$$

In fact, it is also straightforward to show that for $n = 1, 2, 3, 4, 5, 6, \ldots$,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^n e^{-y^2/2} dy = (n-1) \cdot (n-3) \cdot (n-5) \cdots 3 \cdot 1 \cdot \left( \frac{1 + (-1)^n}{2} \right).$$

As for the expected moments, we apply the *Law of the Unconscious Statistician.*

- By definition,

  $$\mathbb{E}(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

  Substituting $y = \frac{x-\mu}{\sigma}$ so that $x = \sigma y + \mu$, $\sigma dy = dx$ transforms the integral into

  $$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-y^2/2} dy &= \sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-y^2/2} dy + \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} dy \\ &= \sigma \cdot 0 + \mu \cdot 1 = \mu \end{aligned}$$

  using the integrals above.

- By definition,

  $$\mathbb{E}(X^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

  Substituting $y = \frac{x-\mu}{\sigma}$ so that $x = \sigma y + \mu$, $\sigma dy = dx$ transforms the integral into

  $$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu)^2 e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma^2 y^2 + 2\sigma\mu y + \mu^2) e^{-y^2/2} dy.$$

  As in the previous part, splitting up the integral into the three separate pieces, and using the integrals computed above, we find

  $$\mathbb{E}(X^2) = \sigma^2 \cdot 1 + 2\sigma\mu \cdot 0 + \mu^2 \cdot 1 = \sigma^2 + \mu^2.$$

- By definition,
$$\mathbb{E}(e^{\theta X}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\theta x} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx.$$
The first step is to combine and simplify the integrand, namely
$$\begin{aligned} e^{\theta x} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \exp\left(\theta x - \frac{(x-\mu)^2}{2\sigma^2}\right) &= \exp\left(\frac{\theta^2\sigma^4 + 2\mu\theta\sigma^2 - (x-\theta\sigma^2-\mu)^2}{2\sigma^2}\right) \\ &= \exp\left(\mu\theta + \frac{\theta\sigma^2}{2}\right) \exp\left(\frac{-(x-\theta\sigma^2-\mu)^2}{2\sigma^2}\right) \end{aligned}$$
where the last equality was obtained by *completing the square*. Substituting this back into the original integral gives
$$\mathbb{E}(e^{\theta X}) = \exp\left(\mu\theta + \frac{\theta\sigma^2}{2}\right) \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-\theta\sigma^2-\mu)^2}{2\sigma^2}\right) dx.$$
To compute this final integral we make the substitution $y = \frac{x-\theta\sigma^2-\mu}{\sigma}$ so that $\sigma dy = dx$. This gives
$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-\theta\sigma^2-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2}\, dy = 1,$$
so that
$$\mathbb{E}(e^{\theta X}) = \exp\left(\mu\theta + \frac{\theta\sigma^2}{2}\right).$$

**(5.2)** If the density of $X$ is
$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2},$$
then
$$\mathbb{E}(|X|) = \int_{-\infty}^{\infty} |x|\, f(x)\, dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{|x|}{1+x^2}\, dx = \frac{1}{\pi} \int_{0}^{\infty} \frac{x}{1+x^2}\, dx$$
where the last equality follows since the integrand is even. Now, we must be extra careful with the improper integral:
$$\int_0^{\infty} \frac{x}{1+x^2}\, dx = \lim_{N\to\infty} \int_0^{N} \frac{x}{1+x^2}\, dx = \lim_{N\to\infty} \int_1^{1+N^2} \frac{1}{2u}\, du = \lim_{N\to\infty} \frac{1}{2}(\ln|1+N^2| - \ln|1|) = \infty.$$
Thus, $X \notin L^1$.

**(5.12)** If $X$ and $Y$ are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$. Hence,
$$\operatorname{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0,$$
so that $X$ and $Y$ are uncorrelated.

**(5.13)**

- We find the density of $Y$ simply using the *Law of Total Probability*:

  $$\begin{aligned}
  &P(Y=0)\\
  &\quad = P(Y=0|X=1)P(X=1) + P(Y=0|X=0)P(X=0) + P(Y=0|X=-1)P(X=-1)\\
  &\quad = 0\cdot 1/4 + 1\cdot 1/2 + 0\cdot 1/4\\
  &\quad = 1/2,
  \end{aligned}$$

  $$\begin{aligned}
  &P(Y=1)\\
  &\quad = P(Y=1|X=1)P(X=1) + P(Y=1|X=0)P(X=0) + P(Y=1|X=-1)P(X=-1)\\
  &\quad = 1\cdot 1/4 + 0\cdot 1/2 + 1\cdot 1/4\\
  &\quad = 1/2.
  \end{aligned}$$

- The joint density of $(X, Y)$ is given by

  $$\begin{aligned}
  &P(X=0, Y=0) = P(Y=0|X=0)P(X=0) = 1\cdot 1/2 = 1/2;\\
  &P(X=0, Y=1) = P(Y=1|X=0)P(X=0) = 0\cdot 1/2 = 0;\\
  &P(X=1, Y=0) = P(Y=0|X=1)P(X=1) = 0\cdot 1/4 = 0;\\
  &P(X=1, Y=1) = P(Y=1|X=1)P(X=1) = 1\cdot 1/4 = 1/4;\\
  &P(X=-1, Y=0) = P(Y=0|X=-1)P(X=-1) = 0\cdot 1/4 = 0;\\
  &P(X=-1, Y=1) = P(Y=1|X=-1)P(X=-1) = 1\cdot 1/4 = 1/4.
  \end{aligned}$$

  Since, for example, $P(X=0, Y=0) = 1/2$, but $P(X=0)P(Y=0) = 1/2\cdot 1/2 = 1/4$, we see that $X$ and $Y$ cannot be independent.

- The possible values of $XY$ are 0, 1, $-1$. Hence,

  $$P(XY=0) = P(X=0, Y=0) = 1/2$$

  and

  $$P(XY=1) = P(X=1, Y=1) = 1/4$$

  using the computations above. By the law of total probability,

  $$P(XY=-1) = 1/4.$$

  (Equivalently, $P(XY=-1) = P(X=-1, Y=1) = 1/4$.) Thus,

  $$\mathbb{E}(XY) = 0\cdot P(XY=0) + 1\cdot P(XY=1) + (-1)\cdot P(XY=-1) = 0 + 1/4 - 1/4 = 0.$$

  Since $\mathbb{E}(X) = 0$ and $\mathbb{E}(Y) = 0$, we see that

  $$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0 - 0 = 0;$$

  whence $X$ and $Y$ are uncorrelated.

**Textbook**

**(1.1(c))** Briefly: The parameter of interest is the *weekly water consumption for single-family dwelling units in the city.* The population, obviously, consists of all *single-family dwelling units in the city.* The inferential objective of the city engineer is to determine the average weekly water consumption for single-family dwelling units in the city This can be done by collecting a random sample from among all single-family dwelling units in the city, and either constructing a confidence interval or conducting a hypothesis test. As city engineer, it should be relatively straightforward to obtain a map of city water lines, and the locations of all single-family dwellings that receive city water. Note, however, that he may not have access to their names.

**(1.5(c))** Reading the histogram, we find that ehe proportion of students who had GPAs less that 2.65 is

$$\frac{3}{30} + \frac{3}{30} + \frac{3}{30} + \frac{7}{30} = \frac{16}{30}.$$

**(1.9)** By definition,

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2 \text{ where } \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i.$$

Notice that $(y_i - \overline{y})^2 = y_i^2 - 2\overline{y}y_i + \overline{y}^2$. Thus,

$$\begin{aligned}
\sum_{i=1}^{n} (y_i - \overline{y})^2 &= \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} (-2)\overline{y}y_i + \sum_{i=1}^{n} \overline{y}^2 \text{ (using c)} \\
&= \sum_{i=1}^{n} y_i^2 - 2\overline{y} \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \overline{y}^2 \text{ (using a)} \\
&= \sum_{i=1}^{n} y_i^2 - 2\overline{y} \sum_{i=1}^{n} y_i + n\overline{y}^2 \text{ (using b).}
\end{aligned}$$

But,

$$\sum_{i=1}^{n} y_i = n\overline{y}$$

so we can substitute that into the above to conclude

$$\sum_{i=1}^{n} y_i^2 - 2\overline{y} \sum_{i=1}^{n} y_i + n\overline{y}^2 = \sum_{i=1}^{n} y_i^2 - 2n\overline{y}^2 + n\overline{y}^2 = \sum_{i=1}^{n} y_i^2 - n\overline{y}^2.$$

Substituting back into this for $\overline{y}$ gives

$$s^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} y_i^2 - n\overline{y}^2 \right] = \frac{1}{n-1} \left[ \sum_{i=1}^{n} y_i^2 - n \left( \frac{1}{n} \sum_{i=1}^{n} y_i \right)^2 \right] = \frac{1}{n-1} \left[ \sum_{i=1}^{n} y_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} y_i \right)^2 \right].$$

**(1.10)** If our data consists of $\{1, 4, 2, 1, 3, 3\}$, then we trivially compute

$$\sum_{i=1}^{6} y_i = 1 + 4 + 2 + 1 + 3 + 3 = 14$$

and

$$\sum_{i=1}^{6} y_i^2 = 1^2 + 4^2 + 2^2 + 1^2 + 3^2 + 3^2 = 40.$$

Thus,

$$s^2 = \frac{1}{6-1}\left[\sum_{i=1}^{6} y_i^2 - \frac{1}{6}\left(\sum_{i=1}^{6} y_i\right)^2\right] = \frac{1}{5}\left[40 - \frac{1}{6}\cdot 14^2\right] = \frac{22}{15}.$$

*Note that writing garbage with decimals is unacceptable here!*

**(1.30)** Suppose that there is a set of $n$ measurements, namely $y_1, y_2, \ldots, y_n$. For each measurement, calculate $|y_i - \overline{y}|$ and determine whether or not $|y_i - \overline{y}| \geq ks$ for a given $k > 1$. Thus, we can write

$$\{y_1, y_2, \ldots, y_n\} = \{y_i : |y_i - \overline{y}| \geq ks\} \cup \{y_i : |y_i - \overline{y}| < ks\}.$$

Suppose that there are $n'$ of the measurements for which $|y_i - \overline{y}| \geq ks$. (Note that $0 \leq n' \leq n$.) This means that $n - n'$ of the measurements fall within $ks$ of the mean, so that the fraction of the measurements which do so is

$$\frac{n - n'}{n} = 1 - \frac{n'}{n}.$$

Our goal, therefore, is to show

$$1 - \frac{n'}{n} \geq 1 - \frac{1}{k^2}.$$

Suppose that

$$A = \{y_i : |y_i - \overline{y}| \geq ks\} \quad \text{and} \quad B = \{y_i : |y_i - \overline{y}| < ks\}.$$

Hence,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2 = \frac{1}{n-1}\left(\sum_A (y_i - \overline{y})^2 + \sum_B (y_i - \overline{y})^2\right) \geq \frac{1}{n-1}\sum_A k^2 s^2 = \frac{n'k^2s^2}{n-1}.$$

The first inequality follows since $\sum_B (y_i - \overline{y})^2 \geq 0$ and since $(y_i - \overline{y})^2 \geq k^2 s^2$ if $y_i \in A$. (Note that there are $n'$ points in $A$.) Therefore, we conclude

$$s^2 \geq \frac{n'k^2s^2}{n-1}.$$

Smplifying gives

$$1 \geq \frac{n'k^2}{n-1} \quad \text{which implies} \quad \frac{1}{k^2} \geq \frac{n'}{n-1}.$$

But, notice that $n'/(n-1) \geq n'/n$. Therefore, $1/k^2 \geq n'/n$ which implies the result.

**(1.33)** Briefly: Lead content readings must be non-negative. Since 0 is only 0.33 standard deviations below the mean, the population can only extend 0.33 standard deviations below the mean. This radically skews the distribution so that it *cannot* be normal. (If this is unclear, draw a picture.)