Answer Sheet for Homework 1

2.6. The table shows that $\Pr(X=0, Y=0) = 0.037$, $\Pr(X=0, Y=1) = 0.622$, $\Pr(X=1, Y=0) = 0.009$, $\Pr(X=1, Y=1) = 0.332$, $\Pr(X=0) = 0.659$, $\Pr(X=1) = 0.341$, $\Pr(Y=0) = 0.046$, $\Pr(Y=1) = 0.954$.

(a) $$E(Y) = \mu_Y = 0 \times \Pr(Y=0) + 1 \times \Pr(Y=1)$$
$$= 0 \times 0.046 + 1 \times 0.954 = 0.954.$$

(b) $$\text{Unemployment Rate} = \frac{\#(\text{unemployed})}{\#(\text{labor force})}$$
$$= \Pr(Y=0) = 1 - \Pr(Y=1) = 1 - E(Y) = 1 - 0.954 = 0.046.$$

(c) Calculate the conditional probabilities first:

$$\Pr(Y=0|X=0) = \frac{\Pr(X=0, Y=0)}{\Pr(X=0)} = \frac{0.037}{0.659} = 0.056,$$
$$\Pr(Y=1|X=0) = \frac{\Pr(X=0, Y=1)}{\Pr(X=0)} = \frac{0.622}{0.659} = 0.944,$$
$$\Pr(Y=0|X=1) = \frac{\Pr(X=1, Y=0)}{\Pr(X=1)} = \frac{0.009}{0.341} = 0.026,$$
$$\Pr(Y=1|X=1) = \frac{\Pr(X=1, Y=1)}{\Pr(X=1)} = \frac{0.332}{0.341} = 0.974.$$

The conditional expectations are

$$E(Y|X=1) = 0 \times \Pr(Y=0|X=1) + 1 \times \Pr(Y=1|X=1)$$
$$= 0 \times 0.026 + 1 \times 0.974 = 0.974,$$
$$E(Y|X=0) = 0 \times \Pr(Y=0|X=0) + 1 \times \Pr(Y=1|X=0)$$
$$= 0 \times 0.056 + 1 \times 0.944 = 0.944.$$

Use the solution to part (b),

Unemployment rate for college graduates = $1 - E(Y|X=1) = 1 - 0.974 = 0.026$

Unemployment rate for non-college graduates = $1 - E(Y|X=0) = 1 - 0.944 = 0.056$

(d) The probability that a randomly selected worker who is reported being unemployed is a college graduate is

$$\Pr(X=1|Y=0) = \frac{\Pr(X=1, Y=0)}{\Pr(Y=0)} = \frac{0.009}{0.046} = 0.196.$$

The probability that this worker is a non-college graduate is

$$\Pr(X=0|Y=0)=1-\Pr(X=1|Y=0)=1-0.196=0.804.$$

(e) Educational achievement and employment status are not independent because they do not satisfy that, for all values of x and y,

$$\Pr(X=x|Y=y)=\Pr(X=x).$$

For example, from part (e) $\Pr(X=0|Y=0)=0.804$, while from the table $\Pr(X=0)=$ 0.659.

2.10. Using the fact that if $Y \quad N\left(\mu_Y, \sigma_Y^2\right)$ then $\frac{Y-\mu_Y}{\sigma_Y} \sim N(0,1)$ and Appendix Table 1, we have

(a) $\Pr(Y \le 3)=\Pr\left(\frac{Y-1}{2}\le\frac{3-1}{2}\right)=\Phi(1)=0.8413.$

(b) $\Pr(Y>0)=1-\Pr(Y\le 0)=1-\Pr\left(\frac{Y-3}{3}\le\frac{0-3}{3}\right)$
$=1-\Phi(-1)=\Phi(1)=0.8413.$

(c) $\Pr(40\le Y\le 52)=\Pr\left(\frac{40-50}{5}\le\frac{Y-50}{5}\le\frac{52-50}{5}\right)$
$=\Phi(0.4)-\Phi(-2)=\Phi(0.4)-[1-\Phi(2)]$
$=0.6554-1+0.9772=0.6326.$

(d) $\Pr(6\le Y\le 8)=\Pr\left(\frac{6-5}{\sqrt{2}}\le\frac{Y-5}{\sqrt{2}}\le\frac{8-5}{\sqrt{2}}\right)$
$=\Phi(2.1213)-\Phi(0.7071)$
$=0.9831-0.7602=0.2229.$

2.13. (a) $E(Y^2)=\mathrm{Var}(Y)+\mu_Y^2=1+0=1; E(W^2)=\mathrm{Var}(W)+\mu_W^2=100+0=100.$

(b) Y and W are symmetric around 0, thus skewness is equal to 0; because their mean is zero, this means that the third moment is zero.

(c) The kurtosis of the normal is 3, so $3 = E(Y-\mu_Y)^4/\sigma_Y^4$; solving yields

$\mathrm{E}(Y^4) = 3;$ a similar calculation yields the results for W.

(d) First, condition on $X = 0$, so that $S = W$:

$$E(S|X=0)=0; E(S^2|X=0)=100, E(S^3|X=0)=0, E(S^4|X=0)=3\times 100^2.$$

Similarly,

$$E(S|X=1)=0; E(S^2|X=1)=1, E(S^3|X=1)=0, E(S^4|X=1)=3.$$

From the law of iterated expectations

$$\begin{aligned}
E(S) &= E(S|X=0)\times \Pr(\mathrm{X}=0) + E(S|X=1)\times \Pr(X=1) = 0 \\
E(S^2) &= E(S^2|X=0)\times \Pr(\mathrm{X}=0) + E(S^2|X=1)\times \Pr(X=1) = 100\times 0.01 + 1\times 0.99 = 1.99 \\
E(S^3) &= E(S^3|X=0)\times \Pr(\mathrm{X}=0) + E(S^3|X=1)\times \Pr(X=1) = 0 \\
E(S^4) &= E(S^4|X=0)\times \Pr(\mathrm{X}=0) + E(S^4|X=1)\times \Pr(X=1) \\
&= 3\times 100^2\times 0.01 + 3\times 1\times 0.99 = 302.97
\end{aligned}$$

(e) $\mu_S = E(S) = 0$, thus $E(S-\mu_S)^3 = E(S^3) = 0$ from part (d). Thus skewness = 0.

Similarly, $\sigma_S^2 = E(S-\mu_S)^2 = E(S^2) = 1.99$, and $E(S-\mu_S)^4 = E(S^4) = 302.97.$

Thus, $\text{kurtosis} = 302.97/(1.99^2) = 76.5$

2.23. X and Z are two independently distributed standard normal random variables, so

$$\mu_X = \mu_Z = 0, \sigma_X^2 = \sigma_Z^2 = 1, \sigma_{XZ} = 0.$$

(a) Because of the independence between X and Z, $\Pr(Z=z|X=x)=\Pr(Z=z)$, and $E(Z|X)=E(Z)=0$. Thus

$$E(Y|X)=E(X^2+Z|X)=E(X^2|X)+E(Z|X)=X^2+0=X^2.$$

(b) $E(X^2)=\sigma_X^2+\mu_X^2=1$, and $\mu_Y=E(X^2+Z)=E(X^2)+\mu_Z=1+0=1.$

(c) $E(XY)=E(X^3+ZX)=E(X^3)+E(ZX)$. Using the fact that the odd moments of a standard normal random variable are all zero, we have $E(X^3)=0$. Using the independence between X and Z, we have $E(ZX)=\mu_Z\mu_X=0$. Thus

$$E(XY)=E(X^3)+E(ZX)=0.$$

(d)
$$\begin{aligned}\operatorname{cov}(XY)&=E[(X-\mu_X)(Y-\mu_Y)]=E[(X-0)(Y-1)]\\&=E(XY-X)=E(XY)-E(X)\\&=0-0=0.\end{aligned}$$

$$\operatorname{corr}(X,Y)=\frac{\sigma_{XY}}{\sigma_X\sigma_Y}=\frac{0}{\sigma_X\sigma_Y}=0.$$

2.26. (a) corr(Y_i,Y_j) = $\operatorname{cov}(Y_i,Y_j)/\sigma_{Y_i}\sigma_{Y_j}=\operatorname{cov}(Y_i,Y_j)/\sigma_Y\sigma_Y=\operatorname{cov}(Y_i,Y_j)/\sigma_Y^2=\rho$, where the first equality uses the definition of correlation, the second uses the fact that Y_i and Y_j have the same variance (and standard deviation), the third equality uses the definition of standard deviation, and the fourth uses the correlation given in the problem. Solving for cov(Y_i, Y_j) from the last equality gives the desired result.

(b)
$$\bar{Y}=\frac{1}{2}Y_1+\frac{1}{2}Y_2\text{, so that } E(\bar{Y})=\frac{1}{2}E(Y)_1+\frac{1}{2}E(Y_2)=\mu_Y$$

$$\operatorname{var}(\bar{Y})=\frac{1}{4}\operatorname{var}(Y_1)+\frac{1}{4}\operatorname{var}(Y_2)+\frac{2}{4}\operatorname{cov}(Y_1,Y_2)=\frac{\sigma_Y^2}{2}+\frac{\rho\sigma_Y^2}{2}$$

(c) $\bar{Y}=\frac{1}{n}\sum_{i=1}^{n}Y_i$, so that $E(\bar{Y})=\frac{1}{n}\sum_{i=1}^{n}E(Y_i)=\frac{1}{n}\sum_{i=1}^{n}\mu_Y=\mu_Y$

$$\begin{aligned}
\operatorname{var}(\bar{Y}) &= \operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{var}(Y_i) + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\operatorname{cov}(Y_i, Y_j) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma_Y^2 + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\rho\sigma_Y^2 \\
&= \frac{\sigma_Y^2}{n} + \frac{n(n-1)}{n^2}\rho\sigma_Y^2 \\
&= \frac{\sigma_Y^2}{n} + \left(1-\frac{1}{n}\right)\rho\sigma_Y^2
\end{aligned}$$

where the fourth line uses $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a = a(1+2+3+ \quad + n-1) = \frac{an(n-1)}{2}$ for any variable a.

(d) When n is large $\frac{\sigma_Y^2}{n} \approx 0$ and $\frac{1}{n} \approx 0$, and the result follows from (c).

3.3. Denote each voter's preference by Y. $Y = 1$ if the voter prefers the incumbent and $Y = 0$ if the voter prefers the challenger. Y is a Bernoulli random variable with probability $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = 1 - p$. From the solution to Exercise 3.2, Y has mean p and variance $p(1-p)$.

(a) $\hat{p} = \frac{215}{400} = 0.5375.$

(b) The estimated variance of $\hat{p}$ is $\widehat{var(\hat{p})} = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.5375(1-0.5375)}{400} = 6.2148\times10^{-4}$. The standard error is $\mathrm{SE}(\hat{p}) = \left(var(\hat{p})\right)^{1/2} = 0.0249$

(c) The computed t-statistic is $t^{act} = \frac{\hat{p}-\mu_{p,0}}{\mathrm{SE}(\hat{p})} = \frac{0.5375-0.5}{0.0249} = 1.506.$ Because of the large sample size $(n = 400)$, we can use Equation (3.14) in the text to get the p-value for the test $H_0 : p = 0.5$ vs. $H_1 : p \neq 0.5$:

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-1.506) = 2\times0.066 = 0.132.$$

(d) Using Equation (3.17) in the text, the p-value for the test $H_0 : p = 0.5$ vs. $H_1 : p > 0.5$ is $p\text{-value} = 1-\Phi(t^{act}) = 1-\Phi(1.506) = 1-0.934 = 0.066.$

(e) Part (c) is a two-sided test and the p-value is the area in the tails of the standard normal distribution outside ± (calculated t-statistic). Part (d) is a one-sided test and the p-value is the area under the standard normal distribution to the right of the calculated t-statistic.

(f) For the test $H_0 : p = 0.5$ vs. $H_1 : p > 0.5,$ we cannot reject the null hypothesis at the 5% significance level. The p-value 0.066 is larger than 0.05. Equivalently the calculated t-statistic 1.506 is less than the critical value 1.64 for a one-sided test with a 5% significance level. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

3.9. Denote the life of a light bulb from the new process by Y. The mean of Y is μ and the standard deviation of Y is $\sigma_Y = 200$ hours. $\bar{Y}$ is the sample mean with a sample size $n = 100$. The standard deviation of the sampling distribution of $\bar{Y}$ is $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}} = \frac{200}{\sqrt{100}} = 20$ hours. The hypothesis test is $H_0 : \mu = 2000$ vs. $H_1 : \mu > 2000$. The manager will accept the alternative hypothesis if $\bar{Y} > 2100$ hours.

(a) The size of a test is the probability of erroneously rejecting a null hypothesis when it is valid.

The size of the manager's test is

$$\begin{aligned}\text{size} &= \Pr(\bar{Y} > 2100 | \mu = 2000) = 1 - \Pr(\bar{Y} \le 2100 | \mu = 2000)\\ &= 1 - \Pr\left(\frac{\bar{Y} - 2000}{20} \le \frac{2100 - 2000}{20} | \mu = 2000\right)\\ &= 1 - \Phi(5) = 1 - 0.999999713 = 2.87 \times 10^{-7},\end{aligned}$$

where $\Pr(\bar{Y} > 2100 | \mu = 2000)$ means the probability that the sample mean is greater than 2100 hours when the new process has a mean of 2000 hours.

(b) The power of a test is the probability of correctly rejecting a null hypothesis when it is invalid. We calculate first the probability of the manager erroneously accepting the null hypothesis when it is invalid:

$$\begin{aligned}\beta &= \Pr(\bar{Y} \le 2100 | \mu = 2150) = \Pr\left(\frac{\bar{Y} - 2150}{20} \le \frac{2100 - 2150}{20} | \mu = 2150\right)\\ &= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062.\end{aligned}$$

The power of the manager's testing is $1 - \beta = 1 - 0.0062 = 0.9938.$

(c) For a test with 5%, the rejection region for the null hypothesis contains those values of the
t-statistic exceeding 1.645.

$$t^{act} = \frac{\overline{Y}^{act} - 2000}{20} > 1.645 \Rightarrow \overline{Y}^{act} > 2000 + 1.645 \times 20 = 2032.9.$$

The manager should believe the inventor's claim if the sample mean life of the new product is greater than 2032.9 hours if she wants the size of the test to be 5%.

3.11. Assume that n is an even number. Then Y is constructed by applying a weight of 1/2 to the n/2 "odd" observations and a weight of 3/2 to the remaining n/2 observations.

$$E(Y) = \frac{1}{n}\left(\frac{1}{2}E(Y_1) + \frac{3}{2}E(Y_2) + \quad \frac{1}{2}E(Y_{n-1}) + \frac{3}{2}E(Y_n)\right)$$
$$= \frac{1}{n}\left(\frac{1}{2}\cdot\frac{n}{2}\cdot\mu_Y + \frac{3}{2}\cdot\frac{n}{2}\cdot\mu_Y\right) = \mu_Y$$
$$\mathrm{var}(Y) = \frac{1}{n^2}\left(\frac{1}{4}\mathrm{var}(Y_1) + \frac{9}{4}\mathrm{var}(Y_2) + \quad \frac{1}{4}\mathrm{var}(Y_{n-1}) + \frac{9}{4}\mathrm{var}(Y_n)\right)$$
$$= \frac{1}{n^2}\left(\frac{1}{4}\cdot\frac{n}{2}\cdot\sigma_Y^2 + \frac{9}{4}\cdot\frac{n}{2}\cdot\sigma_Y^2\right) = 1.25\frac{\sigma_Y^2}{n}.$$