A Justification for the Unbiasedness of $\bar{x}$, s^2, and $\hat{p}$

During the first week of class, I led you through a little Excel exercise to illustrate the three statistics that are unbiased estimators of their parameter values. $\bar{x}$ does a great job of estimating μ, s^2 does the same for σ^2, and $\hat{p}$ the same for p. Remember what "does a great job estimating" means? It means that the average values of these statistics are equal to the population parameters that they are targeting.

Back then, I challenged you to ask, "Why?" Well, here's why, one by one.

Why $\bar{x}$ is an unbiased estimator of μ...

Back in MTH 243, you learned of a neat little idea called expectation. This was a fancy name for the average of a random variable. Well, we can apply that same idea to statistics, for statistics are really just random variables; they have to be, as their values are dependent upon the random sample from which they are calculated.

The notation for mathematical expectation was most likely not emphasized back in MTH 243. However, it's not too hard to follow. The mathematical expectation of any random variable x is written as E[x]. Back in MTH 243, we defined this quantity to be equal to $\Sigma xp(x)$, or, the sum of the products of all of the values of the individual random variables' values and their corresponding probabilities. Thus,

$$E[x] = \Sigma xp(x) = x_1p(x_1) + x_2p(x_2) + x_3p(x_3) + \ldots + x_np(x_n)$$

Here, we'll show that $E[\bar{x}] = \mu$:

$$E[\bar{x}] = E\left[\frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}\right]$$

Replace $\bar{x}$ with what $\bar{x}$ is defined to be...the sum of the data divided by how many data points there are.

$$= \frac{1}{n}\, E[x_1 + x_2 + x_3 + \ldots + x_n]$$

Since the n denominator is a constant (and not a variable), the addition of the terms can happen first, and thus we can factor out the $\frac{1}{n}$ and deal with it at the end of the computation.

$$= \frac{1}{n}\left[E[x_1] + E[x_2] + E[x_3] + \ldots + E[x_n]\right]$$

Expectations distribute...just like sums.

$$= \frac{1}{n}\,[\mu + \mu + \mu + \ldots + \mu]$$

This should make sense, since the expected value of each data point would be the average...that's what an average *is*, right? What you expect. By the way, there are n μ's in those brackets, so...

$$E[\bar{x}] = \frac{1}{n}\,[n\mu] = \mu$$

...and there we are!

Why s^2 is an unbiased estimator of σ^2...

Let's keep this ball a – rollin' and show s^2 unbiasedly estimates σ^2. For starters, you have to recall how the variance is defined: it measures average deviations from the mean of all data points. However, if you remember, variance was the square of the standard deviation (we use the standard deviation much more frequently, in an applied setting), and was written as $\boldsymbol{s}^2 = \dfrac{\Sigma(\boldsymbol{x} - \bar{\boldsymbol{x}})^2}{\boldsymbol{n} - 1}$. Here's the derivation of why you need that ($\boldsymbol{n}$ – 1) denominator instead of just $\boldsymbol{n}$:

$$E[\boldsymbol{s}^2] = E\left[\frac{\Sigma(\boldsymbol{x} - \bar{\boldsymbol{x}})^2}{\boldsymbol{n} - 1}\right]$$

Replace $\boldsymbol{s}^2$ with what $\boldsymbol{s}^2$ is defined to be, for starters.

$$= \frac{1}{\boldsymbol{n} - 1} E[\Sigma(\boldsymbol{x} - \bar{\boldsymbol{x}})^2]$$

Let's factor out that pesky denominator, like we did with the average calculation above.

$$= \frac{1}{\boldsymbol{n} - 1} E[\Sigma(\boldsymbol{x}^2 - 2\boldsymbol{x}\,\bar{\boldsymbol{x}} + \bar{\boldsymbol{x}}^2)]$$

Now, the fun starts...distribute out that beautiful squared binomial.

$$= \frac{1}{\boldsymbol{n} - 1} E[\Sigma\boldsymbol{x}^2 - \Sigma 2\boldsymbol{x}\,\bar{\boldsymbol{x}} + \Sigma\bar{\boldsymbol{x}}^2)]$$

The sum can distribute (it's all addition), and then we have this gorgeousness...

* * * * *

We're going to pause here and look at each of those three sums individually, since they can seem a little daunting, until you get them under a microscope.

1. $\Sigma\boldsymbol{x}^2$ we'll just leave, for now, as $\Sigma\boldsymbol{x}^2$
2. $\Sigma 2\boldsymbol{x}\,\bar{\boldsymbol{x}}$. This is cool. Since 2 and $\bar{\boldsymbol{x}}$ are constants, we can pull them through (factor them out) of the sum, so $\Sigma 2\boldsymbol{x}\,\bar{\boldsymbol{x}} = 2\bar{\boldsymbol{x}}\Sigma\boldsymbol{x}$. Now, remember that $\bar{\boldsymbol{x}} = \dfrac{\Sigma\boldsymbol{x}}{\boldsymbol{n}}$, by definition. That means that $\Sigma\boldsymbol{x} = \boldsymbol{n}\bar{\boldsymbol{x}}$. So, $2\bar{\boldsymbol{x}}\Sigma\boldsymbol{x} = 2\bar{\boldsymbol{x}}(\boldsymbol{n}\bar{\boldsymbol{x}})$, or $2\boldsymbol{n}\bar{\boldsymbol{x}}^2$
3. $\Sigma\bar{\boldsymbol{x}}^2$ is easier written as the square of $\bar{\boldsymbol{x}}$, added to itself $\boldsymbol{n}$ times. So, it's just $\boldsymbol{n}\bar{\boldsymbol{x}}^2$.

* * * * *

Taking all of these back into the computation, we'll have the next step:

$$= \frac{1}{\boldsymbol{n} - 1} E[\Sigma\boldsymbol{x}^2 - 2\boldsymbol{n}\,\bar{\boldsymbol{x}}^2 + \boldsymbol{n}\bar{\boldsymbol{x}}^2)]$$

$$= \frac{1}{\boldsymbol{n} - 1} E[\Sigma\boldsymbol{x}^2 - \boldsymbol{n}\,\bar{\boldsymbol{x}}^2]$$

That was cool, no? It gets better (and a little more intricate, as well).

$$= \frac{1}{\boldsymbol{n} - 1}\Big[\Sigma E[\boldsymbol{x}^2] - E[\boldsymbol{n}\,\bar{\boldsymbol{x}}^2]\Big]$$

This is the "distribution of expectation" idea, from above.

It's time for another aside here. The statistical definition of variance that you're used to seeing (or, at some point, saw) was $\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$. Using the fact that E[$X$] = μ, and that variance is in itself an expectation (it's the expected, or average, deviation, remember?), we can rewrite it as follows:

$$\sigma^2 = \frac{\Sigma(x-\mu)^2}{N} = \mathrm{E}[(x-\mu)^2$$

$$= \mathrm{E}[x^2 - 2x\mu + \mu^2]$$

$$= \mathrm{E}[x^2] - \mathrm{E}[2x\mu] + \mathrm{E}[\mu^2]$$

$$= \mathrm{E}[x^2] - 2\mu\mathrm{E}[x] + \mathrm{E}[\mu^2], \text{ so}$$

$$\sigma^2 = \mathrm{E}[x^2] - 2\mu^2 + \mu^2$$

$$= \mathrm{E}[x^2] - \mu^2$$

From this last line, we can see that $\mathrm{E}[x^2] = \mu^2 + \sigma^2$. Call this equation (1). This is good, because we can also define another quantity we need in the same way: $\mathrm{E}[\bar{x}^2] = \bar{\sigma}^2 + \bar{\mu}^2$. Then, by the central limit theorem (which, if we haven't, we will learn soon in MTH 244), we can rewrite it further as $\mathrm{E}[\bar{x}^2] = \frac{\sigma}{n} + \mu^2$. Call this equation (2). We'll now substitute (1) and (2) back into $\frac{1}{n-1}\left[\Sigma\mathrm{E}[x^2] - \mathrm{E}[n\,\bar{x}^2]\right]$ to see what we get:

$$= \frac{1}{n-1}\left[\Sigma(\mu^2 + \sigma^2) - n\left(\frac{\sigma}{n} + \mu^2\right)\right]$$

Here's the result of the substitution. Now comes the algebra. Yay!

$$= \frac{1}{n-1}\left[n\mu^2 + n\,\sigma^2 - \sigma - n\mu^2\right]$$

$$= \frac{1}{n-1}(n-1)\,\sigma^2$$

$$= \sigma^2$$

So, there you have it! That last step is precisely why you must divide by (n – 1) instead of n in the formula for the sample standard deviation. Don't you feel better now?

Why $\hat{p}$ is an unbiased estimator of p...

If you're still with me, we still have one more proof to do. This one's way easier than that last one, though. We need to prove that $E[\hat{p}] = p$. I'd like to do this one a little differently, if that's OK with you all. Since we talking about proportions, we're really talking about probabilities. The idea of a sample proportion (which we will explore thoroughly in this class) is the idea of a binomial random variable. Remember the binomial distribution?

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

If you recall, the mean of the binomial distribution is $\mu = np$. The meaning of this is helpful...we expect μ successes in the population. Therefore,

$E[\hat{p}] = E\left[\frac{x}{n}\right]$ — Replace $\hat{p}$ with what $\hat{p}$ is defined to be...the number of successes divided by the sample size.

$= \frac{1}{n} E[x]$ — Remember this move from the first proof?

$= \frac{1}{n}\mu$ — Hopefully, you agree with this. If not, reread the line starting with "The meaning of this is helpful" above!

$= \frac{1}{n} np$ — By substitution.

$E[\hat{p}] = p$ — Now...was that so bad?

Well, I feel better. You now have seen why you sit in MTH 244, happily agreeing to do inferential statistics to learn more about these 3 parameters. Don't you feel better, too?

You don't have to answer that.