8. [Practice question, not graded]

	Rain (X=0)	No Rain (X=1)	Total
Long Commute (Y=0)	0.15	0.07	0.22
Short Commute (Y=1)	0.15	0.63	0.78
Total	0.30	.70	1.00

Using the random variables X and Y from Table given above, consider two new random variables W = 3 + 6X and V = 20 – 7Y. Compute:

(a) E(W) and E(V).

(b) σ^2_W and σ^2_V.

(c) $\sigma_{W,V}$ and Corr(W,V).

$X_i = \{0, 1\}$

a) $\mathbb{E}[W] = \mathbb{E}[3 + 6X]$

$= 3 + 6\,\mathbb{E}[X]$

$\mathbb{E}[X] = 0 \times 0.3 + 1 \times 0.7$

$= 0.7$

$= 3 + 6 \cdot 0.7$

$= 7.2$

$\mathbb{E}[V] = \mathbb{E}[20 - 7Y]$

$= 20 - 7\,\mathbb{E}[Y]$

$\mathbb{E}[Y] = 0 \cdot 0.22 + 1 \cdot 0.78$

$= 0.78$

$= 20 - 7 \cdot 0.78$

$= 14.54$

b) $\sigma^2_W = \mathrm{Var}(W)$

$= \mathrm{Var}(3 + 6X)$

$= \mathrm{Var}(6X)$

$= 36\,\mathrm{Var}(X)$

$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$

$= 0.3\,[0 - 0.7]^2$

$+ 0.7\,[1 - 0.7]^2$

$= 0.21$

$= 36 \times 0.21$

$= 7.56$

$\bar{Y}$ $\mathrm{Var}(\bar{Y}) =$

$\sigma^2_V = \mathrm{Var}(V)$

$= \mathrm{Var}(20 - 7Y)$

$= 49\,\mathrm{Var}(Y)$

$\mathrm{Var}(Y) = \mathbb{E}[(Y - \mathbb{E}[Y])^2]$

$= 0.22\,[(0 - 0.78)^2]$

$+ 0.78\,[(1 - 0.78)^2]$

$= 0.1716$

$= 49 \times 0.1716$

$= 8.48$

c) $\sigma_{W,V} = \mathrm{Cov}(W, V)$

$= \mathrm{Cov}(3 + 6X, 20 - 7Y)$

$= \mathrm{Cov}(6X, -7Y)$

$= -42\,\mathrm{Cov}(X, Y)$

$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$

$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

$- \mathbb{E}[X]\mathbb{E}[Y]$

$+ \mathbb{E}[X]\mathbb{E}[Y]$

$= \mathbb{E}[XY] - 2\,(0.7)(0.78)$

$+ 0.7 \cdot 0.78$

$= \mathbb{E}[XY] - 0.7 \cdot 0.78$

$\mathbb{E}[XY] = 0.15\,(0) + 0.07\,(0)(1)$

$+ 0.15\,(0)(1) + 0.63\,(1)(1)$

$= 0.63$

$= 0.63 - 0.7 \cdot 0.78$

$= 0.084$

$\sigma_{W,V} = 0.084 \times (-42)$

$= -3.528$

$\mathrm{Corr}(W,V) = \dfrac{\mathrm{Cov}(W,V)}{\sigma_W \cdot \sigma_V}$

$= \dfrac{-3.528}{\sqrt{7.56} \times \sqrt{8.48}}$

$= -0.442$

Sidenote: $\mathrm{Var}(aX + c)$

$= \mathrm{Var}(aX)$ since c has no variance

$= a^2\,\mathrm{Var}(X)$

$\mathrm{Cov}(aX + b, cY + d)$

$= \mathrm{Cov}(aX, cY)$

$= ac\,\mathrm{Cov}(X, Y)$

Can also regard variance as a special case of the covariance where you are measuring the covariance of a variable w/ itself:

$\mathrm{Cov}(aX + c, aX + c) = \mathrm{Cov}(aX, aX)$

$= a^2\,\mathrm{Cov}(X, X)$

$= a^2\,\mathrm{Var}(X) \equiv \mathrm{Var}(aX + c)$

9. [Practice question, not graded]

The following table gives the joint probability distribution between employment status and college graduation among those either employed or looking for work (unemployed) in the working age US population, based on the 1990 US Census.

	Unemployed (Y=0)	Employed (Y=1)	Total
Non-college grads (X=0)	0.045	0.709	0.754
College grads (X=1)	0.005	0.241	0.246
Total	0.050	0.950	1.000

(a) Compute E(Y).

(b) The unemployment rate is the fraction of the labor force that is unemployed. Show that the unemployment rate is given by 1-E(Y).

(c) Calculate the E(Y|X=1) and E(Y|X=0).

(d) Calculate the unemployment rate for (i) college graduates and (ii) non-college graduates.

(e) A randomly selected member of this population reports being unemployed. What is the probability that this worker is a college graduate? A non-college graduate?

(f) Are educational achievement and employment status independent? Explain.

a) $\mathbb{E}[Y] = 0.05(0) + 0.95(1)$

$= 0.95$

b) Unemployment Rate $= \dfrac{\#\text{ unemployed}}{\#\text{ labor force}}$

$= \dfrac{L \times \mathbb{P}(Y=0)}{L}$

$= \mathbb{P}(Y=0)$

$= 1 - \mathbb{P}(Y=1)$

$= 1 - 0.95$

$= 0.05$

c) $\mathbb{E}[Y] = 0 \cdot \mathbb{P}(Y=0) + 1 \cdot \mathbb{P}(Y=1)$

$\mathbb{E}[Y|X=1] = 0 \cdot \dfrac{\mathbb{P}(Y=0, X=1)}{\mathbb{P}(X=1)} + 1 \cdot \dfrac{\mathbb{P}(Y=1, X=1)}{\mathbb{P}(X=1)}$

$= 0 \cdot \dfrac{0.005}{0.246} + 1 \cdot \dfrac{0.241}{0.246}$

$= 0.9797$

$\mathbb{E}[Y|X=0] = 0 \cdot \dfrac{\mathbb{P}(Y=0, X=0)}{\mathbb{P}(X=0)} + \dfrac{1 \cdot \mathbb{P}(Y=1, X=0)}{\mathbb{P}(X=0)}$

$= 0.9403$

$Y_i = \beta_0 + \beta_1 X + e$

$\mathbb{E}[Y] = \beta_0 + \beta_1 \mathbb{E}[X]$

$\mathbb{E}[Y|X=0] = \hat{\beta}_0$

$\mathbb{E}[Y|X=1] = \hat{\beta}_0 + \hat{\beta}_1$

$\mathbb{E}[Y|X=1] - \mathbb{E}[Y|X=0] = \hat{\beta}_1$

d) Unemployment rate for college graduates

$=$ 1 − employment rate for college graduates

$= 1 - \mathbb{E}[Y|X=1]$

$= 1 - 0.9797$

$= 0.020$

Unemployment rate for non-graduates

$=$ 1 − employment rate

$= 1 - \mathbb{E}[Y|X=0]$

$= 1 - 0.9403$

$= 0.0597$

e) $\mathbb{P}(X=1 | Y=0)$

$= \dfrac{\mathbb{P}(X=1, Y=0)}{\mathbb{P}(Y=0)}$

$= \dfrac{0.005}{0.050}$

$= 0.1$

$\mathbb{P}(X=0 | Y=0)$

$= \dfrac{\mathbb{P}(X=0, Y=0)}{\mathbb{P}(Y=0)}$

$= \dfrac{0.045}{0.050}$

$= 0.9$

f) For X, Y to be independent

$\mathbb{P}(Y=y_i | X=x_i) = \mathbb{P}(Y=y_i) \quad \forall x_i, y_i$

$\mathbb{P}(X=x_i | Y=y_i) = \mathbb{P}(X=x_i)$

Counterexample:

$\mathbb{P}(Y=0 | X=0) = \dfrac{\mathbb{P}(Y=0, X=0)}{\mathbb{P}(X=0)}$

$= \dfrac{0.045}{0.754}$

$= 0.0597$

$\mathbb{P}(Y=0) = 0.05 \neq 0.0597$

10. [Practice question, not graded] SW 2.14 [Hint: Use SW Appendix Table 1.]

In a population E[Y] = 100 and Var(Y) = 43. Use the central limit theorem to answer the following questions:

(a) In a random sample of size n = 100, find Pr($\overline{Y} \leq 101$)

(b) In a random sample of size n = 165, find Pr($\overline{Y} > 98$)

(c) In a random sample of size n = 64, find Pr($101 \leq \overline{Y} \leq 103$)

Central Limit Theorem

- $(Y_1, Y_2, \ldots, Y_n) \sim$ iid

$\uparrow$ independent and identically dist.

- $0 < \sigma_Y^2 = Var(Y) < \infty$

Then $\overline{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{n})$

$\Leftrightarrow \frac{\sqrt{n}(\overline{Y} - \mu_Y)}{\sigma_Y} \sim N(0,1)$

a) $n = 100$: $\frac{\sqrt{100}\,(101 - 100)}{\sqrt{43}}$

$\Phi\left(\frac{\sqrt{100}\,(101-100)}{\sqrt{43}}\right)$

$= \Phi(1.525)$

$= 0.937 =: P(\overline{Y} \leq 101)$

b) $n = 165$

$P(\overline{Y} > 98) = 1 - P(\overline{Y} \leq 98)$

$= 1 - \Phi\left(\frac{\sqrt{165}\,(98-100)}{\sqrt{43}}\right)$

$= 1 - \Phi(-3.917)$

$= 1 - 0$

$= 1$

c) $P(101 \leq \overline{Y} \leq 103)$, $n = 64$

$\Phi\left(\sqrt{64}\,(103-100)\cdot\frac{1}{\sqrt{43}}\right) - \Phi\left(\sqrt{64}\,(101-100)\frac{1}{\sqrt{43}}\right)$

$\approx \Phi(3.660) - \Phi(1.22) \approx 0.111$

11. [Practice question, not graded] SW 3.12

To investigate possible gender discrimination in a firm, a sample of 100 men and 64 women with similar job descriptions are selected at random. A summary of the resulting monthly salaries are:

	Avg. Salary ($\bar{Y}$)	Stand Dev (of Y)	n
Men	$3100	$200	100
Women	$2900	$320	64

(a) What do these data suggest about wage differences in the firm? Do they represent statistically significant evidence that wages of men and women are different? (To answer this question, first state the null and alternative hypothesis; second, compute the relevant t-statistic; and finally, use the p-value to answer the equation.)

(b) Do these data suggest that the firm is guilty of gender discrimination in its compensation politics? Explain.

a) $H_0 : \mu_M = \mu_W \iff \mu_M - \mu_W = 0$

$H_1 : \mu_M - \mu_W \neq 0$

T-statistic : $\dfrac{\bar{Y}_M - \bar{Y}_W}{SE(\bar{Y}_M - \bar{Y}_W)}$

$W = \bar{Y}_M - \bar{Y}_W$

$Var(W) = Var(\bar{Y}_M - \bar{Y}_W)$

$= Var(\bar{Y}_M) + Var(-\bar{Y}_W)$

$= Var(\bar{Y}_M) + Var(\bar{Y}_W)$

$SE(W) = \sqrt{Var(W)}$

$= \sqrt{\dfrac{S_M^2}{100} + \dfrac{S_W^2}{64}}$

$= \sqrt{\dfrac{200^2}{100} + \dfrac{320^2}{64}}$

$= 44.72$

T-statistic $= \dfrac{3100 - 2900}{44.72}$

≈ 4.4722

$p = 2\phi(-4.4722)$

$= 7.74 \times 10^{-6}$

b) See recording, slides on measuring gender discrimination in my recitations folder